MODEL CHECKING PROBLEMS IN MEASUREMENT ERROR MODELS WITH VALIDATION DATA

By

Pei Geng

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics – Doctor of Philosophy

2017

ABSTRACT

MODEL CHECKING PROBLEMS IN MEASUREMENT ERROR MODELS WITH VALIDATION DATA

By

Pei Geng

This thesis addresses some aspects of regression model checking problems when the covariates are observed with measurement errors. Both classical error-in-variables models and Berkson models are investigated when validation data is available.

In Tobit error-in-variables regression models, the response is truncated at a given level while the covariate is collected with errors. In this thesis we assume the density of measurement error to be unknown. Using the calibration idea, a new regression function is derived under the null hypothesis and estimated using the kernel smoothing method and validation data. Then a class of test statistics are constructed using the nonparametric residuals based on kernel regression estimators when validation data is available. The proposed class of tests is shown to be robust to the choices of parameter estimators and consistent against a large class of fixed alternatives. The asymptotic normality of these test statistics is established under the null hypothesis and under a sequence of local alternatives. A practical bandwidth selection strategy is developed. A finite sample simulation study shows superiority of a member of the proposed class of tests over the two existing tests in terms of empirical power. A real data application is presented to validate the current understanding of the data set.

In Berkson models, without specifying the measurement error density, the calibrated regression function is estimated using both the primary data containing the responses and the validation data. A kernel smoothed integrated square distance is defined between the responses and the regression estimator. The parameter estimators are obtained by minimizing the integrated square distance. Further the test statistic is constructed based on the minimized distance. The consistency and asymptotic normality of these estimators are proved. The asymptotic null distribution of the proposed class of test statistics based on the corresponding minimized distances and the test consistency against certain alternatives are also established. A simulation study shows desirable behavior of a member of these minimum distance estimators and tests.

To my dear and beloved parents Jinchen Geng and Jingmiao Liu, aunt Jingzhi Liu and sister Ying Geng.

ACKNOWLEDGMENTS

First and foremost, I wish to convey my greatest and most sincere appreciation to my advisor Professor Hira L. Koul. I am indebted to his constant guidance, encouragement and enthusiasm in research, endless time and prompt replies to my questions and confusion over the years. Without his valuable suggestions and comments, vast and sharp knowledge, passionate and caring attitude, I would not have grown so fast in academia. From Professor Koul, I have learned not only the professional knowledge and critical thinking in statistics, but more importantly, the rigorous and responsible spirit towards science which has built a profound basis for my future work.

Secondly, I would like to thank Dr. Lyudmila Sakhanenko for all her advice and helpful discussions in the past. Her broad knowledge and innovative ideas has greatly enlightened me and helped me overcome the difficulty. I would like to thank Dr. Qing Lu for shedding me the light on the applications of statistics. I also would like to thank Dr. Ping-shou Zhong for serving as a member of my doctoral committee.

Thirdly, I wish to express my sincere gratitude to Professor Yimin Xiao, my MSc advisor Professor Wensheng Wang and my undergraduate professor Dr. Zifeng Yang. Without their encouragement and support, I would not have come to MSU to pursue my doctoral degree.

Last but not in the least, with my deepest affection, I would like to thank my parents Jinchen Geng, Jingmiao Liu, my sister Ying Geng and my aunt Jingzhi Liu for their endless love and support over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

x^T	the transpose of an Euclidean vector x
I_A	the indicator of event A
$\to_d$	convergence in distribution
$\to_p$	convergence in probability
$n \wedge N$	minimum of n and N
$\mathcal{N}_q(\nu, \Sigma)$	q-variate normal distribution with mean vector ν and covariance matrix Σ
KN	Koul and Ni (2004)
KS	Koul and Song (2009)

Chapter 1

Introduction

In the area of statistical inference, regression model checking is a critical topic to study the significant relationships between responses and covariates. Extensive research studies have been focusing on this topic since the early 1990s, as is evidenced by the recent review article of González-Manteiga and Crujeiras (2013). Among the main classes of tests, kernel-based test procedures are important tools to investigate if the regression functions belong to a particular parametric family.

The most of the literature on regression model checking assumes that the covariates are fully observed. However, in reality, many covariates are observed with errors. Instead of observing the true covariate of interest, one observes a surrogate variable. The regression models where covariates are observed with errors are known as measurement error regression models. In general, there are two types of measurement error models: error-in-variables (EIV) models and Berkson models. The monographs of Fuller (1987), Cheng and van Ness (1999) and Carroll, Ruppert, Stefanski and Crainiceanu (2006) provide ample examples of such models and contain systematic analysis of the underlying issues involved in these models. In general, the simple analysis using the error-prone covariates as the true ones causes heavy bias in the estimators of the underlying parameters and loss of power in tests. Hence regression calibration is needed for parameter estimation and hypothesis tests.

In economics and other social sciences, Tobit regression first introduced by Tobin (1958) is a useful model to analyze truncated responses. In Tobit models, the early study mainly focused on parameter estimation. For example, Amemiya (1984) summarizes comprehensive estimation procedures when the regression error follows a Gaussian distribution while Abarin and Wang (2009) proposes the second-order least squares estimation when the error has a general parametric distribution. In Tobit EIV models, Wang (1998) proposed a two-step moment estimators in the linear case when both the covariate and measurement error are normally distributed. When the measurement error distribution is unknown, but with the help of validation data, the least squares estimation procedure proposed by Song (2009) is applicable after the regression model is calibrated to that based on the surrogate variable.

Regarding the model checking problem in Tobit EIV models, existing methods include a score-type test proposed by Song (2009) and a transformation-based distribution-free test proposed by Song (2011). The former test applies to the least squares estimators and the latter test achieves superior performance for one dimensional covariate when the measurement error distribution is known.

However, the measurement error distribution is hardly known in reality. An alternative approach is to conduct statistical inference with an available validation sample. Chapter 2 of this thesis aims to develop a robust model checking procedure in Tobit EIV models with the help of validation data. In the chapter, a kernel-based nonparametric test is proposed for fitting a parametric function to the regression function in Tobit EIV models. Given a consistent parameter estimator, the calibrated regression function is first estimated by the Nadaraya-Watson estimator based on validation data, then a class of test statistics is constructed based on the nonparametric residuals. This class of tests is obtained by adopting the test of Zheng (1996) to the current set up. The proposed tests are shown to be robust to the parameter estimation choices and the consistency and asymptotic distributions of the test statistics are established under the null hypothesis and under certain alternatives. With the two bandwidth parameters involved in these tests, a practical bandwidth selection strategy is developed and applied in a simulation study. The simulation study shows attractive empirical power performance of a member of the proposed class of tests compared to the two existing tests.

When a predicting variable is observed with errors, in many cases, it is more appropriate to assume that the true variable equals the observed plus an error. This is the so called Berkson measurement error models. For example, the levels of a certain pollutant, such as lead, in a place are usually measured at fixed spots while the actual exposure to an individual is subject to the location, time and air condition. Therefore, it is natural to treat the actual exposure as the measured pollutant levels with a small random error term. To ensure the identifiability in the Berkson models, it is often assumed that the density of the measurement error is known. For parameter estimation, Wang (2004) proposed a minimum distance procedure based on the first two moments of responses in nonlinear regressions. As for model checking, the main contribution is made by Koul and Song (2009) where a minimum distance test is constructed based on kernel smoothing technique.

In all the mentioned studies for Berkson models above, authors have assumed that the measurement error density is known or known up to an unknown Euclidean parameter. This is a restrictive assumption as it limits the applicability of the inference procedures. However, the availability of validation data helps to circumvent this assumption.

Regarding the parameter estimation in Berkson models with validation data, a collection of attractive methodologies have been studied. In linear and nonlinear EIV models with validation data, Lee and Sepanski (1995) constructed an estimation procedures based on least squares methods with regression functions replaced by their corresponding wide-sense conditional expectation functions. In linear EIV models, Wang and Rao (2002) developed an estimated empirical loglikelihood based on the validation data and then constructed an estimated empirical likelihood confidence region for the parameters in regression functions. For general Berkson-type models, Du, Zou and Wang (2011) proposed a nonparametric regression function estimation based on kernel smoothing skills.

But the model checking study in Berkson models with validation data seems to be sparse. To fill this gap, in Chapter 3 below, we adopt the minimum distance methodology of Koul and Song (2009) to propose analogous procedures for obtaining the parameter estimators and further perform lack-of-fit hypothesis testing. The regression function given the surrogate variable is nonparametrically estimated based on validation data. Then an integrated square distance between the responses and the regression estimator is defined by means of kernel smoothing and is minimized for parameter estimation. Eventually the minimized distance is used as the test statistic. Both consistency and asymptotic normality of the proposed estimator are established. The asymptotic distributions of the proposed test statistics under the null and the consistency against certain fixed alternative hypotheses are also derived. It is shown that the asymptotic distributions of these test statistics are the same as in the case of known measurement error density while those of the corresponding estimators of the parameters in the null model are affected by the estimation of the regression function using validation data. A finite sample study shows literally no bias in the proposed minimum distance estimators. Empirical levels and power are obtained for different choices of the sample size ratio between primary data and validation data under various alternative hypotheses. The empirical level is well controlled for most of the chosen cases and the empirical power significantly increases as the sample size increases for all the chosen alternatives.

Chapter 2

Model checking in Tobit EIV regression using validation data

2.1 Introduction

In economics and other social sciences, many response variables are observed with lower or upper thresholds. For instance, household expenditure on certain durable goods is zero for some families depending on other factors and positive for other families, hours worked in social science is zero for women who choose not to work and positive for others, and as the third example, the demand of tickets for a game or conference is also limited to the capacity of the event. Regression model with truncated response data was first studied by Tobin (1958). Since then these models are called Tobit regression models. Bhattacharya, Chernoff and Yang (1983) developed a nonparametric Mann-Whitney type estimator of the parameter in linear Tobit models. The survey paper of Amemiya (1984) provides a comprehensive introduction to these regression models with Gaussian errors.

To proceed a bit more precisely, in the Tobit regression model of interest here, the scalar response variable Y^* is observed only when it is positive and is related to the p-dimensional predicting variable vector X by the relation

$$Y^* = \mu(X) + \varepsilon, \qquad Y = Y^* I_{(Y^*>0)}. \tag{2.1}$$

Here Y is the observed response and the scalar random error ε is assumed to have zero mean and to be independent of X so that $\mu(x) = E(Y^*|X = x), x \in \mathbb{R}^p$.

The last two and a half decades have seen an intense research activity on the topic of testing for lack-of-fit of a regression model as is evidenced in a recent review paper of González-Manteiga and Crujeiras (2013). Let $\Theta \subset \mathbb{R}^q$ and $\mathcal{M} := \{m(x,\theta), x \in \mathbb{R}^p, \theta \in \Theta\}$ be a family of known parametric functions. In the lack-of-fit testing problem of interest here we wish to test the hypothesis

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

where $\mathcal{C}$ is a compact subset of $\mathbb{R}^p$.

In the case of no measurement error, i.e., when X is fully observed, there are a few tests for fitting a parametric model to $\mu(x)$ in the above Tobit regression model. Wang (2007) developed a nonparametric test to diagnose nonlinearity in the median Tobit regression model where covariate is non-random, one dimensional, and the function $\mu(x)$ represents the median of the distribution of Y^* at the design variable x. Song (2011) proposed an asymptotically distribution-free test for fitting a parametric model to $\mu(x)$ of (2.1). This test is based on the supremum of the Stute, Thies and Zhu (1998) type transformation of a partial sum process of calibrated residuals and is applicable only when the dimension p of X equals 1. Koul, Song and Liu (2014) adopted the Zheng's (1996) statistics to test H_0 for a large class of the given functions $\mathcal{M}$ and for $p \geq 1$.

The goal here is to develop tests for H_0 in the model (2.1) when there is measurement error in the covariate vector X, where one does not observe X. Instead one observes a surrogate variable Z related to X by the relation

$$Z = X + U, \tag{2.2}$$

where the random error U is distributed with mean 0 and unknown covariance matrix Σ_u. The r.v.'s X, U, ε are assumed to be mutually independent.

Throughout the chapter, the primary data set consists of a random sample of n observations $\{(Y_i, Z_i), i = 1, 2, \cdots, n\}$ obtained from the model (2.1) and (2.2). We further assume that there is a validation data set consisting of N i.i.d. observations $\{(\widetilde{X}_j, \widetilde{Z}_j), j = 1, \cdots, N\}$ on (X, Z) of (2.2), independent of the primary data set.

To proceed further, we recall the testing methodology developed in Koul et al. (2014) (KSL) when there is no measurement error in X. For any r.v. V, let f_V denote its density. Under H_0, the regression function of Y, given X, is

$$q(x, \theta) = E(Y|X = x) = m(x, \theta)Q_{\varepsilon,0}(-m(x, \theta)) + Q_{\varepsilon,1}(-m(x, \theta)),$$

where $Q_{\varepsilon,j}(z) = \int_z^\infty u^j f_\varepsilon(u)du$, $j = 0, 1$. Thus one has the regression model

$$Y = q(X, \theta) + \xi, \qquad E(\xi|X) = 0. \tag{2.3}$$

The function q is monotone as a function of $m(x, \theta)$. Hence the original testing problem is equivalent to testing for $E(Y|X = x) = q(x, \theta)$. As in KSL, in order to ensure the model identifiability, and for simplicity, we assume f_ε to be known. See Remark 2.3.1 for the case when f_ε belongs to a parametric family.

In the error-in-variables model here, the regression model (2.3) is of little help, because X is not observable. Instead we now derive the new regression model, given Z. Let $g(z,\theta) = E(q(X,\theta)|Z=z)$. Direct calculations show that under H_0, $E(Y|Z=z) = g(z,\theta)$, so that we have the regression model

$$Y = g(Z,\theta) + \eta, \qquad E(\eta|Z) = 0. \tag{2.4}$$

The original testing problem is thus transformed to the problem of testing

$$\begin{aligned} &H_0' : E(Y|Z=z) = g(z,\theta), \text{ for some } \theta \in \Theta, \text{ for all } z \in \mathcal{C}, \text{ versus} \\ &H_1' : H_0' \text{ is not true.} \end{aligned} \tag{2.5}$$

Clearly H_0 implies H_0'. The converse need not be true in general. Song (2008) gives a sufficient condition for the equivalence of H_0 and H_0'. It suffices to require the family of densities $f_u(z-\cdot), z \in \mathbb{R}^p$, to be a complete family. From now on, we focus on testing (2.5) under (2.4) given both primary data and validation data.

Existing literature on parametric measurement error Tobit regression model mainly focuses on the parameter estimation. Song (2009) obtained consistent estimators by a modified least square procedure while Wang (1998) proposed method of moments estimators when X, u and ε all follow normal distributions. Both estimators are shown to be $\sqrt{n}$-consistent for the true parameter θ_0 and asymptotically normal. As far as the lack-of-fit testing in these models is concerned, under the assumption that the measurement error density f_U

is known, Song and Yao (2011) generalized the test procedure in Song (2011) to the measurement error Tobit models for the one dimensional covariate while Song (2009) provided a score-type test based on the least square residuals, which were constructed using the validation data. In the current chapter we assume the availability of a validation data set, which is used to estimate g and f_U, thereby avoiding the assumption of having known f_U.

In the next section we describe the proposed tests for this problem. Section 2.3 describes the asymptotic normality of the proposed test statistics under H_0 and under some alternatives along with the needed assumptions. Some parameter estimators under H_0 are described in Section 2.4. Section 2.5.1 reports findings of a finite sample simulation study, which shows some superiority of a member of the proposed class of tests compared to the tests in Song and Yao (2011) and Song (2009) in terms of the empirical power. The proposed test is also applied to a real data example in Section 2.5.2 and the results validate the current understanding of the dataset. All proofs are deferred to the last Section 2.6.

2.2 A class of tests

To describe the proposed class of tests, we need to construct residuals in the model (2.4). Since here the function g is unknown, we need to estimate this function nonparametrically. The validation data is critical to do this. Let K be a kernel density function, w be a window width associated with sample sizes n and N, and set $K_w(x) = K(x/w)/w^p$, $x \in \mathbb{R}^p$. Let

$$W_N(z,\theta) := N^{-1}\sum_{k=1}^{N} K_w(z-\widetilde{Z}_k)q(\widetilde{X}_k,\theta), \quad \tilde{f}(z) := N^{-1}\sum_{k=1}^{N} K_w(z-\widetilde{Z}_k), \;\; z\in\mathbb{R}^p, \theta\in\Theta.$$

Then, for a given θ, a kernel estimator of $g(z, \theta)$ using the validation data set is

$$\hat{g}(z, \theta) := \frac{W_N(z, \theta)}{\tilde{f}(z)}, \quad z \in \mathbb{R}^p. \tag{2.6}$$

Because the validation data is independent of the primary data, $\hat{g}(z, \theta)$ is independent of the primary data, for each θ, so is the kernel density estimator $\tilde{f}$ of the density f_Z of Z.

Let θ_0 be the true value of the parameter θ for which H_0 holds. Let $\hat{\theta}_n$ be a $\sqrt{n}$-consistent estimator of θ_0, and define the residuals

$$\hat{\eta}_i = Y_i - \hat{g}(Z_i, \hat{\theta}_n), \quad i = 1, \cdots, n.$$

Let $h = h_n$ be another sequence of window widths. Based on the idea proposed by Zheng (1996), under H_0, since $E(\eta_i | Z_i = z) = 0$ for all $z \in \mathcal{C}$, we have

$$E[\eta_i E(\eta_i | Z_i) f_Z(Z_i)] = 0, \quad \forall\, i \geq 1, \tag{2.7}$$

while it is strictly positive under H_1. In order to use the empirical version of (2.7) in the primary dataset to form the test statistic, the conditional expectation in the above equation can be estimated by

$$\hat{E}[\eta_i | Z_i] = \sum_{j \neq i, j=1}^{n} \frac{1}{(n-1)h^p} K\Big(\frac{Z_j - Z_i}{h}\Big) \hat{\eta}_j I_{\mathcal{C}}(Z_j) \Big/ \tilde{f}(Z_i). \tag{2.8}$$

Upon multiplying this by $\hat{\eta}_i \tilde{f}(Z_i) I_{\mathcal{C}}(Z_i)$ and then summing up over i leads to the class of test statistics, one for each K,

$$V_n = \frac{1}{n(n-1)h^p} \sum_{i=1}^{n} \sum_{i \neq j=1}^{n} I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K\Big(\frac{Z_i - Z_j}{h}\Big) \hat{\eta}_i \hat{\eta}_j, \tag{2.9}$$

useful in the current set up. The reason for restricting the covariate Z to a compact set $\mathcal{C}$ is to avoid the usual difficulty associated with the vanishing $\tilde{f}(z)$.

We first decompose the residuals as

$$\begin{aligned} \hat{\eta}_i &= Y_i - \hat{g}(Z_i, \hat{\theta}_n) \qquad (2.10) \\ &= [Y_i - g(Z_i, \theta_0)] + [g(Z_i, \theta_0) - \hat{g}(Z_i, \theta_0)] + [\hat{g}(Z_i, \theta_0) - \hat{g}(Z_i, \hat{\theta}_n)] \\ &:= \eta_i - e_i - \delta_i, \quad \text{say.} \end{aligned}$$

Compared to the test statistic in Zheng (1996), it is important to observe that there is an extra nonparametric estimation residual term e_i involved here, which is shown later to contribute to the asymptotic distributions of V_n.

2.3 Asymptotic distributions

In this section we shall derive the asymptotic distributions of V_n under H_0 and under some alternatives.

2.3.1 Asymptotic null distribution

We shall first state the needed assumptions for obtaining the asymptotic null distribution of V_n. Let $\sigma^2(z) := E(\eta^2 | Z = z), z \in \mathbb{R}^p$, $\widetilde{\gamma}^2(\widetilde{z}) := E\big\{\big(q(\widetilde{X}, \theta_0) - g(\widetilde{Z}, \theta_0)\big)^2 | \widetilde{Z} = \widetilde{z}\big\}, \widetilde{z} \in \mathbb{R}^p$,

$\mathcal{N}_0$ denote an open neighborhood of θ_0 and $\|\cdot\|$ the Euclidean norm of a vector or of a matrix. For a given positive integer k, a density kernel K is said to be of order k if $\int u^j K(u)du = 0$, for all $1 \le j \le k-1$ and $\int u^k K(u)du \neq 0$. We are now ready to state our assumptions. All limits are taken under $N \wedge n \to \infty$, unless mentioned otherwise.

(C1) The density function f_Z is continuously differentiable and $\inf_{z\in\mathcal{C}} f_Z(z) > 0$.

