ARTICLE IN PRESS

Journal of Statistical Planning and Inference ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Distribution-free test in Tobit mean regression model

Weixing Song

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

ARTICLE INFO

Article history:
Received 22 March 2010
Received in revised form
25 January 2011
Accepted 7 March 2011

Keywords:
Tobit regression model
Khamaladze-type transformation
Brownian motion
Consistency and local power

ABSTRACT

This paper discusses the problem of fitting a parametric model in Tobit mean regression models. The proposed test is based on the supremum of the Khamaladze-type transformation of a partial sum process of calibrated residuals. The asymptotic null distribution of this transformed process is shown to be the same as that of a time-transformed standard Brownian motion. Consistency of this sequence of tests against some fixed alternatives and asymptotic power under some local nonparametric alternatives are also discussed. Simulation studies are conducted to assess the finite sample performance of the proposed test. The power comparison with some existing tests shows some superiority of the proposed test at the chosen alternatives.

Published by Elsevier B.V.

1. Introduction

The relationship between a predictor X and a response Y is often studied through regression analysis, based on a fully observed data set on these two variables. However, in some cases, Y cannot be observed unless it is above or below some known threshold. This happens in many data sets from economics, epidemiology, and biomedical studies etc. The Tobit regression model proposed by Tobin (1958) is often used to describe the data of this type. The Tobit regression model assumes that a latent variable, say Y^*, is related to a predictor X, in the following fashion:

$$Y^* = m(X) + \varepsilon, \tag{1.1}$$

where $m(x) = E(Y^*|X = x)$. But instead of observing Y^*, one can actually only observe the value $Y = \max(Y^*, y_0)$, where y_0 is a known number. By assuming $m(x)$ has a parametric form $m(x,\theta)$, the existing work on this model mainly focuses on the estimation of the unknown parameter $\theta \in \mathbb{R}^p$ for some known integer $p \geq 1$. Under the assumption of normality of ε, Amemiya (1973) and Heckman (1976, 1979) 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 (1984) based on least absolute deviations. 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 Horowitz and Neumann (1989). Therefore, it is necessary to develop some formal numeric tests to check the adequacy of the selected regression functions.

Wang (2007) 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 and $y_0 = 0$. Under the null hypothesis $m(x) = \beta_0 + \beta_1 x$, for a random sample (X_i, Y_i), $i = 1,2,\ldots,n$, $\varepsilon_i = I(Y_i \leq \max(0, \beta_0 + \beta_1 X_i)) - 1/2$, $i = 1,2,\ldots,n$, are i.i.d. Bernoulli random variables with mean 0 and variance 1/4. Replacing each ε_i with an estimated value $\hat{\varepsilon}_i = I(Y_i \leq \max(0, \hat{\beta}_0 + \hat{\beta}_1 X_i)) - 1/2$, where $\hat{\beta}_0$, $\hat{\beta}_1$ are

E-mail address: weixing@ksu.edu

0378-3758/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jspi.2011.03.012

$\sqrt{n}$-consistent estimators of β_0,β_1, respectively, Wang (2007) considered each distinct covariate as a "category" and constructed a local window around each covariate value, and created a balanced one-way table. The test statistic can be viewed as a generalization of the classical F-test statistic in the context of analysis of variance. 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. Further, it does not require any knowledge of the parametric distribution of the random error. However, a problem that was not resolved in Wang (2007) is the selection of the window width.

In this paper, we develop a lack-of-fit testing procedure for the Tobit mean regression model. That is, instead of assuming the median of ε to be 0, we assume 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. A more general null hypothesis, not limited to linear functions, is considered. The main advantages of the proposed test, compared to Wang (2007)'s test, are that it does not require the selection of window width or any other smoothing parameter, it can test any parametric regression functions rather than only linear ones as in Wang (2007), and the computation of the test statistic is very fast. The limitation is that we have to assume that the density function of the error term is known.

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

In the sequel, B denotes standard Brownian motion on [0,1], and for any random variable V, F_V, f_V denote its cumulative distribution function (CDF) and density function, respectively. For any vector a, we will use $\|a\|$ to denote its Euclidian norm.

2. Test statistic

In the following, we shall assume that ε has a known distribution with density function f_ε and CDF F_ε. Denote the conditional expectation $E(Y|X=x)$ by $g(x)$, and let

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

Then from model (1.1),

$$g(x)=y_0-[y_0-m(x)]K_0(y_0-m(x))+K_1(y_0-m(x)). \tag{2.2}$$

So one can consider a transformed regression model

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

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

$$E(\xi^2|X=x)=m^2(x)K_0(y_0-m(x))+2m(x)K_1(y_0-m(x))+K_2(y_0-m(x))+y_0^2F_\varepsilon(y_0-m(x))-g^2(x), \tag{2.4}$$

which will be denoted by $\tau^2(x)$ in the sequel. Therefore, to test $H_0: m(x)=m(x,\theta)$ for some $\theta\in\mathbb{R}^p$, we can test $H_0: g(x)=g(x,\theta)$ for some $\theta\in\mathbb{R}^p$ in the transformed model (2.3), where $g(x,\theta)$ is the same as $g(x)$ in (2.2) with $m(x)$ being replaced with $m(x,\theta)$. If $F_\varepsilon(x)$ is strictly increasing, then the two hypotheses are equivalent. In fact, let

$$h(x)=y_0-(y_0-x)K_0(y_0-x)+K_1(y_0-x),$$

then $dh(x)/dx=1-F_\varepsilon(y_0-x)>0$ for all $x\in\mathbb{R}$, which implies $h(x)$ is strictly increasing. Without loss of generality, we assume $y_0=0$ in the sequel.

