# MODEL CHECKING IN TOBIT REGRESSION MODEL VIA NONPARAMETRIC SMOOTHING

by

SHAN LIU

B.S., Central University of Finance and Economics, Beijing, China, 2009

---

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2012

Approved by:

Major Professor
Weixing Song

# Abstract

A nonparametric lack-of-fit test is proposed to check the adequacy of the presumed parametric form for the regression function in Tobit regression models by applying Zheng's device with weighted residuals. It is shown that testing the null hypothesis for the standard Tobit regression models is equivalent to test a new null hypothesis of the classic regression models. An optimal weight function is identified to maximize the local power of the test. The test statistic proposed is shown to be asymptotically normal under null hypothesis, consistent against some fixed alternatives, and has nontrivial power for some local nonparametric power for some local nonparametric alternatives. The finite sample performance of the proposed test is assessed by Monte-Carlo simulations. An empirical study is conducted based on the data of University of Michigan Panel Study of Income Dynamics for the year 1975.

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

First and for most, I am heartily thankful to my major professor, Dr. Weixing Song, whose encouragement, guidance and support from the initial to the final stage enabled my understanding of the subject.

I'd like to thank Dr. Paul Nelson and Dr. Weixin Yao as being my committee members.

My gratefulness extends to all the people who supported me in any respect during the completion of the report.

# Dedication

To all the people by my side, through thick and thin;

To all the people in my way, opponent and foe;

To all the fortune I am so blessed to have as the source of joy and wisdom;

To all the mishap I am so reluctant to accept as the grindstone of my will and my persistence.

Also to anyone who finds themselves at a place in life where the question of why seems unanswerable, you are not alone.

# Chapter 1

# Introduction

## 1.1 The Tobit Regression Model

Household expenditures on various categories of goods vary with household income. The expenditures for many categories are zero when the income levels are low. Most households would report no expenditures on automobiles or major household durable goods during any given year. Among those households who made any such expenditure, there would be wide variability in amount. The behavior of consumption is reflected in the Engel curve, where a straight line can't represent the expenditure on durable goods for both low and high incomes. In an effort to quantify the behavior of consumption as described before, while get around the inefficiency of Probit analysis caused by throwing away available information on the value of the dependent variable, Professor James Tobin in 1958 proposed the model of limited dependent variables, the name of which was later known as Tobit model, coined by Goldberger in 1964 inspired by the difference with and similarity to Probit model.

Amemiya wrote a survey in 1984 about the Tobit regression model, in which an elementary utility maximization model was developed to illuminate the phenomenon in question.

To be specific, let

$$\begin{aligned} Y &= \text{a household's expenditure on a durable good,} \\ y_0 &= \text{the price of the cheapest available durable good,} \\ Z &= \text{all the other expenditure,} \\ X &= \text{income.} \end{aligned}$$

A household was assumed to maximize its utility $U(Y, Z)$ under the constraint of the budget $Y + Z \leq X$ and the boundary constraint $y_0 \leq Y$. Suppose $Y^*$ is the solution of the maximization of U subject to $Y + Z \leq X$, but ignoring the other constraint, and assume $Y^* = m(x) + \varepsilon$, where $\varepsilon$ may be interpreted as the collection of all the unobservable variables which affect the utility function. Then the solution Y to the original problem was depicted as following,

$$Y = \begin{cases} Y^* & \text{if } Y^* > y_0, \\ y_0 & \text{if } Y^* \leq y_0. \end{cases}$$

That is, one can actually observe $Y = Y^* \cdot I(Y^* > y_0) + y_0 I(Y^* \leq y_0)$.

We choose $y_0 = 0$ for this report without loss of generality. Then Toibt regression model takes the look of the following,

$$\begin{aligned} Y^* &= m(X) + \varepsilon, \\ Y &= \begin{cases} Y^* & \text{if } Y^* > 0, \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Note that $\{(X, Y)\}$ are observable, while $Y^*$'s are not if $Y^* < 0$. In the following, we denote $\{(x_i, y_i)\}_{i=1}^n$ as the sample from the Tobit regression model, $\varepsilon_i$'s are assumed to be i.i.d with mean 0 and finite variance $\sigma^2$.

By assuming that the regression function $m(X)$ is linear, with the form $m(X) = \alpha + \beta\prime X$ the existing work on the standard Tobit regression model mainly focuses on the estimation of the unknown regression parameters $(\alpha, \beta')'$ and the error variance $\sigma^2$. Under the normality assumption of the error term $\varepsilon$, Amemiya (1973) and Heckman (1976,1979) proposed consistent estimators for $\theta = (\alpha, \beta', \sigma^2)'$, but these estimators are not consistent if the normality assumption is violated. Powell (1984) extended least absolute deviations (LAD) estimation to the Tobit regression model with non-negativity dependent variable and gave conditions under which this estimator is consistent and asymptotically normal. The consistent estimator does not depend upon the functional form of the distribution of the residuals. Nonparametric estimation were tried for this important regression model by Lewbel and Linton (2002), and Zhou (2007), but the hard-to-interpret nature of nonparametric procedure makes the parametric modeling the first choice for practitioners.

Originally designed for the investigating of the relationship between the household expenditures on durable goods and the income, Tobit regression model is now a frequently used tool for modeling censored and truncated variables in a wide range of fields such as econometrics, biometrics, agricultural and engineering. Many empirical examples can be found in Amemiya (1984), Blaylock and Blisard (1993), McConnel and Zetzman (1993), Lichtenberg and Shapiro (1997), Adesina and Zinnah (1993), Ekstrand and Carpenter (1998), Anastasopoulos, Tarko and Mannering (2007) and the references therein. One of the merits of McConnel and Zetzman (1997)'s study of the differences between urban and rural elderly person in the use of hospital, nursing home, and physician services, was the application of Tobit regression model to address the fact that a substantial proportion of individuals were not likely to use the service in question over the designated study period. Nursing home admission in the data set are truncated at six admissions and use of physician services at 25 visits. The analysis revealed that the utilization pattern of hospital, nursing home, and physician services was unrelated to either rural or urban residential location or the availability of health resources in those areas. With the domain knowledge in Chemistry that $NO_3N$'s detectable concentration is 0.1 mg/L and above, echoed by the fact that 57% of community water system (CWS) test showed no detectable $NO_3N$, Lichtenberg and Shapiro (1997) used Tobit regression model in their study of the relationship between land use and $NO_3N$ concentrations in drinking water wells. Although it is the 90's that witnessed the wide application of Tobit regression model, its appeal doesn't fade with the elapse of time.

Tobit regression model has its unique advantage in dealing with inconsistent parameter estimates caused by the inappropriate use of standard ordinary least squares, and is being paid with more and more attention. Tarko and Mannering (2007) introduced Tobit regression model to the transportation arena to understand the determinant factors about the frequency of accidents on roadway segments over some period of time. Since it is likely that many highway segments will have no accidents reported during the analysis period, modeling accident rates with standard ordinary least squares would result in inconsistent parameter estimates. The left-censored accidents rate data on Indiana interstates with a clustering at zero (zero accidents per 100 million vehicle miles traveled) led Tarko and Mannering to the conclusion that many factors relating to pavement condition, roadway geometrics and traffic characteristics have an influence on vehicle accident rate.

For all the literatures I cited above, the regression function m(x) is assumed to be linear.

## 1.2 The Research Objective

The prerequisite of an insightful application of Tobit regression model is the chosen of a suitable parametric form of $m(X)$, the importance of which can never be overemphasized. The outcome of a misidentified regression function is often, if not every time, a misleading conclusion. For instance, Horowitz and Neumann (1989) showed that violation of the linearity assumption can produce inconsistent estimators of the parameters and biased prediction of the survival time in censored regression models. However, so far, the selection of the parametric form of $m(X)$ is still quite judgmental, as some models are chosen for the sake of mathematical convenience, and experience-oriented, as some are chosen based on empirical evidence. Hence, both theoretically and practically, it is necessary to develop formal tests to check the adequacy of the selected regression function.

The tests in Tobit regression model can be roughly divided into two categories. The first one is goodness-of-fit test, focusing on the verifying of the distribution of the error terms, considering the fact that violation of the normality assumption can cause serious estimation problems. Examples in this category include Nelson (1981), Olsen (1980), Lee (1981), and Lin and Schmidt (1984), among others.

The second category is the lack-of-fit test, focusing on checking the adequacy of the pre-specified parametric form for the regression fucntion. Surprisingly, researches in this category are not as thriving as the previous one. What's more, a common property shared by the tests in this category is that all the tests proposed have both their merits and limitations. In order to provide the reader with a full appreciation of the lack-of-fit test available now, I give a brief introduction of every test, along with an elaboration of each test's advantages and disadvantages. 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. By considering each distinct covariate as a category and constructing a local window $W_i$, encompassing the $k_n$ nearest covariate values, around each $x_i$, Wang created an artificial balanced one-way table with $n$ categories, where the responses in the $i$th category are the $\hat{\varepsilon}_j$'s associated with the covariate values belonging to $W_i$, which can be expressed as $\hat{\varepsilon}_i = I(Y_i \leq \max(0, \hat{\beta}_0 + \hat{\beta}_1 X_i)) - 1/2$. The test statistic can be viewed as a generalization of the classical F-test statistic in the context of analysis of variance. Wang (2007)'s test has the advantage of allowing the alternative to be any smooth function and no knowledge about the parametric distribution of the random error is required. However, our simulation study indicates that Wang (2007)'s test is sensitive to the choices of the smoothing parameters. Song (2011) developed a lack-of-fit test procedure for a more general null hypothesis based on the Khamaladze type transformation of a certain marked residual process, assuming that the mean of $\varepsilon$ is 0. Song (2011)'s test circumvents the problem of window width selection or the selection of any other smoothing parameter. Song (2011)'s test is applicable not only to linear regression functions, like Wang (2007)'s, but any parametric regression functions. However, a major limitation pertains to Song (2011)'s test is that the predictor variable X must be one-dimensional. Following a few of the significant works such as Härdle and Mammen (1993), Koul and Ni (2004) in the classic regression models, Song and Zhang (2011) developed a lack-of-fit test 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. Although Song and Zhang (2011)'s test units the beauty of not being limited to linear regression function and being applicable to multidimensional predictors, the computation of this test is relatively tedious.

