EMPIRICAL MINIMUM DISTANCE LACK-OF-FIT TESTS FOR TOBIT REGRESSION MODELS

by

YI ZHANG

B.S., North Dakota State University, 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
2011

Approved by:

Major Professor
Weixing Song

Copyright

Yi Zhang

2011

Abstract

The purpose of this report is to propose and evaluate two lack-of-fit test procedures to check the adequacy of the regression functional forms in the standard Tobit regression models. It is shown that testing the null hypothesis for the standard Tobit regression models amounts testing a new equivalent null hypothesis of the classic regression models. Both procedures are constructed based on the empirical variants of a minimum distance, which measures the squared difference between a nonparametric estimator and a parametric estimator of the regression functions fitted under the null hypothesis for the new regression models. The asymptotic null distributions of the test statistics are investigated, as well as the power for some fixed alternatives and some local hypotheses. Simulation studies are conducted to assess the finite sample power performance and the robustness of the tests. Comparisons between these two test procedures are also made.

Table of Contents

List of Tables

Acknowledgments

First of all, I am heartily thankful to my major professor, Dr. Weixing Song, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

Then I would like to thank Dr. John Boyer and Dr. Juan Du, for being the committee members for this report, making comments, and sharing ideas.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the report.

Dedication

I dedicate this master's report to my family on the other side of the earth - Dad, Mom, Grandpa and Grandma. Without their support, encouragement, understanding, and most of all love, the completion of this work would not be possible.

Also, I dedicate this report to the people at Department of Statistics at Kansas State University. They made my two years at Kansas State University so unforgettable.

Chapter 1

Introduction

> Nothing is absolutely the limit of *nothingness*. It's the *lowest* you can go. It's the end of the line. How can something be less than nothing? If there were something that was less than nothing then nothing would be nothing, it would be something - even though it's just a very little bit of something. But if nothing is *nothing*, then nothing has nothing that's less than *it* is.
>
> E.B White, Charlotte's Web
> (New york: Harper, 1952) P.28.

1.1 The Tobit Regression Model

James Tobin (1958) first suggested the regression models in which the range of the dependent variable had constraints. Tobin named this model the model of *limited dependent variables*, and also demonstrated an example in economics of a household's expenditure on durable goods using a first order regression model which specifically took account of the fact that the value of the expenditure could not be negative. Later on in 1964, the name *Tobit* was coined by the econometrician and economist Arthur Stanley Goldberger inspired by the similarity to and difference from of the *Probit models*. Nowadays, the Tobit regression model is a frequently used tool for modeling censored or truncated variables in not only econometrics research, but also other disciplines such as biometrics and engineering, where the expected value of the dependent variable is its limiting value. It's possible for this expected value to have both positive and negative deviations.

To be specific, let

$Y =$ a household's expenditure on a durable good,

$y_0 =$ the price of the cheapest available durable good,

$Z =$ all the other expenditure,

$X =$ income.

Tobin (1958) considered an utility maximization model in which a household was assumed to maximize utility $U(Y, Z)$ with the budget constraint $Y + Z \leq X$, and the boundary constraint $Y \geq y_0$. Suppose Y^* is the solution of the maximization of U subject to $Y + Z \leq X$ but ignoring the other constraint, assume that $Y^* = m(X, \theta) = \alpha + \beta X + \varepsilon$, where ε could be interpreted as the effects of all other unobservable factors, and $\theta = (\alpha, \beta)'$, then the solution Y to the original problem was depicted by the following, and is termed the Tobit regression model,

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

Without loss of generality, we choose $y_0 = 0$ for this report. In the original model proposed by Tobin (1958), the error term ε was normally distributed, and the threshold y_0 was chosen to be the same for all households.

The *standard Tobit regression model* was one of the five types of Tobit regression models defined by Amemiya (1984). It was expressed as the following:

$$Y_i^* = X_i'\beta + \varepsilon_i, \quad i = 1, 2, \ldots, n,$$
$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0, \\ 0 & \text{otherwise,} \end{cases}$$

where ε_i's were assumed to be i.i.d. copies from $N(0, \sigma^2)$. Note that $\{(X_i, Y_i)\}_{i=1}^n$ were observable, while Y_i^*'s were not if $Y_i^* < 0$. In this report, we mainly work on the standard Tobit regression model.

1.2 The Research Objective and Literature Review

Under the assumption that the regression function $m(X)$ is linear, the existing work on the standard Tobit regression model mainly focuses on the estimation of the unknown regression coefficient θ. Under the normality assumption of the error term ε, Amemiya (1973) and Heckman (1976,1979) proposed consistent estimators for θ, but these estimators were not consistent if the normality assumption failed. A robust estimator of θ was proposed by Powell (1984) based on the least absolute deviations. See Amemiya (1984) for a comprehensive discussion on the estimation issue related to Tobit regression models. In statistical research, the predetermined parametric form of the regression function is formulated either based on some empirical evidence or simply for the sake of mathematical convenience. However, 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 for the parameters and cause biased prediction of the survival time. See Horowitz and Neumann (1989). Therefore, it is theoretically important and practically significant to develop some formal numeric hypothesis tests to check the adequacy of the selected regression functions.

Testing the functional form of the Tobit regression models began in the early 1980's. Among others, Nelson's (1981) test was a general specification test which compared the restricted estimates with the unrestricted estimates of various moments of the dependent variable, and this test was not limited to any specific alternatives. It was the same for Olsen's (1980) test which was derived for the comparisons between the actual and the predicted numbers of zero observations. Ruud's (1982) suggestion followed the same thread and his test was conducted to check the level of significance of the difference between the Tobit and Probit estimates. Lee's (1981) procedure was a test of normality against the alternative of a more general member of the Pearson family. Lin and Schmidt (1984) proposed a relatively simple test for the hypothesis that the Tobit model was correctly specified, against the alternative that different sets of parameters determined the probability of a zero observation and the density of the nonzero observations.

Wang (2007) proposed a simple nonparametric test for checking the nonlinearity in Tobit median regression models, where the median of the random error was assumed to be zero and $y_0 = 0$. Under the null hypothesis of $m(x) = \beta_0 + \beta_1 x$, for a random sample $\{(X_i, Y_i)\}_{i=1}^n$, $\{\varepsilon_i\}_{i=1}^n = I(Y_i \le \max(0, \beta_0 + \beta_1 X_i)) - 1/2$ were i.i.d. Bernoulli random variables with mean zero and variance $1/4$. Replacing each ε_i with the estimated value $\{\hat{\varepsilon}_i\}_{i=1}^n = I(Y_i \le \max(0, \hat{\beta}_0 + \hat{\beta}_1 X_i)) - 1/2$, where $\hat{\beta}_0$ and $\hat{\beta}_1$ were $\sqrt{n}$-consistent estimators of β_0 and β_1, respectively, Wang (2007) considered each distinct covariate as a "category", constructed a local window around each covariate value, and also created a balanced one-way table. The test statistic could be viewed as a generalization of the classical F-test statistic in the context of analysis of variance. More advanced than the other existing methods in the literature, Wang's (2007) test had the advantage of allowing the alternative to be any smooth function. Furthermore, this test did not require any knowledge of the parametric distribution of the random error. However, the window width selection was an issue that was not solved in Wang (2007). Most recently, Song (2011) developed a lack-of-fit test procedure under the assumption that the error term ε had mean zero. The proposed test was from the Khamaladze type transformation of a certain marked residual process, which converged weakly to a time-transformed Brownian motion in a uniform metric scale. Consequently, any test based on a continuous function of this process was asymptotically distribution free, and could be implemented for at least moderate to large samples without resorting to a resampling method. A more general null hypothesis, not limited to linear functions, was taken into consideration. The main advantages of Song's (2011) test, in the comparision with Wang's (2007) test, were the following: Song's (2011) test did not require the window width selection or any other smoothing parameter; it could be used to test any parametric regression functions not limited to those linear functions in Wang's (2007) paper; the computation of the test statistic was less complex and much faster.

However, Song's (2011) test procedure was restricted to one-dimensional predictive variable X. Here, as in the results in such works as Hárdle and Mammen (1993), Zheng (1996), Koul and Ni (2004) in the classic regression models, we attempt to develop some test procedures to check the adequacy of a specific regression functional form which may not limited to linearity, and can be executed on multidimensional predictors as well.

Chapter 2

Minimum Distance Lack-of-Fit Tests in the Tobit Regression Model

The relationship between a predictor X(possibly multidimensional), and a response Y is often studied in regression analysis, based on a fully observed data set $\{(X_i, Y_i)\}_{i=1}^n$. In this chapter, we will formally state the hypotheses in the regression functional form in the standard Tobit model. Two test statistics will be constructed to check the validity of specified regression functional forms. Also, asymptotic distributions of the test statistics will be investigated, and some discussion on the consistency and local power of the tests will be provided.

Consider the following classic regression model $Y = m(X) + \varepsilon$. We would like to test the following hypotheses:

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

where $m(x, \theta)$ is of a parametric form with parameter θ. Many test procedures have been conducted for these model specification hypotheses. An extensive introduction to this topic can be found in Hart (1997) and the references therein. In particular, Koul and Ni (2004) used the minimum distance method to construct lack-of-fit tests of a parametric regression model in this specific set up. In a finite sample comparison of these tests among some other existing tests, they noted that a member of this class asymptotically preserved the level of the test, and had relatively high power against some alternatives. Our report extends Koul and Ni's method to the standard Tobit regression model.