(C2) The regression function $m(x,\theta)$ is differentiable with respect to θ, for each $x \in \mathbb{R}^p$, with its vector of derivatives $\dot{m}(x,\theta)$ satisfying $E\big(\sup_{\theta\in\Theta} |m(X,\theta)|^2 + \sup_{\theta\in\Theta} \|\dot{m}(X,\theta)\|^2\big) < \infty$.

(C3) There exists an estimator $\hat{\theta}_n$ of θ_0, such that $\sqrt{n}\|\hat{\theta}_n - \theta_0\| = O_p(1)$, under H_0.

(C4) For some $\Delta > 0$, $\sup_{z\in\mathbb{R}^p} \sigma^{2+\Delta}(z) f_Z(z) < \infty$. The functions $\sigma^2(z)f_Z(z)$ and $g(z,\theta_0)f_Z(z)$ are continuous and uniformly bounded. The functions $f_Z(z)$ and $g(z,\theta_0)f_Z(z)$ and their first and second derivatives are continuous and uniformly bounded.

(C5) For each $z \in \mathbb{R}^p$, $g(z,\theta)$ is twice continuously differentiable in θ at θ_0. The second differential $\ddot{g}$ satisfies $E\{\sup_{\theta\in\mathcal{N}_0} \|\ddot{g}(Z,\theta)\|^2\} < \infty$.

(C6) $E(\sigma^2(Z))^2 + E(\widetilde{\gamma}^2(\widetilde{Z}))^2 < \infty$.

(C7) K is a continuous and symmetric density function on $\mathbb{R}^p$ with bounded partial derivatives, and of order $k > p/4$.

(C8) $N/n \to \lambda$, $h \to 0$, $w \to 0$, and $w/h \to c,\ 0 \le c < \infty$.

(C9) $(n \wedge N)h^p \to \infty$, $(n \wedge N)w^p \to \infty$.

(C10) With k as in (C7), $nh^{p/2}w^{2k} \to 0$.

Remark 2.3.1. Parametric f_ε. Suppose f_ε belongs to a parametric family of densities with unknown parameter vector ν. Then $Q_{\varepsilon,1}$, $Q_{\varepsilon,0}$ and $g(z,\theta)$ will also depend on ν. Let $\gamma := (\theta^T, \nu^T)^T$, and $\hat{\gamma}$ be a $\sqrt{n}$-consistent estimator of γ under H_0. Then one can apply the above tests with $g(z, \hat{\theta})$ replaced by $g(z, \hat{\gamma})$ throughout. The asymptotic distributions of the thus modified test statistics are not affected by this modification. For example, if $\varepsilon \sim \mathcal{N}_1(0, \sigma^2)$ and $m(x, \theta) = \alpha + \beta x$, then $\gamma = (\theta^T, \sigma)^T = (\alpha, \beta, \sigma)^T$ and

$$\begin{aligned} q(x, \gamma) &= m(x, \theta) Q_{\varepsilon,0}(-m(x, \theta)) + Q_{\varepsilon,1}(-m(x, \theta)) \\ &= (\alpha + \beta x)\Phi\Big(\frac{\alpha + \beta x}{\sigma}\Big) + \sigma\phi\Big(\frac{\alpha + \beta x}{\sigma}\Big), \end{aligned}$$

where Φ and ϕ are the cumulative distribution function and density function of the standard normal distribution. For more estimation details, see Wang (1998) and Amemiya (1984).

Remark 2.3.2. The conditions (C3)–(C5) are essentially used to ensure the $\sqrt{n}$-consistency of the least square parameter estimators in Section 3.3 while the assumptions (C7)–(C10) about K and bandwidths are needed to derive the asymptotic distributions of the proposed test statistics.

Remark 2.3.3. Note that the order k of the kernel function K needs to be larger than $p/4$ in order to obtain a valid bandwidth h since both $nh^p \to \infty$ and $nh^{p/2+2k} \to 0$ should be satisfied. For example, if $p < 8$, (C10) will be automatically true for any symmetric kernel density. However, the test will suffer from the curse of dimensionality since the asymptotic bias of kernel regression estimators is of the order $O(h^k)$. As p increases, the basic assumption $nh^p \to \infty$ requires wider bandwidth. As a consequence, the kernel function order k has to be increased in order to make the bias negligible compared to the asymptotic rate.

To state the main theorem, here we need to introduce

$$K_1 := \int K^2(u) du, \tag{2.11}$$

$$
\begin{aligned}
K_2 &:= \int \Big\{ \int K(u)K(v)\frac{1}{2}\big[K(s+c(u-v)) + K(-s+c(u-v))\big]dudv\Big\}^2 ds \\
\tau_1 &:= \int I_{\mathcal{C}}(z)[\sigma^2(z)]^2 f_Z^2(z)dz\, K_1, \quad \tau_2 := \int I_{\mathcal{C}}(z)[\widetilde{\gamma}^2(z)]^2 f_Z^2(z)dz\, K_2,
\end{aligned}
$$

where c is as in assumption (C8). We are now ready to state the following theorem describing the asymptotic null distribution of V_n. Throughout, $\to_p$ and $\to_D$ denote the convergence in probability and in distribution, respectively.

Theorem 2.3.1. *Under (2.1) and (2.2), the assumptions (C1)–(C10) and under H_0, the following result holds. With λ as in assumption (C8), if $0 < \lambda < \infty$, then*

$$
nh^{p/2}V_n \to_d \mathcal{N}_1(0, 2\tau_1 + (2\tau_2)/\lambda^2). \tag{2.12}
$$

Moreover, τ_1 and τ_2 can be consistently estimated by

$$
\begin{aligned}
\hat{\tau}_1 &:= \frac{1}{n(n-1)} \sum_{i \neq j=1}^{n} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)K_h(Z_i - Z_j)\hat{\eta}_i^2\hat{\eta}_j^2\, K_1, \\
\hat{\tau}_2 &:= \frac{1}{N(N-1)} \sum_{k \neq l=1}^{N} I_{\mathcal{C}}(\widetilde{Z}_k)I_{\mathcal{C}}(\widetilde{Z}_l)K_w(\widetilde{Z}_k - \widetilde{Z}_l)\widetilde{\eta}_k^2\widetilde{\eta}_l^2\, K_2,
\end{aligned}
$$

where $\widetilde{\eta}_k = q(\widetilde{X}_k, \hat{\theta}_n) - \hat{g}(\widetilde{Z}_k, \hat{\theta}_n),\ k = 1, \cdots, N.$

Consequently, the test that rejects the null hypothesis whenever

$$
\Big| nh^{p/2}V_n / \sqrt{2\hat{\tau}_1 + 2\hat{\tau}_2/\lambda^2} \Big| > z_{\alpha/2}
$$

will have the asymptotic size α, where z_α is the upper α quantile of the $\mathcal{N}_1(0,1)$ distribution.

The proof of the above theorem is given in Section 2.6. Here we briefly sketch the idea of the proof, which is also helpful in discussing the case of $\lambda = \infty$. For the sake of brevity, write $\sum_{i\neq j}$ for $\sum_{i=1}^n \sum_{j\neq i=1}^n$, $\sum_{k\neq l}$ for $\sum_{k=1}^N \sum_{l\neq k=1}^N$, and let

$$K_{h,ij} \;:=\; K_h(Z_i - Z_j) = h^{-p} K\big((Z_i - Z_j)/h\big), \quad 1 \le i, j \le n.$$

Then, using the decomposition (2.10), the statistic V_n can be decomposed as

$$\begin{aligned} V_n &= \frac{1}{n(n-1)} \sum_{i\neq j} I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_{h,ij}\, \hat\eta_i\, \hat\eta_j \qquad (2.13) \\ &= \frac{1}{n(n-1)} \sum_{i\neq j} I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_{h,ij} \Big\{ \eta_i\eta_j + e_i e_j + \delta_i\delta_j - 2\eta_i e_j - 2\eta_i\delta_j - 2e_i\delta_j \Big\} \\ &:= V_{n1} + V_{n2} + V_{n3} - 2U_{n1} - 2U_{n2} - 2U_{n3}, \quad \text{say.} \end{aligned}$$

We will first show that only V_{n1} and V_{n2} contribute to the asymptotic variance of $nh^{p/2}V_n$ and the asymptotic mean of $nh^{p/2}V_n$ is 0 under assumed conditions. Then both V_{n1} and V_{n2} are approximated by degenerate U statistics, constructed by the projection of V_{n1} based only on the primary data and that of V_{n2} based only on the validation data. Eventually we obtain that V_n is asymptotically normally distributed with convergence rate $nh^{p/2}$.

In fact, $\tau_1 = E_{\mathcal{C}}\{[\sigma^2(Z)]^2 f_Z(Z)\} K_1 = E_{\mathcal{C}}\{\eta^2 [f_Z(Z) E(\eta^2|Z)]\} K_1$, where $E_{\mathcal{C}}$ denotes the expectation over the compact subset $\mathcal{C}$. The unconditional expectation can be consistently estimated by the sample average

$$\frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i) \hat\eta_i^2 \hat f(Z_i) \{\widehat{E}(\eta_i^2|Z_i)\},$$

where $\hat f$ is a kernel density estimator of f_Z based on $\{Z_i, i = 1, \cdots, n\}$ in the primary data, and the conditional expectation $\widehat{E}$ is the kernel estimator given by

$$\widehat{E}(\eta_i^2|Z_i) = \frac{1}{n-1}\sum_{j\neq i,j=1}^{n} I_{\mathcal{C}}(Z_j)K_{h,ij}\hat{\eta}_j^2/\hat{f}(Z_i).$$

Plugging in the estimated conditional residuals in the sample average, we obtain

$$\hat{\tau}_1 = \frac{1}{n(n-1)}\sum_{j\neq i} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)K_{h,ij}\hat{\eta}_i^2\hat{\eta}_j^2K_1.$$

The parameter τ_2 can be estimated similarly. Actually, both $\{Z_i, i = 1, \cdots, n\}$ and $\{\widetilde{Z}_k, k = 1, \cdots, N\}$ can be used to formulate the kernel density estimator in $\hat{\tau}_1$ to make the estimation more efficient as long as they are i.i.d. copies of Z.

Remark 2.3.4. Alternative consistent estimators of τ_1 and τ_2 are given by

$$\begin{aligned}\tilde{\tau}_1 &:= \frac{1}{n(n-1)}\sum_{i\neq j} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)h^pK_{h,ij}^2\hat{\eta}_i^2\hat{\eta}_j^2, \\ \tilde{\tau}_2 &:= \frac{1}{N(N-1)}\sum_{k\neq l} I_{\mathcal{C}}(\widetilde{Z}_k)I_{\mathcal{C}}(\widetilde{Z}_l)w^pK_{w,kl}^2\tilde{\eta}_k^2\tilde{\eta}_l^2.\end{aligned} \qquad (2.14)$$

A modification of the proofs in Zheng (1996) yields that $\tilde{\tau}_j, j = 1, 2$ are indeed consistent for $\tau_j, j = 1, 2$, respectively. Details are skipped. A simulation study also shows little difference between the two proposed estimation methods when sample size is large.

Remark 2.3.5. ***The case $\lambda = \infty$ and $\lambda = 0$.*** When the validation data size N is much larger than the primary data size n, i.e., when λ is sufficiently large, the regression function $g(Z, \theta_0)$ can be efficiently recovered by means of the validation data, hence the primary data dominates the asymptotic behavior of the tests while the validation data does not play a role in their asymptotic variances. This is justified by letting $N/n \to \infty$, or having $\lambda = \infty$.

From the proof of the above theorem, we see that

$$nh^{p/2}V_{n1} \to_D N(0, 2\tau_1), \text{ and } Nh^{p/2}V_{n2} \to_D N(0, 2\tau_2).$$

Moreover, V_{n1} and V_{n2} are asymptotically independent. All other terms in V_n are asymptotically negligible, compared to these two terms. Hence

$$nh^{p/2}V_n \sim nh^{p/2}V_{n1} + nh^{p/2}V_{n2} = nh^{p/2}V_{n1} + \frac{n}{N}Nh^{p/2}V_{n2} \to_D N(0, 2\tau_1),$$

since now $n/N \to 0$.

On the other hand, when the primary data set is much larger than the validation data set, the asymptotic convergence rate is limited by the validation data sample size. By going through the proof of Theorem 2.3.1 again with $N/n \to 0$, we obtain

Theorem 2.3.2. *Under (2.1), (2.2), assumptions (C1)–(C10), and H_0, as $N/n \to \lambda = 0$, $Nh^{p/2}V_n \to_D N(0, 2\tau_2)$.*

2.3.2 Asymptotic power

In this section we shall investigate the asymptotic power of the proposed tests under the fixed alternative

$$H_\ell : \mu(x) = \ell(x),$$

where $\ell \notin \mathcal{M}$ and $E\ell^2(X) < \infty$. Let $h(Z) = E(\ell(X)|Z)$. Then the relation between Y and Z takes the form

$$Y = h(Z) + \eta.$$

Note that $E\ell^2(X) < \infty$ implies that $Eh^2(Z) < \infty$. Additionally, we assume the following.

(C11) $E[(h(Z) - g(Z, \theta))^2 f_Z^2(Z)]$ has a unique minimizer θ_a.

Under H_ℓ, the decomposition of the residuals (2.10) becomes

$$\begin{aligned} \hat{\eta}_i &= [Y_i - g(Z_i, \theta_a)] + [g(Z_i, \theta_a) - \hat{g}(Z_i, \theta_a)] + [\hat{g}(Z_i, \theta_a) - \hat{g}(Z_i, \hat{\theta}_n)] \\ &= \bar{\eta}_i - \bar{e}_i - \bar{\delta}_i, \end{aligned}$$

where $\bar{\eta}_i = h(Z_i) - g(Z_i, \theta_a)$. Because $\bar{\eta}_i$ is no longer centered at 0 under H_ℓ, V_{n1} is a non-degenerate U statistic. As shown in Lemma 2.6.1, under H_ℓ, the asymptotic property of V_{n1} is still dominated by

$$T_{n1} = \frac{1}{n(n-1)} \sum_{i \neq j} \tilde{\varphi}_2(Z_i, Z_j),$$

where $\tilde{\varphi}_2(Z_i, Z_j) = I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_{h,ij} \bar{\eta}_i \bar{\eta}_j$. T_{n1} is a non-degenerate U statistic as well. By Lemma 3.1 in Zheng (1996),

$$T_{n1} = \frac{2}{n} \sum_{i=1}^{n} E[\tilde{\varphi}_2(Z_i, Z_j) | Z_i] - E[\tilde{\varphi}_2(Z_1, Z_2)] + o_p(1/\sqrt{n}).$$

By the weak law of large numbers, the first term above converges to $2E[\tilde{\varphi}_2(Z_1, Z_2)]$, in probability. Algebra shows that

$$E[\tilde{\varphi}_2(Z_1, Z_2)] = E\{I_{\mathcal{C}}(Z)[h(Z) - g(Z, \theta_a)]^2 f_Z(Z)\} + o(1).$$

Hence

$$T_{n1} \to_p E\Big(I_{\mathcal{C}}(Z)[h(Z) - g(Z,\theta_a)]^2 f_Z(Z)\Big). \tag{2.15}$$

Since T_{n1} is the only leading term in V_{n1}, we have

$$V_{n1} \to_p E\Big(I_{\mathcal{C}}(Z)[h(Z) - g(Z,\theta_a)]^2 f_Z(Z)\Big). \tag{2.16}$$

Under H_ℓ, both $\bar{e}_i$ and $\bar{\delta}_i$ share the properties of e_i and δ_i with θ_0 replaced by θ_a. Thus arguing as under H_0, one can verify that all the terms except V_{n1} in (2.13) are $O_p(1/(\sqrt{nh^{p/2}}))$. Moreover,

$$\begin{aligned}
\tilde{\tau}_1 &= \frac{1}{n(n-1)} \sum_{i \neq j} h^p I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K^2_{h,ij} \hat{\eta}_i^2 \hat{\eta}_j^2 \\
&= \frac{1}{n(n-1)} \sum_{i \neq j} h^p I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K^2_{h,ij} \bar{\eta}_i^2 \bar{\eta}_j^2 + o_p(1) \\
&\to_p \int K^2(u)du \ E\{I_{\mathcal{C}}(Z)\{\sigma^2(Z) + [h(Z) - g(Z,\theta_a)]^2\} f_Z(Z)\} := \bar{\tau}_1.
\end{aligned}$$

Unlike $\tilde{\tau}_1$, $\tilde{\tau}_2$ does not involve any information of Y_i. Instead, since $\tilde{\eta}_k = q(\widetilde{X}_k) - \hat{g}(\widetilde{Z}_k, \hat{\theta}_n)$, the kernel regression estimator $\hat{g}(\widetilde{Z}_k, \hat{\theta}_n)$ always consistently estimates $E(q(\widetilde{X}_k | \widetilde{Z}_k))$, hence $\tilde{\tau}_2 \to_p \tau_2$ under H_ℓ. Now we state the asymptotic property of the proposed test under H_ℓ.

Theorem 2.3.3. *Under the conditions (C1)–(C10) and (C11), and under the alternative hypothesis H_ℓ, for finite $0 < \lambda < \infty$, we obtain*

$$\frac{V_n}{\sqrt{2\hat{\tau}_1 + 2\hat{\tau}_2/\lambda^2}} \to_p \frac{E\{I_{\mathcal{C}}(Z)[h(Z) - g(Z,\theta_a)]^2 f_Z(Z)\}}{\sqrt{2\bar{\tau}_1 + 2\tau_2/\lambda^2}} > 0.$$

The standardized test statistic $nh^{p/2}V_n/\sqrt{2\hat{\tau}_1 + 2\hat{\tau}_2/\lambda^2} \to_p \infty$, under H_ℓ. Hence the proposed test is consistent against the alternatives H_ℓ.

Now we consider a sequence of local alternatives:

$$H_a : E(Y_i^*|X_i) = m(X_i, \theta_0) + b_n\, a(X_i),$$

where $a(\cdot) \notin \mathcal{M}$, $a(\cdot)$ is continuously differentiable, $E[a(X)]^2 < \infty$, and $b_n \to 0$.

Theorem 2.3.4. *Under (C1)–(C11), H_a with $b_n = (nh^{p/2})^{-1/2}$ and $0 < \lambda < \infty$, we have*

$$\frac{nh^{p/2}V_n}{\sqrt{2\hat{\tau}_1 + (2\hat{\tau}_2)/\lambda^2}} \to_D N(\gamma, 1),$$

where $\gamma = E\{I_{\mathcal{C}}(Z)E[a(X)|Z]^2 f_Z(Z)\}/\sqrt{2\tau_1 + (2\tau_2)/\lambda^2}$.