There exist many lack-of-fit testing procedures for checking the adequacy of the parametric regression functions. Hart (1997)'s monograph is a very good reference on this topic. Among many existing methods for testing H_0, the one based on the cumulative marked residual process

$$T_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-g(X_i,\theta)}{\tau(X_i,\theta)} I(X_i\le x).$$

has been receiving much attention in recent years, where (X_i, Y_i), $i=1,2,\dots,n$ is a random sample from the Tobit mean regression model (1.1), $\tau(x,\theta)$ is the same as $\tau(x)$ with $m(x)$ being replaced by $m(x,\theta)$. If $\theta=\theta_0$, the value of the parameter under H_0 which is known, and one can show that $T_n(x)\Longrightarrow B\circ F_X(x)$ in $D[-\infty,\infty]$ and uniform metric. Therefore, if F_X and all parameters are known, any reasonable continuous functional of $T_n(x)$ might be used to test the hypothesis. For example, one can reject H_0 whenever $\sup_{x\in\mathbb{R}}|T_n(x)/\sqrt{F_X(x)}|$ exceeds certain critical value which can be obtained from the distribution of $\sup_{0\le t\le 1}|B(t)|$. More about the motivation of this test statistic can be found in Stute et al. (1998) (STZ).

If θ is unknown, then one can replace it with a consistent estimator, say $\hat\theta_n$. Denote the resulting process by $\hat T_n(x)$. One can show, however, that $\hat T_n(x)$ will not be distribution free. In fact, if $\hat\theta_n$ is $\sqrt{n}$-consistent, the limiting process will depend

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

on F_X, the derivative of g with respect to θ, and the conditional variance $\tau^2(x,\theta)$. Although we can show that the limiting process will be a Gaussian process with mean 0, the covariance matrix has a complicated form, which makes the testing procedure hard to implement in real applications. The same phenomena occur in the lack-of-fit test in the classical regression models or the measurement error models. See STZ or Koul and Song (2009) and the references therein.

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

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

where

$$\hat{e}_i=\frac{Y_i-g(X_i,\hat{\theta}_n)}{\tau(X_i,\hat{\theta}_n)},\quad \hat{l}(X_i)=\frac{\dot{g}(X_i,\hat{\theta}_n)}{\tau(X_i,\hat{\theta}_n)},\quad \hat{M}_x=\frac{1}{n}\sum_{k=1}^{n}\hat{l}(X_k)\hat{l}'(X_k)I(X_k\ge x),$$

and $\hat{\theta}_n$ is any $\sqrt{n}$-consistent estimator of θ_0.

To derive the asymptotic distribution of $\hat{W}_n(x)$, the following assumptions are needed, where θ_0 denotes the value of θ under H_0.

(v). $\tau^2(x,\theta_0)$ is bounded below from 0.

(m1). $m(x,\theta)$ is differentiable with respect to $\theta, Em^4(X,\theta_0)<\infty, E\|\dot{m}(X,\theta_0)\|^4<\infty, Ef_\varepsilon^2(m(X,\theta_0))<\infty$ and $Em^2(X,\theta_0)\|\dot{m}(X,\theta_0)\|^2 f_\varepsilon(m(X,\theta_0))<\infty$. For any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_0,

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

(m2). For every x, there exists a $p\times p$ square matrix $\ddot{m}(x;\theta_0)$, having finite expectation, and a nonnegative function $k(x;\theta_0)$, with $Ek(X,\theta_0)<\infty$ and satisfying the following: $\forall\delta>0, \exists\zeta>0$ such that $\|\theta-\theta_0\|\le\zeta$ implies

$$\|\dot{m}(x;\theta)-\dot{m}(x;\theta_0)-\ddot{m}(x;\theta_0)(\theta-\theta_0)\|\le\delta k(x;\theta_0)\|\theta-\theta_0\|,\quad\forall x.$$

(K). For any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ_0, and $j=0,1,2$,

$$\max_{1\le i\le n}|K_j(-m(X_i,\hat{\theta}_n))-K_j(-m(X_i,\theta_0))+\dot{K}_j(-m(X_i,\theta_0))(\hat{\theta}_n-\theta_0)'\dot{m}(X_i,\theta_0)|=o_p(1/\sqrt{n}).$$

(M). For all $x<\infty, M_x=El(X,\theta_0)l'(X,\theta_0)I(X\ge x)$ is positive definite, where $l(x,\theta_0)=\dot{g}(x,\theta_0)/\tau(x,\theta_0)$.

Condition (v) is a technical assumption ensuring the standardized residuals to be well defined. It seems restrictive but does hold in some cases. For example, if X and θ have compact supports and τ is a continuous function. Another way to circumvent this condition is to restrict the residual process to a finite interval. (m1), (m2) and (K) require certain smoothness of m and $K_j, j=0,1,2$ as functions of θ. These conditions are satisfied if either m, as a function of θ, has bounded second derivative, or the r.v. X has a compact support. Condition (M) is a technical assumption to ensure that certain matrices used in the martingale transformation are invertible.

The following theorem gives the weak convergence result of the process $\hat{W}_n(x)$.

Theorem 2.1. *Suppose (v), (m1), (m2), (K), and (M) hold. Then under H_0, for every $x_0<\infty, \hat{W}_n(x)\Rightarrow B\circ F_X(x)$, in $D([-\infty,x_0])$ and uniform metric.*

As in STZ, it is recommended to apply the above result with x_0 equal to the 99th percentile of $\hat{F}_X$. Consequently, the test that rejects H_0 whenever $\sup_{x\le x_0}|\hat{W}_n(x)/\sqrt{0.99}|>b_\alpha$ will be of the asymptotic size α, where b_α is such that $P(\sup_{0\le u\le 1}|B(u)|>b_\alpha)=\alpha$. The restriction of the weak convergence of $\hat{W}_n(x)$ over $[-\infty,x_0]$ is a technical one. See more discussion on this issue in the proof of the theorem. The choice of x_0 introduces some subjectiveness into our test. In real applications, our recommendation is to choose a large x_0 to cover the majority of the X range.