It is quite obvious that a test procedure with both the computational convenience and the flexibility of not being limited to linear regression function is in need. In an attempt to develop such a test, we seek inspiration from Zheng (1996)'s consistent test of functional form of nonlinear regression models via nonparametric estimation.

As early as 1996, Zheng realized the importance of lack-of-fit test of function form in econometrics and proposed such a test facilitated by nonparametric estimation techniques. For the sake of completeness, the essence of Zheng (1996)'s consistent test is presented here.

It is held that for observations $\{(X, Y)\}$, where $X$ is a $m \times 1$ vector and $Y$ is a scalar, if $E[|Y|] < \infty$, there exists a Borel measurable function $g$ such that $E(Y|X = x) = g(x)$ where $x \in R^m$. When it comes to a parametric regression model, $g(x)$ is assumed to belong to a parametric family of known real functions $f(x, \theta)$ on $R^m \times \Theta$ where $\Theta \subset R^l$. To justify the use of a parametric model, a lack-of-fit test is needed. The null hypothesis to be tested is that the parametric model is correct:

$$H_0 : Pr[E(Y|X) = f(X, \theta_0)] = 1 \text{ for some } \theta_0 \in \Theta, \tag{1.1}$$

while, without a specific form, the alternative hypothesis, encompassing all possible departures from the null model, is that the null hypothesis is false:

$$H_a : Pr[E(Y|X) = f(X, \theta)] < 1 \text{ for all } \theta \in \Theta,$$

where $\theta_0$ is defined as $\theta_0 = \arg\min_{\theta\in\Theta} E[Y - f(X, \theta)]^2$.

Denote $\varepsilon_i \equiv Y - f(X, \theta_0)$ and let $p(\cdot)$ be the density function of $X$. Under the null hypothesis, since $E(\varepsilon_i|X) = 0$, we have $E[\varepsilon_i E(\varepsilon_i|X)p(X)] = 0$.

Rewrite $E[\varepsilon_i E(\varepsilon_i|X)p(X)]$ as $E[E(\varepsilon_i|X)^2 p(X)]$. Under the alternative hypothesis, since $E(\varepsilon_i|X) = g(X) - f(X,\theta_0)$, we have $E[\varepsilon_i E(\varepsilon_i|X)p(X)] = E[E(\varepsilon_i|X)^2 p(X)] = E\{[g(X) - f(X,\theta_0)]^2 p(X)\} > 0$.

The unknown functions $g$ and $p$ can be estimated by various nonparametric methods. Here Zheng (1996) used analytically simpler kernel regression and density methods to estimate $g$ and $p$. A kernel estimator of the regression function $E(\varepsilon_i|X)$ can be written in the form

$$\hat{E}(\varepsilon_i|X) = \frac{1}{(n-1)\hat{p}(X)} \sum_{j=1,j\neq i}^{n} \frac{1}{h^m} K\left(\frac{X_i - X_j}{h}\right) \varepsilon_j,$$

where $\hat{p}$ is a kernel estimator of the density function of $p$, with

$$\hat{p}(x_i) = \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} \frac{1}{h^m} K\left(\frac{X_i - X_j}{h}\right),$$

where K is a kernel function, $h$, depending on sample size n, is a bandwidth parameter.

Let $e_i = Y_i - f(X_i, \hat{\theta})$, where $\hat{\theta}$ is any $\sqrt{n}$-consistent estimator of $\theta_0$. Zheng (1996) used the the sample analogues of $E[\varepsilon_i E(\varepsilon_i|X_i)p(X_i)]$ to form a test statistics:

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{i=1}^{n} [e_i \hat{E}(\varepsilon_i|X_i)\hat{p}(X_i)] \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \frac{1}{h^m} K\left(\frac{X_i - X_j}{h}\right) e_i e_j. \end{aligned}$$

Under the null hypothesis (1.1), Zheng proved that $nh^{m/2} V_n \xrightarrow{d} N(0, \Sigma)$, where $\Sigma$ is the asymptotic variance of $nh^{m/2} V_n$, which can be consistently estimated by $\hat{\Sigma}$ as

$$\hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \frac{1}{h^m} K^2\left(\frac{X_i - X_j}{h}\right) e_i^2 e_j^2.$$

Finally, Zheng (1996) defined the standardized version of the test statistic $T_n$ as

$$T_n = \sqrt{\frac{n-1}{n}} \cdot \frac{nh^{m/2} V_n}{\sqrt{\hat{\Sigma}}} = \frac{\sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K(\frac{X_i - X_j}{h}) e_i e_j}{[2\sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} K^2(\frac{X_i - X_j}{h}) e_i^2 e_j^2]^{1/2}}$$

The asymptotic null distribution of $V_n$, along with its power against fixed alternatives and local alternatives, can be derived based on the central limit theorem for degenerated U-statistics developed by Hall (1984).

So far, two fundamental components of this report, Tobit regression model and Zheng (1996)'s consistent test of function form, are presented. With the inspiration of Zheng (1996)'s test, we propose a new lack-of-fit test for Tobit regression model, with a simpler form of test statistic centered asymptotically at 0.

Throughout this paper, we will use $f_v$, $F_v$ to denote the density and the cumulative distribution function (CDF) of a random variable $v$, use $\Rightarrow$ to denote the convergence in distribution.

# Chapter 2

# Model Checking in Tobit Regression Model via Nonparametric Smoothing

In the previous chapter, we elaborate the main idea of the two fundamental components of this report, Tobit regression model and Zheng (1996)'s consistent test of function form. However, under Tobit regression model, $Y^*$'s are not always observable. Certain adjustments are needed to accommodate Tobit regression model to the setup of Zheng (1996)'s consistent test. An optimal weight function is defined to maximize the local power of the test.

Test statistics, assumptions and main results are presented in this chapter. The later part of this chapter is devoted to the proofs of the main results.

## 2.1 Test Statistics and Assumptions

With a brief review of Tobit regression model,

$$\begin{aligned} Y^* &= m(X) + \varepsilon, \\ Y &= \begin{cases} Y^* & \text{if } Y^* > 0, \\ 0 & \text{otherwise,} \end{cases} \end{aligned}$$

we realize that as the solution $Y^*$'s for the utility function can only be observed when $Y^* > 0$, it is not feasible to apply Zheng (1996)'s method directly to test if $m(X) = m(X, \theta)$ holds in the Tobit regression model. Therefore, we have to build a new regression model which certain dependence between the observable quantity and the fitted regression function $m(X)$ is reflected. A natural way of finding such a dependence is to consider the conditional expectation $E(Y|X=x)$. Bear in mind that $E(Y|X=x) = E[(m(X)+\varepsilon) \cdot I(Y^* > 0)|X=x] = \int_{-m(x)}^{\infty}(m(x)+u)f_\varepsilon(u)du = \int_{-m(x)}^{\infty} m(x)f_\varepsilon(u)du + \int_{-m(x)}^{\infty} uf_\varepsilon(u)du = m(x)\int_{-m(x)}^{\infty} f_\varepsilon(u)du + \int_{-m(x)}^{\infty} uf_\varepsilon(u)du$. Let $Q_j(x) = \int_x^\infty u^j f_\varepsilon(u)du$, $j = 0, 1$, then $E(Y|X = x) = m(x)Q_0(-m(x)) + Q_1(-m(x))$.

Considering the following regression model

$$Y = m(X)Q_0(-m(X)) + Q_1(-m(X)) + \xi = g(X) + \xi, \tag{2.1}$$

It is easily seen that $\xi$ and $g(X)$ are uncorrelated.

Throughout this paper, we shall assume that the density function $f_\varepsilon$ is known for the sake of simplicity, readability and model identifiability, but $f_\varepsilon$ doesn't have to be normally distributed. A more realistic assumption should be that $f_\varepsilon$ has a known form with mean 0 and unknown variance $\sigma_\varepsilon^2$. In this case, $Q_0$ and $Q_1$ are also functions of $\sigma_\varepsilon^2$. It can be shown that even if more regularity conditions are imposed to the model, the test procedures proposed in this report are still applicable. An attractive feature of (2.1) is that, as a function of $m(x)$, g is strictly monotone provided that $F_\varepsilon$ is strictly increasing, the prove of which will be shown in the third section of this chapter. Therefore, to test $H_0 : m(x) = m(x, \theta)$, it is equivalent to test

$$H_0 : g(x) = g(x, \theta) \text{ for some } \theta \in \Theta, \text{ versus } H_1 : H_0 \text{ is not true} \tag{2.2}$$

for (2.1), where $g(x, \theta)$ is the same as $g(x)$ with $m(x)$ replaced by $m(x, \theta)$.

Note that $E(I(Y = 0)|X = x) = Pr(\varepsilon < -m(x)) = \int_{-\infty}^{-m(x)} f_\varepsilon(u)du = 1-\int_{-m(x)}^{\infty} f_\varepsilon(u)du = 1 - Q_0(-m(x))$. As a function of $m(x)$, $I(Y = 0) = 1 - Q_0(-m(X)) + \eta$ is also strictly monotone provided that $F_\varepsilon$ is strictly increasing. One might consider the use of the regression model $I(Y = 0) = 1 - Q_0(-m(X)) + \eta$. Since only the truncated information of $Y^*$, instead of the values for the whole dataset, are used to form the test, the corresponding test procedure may not be as powerful as the one based on (2.1), an inspect confirmed by the simulation studies.