To be specific, Koul and Ni (2004) considered the following tests of H_0 in (2.1) where the design was random and observable, and the errors were heteroscedastic. For any kernel density K, let $K_h(x) := K(x/h)/h^d$, $h > 0$, $x \in \mathbb{R}^d$. Define, $\hat{f}_w(x) := \frac{1}{n}\sum_{j=1}^n K_w^*(x - X_j)$, $w = w_n \sim (\log n/n)^{1/(d+4)}$,

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

and

$$\hat{\theta}_n := \mathrm{argmin}_{\theta\in\Theta} T_n(\theta),$$

where K and K^* are kernel density functions (possibly different), $h = h_n$ and $w = w_n$ are the window widths depending on the sample size n, Θ is a compact subset of $\mathbb{R}^d$, and G was a σ-finite measure on the compact subset C of $\mathbb{R}^d$. They proved the consistency and the asymptotic normality of this estimator, and that the asymptotic null distribution of $D_n := nh_n^{d/2}(T_n(\hat{\theta}_n) - \hat{C}_n)/\hat{\Gamma}_n^{1/2}$ is standard normal, where

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

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

The test based on D_n is preferable to the tests developed by Härdle and Mammen (1993), and Zheng (1996). Unlike the concepts from other related papers, Koul and Ni's does not require the null regression function to be twice continuously differentiable in the parameter vectors, nor do their proofs require any rate of uniform consistency of nonparametric regression function estimators. Moreover, the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta)$ and D_n are made feasible by utilizing different window widths for the estimation of the numerator and denominator in the nonparametric regression function. As a consequence of the asymptotic normality result, it was unnecessary to use any resampling techniques to implement these tests for large samples.

Thus, these findings motivate us to form lack-of-fit tests in the standard Tobit regression models based on the minimized distances mentioned above, and modify their procedures because the response Y^* in the standard Tobit regression models is not always observable. Also, the implementations of T_n, $\hat{C}_n$ and $\hat{\Gamma}_n$ are not feasible because of the integration calculations involved. More specifically, these integrations usually do not have tractable forms, so approximating these integrations forces us utilize some advanced numerical integration techniques. In addition, the kernel estimator $\hat{f}_w(x)$ usually takes small values at the boundary of x, which makes the computation very unstable. In the next section, an empirical minimum distance test procedure is presented which overcomes all the issues mentioned above.

2.1 Empirical MD Test Statistics and Assumptions

Since Y_i^*'s in $\{(X_i, Y_i^*)\}_{i=1}^n$ are not always observable with the constraint of nonnegative values, one has to construct a test statistic based on the observations $\{(X_i, Y_i)\}_{i=1}^n$. Therefore we would like to relate the unobserved values to the observed ones. A natural way to find such relationships is to consider the regression of Y, or some other quantities related to Y, against X. Thus, denote the density function of ε by f_ε, and we define

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

Then we are able to derive the following equations easily:

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

and

$$E(I(Y = 0)|X = x) = F_\varepsilon(-m(x)) \text{ or } 1 - Q_0(-m(x)). \tag{2.3}$$

Thus, we may consider the following two regression models based on (2.2) and (2.3), respectively,

$$R1: \quad Y = m(x)Q_0(-m(x)) + Q_1(-m(x)) + \xi = g_1(x) + \xi, \tag{2.4}$$

$$R2: \quad I(Y = 0) = F_\varepsilon(-m(x)) + \eta = g_2(x) + \eta, \tag{2.5}$$

where both ξ and η are uncorrelated with X. As a function of $m(x)$, $g_1(x)$ and $g_2(x)$ are strictly monotonic provided that $F_\varepsilon(x)$ is strictly increasing. This can be easily verified by checking the derivatives of $g_1(x)$, $g_2(x)$, as functions of $m(x)$. Therefore, to test $H_0 : m(x) = m(x, \theta)$, it is equivalent to test

$$H_0 : g_1(x) = g_1(x, \theta) \quad \text{for some} \theta \in \Theta, \qquad \text{versus} \qquad H_1 : H_0 \text{ is not true} \tag{2.6}$$

for the regression model R1, or,

$$H_0 : g_2(x) = g_2(x, \theta) \quad \text{for some} \theta \in \Theta, \qquad \text{versus} \qquad H_1 : H_0 \text{ is not true} \tag{2.7}$$

for the regression model R2, where $g_1(x, \theta)$ and $g_2(x, \theta)$ are expressed in the same way as $g_1(x)$ and $g_2(x)$ replacing $m(x)$ by $m(x, \theta)$.

Hence, let K be a symmetric density function and h be a sequence of positive numbers depending on the sample size n. As before, denote $K_h(x) = h^{-d}K(x/h)$. The univariate kernel estimations of the regression functions $g_1(x)$ and $g_2(x)$ can be written as:

$$\hat{g}_1(x) = \frac{\sum_{i=1}^n K_h(x - X_i)Y_i}{\sum_{i=1}^n K_h(x - X_i)}, \quad \hat{g}_2(x) = \frac{\sum_{i=1}^n K_h(x - X_i)I(Y_i = 0)}{\sum_{i=1}^n K_h(x - X_i)}.$$

Let $W(x)$ be a weight function that may depend on the sample. Then the following minimum distances D_{1n}, D_{2n} can be used for testing the hypotheses (2.6) and (2.7), respectively, where

$$D_{kn} = = \int \left[\hat{g}_k(x) - g_k(x, \hat{\theta}_n)\right]^2 dW(x), k = 1, 2,$$

which is similar to Härdle and Mammen's (1993) test statistic. As noted in Koul and Ni (2004), the nonparametric estimator $\hat{g}_k(x)$ has a non-negligible bias, and the lack-fit-test statistic D_{kn} may not have a desirable asymptotic null distribution. Therefore, by means of mimicking Koul and Ni (2004)'s procedure, we also make the following modification to test

(2.6) and (2.7), respectively:

$$S_{1n} = = \int \left[\frac{\sum_{j=1}^{n} K_h(x - X_j)(Y_j - g_1(X_j, \hat{\theta}_n))}{\sum_{j=1}^{n} K_h(x - X_j)} \right]^2 dG(x), \tag{2.8}$$

and

$$S_{2n} = = \int \left[\frac{\sum_{j=1}^{n} K_h(x - X_j)(I(Y_j = 0) - g_2(X_j, \hat{\theta}_n))}{\sum_{j=1}^{n} K_h(x - X_j)} \right]^2 dG(x). \tag{2.9}$$

In addition, we would like to avoid the possible instability incurred by the small values of the kernel density estimator $\hat{f}_X$ in the denominators, and reduce the complexity which may arise from the potential intractable integration. Here, we successfully tackle this task by choosing the weight function $G(x)$ such that $dG(x) = \hat{f}_X^2(x)dF_n(x)$.[1] Accordingly, S_{1n} and S_{2n} then become

$$D_{1n} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{n} \sum_{j=1}^{n} K_h(X_i - X_j)(Y_j - g_1(X_j, \hat{\theta}_n)) \right]^2, \tag{2.10}$$

and

$$D_{2n} = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{1}{n} \sum_{j=1}^{n} K_h(X_i - X_j)(I(Y_j = 0) - g_2(X_j, \hat{\theta}_n)) \right]^2, \tag{2.11}$$

Now, we desire an appropriate standardization for $D_{kn}, k = 1, 2$ to determine the asymptotic distribution under the null hypothesis. Let T_{kn} be the standardized test statistic. It will be proved later that $T_{kn} = nh^{d/2}\hat{\Gamma}_{kn}^{-1/2}(D_{kn} - \hat{C}_{kn})$, where

$$\hat{C}_{kn} = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h^2(X_i - X_j)e_{ki}^2,$$

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

and $e_{1i} = Y_i - g_1(X_i, \hat{\theta}_n)$, $e_{2i} = I(Y_i = 0) - g_2(X_i, \hat{\theta}_n)$. Based on our knowledge, larger values of $|T_{kn}|$ should be the evidence to reject H_0 in (2.6) and (2.7). The thresholds of the

[1] F_n denotes the empirical CDF of the sample from X.

tests will be determined by the asymptotic distribution of T_{kn} under the null hypothesis, and details will be discussed in the next section.

The following is a list of assumptions needed so as to derive the asymptotic results of the test statistics:

(C1). The random error ε satisfies $E(\varepsilon) = 0$, and $E(\varepsilon^4) < \infty$.

(C2). $\tau^2(x) = E[(Y - g_k(X))^2 | X = x]$, and $\sigma^4(x) = E[(Y - g_k(X))^4 | X = x]$ are uniformly continuous in x for $k = 1, 2$.

(C3). The density function of X is uniformly continuous and bounded.

(C4). $g_k(x, \theta), k = 1, 2$ are continuous and differentiable with respect to x and $E(\|g_k'(X)\|^4) < \infty$. Then for any $\sqrt{n}$-consistent estimator $\hat{\theta}_n$ of θ,

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

(C5). The kernel density function $K(x)$ is continuous and symmetric around 0.

(C6). $h \to 0$, $nh^{2d} \to \infty$ as $n \to \infty$.

(C7). $h \to 0$, $nh^{5d/2} \to \infty$ as $n \to \infty$.

2.2 Main Results

The following theorem states the asymptotic null distribution of the test statistics T_{kn}, $k = 1, 2$.

Theorem 2.2.1. *Suppose (C1)-(C6) hold. Under the null hypothesis H_0 in (2.6) and (2.7),*

$$T_{kn} = nh^{d/2} \hat{\Gamma}_{kn}^{-1/2} (D_{kn} - \hat{C}_{kn}) \Longrightarrow N(0, 1), \quad k = 1, 2.$$

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

It is logical that consistency is a basic and reasonable requirement for any test procedure. Ideally, for some fixed alternatives, a consistent test should have power that tends to be one as the sample size approaches ∞.