3. Consistency and local power

The ability to detect any deviation from the null hypothesis is often referred as the consistency. For a fixed alternative, a consistent test should have power that tends to 1 as the sample size goes to ∞. In this section, we show that our test is consistent. To this end, consider a class of fixed alternative hypotheses: H_a: $m(x)=m_1(x)$ such that $Em_1^2(X)<\infty$ and $m_1(x)\neq m(x,\theta)$ for any θ.

In the previous section, we assume that estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent. 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 classical regression set up, 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 the following, we simply assume that $\sqrt{n}(\hat{\theta}_n-\theta_a)=O_p(1)$ under the alternative H_a for some $\theta_a\in\mathbb{R}^p$ which may be different from θ_0. We will not justify this assumption rigorously here.

Now define

$$d_1(x)=E\left[\frac{m_1(X)-m(X,\theta_a)+|m_1(X)+\varepsilon|-|m(X,\theta_a)+\varepsilon|}{2\tau(X,\theta_a)}\right]I(X\leq x), \tag{3.1}$$

$$\rho(x)=E\left[\frac{m_1(X)-m(X,\theta_a)+|m_1(X)+\varepsilon|-|m(X,\theta_a)+\varepsilon|}{2\tau(X,\theta_a)}\right]l(X,\theta_a)I(X\leq x), \tag{3.2}$$

$$d_2(x)=El'(X)M_X^{-1}\rho(X)I(X\leq x). \tag{3.3}$$

Then we have the following result:

Theorem 3.1. *Suppose all the conditions in Theorem* 2.1 *hold with* θ_0 *replaced by* $\theta_a, d(x_0)=\sup_{x\leq x_0}|d_1(x)-d_2(x)|>0$. *Then for any* $0<\alpha<1$, *the test that rejects* H_0 *whenever* $\sup_{x\leq x_0}|\hat{W}_n(x)/\sqrt{\hat{F}_X(x_0)}|>b_\alpha$ *is consistent.*

Sometimes it is desirable to investigate the performance of a test statistic at local alternatives, since the 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 $H_{\text{Loc}}: m(x)=m(x,\theta_0)+\delta(x)/\sqrt{n}$. We assume the estimators used in the test statistic are such that $\sqrt{n}(\hat{\theta}-\theta_0)=O_p(1)$.

Define

$$d_1^L(x)=E\frac{\delta(X)}{\tau(X,\theta_0)}I(X\leq x)I(m(X,\theta_0)+\varepsilon>0), \tag{3.4}$$

$$\rho(x)=E\frac{\delta(X)l(X)}{\tau(X,\theta_0)}I(X\geq x)I(m(X,\theta_0)+\varepsilon>0), \tag{3.5}$$

$$d_2^L(x)=El'(X)M_X^{-1}\rho(X)I(X\leq x). \tag{3.6}$$

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}, $\hat{W}_n(x)\Longrightarrow B\circ F_X(x)+d_1^L(x)-d_2^L(x)$ *weakly in* $D[-\infty,x_0]$ *and in uniform metric.*

4. Simulation study

A simulation study is conducted in this section to evaluate the finite sample behavior of the proposed testing procedure. For comparison purposes, we also include the simulation results from the classical asymptotic Wald test and Wang's nonparametric test from Wang (2007). Wang's test is designed specifically for the Tobit median regression model, but it is still applicable to the cases for which the error terms follow a distribution symmetric around 0.

The data are generated from the following models which are also used in Wang (2007): $Y_i^*=m(X_i)+\varepsilon_i, Y_i=\max\{Y_i^*,0\}, i=1,2,\ldots,n$, where the regression function $m(x)$ is chosen to be

$$m(x)=\alpha+\beta x+ax^2,\quad a=0,1,2,3,$$

$$m(x)=\alpha+\beta x+b\sin^2(2\pi x),\quad b=1,2,3,$$

$$m(x)=\alpha+cx^2\sin^2(2\pi x),\quad c=2,3,4$$

and $X_i\sim\text{Uniform}(-1,1), \varepsilon_i\sim N(0,1)$, $i=1,2,\ldots,n$. The sample size is 100. Under the null hypothesis $H_0: m(x)=\alpha+\beta x$, we have $g(x)=(\alpha+\beta x)\Phi(\alpha+\beta x)+\phi(\alpha+\beta x)$, and

$$\tau^2(x)=(\alpha+\beta x)^2\Phi(\alpha+\beta x)+(\alpha+\beta x)\phi(\alpha+\beta x)+\Phi(\alpha+\beta x)-g^2(x),$$

and

$$\frac{\partial g(x)}{\partial \alpha} = \Phi(\alpha+\beta x), \quad \frac{\partial g(x)}{\partial \beta} = x\Phi(\alpha+\beta x),$$

where ϕ,Φ are the density function and CDF of standard normal random variable, respectively.

In the simulation, we choose $\alpha = 0.6, \beta = 1$. The case $a=0$ corresponds to the null hypothesis. All other models are considered as alternative models. x_0 is chosen to be the 99th percentile of the empirical distribution function $\hat{F}$. For the significance level $\alpha = 0.05$, the critical value b_α obtained from the distribution of $\sup_{0 \le u \le 1}|B(u)|$ is 2.24241. See, e.g. Khmaladze and Koul (2004). We repeated the testing procedure 1000 times and the empirical size and power are computed by using $\#\{\sup_{x \le x_0}|\hat{W}_n(x)/\sqrt{0.99}| \ge b_\alpha\}/1000$. The following table shows the simulation results, where DF Test denotes the proposed Distribution-Free testing procedure.

Compared to Wang's test (the simulation result of $k_n=9$ is included here, and k_n is the window width defined in Wang, 2007) and the Wald test, the empirical level of our test is less than the nominal level 0.05, which means that the proposed test is conservative. If the true model is $x+b\sin^2(2\pi x)$, Wang's test has the highest power. For the case of $cx^2\sin^2(2\pi x)$, our test is slightly better than Wang's test and much better than Wald test. However, when the true model is quadratic $x+ax^2$, $a=1,2,3$, the superiority of our test is clear.