Throughout this paper, we will only focus on the model (2.1), but a brief discussion about the performance of the test based on the model $E(I(Y=0)|X=x) = 1 - Q_0(-m(x))$ will be mentioned in Chapter 3.

Let $\hat{\theta}$ be any $\sqrt{n}$-consistent estimate of $\theta_0$, the true value of $\theta$ under the null hypothesis, $K$ be a symmetric density function and $h$ be a sequence of positive numbers depending on the sample size n, $d$ be the dimension of the predictor variable, $w(x)$ be an positive measurable function and $\hat{\xi}_i = Y_i - g(X_i, \hat{\theta})$. Instead of only using the residual $\hat{\xi}_i$ as Zheng did in his paper in 1996, we use a weighted residual $\hat{\xi}_i w(X_i)$ to construct the test statistic, hoping to increase the power of the test by choosing an optimal weight function. Applying Zheng (1996)'s idea, we can use the following statistic to construct a test for the hypotheses in (2.2), hence the hypotheses $H_0 : m(x) = m(x, \theta)$ versus $H_a : H_0$ is not true:

$$V_n = \frac{1}{n(n-1)h^d} \sum_{i \neq j} K\left(\frac{X_i - X_j}{h}\right) \hat{\xi}_i \hat{\xi}_j w(X_i) w(X_j). \tag{2.3}$$

We shall show that $nh^{d/2}V_n$ is asymptotically normal with mean 0 and variance

$$\sigma^2 = 2 \int K^2(u)du \int (\tau^2(x))^2 f_X^2(x) w^4(x) dx, \tag{2.4}$$

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

$$\hat{\sigma}^2 = \frac{2}{n(n-1)} \sum_{i \neq j}^{n} \frac{1}{h^d} K^2\left(\frac{X_i - X_j}{h}\right) \hat{\xi}_i^2 \hat{\xi}_j^2 w^2(X_i) w^2(X_j). \tag{2.5}$$

Hence, the test statistic for testing the hypotheses (2.2) is $T_n = nh^{d/2}V_n/\hat{\sigma}$.

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

(C1). The random error $\varepsilon$ has a bounded density function, $E(\varepsilon) = 0$, and $E(\varepsilon^4) < \infty$; $\varepsilon$ and $X$ are independent.

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

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

(C4). $m(x,\theta)$ is continuously differentiable with respect to $\theta$, and the derivative $\dot{m}(x,\theta)$ satisfies $E \parallel \dot{m}(X;\theta) \parallel^4 < \infty$; for any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of $\theta_0$, the true value of $\theta$ under the null hypotheses,

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

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

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

(C7). $w(x)$ is continuous and $E[(\tau^2(X))^4 + \parallel \dot{m}(X,\theta_0) \parallel^4] w^8(x) < \infty$.

Conditions (C2) and (C3) are the same as the Assumption 1 in Zheng (1996), and are very typical in nonparametric smoothing literature. Condition (C4) plays a similar role as the Assumption 2 in Zheng (1996) to guarantee the negligibility of the higher order term in some Taylor expansions used when showing the asymptotics of test statistics. The kernel function in Condition (C5) and the bandwidth $h$ in Condition (C6) are the most commonly used ones in nonparametric literature.

## 2.2 Main Results

We shall assume that there always exists such a $\sqrt{n}$-consistent estimator for the parameter $\theta$ in the regression function under the null hypothesis. The theorem below states the asymptotic null distribution of the test statistics $T_n$.

**Theorem 1.** *Suppose (C1)-(C7) hold. Then under the null hypotheses $H_0$ in (2.2), $T_n = nh^{d/2}V_n/\hat{\sigma} \Rightarrow N(0,1)$, where $V_n$ is defined in (2.3) and $\hat{\sigma}$ in (2.5).*

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

For fixed alternatives, a reasonable test should have power approaching to 1 as the sample size goes to $\infty$. That is, a desirable test should be consistent. For this purpose, consider a class of fixed alternative hypotheses:

$$H_a : E(Y^*|X = x) = m(x), \; m(x) \neq m(x, \theta) \text{ for any } \theta, \tag{2.6}$$

such that $Em^2(X) < \infty$.

To show the consistency of the proposed procedure, we have to consider the asymptotic behavior of $\hat{\theta}_n$ under the alternative hypothesis. In the classic 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. Similarly, for the regression model (2.1), if we define $\theta_a = \text{argmin}_\theta E[Y - g(X, \theta)]^2$, under the alternative hypothesis and some regularity conditions on regression functions, one can show that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$. We will not pursue a rigorous verification of this claim here. The relevant discussion can be found in Zheng (1996).

**Theorem 2.** *Suppose all the conditions in Theorem 1 hold with $\theta_0$ replaced by $\theta_a$, $E[g(X) - g(X, \theta_a)]^2 f_X(X) > 0$. Then for any $0 < \alpha < 1$, the test that rejects $H_0$ in (2.2) whenever $|T_n| > z_{1-\alpha/2}$ is consistent against the alternatives $H_a$ in (2.6).*

Sometimes it is desirable to investigate the performance of a test statistic at local alternatives, in particular, when we have to determine the sample sizes needed to achieve a desired power. For this purpose, let $\delta(x)$ be a continuous function such that $E\delta^2(X)w^2(X) < \infty$, and consider the following sequence of local alternatives

$$H_{LOC} : m(x) = m(x, \theta_0) + \delta(x)/\sqrt{nh^{d/2}}. \tag{2.7}$$

We keep on assuming that the estimators $\hat{\theta}_n$ used in the test statistic satisfies $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ without a rigorous justification. We have

**Theorem 3.** *Suppose all the conditions in Theorem 1 hold, then under $H_{LOC}$ in (2.7), $T_n \Rightarrow N(\mu, 1)$, where $\mu = EQ_0^2(-m(X,\theta_0))\delta^2(X)w^2(X)f_X(X)/\sigma$ and $\sigma$ is defined as in (2.4).*

From Theorem 3, we conclude that the asymptotic power of the test proposed is $1 - \Phi(z_{\alpha/2} - \mu) + \Phi(-z_{\alpha/2} - \mu)$, which is an increasing function of $\mu$. Thus, the weight function $w$ that maximizes the power is the one that maximizes $\mu$. But

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

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

If the regression model $I(Y = 0) = 1 - Q_0(-m(X)) + \eta$ is used for model checking, and the first order derivative of the density function $f_\varepsilon$ of $\varepsilon$ is bounded, then under the local alternative hypothesis (2.7), one can show that $T_n \Rightarrow N(\mu, 1)$ with $\mu = Ef_\varepsilon^2(-m(X,\theta_0))\delta^2(X)w^2(X)f_X(X)/\sigma$ and $\sigma$ is defined in (2.4). It is easily seen that the optimal weight function is $w(x) = f_\varepsilon(-m(x,\theta_0))\delta(x)/\tau^2(x)$.

## 2.3 Proofs of the Main Results

**Lemma 1.** *As a function of $m(x)$, $g(x)$ is strictly monotone provided that $F_\varepsilon$ is strictly increasing.*

*Proof of the Lemma 1.* Let $h(x) = xQ_0(-x) + Q_1(-x)$, take the derivative of $h(x)$:

$$\begin{aligned}
h'(x) &= Q_0(x) + xQ_0'(x) - Q_1'(x) \\
&= \int_x^\infty f(u)du + xf(x) - xf(x) \\
&= 1 - F(x) > 0
\end{aligned}$$

So $h(x)$ is strictly monotone provided that $F_\varepsilon$ is strictly increasing. In our case, let $m(X) = x$, $g(X)$, as a function of $m(x)$ is strictly monotone provided that $F_\varepsilon$ is strictly increasing. This property is essential to the construction of our test statistics. $\square$

The proof of Theorem 1 is facilitated by the lemmas stated below.

**Lemma 2.** *Under condition (C1),* $g(x) = xQ_0(-x) + Q_1(-x) = \int_{-x}^{\infty}[1 - F_\varepsilon(u)]du.$

*Proof of the Lemma 2.* (C1) implies the existence of $E|\varepsilon|$, hence $lim_{x\to\infty} x[1 - F_\varepsilon(x)] = 0$. Note that

$$\begin{aligned} g(x) &= xQ_0(-x) + Q_1(-x) \\ &= x[1 - F_\varepsilon(-x)] + \int_{-x}^{\infty} u f_\varepsilon(u)du \\ &= x[1 - F_\varepsilon(-x)] - \int_{-x}^{\infty} ud[1 - F_\varepsilon(u)] \\ &= x[1 - F_\varepsilon(-x)] - u[1 - F_\varepsilon(u)] \,|_{-x}^{\infty} + \int_{-x}^{\infty} [1 - F_\varepsilon(u)]du \\ &= \int_{-x}^{\infty} [1 - F_\varepsilon(u)]du \end{aligned}$$

Hence the lemma. $\square$

**Lemma 3.** *From condition (C1), for each x,*

$$|g(-m(x,\theta_1)) - g(-m(x,\theta_2)) - [m(x,\theta_1) - m(x,\theta_2)]Q_0(-m(x,\theta_2))| \le B[m(x,\theta_1) - m(x,\theta_2)]^2$$

*for some constant B.*

*Proof of the Lemma 3.* A Taylor expansion of $g$ function and an application of Lemma 2, together with the boundedness of $f_\varepsilon$, imply the result. $\square$

To find out the asymptotic distribution of the test statistic, we also need Theorem 1 of Hall (1984) which is reproduced here for the sake of completeness.