Now, let's consider a class of fixed alternative hypotheses:

$H_a : E(Y^*|X = x) = m(x)$ such that $E(m^2(X)) < \infty$ and $m(x) \not\equiv m(x, \theta)$ for any θ.

Under the null hypothesis, we assume that estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for the true parameter θ_0. A question of interest in its own right come to our minds: would this estimator still have the similar property under H_a? In the classic regression set up, Jennrich (1969) and White (1981, 1982) showed that the nonlinear least squares estimator converged in probability under some mild regularity conditions, and was asymptotically normal even in the presence of model misspecification. In the rest of the report, we simply assume that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$ under the alternative H_a for some $\theta_a \in \mathbb{R}^p$, but we will not justify this assumption rigorously here.

Theorem 2.2.2. *Suppose all the conditions in Theorem 2.2.1 hold with θ_0 replaced by θ_a, $\int [g_k(x) - g_k(x; \theta_a)]^2 f_X^3(x)dx > 0$. For any $0 < \alpha < 1$, the test that rejects H_0 in (2.3) and (2.6) whenever $|T_{nk}| > z_{1-\alpha/2}$ is consistent for $k = 1, 2$.*

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

$$H_{\text{Loc}} : m(x) = m(x, \theta_0) + \delta(x)/\sqrt{nh^{d/2}}. \tag{2.12}$$

It is reasonable that we assume the estimators used in the test statistic satisfies that $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$.

Surprisingly, although we can show that the proposed tests have non-trivial power for the local alternatives, the conditions imposed on the bandwidth are still different from those of the asymptotic distributions under H_{Loc}. We must have the properties that $nh^{2d} \to \infty$ for T_{1n}, and $nh^{5d/2} \to \infty$ for T_{2n}. This indicates that more work on smoothing should be done if the response variables in the transformed regression models are dichotomous. In fact, define

$$\mu_1 = E[\delta^2(X)Q_0^2(-m(X,\theta_0))f_X^2(X)], \quad \mu_2 = E[\delta^2(X)f_\varepsilon^2(-m(X,\theta_0))f_X^2(X)].$$

Then we obtain the following theorem:

Theorem 2.2.3. *Suppose all the conditions in Theorem 2.2.1 hold, and the bandwidth h satisfies (C6) for T_{1n}, and (C7) for T_{2n} . Under H_{Loc} in (2.12), $T_{kn} \Longrightarrow N(\mu_k, 1), k = 1, 2.$*[2]

2.3 Proofs of Main Results

This section contains the proofs of all the main results proposed in Section 2.1 and 2.2.

Here, we will start with the proofs of (2.2) and (2.3) with the general case of y_0.

Proof of (2.2):

$$\begin{aligned}
g_1(x) &= E\left[Y \mid X = x\right] = E\left[YI\left(Y > y_0\right) \mid X = x\right] + E\left[YI\left(Y = y_0\right) \mid X = x\right] \\
&= E\left[Y^*I\left(Y^* > y_0\right) \mid X = x\right] + E\left[y_0 I\left(Y^* \le y_0\right) \mid X = x\right] \\
&= E\left[\left(m\left(x\right) + \varepsilon\right) I\left(m\left(x\right) + \varepsilon > y_0\right)\right] + y_0 E\left[I\left(m\left(x\right) + \varepsilon \le y_0\right)\right] \\
&= m\left(x\right) E\left[I\left(\varepsilon > y_0 - m\left(x\right)\right)\right] + E\left[\varepsilon I\left(\varepsilon > y_0 - m\left(x\right)\right)\right] + y_0 E\left[I\left(\varepsilon \le y_0 - m\left(x\right)\right) \mid x\right] \\
&= m\left(x\right) P\left(\varepsilon > y_0 - m\left(x\right)\right) + E\left[\varepsilon I\left(\varepsilon > y_0 - m\left(x\right)\right)\right] + y_0 P\left(\varepsilon \le y_0 - m\left(x\right)\right) \\
&= m\left(x\right) \int_{y_0 - m(x)}^{\infty} f_\varepsilon\left(u\right) du + \int_{y_0 - m(x)}^{\infty} u f_\varepsilon\left(u\right) du + y_0 \int_{-\infty}^{y_0 - m(x)} f_\varepsilon\left(u\right) du \\
&= m\left(x\right) Q_0\left(y_0 - m\left(x\right)\right) + Q_1\left(y_0 - m\left(x\right)\right) + y_0 F_\varepsilon\left(y_0 - m\left(x\right)\right).
\end{aligned}$$

■

Proof of (2.3):

$$\begin{aligned}
g_2(x) &= E\left(I\left[Y = y_0\right] \mid X = x\right) = P\left(Y = y_0 \mid X = x\right) \\
&= P\left(Y^* \le y_0 \mid X = x\right) = P\left(m\left(x\right) + \varepsilon \le y_0\right) \\
&= P\left(\varepsilon \le y_0 - m\left(x\right)\right) = F_\varepsilon\left(-m\left(x\right)\right) + y_0).
\end{aligned}$$

[2] The notation "$\Longrightarrow$" stands for convergence in distribution in this report.

■

Next, we are going to demonstrate the proof of Theorem 2.2.1. For the sake of simplicity, throughout this chapter, we will use $g(x), g(x,\theta)$ instead of $g_k(x), g_k(x,\theta)$ for the regression functions. The actual case under discussion will be clear from the context. Other similar quantities should be understood in the same way.

Let $\xi_i = Y_i - g_1(X_i, \theta_0)$ or $I(Y_i = 0) - g_2(X_i, \theta_0)$,

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

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

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

The proof of Theorem 2.2.1 is facilitated by the following two lemmas:

Lemma 2.3.1. *Let Z_i, $i = 1, 2, \ldots, n$ be i.i.d. random vectors, and let*

$$U_n = \sum_{1\le i<j\le n} H_n(Z_i, Z_j), \quad M_n(x,y) = E[H_n(Z_1, x)H_n(Z_1, y)],$$

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

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

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

Lemma 2.3.2. *Suppose the conditions (C1)-(C3), (C5) and (C6) hold, then under the null hypotheses H_0 in (2.6) and (2.7) $nh^{d/2}(\tilde{D}_n - \tilde{C}_n) \Longrightarrow N(0, \Gamma)$.*

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

$$A_{n1} = \frac{1}{n^2} \sum_{j \neq k} \left[\frac{1}{n} \sum_{i \neq j,k} K_h(X_i - X_j) K_h(X_i - X_k) \right] \xi_j \xi_k$$
$$A_{n2} = \frac{2K_h(0)}{n^3} \sum_{j \neq k} K_h(X_j - X_k) \xi_j \xi_k.$$

Note that $E(A_{2n}) = 0$, and

$$\begin{aligned}
\mathrm{Var}(A_{2n}) &= \frac{4K_h^2(0)}{n^6} E \left[\sum_{j \neq k} K_h(X_j - X_k) \xi_j \xi_k \right]^2 \\
&= \frac{16K_h^2(0)}{n^6} \sum_{j<k} E[K_h^2(X_j - X_k)] \xi_j^2 \xi_k^2 \\
&= \frac{8K^2(0) n(n-1)}{n^6 h^{4d}} E[K^2 \left(\frac{X_1 - X_2}{h} \right)] \xi_1^2 \xi_2^2 \\
&= \frac{8K^2(0) n(n-1)}{n^6 h^{4d}} E[K^2 \left(\frac{X_1 - X_2}{h} \right)] \tau^2(X_1) \tau^2(X_2).
\end{aligned}$$

We obtain $\mathrm{Var}(A_{2n}) = O(1/n^4 h^{3d})$ by the continuity of $\tau^2(x)$. Thus,

$$nh^{d/2} A_{2n} = O_p(1/(n^3 h^{5d/2})) = o_p(1).$$

Also, let

$$H(X_j, X_k, h) = E[K_h(X_i - X_j) K_h(X_i - X_k) | X_j, X_k]$$

which is a symmetric function of X_j and X_k. Then A_{n1} can be written as the sum of the following two terms:

$$A_{n11} = \frac{1}{n^3} \sum_{j \neq k} \left[\sum_{i \neq j,k} K_h(X_i - X_j) K_h(X_i - X_k) - h^{-d} H(X_j, X_k, h) \right] \xi_j \xi_k,$$
$$A_{n12} = \frac{1}{n^2 h^d} \sum_{j \neq k} H(X_j, X_k, h) \xi_j \xi_k.$$

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

$E(A_{n11}) = 0$, and

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

In addition, $E[G^2(X_1, X_2, h)]\tau^2(X_1)\tau^2(X_2)$ equals

$$
\begin{aligned}
& E\left[\frac{1}{n}\sum_{i=3}^{n} K_h(X_i - X_1)K_h(X_i - X_2) - h^{-d}H(X_1, X_2, h)\right]^2 \tau^2(X_1)\tau^2(X_2) \\
&\leq E\left[\frac{1}{n}\sum_{i=3}^{n} K_h(X_i - X_1)K_h(X_i - X_2) - \frac{n-2}{nh^d}H(X_1, X_2, h) - \frac{2}{nh^d}H(X_1, X_2, h)\right]^2 \tau^2(X_1)\tau^2(X_2) \\
&\leq 2\left(\frac{n-2}{n}\right)^2 E\left[\frac{1}{n-2}\sum_{i=3}^{n} K_h(X_i - X_1)K_h(X_i - X_2) - h^{-d}H(X_1, X_2, h)\right]^2 \tau^2(X_1)\tau^2(X_2) \\
& + \frac{8}{n^2h^{2d}}E[H^2(X_1, X_2, h)]\tau^2(X_1)\tau^2(X_2). \qquad (2.13)
\end{aligned}
$$