$h(x)$	Model	DF Test	Wang's test	Wald test
$x+ax^2$	$a=0$	0.024	0.058	0.046
	$a=1$	0.619	0.170	0.560
	$a=2$	0.996	0.752	0.968
	$a=3$	1.000	0.986	1.000
$x+b\sin^2(2\pi x)$	$b=1$	0.060	0.166	0.068
	$b=2$	0.172	0.792	0.076
	$b=3$	0.408	0.998	0.086
$cx^2\sin^2(2\pi x)$	$c=2$	0.364	0.368	0.374
	$c=3$	0.750	0.728	0.588
	$c=4$	0.951	0.942	0.754

5. Proofs

To prove Theorem 2.1, we need two lemmas. Throughout this section, $u_p(1)$ denotes a sequence of stochastic processes that tends to zero in probability, uniformly over its time domain.

Lemma 5.1. *Suppose ξ and U are random variables with $E(\xi|U)=0, 0<E\xi^2<\infty$. Let $\sigma^2(u)=E(\xi^2|U=u), L(u)=E\sigma^2(U)I[U\le u], u\in\mathcal{I}$, a subset of $\mathbb{R}$. Let $(\xi_i,U_i), 1\le i\le n$ be i.i.d. copies of (ξ,U). Define*

$$U_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I[U_i \le u], \quad u \in \overline{R}.$$

Assume L to be continuous. Then $U_n \Longrightarrow B\circ L$ in $D(\mathcal{I})$ and uniform metric.

The proof of this lemma uses Theorems 12.6, 15.5, in Billingsley (1968). Details are similar to those appearing in STZ.

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 the matrix $C_u := E\ell(U)\ell'(U)I(U\ge u)$ is positive definite for all $u\in\mathbb{R}$. For a real valued function $\gamma\in L_2(\mathbb{R},G)$ define the transforms

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

The following lemma is from Proposition 4.1 of Khmaladze and Koul (2004) and Lemma 9.1 of Koul (2006), which in turn has origin in Khmaladze (1988).

Lemma 5.2. *Under the above set up,*

$$E\mathcal{K}_\gamma(U))\ell'(U) = 0, \quad \forall\gamma\in L_2(\mathbb{R},G), \tag{5.1}$$

$$E\mathcal{K}\gamma_1(U)\mathcal{K}_{\gamma_2}(U) = E\gamma_1(U)\gamma_2(U), \quad \forall\gamma_1,\gamma_2\in L_2(\mathbb{R},G). \tag{5.2}$$

Proof of Theorem 2.1. Denote the first term in $\hat{W}_n(x)$ by $W_{n1}(x)$, and the second by $W_{n2}(x)$. Let $\Delta_n(X_i)=\tau(X_i,\theta_0)/\tau(X_i,\hat{\theta}_n)-1$, and $\xi_i=(Y_i-g(X_i,\theta_0))/\tau(X_i,\theta_0)$. Then $W_{n1}(x)$ is the sum of four terms:

$$W_{n11}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I(X_i\le x),$$

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

$$W_{n12}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\Delta_n(X_i)I(X_i\le x),$$

$$W_{n13}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\theta_0)-g(X_i,\hat\theta_n)}{\tau(X_i,\theta_0)}I(X_i\le x),$$

$$W_{n14}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\theta_0)-g(X_i,\hat\theta_n)}{\tau(X_i,\hat\theta_n)}\Delta_n(X_i)I(X_i\le x).$$

By Lemma 5.1 and 5.2, $W_{n11}(x)\Longrightarrow B\circ F_X(x)$ uniformly in $D[-\infty,\infty]$ and uniform metric. To deal with $W_{n12}(x)$, we have to know the asymptotics of $\Delta_n(X_i)$. In fact, one can show that

$$\max_{1\le i\le n}|\tau^2(X_i,\hat\theta_n)-\tau^2(X_i,\theta_0)|=o_p(1), \tag{5.3}$$

$$\max_{1\le i\le n}\left|\frac{1}{\tau(X_i,\hat\theta_n)[\tau(X_i,\hat\theta_n)-\tau(X_i,\theta_0)]}-\frac{1}{2\tau^2(X_i,\theta_0)}\right|=o_p(1). \tag{5.4}$$

In fact, after some algebra, $\tau^2(X_i,\theta)$ can be written as

$$m^2(X_i,\theta)F(-m(X_i,\theta))+K_2(-m(X_i,\theta))-[K_1(-m(X_i,\theta))-m(X_i,\theta)F(-m(X_i,\theta))]^2.$$

Therefore,

$$\begin{aligned}\tau^2(X_i,\hat\theta_n)-\tau^2(X_i,\theta_0)&=m^2(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))-m^2(X_i,\theta_0)F(-m(X_i,\theta_0))+K_2(-m(X_i,\hat\theta_n))-K_2(-m(X_i,\theta_0))\\&\quad-[K_1(-m(X_i,\hat\theta_n))-m(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))]^2+[K_1(-m(X_i,\theta_0))-m(X_i,\theta_0)F(-m(X_i,\theta_0))]^2.\end{aligned} \tag{5.5}$$

By (m1) and (K), easy to see that

$$\max_{1\le i\le n}|m^2(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))-m^2(X_i,\theta_0)F(-m(X_i,\theta_0))|=o_p(1).$$

Similarly, one can show $\max_{1\le i\le n}|K_2(-m(X_i,\hat\theta_n))-K_2(-m(X_i,\theta_0))|=o_p(1)$ and

$$\max_{1\le i\le n}|[K_1(-m(X_i,\hat\theta_n))-m(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))]^2-[K_1(-m(X_i,\theta_0))-m(X_i,\theta_0)F(-m(X_i,\theta_0))]^2|=o_p(1).$$

Hence (5.3) holds. Using (5.3), (v), and the fact that the left hand side of (5.4) is bounded above by the sum

$$\max_{1\le i\le n}\left|\frac{\tau^2(X_i,\hat\theta_n)-\tau^2(X_i,\theta_0)}{2\tau^2(X_i,\hat\theta_n)\tau^2(X_i,\theta_0)[\tau(X_i,\hat\theta_n)+\tau(X_i,\theta_0)]}\right|,$$

and

$$\max_{1\le i\le n}\left|\frac{\tau^2(X_i,\hat\theta_n)-\tau^2(X_i,\theta_0)}{2\tau^2(X_i,\hat\theta_n)\tau^2(X_i,\theta_0)[\tau(X_i,\hat\theta_n)+\tau(X_i,\theta_0)]^2}\right|.$$

One can show (5.4) holds.