**Lemma 4.** *Let $Z_i, i = 1, 2, ..., n$ be i.i.d. random vectors, and let $U_n = \sum_{1 \le i < j \le n} H_n(Z_i, Z_j)$, $M_n(x, y) = EH_n(Z_1, x)H_n(Z_1, y)$, where $H_n$ is a sequence of measurable functions symmetric under permutation, with*

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

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

*Proof of Theorem 1.* Let $\xi_i = Y_i - g(X_i, \theta_0)$. Then $\hat{\xi}_i = Y_i - g(X_i, \hat{\theta}_n) = \xi_i - [g(X_i, \hat{\theta}_n) - g(X_i, \theta_0)]$. Hence $V_n$ (2.3) can be written as the sum of the following three terms

$$\begin{aligned}
V_{1n} &= \frac{1}{n(n-1)} \sum_{i \ne j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i \xi_j w(X_i) w(X_j), \\
V_{2n} &= -\frac{2}{n(n-1)} \sum_{i \ne j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i [g(X_j, \hat{\theta}_n) - g(X_j, \theta_0)] w(X_i) w(X_j), \\
V_{3n} &= \frac{2}{n(n-1)} \sum_{i \ne j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) [g(X_i, \hat{\theta}_n) - g(X_i, \theta_0)][g(X_j, \hat{\theta}_n) - g(X_j, \theta_0)] w(X_i) w(X_j).
\end{aligned}$$

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

$$H_n(Z_i, Z_j) = \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i \xi_j w(X_i) w(X_j).$$

Under the null hypothesis, we have $E[H_n(Z_1, Z_2)|Z_1] = 0$, so $V_{1n}$ is a degenerate U-statistic. To apply Lemma 4 to show the asymptotic normality of $V_{1n}$, a series of conditions needs to be verified. For this purpose, we have to investigate the asymptotic behaviors of $EM_n^2(Z_1, Z_2)$, $EH_n^4(Z_1, Z_2)$, and $EH_n^2(Z_1, Z_2)$, where $M_n(x, y) = EH_n(Z_1, x)H_n(Z_1, y)$ is defined as in

Lemma 4. For $EM_n^2(Z_1, Z_2)$, we have

$$
\begin{aligned}
& EM_n^2(Z_1, Z_2) \\
= \;& E(E[H_n(Z_3, Z_1)H_n(Z_3, Z_2)|Z_1, Z_2])^2 \\
= \;& E(E\left[\frac{1}{h^{2d}}K\left(\frac{X_3 - X_1}{h}\right)K\left(\frac{X_3 - X_2}{h}\right)w(X_1)w(X_2)w^2(X_3)\xi_1\xi_2\xi_3^2|Z_1, Z_2\right])^2 \\
= \;& \frac{1}{h^{4d}}E(\xi_1\xi_2 w(X_1)w(X_2)\int K(\frac{x_3 - X_1}{h})K\left(\frac{x_3 - X_2}{h}\right)w^2(x_3)\tau^2(x_3)f_X(x_3)dx_3)^2 \\
= \;& \frac{1}{h^{2d}}E(\xi_1\xi_2 w(X_1)w(X_2)\int K(u)K\left(u + \frac{X_1 - X_2}{h}\right)w^2(X_1 + hu)\tau^2(X_1 + hu)f_X(X_1 + hu)du)^2 \\
= \;& \frac{1}{h^{2d}}\iint \tau^2(x_1)\tau^2(x_2)w^2(x_1)w^2(x_2)[\int K(u)K\left(u + \frac{x_1 - x_2}{h}\right)w^2(x_1 + hu)\cdot \\
& \tau^2(x_1 + hu)f_X(x_1 + hu)du]^2 f_X(x_1)f_X(x_2)dx_1dx_2 \\
= \;& \frac{1}{h^{d}}\iint \tau^2(x_1)\tau^2(x_1 - hv)w^2(x_1)w^2(x_1 - hv)[\int K(u)K(u + v)w^2(x_1 + hu)\cdot \\
& \tau^2(x_1 + hu)f_X(x_1 + hu)du]^2 f_X(x_1)f_X(x_1 - hv)dvdx_1dx_2 \\
= \;& \frac{1}{h^{d}}\int[\int K(u)K(u + v)du]^2 dv\int[\tau^2(x)]^4 w^8(x)f_X^4(x)dx + o(1/h^{d\cdot})
\end{aligned}
$$

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

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

Similarly, for $EH_n^4(Z_1, Z_2)$, we have

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

Therefore, from (C6), we obtain

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

By Lemma 4, we show that

$$nh^{d/2}V_{1n} \Rightarrow N(0, \sigma^2), \tag{2.8}$$

where $\sigma^2$ is defined in (2.4).

Now let's consider $V_{2n}$. By adding and subtracting $[m(X_i,\hat\theta_n)-m(X_i,\theta_0)]Q_0(-m(X_i,\theta_0))$ from $g(-m(X_i,\hat\theta_n))-g(-m(X_i,\theta_0))$, and denoting

$$\begin{aligned} d_{ni} &= g(X_j,\hat\theta_n)-g(X_j,\theta_0)-[m(X_i,\hat\theta_n)-m(X_i,\theta_0)]Q_0(-m(X_i,\theta_0)), && (2.9)\\ \delta_{ni} &= m(X_i,\hat\theta_n)-m(X_i,\theta_0)-(\hat\theta_n-\theta_0)'\dot m(X_i,\theta_0), && (2.10)\end{aligned}$$

$V_{2n}$ can be written as the sum of the following two terms

$$\begin{aligned} V_{2n1} &= -\frac{2}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)\xi_i d_{nj}w(X_i)w(X_j),\\ V_{2n2} &= -\frac{2}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)\xi_i[m(X_j,\hat\theta_n)-m(X_j,\theta_0)]Q_0(-m(X_j,\theta_0))w(X_i)w(X_j).\end{aligned}$$

Applying Lemma 3 for $\theta_1=\hat\theta_n,\theta_2=\theta_0$, and $x=X_i, i=1,2,...,n$, we have

$$\begin{aligned} |V_{2n1}| &\leq \frac{2B}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|[m(X_j,\hat\theta_n)-m(X_j,\theta_0)]^2w(X_i)w(X_j)\\ &\leq \frac{4B}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|\delta_{nj}^2w(X_i)w(X_j)\\ &\quad+\frac{4B}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|[(\hat\theta_n-\theta_0)'\dot m(X_j,\theta_0)]^2w(X_i)w(X_j)\\ &= A_{n1}+A_{n2}.\end{aligned}$$

According to Condition (C4), $A_{n1}$ is bounded above by

$$O_p(\frac{1}{n^2})\cdot\frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|w(X_i)w(X_j).$$

Note that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|w(X_i)w(X_j)=O_p(1).$$

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

For $A_{n2}$, we have

$$nh^{d/2}A_{n2}\leq nh^{d/2}\parallel\hat\theta_n-\theta_0\parallel^2\cdot\frac{4B}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)|\xi_i|\parallel\dot m(X_j,\theta_0)\parallel^2w(X_i)w(X_j)=o_p(1)$$

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

$$\frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) |\xi_i| \parallel \dot{m}(X_j, \theta_0) \parallel^2 w(X_i) w(X_j) = O_p(1).$$

Hence,

$$nh^{d/2} V_{2n1} = o_p(1). \tag{2.11}$$

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

$$\begin{aligned} B_{n1} &= -\frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i \delta_{nj} Q_0(-m(X_j, \theta_0)) w(X_i) w(X_j), \\ B_{n2} &= -\frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i (\hat{\theta}_n - \theta_0)' \dot{m}(X_j, \theta_0) Q_0(-m(X_j, \theta_0)) w(X_i) w(X_j). \end{aligned}$$

By Condition (C4),

$$|B_{n1}| \leq \sup_{1 \leq i \leq n} |\delta_{ni}| \cdot \frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) |\xi_i| w(X_i) w(X_j) = O_p(1/n),$$

thus, $nh^{d/2} B_{n1} = o_p(1)$.

As for $B_{n2}$, it is easily seen that

$$|B_{n2}| \leq \parallel \hat{\theta}_n - \theta_0 \parallel \cdot \parallel \frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \xi_i m(X_j, \theta_0) Q_0(-m(X_j, \theta_0)) \parallel w(X_i) w(X_j).$$

Using the similar method as in proving Lemma 3.3b in Zheng (1996), one can show that the second norm in the above inequality is the order of $O_p(1/\sqrt{n})$, which, together with the $\sqrt{n}$-consistency of $\hat{\theta}_n$, implies $nh^{d/2} B_{n2} = o_p(1)$. Thus,

$$nh^{d/2} V_{2n2} = o_p(1). \tag{2.12}$$

From (2.11) and (2.12), we obtain

$$nh^{d/2} V_{2n} = o_p(1). \tag{2.13}$$

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

To show that $\hat{\sigma}^2$ defined in (2.5), it is sufficient to show that

$$\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)(\hat{\xi}_i^2\hat{\xi}_j^2-\xi_i^2\xi_j^2)w^2(X_i)w^2(X_j)=o_p(1), \tag{2.14}$$

$$\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i^2\xi_j^2w^2(X_i)w^2(X_j)=\sigma^2+o_p(1). \tag{2.15}$$