Conditional on (X_1, X_2), and from the continuity of $f(x)$ and $\tau^2(x)$,

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

Thus, the first term in (2.13) has the order of $O_p(1/(nh^{2d}))$. Similarly, by the continuity of

$f(x)$ and $\tau^2(x)$, we have

$$
\begin{aligned}
&E[H^2(X_1, X_2, h)]\tau^2(X_1)\tau^2(X_2) \\
&= \iint \left(\int K(y)K(y + (x_1 - x_2)/h) f(x_1 + hy) dy \right)^2 \tau^2(x_1)\tau^2(x_2) f(x_1) f(x_2) dx_1 dx_2 \\
&= h^d \iint \left(\int K(y)K(y + u) f(x_1 + hy) dy \right)^2 \tau^2(x_1)\tau^2(x_1 - uh) f(x_1) f(x_1 - uh) dx_1 du \\
&= O(h^d).
\end{aligned}
$$

Hence, the second term in (2.13) is the order of $O_p(n^{-2}h^{-d})$. As a result

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

which implies

$$
nh^{d/2}A_{n1} = O_p\left(\frac{1}{\sqrt{nh^d}}\right) + O_p\left(\frac{1}{n}\right) = o_p(1).
$$

Eventually,

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

For the next step, denote $Z_j = (X_j', \xi_j)'$, and $H_n(Z_j, Z_k) = n^{-1}h^{-d/2}H(X_j, X_k, h)\xi_j\xi_k$, such that

$$
nh^{d/2}(\tilde{D}_n - \tilde{C}_n) = 2 \sum_{1 \leq j < k \leq n} H_n(Z_j, Z_k) + o_p(1). \tag{2.14}
$$

Note that $H_n(x, y)$ is symmetric, and $E[H_n(Z_1, Z_2)|Z_1] = 0$.Then for each n, by the continuity of $f(x)$ and $\tau^2(x)$,

$$
\begin{aligned}
&E[H_n^2(Z_j, Z_k)] \qquad (2.15) \\
&= \frac{1}{n^2h^d} \iint H^2(x, y, h)\tau^2(x)\tau^2(y) f(x) f(y) dx dy \\
&= \frac{1}{n^2h^d} \iint \left[\int K(u)K(u + (x - y)/h) f(x + hu) du \right]^2 \tau^2(x)\tau^2(y) f(x) f(y) dx dy \\
&= \frac{1}{n^2} \iint \left[\int K(u)K(u + v) f(x + hu) du \right]^2 \tau^2(x)\tau^2(x - vh) f(x) f(x - vh) dx dv \\
&= \frac{1}{n^2} \int \left[\int K(u)K(u + v) du \right]^2 dv \cdot \int \tau^4(x) f^4(x) dx + o\left(\frac{1}{n^2}\right) < \infty.
\end{aligned}
$$

Hence, in view of Lemma 2.3.1, it suffices to verify that

$$\frac{E[M_n^2(Z_1, Z_2)]}{\{E[H_n^2(Z_1, Z_2)]\}^2} \to 0, \qquad \frac{H_n^4(Z_1, Z_2)]}{n\{E[H_n^2(Z_1, Z_2)]\}^2} \to 0. \tag{2.16}$$

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

$$M_n(t,s) = E[H_n^2(Z,t)H_n^2(Z,s)] = \frac{1}{n^2h^d}E[H(X,t_1,h)H(X,s_1,h)]\xi^2 t_2 s_2.$$

Note that

$$\begin{aligned}
&E[H(X,t_1,h)H(X,s_1,h)]\xi^2 = E[H(X,t_1,h)H(X,s_1,h)]\tau^2(X) \\
&= \iiint K(x)K(y)K(x+(u-t_1)/h)K(y+(u-s_1)/h)f(u+hx)f(u+hy)\tau^2(u)f(u)dxdydu.
\end{aligned}$$

Change the variables $(u-t_1)/h$ into v. The integration can be written as $h^d B_h(t_1, s_1)$ with $B_h(t_1, s_1)$ expressed in the following integral

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

Therefore,

$$\begin{aligned}
E[M_n^2(Z_1,Z_2)] &= \frac{1}{n^4}E[B_h^2(X_1,X_2)]\xi_1^2\xi_2^2 = \frac{1}{n^4}E[B_h^2(X_1,X_2)]\tau^2(X_1)\tau^2(X_2) \\
&= \frac{1}{n^2}\int B_h^2(t_1,s_1)\tau^2(t_1)\tau^2(s_1)f(t_1)f(s_1)dt_1ds_2 \\
&= \frac{h^d}{n^2}\int B_h^2(t_1,t_1+wh)\tau^2(t_1)\tau^2(t_1+wh)f(t_1)f(t_1+wh)dt_1dw.
\end{aligned}$$

With the continuity property of $f(x)$ and $\tau^2(x)$, and the fact that

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

we obtain $E[M_n^2(Z_1, Z_2)] = O(h^d/n^4)$, which implies

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

Similarly,

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

Therefore,

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

This completes the proof of (2.16). Then by (2.15),

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

The theorem is then proved by Lemma 2.3.1. ■

This part of the report contains Lemma 2.3.3 and its proof. Also, we will use this lemma to prove Equations 2.17, 2.18, and 2.19.

Lemma 2.3.3. *Suppose (C1)-(C3), (C5) and (C6) hold. Then for a continuous function* $L(x) \in L_2(F)$

$$
\frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)L(X_j)\right]^2 = O_p(1), \tag{2.17}
$$

$$
\frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)\xi_j\right]^2 = O_p(1/nh^d), \tag{2.18}
$$

$$
\frac{1}{n}\sum_{i=1}^n \left[\frac{1}{n}\sum_{j=1}^n K_h(X_i - X_j)\xi_j L(X_i)\right] = O_p(1/\sqrt{n}). \tag{2.19}
$$

Proof: Note that

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

where the right hand side is bounded above by

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

The first term is the order of $O(1/(nh^d)^2)$. According to the assumption, it is $o_p(1)$. For any $j\neq 1$, denote

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

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

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

The continuity of $f(x)$ and $L(x)$ implies $E[J^2(X_1,h)]=O(1)$. While the first term above is further bounded by

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

This proves (2.17).

Similar argument implies that the left-hand side of (2.18) is bounded above by

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

Thus, (2.18) can be obtained by the continuity of $f(x)$ and $\tau^2(x)$.

To show (2.19), note that the left-hand side of (2.19) can be written as

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

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

$$\begin{aligned}
&\frac{n(n-1)}{2n^4}E[K_h(X_1-X_2)]\tau^2(X_2)L^2(X_1)\\
+&\frac{n(n-1)(n-2)}{n^4}E[K_h(X_1-X_2)K_h(X_3-X_2)]\tau^2(X_2)L(X_1)L(X_2).
\end{aligned}$$

The continuity of $f(x)$, $\tau^2(x)$ and $L(x)$ implies that the first term is the order of $O(1/n^2h^d)$, and the second term is $O(1/n)$. In summary, the left-hand side of (2.19) has the order of

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

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

Before we prove Theorem 2.2.2, we would like to state three lemmas with their proofs:

Lemma 2.3.4. *Suppose (C1)-(C6) hold. Under the null hypotheses H_0 in (2.6) and (2.7),*

$$nh^{d/2}(D_n-\tilde{D}_n)\to 0$$

in probability.

Proof: By subtracting and adding $g(X_j,\theta_0)$ from $Y_j - g(X_j,\hat{\theta}_n)$, D_n can be written as in this linear form $\tilde{D}_n + B_{n1} - 2B_{n2}$, where

$$\begin{aligned}
B_{n1}&=\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]\right]^2,\\
B_{n2}&=\frac{1}{n}\sum_{i=1}^{n}\left(\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)[g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)]\right]\left[\frac{1}{n}\sum_{j=1}^{n}K_h(X_i-X_j)\xi_j\right]\right).
\end{aligned}$$

Denote

$$\delta_{jn}=g(X_j,\hat{\theta}_n)-g(X_j,\theta_0)-(\hat{\theta}_n-\theta_0)'\dot{g}(X_j,\theta_0), \tag{2.20}$$

then B_{n1} can be shown bounded above by

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

From the assumption of the $\sqrt{n}$-consistency of $\hat{\theta}_n$, and together with Lemma 2.3.3, we obtain

$$B_{n1} = o_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n}\right),$$

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

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

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

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

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

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

The first factor is $o_p(1/\sqrt{n})$ according to the assumption. From Lemma 2.3.3, the second term is $O_p(1)$, and the third is $O_p(1/\sqrt{nh^d})$. Thus, $nh^{d/2}B_{n21} = o_p(1)$.

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

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

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

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

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

By (2.19), $nh^{d/2}B'_{n22} = O_p(1/n\sqrt{n}h^{d/2}) = o_p(1)$. Define $P_h(x) = E[K_h(x - X)\dot{g}(X, \theta_0)]$. B''_{n22} can be written as $O_p\left(1/\sqrt{n}\right)[A_{n1} + A_{n2}]$ with

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

and

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

Similar to the argument in the proof of (2.18), it can be shown that

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

Therefore, by the Cauchy-Schwarz inequality and (2.18), we have

$$A_{n1} = O_p\left(\frac{1}{nh^d}\right), \quad A_{n2} = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Hence

$$nh^{d/2} B_{n22''} = nh^{d/2} \left[O_p\left(\frac{1}{n\sqrt{nh^d}}\right) + O_p\left(\frac{1}{n}\right) \right] = o_p(1)$$

and

$$nh^{d/2} B_{n22} = o_p(1)$$

. Therefore, the lemma is proved. ■

Lemma 2.3.5. *Suppose all the conditions in Lemma 2.3.4 hold. Then under the null hypotheses H_0 in (2.6) and (2.7),*

$$nh^{d/2}(\hat{C}_n - \tilde{C}_n) \to 0$$

in probability.