2.4 Estimation of θ_0

To perform the proposed testing procedure, we first need to obtain a $\sqrt{n}$-consistent estimator θ_0. Song (2009) uses the least square estimator

$$\hat{\theta}_{OLS} = \arg\min_\theta \sum_{i=1}^n I_{\mathcal{C}}(Z_i)\big(Y_i - \hat{g}_{w_s}(Z_i, \theta)\big)^2, \qquad (2.17)$$

where the Nadaraya-Watson regression estimator $\hat{g}_{w_s}$ employs a bandwidth w_s. In order to assess the performance of the proposed test under different estimation procedures, we also introduce the weighted least square estimators. In particular, we choose the weight function $\tilde{f}(Z_i)$ for each observation Z_i to avoid the unstableness of kernel regression estimation with

$\tilde{f}(Z_i)$ close to 0. In other words, we considered the weighted least squares estimator

$$\hat{\theta}_{WLS} \;=\; \arg\min_{\theta} \sum_{i=1}^{n} \big(Y_i - \hat{g}_{w_S}(Z_i,\theta)\big)^2 [\tilde{f}(Z_i)]^2. \tag{2.18}$$

An argument similar to the one used in Song (2009) shows that $\hat{\theta}_{WLS}$ is also asymptotically normal with convergence rate $\sqrt{n}$. In addition, when all r.v.'s X, u and ε are Gaussian in the above linear errors-in-variables Tobit model, Wang (1998) obtained two-step moment estimators $\hat{\theta}_{TME}$. In order to conveniently apply the estimator in the next section, we briefly describe it here. Assume

$$Y^* = \alpha + \beta^T X + \varepsilon, \quad Y = \max\{Y^*, 0\}, \quad Z = X + u,$$

$$X \sim N(\mu_X, \Sigma_X), \quad u \sim N(0, \Sigma_u), \quad \varepsilon \sim N(0, \Sigma_\varepsilon),$$

where X, u and ε are mutually independent and $\Delta := \Sigma_X^{-1}\Sigma_u$ is known. Under the normality assumption, the first and second moments can be calculated

$$\mu_{Y^*} = \alpha + \beta^T \mu_X, \quad \mu_X = \mu_Z, \tag{2.19}$$

$$\sigma^2_{Y^*} = \beta^T \Sigma_X \beta + \sigma^2_\varepsilon, \quad \sigma_{ZY^*} = \Sigma_X \beta, \quad \Sigma_Z = \Sigma_X + \Sigma_u. \tag{2.20}$$

One can construct estimating equations by the substitution of sample moments in the above euqations in order to obtain the parameter estimation. However the moments of Y^* are also needed. But algebra shows that

$$E(Y) = \Phi(\delta)E(Y|Y>0), \quad \delta := \mu_{Y^*}/\sigma_{Y^*} \tag{2.21}$$

$$E(Y|Y>0)=\mu_{Y*}+\sigma_{Y*}\phi(\delta)/\Phi(\delta) \tag{2.22}$$

$$E(ZY|Y>0)=\sigma_{ZY*}+\mu_Z E(Y|Y>0). \tag{2.23}$$

Let $\hat{\mu}_Y$ denote the overall sample mean of Y_i's responses and $\hat{\mu}_{Y+}$ be the mean of the positive Y_i's only. Using sample moments in equation (2.21), we obtain

$$\hat{\delta}=\Phi^{-1}(\hat{\mu}_Y/\hat{\mu}_{Y+}).$$

Then combining this with (2.22), we obtain the estimates of μ_{Y*} and σ_{Y*} as follows.

$$\hat{\mu}_{Y*}=\hat{\delta}\hat{\mu}_{Y+}/[\hat{\delta}+\phi(\hat{\delta})/\Phi(\hat{\delta})], \qquad \hat{\sigma}_{Y*}=\hat{\mu}_{Y*}/\hat{\delta}.$$

By equation (2.23), one can further estimate σ_{ZY*} by

$$\hat{\sigma}_{ZY*}=\hat{\mu}_{ZY*}-\hat{\mu}_Z\hat{\mu}_{Y+}.$$

Then plugging $\hat{\mu}_{Y*}, \hat{\sigma}_{Y*}, \hat{\sigma}_{ZY*}$ in (2.19) and (2.20), and using $\Delta=\Sigma_X^{-1}\Sigma_u$, which is known, we obtain the following estimators of α and β.

$$\hat{\beta}_{TME}=\hat{\Sigma}_X\hat{\sigma}_{ZY*}, \quad \hat{\Sigma}_X=\hat{\Sigma}_Z(I+\Delta)^{-1}, \quad \hat{\alpha}_{TME}=\hat{\mu}_{Y*}-\hat{\mu}_X\hat{\beta}, \quad \hat{\mu}_X=\hat{\mu}_Z.$$

Computationally, $\hat{\theta}_{TME}=(\hat{\alpha}_{TME},\hat{\beta}_{TME})^T$ is more efficient due to the closed form while the other two estimators require optimization. We use all three estimators in the simulation study of the next section.

2.5 Data analysis

2.5.1 Simulations

In this section we present the findings of a Monte Carlo simulation study. In this study we used the three estimators of θ_0 mentioned in Section 3.3. Let $V_{\hat{\theta}}$ denote the corresponding test statistic with estimator of θ_0 equal to $\hat{\theta}$. We chose $p = 1$, 2. The empirical level of the $V_{\hat{\theta}}$ test is seen to be robust against these choices of the estimators for $p = 1$. In all simulations we set $N = 2n$ for convenience. All the results are obtained by generating 1000 replications.

In the case of one-dimensional covariate, i.e., when $p = 1$, we compared the power performance of the $V_{\hat{\theta}}$ test with the two existing methods: one is the W_n test of Song and Yao (2011) based on Stute, Thies and Zhu (1998) type transformation of a partial sum residual process and the other is the score-type test S_n in Song (2009). Although Song's score-type test does not directly apply to the Tobit model, it can be successfully adapted to the current model when the testing problem is transformed to (2.5). In order to clearly see how the simulation is implemented, we briefly describe both existing methods here. Regarding the score type test, S_n is defined as

$$S_n = \frac{1}{n}\sum_{i=1}^{n} I_{\mathcal{C}}(Y_i - \hat{g}_{w_s}(Z_i))W(Z_i),$$

where $W(\cdot)$ is a weight function of the covariate Z. Specifically, we used the uniform weight function $W(Z_i) \equiv 1$ in the simulation.

As for the transformed test W_n, first define the stochastic process

$$W_n(z) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{e}_i I(c \le Z_i \le z) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{e}_i\Big[\frac{1}{n}\sum_{j=1}^{n} \hat{l}^T(Z_j)\hat{M}_j^{-1} I(c \le Z_j \le Z_i \wedge z)\Big]\hat{l}(Z_i),$$

where

$$
\begin{aligned}
\hat{e}_i &= \frac{Y_i - g(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \quad \hat{l}(Z_i) = \frac{\dot{g}(Z_i, \hat{\theta}_n)}{\tau(Z_i, \hat{\theta}_n)}, \\
\hat{M}_j &= \frac{1}{n} \sum_{k=1}^{n} \hat{l}(Z_k) l^T(Z_k) I(Z_k \geq Z_j \geq c), \\
\tau(z, \theta) &= E((Y - g(Z))^2 | Z = z) \\
&= \sigma^2 (\tilde{\alpha} + \tilde{\beta} z)^2 \Phi(\tilde{\alpha} + \tilde{\beta} z) + \sigma^2 (\tilde{\alpha} + \tilde{\beta} z) \phi(\tilde{\alpha} + \tilde{\beta} z) + \sigma^2 \Phi(\tilde{\alpha} + \tilde{\beta} z) - g(z, \theta)^2, \\
g(z, \theta) &= \sigma(\tilde{\alpha} + \tilde{\beta} z) \Phi(\tilde{\alpha} + \tilde{\beta} z) + \sigma \phi(\tilde{\alpha} + \tilde{\beta} z), \\
\tilde{\alpha} &= (\alpha + (1 - \Delta) \beta \mu_x)/\sigma, \quad \tilde{\beta} = (1 - \Delta) \beta / \sigma, \\
\sigma^2 &= \sigma_\varepsilon + \Delta \beta^2 \sigma_u^2, \quad \Delta = \sigma_x^2 / (\sigma_x^2 + \sigma_u^2).
\end{aligned}
$$

One rejects H_0 if

$$
\sup_{c \leq z \leq z_0} |W_n(z) / \sqrt{\hat{F}_Z(z_0) - \hat{F}_Z(c)}| > b_\alpha,
$$

where b_α is the value such that $P(\sup_{0 \leq u \leq 1} |B(u)| > b_\alpha) = \alpha$ and $B(u)$ is the standard Brownian motion. For the nominal level 0.05, $b_\alpha = 2.242$. We used $c = \min\{Z_i\}$ and z_0 is the 95th quantile of Z_i.

The simulation results show that in terms of the empirical power, the V test with $\hat{\theta}_{TME}$ outperforms all other tests for small and moderate sample sizes, and it behaves as well as the W_n test for the large sample size. Moreover, all the three $V_{\hat{\theta}}$ tests corresponding to the three different estimators $\hat{\theta}$ of θ_0 produce higher power than the corresponding score-type tests at the chosen alternatives. In addition, the empirical powers of the proposed $V_{\hat{\theta}}$ test under the three specified estimation options match the estimation performance in the sense that more accurate estimator $\hat{\theta}$ leads to higher empirical power of the corresponding $V_{\hat{\theta}}$ test. The estimators of θ_0 used in W_n and S_n tests are $\hat{\theta}_{TME}$ and $\hat{\theta}_{OLS}$, respectively, as was done in Song and Yao (2011) and Song (2009). We used the kernel density $K(u) = 0.75(1-u^2)I(|u| \leq 1)$ of order $k = 2$ for all the kernel density related tests.

In the case $p = 2$, the only other available test is the Song's score-type test. However, in Song's test, the bandwidth w_s should satisfy both conditions $Nw_s^{2p}/\log(N) \to \infty$ and $Nw_s^{2k} \to 0$, where k is a positive even integer specifying the order of the symmetric kernel density K. When p = 2, the above bandwidth conditions require k to be larger than 2, which means that symmetric densities like the above kernel $K(u)$ can not be used here. If we take $k = 4$, K will take negative values at certain points. In a simulation, using the fourth order density kernel, $K_{(4)}(u,v) = (0.086 - 0.2u^2)(0.086 - 0.2v^2)K(u)K(v)/0.046^2$ as introduced in Jones and Signorini (1997), the least square estimator of θ_0 showed large bias and mean square error, which in turn makes the score-type test difficult to implement. For these reasons we only report the finite sample behavior of the $V_{\hat{\theta}}$ tests. We find both the empirical levels and powers under the chosen alternatives are satisfying.

Bandwidth selection: As is evident, the implementation of the proposed tests requires the selection of the two bandwidths. One is the kernel regression bandwidth w and the other is the bandwidth h used in forming the test statistics V_n. It is thus important to provide a practical strategy for the selection of these bandwidths.

Given a consistent estimator $\hat{\theta}$, we propose to obtain the optimal w, denoted by w_b, by minimizing the mean square error of the kernel regression estimator $\hat{g}_w$ as follows.

$$MSE_1(w) := \frac{1}{n}\sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)(Y_i - \hat{g}_w(Z_i, \hat{\theta}))^2, \quad w_b := \arg\min_w MSE_1(w).$$

Note that no cross validation is needed since the Nadaraya-Watson estimator $\hat{g}_w$ is constructed based on the independent validation data $\{(q(\widetilde{X}_k), \widetilde{Z}_k), k = 1, ..., N\}$ instead of $\{(Y_j, Z_j), j = 1, ..., n\}$.

Regarding the bandwidth h, recall that h was originally used to estimate $E(\eta_i|Z_i)$ by the estimator given in (2.8). Since, under H_0, $E(\eta|Z = z) \equiv 0$, we propose to obtain an optimal h by minimizing the mean square error

$$MSE_2(h) = \frac{1}{n}\sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)(\hat{E}[\eta_i|Z_i])^2 = \frac{1}{n}\sum_{i=1}^{n}\Big(\sum_{j\neq i, j=1}^{n} \frac{I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)}{(n-1)h^p} K\Big(\frac{Z_j - Z_i}{h}\Big)\hat{\eta}_j \Big/ \tilde{f}(Z_i)\Big)^2.$$

To satisfy that $h \to 0$ and $w/h \to c < \infty$ in (C8), we enforce the constraint $0.1w_b \leq h \leq 10w_b$ in the above minimization. In our simulation, we applied the grid search of bandwidths starting from $0.1w_b$ with step 0.02 to obtain the optimal bandwidth h_{opt}. In some cases, a grid search study showed that $MSE_2(h)$ decays slowly to 0, for all sufficient large values of h. This will cause the chosen bandwidth to be much larger. Hence, we set a threshold 0.05 for MSE_2 to avoid choosing too large a bandwidth. To summarize,

$$h_{opt} := \min\big\{h:\ h = \text{argmin}_{0.1w_b \leq h \leq 10w_b} \max\{MSE_2(h), 0.05\}\big\}.$$

Simulation 1: $p = 1$. In this simulation, the data were generated from the model (2.1) and under H_0, where $m(x, \theta) = \alpha + \beta x$, $\theta = (\alpha, \beta)^T$, with $\alpha = 1, \beta = 0.6$, and $Z = X + u$, where $X \sim \mathcal{N}_1(0, 2^2)$, $\varepsilon \sim \mathcal{N}_1(0, 1)$ and $u \sim N(0, 0.5^2)$ so that the ratio $\sigma_u^2/\sigma_x^2 = 1/16$ in the TME procedure of Wang (1998). The truncation rate is approximately 26%. Moreover, here $q(x, \theta) = (\alpha + \beta x)\Phi((\alpha + \beta x)) + \phi((\alpha + \beta x))$, where Φ and ϕ are distribution function and density function of the $\mathcal{N}_1(0, 1)$ r.v., respectively.

Following the estimation procedure in Song (2009), w_s was set as $N^{-1/3}$ to obtain estimators $\hat{\theta}_{OLS}$ and $\hat{\theta}_{WLS}$. Table 2.1 reports the absolute bias and square root of mean square error (RMSE) (in parenthesis) of the three estimators of θ_0 mentioned in Section 3.3. As expected, both bias and RMSE decrease as the sample size n increases for the three estimators, which indicates the consistency of the estimators. Moreover, under the Gaussian scenario, $\hat{\theta}_{TME}$ is seen to be superior among the three and $\hat{\theta}_{WLS}$ is the least favorable. In the power analysis section, we will see that the empirical power of the $V_{\hat{\theta}}$-test shows a similar pattern.

n, N	$(50, 100)$	$(100, 200)$	$(200, 400)$	$(300, 600)$
$\hat{\alpha}_{OLS}$	0.019(0.196)	0.005(0.129)	0.004(0.090)	0.002(0.076)
$\hat{\alpha}_{WLS}$	0.018(0.237)	0.010(0.152)	0.005(0.104)	0.0001(0.087)
$\hat{\alpha}_{TME}$	0.008(0.166)	0.004(0.116)	0.004(0.082)	0.001(0.069)
n, N	$(50, 100)$	$(100, 200)$	$(200, 400)$	$(300, 600)$
$\hat{\beta}_{OLS}$	0.021(0.142)	0.003(0.083)	0.004(0.057)	0.001(0.049)
$\hat{\beta}_{WLS}$	0.023(0.183)	0.011(0.113)	0.004(0.081)	0.002(0.068)
$\hat{\beta}_{TME}$	0.005(0.103)	0.004(0.069)	0.0006(0.049)	0.0003(0.041)

Table 2.1: Comparison of estimators: absolute bias and RMSE in parenthesis

In the discussion below, we write V_{OLS}, V_{WLS} and V_{TME} for $V_{\hat{\theta}}$ when $\hat{\theta}$ equals the $\hat{\theta}_{OLS}, \hat{\theta}_{WLS}$ and $\hat{\theta}_{TME}$ of the previous section, respectively.

To implement the proposed test, the set $\mathcal{C}$ was used to be the overlap interval of $\{Z_i, i = 1, ..., n\}$ and $\{\widetilde{Z}_k, k = 1, ..., N\}$. In other words, $\mathcal{C}$ is chosen as the interval $[a, b]$, where $a = \max\{\min\{Z_i\}, \min\{\widetilde{Z}_k\}\}$ and $b = \min\{\max\{Z_i\}, \max\{\widetilde{Z}_k\}\}$. As mentioned in the main theorem, there are two options for estimating the asymptotic variances. To simplify the computation, we used the estimators given at (2.14). Applying the above bandwidth selection scheme, Table 2.2 shows that the empirical levels of all the V tests are well controlled for the large sample sizes.

To investigate the power performance, we compared the proposed test with the W_n and

(n,N)	(50,100)	(100,200)	(200,400)	(300,600)	(400,800)	(500,1000)
V_{OLS}	0.008	0.011	0.030	0.029	0.043	0.046
V_{WLS}	0.022	0.019	0.036	0.031	0.041	0.051
V_{TME}	0.011	0.014	0.033	0.034	0.049	0.048

Table 2.2: Empirical levels for $p = 1$ at nominal level 0.05

S_n tests mentioned above. We performed a finite sample power comparison by generating data under the model (2.1) and the alternatives $H_1 : \mu(x) = m(x, \theta, b)$, for all $x \in \mathcal{C}$ and some $b \in \mathbb{R}$, where $m(x, \theta, b) = 1 + 0.6x + b\,\sin(x)$, $b \in \mathbb{R}$.

Table 2.3 displays the empirical power of the three types of tests for increasing sample sizes. One can see that the empirical power of all tests increases as the sample sizes n, N and the nonlinear effect b increase. For small and moderate sample sizes, V_{TME} performs the best, and both W_n and V_{TME} achieve the highest power for the large sample size among the five tests. All the three V tests outperform the score-type test for the larger sample sizes. Among the three V tests, V_{TME} performs the best, followed by V_{OLS} and V_{WLS}. This finding also matches the behavior of the three estimators of θ_0 presented in Table 2.1.

(n,N)	b	W_n	S_n	V_{OLS}	V_{WLS}	V_{TME}
$(100, 200)$	0	0.059	0.039	0.011	0.019	0.014
	0.5	0.213	0.119	0.115	0.08	0.218
	1	0.382	0.306	0.448	0.300	0.743
$(300, 600)$	0	0.067	0.049	0.029	0.031	0.034
	0.5	0.563	0.317	0.592	0.400	0.643
	1	0.927	0.504	0.983	0.966	0.984
$(500, 1000)$	0	0.077	0.058	0.046	0.051	0.048
	0.5	0.822	0.394	0.805	0.640	0.834
	1	0.991	0.583	0.984	0.979	0.983

Table 2.3: Empirical power comparison for $p = 1$ at nominal level 0.05

Simulation 2: p=2. We conducted a brief simulation study for the bivariate predicting variables case. Here, both primary sample $\{(Y_i, Z_i), i = 1, ..., n\}$ and validations sample

$\{(\widetilde{X}_k, \widetilde{Z}_k), k = 1, ..., N\}$ are generated from the model

$$Y^* \;=\; \alpha + \beta_1 X_1 + \beta_2 X_2 + b\,(X_1^2 + X_2^2) + \varepsilon, \qquad Y = Y^* I(Y^* > 0), \qquad Z = X + u,$$

where $\alpha = \beta_1 = \beta_2 = 1$, $\varepsilon \sim \mathcal{N}_1(0, 0.5^2)$, $X = (X_1, X_2)^T \sim \mathcal{N}_2(0, \Sigma_x)$, $\Sigma_x = (\sigma_{ij})_{2\times 2}$, $\sigma_{11} = \sigma_{22} = 1$, $\sigma_{12} = 0.5$ and $u \sim \mathcal{N}_2(0, \Sigma_u), \Sigma_u = 0.5^2 I_2$. Here I_2 is a 2×2 identity matrix. $b = 0$ corresponds to the null model. The Gaussian distribution assumption and known covariance ratio $\Sigma_x^{-1}\Sigma_u$ suggests the use of V_{TME}-test. The compact set $\mathcal{C}$ is now a rectangle with sides chosen in a similar way as in the case of $p = 1$. We used the kernel function $K(u, v) = K(u)K(v)$ of order $k = 2$, where K is the same as in the case of $p = 1$, and the bandwidths were selected by the above MSE criteria. In Table 2.4, the empirical level is seen to be slightly liberal for the larger sample sizes and the empirical power increases as b, n and N increase.

b	(100,200)	(200,400)	(300,600)	(400,800)	(500,1000)
0	0.032	0.047	0.051	0.055	0.068
0.1	0.051	0.060	0.096	0.154	0.164
0.3	0.156	0.269	0.502	0.683	0.782
0.5	0.385	0.758	0.896	0.975	0.989

Table 2.4: Empirical level and power of V_{TME} test for $p = 2$ at nominal level 0.05

2.5.2 Real data application

The enzyme reaction speed data was originally collected in 1974 to study the relationship between the initial rate of enzyme reaction and the concentration of UDP-galactose. The data set has been analyzed by both Stute, Xue and Zhu (2007) and Du et al. (2011). The primary sample contains $n = 30$ observations of (Y_i, Z_i), where Y_i is the initial rate of reaction speed and Z_i denotes the basal density of UDP-galactose, for the ith individual, $1 \leq i \leq 30$. The basal density can be measured in the two ways described in Du et al. (2011): as a simple chemical treatment and by an expensive precision machine. The former treatment produces surrogate observation Z while the latter treatment serves accurate observation X. A validation data consisting of $N = 10$ pairs of basal density were obtained. We manually truncated the responses at 125 with truncation rate 27% to apply the proposed test. It is commonly believed that the Michaelis-Menten model given by

$$m(x, \theta) = \theta_1 x/(\theta_2 + x),$$

is an appropriate model for this data set.

In the estimation step, we adopted the *ad hoc* bandwidth $w_s = \hat{\sigma}_{\tilde{Z}} N^{-1/3}$ as recommended in Sepanski and Carroll (1993) and obtained both $\hat{\theta}_{OLS}$ and $\hat{\theta}_{WLS}$. Besides these two estimators, the empirical likelihood estimator $\hat{\theta}_{EL}$ obtained in Stute, Xue and Zhu (2007) is also presented in Table 2.5. Then in the test step, we continued applying the bandwidth selection method introduced in Section 3.5. The parameter estimators, optimal bandwidths and p-values of the V tests are presented in Table 2.5 and the estimated curves by both least square methods and empirical likelihood in Stute, Xue and Zhu (2007) are displayed in Figure 2.1. None of the V tests using the three estimators are significant, which validates the current understanding that the above Michaelis-Menten model is proper for the data set.