Adding and subtracting $\xi_i$ from $\hat{\xi}_i$, $\xi_j$ from $\hat{\xi}_j$, the term on the left hand side of (2.14) can be written as the sum of the following five terms,

$$\begin{aligned}
&\frac{2}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)(\hat{\xi}_i-\xi_i)^2(\hat{\xi}_j-\xi_j)^2w^2(X_i)w^2(X_j),\\
&\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j(\hat{\xi}_i-\xi_i)(\hat{\xi}_j-\xi_j)w^2(X_i)w^2(X_j),\\
&\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i(\hat{\xi}_i-\xi_i)(\hat{\xi}_j-\xi_j)^2w^2(X_i)w^2(X_j),\\
&\frac{4}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i^2(\hat{\xi}_j-\xi_j)^2w^2(X_i)w^2(X_j),\\
&\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2(\hat{\xi}_i-\xi_i)w^2(X_i)w^2(X_j).
\end{aligned} \tag{2.16}$$

We only show that (2.16) is the order of $o_p(1)$. Note that

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

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

$$-\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2d_{ni}w^2(X_i)w^2(X_j), \tag{2.17}$$

$$-\frac{8}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2\delta_{ni}Q_0(-m(X_i,\theta_0))w^2(X_i)w^2(X_j), \tag{2.18}$$

$$-\frac{8(\hat{\theta}_n-\theta_0)'}{n(n-1)}\sum_{i\neq j}^{n}\frac{1}{h^d}K^2\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j^2\dot{m}(X_i,\theta_0)Q_0(-m(X_i,\theta_0))w^2(X_i)w^2(X_j) \tag{2.19}$$

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

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

If the last quantity is bounded, we have

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

Since $\sup_{1\leq i\leq n}|d_{ni}|$ and $\sup_{1\leq i\leq n}|\delta_{ni}|$ are both negligible, it is easily seen that (2.17), (2.18) and (2.19) are all of the order of $o_p(1)$.

The proof of (2.15) is similar to the proof of Lemma 3.3e in Zheng (1996), and this completes the proof of the theorem. □

*Proof of Theorem 2.* The proof is similar to that of Theorem 1. We only outline the main steps here for the sake of brevity.

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

$$
\begin{aligned}
V_{1n}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)\xi_i\xi_j w(X_i)w(X_j), \\
V_{2n}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)\xi_i[g(X_j)-g(X_j,\hat\theta_n)]w(X_i)w(X_j), \\
V_{3n}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)[g(X_i)-g(X_i,\hat\theta_n))][g(X_j)-g(x_j,\hat\theta_n)]w(X_i)w(X_j).
\end{aligned}
$$

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

$$\begin{aligned}
V_{3n1}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)[g(X_i)-g(X_i,\theta_a)][g(X_j)-g(X_j,\theta_a)]w(X_i)w(X_j),\\
V_{3n2}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)[g(X_i)-g(X_i,\theta_a)][g(X_j,\theta_a)-g(X_j,\hat\theta_n)]w(X_i)w(X_j),\\
V_{3n3}^a &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}K\left(\frac{X_i-X_j}{h}\right)[g(X_i,\theta_a)-g(X_i,\hat\theta_n)][g(X_j,\theta_a)-g(X_j,\hat\theta_n)]w(X_i)w(X_j).
\end{aligned}$$

One can show that

$$V_{3n1}^a = \int K^2(u)du\cdot\int [g(x)-g(x,\theta_a)]^2\, w^2(x)f_X^2(x)dx + o_p(1),$$

and $V_{3n2}^a = o_p(1), V_{3n3}^a = o_p(1), V_{2n}^a = o_p(1)$. Eventually, one can show that

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

Finally, we can show that

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

This completes the proof. □

*Proof of Theorem 3.* Now, define $Y_i^{*L} = m(X_i,\theta_0)+\varepsilon_i$, $Y_i^L = max\{Y_i^{*L},0\}$, and $W_i = Y_i - Y_i^L$. The elementary inequality $\max\{a,0\} = (a+|a|)/2$ implies that $W_i = [\delta(X_i)+\Delta_n(X_i)]/2\sqrt{nh^{d/2}}$ with

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

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

$$\begin{aligned}
V_{1n}^L &= \frac{1}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)\hat\xi_i^L\hat\xi_j^L w(X_i)w(X_j),\\
V_{2n}^L &= \frac{2}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)\hat\xi_i^L W_j w(X_i)w(X_j),\\
V_{3n}^L &= \frac{1}{n(n-1)h^d}\sum_{i\neq j}K\left(\frac{X_i-X_j}{h}\right)W_iW_j w(X_i)w(X_j).
\end{aligned}$$

Similar to the proof of Theorem 1, $nh^{d/2}V_{1n}^L \Rightarrow N(0,\sigma^2)$, where $\sigma^2$ is defined in (2.4).

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

$$\begin{aligned}
& \frac{1}{h^d}EK\left(\frac{X_1-X_2}{h}\right)W_1W_2w(X_1)w(X_2) \\
= & \frac{1}{nh^{3d/2}}\cdot \\
& \iint K\left(\frac{x_1-x_2}{h}\right)Q_0(-m(x_1,\theta_0))Q_0(-m(x_2,\theta_0))\delta(x_1)\delta(x_2)w(x_1)w(x_2)f_X(x_1)f_X(x_2)dx_1dx_2 \\
= & \frac{1}{nh^{d/2}}\cdot \\
& \iint K(u)Q_0(-m(x+hu,\theta_0))Q_0(-m(x,\theta_0))\delta(x+hu)\delta(x)w(x+hu)w(x)f_X(x+hu)f_X(x)dxdu \\
= & \frac{1}{nh^{d/2}}\int Q_0^2(-m(x,\theta_0))\delta^2(x)w^2(x)f_X^2(x)dx+o\left(\frac{1}{nh^{d/2}}\right).
\end{aligned}$$

The last integral is exactly the $\mu$ defined in Theorem 3. Therefore,

$$nh^{d/2}EV_{3n}^L \to \mu \tag{2.20}$$

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

$$S_{ij}=K\left(\frac{X_i-X_j}{h}\right)W_iW_jw(X_i)w(X_j)-EK\left(\frac{X_i-X_j}{h}\right)W_iW_jw(X_i)w(X_j).$$

A simple but tedious derivation leads to

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

Note that $|ES_{12}S_{13}|\le ES_{12}^2$ by Cauchy-Schwartz inequality, $\mathrm{Var}(V_{3n}^L)$ is bounded above by

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

One can show that

$$\begin{aligned}
ES_{12}^2 &= E[K\left(\frac{X_1-X_2}{h}\right)W_1W_2w(X_1)w(X_2)-EK\left(\frac{X_1-X_2}{h}\right)W_1W_2w(X_1)w(X_2)]^2 \\
&\le EK^2\left(\frac{X_1-X_2}{h}\right)W_1^2W_2^2w^2(X_1)w^2(X_2)
\end{aligned}$$

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

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

Therefore,

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

which implies

$$V_{3n}^L - EV_{3n}^L = O_p\left(\frac{1}{\sqrt{n^3h^{2d}}}\right). \tag{2.21}$$

From (2.20) and (2.21), one can get, in probability,

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

Similarly, one can show that $nh^{d/2}V_{2n}^L = o_p(1)$, and $\hat{\sigma}^2 \to \sigma^2$ in probability. Details are omitted for the sake of brevity. Summarizing the above arguments, we finish the proof of Theorem 3. $\square$

# Chapter 3

# Numerical Studies

## 3.1 Simulation Studies

Two sets of Monte Carlo simulations, one-dimensional and two-dimensional linear regression functions serving as the models under null hypothesis, are conducted in this section to assess the finite sample performance of the test proposed. A variety of quadratic components, $\gamma$=0, 0.1, 0.2, 0.3, 0.5, to be specific, are added to the linear terms, serving as the alternative models. For both sets of the simulation, models with $\gamma = 0$ are used to study the empirical size, while models with $\gamma$=0.1, 0.2, 0.3, 0.5 are used to study the empirical powers. Sample sizes are set at $n = 100, 300, 500, 800, 1000$. For each set of the Monte Carlo simulation, two sets of models

$$\begin{aligned}\text{Model I} &: \quad E(Y|X = x) = m(x)Q_0(-m(x)) + Q_1(-m(x)),\\ \text{Model II} &: \quad E(I(Y = 0)|X = x) = F_\varepsilon(-m(x)),\end{aligned}$$

are studied under all the combinations of sample size, $n$ and quadratic components, $\gamma$, by repeating the tests 1000 times. The empirical level and power are determined by $\#\{|T_n| \geq 1.96\}/1000$. The simulation setups are exactly the same as in Song and Zhang (2011)'s.

Ideally, the optimal weight functions should be used in the model checking. However, the optimal weight functions depend on the true departures of the alternative models from the null models, which, in real application, are rarely known, although some approximations could be obtained by exploratory data analysis. Therefore, in the following simulation studies, we will assess the tests proposed using the noninformative weight function $w(x) = 1$. For comparison purpose, simulation with an optimal weight functions are also considered.

*Simulation 1:* The data are generated from the one-dimensional linear regression model

$$Y^* = \alpha + \beta X + \gamma X^2 + \varepsilon, Y = max\{Y^*, 0\}. \tag{3.1}$$

In the simulation, with $X \sim N(0,1), \varepsilon \sim N(0, \sigma_\varepsilon^2)$, 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. Theoretically, under current settings $P(\varepsilon \leq -1 - X) \approx 24\%$ observations of $Y^*$ are truncated below 0 when $\gamma = 0$. The vglm function in the R package VGAM is used to calculate the estimates of all unknown parameters.