Proof: First we claim that, for any nonnegative continuous function $L(x)$ such that $E[L(X)] < \infty$,

$$\frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} K_h^2(X_i - X_j) L(X_i) = O_p\left(\frac{1}{nh^d}\right). \tag{2.21}$$

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

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

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

By adding and subtracting $g(X_j;\theta_0)$ from $Y_j - g(X_j,\hat\theta_n)$, $\hat C_n - \tilde C_n$ can be written as the sum of the following two terms:

$$C_{n1} = \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n K_h^2(X_i - X_j)[g(X_j;\hat\theta_n) - g(X_j;\theta_0)]^2,$$
$$C_{n2} = -\frac{2}{n^3}\sum_{i=1}^n\sum_{j=1}^n K_h^2(X_i - X_j)\xi_j[g(X_j;\hat\theta_n) - g(X_j;\theta_0)].$$

Recall the notation δ_{jn} in (2.20), one can show that C_{n1} is bounded above by

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

Applying (2.21) with $L(x) = 1$ and $\|\dot g(x;\theta_0)\|^2$, together with our condition on δ_{in} and the $\sqrt{n}$-consistency of $\hat\theta_n$, this upper bound is the order of $O_p(1/n^2h^d)$. Hence, $nh^{d/2}C_{n1} = o_p(1)$.

Similarly, rewrite $-C_{n2}$ as in the following way

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

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

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

Similar to the proof of (2.17), we can show that

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

Therefore, by the assumption of δ_{jn} and the finiteness of $E(\xi^2)$, the first term is the order of $O_p(1/n^2h^d)$. Finally, similar to the proof of (2.19), it can be shown that

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

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

Lemma 2.3.6. *Suppose all the conditions in Lemma 2.3.4 hold. Then under the null hypotheses H_0 in (2.6) and (2.7),*

$$\hat{\Gamma}_n - \tilde{\Gamma}_n \to 0, \quad \tilde{\Gamma}_n \to \Gamma$$

in probability.

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

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

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

At this point, we state the following claims:

$$\frac{h^d}{n^4} \sum_{i \neq j} \left[\sum_{k=1}^{n} K_h(X_k - X_i) K_h(X_k - X_j) |\xi_i \xi_j| \right]^2 = O_p(1), \tag{2.22}$$

$$\frac{h^d}{n^4} \sum_{i \neq j} \left[\sum_{k=1}^{n} K_h(X_k - X_i) K_h(X_k - X_j) |\xi_i| \right]^2 = O_p(1), \tag{2.23}$$

$$\frac{h^d}{n^4} \sum_{i \neq j} \left[\sum_{k=1}^{n} K_h(X_k - X_i) K_h(X_k - X_j) \right]^2 = O_p(1). \tag{2.24}$$

For the sake of brevity, we only demonstrate the proof for (2.23), (2.22) and (2.24) can be similarly argued. Take the following expectation

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

By the continuity of $\tau^2(x)$ and $f(x)$, the first term on the right-hand side is $O(n^2/h^{3d})$. The second term equals

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

Therefore, the left-hand side of (2.23) has the order of

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

which is the desired result. Note that $E\|\dot g(X;\theta_0)\|^2 < \infty$ which implies $\|\hat\theta_n - \theta_0\| \cdot \max_{1\le i\le n}\|\dot g(X_i;\theta_0)\| = o_p(1)$, combined with the assumption that $\sup_{1\le i\le n}\|\delta_{in}\| = o_p(1)$, we have $\sup_{1\le i\le n}|t_i| = o_p(1)$. Therefore, $\Gamma_{n1} = o_p(1)$ by (2.23), and (2.24), and we obtain the fact that $t_i, 1\le i\le n$ are free from x. It further implies that $\Gamma_{n2} = o_p(1)$ by (2.22) by applying the Cauchy-Schwarz inequality to the double sum. Hence $\hat\Gamma_n - \tilde\Gamma_n = o_p(1)$.

To show $\tilde\Gamma_n \to \Gamma$ in probability, first note that

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

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

$$
\begin{aligned}
\tilde\Gamma_{n1} &= \frac{8h^d}{n^4}\sum_{i\neq j} K_h^2(X_i - X_j)\xi_i^2\xi_j^2, \\
\tilde\Gamma_{n2} &= \frac{2h^d}{n^4}\sum_{i\neq j}\left[\sum_{k\neq i,j} K_h(X_k - X_i)K_h(X_k - X_j)\xi_i\xi_j\right]^2, \\
\tilde\Gamma_{n3} &= \frac{8h^d}{n^4}\sum_{i\neq j}\left[(K_h(X_i - X_j)\xi_i\xi_j)\sum_{k\neq i,j} K_h(X_k - X_i)K_h(X_k - X_j)\xi_i\xi_j\right].
\end{aligned}
$$

Taking the expectation of $\tilde{\Gamma}_{n1}$ gives $\tilde{\Gamma}_{n1} = O_p(1/n^2h^{2d}) = o_p(1)$. Recall the notation $G(x, y)$, $H(x, y, h)$ in the proof of Lemma (2.3.2), $\tilde{\Gamma}_{n2}$ can be written as

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

From the proof of Lemma 2.3.2,

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

which implies $\tilde{\Gamma}_{n21} = o_p(1)$. Recall the notation $H_n(Z_i, Z_j)$ in the proof of Lemma 2.3.2, where $Z_i = (X_i\,\xi_i)$, $1 \le i \le n$, $\tilde{\Gamma}_{n23}$ is simply $2\sum_{i\neq j} H_n^2(Z_i, Z_j)$. By the Cauchy-Schwarz inequality, and the fact that the variance is bounded above by the second moment,

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

The proof in Lemma 2.3.2 shows that the upper bound is $O((nh^d)^{-1})$. Also from the proof in Lemma 2.3.2, we know that, as $n \to \infty$,

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

so

$$\tilde{\Gamma}_{n23} = 2\sum_{i\neq j} H_n^2(Z_i, Z_j) \to \Gamma.$$

By using the Cauchy-Schwarz inequality and the results of $\tilde{\Gamma}_{n21}$, $\tilde{\Gamma}_{n23}$, we can show that $\tilde{\Gamma}_{n22} = o_p(1)$. Finally, by using the Cauchy-Schwarz inequality again and the results of $\tilde{\Gamma}_{n1}$, $\tilde{\Gamma}_{n2}$, $\tilde{\Gamma}_{n3} = o_p(1)$. This completes the proof of the lemma. ■

The rest of the Apprendix includes the proofs of the last two theorems:

Proof of Theorem 2.2.2. We only prove the theorem for T_{1n}. A similar argument applies to T_{2n}. For the sake of convenience, we use $g(x)$ instead of $g_1(x)$ to denote the regression function in the transformed regression model.

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

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

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

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

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

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

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

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

From (C1), and similar to the proof of $nh^{d/2}B_{n1} = o_p(1)$ in Lemma 2.3.4, one can show that $nh^{d/2}D_{n22} = o_p(1)$. By the Cauchy-Schwarz inequality, $nh^{d/2}|D_{n23}| \le nh^{d/2}\sqrt{D_{n21}}\sqrt{D_{n22}}$. Using Lemma 2.3.2 with $L(x) = g(x) - g(x, \theta_a)$, $D_{n21} = O_p(1)$, therefore

$$nh^{d/2}|D_{n23}| = O_p(1) \cdot \sqrt{nh^{d/2}} \cdot \sqrt{nh^{d/2}D_{n22}} = o_p(\sqrt{nh^{d/2}}).$$

In fact, it can be shown that

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

with $L(x) = g(x) - g(x, \theta_a)$. By Cauchy-Schwarz inequality again, $D_{n3} \le 2\sqrt{D_{n1}} \cdot \sqrt{D_{n2}}$. Note that $D_{n1} = O_p(1/(nh^d))$ by (2.18), so $nh^{d/2}D_{n3} = o_p(nh^{d/2})$.

Now, let's consider $\hat{C}_n$. By adding and subtracting $g(X_j)$ in e_j, $\hat{C}_n$ can be written as the sum of the following three terms

$$\hat{C}_{n1} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h^2(X_i - X_j)[Y_j - g(X_j)]^2,$$
$$\hat{C}_{n2} = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h^2(X_i - X_j)[g(X_j) - g(X_j, \hat{\theta}_n)]^2,$$
$$\hat{C}_{n3} = \frac{2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n} K_h^2(X_i - X_j)[Y_j - g(X_j)][g(X_j) - g(X_j, \hat{\theta}_n)].$$

Note that $\hat{C}_{n2}$ is bounded above by

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

By (2.21), the first term is the order of $O_p(1/(nh^d))$, and by a similar argument to C_{n1} in Lemma 2.3.5, the second term is the order of $O_p(1/(n^2h^d))$. Hence, $nh^{d/2}C_{n2} = o_p(nh^{d/2})$. Similarly, one can show that $\hat{C}_{n3} = O_p(1/n^3h^{2d}) + O_p(1/n^{3/2})$ by (2.22) and the argument on C_{n2} in Lemma 2.3.5. This implies $nh^{d/2}C_{n3} = o_p(nh^{d/2})$ and

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

Hence, the theorem is proved if one can show that $\hat{\Gamma}_n = \Gamma + O_p(1)$. The argument is routine and omitted here for the sake of brevity. ■

Proof of Theorem 2.2.3: We will consider T_{1n} first.