Now, let us consider $W_{n12}(x)$. Note that $W_{n12}(x)$ can be written as the sum of two terms

$$A_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[\tau^2(X_i,\theta_0)-\tau^2(X_i,\hat\theta_n)]\xi_iI(X_i\le x)}{2\tau^2(X_i,\theta_0)},$$

$$B_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{1}{\tau(X_i,\hat\theta_n)[\tau(X_i,\hat\theta_n)+\tau(X_i,\theta_0)]}-\frac{1}{2\tau^2(X_i,\theta_0)}\right)\cdot[\tau^2(X_i,\theta_0)-\tau^2(X_i,\hat\theta_n)]\xi_iI(X_i\le x).$$

From (5.5), $A_n(x)$ is the sum of three terms:

$$A_{n1}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[m^2(X_i,\theta_0)F(-m(X_i,\theta_0))-m^2(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))]\xi_iI(X_i\le x)}{2\tau^2(X_i,\theta_0)},$$

$$A_{n2}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[K_2(-m(X_i,\theta_0))-K_2(-m(X_i,\hat\theta_n))]\xi_iI(X_i\le x)}{2\tau^2(X_i,\theta_0)},$$

$$\begin{aligned}A_{n3}(x)=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[K_1(-m(X_i,\hat\theta_n))-m(X_i,\hat\theta_n)F(-m(X_i,\hat\theta_n))]^2\xi_iI(X_i\le x)}{2\tau^2(X_i,\theta_0)}\\&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[K_1(-m(X_i,\theta_0))-m(X_i,\theta_0)F(-m(X_i,\theta_0))]^2\xi_iI(X_i\le x)}{2\tau^2(X_i,\theta_0)}.\end{aligned}$$

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

All three terms are $u_p(1)$ by (m1) and (K). Hence $A_n(x)=u_p(1)$. By (5.4), we can also show that $B_n(x)=u_p(1)$, which, together with $A_n(x)=u_p(1)$, we have $W_{n12}(x)=u_p(1)$.

Adding and subtracting $m(X_i,\theta_0)$ from $m(X_i,\hat\theta_n)$, $K_0(-m(X_i,\theta_0))$ from $K_0(-m(X_i,\hat\theta_n))$, and $K_1(-m(X_i,\theta_0))$ from $K_1(-m(X_i,\hat\theta_n))$, $W_{n13}(x)$ can be written as the sum of four terms:

$$B_{n1}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[m(X_i,\hat\theta_n)-m(X_i,\theta_0)][K_0(-m(X_i,\hat\theta_n))-K_0(-m(X_i,\theta_0))]I(X_i\le x)}{\tau(X_i,\theta_0)},$$

$$B_{n2}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{m(X_i,\theta_0)[K_0(-m(X_i,\hat\theta_n))-K_0(-m(X_i,\theta_0))]I(X_i\le x)}{\tau(X_i,\theta_0)},$$

$$B_{n3}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[m(X_i,\hat\theta_n)-m(X_i,\theta_0)]K_0(-m(X_i,\theta_0))I(X_i\le x)}{\tau(X_i,\theta_0)},$$

$$B_{n4}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[K_1(-m(X_i,\hat\theta_n))-K_1(-m(X_i,\theta_0))]I(X_i\le x)}{\tau(X_i,\theta_0)}.$$

Using (m1) and (K), one can show that $B_{n1}(x)=u_p(1)$, and

$$B_{n2}(x)=\sqrt{n}(\hat\theta_n-\theta_0)'E\left[\frac{m(X,\theta_0)\dot K_0(-m(X,\theta_0))\dot m(X,\theta_0)I(X\le x)}{\tau(X,\theta_0)}\right]+u_p(1),$$

$$B_{n3}(x)=-\sqrt{n}(\hat\theta_n-\theta_0)'E\left[\frac{\dot m(X,\theta_0)K_0(-m(X,\theta_0))I(X\le x)}{\tau(X,\theta_0)}\right]+u_p(1),$$

$$B_{n4}(x)=\sqrt{n}(\hat\theta_n-\theta_0)'E\left[\frac{\dot K_1(-m(X,\theta_0))\dot m(X,\theta_0)I(X\le x)}{\tau(X,\theta_0)}\right]+u_p(1).$$

Note that

$$\dot g(x,\theta)=\dot m(x,\theta)K_0(-m(x,\theta))-m(x,\theta)\dot K_0(-m(x,\theta))\dot m(x,\theta)-\dot K_0(-m(x,\theta))\dot m(x,\theta),$$

we have

$$W_{n13}(x)=-\sqrt{n}(\hat\theta_n-\theta_0)'E\left[\frac{\dot g(X,\theta_0)I(X\le x)}{\tau(X,\theta_0)}\right]+u_p(1)=-\sqrt{n}(\hat\theta_n-\theta_0)'El(X,\theta_0)I(X\le x).$$

By (5.5), (m1), (K), and the $\sqrt{n}$-consistency of $\hat\theta_n$, one can also show that $W_{n14}(x)=o_p(1)$. Therefore, we have

$$W_{n1}(x)=W_{n11}(x)-\sqrt{n}(\hat\theta_n-\theta_0)'El(X,\theta_0)I(X\le x)+u_p(1). \tag{5.6}$$

Denote

$$\hat U_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat e_i\hat l_i I(X_i\ge x),\quad U_n(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_i l_i I(X_i\ge x).$$