First, the uninformative weight function $w(x) = 1$ is considered. For Model I, the simulation result presented on the left part of Table 3.1 shows that the empirical levels are less than the nominal levels without any exception, 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.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.006 | 0.004 | 0.005 | 0.009 | 0.006 | 0.048 | 0.053 | 0.056 | 0.051 | 0.052 |
| $\gamma$=0.1 | 0.016 | 0.049 | 0.091 | 0.200 | 0.320 | 0.070 | 0.162 | 0.172 | 0.324 | 0.474 |
| $\gamma$=0.2 | 0.086 | 0.447 | 0.778 | 0.965 | 0.995 | 0.268 | 0.682 | 0.860 | 0.979 | 0.998 |
| $\gamma$=0.3 | 0.353 | 0.924 | 0.996 | 1.000 | 1.000 | 0.557 | 0.978 | 1.000 | 1.000 | 1.000 |
| $\gamma$=0.5 | 0.903 | 1.000 | 1.000 | 1.000 | 1.000 | 0.967 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 3.1**: *d=1. Empirical powers based on Model I, left panel based on normal simulation, right panel based on bootstrap*

In general, bootstrap provides 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 800 times, the critical values are then obtained by finding out the upper 97.5th percentile and lower 2.5th percentile of these 800 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 percentile and bigger than the 97.5th percentile. The right part of Table 3.2 reports the simulation results. As expected, all the empirical levels are very close to the nominal level 0.05, and the powers are much larger than the ones reported on the left part of Table 3.1.

We also did a simulation study based on Model II. Under this condition, the same test statistic is used with $\hat{\xi}_i$ being replace by $I(Y_i = 0) - F_\varepsilon(-m(X_i, \hat{\theta}))$. The simulation result is presented in Table 3.2. The left part of Table 3.2 is the simulation results based on the critical values from Theorem 1, while the right part of Table 3.2 is the simulation results based on the bootstrap critical values. As discussed in Chapter 2, the test is much less powerful than the one based on Model I.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.010 | 0.006 | 0.011 | 0.014 | 0.012 | 0.050 | 0.039 | 0.042 | 0.048 | 0.054 |
| $\gamma$=0.1 | 0.008 | 0.014 | 0.035 | 0.058 | 0.089 | 0.050 | 0.045 | 0.051 | 0.102 | 0.128 |
| $\gamma$=0.2 | 0.015 | 0.095 | 0.191 | 0.365 | 0.479 | 0.075 | 0.138 | 0.220 | 0.539 | 0.654 |
| $\gamma$=0.3 | 0.029 | 0.225 | 0.528 | 0.829 | 0.927 | 0.117 | 0.371 | 0.597 | 0.924 | 0.972 |
| $\gamma$=0.5 | 0.136 | 0.710 | 0.970 | 0.999 | 1.000 | 0.301 | 0.811 | 0.977 | 1.000 | 1.000 |

**Table 3.2**: *d=1. Empirical powers based on Model II, left panel based on normal simulation, right panel based on bootstrap*

For comparison purpose, we also conduct simulation studies based on the optimal weight functions. Let $z = (\alpha + \beta x)/\sigma_\varepsilon$. For model I, the optimal weight function is $w(x) = \Phi(z)x^2/\tau^2(x)$, and $\tau^2(x) = \sigma_\varepsilon^2[1 + z^2]\Phi(z) + \sigma_\varepsilon^2 z\phi(z) - \sigma_\varepsilon^2[z\Phi(z) + \phi(z)]^2$. For model II, the optimal weight function is $w(x) = \phi(z)x^2/\tau^2(x)$, and $\tau^2(x) = \Phi(z)(1 - \Phi(z))$. Again, $\alpha$, $\beta$ and $\sigma_\varepsilon$ are estimated as above. Simulation results are presented in Table 3.3. The left part is for the test based on model I, and the right part is for the test based on the model II. It is easily seen that the empirical levels are closer to 0.05 than the ones reported in Table 3.1 and 3.2, and as expected, the empirical powers are larger than the corresponding ones in Table 3.1 and 3.2 for large sample sizes. Recall that the optimal weight functions are indeed "optimal" asymptotically, so it would be no surprise to us when the empirical powers are less than the ones reported in Table 3.1 and 3.2 for small sample cases. Also, the test based on the model I is more powerful than the one based on the model II.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma = 0$ | 0.007 | 0.014 | 0.014 | 0.016 | 0.024 | 0.036 | 0.018 | 0.028 | 0.024 | 0.026 |
| $\gamma = 0.1$ | 0.016 | 0.078 | 0.213 | 0.387 | 0.533 | 0.019 | 0.057 | 0.117 | 0.207 | 0.298 |
| $\gamma = 0.2$ | 0.077 | 0.584 | 0.877 | 0.988 | 0.998 | 0.065 | 0.358 | 0.654 | 0.911 | 0.958 |
| $\gamma = 0.3$ | 0.267 | 0.942 | 0.991 | 0.996 | 0.998 | 0.157 | 0.763 | 0.963 | 1.000 | 1.000 |
| $\gamma = 0.5$ | 0.715 | 0.987 | 0.994 | 0.999 | 1.000 | 0.611 | 0.994 | 1.000 | 1.000 | 1.000 |

**Table 3.3**: *d=1. Empirical powers with optimal weights, left panel based on Model I, right panel based on Model II*

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

$$Y^* = \alpha + \beta_1 X_1 + \beta_2 X_2 + \gamma(X_1^2 + X_2^2) + \varepsilon, Y = max\{Y^*, 0\}. \tag{3.2}$$

In the simulation, $(X_1, X_2)$ is from a bivariate normal distribution with $\mathbf{0}$ mean vector, and identity covariance matrix with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. 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. Theoretically, under current settings, $P(\varepsilon \le -1 - X_1 - X_2) \approx 28\%$ observations of $Y^*$ are truncated below 0 when $\gamma = 0$. The censReg function in the R package censReg is used to estimate all unknown parameters.

For Model I, the simulation result presented on the left part of Table 3.4 preserves to be conservative. The power enhances with the increase of sample size. Similar to the one-dimensional case, we conduct a parametric bootstrap simulation and the results are shown on the right part of Table 3.4. Clearly, the nominal level 0.05 is well preserved in the bootstrap simulation and the power is much larger than the one shown on the left part of Table 3.4.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.002 | 0.008 | 0.010 | 0.009 | 0.019 | 0.054 | 0.047 | 0.050 | 0.045 | 0.046 |
| $\gamma$=0.1 | 0.034 | 0.158 | 0.283 | 0.529 | 0.672 | 0.111 | 0.221 | 0.388 | 0.640 | 0.802 |
| $\gamma$=0.2 | 0.247 | 0.838 | 0.985 | 1.000 | 1.000 | 0.423 | 0.903 | 0.994 | 1.000 | 1.000 |
| $\gamma$=0.3 | 0.690 | 0.999 | 1.000 | 1.000 | 1.000 | 0.851 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\gamma$=0.5 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 3.4**: *d=2, $\rho(X_1, X_2) = 0.2$, Empirical powers based on Model I, left panel based on normal simulation, right panel based on bootstrap*

The simulation result for Model II is presented in Table 3.5. The formulation of Table 3.5 is the same as the one in Table 3.4. It is easily seen that the test is much less powerful than the one based on Model I.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.008 | 0.012 | 0.006 | 0.014 | 0.013 | 0.050 | 0.053 | 0.050 | 0.052 | 0.042 |
| $\gamma$=0.1 | 0.010 | 0.023 | 0.038 | 0.081 | 0.087 | 0.046 | 0.066 | 0.086 | 0.127 | 0.128 |
| $\gamma$=0.2 | 0.021 | 0.090 | 0.207 | 0.396 | 0.519 | 0.070 | 0.211 | 0.247 | 0.479 | 0.536 |
| $\gamma$=0.3 | 0.036 | 0.235 | 0.494 | 0.798 | 0.904 | 0.091 | 0.371 | 0.605 | 0.882 | 0.951 |
| $\gamma$=0.5 | 0.102 | 0.622 | 0.948 | 1.000 | 1.000 | 0.193 | 0.792 | 0.968 | 1.000 | 1.000 |

**Table 3.5**: *d=2, $\rho(X_1, X_2) = 0.1$, Empirical powers based on Model II, left panel based on normal simulation, right panel based on bootstrap*

We also did some simulation studies when $X_1$ and $X_2$ are correlated with $\rho(X_1, X_2) = 0.2$ and $\rho(X_1, X_2) = 0.5$. Table 3.6 is for the occasion when $\rho(X_1, X_2) = 0.2$. The left part of this table is the result of the simulation study based on Model I, while the other part is based on Model II.