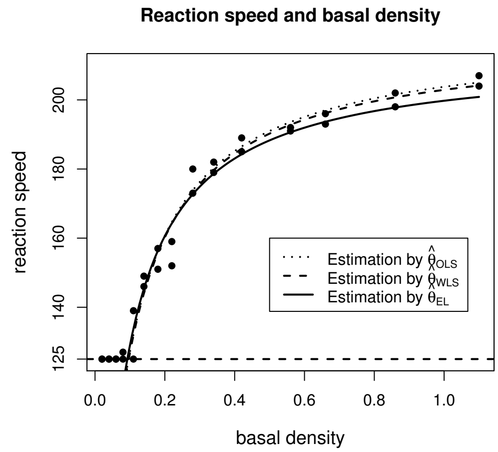

Figure 2.1: Estimated regression functions using three estimation methods

	$(\hat\theta_1, \hat\theta_2)$	w_b	h_{opt}	p-value
V_{OLS}	$(217.37, 0.071)$	0.12	0.49	0.464
V_{WLS}	$(218.41, 0.072)$	0.12	0.45	0.592
V_{EL}	$(212.70, 0.065)$	0.12	0.29	0.462

Table 2.5: Estimation and testing results of the enzyme reaction dataset

2.6 Proofs

Recall the notation given in (2.10). Throughout this section, f stands for the density f_Z of Z. We begin by listing some of the important facts about the first and second moments of the three parts of the residuals below, where $\bar\theta$ is a vector such that $\|\bar\theta - \theta_0\| \le \|\hat\theta_n - \theta_0\|$.

$$E(\eta_i|Z_i) = 0,\ E(\eta_i^2|Z_i) = \sigma^2(Z_i),\ \text{for all } 1 \le i \le n. \tag{2.24}$$

$$E(e_i|Z_i) = O(w^k),\ E(e_i^2|Z_i) = O(1/(Nw^p) + w^{2k}),\ \text{uniformly in } i \text{ for } Z_i \in \mathcal{C}. \tag{2.25}$$

$$\delta_i = \hat g(Z_i, \theta_0) - \hat g(Z_i, \hat\theta_n) = \dot{\hat g}(Z_i, \theta_0)(\theta_0 - \hat\theta_n) + \frac{1}{2}(\theta_0 - \hat\theta_n)^T \ddot{\hat g}(Z_i, \bar\theta)(\theta_0 - \hat\theta_n). \tag{2.26}$$

Fact (2.24) is assumption (C3). Fact (2.25) follows from Theorem 2.2.1 of Bierens (1987) pertaining to Nadaraya-Watson regression estimators while the claim (2.26) follows from Taylor expansions of $\hat g$ at θ_0. Intuitively, both e and δ are asymptotically negligible compared to η, however $\{e_i, i = 1, \cdots, n\}$ are not independent since they are all based on the validation data set $(\widetilde X_k,\ \widetilde Z_k)_{k=1,\cdots,N}$. Hence we need to study those terms that involve

$\{e_i, i = 1, \cdots, n\}$.

In the sequel,

$$D_i = (Z_i, \eta_i), \ \ 1 \le i \le n; \qquad \widetilde{D}_k = (\widetilde{Z}_k, \widetilde{X}_k), \ \ 1 \le k \le N.$$

We have

Lemma 2.6.1. *Under H_0 and (C1)–(C10),*

$$nh^{p/2} V_{n1} \to_d \mathcal{N}_1(0, 2\tau_1), \tag{2.27}$$

where τ_1 is defined at (2.11).

Proof. Recall (2.13). Rewrite

$$V_{n1} \ \ = \ \ \frac{1}{n(n-1)} \sum_{i \ne j} I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_h(Z_i - Z_j) \eta_i \eta_j.$$

Define

$$H_n(D_i, D_j) = I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_h(Z_i - Z_j) \eta_i \eta_j.$$

It can be seen that V_{n1} is a degenerate statistic since $E[H_n | D_i] = 0$. The result is immediate by a slight modification of the proof in Zheng (1996). □

The proofs of Lemma 2.6.2 and 2.6.3 use Lemmas 2.6.4 and 2.6.5 which are given in the end of the section.

Lemma 2.6.2. *Assume (C1)–(C10), H_0 hold and $0 < \lambda < \infty$. Then, with τ_2 is as in (2.11),*

$$Nh^{p/2} V_{n2} \to_d \mathcal{N}_1(0, 2\tau_2). \tag{2.28}$$

Proof. Let $I_{ij} = I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)$. In the proof here the indices i, j vary from 1 to n. Let

$$\widetilde{V}_{n2} = \frac{1}{n(n-1)N^2}\sum_{i\neq j}\sum_{k=1}^{N}\sum_{l=1}^{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}K_{w,ik}K_{w,jl}[q(\widetilde{X}_k)-g(Z_i)][q(\widetilde{X}_l)-g(Z_j)].$$

Direct calculations show that under (C1) and (C7)–(C9),

$$\sup_{z\in\mathcal{C}}\left|\frac{\tilde{f}(z)}{f(z)}-1\right| = o_p(1). \tag{2.29}$$

Now rewrite

$$\begin{aligned}
V_{n2} &= \frac{1}{n(n-1)}\sum_{i\neq j} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)K_h(Z_i-Z_j)e_ie_j \\
&= \frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}\Big(\frac{g(Z_i)-W_N(Z_i)}{\tilde{f}(Z_i)}\Big)\Big(\frac{g(Z_j)-W_N(Z_j)}{\tilde{f}(Z_j)}\Big) \\
&= \frac{1}{n(n-1)N^2}\sum_{i\neq j}\frac{I_{ij}K_{h,ij}}{\tilde{f}(Z_i)\tilde{f}(Z_j)}\sum_{k=1}^{N}\sum_{l=1}^{N}K_{w,ik}K_{w,jl}[q(\widetilde{X}_k)-g(Z_i)][q(\widetilde{X}_l)-g(Z_j)] \\
&= \widetilde{V}_{n2}+o_p(\widetilde{V}_{n2}).
\end{aligned}$$

The above fact follows from (2.29).

We shall now analyze $\widetilde{V}_{n2}$. Rewrite

$$\begin{aligned}
\widetilde{V}_{n2} &= \frac{1}{n(n-1)N}\sum_{i\neq j}\sum_{k=1}^{N}\frac{1}{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}K_{w,ik}K_{w,jk}[q(\widetilde{X}_k)-g(Z_i)][q(\widetilde{X}_k)-g(Z_j)] \\
&\quad +\frac{1}{n(n-1)N^2}\sum_{i\neq j}\sum_{k\neq l}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}K_{w,ik}K_{w,jl}[q(\widetilde{X}_k)-g(Z_i)][q(\widetilde{X}_l)-g(Z_j)] \\
&= V_{n21}+V_{n22}.
\end{aligned}$$

For V_{n21}, define the symmetric function

$$\psi_1(D_i, D_j, \widetilde{D}_k) = \frac{1}{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}K_{w,ik}K_{w,jk}[q(\widetilde{X}_k) - g(Z_i)][q(\widetilde{X}_k) - g(Z_j)]$$

and $L(Z_i, Z_j, \widetilde{Z}_k) = E[(q(\widetilde{X}_k) - g(Z_i))^2(q(\widetilde{X}_k) - g(Z_j))^2|Z_i, Z_j, \widetilde{Z}_k]$. Note that this kernel depends on both n and N, but this dependence is not exhibited for the sake of brevity. In order to apply Lemma 2.6.4, we need to calculate variances of all projections of ψ_1. Rigorous calculation shows that

$$\begin{aligned}
& \mathrm{Var}(\psi_1) \qquad (2.30)\\
\le\ & E\psi_1^2 = E\Big\{\frac{1}{N^2}\frac{I_{ij}K^2_{h,ij}}{f^2(Z_i)f^2(Z_j)}K^2_{w,ik}K^2_{w,jk}[q(\widetilde{X}_k) - g(Z_i)]^2[q(\widetilde{X}_k) - g(Z_j)]^2\Big\}\\
=\ & \frac{1}{N^2}E\Big\{\frac{I_{ij}K^2_{h,ij}}{f^2(Z_i)f^2(Z_j)}K^2_{w,ik}K^2_{w,jk}L(Z_i, Z_j, \widetilde{Z}_k)\Big\}\\
=\ & \frac{1}{N^2}E\Big\{\frac{I_{ij}K^2_{h,ij}}{f^2(Z_i)f^2(Z_j)}E[K^2_{w,ik}K^2_{w,jk}L(Z_i, Z_j, \widetilde{Z}_k)|Z_i, Z_j]\Big\}\\
=\ & \frac{1}{N^2w^{3p}}E_{\mathcal{C}}\Big\{\frac{K^2_{h,ij}}{f^2(Z_i)f^2(Z_j)}\int K^2(u)K^2\Big(\frac{Z_j - Z_i}{w} + u\Big)L(Z_i, Z_j, Z_i - wu)f(Z_i - wu)du\Big\}\\
=\ & O\Big(\frac{1}{N^2w^{3p}h^{2p}}E_{\mathcal{C}}\Big\{K^2\Big(\frac{Z_j - Z_i}{h}\Big)L(Z_i, Z_j, Z_i)f(Z_i)\int K^2(u)K^2\Big(\frac{Z_j - Z_i}{w} + u\Big)du\Big\}\Big)\\
=\ & O\Big(\frac{1}{N^2w^{3p}h^p}\Big).
\end{aligned}$$

In the above derivations we used assumption (C1) that guarantees that the density f is bounded from below. Next, consider

$$\begin{aligned}
& E(\psi_1|D_i, D_j)\\
=\ & \frac{1}{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}E\{K_{w,ik}K_{w,jk}[q(\widetilde{X}_k) - g(Z_i)][q(\widetilde{X}_k) - g(Z_j)]|D_i, D_j\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}E\{K_{w,ik}K_{w,jk}[\mu_2(\widetilde{Z}_k)-(g(Z_i)+g(Z_j))g(\widetilde{Z}_k)+g(Z_i)g(Z_j)]|D_i,D_j\} \\
&= \frac{1}{N}\frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}\int\frac{1}{w^{2p}}K\Big(\frac{Z_i-x}{w}\Big)(\mu_2(x)-(g(Z_i)+g(Z_j))g(x)+g(Z_i)g(Z_j))f(x)dx \\
&= O_p\Big(\frac{1}{Nw^p}\frac{I_{ij}K_{h,ij}\widetilde{\sigma}^2(Z_i)}{f(Z_i)f(Z_j)}\int K(u)K\Big(\frac{Z_j-Z_i}{w}+u\Big)du\Big).
\end{aligned}
$$

Furthermore, uniformly in i,j,

$$
\mathrm{Var}\big(E(\psi_1|D_i,D_j)\big)\le E\big(E(\psi_1|D_i,D_j)\big)^2=O\Big(\frac{1}{N^2w^{2p}h^p}\Big).
$$

Similar arguments imply that uniformly in $1\le i\le n$ and $1\le k\le N$,

$$
\begin{aligned}
\mathrm{Var}(E(\psi_1|D_i)) &= O\Big(\frac{1}{N^2w^{2p}}\Big), \qquad \mathrm{Var}(E(\psi_1|\widetilde{D}_k))=O\Big(\frac{1}{N^2w^{2p}}\Big), \\
\mathrm{Var}(E(\psi_1|D_i,\widetilde{D}_k)) &= O\Big(\frac{1}{N^2w^ph^{2p}}\Big).
\end{aligned}
$$

Assumptions (C9), (C10) and $0<\lambda<\infty$ together with Lemma 2.6.4 yield

$$
\mathrm{Var}(Nh^{p/2}V_{n21})=O\Big(\frac{2}{(Nw^p)^3}+\frac{2}{(Nw^p)^2}+\frac{4}{(Nw^p)(Nh^p)}+\frac{4nh^p}{(Nw^p)^2}+\frac{nh^p}{N^2w^{2p}}\Big)=o(1).
$$

Moreover, direct calculations show that $E[V_{n21}]=E\psi_1=o(1/N)$. Hence,

$$
\begin{aligned}
E[Nh^{p/2}V_{n21}]^2 &= \mathrm{Var}(Nh^{p/2}V_{n21})+E^2[Nh^{p/2}V_{n21}]=o(1), \qquad (2.31)\\
V_{n21} &= o_p(1/(Nh^{p/2})).
\end{aligned}
$$

In order to analyze the variance of V_{n22}, for $i\ne j, 1\le i,j\le n$, $k\ne l, 1\le k,l\le N$, we define the following symmetric function

$$
\begin{aligned}
\psi_2(D_i, D_j, \widetilde{D}_k, \widetilde{D}_l) &= \frac{I_{ij}K_{h,ij}}{f(Z_i)f(Z_j)}\{K_{w,ik}K_{w,jl}[q(\widetilde{X}_k) - g(Z_i)][q(\widetilde{X}_l) - g(Z_j)] \\
&\quad + K_{w,il}K_{w,jk}[q(\widetilde{X}_l) - g(Z_i)][q(\widetilde{X}_k) - g(Z_j)]\}/2,
\end{aligned}
$$

i.e., ψ_2 is symmetric within the blocks (D_i, D_j) and $(\widetilde{D}_k, \widetilde{D}_l)$. Then we have

$$
V_{n22} = \frac{1}{n(n-1)N^2} \sum_{i \neq j=1}^{n} \sum_{k \neq l=1}^{N} \psi_2(D_i, D_j, \widetilde{D}_k, \widetilde{D}_l).
$$

In order to apply Lemma 2.6.5, we need to calculate the variances of all projections of ψ_2. Computations similar to the above used for analyzing $\text{Var}(V_{n21})$ yield the following facts.

$$
\begin{aligned}
&\text{Var}(\psi_2) = O\Big(\frac{1}{h^p w^{2p}}\Big), \qquad \text{Var}(E(\psi_2|D_i)) = O(w^{4k}), \qquad (2.32)\\
&\text{Var}(E(\psi_2|\widetilde{D}_k)) = O(w^{2k}), \qquad \text{Var}(E(\psi_2|D_i, D_j)) = O\Big(\frac{w^{4k}}{h^p}\Big),\\
&\text{Var}(E(\psi_2|D_i, \widetilde{D}_k)) = O\Big(\frac{w^{2k}}{w^p}\Big), \qquad \text{Var}(E(\psi_2|\widetilde{D}_k, \widetilde{D}_l)) = O\Big(\frac{1}{h^p}\Big),\\
&\text{Var}(E(\psi_2|D_i, D_j, \widetilde{D}_k)) = O\Big(\frac{w^{2k}}{h^p w^p}\Big), \qquad \text{Var}(E(\psi_2|D_i, \widetilde{D}_k, \widetilde{D}_l)) = O\Big(\frac{1}{h^{2p}}\Big).
\end{aligned}
$$

Given the above projection variances, (C8)–(C10), $0 < \lambda < \infty$ and Lemma 2.6.5 imply that

$$
\begin{aligned}
\text{Var}(V_{n22}) &= O\Big(\frac{4}{n^2N^2h^pw^{2p}} + \frac{4w^{4k}}{n} + \frac{4w^{2k}}{N} + \frac{2w^{4k}}{n^2h^p} + \frac{2}{N^2h^p} + \frac{16w^{2k}}{nNw^p} \qquad (2.33)\\
&\quad + \frac{8}{nN^2h^p} + \frac{8}{nN^2h^{2p}} + \frac{8w^{2k}}{n^2Nh^pw^p}\Big)\\
&= \frac{2}{N^2}Var(E(\psi_2|\widetilde{D}_k, \widetilde{D}_l)) + o\Big(\frac{1}{N^2h^p}\Big) = O\Big(\frac{1}{N^2h^p}\Big).
\end{aligned}
$$

In fact, the variance term of $E(\psi_2|\widetilde{D}_k, \widetilde{D}_l)$ dominates the variance of V_{n22}. The facts (2.31) and (2.33) in turn imply that

$$V_{n2} = V_{n22} + o_p(1/(Nh^{p/2})). \tag{2.34}$$

To further investigate V_{n22}, we study the projection of ψ_2 on the validation data set $E(\psi_2|\widetilde{D}_k, \widetilde{D}_l)$. From the format of ψ_2, we only need to study the projection when $\widetilde{Z}_k \in \mathcal{C}$ and $\widetilde{Z}_l \in \mathcal{C}$. In fact, for fixed $\widetilde{Z}_k$, suppose $\widetilde{Z}_k \notin \mathcal{C}$, there is small enough r such that $\mathcal{N}_r(\widetilde{Z}_k) \cap \mathcal{C} = \emptyset$ where $\mathcal{N}_r(\widetilde{Z}_k)$ is the neighborhood of $\widetilde{Z}_k$ within radius of Mr and M is the boundary of the density support of K on each coordinate. It leads to $K_{w,ik} = 0$, then $\psi_2 = 0$. For kernel function with density support on $\mathbb{R}^p$ such as normal density, one can argue with large enough M_K such that the kernel density is arbitrarily small outside of $[-M_K, M_K]^p$. Details are skipped. Hence asymptotically, change of variables and Taylor expansion yield

$$\begin{aligned}
&E(\psi_2|\widetilde{D}_k, \widetilde{D}_l)\\
&= I_{\mathcal{C}}(\widetilde{Z}_k)I_{\mathcal{C}}(\widetilde{Z}_l)\Big\{[q(\widetilde{X}_k) - g(\widetilde{Z}_k)][q(\widetilde{X}_l) - g(\widetilde{Z}_l)] \int K(u)K(v)\widetilde{K}_{h,kl}(u,v)dudv\\
&\quad + w^k C_1 \int K(u)K(v)\widetilde{K}_{h,kl}\{[q(\widetilde{X}_k) - g(\widetilde{Z}_k)]g^{(k)}(\widetilde{Z}_l)v^k\\
&\qquad\qquad + [q(\widetilde{X}_l) - g(\widetilde{Z}_l)]g^{(k)}(\widetilde{Z}_k)u^k\}dudv\Big\} + O_p(w^{2k})\\
&:= \widetilde{\psi}_2(\widetilde{D}_k, \widetilde{D}_l) + R_2'(\widetilde{D}_k, \widetilde{D}_l) + O_p(w^{2k}),
\end{aligned}$$

where $\widetilde{K}_{h,kl}(u,v) = 1/2\{K((\widetilde{Z}_k - \widetilde{Z}_l)/h + w(u-v)/h)/h^p + K((\widetilde{Z}_l - \widetilde{Z}_k)/h + w(u-v)/h)/h^p\}$, $C_1 = 1/k!$. Notice that $\widetilde{\psi}_2(\widetilde{D}_k, \widetilde{D}_l)$ is the leading term, the other two terms are negligible as $w \to 0$. Hence (2.33) can be rewritten as

$$\text{Var}(V_{n22}) = \frac{2}{N^2} Var[\widetilde{\psi}_2(\widetilde{D}_k, \widetilde{D}_l)] + o\Big(\frac{1}{N^2h^p}\Big) = O\Big(\frac{1}{N^2h^p}\Big). \tag{2.35}$$

Next, we will show that V_{n22} is asymptotically equivalent to $\widetilde{V}_{n22}$ defined below:

$$\begin{aligned}\widetilde{V}_{n22} &= \frac{1}{N^2}\sum_{k\neq l}\{\widetilde{\psi}_2(\widetilde{D}_k,\widetilde{D}_l)+R_2'(\widetilde{D}_k,\widetilde{D}_l)\}+O_p(w^{2k}) \\ &:= T_{n2}+T_{n2}'+O_p(w^{2k}).\end{aligned} \tag{2.36}$$

It can be seen that $\widetilde{V}_{n22}$ is the projection of V_{n22} on the validation data. Hence $E\{(V_{n22}-\widetilde{V}_{n22})\widetilde{V}_{n22}\}=0$, and

$$\mathrm{Var}(V_{n22}-\widetilde{V}_{n22})=\mathrm{Var}(V_{n22})-\mathrm{Var}(\widetilde{V}_{n22}). \tag{2.37}$$

Now we will prove that T_{n2} dominates $\widetilde{V}_{n22}$ by showing the asymptotic properties of each term. The last term in (2.36) is negligible with asymptotic rate $Nh^{p/2}$ since $Nh^{p/2}w^{2k}\to 0$ by assumption (C10).