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

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

with

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

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

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

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

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

With j equal to i or not, we write $D_{n2} = D_{n21} + D_{n22} + D_{n23}$, where

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

and

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

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

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

and

$$nh^{d/2}D_{n21} = nh^{d/2}\cdot\frac{1}{n^2h^{2d}}\cdot O_p\left(\frac{1}{nh^{d/2}}\right) = O_p\left(\frac{1}{n^2h^{2d}}\right) = o_p(1).$$

Define $\overline{W}_{Ki} = E[K_h(X_i - X)W|X_i]$. Then D_{n22} can be written as

$$D_{n22} = \left(\frac{n}{n-1}\right)^2 (D_{n221} + D_{n222} + D_{n223}) \tag{2.25}$$

where

$$D_{n221} = \frac{1}{n}\sum_{i=1}^{n}\overline{W}_{Ki}^2,$$

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

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

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

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

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

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

and this implies

$$nh^{d/2}D_{222} = O_p\left(\frac{1}{nh^d}\right) = o_p(1). \tag{2.26}$$

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

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

with the assumption that $E[\delta^4(X)f_X^4(X)<\infty]$. Therefore,

$$D_{n221}-E(D_{n221}) = O_p\left(\frac{1}{\sqrt{n^2h^{d/2}}}\right) \quad \text{or} \quad nh^{d/2}D_{221} = nh^{d/2}E(\overline{W}_{K1})+o_p(1). \tag{2.27}$$

A routine calculation shows that $nh^{d/2}E(\overline{W}_{K1}^2)=\int\delta^2(u)H_0^2(-m(u;\theta_0))f^3(u)du+o_p(1)$. By the Cauchy-Schwarz inequality, $nh^{d/2}D_{n223}=o_p(1)$, $\quad nh^{d/2}D_{n23}=o_p(1)$. Hence for D_{n2}, we obtain

$$nh^{d/2}D_{n2} = \int \delta^2(u)H_0^2(-m(u;\theta_0))f^3(u)du + o_p(1).$$

Finally, let's consider D_{n3}. By adding and subtracting $g(X_j;\theta_0)$ from e_i^L, $\overline{W}_{Ki}$ from $K_h(X_i-X_j)W_j$, we write

$$D_{n3} = 2(D_{n31}+D_{n32}+D_{n33}+D_{n34}) \tag{2.28}$$

with

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

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

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

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

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

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

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

Similarly, using the $\sqrt{n}$-consistency of $\hat{\theta}_n$, (C1), (2.19), and the Cauchy-Schwarz inequality, one can show that $nh^{d/2}D_{n3k} = o_p(1)$ for $k = 2, 3, 4$. Also, using the similar argument to the null case, one can show that $\Gamma_n \to \Gamma$ in probability, and $nh^{d/2}(C_n - C_n^L) = o_p(1)$, where C_n^L is the same as C_n with ξ_i replaced with ξ^L. This completes the proof of Theorem 2.2.3 for T_{n1}.

The proof for T_{n2} is similar to the argument for T_{n1}. Only the difference is discussed here. Break down e_i to $e_i = e_i^L + W_i$ with $e_i^L = I(Y_i^L = 0) - g(X_i;\hat{\theta}_n)$ and $W_i = I(Y_i = 0) - I(Y_i^L = 0)$, then D_n is the sum of three terms:

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

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

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

Further, D_{n2} can be written as the sum

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

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

Note that $n^{-1}E[\sum_{i=1}^{n} W_i^2]$ equals

$$E[I(Y^L = 0) - I(Y = 0)]^2 = P(Y^* \le 0) + P(Y^{*L} \le 0) - 2P(Y^* \le 0, Y^{*L} \le 0)$$

$$= P[m(X, \theta_0) + b_n\delta(X) + \varepsilon \le 0] - P[m(X, \theta_0) + b_n\delta(X) + \varepsilon \le 0, m(X, \theta_0) + \varepsilon \le 0]$$

$$+P[m(X, \theta_0) + \varepsilon \le 0] - P[m(X, \theta_0) + b_n\delta(X) + \varepsilon \le 0, m(X, \theta_0) + \varepsilon \le 0]$$

$$= \int_{\delta(x)\le 0} [F_\varepsilon[-m(x, \theta_0) - b_n\delta(X)] - F_\varepsilon[-m(x;\theta_0)]]f_X(x)dx$$

$$+ \int_{\delta(x)>0} [F_\varepsilon[-m(x;\theta_0)] - F_\varepsilon[-m(x, \theta_0) - b_n\delta(X)]F_\varepsilon[-m(x;\theta_0)]]f_X(x)dx.$$

If we can assume that

$$F_\varepsilon[-m(x, \theta_0) - b_n\delta(X)] - F_\varepsilon[-m(x;\theta_0)] = -b_n\delta(x)f_\varepsilon(-m(x, \theta_0)) + u_n(x),$$

$$\int |u_n(x)|f_X(x)dx = o(b_n),$$

then

$$E\left[\frac{1}{n}\sum_{i=1}^{n} W_i^2\right] = b_n \int [I(\delta(x) \ge 0) - I(\delta(x) < 0)]\delta(x)f_\varepsilon(-m(x;\theta_0))f_X(x)dx + o(b_n).$$

Therefore,

$$nh^{d/2}E(D_{n21}) = nh^{d/2} \cdot \frac{K^2(0)}{n^2h^{2d}} \cdot O(b_n) = O\left(\frac{1}{(nh^{d+d/6})^3}\right) = o(1)$$

which implies $nh^{d/2}D_{n21} = o_p(1)$.

For D_{n22}, one can argue in a similar way as in (2.25) and write D_{n22} as $(n/(n-1))^2(D_{n221} + D_{n222} + D_{n223})$. Same as before, a conditional argument leads to $E(D_{n222}) \le E[K_h(X_1 - X_2)W_2^2]/(n-1)$, but $E[K_h^2(X_1 - X_2)W_2^2]$ equals

$$E\left[\frac{1}{h^{2d}}K^2\left(\frac{X_1 - X_2}{h}\right)[I(m(X_1;\theta_0) + b_n\delta(X_1) + \varepsilon_1 \le 0) - I(m(X_1;\theta_0) + \varepsilon_1 \le 0)]^2\right]$$

$$= E\left[\frac{1}{h^{2d}}K^2\left(\frac{X_1 - X_2}{h}\right)E\left([I(m(X_1;\theta_0) + b_n\delta(X_1) + \varepsilon_1 \le 0) - I(m(X_1;\theta_0) + \varepsilon_1 \le 0)]^2|X_1\right)\right].$$

This implies $E(D_{n222}) = O(b_n/(nh^d))$ and $D_{n222} = O_p(b_n/(nh^d))$, hence

$$nh^{d/2}D_{222} = nh^{d/2} \cdot O_p(b_n/(nh^d)) = O_p(1/\sqrt{nh^{3d/2}}) = o_p(1). \tag{2.29}$$

Note that the result above is different from (2.26). A similar argument as that in deriving (2.27) leads to

$$nh^{d/2}D_{221} = E[\delta^2(X)f_\varepsilon^2(-m(X;\theta_0))f_X^2(X)] + o_p(1).$$

By the Cauchy-Schwarz inequality, we have $nh^{d/2}D_{n223} = o_p(1)$ and $nh^{d/2}D_{n22} = o_p(1)$. Thus $nh^{d/2}D_{n2} \to E[\delta^2(X)f_\varepsilon^2(-m(X;\theta_0))f_X^2(X)] + o_p(1)$.

The discussion on D_{n3} can be processed in a similar way to that in (2.28). Applying the Cauchy-Schwarz inequality to D_{n31}, and using (2.29), it is obtained that

$$nh^{d/2}D_{n31} = O_p\left(\frac{1}{(nh^{5d/2})^{1/4}}\right)$$

which is $o_p(1)$ because $nh^{5d/2} \to \infty$.

The rest of the proof is similar to the null case hence omitted here for the sake of brevity.

■

Chapter 3

Simulation Studies

Two Monte Carlo simulations are conducted in this section to assess the finite sample performance of the proposed tests. We choose linear regression functions ($d = 1$ and $d = 2$) in the null models, a variety of quadratic components are added to the linear terms to serve as the alternative models. The significance level is chosen to be 0.05 for all simulations. For each scenario and various sample sizes, we repeat the tests 1000 times, the empirical level and power are calculated by $\#\{|T_{kn}| \geq 1.96\}/1000, k = 1, 2$, respectively. The `vglm` function in the R package `VGAM` is used to calculate the estimates of all unknown parameters.

Simulation 1: The data are generated from the models

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

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

First we assume σ_ε^2 is known and it takes value 1. The simulation results are presented in the left part of Table 3.1. For each model, the first row is for the test based T_{1n}, and the second row is for the test based on T_{2n}. The simulation shows that the empirical levels for both tests are all less than the nominal levels in all the chosen cases, hence the proposed tests are quite conservative. This is very common for nonparametric smoothing tests. Both tests have small powers against the alternative models for small sample sizes, but the power improves as the sample sizes gets larger. One may bootstrap the proposed tests to see their finite performance, since the bootstrap often provides a more accurate approximation to the distribution of the test statistic than the asymptotic normal distribution. Clearly, the power for test based on T_{1n} is much greater than the one based on T_{2n}, which is not surprising based on the argument given in Section 2.2. In fact, one can also show that, in the current setup,

$$\mu_1 = \int x^4[1-\Phi(-1-x)]^2\phi^3(x)dx = 0.0177 > 0.0021 = \int x^4\phi^2(1+x)\phi^3(x)dx = \mu_2.$$

The simulation results for unknown σ_ε^2 are presented in the right part of Table 3.1. The true value of σ_ε^2 is still 1. Similar pattern are obtained, and the power for test based on T_{1n} again is much greater than the one based on T_{2n}.