Table 3.7 is for the occasion when $\rho(X_1, X_2) = 0.5$. The left part of this table is the result of the simulation study based on Model I, while the other part is based on Model II.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.003 | 0.008 | 0.009 | 0.008 | 0.019 | 0.004 | 0.013 | 0.009 | 0.019 | 0.010 |
| $\gamma$=0.1 | 0.025 | 0.125 | 0.240 | 0.474 | 0.589 | 0.007 | 0.020 | 0.029 | 0.060 | 0.066 |
| $\gamma$=0.2 | 0.219 | 0.786 | 0.978 | 1.000 | 1.000 | 0.016 | 0.074 | 0.140 | 0.309 | 0.423 |
| $\gamma$=0.3 | 0.666 | 0.999 | 1.000 | 1.000 | 1.000 | 0.033 | 0.194 | 0.430 | 0.750 | 0.881 |
| $\gamma$=0.5 | 0.997 | 1.000 | 1.000 | 1.000 | 1.000 | 0.118 | 0.650 | 0.970 | 1.000 | 1.000 |

**Table 3.6**: *d=2, $\rho(X_1, X_2) = 0.2$, Empirical powers, left panel based on Model I, right panel based on Model II*

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.002 | 0.009 | 0.007 | 0.007 | 0.016 | 0.002 | 0.011 | 0.007 | 0.013 | 0.015 |
| $\gamma$=0.1 | 0.025 | 0.097 | 0.202 | 0.425 | 0.535 | 0.007 | 0.020 | 0.035 | 0.053 | 0.053 |
| $\gamma$=0.2 | 0.185 | 0.751 | 0.967 | 1.000 | 1.000 | 0.013 | 0.058 | 0.109 | 0.265 | 0.358 |
| $\gamma$=0.3 | 0.683 | 0.998 | 1.000 | 1.000 | 1.000 | 0.029 | 0.165 | 0.358 | 0.684 | 0.828 |
| $\gamma$=0.5 | 0.995 | 1.000 | 1.000 | 1.000 | 1.000 | 0.181 | 0.756 | 0.990 | 1.000 | 1.000 |

**Table 3.7**: *d=2, $\rho(X_1, X_2) = 0.5$, Empirical powers, left panel based on Model I, right panel based on Model II*

*Remark 3.1:* For the purpose of comparison, a simulation study for simple null hypotheses is conducted by using the same setups as in Simulation 1 and 2 but assuming that $\alpha = \beta_1 = \beta_2 = \sigma_\varepsilon^2 = 1$ are all known in the null models. The results of the simulation under this circumstance are shown in Table 3.8. The empirical level is much closer to the nominal level 0.05 in all cases.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.038 | 0.043 | 0.046 | 0.056 | 0.050 | 0.047 | 0.057 | 0.053 | 0.044 | 0.055 |
| $\gamma$=0.1 | 0.075 | 0.126 | 0.189 | 0.326 | 0.433 | 0.117 | 0.302 | 0.524 | 0.736 | 0.844 |
| $\gamma$=0.2 | 0.195 | 0.557 | 0.847 | 0.976 | 0.998 | 0.438 | 0.962 | 0.998 | 1.000 | 1.000 |
| $\gamma$=0.3 | 0.454 | 0.958 | 0.999 | 1.000 | 1.000 | 0.859 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\gamma$=0.5 | 0.922 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 3.8**: *Simulation results for the simple hypotheses, left panel based on Model I, right panel based on Model II*

*Remark 3.2 :* Under the same setup as described in Remark 3.1, we also did the simulation when $\rho(X_1, X_2) = 0.2$ and $\rho(X_1, X_2) = 0.5$, respectively. The results of the simulation under this circumstance are shown in Table 3.9. The empirical level is much closer to the nominal level 0.05 in all cases.

| | 100 | 300 | 500 | 800 | 1000 | 100 | 300 | 500 | 800 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma$=0 | 0.045 | 0.060 | 0.056 | 0.047 | 0.057 | 0.050 | 0.058 | 0.064 | 0.052 | 0.055 |
| $\gamma$=0.1 | 110 | 0.298 | 0.508 | 0.718 | 0.824 | 0.115 | 0.323 | 0.545 | 0.757 | 0.860 |
| $\gamma$=0.2 | 0.429 | 0.951 | 0.998 | 1.000 | 1.000 | 0.456 | 0.954 | 0.998 | 1.000 | 1.000 |
| $\gamma$=0.3 | 0.833 | 1.000 | 1.000 | 1.000 | 1.000 | 0.820 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\gamma$=0.5 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 |

**Table 3.9**: *Simulation results for the simple hypotheses, left panel when $\rho(X_1, X_2) = 0.2$, right panel when $\rho(X_1, X_2) = 0.5$*

*Remark 3.3 :* We also made a comparison study with Song and Zhang (2011)'s test procedure, and the results are mixed. When we rely on R packages to estimate the parameter, for Model I, the proposed test is slightly powerful than Song and Zhang (2011)'s when $d = 1$, but less powerful when $d = 2$. The conclusion reverses if bootstrap critical values are used. Song and Zhang (2011)'s test is more powerful than the test proposed in this report. For Model II, the test proposed in this report outperforms Song and Zhang (2011)'s for both of the cases. For simple null hypothesis with $\alpha = \beta_1 = \beta_2 = \sigma_\varepsilon^2 = 1$ all known, the result of the comparison with Song and Zhang (2011)'s test doesn't change much.

## 3.2 A Real Data Application

Thomas A. Mroz in 1987 undertook a systematic analysis of several theoretic and statistical assumptions used in many empirical models of female labor supply. The data for the analysis came from the University of Michigan Panel Study of Income Dynamics for the year 1975, which has been cited multiple time either for the purpose of research, or for the purpose of academic demonstration.

The sample consists of 753 married white women between the ages of 30 and 60 in 1975, with 428 working at some time during the year, while the remaining 325 observations are women who did not work for pay in 1975. The dependent variable, the wife's annual hours of work, is the product of the number of weeks the wife worked for money in 1975 and the average number of hours of work per week during the weeks she worked. The histogram of wife's annual hours of work of Figure 1 clearly shows that a large amount of women didn't participate in any work in the labor market, a reason why Tobit regression model is needed.

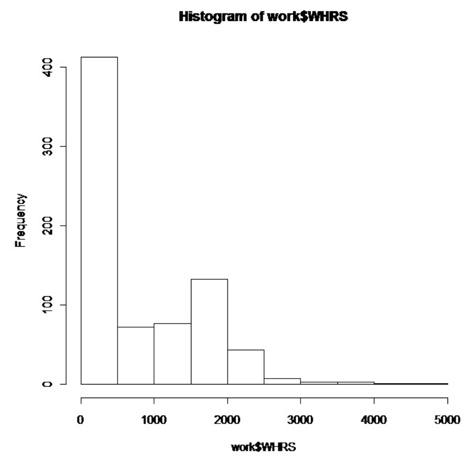


**Figure 3.1**: *Histogram of wife's annual hours of work*

Instead of using all of the 17 variables in the original dataset, we focus on two of them, wife's educational attainment, in years, denoted as WE, and actual years of wife's previous labor market experience, denoted as AX.

We treat both wife's educational attainment and previous labor market experience as independent variables to explain wife's annual hours of work. The Tobit regression takes the form:

$$\begin{aligned} Y_i^* &= m(X) + \varepsilon_i = -1364 + 0.7644 x_{WE} + 0.6981 x_{AX} + \varepsilon_i, i = 1, 2, ..., n, \\ Y_i &= \begin{cases} Y_i^* & \text{if } Y_i^* > 0, \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

We apply the test proposed in this report to check if there is adequacy evidence to explain wife's wife's annual hours of work by using educational attainment and previous labor market. The hypothesis is:

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

We get 4.85 as the test statistic, the corresponding p-value of which is 1.239401e-06. Since p-value is less than 0.05, there is significant evidence to show the inappropriate use of the linear function. Thus, it's not appropriate to use the linear function to express the relationship between wife's annual hours of work and wife's education level and previous labor market.

# Chapter 4

# Conclusion

We proposed a nonparametric lack-of-fit test to check the adequacy of the presumed parametric form for the regression function in Tobit regression models by applying Zheng's device with weighted residuals. It is shown that testing the null hypothesis for the standard Tobit regression models is equivalent to testing a new null hypothesis of the classic regression models, one of which was built based on the whole data set, while the other one of which was built based on the part that has been truncated. An optimal weight function is identified to maximize the local power of the test.The proposed test statistic is shown to be asymptotically normal under null hypothesis, consistent against some fixed alternatives, and has nontrivial power for some local nonparametric power for some local nonparametric alternatives. The applicability of the test proposed is verified by the performance of the finite sample Monte-Carlo simulations.

# Bibliography


[1] Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica,* **41**, 997-1016.

[2] Amemiya, T. (1984). Tobit models: A survey. *J. of Econometrics*, **24**(1-2), 3-61.

[3] Ekstrand, C.,Carpenter, T.E. (1998). Using a tobit regression model to analyse risk factors for foot-pad dermatitis in commercially grown broilers. *Preventive Veterinary Medicine,* **37**(1-4), 219-228

[4] Anastasopoulos, P., Tarko, A., Mannering, F., (2008). Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis and Prevention* **40**(2), 768-775 Andrew P. Tarko, Panagiotis Ch. Anastasopoulos, Fred L. Mannering (2007). Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis and Prevention,* **40**(2008)768-775

[5] Goldberger, A.S. (1964). *Econometric theory.* Wiley, New York.

[6] Hall, P. J. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivariate Anal.,* **14**(1), 1-16

[7] Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann, Statist.,* **21**(4), 1926-1947

[8] Heckman, J. 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**(4), 475-492

[9] Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica,* **47**(1),153-162

[10] Horowitz, J. L., Neumann, G. R. (1989). Specification testing in censored regression models: parametric and semiparametric methods. *J. Appl. Econometrics,* **4,** supplement issue, 61-86.

[11] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.,* **40**(2),633-643.

[12] Koul, Hira L., Ni, P. (2004). Minimum distance regression model checking, *J. Stat. Plann. Inference,* **119**(1), 109-141.

[13] Lee, L. F. (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.

[14] Lewbel, A. and Linton, O. B. (2002). Nonparametric censored and truncated regression. *Econometrica,* **70,** 765-779.

[15] Lichtenberg, E., Shapiro, L. K. (1997). Agriculture and nitrate concentration in Maryland community water system wells. *J. Environ. Qual.,* **26**(1), 145-153.

[16] Lin, T. F. and Schmidt, P. (1984). A test of the tobit specification against an alternative suggested by Cragg. *The Review of Economics and Statistics,* **66**(1), 174-177.

[17] McConnel, C. E., Zetzman, M. R. (1993). Urban/rural differences in health service utilization by elderly persons in the United States. *J. Rural Health,* **9**(4), 270-280.

[18] Nelson, F. D. (1981). A test for misspecification in the censored normal model. *Econometrica,* **49**(5), 1317-1330.