First, note that T_{n2} is a degenerate U statistic. After verifying the conditions in Theorem 1 of Hall (1984), we apply the theorem and obtain that

$$\frac{NT_{n2}}{\{2E\widetilde{\psi}_2^2(\widetilde{D}_1,\widetilde{D}_2)\}^{1/2}}\to_d \mathcal{N}_1(0,1). \tag{2.38}$$

Since $E(\widetilde{\psi}_2(\widetilde{D}_1,\widetilde{D}_2))=0$, (2.32) further implies that $E\widetilde{\psi}_2^2(\widetilde{D}_1,\widetilde{D}_2)=\mathrm{Var}[\widetilde{\psi}_2(\widetilde{D}_1,\widetilde{D}_2)]=O(1/h^p)$. Moreover, (2.38) implies that

$$Var(T_{n2})=\frac{2}{N^2}\mathrm{Var}(\widetilde{\psi}_2(\widetilde{D}_1,\widetilde{D}_2))+o\Big(\frac{1}{N^2h^p}\Big)=O\Big(\frac{1}{N^2h^p}\Big). \tag{2.39}$$

Second, note that T_{n2}' is a non-degenerate U statistic with mean 0. By applying the central limit theorem for non-degenerate U statistics presented in Serfling (1981), we obtain

$$\frac{\sqrt{N}T'_{n2}}{\{4\mathrm{Var}(E(R'_2|\widetilde{D}_1))\}^{1/2}} \to_d \mathcal{N}_1(0,1).$$

Straightforward calculation indicates that $\mathrm{Var}(E(R'_2|\widetilde{D}_1)) = O(w^{2k})$. Then, as $n \wedge N \to 0$,

$$\mathrm{Var}(Nh^{p/2}T'_{n2}) = O(N^2h^pw^{2k}/N) = O(Nh^pw^{2k}) = o(1),$$

under the condition that $nh^{p/2}w^{2k} \to 0$ and $0 < \lambda < \infty$. Therefore

$$T'_{n2} = o_p(1/(Nh^{p/2})).$$

This fact combined with (2.39) and (2.36) yield

$$\mathrm{Var}(\widetilde{V}_{n22}) = \frac{2}{N^2}\mathrm{Var}[\widetilde{\psi}_2(\widetilde{D}_1, \widetilde{D}_2)] + o\Big(\frac{1}{N^2h^p}\Big). \tag{2.40}$$

The results (2.35), (2.37) and (2.40) together imply that

$$\mathrm{Var}(V_{n22} - \widetilde{V}_{n22}) = o\Big(\frac{1}{N^2h^p}\Big).$$

Hence

$$V_{n2} = V_{n22} + o_p(1/(Nh^{p/2})) = \widetilde{V}_{n22} + o_p(1/(Nh^{p/2})) = T_{n2} + o_p(1/(Nh^{p/2})). \tag{2.41}$$

Given the asymptotic results in (2.38), we have

$$\begin{aligned} h^p E\widetilde{\psi}_2(\widetilde{D}_1, \widetilde{D}_2) \;\to\; & \int\Big\{\int K(u)K(v)[K(s+c(u-v))+K(-s+c(u-v))]/2dudv\Big\}^2 ds \\ & \times\int I_{\mathcal{C}}(x)(\widetilde{\gamma}^2(x))^2 f^2(x)dx = \tau_2. \end{aligned}$$

By connecting the above limiting variance with (2.38) and (2.41), eventually we obtain that

$$Nh^{p/2}T_{n2} \to_1 \mathcal{N}_1(0, 2\tau_2), \text{ hence } Nh^{p/2}V_{n2} \to_d \mathcal{N}_1(0, 2\tau_2). \tag{2.42}$$

This in turn completes the proof of Lemma 2.6.2. □

For the next lemma recall the decomposition (2.13).

Lemma 2.6.3. *Under assumptions (C1)–(C10) and* $0 < \lambda < \infty$*, the following holds when* H_0 *is true.*

$$V_{n3} = o_p(1/(nh^{p/2})); \quad U_{nj} = o_p(1/(nh^{p/2})), \; j = 1, 2, 3.$$

Proof. The proof of the claim about V_{n3} is similar to that of Lemma 6.2. Rewrite

$$\begin{aligned} V_{n3} &= \frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}\delta_i\delta_j \\ &= \frac{1}{n(n-1)}\sum_{i\neq j}\frac{I_{ij}K_{h,ij}}{\tilde{f}(Z_i)\tilde{f}(Z_j)}[W_N(Z_i,\theta_0)-W_N(Z_i,\hat{\theta})][W_N(Z_j,\theta_0)-W_N(Z_j,\hat{\theta})] \\ &= \frac{1}{n(n-1)N^2}\sum_{i\neq j}\sum_{k,l}\frac{I_{ij}K_{h,ij}}{\tilde{f}(Z_i)\tilde{f}(Z_j)}K_{w,ik}K_{w,jl}[q(\widetilde{X}_k,\hat{\theta})-q(\widetilde{X}_k,\theta_0)] \\ &\qquad\qquad\times[q(\widetilde{X}_l,\hat{\theta})-q(\widetilde{X}_l,\theta_0)] \\ &= \widetilde{V}_{n3} + o_p(\widetilde{V}_{n3}), \end{aligned}$$

where

$$\widetilde{V}_{n3} = \frac{1}{n(n-1)} \sum_{i \neq j} \sum_{k,l} \frac{I_{\mathcal{C}}(Z_i) I_{\mathcal{C}}(Z_j) K_{h,ij}}{f(Z_i) f(Z_j)} [q(\widetilde{X}_k, \hat{\theta}) - q(\widetilde{X}_k, \theta_0)][q(\widetilde{X}_l, \hat{\theta}) - q(\widetilde{X}_l, \theta_0)].$$

Furthermore, $\widetilde{V}_{n3}$ is decomposed as the sum of the following two terms.

$$\begin{aligned} \widetilde{V}_{n3} &= \frac{1}{n(n-1)N^2} \sum_{i \neq j} \sum_{k=1}^{N} \frac{I_{ij} K_{h,ij}}{f(Z_i) f(Z_j)} K_{w,ik} K_{w,jk} [q(\widetilde{X}_k, \hat{\theta}_n) - q(\widetilde{X}_k, \theta_0)]^2 \\ &\quad + \frac{1}{n(n-1)N^2} \sum_{i \neq j} \sum_{k \neq l} \frac{I_{ij} K_{h,ij}}{f(Z_i) f(Z_j)} K_{w,ik} K_{w,jl} \\ &\qquad \times [q(\widetilde{X}_k, \hat{\theta}_n) - q(\widetilde{X}_k, \theta_0)][q(\widetilde{X}_l, \hat{\theta}_n) - q(\widetilde{X}_l, \theta_0)] \\ &= V_{n31} + V_{n32}, \quad \text{say.} \end{aligned}$$

Similar to the analysis of V_{n21}, define the symmetric function

$$\phi_1(D_i, D_j, \widetilde{D}_k) = \frac{I_{ij} K_{h,ij}}{f(Z_i) f(Z_j)} K_{w,ik} K_{w,jk} [q(\widetilde{X}_k, \hat{\theta}_n) - q(\widetilde{X}_k, \theta_0)]^2.$$

Because $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ and by the Taylor expansion, we have

$$q(\widetilde{X}_k, \hat{\theta}_n) - q(\widetilde{X}_k, \theta_0) = O_p(1/\sqrt{n}).$$

Then we can easily check that $V_{n31} = o_p(1)$. Following the routine argument showed in the proof of Lemma 3.3d in Zheng (1996), we obtain that $V_{n32} = o_p(1)$ under H_0 and (C2), (C3), (C5), (C7)–(C9).

Similarly, U_{n1} can be written as

$$
\begin{aligned}
U_{n1} &= \frac{1}{n(n-1)N^2}\sum_{i\neq j,k,l} I_{ij}K_{h,ij}K_{w,ik}K_{w,jl}\eta_i(q(\widetilde{X}_l)-g(Z_j)) \\
&= \frac{1}{n(n-1)N^2}\sum_{i\neq j,k} I_{ij}K_{h,ij}K_{w,ik}K_{w,jk}\eta_i(q(\widetilde{X}_k)-g(Z_j)) \\
&\quad + \frac{1}{n(n-1)N^2}\sum_{i\neq j,k\neq l} I_{ij}K_{h,ij}K_{w,ik}K_{w,jl}\eta_i(q(\widetilde{X}_l)-g(Z_j)) \\
&= U_{n11}+U_{n12}, \qquad \text{say.}
\end{aligned}
$$

Analogous to the analysis of V_{n1} and V_{n2}, similar results can be derived for U_{n1} as follows:

$$
U_{n11} = o_p(1/(nh^{p/2})),
$$

and U_{n12} can be formulated as a non-degenerate U statistic with the kernel function

$$
\begin{aligned}
\phi_2(D_i, D_j, \widetilde{D}_k, \widetilde{D}_l) &= I_{ij}K_{h,ij}K_{w,ik}K_{w,jl}[\eta_i(q(\widetilde{X}_l)-g(Z_j))+\eta_j(q(\widetilde{X}_k)-g(Z_i))]/4 \\
&\quad + I_{ij}K_{h,ij}K_{w,il}K_{w,jk}[\eta_i(q(\widetilde{X}_k)-g(Z_j))+\eta_j(q(\widetilde{X}_l)-g(Z_i))]/4.
\end{aligned}
$$

By the central limit theorem of non-degenerate U statistics, we can see that $\sqrt{n}U_{n12} = O_p(w^k)$. Thus

$$
nh^{p/2}U_{n1} = O_p\Big(\sqrt{nh^p}\cdot\sqrt{n}U_{n12}\Big) = O_p(\sqrt{nh^pw^{2k}}) = o_p(1)
$$

under the assumption (C10). The proofs of the claims pertaining to U_{n2} and U_{n3} are similar. Details are omitted for the sake of brevity. □

Proof of Theorem 2.3.4. Similar to the proof of V_n under H_0, we can show that under H_a

$$\begin{aligned}
& V_n \qquad (2.43)\\
= & \frac{1}{n(n-1)}\sum_{i\neq j} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)K_{h,ij}\hat{\eta}_i\hat{\eta}_j\\
= & \frac{1}{n(n-1)}\sum_{i\neq j} I_{\mathcal{C}}(Z_i)I_{\mathcal{C}}(Z_j)K_{h,ij}\bar{\eta}_i\bar{\eta}_j\\
& +\frac{1}{N^2}\sum_{k\neq l} I_{\mathcal{C}}(\widetilde{Z}_k)I_{\mathcal{C}}(\widetilde{Z}_l)\Big\{[q(\widetilde{X}_k)-g(\widetilde{Z}_k)][q(\widetilde{X}_l)-g(\widetilde{Z}_l)]\int K(u)K(v)\widetilde{K}_{h,kl}(u,v)dudv\Big\}\\
& +o_p(1/(nh^{p/2}))\\
= & \bar{T}_{n1}+\bar{T}_{n2}+o_p(1/(nh^{p/2})), \qquad \text{say.}
\end{aligned}$$

One can verify that $\bar{T}_{n1}$ and $\bar{T}_{n2}$ are the leading terms of V_{n1} and V_{n2}, respectively, as derived in Lemma 2.6.2. Rewrite $\bar{\eta}_i = \eta_i + b_n\, E(a(X_i)|Z_i)$, where $E(\eta_i|Z_i) = 0$.

$$\begin{aligned}
\bar{T}_{n1} &= \frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}\bar{\eta}_i\bar{\eta}_j\\
&= \frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}[\eta_i + b_n\, E(a(X_i)|Z_i)][\eta_j + b_n\, E(a(X_j)|Z_j)]\\
&= \frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}\eta_i\eta_j + b_n\frac{2}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}\eta_i E[a(X_j)|Z_j]\\
&\quad +b_n^2\frac{1}{n(n-1)}\sum_{i\neq j} I_{ij}K_{h,ij}E[a(X_i)|Z_i]E[a(X_j)|Z_j]\\
&:= W_1 + b_nW_2 + b_n^2W_3.
\end{aligned}$$

W_1 is a degenerate two sample U statistic, hence

$$nh^{p/2}W_1 \to_d \mathcal{N}_1(0, 2\tau_1).$$

After symmetrization, W_2 can be written as a non-degenerate U statistic, hence $\sqrt{n}W_2 = O_p(1)$, furthermore,

$$nh^{p/2}b_nW_2 = h^{p/4}(\sqrt{n}W_2) \to_p 0.$$

A similar argument as (2.15) indicates that

$$W_3 \to_p E\{I_{\mathcal{C}}(Z)E[a(X)|Z]^2 f(Z)\}.$$

Hence

$$nh^{p/2}\bar{T}_{n1} \to_d \mathcal{N}_1(E\{I_{\mathcal{C}}(Z)E[a(X)|Z]^2 f(Z)\}, 2\tau_1). \tag{2.44}$$

As for $\bar{T}_{n2}$, the result of T_{n2} in (2.42) still holds since $\bar{T}_{n2}$ only involves the validation data and it is irrelevant to the hypothesis of the regression model, i.e.,

$$nh^{p/2}\bar{T}_{n2} \to_d \mathcal{N}_1(0, 2\tau_2). \tag{2.45}$$

Note that $\bar{T}_{n1}$ and $\bar{T}_{n2}$ are independent since they are constructed based on independent samples. Combining (2.43), (2.44) and (2.45), we obtain that

$$nh^{p/2}V_n \to_d \mathcal{N}_1(E\{I_{\mathcal{C}}(Z)E[a(X)|Z]^2 f(Z)\}, 2\tau_1 + (2\tau_2)/\lambda^2).$$

This completes the proof of Theorem 2.3.4. □

Lemma 2.6.4. *Let* $\{D_i, i = 1, \cdots, n\}$ *be a set of i.i.d. r.v.'s and* $\{\widetilde{D}_k, k = 1, \cdots, N\}$ *be* *another set of i.i.d. r.v.'s, which is independent of $\{D_i\}$. Define the two sample U statistic*

$$T = \frac{1}{n(n-1)N} \sum_{i \neq j=1}^{n} \sum_{k=1}^{N} \varphi_n(D_i, D_j, \widetilde{D}_k),$$

where φ_n is a symmetric function with regard to permutation of (D_i, D_j) and square integrable for each n. Then

$$\begin{aligned} \mathrm{Var}(T) &= O\Big(\frac{2}{n^2 N}\mathrm{Var}(\varphi_n) + \frac{4}{n}\mathrm{Var}(E(\varphi_n|D_1)) + \frac{1}{N}\mathrm{Var}(E(\varphi_n|\widetilde{D}_1)) \\ &\quad + \frac{2}{n^2}\mathrm{Var}(E(\varphi_n|D_1, D_2)) + \frac{4}{nN}\mathrm{Var}(E(\varphi_n|D_1, \widetilde{D}_1))\Big). \end{aligned} \tag{2.46}$$

Proof. Algebra shows that

$$\begin{aligned} &\mathrm{Var}(Nn(n-1)T) \\ &= E\Big\{\sum_{i\neq j,k} [\varphi_n(D_i, D_j, \widetilde{D}_k) - E\varphi_n(D_i, D_j, \widetilde{D}_k)]\Big\}^2 \\ &= \sum_{i\neq j,k}\sum_{s\neq t,l} E\Big\{[\varphi_n(D_i, D_j, \widetilde{D}_k) - E\varphi_n(D_i, D_j, \widetilde{D}_k)][\varphi_n(D_s, D_t, \widetilde{D}_l) - E\varphi_n(D_s, D_t, \widetilde{D}_l)]\Big\} \\ &= \Big\{\sum_{\{s,t\}=\{i,j\},l=k} + \sum_{\{s,t\}=\{i,j\},k\neq l} + 4\sum_{s=i,t\neq j,k=l} + 4\sum_{s=i,t\neq j,k\neq l} \\ &\quad + \sum_{s\neq i,t\neq j,k=l} + \sum_{s\neq i,t\neq j,k\neq l}\Big\} E\{[\varphi_n(D_i, D_j, \widetilde{D}_k) - E\varphi_n][\varphi_n(D_s, D_t, \widetilde{D}_l) - E\varphi_n]\} \\ &= 2n(n-1)N\mathrm{Var}(\varphi_n) + 2n(n-1)N(N-1)\mathrm{Var}(E(\varphi_n|D_1, D_2)) \\ &\quad + 4n(n-1)(n-2)N\mathrm{Var}(E(\varphi_n|D_1, \widetilde{D}_1)) + 4n(n-1)(n-2)N(N-1)\mathrm{Var}(E(\varphi_n|D_1)) \\ &\quad + n(n-1)(n-2)(n-3)N\mathrm{Var}(E(\varphi_n|\widetilde{D}_1)). \end{aligned}$$

The claim (2.46) follows from this identity upon dividing both sides by $\big(Nn(n-1)\big)^2$ and using the fact that $(n-k)/n \to 1$, and $(N-k)/N \to 1$, for $k = 1, 2, 3$. □

Furthermore, define

$$S = \frac{1}{n(n-1)} \sum_{i \neq j=1}^{n} \frac{1}{N(N-1)} \sum_{k \neq l=1}^{N} \psi_n(D_i, D_j, \widetilde{D}_k, \widetilde{D}_l),$$

where ψ_n is square integrable and symmetric with regard to permutation of (D_i, D_j) as well as $(\widetilde{D}_k, \widetilde{D}_l)$, i.e., $\psi_n(D_i, D_j, \cdot, \cdot) = \psi_n(D_j, D_i, \cdot, \cdot)$ and $\psi_n(\cdot, \cdot, \widetilde{D}_k, \widetilde{D}_l) = \psi_n(\cdot, \cdot, \widetilde{D}_l, \widetilde{D}_k)$, for each n. An argument similar to the one used in Lemma 2.6.4 yields the following lemma.

Lemma 2.6.5. *Suppose $\{D_i, 1 \leq i \leq n\}$ and $\{\widetilde{D}_k, 1 \leq k \leq N\}$ are the two independent random samples and S is the two sample statistic defined above. Then*

$$\begin{aligned} \mathrm{Var}(S) &= O\Big(\frac{4}{n^2N^2}\mathrm{Var}(\psi_n) + \frac{4}{n}\mathrm{Var}(E(\psi_n|D_i)) + \frac{4}{N}\mathrm{Var}(E(\psi_n|\widetilde{D}_k)) \\ &\quad + \frac{2}{n^2}\mathrm{Var}(E(\psi_n|D_i, D_j)) + \frac{2}{N^2}\mathrm{Var}(E(\psi_n|\widetilde{D}_k, \widetilde{D}_l)) + \frac{16}{nN}\mathrm{Var}(E(\psi_n|D_i, \widetilde{D}_k)) \\ &\quad + \frac{8}{nN^2}\mathrm{Var}(E(\psi_n|D_i, \widetilde{D}_k, \widetilde{D}_l)) + \frac{8}{n^2N}\mathrm{Var}(E(\psi_n|D_i, D_j, \widetilde{D}_k))\Big). \end{aligned}$$

Chapter 3

Minimum distance model checking in Berkson models

3.1 Introduction

In statistical data analysis, the data is often collected subject to measurement error. One typical way to treat the measurement error is the errors-in-variables model which assumes that the real observation Z is a surrogate of the true unobserved variable X, i.e., $Z = X + \eta$, where η is the measurement error variable. Regression models with measurement error in covariates has received broad attention in the literature over the last century. In the last three decades it has been the focus of numerous researchers, as is evidenced in the three monographs by Fuller (1987), Cheng and Van Ness (1999) and Carroll, Ruppert, Stefanski and Crainiceanu (2006), and the references therein. However, as Berskon (1950) argued that in many situations it is more appropriate to treat the true unobserved variable X as the observed variable Z plus an error, i.e., $X = Z + \eta$. For instance, in economics, the household income is usually not precisely collected due to the survey design or data sensitivity. It was described in Kim, Chao and Härdle (2016) that when the income data were collected by asking individuals which salary range categories they belong to, such as between 5,000 USD and 9,999 USD, then the midpoint of the range interval was used in analysis. In this case, it is sensible to assume that the true income fluctuates around the midpoint observation subject to errors. Another example is an individual's exposure to some contaminant in epidemiological study, for instance, the atmospheric particulate matter that have a diameter less than 2.5 micrometers (PM2.5). Usually the concentration of PM2.5 in an area is reported hourly or daily as an average measurement, however, the true exposure for an individual relies on the specific location and the time of the day. This type of data also favors the Berkson error model. More examples can be found in Du et al. (2011) and Carroll et al. (2006).

Proceeding a bit more precisely, in the Berkson measurement error regression model of interest here one has the triple X, Y, Z, obeying the relations

$$Y = \mu(X) + \varepsilon, \qquad X = Z + \eta. \tag{3.1}$$

Here Y is a scalar response variable and ε is an error variable with $E\varepsilon = 0$. The random vectors X, Z, η are p-dimensional, with X being the true unobservable covariate vector, Z representing an observation on X and η denoting the measurement error having $E\eta = 0$. For identifiability reasons, the three r.v.'s ε, X, η are assumed to be mutually independent. Thus $\mu(x) = E(Y|X = x)$, for all $x \in \mathbb{R}^p$.