Then $W_{n2}(x)$ can be written as $W_{n2}(x)=\int_{z\le x}\hat l'(z)\hat M_z^{-1}\hat U_n(z)d\hat F_X(x)$. Recall the notation $\Delta_n(X_i)=\tau(X_i,\theta_0)/\tau(X_i,\hat\theta_n)-1$, one can write $\hat U_n(x)$ as the sum of the following eight terms

$$\hat U_{n1}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\dot g(X_i,\hat\theta_n)-\dot g(X_i,\theta_0)}{\tau(X_i,\theta_0)}e_i\Delta_n(X_i)I(X_i\ge x),$$

$$\hat U_{n2}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\dot g(X_i,\theta_0)}{\tau(X_i,\theta_0)}e_i\Delta_n(X_i)I(X_i\ge x),$$

$$\hat U_{n3}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(g(X_i,\hat\theta_n)-g(X_i,\theta_0))(\dot g(X_i,\hat\theta_n)-\dot g(X_i,\theta_0))}{\tau^2(X_i,\theta_0)}\Delta_n(X_i)I(X_i\ge x),$$

$$\hat U_{n4}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(g(X_i,\hat\theta_n)-g(X_i,\theta_0))\dot g(X_i,\theta_0)}{\tau^2(X_i,\theta_0)}\Delta_n(X_i)I(X_i\ge x),$$

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

$$\hat{U}_{n5}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)}{\tau(X_i,\theta_0)}e_iI(X_i\geq x),$$

$$\hat{U}_{n6}(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\theta_0)}{\tau(X_i,\theta_0)}e_iI(X_i\geq x),$$

$$\hat{U}_{n7}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(g(X_i,\hat{\theta}_n)-g(X_i,\theta_0))(\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0))}{\tau^2(X_i,\theta_0)}I(X_i\geq x),$$

$$\hat{U}_{n8}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(g(X_i,\hat{\theta}_n)-g(X_i,\theta_0))\dot{g}(X_i,\theta_0)}{\tau^2(X_i,\theta_0)}I(X_i\geq x).$$

From condition (m1), (m2), (K), and the $\sqrt{n}$-consistency of $\hat{\theta}_n$, one can show, by a tedious but straightforward argument, that all $\hat{U}_{nj}(x)=u_p(1)$ except for $j=6,8$. $\hat{U}_{n6}$ is simply $U_n(x)$, and

$$\hat{U}_{n8}(x)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{g(X_i,\hat{\theta}_n)-g(X_i,\theta_0)-(\hat{\theta}_n-\theta_0)'\dot{g}(X_i,\theta_0)}{\tau^2(X_i,\theta_0)}\right]\dot{g}(X_i,\theta_0)I(X_i\geq x)-\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\theta_0)\dot{g}'(X_i,\theta_0)}{\tau^2(X_i,\theta_0)}I(X_i\geq x)\sqrt{n}(\hat{\theta}_n-\theta_0).$$

The first term on the right is $u_p(1)$, and

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\theta_0)\dot{g}'(X_i,\theta_0)}{\tau^2(X_i,\theta_0)}I(X_i\geq x)=M_x\sqrt{n}(\hat{\theta}_n-\theta_0)+u_p(1). \tag{5.7}$$

The law of large numbers implies the pointwise convergence in (5.7), the uniformity is obtained with the aid of a Glivenko–Cantelli type argument. Hence

$$\sup_{x\in\overline{\mathbb{R}}}\|\hat{U}_n(x)-U_n(x)+M_x\sqrt{n}(\hat{\theta}_n-\theta_0)\|=o_p(1). \tag{5.8}$$

Adding and subtracting $\dot{g}(X_i,\theta_0)$ from $\dot{g}(X_i,\hat{\theta}_n)$, $\hat{M}_x$ can be written as the sum of four terms

$$\hat{M}_{x,1}=\frac{1}{n}\sum_{i=1}^{n}\frac{[\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)][\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)]'}{\tau^2(X_i,\hat{\theta}_n)}I(X_i\geq x),$$

$$\hat{M}_{x,2}=\frac{1}{n}\sum_{i=1}^{n}\frac{[\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)]\dot{g}'(X_i,\theta_0)}{\tau^2(X_i,\hat{\theta}_n)}I(X_i\geq x),$$

$$\hat{M}_{x,3}=\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\theta_0)[\dot{g}(X_i,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)]'}{\tau^2(X_i,\hat{\theta}_n)}I(X_i\geq x),$$

$$\hat{M}_{x,4}=\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\theta_0)\dot{g}'(X_i,\theta_0)}{\tau^2(X_i,\hat{\theta}_n)}I(X_i\geq x).$$

Routine argument implies $\hat{M}_{x,j}=u_p(1)$ for$j=1,2,3$, and $\hat{M}_{x,4}=M_x+u_p(1)$. Therefore, $\sup_{x\in\overline{\mathbb{R}}}\|\hat{M}_x-M_x\|=o_p(1)$. Since $M_x>0$ for all $x\in\mathbb{R}$, so for any $x_0<\infty$, we have $\sup_{x\leq x_0}\|\hat{M}_x^{-1}-M_x^{-1}\|=o_p(1)$. Therefore, for any finite x,

$$\begin{aligned}W_{n2}(x)&=\int_{z\leq x}\frac{\dot{g}(z,\hat{\theta}_n)-\dot{g}(X_i,\theta_0)+\dot{g}(X_i,\theta_0)}{\tau(z,\theta_0)}[\hat{M}_z^{-1}-M_z^{-1}+M_z^{-1}]\cdot[\hat{U}_n(z)-U_n(z)+U_n(z)](\Delta_n(z)+1)\,d\hat{F}_X(z)\\&=\int_{z\leq x}\frac{\dot{g}'(z,\theta_0)}{\tau(z,\theta_0)}M_z^{-1}U_n(z)\,d\hat{F}_X(z)+\int_{z\leq x}\frac{\dot{g}'(z,\theta_0)}{\tau(z,\theta_0)}M_z^{-1}[\hat{U}_n(z)-U_n(z)]\,d\hat{F}_X(z)+u_p(1).\end{aligned}$$