[19] Olsen, R. J. (1980). Approximating a truncated normal regression with the method of moments. *Econometrica,* **48**(5), 1099-1106.

[20] Powell, J. L. (1984). Least abosulte deviations estimation for the censored regression model. *J. Econometrics,* **25**(3),303-325.

[21] Song, W. X. (2011). Distribution-free test in Toibt mean regression model. *J. Stat. Plann. Inference,* **141**(8), 2891-2901.

[22] Song, W. X. (2011). Empirical $L_2$-Distance Lack-of-Fit Tests For Tobit Regression Models. Submitted to JMVA.

[23] Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica,* **26**(1), 24-36.

[24] Thomas A.Mroz (1987) The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica,* 765-799.

[25] Wang, L. (2007). A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model. *Stat. and Prob. Letters,* **77**(10), 1034-1042.

[26] White, H. (1981). Consequences and detection of misspecified nonlinear models. JASA, **76**(374), 419-433.

[27] White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrics,* **50**(1), 1-25.

[28] Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics,* **75**(2), 263-289.

[29] Zhou, X. B. (2007). Semiparametric and Nonparametric Estimation of Toibt Models. PhD thesis, Department of Economics, Hong Kong University of Science and Technology.

# Appendix A

# R-Programs

The first part of our simulation is on Tobit linear regression with one predictor, composite hypothesis. Note that this is the code when the $\sigma_\varepsilon$ is unknown and estimated, but we also did simulations when $\sigma_\varepsilon$ is known.

```
library(VGAM)


set.seed(987654)
table1=matrix(c(1:25),nrow=5,ncol=5)
colnames(table1)<-c("n=100","n=300","n=500","n=800","n=1000")
rownames(table1)<-c("m0=0","m1=0.1","m2=0.2","m3=0.3","m4=0.5")
m<-c(100,300,500,800,1000)
r<-c(0,0.1,0.2,0.3,0.5)
 for (j in 1:length(m)){
 for (k in 1:length(r)){
 freq=0;
 for(i in seq(1000))
 {
  # Generating Data
  n=m[j]
  sx=1
```

```
se=1

a=1

b=1

x=rnorm(n,0,sx)

e=rnorm(n,0,se)

ystar=a+b*x+r[k]*x^2+e

y=pmax(ystar,0)


# Estimation of regression parameters

fit=vglm(y~x, tobit(Lower=0, Upper=Inf))

a=fit@coefficients[1]

b=fit@coefficients[3]

sig=exp(fit@coefficients[2])


# Test Statistics

res=y-((a+b*x)*pnorm((a+b*x)/sig)+dnorm((a+b*x)/sig)*sig)

res2=res^2

res4=res^4

h=n^(-1/5)

K=function(u){dnorm(u)}

xdif=kronecker(x,x,"-")

A1=matrix(K(xdif/h)/h,nrow=n)

A2=matrix((K(xdif/h))^2/h,nrow=n)

Vn=(t(res)%*%A1%*%res-sum(diag(A1)*res2))/(n*(n-1))

Sn=2*(t(res2)%*%A2%*%res2-sum(diag(A2)*res4))/(n*(n-1))

Tn=n*sqrt(h)*Vn/sqrt(Sn)

freq=freq+(abs(Tn)>=1.96)
```

```
  }
 table1[k,j]=freq/1000
 }
}
table1
```

The second part of our simulation is on Tobit linear regression with two predictors. Note that this is the code when the two predictors are independent. We also did simulations when the two predictors are not independent.

```
library(VGAM)
library(maxLik)
library(miscTools)
library(censReg)
library(mvtnorm)

set.seed(98765432)
table1=matrix(c(1:25),nrow=5,ncol=5)
colnames(table1)<-c("n=100","n=300","n=500","n=800","n=1000")
rownames(table1)<-c("m0=0","m1=0.1","m2=0.2","m3=0.3","m4=0.5")
m<-c(100,300,500,800,1000)
r<-c(0,0.1,0.2,0.3,0.5)
 for (j in 1:length(m)){
 for (k in 1:length(r)){
 freq=0;
 for(i in seq(1000))
 {

 #Generating Data
```

```
n=m[j]
sx=matrix(c(1,0,0,1),2)
se=1
a=1
b1=1
b2=1
xno=rmvnorm(n,mean=c(0,0),sx)
x1=xno[ ,1]
x2=xno[ ,2]
e=rnorm(n,0,se)
ystar=a+b1*x1+b2*x2+r[k]*(x1^2+x2^2)+e
y=pmax(ystar,0)
data=data.frame(cbind(y,x1,x2))

#Estimation of regression parameters
estimation <- censReg( y ~ x1 + x2, data = data )
a=summary(estimation)$estimate[1]
b1=summary(estimation)$estimate[2]
b2=summary(estimation)$estimate[3]
sigma=exp(summary(estimation)$estimate[4])

#Test Statistics
res=y-((a+b1*x1+b2*x2)*pnorm((a+b1*x1+b2*x2)/sigma)+dnorm((a+b1*x1+b2*x2)/sigma)*sigma)
res2=res^2
res4=res^4
h=n^(-1/7)
K=function(u,v){dnorm(u)*dnorm(v)}
```

```
 x1dif=kronecker(x1,rep(1,n))-kronecker(rep(1,n),x1)
 x2dif=kronecker(x2,rep(1,n))-kronecker(rep(1,n),x2)
 A1=matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n)
 A2=matrix((K(x1dif/h,x2dif/h))^2/h^2,nrow=n)
 Vn=(t(res)%*%A1%*%res-sum(diag(A1)*res2))/(n*(n-1))
 Sn=2*(t(res2)%*%A2%*%res2-sum(diag(A2)*res4))/(n*(n-1))
 Tn=n*h*Vn/sqrt(Sn)
 freq=freq+(abs(Tn)>=1.96)
 }
 table1[k,j]=freq/1000
 }
}
table1
```

The last part of our simulations is Tobit linear regression, simple hypotheses. The codes below are the one predictor case and the two predictor case.

(a) One Predictor Case:

```
library(VGAM)
set.seed(987654)

table1=matrix(c(1:25),nrow=5,ncol=5)
colnames(table1)<-c("n=100","n=300","n=500","n=800","n=1000")
rownames(table1)<-c("m0=0","m1=0.1","m2=0.2","m3=0.3","m4=0.5")
m<-c(100,300,500,800,1000)
r<-c(0,0.1,0.2,0.3,0.5)
 for (j in 1:length(m)){
 for (k in 1:length(r)){
 freq=0;
```

```
for(i in seq(1000))
{

# Generating Data
n=m[j]
sx=1
se=1
a=1
b=1
x=rnorm(n,0,sx)
e=rnorm(n,0,se)
ystar=a+b*x+r[k]*x^2+e
y=pmax(ystar,0)

# Test Statistics
res=y-((a+b*x)*pnorm((a+b*x)/se)+dnorm((a+b*x)/se)*se)
res2=res^2
res4=res^4
h=n^(-1/5)
K=function(u){dnorm(u)}
xdif=kronecker(x,x,"-")
A1=matrix(K(xdif/h)/h,nrow=n)
A2=matrix((K(xdif/h))^2/h,nrow=n)
Vn=(t(res)%*%A1%*%res-sum(diag(A1)*res2))/(n*(n-1))
Sn=2*(t(res2)%*%A2%*%res2-sum(diag(A2)*res4))/(n*(n-1))
Tn=n*sqrt(h)*Vn/sqrt(Sn)
freq=freq+(abs(Tn)>=1.96)
```

```
 }
 table1[k,j]=freq/1000
 }
}
table1
```

(b)Two Predictor Case:

```
library(VGAM)
library(maxLik)
library(miscTools)
library(censReg)
library(mvtnorm)
set.seed(987654)

table1=matrix(c(1:25),nrow=5,ncol=5)
colnames(table1)<-c("n=100","n=300","n=500","n=800","n=1000")
rownames(table1)<-c("m0=0","m1=0.1","m2=0.2","m3=0.3","m4=0.5")
m<-c(100,300,500,800,1000)
r<-c(0,0.1,0.2,0.3,0.5)
for (j in 1:length(m)){
for (k in 1:length(r)){
 freq=0;
 for(i in seq(1000))
 {
 # Generating Data
 n=m[j]
 sx=matrix(c(1,0,0,1),2)
 se=1
```

```
a=1
b1=1
b2=1
xno=rmvnorm(n,mean=c(0,0),sx)
x1=xno[ ,1]
x2=xno[ ,2]
e=rnorm(n,0,se)
ystar=a+b1*x1+b2*x2+r[k]*(x1^2+x2^2)+e
y=pmax(ystar,0)

#Test Statistics
res=y-((a+b1*x1+b2*x2)*pnorm((a+b1*x1+b2*x2)/se)+dnorm((a+b1*x1+b2*x2)/se)*se)
res2=res^2
res4=res^4
h=n^(-1/7)
K=function(u,v){dnorm(u)*dnorm(v)}
x1dif=kronecker(x1,rep(1,n))-kronecker(rep(1,n),x1)
x2dif=kronecker(x2,rep(1,n))-kronecker(rep(1,n),x2)
A1=matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n)
A2=matrix((K(x1dif/h,x2dif/h))^2/h^2,nrow=n)
Vn=(t(res)%*%A1%*%res-sum(diag(A1)*res2))/(n*(n-1))
Sn=2*(t(res2)%*%A2%*%res2-sum(diag(A2)*res4))/(n*(n-1))
Tn=n*h*Vn/sqrt(Sn)
freq=freq+(abs(Tn)>=1.96)
}
table1[k,j]=freq/1000
}
```

```
}
table1
```