Let $\Theta \subset \mathbb{R}^q$ be a compact set, $\{m_\theta(x); \theta \in \Theta, x \in \mathbb{R}^p\}$ be a family of given functions and $\mathcal{C}$ be a compact subset in $\mathbb{R}^p$. The problem of interest here is to test

$$H_0 : \mu(x) = m_{\theta_0}(x), \quad \text{for some } \theta_0 \in \Theta \text{ and all } x \in \mathcal{C}, \quad \text{versus}$$
$$H_1 : H_0 \text{ is not true,}$$

based on the primary sample $\{(Z_i, Y_i), i = 1, ..., n\}$ and an independent validation sample

$\{(\widetilde{Z}_k, \widetilde{X}_k), k = 1, ..., N\}$, all satisfying (3.1). Then the empirical version of η are naturally obtained by $\widetilde{\eta}_k := \widetilde{X}_k - \widetilde{Z}_k, 1 \leq k \leq N$.

The literature contains several references that address the estimation of the underlying parameters in the model (3.1). In the case $\mu(x)$ is linear, Berkson (1950) showed that the ordinary least squares estimators continue to be unbiased and consistent for the underlying parameters. For polynomial regression, Huwang and Huang (2000) use the method of moments based on the first two conditional moments of Y, given Z, when Z, ε, η are Gaussian, to produce consistent estimators of the underlying parameters. Relaxing the normality error distribution assumption to a parametric density family, Wang (2004) developed a minimum distance approach based on the first two conditional moments of the response variable to consistently estimate more general parametric regression functions. In the case the measurement error density f_η is known, Delaigle, Hall and Qiu (2006) constructed nonparametric estimators of $\mu(x)$ by means of trigonometric series and deconvolution techniques. In the case f_η is unknown but validation data is available, Du et al. (2011) used integrated local linear smoothing and Fourier transformation to formulate a nonparametric estimator of $\mu(x)$. Schennach (2013) obtained a sieve-based nonparametric regression estimator with the help of instrumental variable without assuming f_η to be known.

Relatively, the literature is scant on the above testing problem. Koul and Song (2009) are the first authors to address this problem. Assuming f_η is known, they proposed parameter estimation by minimizing an integrated square distance between a nonparametric regression function estimator and the model being fitted and then utilized the minimized distance to implement the hypothesis test. In the current chapter, we extend this methodology to the case when f_η is unknown, but when validation data is available.

A surprising finding is that the asymptotic distributions of the minimum distance (m.d.)

test statistics in the case of unknown f_η is the same as in the case of known f_η. The asymptotic distributions of the corresponding m.d. estimators of the null model parameters are affected by not knowing f_η in general. Exceptions are provided by the linear models when the set $\mathcal{C}$ and the integrating measure used in the definition of the above mentioned distances are symmetric around zero.

This chapter is organized as follows. Section 3.2 describes the proposed m.d. estimators and test statistics and the needed assumptions for the derivation of their consistency and asymptotic normality. Section 3.3 establishes the consistency and asymptotic normality of the m.d. estimators while in Section 3.4 we state the main results of the proposed tests under the null and certain fixed alternative hypotheses and provide sketches of the proofs. It is worth mentioning that the variation in validation data contributes to the asymptotic distributions of the proposed m.d. estimators of the null model parameters but not to the asymptotic distributions of the m.d. test statistics. Section 3.5 reports findings of a Monte Carlo study that assesses some finite sample properties of an estimator and a test in the proposed classes of these inference procedures. Some of the proofs are relegated to the last Section 3.6 of the chapter.

3.2 A class of tests

This section describes a class of the proposed tests and estimators of the null model parameters along with the needed assumptions. To overcome the difficulty created by not observing X we use the calibration idea as used in Koul and Song (2009). Accordingly, assume $E|\mu(X)| < \infty$, $E|m_\theta(X)| < \infty$, for all $\theta \in \Theta$ and $z \in \mathbb{R}^p$, define

$$\begin{aligned} H(z) &:= E\big[\mu(X)\big|Z = z\big] = \int \mu(x) f_\eta(x-z)dx = \int \mu(y+z) f_\eta(y)dy, \\ H_\theta(z) &:= E[m_\theta(X)|Z = z] = \int m_\theta(x) f_\eta(x-z)dx = \int m_\theta(y+z) f_\eta(y)dy. \end{aligned}$$

Then the original model can be transformed to

$$Y = H(Z) + \xi, \qquad E(\xi|Z) = 0, \tag{3.2}$$

and the hypothesis testing becomes

$$\begin{aligned} &H_0' : H(z) = H_{\theta_0}(z), \text{ for some } \theta_0 \in \Theta \text{ and all } z \in \mathcal{C}, \text{ vs.} \\ &H_1' : H_0' \text{ is not true.} \end{aligned}$$

To proceed further, let $w \equiv w_n = c(\log n/n)^{1/(p+4)}$, $c > 0$, and $h \equiv h_n$ be two bandwidth sequences associated with sample sizes n and N, K be a density kernel and G be a nondecreasing right continuous real valued function on $\mathbb{R}$ and define

$$K_{hi}(z) = \frac{1}{h^p} K\Big(\frac{z - Z_i}{h}\Big),\ \hat{f}_w(z) = \frac{1}{n}\sum_{i=1}^n K_{wi}(z),\ \widehat{H}_\theta(z) = N^{-1}\sum_{k=1}^N m_\theta(z + \widetilde{\eta}_k), \tag{3.3}$$

$$M_n(\theta) = \int \Big[\frac{1}{n\hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z)[Y_i - H_\theta(Z_i)]\Big]^2 dG(z), \quad \tilde{\theta}_n = \text{argmin}_\theta M_n(\theta),$$

$$\widehat{M}_n(\theta) = \int \Big[\frac{1}{n\hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z)[Y_i - \widehat{H}_\theta(Z_i)]\Big]^2 dG(z), \quad \hat{\theta}_n = \text{argmin}_\theta \widehat{M}_n(\theta),$$

$$W_n(\theta) = \int \Big[\frac{1}{n\hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z)[\widehat{H}_\theta(Z_i) - H_\theta(Z_i)]\Big]^2 dG(z).$$

Note that the density estimator $\hat{f}_w$ is based on a bandwidth w that is different from the bandwidth h employed in the numerator of the regression function estimator. This plausible scheme was proposed in Koul and Ni (2004) (KN) in order to have an $nh^{p/2}$-consistent estimator of the asymptotic bias in $M_n(\tilde{\theta}_n)$.

In the case f_η is known then H_θ is a known parametric function and Koul and Song (2009) (KS) proposed the minimum distance testing procedure based on $M_n(\tilde{\theta}_n)$. However, this method is not feasible without the knowledge of f_η, which renders the regression function H_θ to be unknown also. But, with the availability of validation sample $\{\widetilde{X}_k, \widetilde{Z}_k\}$, where $\widetilde{\eta}_k := \widetilde{X}_k - \widetilde{Z}_k, 1 \le k \le N$ is a random sample from f_η, we are able to estimate $H_\theta(z)$ by $\widehat{H}_\theta(z)$ defined above. This then leads to the class, one for each G and K, of m.d. test statistics $\widehat{M}_n(\hat{\theta}_n)$.

We shall now present the needed assumptions for establishing the consistency and asymptotic normality of $\hat{\theta}_n$ and $\widehat{M}_n(\hat{\theta}_n)$. Many of these assumptions are the same as in KS. Define, for $x, y \in \mathbb{R}^p$ and $\theta \in \Theta$,

$$\sigma_\theta(x,y) := \mathrm{Cov}\big(m_\theta(x+\eta), m_\theta(y+\eta)\big), \quad \sigma_\theta^2(x) := \sigma_\theta(x,x) = \mathrm{Var}(m_\theta(x+\eta)).$$

(A1) $\{(Y_i, Z_i), Z_i \in \mathbb{R}^p, i = 1, ..., n\}$ is an i.i.d. sample with regression function $H(z) = E(Y|Z=z)$ satisfying $\int H^2 dG < \infty$, where G is a σ-finite measure with continuous Legesgue density g on $\mathcal{C}$ while $\{(\widetilde{Z}_k, \widetilde{X}_k), \widetilde{Z}_k \in \mathbb{R}^p, \widetilde{X}_k \in \mathbb{R}^p, k = 1, ..., N\}$ is an i.i.d. sample from Berkson measurement error model $X = Z + \eta$.

(A2) $0 < \sigma_\varepsilon^2 := \mathrm{Var}(\varepsilon) < \infty$, $\tau^2(z) = E[(m_{\theta_0}(X) - H_{\theta_0}(Z))^2 | Z = z]$ is a.e. (G) continuous on $\mathcal{C}$.

(A3) Both $E|\varepsilon|^{2+\delta}$ and $E|(m_{\theta_0}(X) - H_{\theta_0}(Z)|^{2+\delta}$ are finite for some $\delta > 0$.

(A4) Both $E|\varepsilon|^4$ and $E|(m_{\theta_0}(X) - H_{\theta_0}(Z)|^4$ are finite.

(A5) $\int \sigma_\theta^2(z)dG(z) < \infty$, for all $\theta \in \Theta$.

(F1) The density f_Z is uniformly continuous and bounded away from 0 in $\mathcal{C}$.

(F2) The density f_Z is twice continuously differentiable in $\mathcal{C}$.

(H1) $m_\theta(x)$ is a.e. continuous in x, for every $\theta \in \Theta$.

(H2) The parametric function family $H_\theta(z)$ is identifiable with respect to θ, i.e, $H_{\theta_1}(z) = H_{\theta_2}(z)$ a.e. in z implies $\theta_1 = \theta_2$.

(H3) For some positive continuous function r on $\mathcal{C}$, and for some $0 < \beta \leq 1$, $|H_{\theta_1}(z) - H_{\theta_2}(z)| \leq \|\theta_1 - \theta_2\|^\beta r(z)$, for all $\theta_1, \theta_2 \in \Theta$ and $z \in \mathcal{C}$.

(H4) For each x, $m_\theta(x)$ is differentiable with respect to θ in a neighborhood of θ_0 with the derivative vector $\dot{m}_\theta(x)$ such that for every sequence $0 < \delta_n \to 0$,

$$\sup_{i,\theta} \frac{\left|N^{-1}\sum_{k=1}^N [m_\theta(Z_i + \widetilde{\eta}_k) - m_{\theta_0}(Z_i + \widetilde{\eta}_k) - (\theta - \theta_0)^T \dot{m}_{\theta_0}(Z_i + \widetilde{\eta}_k)]\right|}{\|\theta - \theta_0\|} = o_p(1),$$

where the supremum is taken over $1 \leq i \leq n, \|\theta - \theta_0\| \leq \delta_n$.

(H5) The vector function $\dot{m}_{\theta_0}(x)$ is continuous in $x \in \mathcal{C}$ and for every $\epsilon > 0$, there are n_ϵ and N_ϵ such that for every $0 < a < \infty$, and for all $n > n_\epsilon, N > N_\epsilon$,

$$P\Big(\max_{1\leq i\leq n, 1\leq k\leq N, (nh^p)^{1/2}\|\theta-\theta_0\|\leq a} h^{-p/2}\|\dot{m}_\theta(Z_i + \widetilde{\eta}_k) - \dot{m}_{\theta_0}(Z_i + \widetilde{\eta}_k)\| \geq \epsilon\Big) \leq \epsilon.$$

(H6) $\int \|\dot{H}_{\theta_0}\|^2 dG < \infty$ and $\Sigma_0 = \int \dot{H}_{\theta_0}\dot{H}_{\theta_0}^T dG$ is positive definite.

(K) The density kernel K is positive symmetric and square integrable on $[-1,1]^p$.

(W1) $nh^{2p} \to \infty$ and $N/n \to \lambda$, $\lambda > 0$.

(W2) $h \sim n^{-a}$, where $0 < a < \min(1/2p, 4/(p(p+4)))$.

We state some facts that will be often used in the proofs below. Note that (H4) implies that for every $0 < a < \infty$,

$$\sup_{i,\theta} \frac{\left| N^{-1} \sum_{k=1}^{N} \left[m_\theta(Z_i + \widetilde{\eta}_k) - m_{\theta_0}(Z_i + \widetilde{\eta}_k) - (\theta - \theta_0)^T \dot{m}_{\theta_0}(Z_i + \widetilde{\eta}_k) \right] \right|}{\|\theta - \theta_0\|} = o_p(1), \quad (3.4)$$

where the supremum is taken over $1 \leq i \leq n, (nh^p)^{1/2}\|\theta - \theta_0\| \leq a$. From Mack and Silverman (1982) we obtain that under (F1), (K1), (W1) and (W2),

$$\sup_{z \in \mathcal{C}} |\hat{f}_h(z) - f_Z(z)| = o_p(1), \quad \sup_{z \in \mathcal{C}} |\hat{f}_w(z) - f_Z(z)| = o_p(1), \quad \sup_{z \in \mathcal{C}} \left| \frac{f_Z^2(z)}{\hat{f}_w^2(z)} - 1 \right| = o_p(1). \quad (3.5)$$

Theorem 2.2 part (2) in Bosq (1998) yields that under assumptions (F2) and (K1),

$$(\log_k n)^{-1} (n/\log n)^{2/(p+4)} \sup_{z \in \mathcal{C}} |\hat{f}_w(z) - f_Z(z)| \to 0, \text{ a.s., } \forall \text{ integer } k > 0. \quad (3.6)$$

We also recall the following facts from KN and KS. Let $d\varphi = f_Z^{-2} dG$, $d\hat{\varphi} = \hat{f}_w^{-2} dG$. For any continuous function $\alpha(z)$ with $\int |\alpha(z)| d\varphi(z) < \infty$, (3.5) implies

$$\left| \int \alpha(z) d\hat{\varphi}(z) - \int \alpha(z) d\varphi(z) \right| \leq \sup_{z \in \mathcal{C}} \left| \frac{f_Z^2(z)}{\hat{f}_w^2(z)} - 1 \right| \int |\alpha(z)| d\varphi(z) = o_p\Big(\int |\alpha(z)| d\varphi(z) \Big).$$

Hence

$$\int \alpha(z) d\hat{\varphi}(z) = \int \alpha(z) d\varphi(z) + o_p\Big(\int |\alpha(z)| d\varphi(z) \Big). \quad (3.7)$$

From (3.9) in KN, for any function $\alpha(z)$ as above, (F1), (K1) and (W1) imply

$$\int E\Big\{\frac{1}{n}\sum_{i=1}^{n} K_h(z-Z_i)\alpha(Z_i)\Big\}^2 d\varphi(z) = \int \alpha^2 dG + o(1) = O(1). \tag{3.8}$$

In the sequel, we shall not exhibit the set $\mathcal{C}$ in the integrals. All the integrals with respect to the measure G are supposed to be over this set, unless specified otherwise.

3.3 Estimation of θ_0

In this section, we establish the consistency and asymptotic normality of $\hat{\theta}_n$ under H_0.

To begin with, consider the following decomposition that shows a connection between $\widehat{M}_n(\theta)$ and $M_n(\theta)$, where $W_n(\theta)$ is as in (3.3).

$$\begin{aligned} \widehat{M}_n(\theta) &= \int \Big[\frac{1}{n\hat{f}_w(z)}\sum_{i=1}^{n} K_{hi}(z)[Y_i - H_\theta(Z_i) + H_\theta(Z_i) - \widehat{H}_\theta(Z_i)]\Big]^2 dG(z) \\ &= M_n(\theta) + W_n(\theta) + 2R_n(\theta), \end{aligned} \tag{3.9}$$

where $R_n(\theta)$ is the cross product term. We can see that the validation data is involved through the extra terms W_n and R_n. The following lemma about W_n is found to be useful in deriving various results in the sequel. Its proof is given in the last Section 3.6 of the chapter. Let K_1 be as in (2.11) and let

$$\gamma(\theta) := \int \sigma_\theta^2(x,y)dG(x)dG(y), \qquad A_N(\theta) = \frac{1}{N}\int \sigma_\theta^2(z)dG(z).$$

Lemma 3.3.1. *Suppose (A1), (A2), (A5), (F1), (H1), (K), and (W1) hold. Then for every*

$\theta \in \Theta$ *for which* $\mu(x) = m_\theta(x), x \in \mathcal{C}$, *we have*

$$N\left(W_n(\theta) - A_N(\theta)\right) \to \mathcal{N}_1(0, \gamma(\theta)). \tag{3.10}$$

3.3.1 Consistency of $\hat{\theta}_n$

We first establish the consistency of the proposed m.d. parameter estimators $\hat{\theta}_n$. Many details below are similar to those in KN and KS. Recall $\mu(x) = E(Y|X = x)$. Let $H(z) = E(\mu(X)|Z = z)$, and define

$$\begin{aligned}
&\rho(\nu, H_\theta) = \int (\nu - H_\theta)^2 dG, \\
&T(\nu) = \text{argmin}_\theta \int (\nu - H_\theta)^2 dG = \text{argmin}_\theta\, \rho(\nu, H_\theta), \;\; \nu \in L_2(G).
\end{aligned}$$

Lemma 3.3.2. *Suppose (A1), (A2), (A5), (F1), (H1), (H3), (K) and (W1) hold. If, in addition* $T(H)$ *is unique, then*

$$\hat{\theta}_n = T(H) + o_p(1).$$

The proof is deferred to the last Section 3.6 of the chapter. Assumption (H2), Lemmas 3.3.1 and 3.3.2 immediately imply the consistency of the proposed estimators $\hat{\theta}_n$ as stated in the following theorem.

Theorem 3.3.1. *Suppose (A1), (A2), (A5), (F1), (H1)–(H3), (K), (W1) and* H_0 *hold. Then* $\hat{\theta}_n \to_p \theta_0$.

3.3.2 Asymptotic normality of $\hat{\theta}_n$

Here we present the asymptotic normality result about $\hat{\theta}_n$ under H_0.

Theorem 3.3.2. *Suppose (A1)–(A3), (A5), (F1), (F2), (H1)–(H6), (K), (W1), (W2) and H_0 hold, then under H_0,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \mathcal{N}_q\Big(0, \Sigma_0^{-1}(\Sigma_1 + \lambda^{-1}\Sigma_2)\Sigma_0^{-1}\Big),$$

where Σ_0 is given in (H6) and

$$\begin{aligned} \Sigma_1 &= \int \frac{(\sigma_\varepsilon^2 + \tau^2(u))\dot{H}_{\theta_0}(u)\dot{H}_{\theta_0}^T(u)g^2(u)}{f_Z(u)}du, \\ \Sigma_2 &= \int \sigma_{\theta_0}(x,y)\dot{H}_{\theta_0}(x)\dot{H}_{\theta_0}^T(y)dG(x)dG(y). \end{aligned} \tag{3.11}$$

This theorem shows that $\hat{\theta}_n$ is $\sqrt{n}$-consistent for θ_0 and the asymptotic covariance matrix is mainly determined by the two terms Σ_1 and Σ_2. The matrix Σ_1 represents the variation in Berkson measurement error model when f_η is known as in KS while Σ_2 represents the contribution due to the estimation of H_θ by $\widehat{H}_\theta$ using the validation data. Moreover, the covariance tends to decay as N/n increases. When $N/n \to \infty$, in other words, when the validation sample size N is sufficiently large, compared to the primary sample size n, not surprisingly the above asymptotic covariance degenerates to the case as if f_η is known.

Remark 3.3.1. Here we verify that the quantities Σ_1 and Σ_2 in the asymptotic variance-covariance matrix are well defined under the given assumptions. Given (A2) and the compactness of $\mathcal{C}$, $\tau^2(u)$ is bounded on $\mathcal{C}$. Assumption (H6) further implies that Σ_1 is finite and positive definite.

Next, consider Σ_2. The Cauchy-Schwarz inequality implies that $\sigma_\theta(x,y) \le \sigma_\theta(x)\sigma_\theta(y)$ for all $x, y \in \mathbb{R}, \theta \in \Theta$, and that for any $a \in \mathbb{R}^q$,

$$\begin{aligned}
|a^T \Sigma_2 a| &\le \int \sigma_{\theta_0}(x,y) \big| a^T \dot{H}_{\theta_0}(x) \dot{H}^T_{\theta_0}(y) a \big| dG(x) dG(y) \\
&= \int \sigma_{\theta_0}(x) \sigma_{\theta_0}(y) \| a^T \dot{H}_{\theta_0}(x) \| \| a^T \dot{H}_{\theta_0}(y) \| dG(x) dG(y) \\
&= \Big(\int \sigma_{\theta_0}(x) \| a^T \dot{H}_{\theta_0}(x) \| dG(x) \Big)^2 \le \|a\|^2 \int \sigma^2_{\theta_0}(x) dG(x) \int \|H_{\theta_0}(x)\|^2 dG(x).
\end{aligned}$$

Hence assumptions (A5) and (H6) ensure that the entries of Σ_2 exist and are finite. Moreover, as seen in the proofs below, Σ_2 is a positive definite covariance matrix.