From Lemma 6.6.4 in Koul (2002), The first term equals $\int_{z\leq x}l'(z,\theta_0)M_z^{-1}U_n(z)\,dF_X(z)+u_p(1)$. The second term, from (5.8), equals

$$-\int_{z\leq x}l'(z,\theta_0)M_z^{-1}M_z\,d\hat{F}_X(z)\sqrt{n}(\hat{\theta}_n-\theta_0)+u_p(1)=-El'(X,\theta_0)I(X\leq x)\sqrt{n}(\hat{\theta}_n-\theta_0)+u_p(1).$$

Therefore,

$$\sup_{x\leq x_0}|W_{n2}(x)-\int_{z\leq x}l'(z,\theta_0)M_z^{-1}U_n(z)dF_X(z)+El'(X,\theta_0)I(X\leq x)\sqrt{n}(\hat{\theta}_n-\theta_0)|=o_p(1),$$

which, together with (5.6), implies the desired result. □

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

Proof of Theorem 3.1. Let $Y_i^{*a}=m(X_i,\theta_a)+\varepsilon_i$ and $Y_i^a=\max\{Y_i^{*a},0\}$, then the residual $\hat{e}_i$ can be written as the sum of $\hat{e}_i^a$ and $(Y_i-Y_i^a)/\tau(X_i,\hat{\theta}_n)$, where $\hat{e}_i^a=(Y_i^a-g(X_i,\hat{\theta}_n))/\tau(X_i,\hat{\theta}_n)$. Then $\hat{W}_n(x)$ can be written as the sum $W_n^a(x)+R_n^a(x)$, where

$$W_n^a(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{e}_i^a\left[I(X_i\le x)-\int_{z\le x}\hat{l}'(z)\hat{M}_z^{-1}I(X_i\ge z)\,d\hat{F}_X(z)\hat{l}(X_i)\right],$$

$$R_n^a(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-Y_i^a}{\tau(X_i,\hat{\theta}_n)}\left[I(X_i\le x)-\int_{z\le x}\hat{l}'(z)\hat{M}_z^{-1}I(X_i\ge z)\,d\hat{F}_X(z)\hat{l}(X_i)\right].$$

From the assumption that $\sqrt{n}(\hat{\theta}_n-\theta_a)=O_p(1)$ under the alternative hypothesis, similar arguments lead to validity of (5.3) and (5.4) with θ_0 replaced with θ_a. Therefore, $W_n^a(x)\Longrightarrow B\circ F_X$ can be proved by exactly same thread as in proving $\hat{W}_n(x)\Longrightarrow B\circ F_X$ in the null case.

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

$$R_{n1}^a(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-Y_i^a}{\tau(X_i,\hat{\theta}_n)}I(X_i\le x),$$

$$R_{n2}^a(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-Y_i^a}{\tau(X_i,\hat{\theta}_n)}\int_{z\le x}\hat{l}'(z)\hat{M}_z^{-1}I(X_i\ge z)\,d\hat{F}_X(z)\hat{l}(X_i).$$

Using the elementary equality $\max\{a,0\}=(a+|a|)/2$,

$$n^{-1/2}R_{n1}^a(x)=\frac{1}{n}\sum_{i=1}^n \frac{(Y_i^*-Y_i^{*a})+(|Y_i^*|-|Y_i^{*a}|)}{2\tau(X_i,\hat{\theta}_n)}I(X_i\le x).$$

Note that

$$Y_i^*-Y_i^{*a}=m_1(X_i)-m(X_i,\theta_a),\quad |Y_i^*|-|Y_i^{*a}|=|m_1(X_i)+\varepsilon_i|-|m(X_i,\theta_a)+\varepsilon_i|,$$

one can show that $n^{-1/2}R_{n1}^a(x)=d_1(x)+u_p(1)$, where $d_1(x)$ is defined in (3.1).

Denote

$$\hat{V}_n(x)=\frac{1}{n}\sum_{i=1}^n \frac{Y_i-Y_i^a}{\tau(X_i,\hat{\theta}_n)}\hat{l}(X_i)I(X_i\ge x),$$

then $n^{-1/2}R_{n2}^a(x)=\int_{z\le x}\hat{l}(z)\hat{M}_z^{-1}\hat{V}_n(z)\,d\hat{F}_X(z)$. From the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and a Glivenko–Cantelli type argument, one can show that $\hat{V}_n(x)=\rho(x)+u_p(1)$, where $\rho(x)$ is given in (3.2). Then a routine argument leads to $n^{-1/2}R_{n2}^a(x)=d_2(x)+u_p(1)$, where $d_2(x)$ is given in (3.3). Thus

$$\sup_{x\le x_0}|n^{-1/2}R_n^a(x)-[d_1(x)-d_2(x)]|=o_p(1). \tag{5.9}$$

Finally, the consistency of our test is derived by combining (5.9), the inequality

$$\sup_{x\le x_0}|\hat{W}_n(x)|\ge\sqrt{n}\sup_{x\le x_0}|n^{-1/2}\hat{R}_n^a(x)|-\sup_{x\le x_0}|\hat{W}_n^a(x)|,$$

and the condition $d(x_0)=\sup_{x\le x_0}|d_1(x)-d_2(x)|>0$. □

Proof of Theorem 3.2. Let $Y_i^{*L}=m(X_i,\theta_0)+\varepsilon_i$, $Y_i^L=\max\{Y_i^{*L},0\}$, then we can written $\hat{e}_i$ as $\hat{e}_i^L+(Y_i-Y_i^L)/\tau(X_i,\hat{\theta}_n)$, where $\hat{e}_i^L=(Y_i^L-g(X_i,\hat{\theta}_n))/\tau(X_i,\hat{\theta}_n)$. Similar to the proof of Theorem 3.1, one can written $\hat{W}_n(x)=W_n^L(x)-R_n^L(x)$ with $\hat{e}_i^a$ replaced by $\hat{e}_i^L$, and Y_i^a by Y_i^L. One can show that $W_n^L(x)\Longrightarrow B\circ F_X(x)$. To deal with $R_n^L(x)$, we can rewrite it as $R_{n1}^L(x)-R_{n2}^L(x)$,where

$$R_{n1}^L(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-Y_i^L}{\tau(X_i,\hat{\theta}_n)}I(X_i\le x),$$

$$R_{n2}^L(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i-Y_i^L}{\tau(X_i,\hat{\theta}_n)}\int_{z\le x}\hat{l}'(z)\hat{M}_z^{-1}I(X_i\ge z)\,d\hat{F}_X(z)\hat{l}(X_i).$$