	$\sigma_\varepsilon^2 = 1$ is known					σ_ε^2 is unknown				
	100	300	500	800	1000	100	300	500	800	1000
$\gamma = 0$	0.002	0.002	0.004	0.008	0.004	0.002	0.001	0.002	0.005	0.003
	0.011	0.007	0.005	0.014	0.013	0.006	0.005	0.004	0.008	0.009
$\gamma = 0.1$	0.011	0.036	0.095	0.173	0.248	0.009	0.041	0.102	0.196	0.280
	0.006	0.014	0.029	0.068	0.092	0.003	0.008	0.009	0.031	0.052
$\gamma = 0.2$	0.081	0.353	0.693	0.914	0.976	0.098	0.420	0.775	0.961	0.991
	0.014	0.092	0.200	0.396	0.526	0.003	0.038	0.073	0.158	0.255
$\gamma = 0.3$	0.274	0.860	0.989	1.000	1.000	0.346	0.915	0.998	1.000	1.000
	0.052	0.266	0.562	0.829	0.941	0.012	0.063	0.178	0.392	0.571
$\gamma = 0.5$	0.792	1.000	1.000	1.000	1.000	0.882	1.000	1.000	1.000	1.000
	0.197	0.808	0.982	1.000	1.000	0.032	0.214	0.501	0.830	0.923

Table 3.1: *Tobit linear regression with one predictor, composite hypotheses*

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

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

In the simulation, (X_1, X_2) is from a bivariate normal distribution with 0 mean vector, and identity covariance matrix, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, the true regression parameters are chosen to be

$\alpha = \beta_1 = \beta_2 = 1$. We choose the product of two standard normal density functions as the kernel function, and $h = n^{-1/7}$ as the bandwidth. Data from the model with $\gamma = 0$ are used to study the empirical size, while data from the models with $\gamma = 0.1, 0.2, 0.3, 0.5$ are used to study the empirical powers. In the current setup, we can see that theoretically $P(\varepsilon \leq 1 + X_1 + X_2) \approx 28\%$ observations of Y^* are truncated below at 0.

The simulation results for $\sigma_\varepsilon^2 = 1$ are presented in the left part of Table 3.2, and for unknown σ_ε^2 are present in the right part of Table 3.2. The configuration of Table 3.3 is the same as Table 3.1. The tests again appear to be extremely conservative, the power increases with increasing sample sizes, and the test based on T_{1n} outperforms the one based on T_{2n}. We also did some simulation studies when X_1 and X_2 are weakly and moderately correlated. The results are not reported here because of their similarity to Table 3.2.

	$\sigma_\varepsilon^2 = 1$ is known					σ_ε^2 is unknown				
	100	300	500	800	1000	100	300	500	800	1000
$\gamma = 0$	0.002	0.005	0.006	0.005	0.007	0.002	0.004	0.006	0.005	0.005
	0.013	0.007	0.005	0.012	0.012	0.006	0.003	0.005	0.008	0.011
$\gamma = 0.1$	0.021	0.157	0.317	0.574	0.740	0.022	0.185	0.358	0.621	0.792
	0.016	0.022	0.055	0.104	0.130	0.011	0.011	0.033	0.062	0.081
$\gamma = 0.2$	0.233	0.829	0.990	1.000	1.000	0.289	0.896	0.995	1.000	1.000
	0.029	0.161	0.343	0.613	0.739	0.014	0.053	0.143	0.286	0.411
$\gamma = 0.3$	0.661	0.999	1.000	1.000	1.000	0.750	0.999	1.000	1.000	1.000
	0.072	0.436	0.810	0.965	0.991	0.021	0.132	0.328	0.607	0.716
$\gamma = 0.5$	0.995	1.000	1.000	1.000	1.000	0.998	1.000	1.000	1.000	1.000
	0.299	0.941	0.999	1.000	1.000	0.042	0.246	0.526	0.852	0.949

Table 3.2: *Tobit linear regression with two predictors, composite hypothese*

Remark: For comparison purposes, a simulation study for simple null hypotheses is also conducted. Using the same setups as in Simulation 1 and 2, but assuming that $\alpha = \beta = \beta_1 = \beta_2 = \sigma_\varepsilon^2 = 1$ are all known in the null models, we obtain the simulation results as shown in Table 3.3. Similar patterns as in Table 3.1 and 3.2 are also appeared in Table 3.3, but the nominal level 0.05 is kept very well in all cases. *Remark:* All the simulation results show that the test based on T_{1n} is more powerful than the one based on T_{2n}. This, together with the reasoning we made in Section 2.2, may suggest that we should use T_{1n} instead of T_{2n} to check the adequacy of the regression functions in the Tobit regression model.

	Model (3.1): $d = 1$					Model (3.2): $d = 2$				
	100	300	500	800	1000	100	300	500	800	1000
$\gamma = 0$	0.036	0.044	0.062	0.056	0.046	0.046	0.043	0.053	0.038	0.041
	0.058	0.056	0.050	0.059	0.052	0.054	0.039	0.055	0.045	0.043
$\gamma = 0.1$	0.054	0.075	0.126	0.193	0.195	0.086	0.189	0.292	0.435	0.533
	0.064	0.101	0.099	0.125	0.131	0.102	0.126	0.146	0.199	0.226
$\gamma = 0.2$	0.114	0.280	0.470	0.738	0.852	0.332	0.761	0.938	0.998	1.000
	0.113	0.206	0.297	0.455	0.554	0.198	0.348	0.529	0.743	0.831
$\gamma = 0.3$	0.244	0.673	0.921	0.996	0.999	0.677	0.991	1.000	1.000	1.000
	0.194	0.391	0.625	0.868	0.948	0.352	0.681	0.903	0.987	0.997
$\gamma = 0.5$	0.689	0.997	0.999	1.000	1.000	0.990	1.000	1.000	1.000	1.000
	0.433	0.856	0.993	1.000	1.000	0.690	0.989	1.000	1.000	1.000

Table 3.3: *Tobit linear regression with one and two predictors, simple hypotheses*

Chapter 4

Conclusion

In this report, we proposed two empirical minimum distance lack-of-fit test procedures to check the adequacy of the regression functional forms in the standard Tobit regression models. One procedure was applied to all the dataset, while the other one only focused on a binary response, but both of the two procedures are derived from a nonparametric estimator and a parametric estimator by the means of the minimum distance technique. More specifically, we constructed our test statistics from the squared difference between a nonparametric Kernel smoothing estimator and an estimate of our selected function fitted under the null hypothesis. The proposed test statistics are shown to be asymptotically normal, consistent against some fixed alternatives, and have nontrivial power for some local nonparametric alternatives. Simulation studies are conducted to assess the finite sample performance of the proposed tests. As a result, we recommend using the test procedure based on the statistic T_{1n} which uses all the data information.

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

[4] Hart, J. (1997). *Nonparametric Smoothing and Lack-of-fit Test.* Springer, New York.

[5] 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.

[6] Heckman, J. (1979). Sample bias as a specification error. *Econometrica* **47** 153–162.

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

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

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

[10] 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.

[11] 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.

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

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

[14] Powell, J.L. (1984). Least absolute deviations estimation for the censored regression model. *J. Econometrics* **25** 303–325.

[15] Rudd, P.A. (1982). A score test of consistency. unpublished manuscript, University of California, Berkley.

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

[17] Tobin,J. 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica* **26** 24–36.

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

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

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

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

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 σ_ε is unknown and estimated, but we also did simulations when σ_ε is known.

```
library(VGAM)
  set.seed(987654)
  result1=result2=matrix(rep(0,25),nrow=5);
  i=1
  for(delta in c(0,0.1,0.2,0.3,0.5))
   {
     j=1;
    for(n in c(100,300,500,800,1000))
     {
      # Generating data
        freq1=freq2=0
        for(k in seq(1000))
        {
          a=1
          b=1
          sx=1
```

```
se=1
x=rnorm(n,0,sx)
e=rnorm(n,0,se)
ystar=a+b*x+delta*x^2+e
y=pmax(ystar,0)

# Estimating a, b, se;
fit=vglm(y~x, tobit(Lower=0, Upper=Inf))
a=fit@coefficients[1]
b=fit@coefficients[3]

# g1(x)
g1=(a+b*x)*pnorm((a+b*x)/se)+se*dnorm((a+b*x)/se)
# g2(x)
g2=pnorm(-(a+b*x),0,se)

# Compute the residuals
e1=y-g1
e2=I(y==0)-g2

# Compute kernel density
h=n^{-1/5}
K=function(u){dnorm(u)}
xdif=kronecker(x,rep(1,n))-kronecker(rep(1,n),x)
fhat=apply(matrix(K(xdif/h)/h,nrow=n),2,mean)

# Compute nonparametric estimator
```

```
  hg1=apply(diag(e1)%*%matrix(K(xdif/h)/h,nrow=n),2,mean)
  hg2=apply(diag(e2)%*%matrix(K(xdif/h)/h,nrow=n),2,mean)

 # Compute Dn
  D1n=mean(hg1^2)
  D2n=mean(hg2^2)

 # Compute Cn
  fhat2=apply(matrix((K(xdif/h)/h)^2,nrow=n),2,mean)
  C1n=sum(fhat2*e1^2)/n^2
  C2n=sum(fhat2*e2^2)/n^2

# Compute Gn
  Mn=matrix(K(xdif/h)/h,nrow=n)
  An=(Mn%*%Mn)^2
  G1n=2*(t(e1^2)%*%An%*%(e1^2)/n^4-sum(e1^4*(fhat2)^2)/n^2)*h
  G2n=2*(t(e2^2)%*%An%*%(e2^2)/n^4-sum(e2^4*(fhat2)^2)/n^2)*h
  T1n=n*h^(1/2)*(D1n-C1n)/sqrt(G1n)
  T2n=n*h^(1/2)*(D2n-C2n)/sqrt(G2n)
  freq1=freq1+(abs(T1n)>=1.96)
  freq2=freq2+(abs(T2n)>=1.96)
 }
 result1[i,j]=freq1/1000
 result2[i,j]=freq2/1000
 j=j+1
 }
 i=i+1
```

```
 }
dimnames(result1)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
dimnames(result2)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
result1
result2
```