Now we describe some parametric function families along with the corresponding Σ_2 that satisfy the assumptions (H3)–(H5).

Example 3.3.1. The linear and polynomial cases. Suppose $q = p$, $m_\theta(x) = \theta^T x$, $\theta, x \in \mathbb{R}^p$. Then $H_\theta(z) = \theta^T z$ is a known function. In this case there is no need to estimate this function and one can also use $\tilde{\theta}_n$ as a m.d. estimator of θ. See Remark 3.3.2 for an asymptotic equivalence between $\hat{\theta}_n$ and $\tilde{\theta}_n$.

In the polynomial regression of order p, $q = p + 1$ and $m_\theta(x) = \theta^T \ell(x)$, $x \in \mathbb{R}$, where $\theta = (\theta_1, ..., \theta_{p+1})^T$ and $\ell(x) := (1, x, ..., x^p)^T$ such that $E\|\ell(X)\| < \infty$, where $\|\cdot\|$ denotes the Euclidean norm. Then

$$L(z) := E(\ell(X)|Z = z) = (1, z, E(z+\eta)^2, ..., E(z+\eta)^p)^T, \quad H_\theta(z) = \theta^T L(z).$$

This model is a simple deviation from the linear model and one already sees the need to estimate $H_\theta(z)$. Given the validation data, an estimate of $H_\theta(z)$ in this case is given by

$$\widehat{H}_\theta(z) = \frac{1}{N}\sum_{k=1}^{N} m_\theta(z+\widetilde{\eta}_k) = \frac{1}{N}\sum_{k=1}^{N}\left[\theta_1 + \theta_2(z+\widetilde{\eta}_k) + \theta_3(z+\widetilde{\eta}_k)^2 + \ldots + \theta_{p+1}(z+\widetilde{\eta}_k)^p\right].$$

Here (H3) is satisfied with $r = L$. Furthermore, $\dot{m}_\theta(x) = \ell(x)$ and $\dot{H}_\theta(z) = L(z)$ for all $\theta \in \Theta$. Therefore, similar to the linear case, (H4) and (H5) hold. Moreover,

$$\sigma_\theta(x,y) = \theta^T\left[E\ell(x+\eta)\ell^T(y+\eta) - L(x)L^T(y)\right]\theta,$$
$$\Sigma_2 = \int \theta_0^T[E\ell(x+\eta)\ell^T(y+\eta) - L(x)L^T(y)]\theta_0\, L(x)L^T(y)dG(x)dG(y).$$

Example 3.3.2. The nonlinear case. In biochemistry, one of the well known models for enzyme kinetics relates enzyme reaction rate to the concentration of a substrate x by the formula $\alpha_0 x/(\theta+x)$, $\alpha_0 > 0$, $\theta > 0$, $x > 0$. This is the so called Michaelis–Menten model. The ratio $\gamma_0 = \alpha_0/\theta$ is defined as the catalytic efficiency that measures how efficiently an enzyme converts a substrate into product. With γ_0 known, the function can be written as

$$m_\theta(x) := \frac{\gamma_0\theta x}{\theta+x}, \quad \theta > 0, \quad x > 0. \tag{3.12}$$

We will verify that this nonlinear function satisfies (H3)–(H5).

Regarding (H3), as shown in KS, one sufficient condition is that the regression function $m_\theta(x)$ satisfies the same condition in (H3). In this case, direct calculation shows that

$$|m_{\theta_1}(x) - m_{\theta_0}(x)| = \frac{\gamma_0 x^2|\theta_1-\theta_0|}{(\theta_0+x)(\theta_1+x)} \le \gamma_0|\theta_1-\theta_0|.$$

Hence (H3) holds for (3.12).

Furthermore, suppose for each $x \in \mathbb{R}^p$, the $q \times q$ matrix $\ddot{m}_\theta(x) := \partial^2 m_\theta(x)/\partial\theta^2$ exists for all θ in a neighborhood U_0 of θ_0 and $\|\ddot{m}_\theta(x)\| \leq C$, for all $\theta \in U_0$ and $x \in \mathbb{R}^p$, where the constant C may depend on θ_0. Then, (H4) holds because by the Mean Value Theorem, with probability 1, for all $1 \leq i \leq n$, $N \geq 1$, $\|\theta - \theta_0\| \leq \delta_n$,

$$\frac{\big|N^{-1}\sum_{k=1}^{N}[m_\theta(Z_i+\widetilde{\eta}_k) - m_{\theta_0}(Z_i+\widetilde{\eta}_k) - (\theta-\theta_0)^T\dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k)]\big|}{\|\theta-\theta_0\|} \leq C\|\theta-\theta_0\| \leq C\delta_n.$$

In particular, for the function (3.12), $p = 1 = q$, the second derivative of the function $\ddot{m}_\theta(x) = -2\gamma_0 x^2/(\theta+x)^3$ is bounded for $\theta > 0$ and $x > 0$, so (H4) holds in this case.

As for (H5), with $\sqrt{nh^p}|\theta-\theta_0| \leq a$ and θ_1^* falling between θ and θ_0, we have

$$\begin{aligned}
&\sup_{i,k,\theta} h^{-p/2}|\dot{m}_\theta(Z_i+\widetilde{\eta}_k) - \dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k)| = \sup_{i,k,\theta^*} h^{-p/2}|\ddot{m}_{\theta^*}(Z_i+\widetilde{\eta}_k)(\theta-\theta_0)| \\
&\leq \sup_\theta Ch^{-p/2}|\theta-\theta_0| = O_p(h^{-p/2}/\sqrt{nh^p}) = O_p\big(1/(\sqrt{n}h^p)\big) = o_p(1),
\end{aligned}$$

where C is the upper bound for the second derivative $\ddot{m}_\theta(x)$. Therefore (H5) is satisfied.

Another nonlinear example is the exponential function $m_\theta(x) = e^{\theta x}$, $\theta, x \in \mathbb{R}$. In practice, it is reasonable to assume that both Θ and the domain of X are bounded subsets in $\mathbb{R}$, i.e., $|\theta| \leq C_1$ and $|x| \leq C_2$. Again it suffices to verify that the condition in (H3) holds with $H_\theta(z)$ replaced by $m_\theta(x)$. With θ^* falling between θ_1 and θ_2, we obtain

$$\begin{aligned}
|m_{\theta_2}(x) - m_{\theta_1}(x)| &= |\dot{m}_{\theta^*}(x)(\theta_2-\theta_1)| = |xe^{\theta^* x}(\theta_2-\theta_1)| \\
&\leq (|x|e^{C_1|x|})|\theta_2-\theta_1| := r(x)|\theta_2-\theta_1|.
\end{aligned}$$

Therefore (H3) holds for the exponential regression function. Moreover, the second derivative

$\ddot{m}_\theta(x) = x^2 e^{\theta x}$ is bounded by the constant $C_1^2 e^{C_1 C_2}$. Hence the argument similar to that for (3.12) yields that the exponential function also satisfies (H4) and (H5).

Next, we provide a sketch of the proof of Theorem 3.3.2. The most of the details of the proof are the same as in KN and KS. So we shall be briefly indicating only the major differences.

Proof of Theorem 3.3.2. We first show that

$$nh^p \|\hat{\theta}_n - \theta_0\|^2 = O_p(1). \tag{3.13}$$

Define

$$\widetilde{D}_n(\theta) = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)(\widehat{H}_\theta(Z_i) - H_{\theta_0}(Z_i))\Big]^2 d\hat{\varphi}(z), \tag{3.14}$$

$$\widehat{D}_n(\theta) = \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)(\widehat{H}_\theta(Z_i) - \widehat{H}_{\theta_0}(Z_i))\Big]^2 d\hat{\varphi}(z). \tag{3.15}$$

We shall shortly prove the following two facts.

$$nh^p \widehat{D}_n(\hat{\theta}_n) = O_p(1). \tag{3.16}$$

For any $0 < a < \infty$, there exist n_a and N_a such that

$$P\Big(\widehat{D}_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2 \geq a + \inf_{\|b\|=1} b^T \Sigma_0 b\Big) > 1 - a \qquad \forall\, n > n_a,\, N > N_a, \tag{3.17}$$

where Σ_0 is defined in (H6). Then, as in KS, (3.13) follows from (3.16), (3.17) and the relation

$$nh^p \widehat{D}_n(\hat{\theta}_n) = [nh^p \|\hat{\theta}_n - \theta_0\|^2][\widehat{D}_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2].$$

Proof of (3.16). Subtracting and adding Y_i to the ith summand in (3.14) and the triangular inequality yield

$$\widetilde{D}_n(\hat{\theta}_n) \le 2(\widehat{M}_n(\hat{\theta}_n) + M_n(\theta_0)) \le 2(\widehat{M}_n(\theta_0) + M_n(\theta_0)),$$

because $\hat{\theta}_n$ is the minimizer of $\widehat{M}_n$. From (3.4) of KS, we obtain $nh^p M_n(\theta_0) = O_p(1)$. Lemma 3.3.1 and the decomposition (3.9) of $\widehat{M}_n$ imply that $nh^p \widehat{M}_n(\theta_0) = O_p(1)$. Therefore

$$nh^p \widetilde{D}_n(\hat{\theta}_n) = O_p(1). \tag{3.18}$$

Next, subtracting and adding $H_{\theta_0}(Z_i)$ to the ith summand in (3.15) and the triangular inequality yield

$$\widehat{D}_n(\hat{\theta}_n) \;\; \le \;\; 2(W_n(\theta_0) + \widetilde{D}_n(\hat{\theta}_n)).$$

Lemma 3.3.1 implies that $N\,W_n(\theta_0) = O_p(1)$. This fact, (3.18) and (W1) yield (3.16).

To prove (3.17), define

$$u_n = \hat{\theta}_n - \theta_0, \quad d_{nik} = m_{\hat{\theta}_n}(Z_i + \widetilde{\eta}_k) - m_{\theta_0}(Z_i + \widetilde{\eta}_k) - u_n^T \dot{m}_{\theta_0}(Z_i + \widetilde{\eta}_k), \tag{3.19}$$

$$\dot{\hat{\mu}}_n(z,\theta) = \int \frac{1}{n} \sum_{i=1}^n K_{hi}(z) \dot{\hat{H}}_\theta(Z_i) d\hat{\varphi}(z) = \int \frac{1}{nN} \sum_{i=1}^n \sum_{k=1}^N K_{hi}(z) \dot{m}_\theta(Z_i + \widetilde{\eta}_k) d\hat{\varphi}(z),$$

$$D_{n1} = \int \Big\{ \frac{1}{nN} \sum_{i=1}^n \sum_{k=1}^N K_{hi}(z) \Big[\frac{d_{nik}}{\|u_n\|} \Big] \Big\}^2 d\hat{\varphi}(z), \qquad D_{n2} = \int \Big\{ \frac{u_n^T \dot{\hat{\mu}}_n(z,\theta)}{\|u_n\|} \Big\}^2 d\hat{\varphi}(z).$$

Then by the Cauchy-Schwarz inequality,

$$\begin{aligned}\frac{\widehat{D}_n(\hat{\theta}_n)}{\|u_n\|^2} &= \int \Big\{\frac{1}{nN}\sum_{i=1}^{n}\sum_{k=1}^{N} K_{hi}(z)\Big[\frac{d_{nik}}{\|u_n\|} + \frac{u_n^T \dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k)}{\|u_n\|}\Big]\Big\}^2 d\hat{\varphi}(z) \\ &\geq D_{n1} + D_{n2} - 2\sqrt{D_{n1}D_{n2}}.\end{aligned}$$

By (3.7) and (3.8),

$$\int \Big\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\Big\}^2 d\hat{\varphi}(z) = O_p(1). \tag{3.20}$$

The consistency of $\hat{\theta}_n$, (H4) and (3.20) in turn imply

$$D_{n1} \leq \max_i \Big\{N^{-1}\sum_{k=1}^{N}\frac{d_{nik}}{\|u_n\|}\Big\}^2 \int \Big\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\Big\}^2 d\hat{\varphi}(z) = o_p(1). \tag{3.21}$$

An argument similar to the one used in KN in the analysis of the analog of D_{n2} yields (3.17) for D_{n2}, thereby completing the proof of (3.13).

Now we provide a sketch to derive the asymptotic variance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Proceeding as in KN and KS, $\hat{\theta}_n$ is the root of the score equation

$$\dot{\widehat{M}}_n(\theta) = -2\int \Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)(Y_i - \widehat{H}_\theta(Z_i))\Big]\Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\dot{\widehat{H}}_\theta(Z_i))\Big] d\hat{\varphi}(z) = 0. \tag{3.22}$$

Arguing as in Lemma 4.2 of KN pertaining to g_{n1}, the above equation becomes

$$\dot{\widehat{M}}_n(\theta) = -2\int \Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)(Y_i - \widehat{H}_\theta(Z_i))\Big]\Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\dot{H}_\theta(Z_i))\Big] d\hat{\varphi}(z) = 0. \tag{3.23}$$

Define

$$\dot{\mu}_n(z,\theta) = \frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\dot{H}_\theta(Z_i), \quad \dot{\mu}_h(z,\theta) = E[K_h(z-Z)\dot{H}_\theta(Z)], \quad \xi_i := Y_i - H_{\theta_0}(Z_i),$$

$$U_{n1}(z) = \frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\xi_i, \qquad U_{n2}(z) = \frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)(\widehat{H}_{\theta_0}(Z_i) - H_{\theta_0}(Z_i)),$$

$$S_n = \int U_{n1}(z)\dot{\mu}_h(z,\theta_0)d\varphi(z), \qquad T_n = \int U_{n2}(z)\dot{\mu}_h(z,\theta_0)d\varphi(z),$$

$$V_n(z,\theta) = \frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)(\widehat{H}_\theta(Z_i) - \widehat{H}_{\theta_0}(Z_i)), \qquad \Sigma_0 = \int \dot{H}_{\theta_0}(x)\dot{H}^T_{\theta_0}(x)dG(x).$$

Then the equation (3.23) is equivalent to

$$\int [U_{n1}(z) - U_{n2}(z)]\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z) = \int V_n(z,\hat{\theta}_n)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z). \tag{3.24}$$

A major difference between the proofs in KN, KS and here is the presence of the additional term $\int U_{n2}(z)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z)$ in (3.24) due to the estimation of $H_{\theta_0}(z)$ by $\widehat{H}_{\theta_0}(z)$. A slight modification of the arguments in the proofs of Lemmas 4.1–4.3 of KN yield

$$\sqrt{n}\int U_{n1}(z)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z) = \sqrt{n}S_n + o_p(1), \qquad \sqrt{n}S_n \to_d \mathcal{N}_q(0,\Sigma_1),$$

$$\sqrt{n}\int U_{n2}(z)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z) = \sqrt{n}T_n + o_p(1).$$

It thus remains to investigate the asymptotic property of T_n. For that purpose, define

$$\phi_T(Z_i,\widetilde{\eta}_k) := \int K_{hi}(z)[m_{\theta_0}(Z_i+\widetilde{\eta}_k) - H_{\theta_0}(Z_i)]\dot{\mu}_h(z,\theta_0)d\varphi(z), \quad 1\le i\le n, 1\le k\le N,$$

$$\ell(x) := \int [m_{\theta_0}(z+x) - H_{\theta_0}(z)]\dot{\mu}_h(z,\theta_0)f_Z(z)d\varphi(z).$$

Then

$$\sqrt{n}T_n = \frac{1}{\sqrt{n}N}\sum_{i=1}^{n}\sum_{k=1}^{N}\phi_T(Z_i, \widetilde{\eta}_k).$$

The statistic T_n is a two sample U statistic with kernel function ϕ_T. We shall be using Theorem B.1 in Sepanski and Lee (1995) to derive asymptotic distribution of T_n and some other statistics. For the sake of completeness we include statement of this theorem as Lemma 3.6.1 in the last Section 3.6.

In order to apply Lemma 3.6.1 to T_n, we need to identify the limits of projections of ϕ_T, i.e., $\lim_{n\to\infty} E(\phi_T|Z_1)$ and $\lim_{n\to\infty} E(\phi_T|\widetilde{\eta}_1)$ as well as their corresponding variances. Algebra shows that $E(\phi_T|Z_1) \equiv 0$, $E(\phi_T|\widetilde{\eta}_1) \to_p \ell(\widetilde{\eta}_1)$, $\mathrm{Var}(\ell(\widetilde{\eta}_1)) = \Sigma_2$, where Σ_2 is as in (3.11). Applying Lemma 3.6.1 in Sepanski and Lee (1995) yields that

$$\sqrt{n}T_n \to_d \mathcal{N}_q(0, \Sigma_2/\lambda),$$

where $\lambda > 0$ is as in assumption (W1). Note that the asymptotic property of T_n is dominated by the behavior of $E(\phi_T|\widetilde{\eta}_1)$, the projection of ϕ_T on the validation sample space and S_n is constructed only based on the primary sample (Y_i, Z_i). Hence S_n and T_n are asymptotically independent. Therefore the left hand side of (3.24) is asymptotically normally distributed with convergence rate $\sqrt{n}$ and variance-covariance matrix $\Sigma_1 + (\Sigma_2/\lambda)$.

Now we will show that the right hand side of (3.24) equals $\Omega_n(\hat{\theta}_n - \theta_0)$, where

$$\Omega_n = \Sigma_0 + o_p(1). \tag{3.25}$$

Let $e_n := u_n/\|u_n\|$,

$$\widetilde{V}_n := \int \frac{1}{nN}\sum_{i=1}^{n}\sum_{k=1}^{N}\Big[K_{hi}(z)\frac{d_{nik}}{\|u_n\|}\Big]\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z), \quad L_n := \int \dot{\mu}_n(z,\hat{\theta}_n)\dot{\mu}_n^T(z,\theta_0)d\hat{\varphi}(z).$$

Then the right hand side of (3.24) can be rewritten as

$$\int V_n(z,\hat{\theta}_n)\dot{\mu}_n(z,\hat{\theta}_n)d\hat{\varphi}(z) = [\widetilde{V}_n e_n^T + L_n]u_n.$$

Argue as in KN to show that (H4) implies $\widetilde{V}_n = o_p(1)$ and $L_n = \Sigma_0 + o_p(1)$. Moreover, e_n being a unit vector, we obtain $\|\widetilde{V}_n e_n^T\| = o_p(1)$. This completes the sketch of the proof of (3.25), thereby that of Theorem 3.3.2.

Remark 3.3.2. Connection between $\hat{\theta}_n$ and $\tilde{\theta}_n$ in linear regression. Here we shall investigate a relation between the estimators $\hat{\theta}_n$ and $\tilde{\theta}_n$ in the linear model. Assume

$$\mu(x) = m_\theta(x) = \theta^T x, \quad x \in \mathcal{C} \subset \mathbb{R}^p, \quad \text{for some } \theta \in \Theta \subset \mathbb{R}^p. \tag{3.26}$$

Then $H_\theta(z) = \theta^T z$ and a closed form of $\hat{\theta}_n$ can be derived by taking derivative of $\widehat{M}_n(\theta)$ and solving the equation $\partial\widehat{M}_n(\theta)/\partial\theta = 0$, i.e., $B_n\hat{\theta}_n = A_n$, where

$$\begin{aligned} A_n &= \int \Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(z)Y_i\Big]\Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(z)(Z_i+\bar{\eta})\Big]d\hat{\varphi}(z), \\ B_n &= \int \Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(z)(Z_i+\bar{\eta})\Big]\Big[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(z)(Z_i+\bar{\eta})^T\Big]d\hat{\varphi}(z), \end{aligned}$$

with $\bar{\eta} = N^{-1} \sum_{k=1}^{N} \widetilde{\eta}_k$. Similarly, $\widetilde{B}_n \tilde{\theta}_n = \widetilde{A}_n$, where

$$\begin{aligned}
\widetilde{A}_n &= \int \Big[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(z) Y_i \Big] \Big[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(z) Z_i \Big] d\hat{\varphi}(z), \\
\widetilde{B}_n &= \int \Big[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(z) Z_i \Big] \Big[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(z) Z_i^T \Big] d\hat{\varphi}(z).
\end{aligned}$$

Roughly speaking, because $\bar{\eta} \to_p 0$, $A_n - \widetilde{A}_n = o_p(1)$, $B_n - \widetilde{B}_n = o_p(1)$ and hence $\hat{\theta}_n - \tilde{\theta}_n \to_p 0$. Furthermore, under some specific conditions, both $\hat{\theta}_n$ and $\tilde{\theta}_n$ can achieve the same asymptotic efficiency. We present two such assumptions here.