Note that

$$Y_i-Y_i^L=\frac{1}{2\sqrt{n}}[\delta(X_i)+|\sqrt{n}m(X_i,\theta_0)+\delta(X_i)+\varepsilon_i\sqrt{n}|-|\sqrt{n}m(X_i,\theta_0)+\varepsilon_i\sqrt{n}|],$$

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

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

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012

$$R^L_{n12}(x) = \frac{1}{n}\sum_{i=1}^n \frac{|\sqrt{n}m(X_i,\theta_0)+\delta(X_i)+\varepsilon_i\sqrt{n}|-|\sqrt{n}m(X_i,\theta_0)+\varepsilon_i\sqrt{n}|}{2\tau(X_i,\hat{\theta}_n)} I(X_i \le x).$$

It is easy to show that

$$R^L_{n11}(x) = E\frac{\delta(X)}{2\tau(X,\theta_0)} I(X \le x) + u_p(1). \tag{5.10}$$

One can show that

$$R^L_{n12}(x) = \frac{1}{n}\sum_{i=1}^n \frac{|\sqrt{n}m(X_i,\theta_0)+\delta(X_i)+\varepsilon_i\sqrt{n}|-|\sqrt{n}m(X_i,\theta_0)+\varepsilon_i\sqrt{n}|}{2\tau(X_i,\theta_0)} I(X_i \le x) + u_p(1).$$

For convenience, denote the first term on the right hand side as $\tilde{R}^L_{n12}(x)$. Note that for all i,

$$\big||\sqrt{n}m(X_i,\theta_0)+\delta(X_i)+\varepsilon_i\sqrt{n}|-|\sqrt{n}m(X_i,\theta_0)+\varepsilon_i\sqrt{n}|\big| \le |\delta(X_i)|. \tag{5.11}$$

It is easy to see that $|R^L_{n12}(x)-\tilde{R}^L_{n12}(x)|$ is bounded above by

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

which is $u_p(1)$ from the fact

$$\sup_{1\le i\le n}\left|\frac{1}{\tau(X_i,\hat{\theta}_n)}-\frac{1}{\tau(X_i,\theta_0)}\right| = o_p(1),$$

which can be shown similarly as in proving (5.4).

By (5.11), one can show that $E[\tilde{R}^L_{n12}(x)-E\tilde{R}^L_{n12}(x)]^2 = u(1)$. Hence $\tilde{R}^L_{n12}(x) = E\tilde{R}^L_{n12}(x)+u_p(1)$. By Lebesgue dominated convergence theorem, we can show that

$$\lim_{n\to\infty} E\frac{[|\sqrt{n}m(X,\theta_0)+\delta(X)+\varepsilon\sqrt{n}|-|\sqrt{n}m(X,\theta_0)+\varepsilon\sqrt{n}|]I(X\le x)}{2\tau(X,\theta_0)}$$
$$= E\frac{\delta(X)}{2\tau(X,\theta_0)} I(X\le x)[I(m(X,\theta_0)+\varepsilon>0)-I(m(X,\theta_0)+\varepsilon<0)].$$

This, together with (5.10), implies $R^L_{n1}(x)=d^L_1(x)+u_p(1)$, where $d^L_1(x)$ is given by (3.4).

Let

$$\hat{V}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{Y_i - Y_i^L}{\tau(X_i,\hat{\theta}_n)} \hat{l}(X_i) I(X_i \ge x).$$

Then $R^L_{n2}(x) = \int_{z\le x} \hat{l}(z)\hat{M}_z^{-1}\hat{V}_n(z)\, d\hat{F}_X(z)$. From the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and a Glivenko–Cantelli type argument, one can show that $\hat{V}_n(x) = \rho(x)+u_p(1)$, where $\rho(x)$ is given in (3.5). Then a routine argument leads to $n^{-1/2}R^L_{n2}(x)=d^L_2(x)+u_p(1)$, where $d^L_2(x)$ is given in (3.6). Therefore, we obtain

$$\hat{W}_n(x) = B\circ F_X(x) + d^L_1(x) - d^L_2(x) + u_p(1),$$

which implies the desired result in Theorem 3.2.

Acknowledgments

The author gratefully acknowledges the editor and two referees for their helpful comments which improved the presentation of the paper.

References

Amemiya, T., 1973. Regression analysis when the dependent variable is truncated normal. Econometrica 41, 997–1016.
Billingsley, P., 1968. Convergence of Probability Measure. Wiley.
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.
Horowitz, J.L., Neumann, G.R., 1989. Specification testing in censored regression models: parametric and semiparametric methods. J. Appl. Econom. 4, 61–86.
Jennrich, R.J., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Khamaladze, 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, second ed. Springer.
Koul, H.L., 2006. Model diagnostics via martingale transforms: a brief review. In: Frontiers in Statistics. Imperial College Press, London, pp. 183–206.
Koul, H.L., Song, W., 2009. Minimum distance regression model checking with Berkson error. Ann. Statist. 37, 132–156.
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.
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. Econometrics 50, 1–25.

Please cite this article as: Song, W., Distribution-free test in Tobit mean regression model. J. Statist. Plann. Inference (2011), doi:10.1016/j.jspi.2011.03.012