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(mvtnorm)
cvx=0;
set.seed(987654)
result1=result2=matrix(rep(0,25),nrow=5)
   i=1;
   for(n in c(100,300,500,800,1000))
    {
     j=1;
     for(delta in c(0,0.3,0.5,0.7,1))
      {
     # Generating data
      freq1=freq2=0
      for(k in seq(500))
      {
        se=1
        sx=1
        Sig=matrix(c(sx,cvx,cvx,sx),2)
```

```
x=rmvnorm(n,mean=c(0,0),Sig)
e=rnorm(n,0,1)
a=1
b1=1
b2=1
ystar=a+b1*x[,1]+b2*x[,2]+delta*(x[,1]^2+x[,2]^2)+e
y=pmax(ystar,0)

# Estimating a, b1, b2, se
 fit=vglm(y~x[,1]+x[,2],family=tobit(Lower=0, Upper=Inf))
 a=fit@coefficients[1]
 b1=fit@coefficients[3]
 b2=fit@coefficients[4]
 se=exp(fit@coefficients[2])

# g1(x)
 g1=(a+b1*x[,1]+b2*x[,2])*pnorm((a+b1*x[,1]+b2*x[,2])/se)+
     se*dnorm((a+b1*x[,1]+b2*x[,2])/se)
# g2(x)
g2=pnorm(-(a+b1*x[,1]+b2*x[,2]),0,se)

# Compute the residuals
e1=y-g1
e2=I(y==0)-g2

# Compute kernel density
h=n^{-1/7}
```

```
    x1dif=kronecker(x[,1],rep(1,n))-kronecker(rep(1,n),x[,1])
    x2dif=kronecker(x[,2],rep(1,n))-kronecker(rep(1,n),x[,2])
K=function(u,v){dnorm(u)*dnorm(v)}

   # Compute nonparametric estimator
    hg1=apply(diag(e1)%*%matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n),2,mean)
    hg2=apply(diag(e2)%*%matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n),2,mean)

  # Compute Dn
    D1n=mean(hg1^2)
    D2n=mean(hg2^2)

  # Compute Cn
    fhat2=apply(matrix((K(x1dif/h,x2dif/h)/h^2)^2,nrow=n),2,mean)
    C1n=sum(fhat2*e1^2)/n^2
    C2n=sum(fhat2*e2^2)/n^2

 # Compute Gn
    Mn=matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n)
    An=(Mn%*%Mn)^2
    G1n=2*(t(e1^2)%*%An%*%(e1^2)/n^4-sum(e1^4*(fhat2)^2)/n^2)*h^2
    G2n=2*(t(e2^2)%*%An%*%(e2^2)/n^4-sum(e2^4*(fhat2)^2)/n^2)*h^2
    T1n=n*h*(D1n-C1n)/sqrt(G1n)
    T2n=n*h*(D2n-C2n)/sqrt(G2n)
    freq1=freq1+(abs(T1n)>=1.96)
    freq2=freq2+(abs(T2n)>=1.96)
    }
```

```
        result1[i,j]=freq1/500
        result2[i,j]=freq2/500
        j=j+1
      }
      i=i+1
    }
  dimnames(result1)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
  dimnames(result2)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
  result1
  result2
```

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)
result1=result2=matrix(rep(0,25),nrow=5);
i=1
for(delta in c(0,0.1,0.2,0.3,0.5))
 {
   j=1;
  for(n in c(100,300,500,800,1000))
   {
    # Generating data
      freq1=freq2=0
      for(k in seq(1000))
```

```
{
  a=1
  b=1
  sx=1
  se=1
  x=rnorm(n,0,sx)
  e=rnorm(n,0,se)
  ystar=a+b*x+delta*x^2+e
  y=pmax(ystar,0)

  # g1(x)
  g1=(a+b*x)*pnorm((a+b*x)/se)+se*dnorm((a+b*x)/se)
  # g2(x)
  g2=pnorm(-(a+b*x),0,se)

  # Compute the residuals
  e1=y-g1
  e2=I(y==0)-g2

  # Compute kernel density
  h=n^{-1/5}
  K=function(u){dnorm(u)}
  xdif=kronecker(x,rep(1,n))-kronecker(rep(1,n),x)
  fhat=apply(matrix(K(xdif/h)/h,nrow=n),2,mean)

 # Compute nonparametric estimator
  hg1=apply(diag(e1)%*%matrix(K(xdif/h)/h,nrow=n),2,mean)
```

```
    hg2=apply(diag(e2)%*%matrix(K(xdif/h)/h,nrow=n),2,mean)

  # Compute Dn
    D1n=mean(hg1^2)
    D2n=mean(hg2^2)

  # Compute Cn
    fhat2=apply(matrix((K(xdif/h)/h)^2,nrow=n),2,mean)
    C1n=sum(fhat2*e1^2)/n^2
    C2n=sum(fhat2*e2^2)/n^2

 # Compute Gn
    Mn=matrix(K(xdif/h)/h,nrow=n)
    An=(Mn%*%Mn)^2
    G1n=2*(t(e1^2)%*%An%*%(e1^2)/n^4-sum(e1^4*(fhat2)^2)/n^2)*h
    G2n=2*(t(e2^2)%*%An%*%(e2^2)/n^4-sum(e2^4*(fhat2)^2)/n^2)*h
    T1n=n*h^(1/2)*(D1n-C1n)/sqrt(G1n)
    T2n=n*h^(1/2)*(D2n-C2n)/sqrt(G2n)
    freq1=freq1+(abs(T1n)>=1.96)
    freq2=freq2+(abs(T2n)>=1.96)
  }
  result1[i,j]=freq1/1000
  result2[i,j]=freq2/1000
  j=j+1
  }
  i=i+1
}
```

```
   dimnames(result1)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
   dimnames(result2)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
   result1
   result2
```

(b) Two Predictor Case:

```
library(VGAM)
library(mvtnorm)
cvx=0;
set.seed(987654)
result1=result2=matrix(rep(0,25),nrow=5)
   i=1;
   for(n in c(100,300,500,800,1000))
    {
     j=1;
     for(delta in c(0,0.3,0.5,0.7,1))
      {
     # Generating data
      freq1=freq2=0
      for(k in seq(500))
      {
        se=1
        sx=1
        Sig=matrix(c(sx,cvx,cvx,sx),2)
        x=rmvnorm(n,mean=c(0,0),Sig)
        e=rnorm(n,0,1)
        a=1
        b1=1
```

```
b2=1
ystar=a+b1*x[,1]+b2*x[,2]+delta*(x[,1]^2+x[,2]^2)+e
y=pmax(ystar,0)

# g1(x)
 g1=(a+b1*x[,1]+b2*x[,2])*pnorm((a+b1*x[,1]+b2*x[,2])/se)+
     se*dnorm((a+b1*x[,1]+b2*x[,2])/se)

# g2(x)
g2=pnorm(-(a+b1*x[,1]+b2*x[,2]),0,se)

# Compute the residuals
e1=y-g1
e2=I(y==0)-g2

# Compute kernel density
h=n^{-1/7}
x1dif=kronecker(x[,1],rep(1,n))-kronecker(rep(1,n),x[,1])
x2dif=kronecker(x[,2],rep(1,n))-kronecker(rep(1,n),x[,2])
K=function(u,v){dnorm(u)*dnorm(v)}

# Compute nonparametric estimator
hg1=apply(diag(e1)%*%matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n),2,mean)
hg2=apply(diag(e2)%*%matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n),2,mean)

# Compute Dn
D1n=mean(hg1^2)
```

```
      D2n=mean(hg2^2)

    # Compute Cn
      fhat2=apply(matrix((K(x1dif/h,x2dif/h)/h^2)^2,nrow=n),2,mean)
      C1n=sum(fhat2*e1^2)/n^2
      C2n=sum(fhat2*e2^2)/n^2

   # Compute Gn
      Mn=matrix(K(x1dif/h,x2dif/h)/h^2,nrow=n)
      An=(Mn%*%Mn)^2
      G1n=2*(t(e1^2)%*%An%*%(e1^2)/n^4-sum(e1^4*(fhat2)^2)/n^2)*h^2
      G2n=2*(t(e2^2)%*%An%*%(e2^2)/n^4-sum(e2^4*(fhat2)^2)/n^2)*h^2
      T1n=n*h*(D1n-C1n)/sqrt(G1n)
      T2n=n*h*(D2n-C2n)/sqrt(G2n)
      freq1=freq1+(abs(T1n)>=1.96)
      freq2=freq2+(abs(T2n)>=1.96)
      }
      result1[i,j]=freq1/500
      result2[i,j]=freq2/500
      j=j+1
    }
    i=i+1
   }
dimnames(result1)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
dimnames(result2)=list(c("M0","M1","M2","M3","M4"),c(100,300,500,800,1000))
result1
result2
```