(A6) $E\eta^2 < \infty$. $\tau_1(z) := E\big(|\varepsilon| \big| Z = z\big)$ is a.e. (G) continuous.

(A7) $\nu_G := \int z dG(z) = 0$, $\int zz^T dG(z)$ is positive definite.

Proposition 3.3.1. *Suppose* (3.1) *and* (3.26) *hold with* $\theta = \theta_0$. *In addition suppose (A1), (F1), (K), (W1), (A6) and (A7) hold, then* $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) \to_p 0$.

Proof. For the transparency of the exposition, we give details for the case $p = 1$ only. Then $\widetilde{B}_n = \int \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) Z_i\big]^2 d\hat{\varphi}(z)$. By (3.5), (3.7), (3.8) and direct calculations, $\widetilde{B}_n = \kappa_G + o_p(1)$, where $\kappa_G = \int z^2 dG(z)$. By (A7), $\kappa_G > 0$. Then $\tilde{\theta}_n = \widetilde{B}_n^{-1} \widetilde{A}_n$ is well defined for all sufficiently large n and the consistency of $\tilde{\theta}_n$ yields that $\widetilde{A}_n = O_p(1)$. We shall shortly show that

$$\text{(a)} \quad \sqrt{n}(A_n - \widetilde{A}_n) = o_p(1), \qquad \text{(b)} \quad B_n = \widetilde{B}_n + o_p(n^{-1/2}). \tag{3.27}$$

Then for all sufficiently large n, $\hat{\theta}_n = B_n^{-1} A_n$ and

$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = \frac{\sqrt{n}(A_n \widetilde{B}_n - \widetilde{A}_n B_n)}{B_n \widetilde{B}_n} = \frac{\sqrt{n}[A_n \widetilde{B}_n - \widetilde{A}_n(\widetilde{B}_n + o_p(n^{-1/2}))]}{\widetilde{B}_n(\widetilde{B}_n + o_p(n^{-1/2}))}$$

$$= \quad \frac{\sqrt{n}(A_n - \widetilde{A}_n)\widetilde{B}_n - o_p(\widetilde{A}_n)}{\kappa_G^2 + o_p(1)} = o_p(1).$$

To prove (3.27)(a), rewrite

$$\sqrt{n}(A_n - \widetilde{A}_n) = \sqrt{n}\bar{\eta} \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)Y_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)\Big] d\hat{\varphi}(z) := \sqrt{n}\bar{\eta}\,\mathcal{A}_n.$$

By (A6) and CLT, $\sqrt{n}\bar{\eta} = O_p(1)$. It thus suffices to show that $\mathcal{A}_n = o_p(1)$. Let $\mathcal{A}_n^*$ denote the $\mathcal{A}_n$ with $\hat{\varphi}$ replaced by φ. Then the facts (3.7), $E(|Y|\big|Z = z) \le |\theta_0 z| + \tau_1(z)$, assumption (A6) and rigorous calculation yield that

$$|\mathcal{A}_n - \mathcal{A}_n^*| = o_p\Big(\int \frac{1}{n^2}\sum_{i,j=1}^n |K_{hi}(z)K_{hj}(z)Y_i| d\varphi(z)\Big) = o_p(1).$$

Now we rewrite

$$\mathcal{A}_n^* = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(z)Y_i d\varphi(z) + \frac{1}{n^2}\sum_{i=1}^n \sum_{j\neq i=1}^n \int K_{hi}(z)K_{hj}(z)Y_i d\varphi(z) := \mathcal{A}_{n1} + \mathcal{A}_{n2}.$$

Calculation of moments shows that $E\mathcal{A}_{n1} = O((nh)^{-1})$, $E\mathcal{A}_{n2} = \theta_0\nu_G + o(1)$, $\mathrm{Var}(\mathcal{A}_{n1}) = O(n^{-3}h^{-2})$ and $\mathrm{Var}(\mathcal{A}_{n2}) = O(n^{-1})$. Hence $\mathcal{A}_n^* = \theta_0\nu_G + o_p(1)$, and (A7) implies (3.27)(a).

Now we prove (3.27)(b). Let

$$\mathcal{B}_n := \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)Z_i\Big]\Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)\Big] d\hat{\varphi}(z).$$

Then, by (3.8),

$$B_n - \widetilde{B}_n \quad = \quad 2\bar{\eta}\mathcal{B}_n + \bar{\eta}^2 \int \Big[\frac{1}{n}\sum_{i=1}^n K_{hi}(z)\Big]^2 d\hat{\varphi}(z) = 2\bar{\eta}\,\mathcal{B}_n + O_p(n^{-1}).$$

Argue as in the analysis of $\mathcal{A}_n$ to obtain that $\mathcal{B}_n = \nu_G + o_p(1)$. This fact and $\sqrt{n}\bar{\eta} = O_p(1)$ imply that $\sqrt{n}(B_n - \widetilde{B}_n) = 2(\sqrt{n}\bar{\eta})\nu_G + O_p(n^{-1/2})$, which together with (A7) imply (3.27)(b). This also completes the proof of the lemma. $\square$

3.4 Testing

In this section we establish the asymptotic behavior of the proposed tests associated with $\widehat{M}_n(\hat{\theta}_n)$ under the null and certain fixed alternative hypotheses. Let

$$\begin{aligned}
&\xi_i = Y_i - H_{\theta_0}(Z_i), \qquad \hat{\xi}_i = Y_i - \widehat{H}_{\hat{\theta}_n}(Z_i), \\
&\widetilde{C}_n = \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\xi_i^2 d\varphi(z), \quad \widetilde{\Gamma}_n = \frac{2h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\xi_i\xi_j d\varphi(z)\Big)^2, \\
&\widehat{C}_n = \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\hat{\xi}_i^2 d\hat{\varphi}(z), \quad \widehat{\Gamma}_n = \frac{2h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\hat{\xi}_i\hat{\xi}_j d\hat{\varphi}(z)\Big)^2.
\end{aligned}$$

Because $\xi = Y - H_{\theta_0}(Z) = \varepsilon + m_{\theta_0}(X) - H_{\theta_0}(Z)$ and because Z, η and ε are mutually independent, $E(\xi^2|Z=z) = \sigma_\varepsilon^2 + \tau^2(z)$, where τ^2 is as in (A2). Since $\mathcal{C}$ is compact, and by (A2), τ^2 is continuous, we obtain $\int E\big(\xi^2|Z=z\big)dG(z) < \infty$.

The following theorem gives the main result of this section.

Theorem 3.4.1. *Suppose (A1), (A2), (A4), (A5), (F1)–(F2), (K), (H1)–(H6), (W1) and (W2) hold. Then, under H_0, $nh^{p/2}\widehat{\Gamma}_n^{-1/2}\big(\widehat{M}_n(\hat{\theta}_n) - \widehat{C}_n\big) \to_d \mathcal{N}_1(0,1)$.*

Consequently, the null hypothesis is rejected by the test if $\widehat{\mathcal{T}}_n := nh^{p/2}\widehat{\Gamma}_n^{-1/2}|\widehat{M}_n(\hat{\theta}_n) - \widehat{C}_n| > z_{\alpha/2}$ with the asymptotic size $\alpha > 0$, where z_α is the upper 100αth percentile of the standard normal distribution.

The theorem shows that the ratio parameter N/n does not play a role in the limiting null distribution. This finding is also reflected in the finite sample simulation study through the empirical level and power with different choices of N/n.

Here we provide a sketch of the proof of the above theorem. Rewrite

$$\begin{aligned}
&\widehat{M}_n(\hat{\theta}_n)\\
=\ & \int \Big\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(z)\big[Y_i - H_{\theta_0}(Z_i) + H_{\theta_0}(Z_i) - \widehat{H}_{\theta_0}(Z_i) + \widehat{H}_{\theta_0}(Z_i) - \widehat{H}_{\hat{\theta}_n}(Z_i)\big]\Big\}^2 d\hat{\varphi}(z)\\
=\ & \int [U_{n1}(z) - U_{n2}(z) - V_n(z,\hat{\theta}_n)]^2 d\hat{\varphi}(z)\\
=\ & \int [U_{n1}(z) - U_{n2}(z)]^2 d\hat{\varphi}(z) + \int [V_n(z,\hat{\theta}_n)]^2 d\hat{\varphi}(z)\\
& \qquad\qquad -2\int [U_{n1}(z) - U_{n2}(z)] V_n(z,\hat{\theta}_n) d\hat{\varphi}(z)\\
=:\ & J_n + \widehat{D}_n(\hat{\theta}_n) - 2K_n(\hat{\theta}_n), \qquad \text{say.}
\end{aligned}$$

The following three lemmas are needed for the proof of Theorem 3.4.1.

Lemma 3.4.1. *Suppose assumptions (A1), (A2), (A4), (A5), (F1)–(F2), (K), (H1)–(H6), (W1), (W2) and H_0 hold. Then*

$$nh^{p/2}\widetilde{\Gamma}_n^{-1/2}(J_n - \widetilde{C}_n) \to_d \mathcal{N}_1(0,1). \tag{3.28}$$

Lemma 3.4.2. *Under the assumptions of Lemma 3.4.1, the following holds.*

$$\text{(a)}\quad nh^{p/2}\widehat{D}_n(\hat{\theta}_n) = o_p(1), \qquad \text{(b)}\quad nh^{p/2}K_n(\hat{\theta}_n) = o_p(1). \tag{3.29}$$

Lemma 3.4.3. *Suppose assumptions (A1), (A2), (F1), (K), (H1)–(H6), (W1) with $\lambda < \infty$,* *(W2) and H_0 hold. Then*

$$\text{(a)}\quad nh^{p/2}(\widehat{C}_n - \widetilde{C}_n) = o_p(1). \qquad \text{(b)}\quad \widehat{\Gamma}_n - \widetilde{\Gamma}_n = o_p(1). \tag{3.30}$$

The above three lemmas yield the asymptotic normality of $\widehat{M_n}(\hat{\theta}_n)$ in Theorem 3.4.1 in a routine fashion. Here we provide the proofs of these lemmas.

Proof of Lemma 3.4.1. Let J_n^* denote the J_n with $\hat{\varphi}$ replaced by φ. Algebra shows that

$$EJ_n^* = E\int U_{n1}^2(z)d\varphi(z) + E\int U_{n2}^2(z)d\varphi(z) = O((nh^p)^{-1}) + O(N^{-1}) = O\big((nh^p)^{-1}\big).$$

Then, by (3.6) and (W2),

$$\begin{aligned} nh^{p/2}\big|J_n - J_n^*\big| &\leq nh^{p/2}J_n^* \sup_{z\in\mathcal{C}}\left|\frac{f^2(z)}{\hat{f}_w^2(z)} - 1\right| \\ &= nh^{p/2}O_p((nh^p)^{-1})O_p\Big(\log_k(n)\big(\frac{\log(n)}{n}\big)^{p/(p+4)}\Big) = o_p(1), \end{aligned}$$

Therefore,

$$J_n = J_n^* + o_p((nh^{p/2})^{-1}) = O_p((nh^p)^{-1}). \tag{3.31}$$

It thus suffices to prove (3.28) with J_n replaced by J_n^*. To proceed further, define for $1 \leq i,j \leq n$, $1 \leq k,l \leq N$, $i \neq j$, $k \neq l$,

$$\begin{aligned} \Delta_{ik} &= m_{\theta_0}(Z_i + \widetilde{\eta}_k) - H_{\theta_0}(Z_i), \qquad \mathcal{D}_i = (Z_i, \xi_i), \qquad (3.32) \\ \psi_1(\mathcal{D}_i, \mathcal{D}_j, \widetilde{\eta}_k, \widetilde{\eta}_l) &= \frac{1}{2}\int K_{hi}(z)K_{hj}(z)[(\xi_i - \Delta_{ik})(\xi_j - \Delta_{jl}) + (\xi_j - \Delta_{jk})(\xi_i - \Delta_{il})]d\varphi(z), \end{aligned}$$

$$
\begin{aligned}
\psi_2(\mathcal{D}_i, \widetilde{\eta}_k, \widetilde{\eta}_l) &= \int K_{hi}^2(z)(\xi_i - \Delta_{ik})(\xi_i - \Delta_{il}) d\varphi(z), \\
\psi_3(\mathcal{D}_i, \mathcal{D}_j, \widetilde{\eta}_k) &= \int K_{hi}(z) K_{hj}(z)(\xi_i - \Delta_{ik})(\xi_j - \Delta_{jk}) d\varphi(z), \\
\psi_4(\mathcal{D}_i, \widetilde{\eta}_k) &= \int K_{hi}^2(z)(\xi_i - \Delta_{ik})^2 d\varphi(z).
\end{aligned}
$$

Rewrite

$$
\begin{aligned}
& J_n^* \\
=& \int \Big\{ \frac{1}{nN} \sum_{i=1}^{n} \sum_{k=1}^{N} K_{hi}(z)(\xi_i - \Delta_{ik}) \Big\}^2 d\varphi(z) \\
=& \frac{1}{n^2 N^2} \sum_{i,j=1}^{n} \sum_{k,l=1}^{N} \int K_{hi}(z) K_{hj}(z)(\xi_i - \Delta_{ik})(\xi_j - \Delta_{jl}) d\varphi(z) \\
=& \frac{1}{n^2 N^2} \Big\{ \sum_{i \neq j, k \neq l} + \sum_{i=j, k \neq l} + \sum_{i \neq j, k=l} + \sum_{i=j, k=l} \Big\} \int K_{hi}(z) K_{hj}(z)(\xi_i - \Delta_{ik})(\xi_j - \Delta_{jl}) d\varphi(z) \\
=:& J_{n1} + J_{n2} + J_{n3} + J_{n4}.
\end{aligned}
$$

All these four quantities are similar to the two sample U statistics. We will show that only J_{n2} contributes to the asymptotic expectation and only J_{n1} contributes to the asymptotic variance in the limiting distribution. Note that $E(\Delta_{ik}|Z_i) \equiv 0$ and $E(\xi_i|Z_i) \equiv 0$, a.s. Hence

$$
\begin{aligned}
E J_{n1} &= 0, \qquad (3.33) \\
E(J_{n2} - \widetilde{C}_n) &= E \Big\{ \frac{1}{n^2 N^2} \sum_{i=1}^{n} \sum_{k \neq l} E(\psi_2(\mathcal{D}_i, \widetilde{\eta}_k, \widetilde{\eta}_l)|\mathcal{D}_i) - \widetilde{C}_n \Big\} \\
&= \frac{1}{Nn} \int E\big[K_{h1}^2(z)\xi_1^2\big] d\varphi(z) \Big] = O((Nnh^p)^{-1}), \\
E J_{n3} &= \frac{n-1}{nN} E \int K_{h1}(z) K_{h2}(z) \Delta_{11} \Delta_{21} d\varphi(z) = O(N^{-1}), \\
E J_{n4} &= O((Nnh^p)^{-1}).
\end{aligned}
$$

Now we investigate the variances of $J_{nj}, j = 1, 2, 3, 4$, using Lemmas 2.6.4 and 2.6.5. We verify that J_{n1} is the only leading term. Note that

$$J_{n1} = \frac{1}{n^2N^2} \sum_{i \neq j, k \neq l} \psi_1(\mathcal{D}_i, \mathcal{D}_j, \widetilde{\eta}_k, \widetilde{\eta}_l).$$

In order to apply Lemma 2.6.5, we first calculate the projections of ψ_1:

$$\begin{aligned}
E(\psi_1|\mathcal{D}_i, \mathcal{D}_j) &= \int K_{hi}(z)K_{hj}(z)\xi_i\xi_j d\varphi(z), \\
E(\psi_1|\widetilde{\eta}_k, \widetilde{\eta}_l) &= \int (m_{\theta_0}(z + \widetilde{\eta}_k) - H_{\theta_0}(z))(m_{\theta_0}(z + \widetilde{\eta}_l) - H_{\theta_0}(z))f^2(z)d\varphi(z) + o_p(1), \\
E(\psi_1|\mathcal{D}_i, \mathcal{D}_j, \widetilde{\eta}_k) &= \frac{1}{2}\int K_{hi}(z)K_{hj}(z)[(\xi_i - \Delta_{ik})\xi_j + (\xi_j - \Delta_{jk})\xi_i]d\varphi(z), \\
E(\psi|\mathcal{D}_i, \widetilde{\eta}_k, \widetilde{\eta}_j) &= \frac{1}{2}\int K_{hi}(z)[(\xi_i - \Delta_{ik})(m_{\theta_0}(z + \widetilde{\eta}_k) - H_{\theta_0}(z)) \\
&\quad + (\xi_i - \Delta_{il})(m_{\theta_0}(z + \widetilde{\eta}_l) - H_{\theta_0}(z))]f(z)d\varphi(z) + o_p(1).
\end{aligned}$$

All other projections vanish. We also verify the variances of the above projections

$$\begin{aligned}
&\mathrm{Var}(\psi_1) = O(h^{-p}), \quad \mathrm{Var}\big(E(\psi_1|\mathcal{D}_i, \mathcal{D}_j)\big) = O(h^{-p}), \quad \mathrm{Var}\big(E(\psi_1|\widetilde{\eta}_k, \widetilde{\eta}_l)\big) = O(1), \\
&\mathrm{Var}\big(E(\psi_1|\mathcal{D}_i, \mathcal{D}_j, \widetilde{\eta}_k)\big) = O(h^{-p}), \quad \mathrm{Var}\big(E(\psi_1|\mathcal{D}_i, \widetilde{\eta}_k, \widetilde{\eta}_l)\big) = O(1).
\end{aligned}$$

Therefore, Lemma 2.6.5 implies that

$$\mathrm{Var}(J_{n1}) = O\Big(\frac{1}{n^2N^2h^p} + \frac{1}{n^2h^p} + \frac{1}{N^2} + \frac{1}{n^2Nh^p} + \frac{1}{nN^2}\Big) = O\Big(\frac{1}{n^2h^p}\Big).$$

Furthermore, it is seen that only the variance term associated with $E(\psi_1|\mathcal{D}_i, \mathcal{D}_j)$ dominates the variance of J_{n1} and all other projection variances are $o(1/(n^2h^p))$. Thus, if we let

$$\widetilde{J}_{n1} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i=1}^{n} E(\psi_1 | \mathcal{D}_i, \mathcal{D}_j),$$

then

$$nh^{p/2}(J_{n1} - \widetilde{J}_{n1}) = o_p(1).$$

From Lemma 5.1 in KN, we obtain $nh^{p/2}\widetilde{\Gamma}_n^{-1/2}\widetilde{J}_{n1} \to_d \mathcal{N}_1(0,1)$. Hence

$$nh^{p/2}\widetilde{\Gamma}_n^{-1/2} J_{n1} \to_d \mathcal{N}_1(0,1). \tag{3.34}$$

By using arguments similar to those used in the proof of Lemma 3.3.1, one can verify

$$\mathrm{Var}(nh^{p/2}J_{n2}) = o(1), \quad \mathrm{Var}(nh^{p/2}J_{n3}) = o(1), \quad \mathrm{Var}(nh^{p/2}J_{n4}) = o(1).$$

Combining these facts with the expectation results in (3.33), we have

$$J_{n2} = \widetilde{C}_n + o_p(1/(nh^{p/2})), \quad J_{n3} = o_p(1/(nh^{p/2})), \quad J_{n4} = o_p(1/(nh^{p/2})).$$

Therefore, (3.31) and these facts above imply

$$nh^{p/2}\widetilde{\Gamma}_n^{-1/2}(J_n - \widetilde{C}_n) = nh^{p/2}\widetilde{\Gamma}_n^{-1/2} J_{n1} + o_p(1).$$

This fact together with (3.34) yield the conclusion (3.28).

Proof of Lemma 3.4.2. Recall the notation from (3.19). We have

$$\begin{aligned}\widehat{D}_n(\hat\theta_n) &= \int\Big\{\frac{1}{nN}\sum_{i=1}^n\sum_{k=1}^N K_{hi}(z)\big[d_{nik}+u_n^T\dot m_{\theta_0}(Z_i+\widetilde\eta_k)\big]\Big\}^2 d\hat\varphi \qquad (3.35)\\ &\le 2\int\Big[\frac{1}{nN}\sum_{i=1}^n\sum_{k=1}^N K_{hi}(z)d_{nik}\Big]^2 d\hat\varphi(z)+2\int\big[u_n^T\dot{\hat\mu}_n(z,\theta_0)\big]^2 d\hat\varphi(z)\\ &= 2\|u_n\|^2\big[D_{n1}+D_{n2}\big].\end{aligned}$$

The Cauchy-Schwarz inequality and (3.5) imply that

$$D_{n2}\le\int\|\dot{\hat\mu}_n(z,\theta_0)\|^2 d\hat\varphi(z)=\int[\dot{\hat\mu}_n(z,\theta_0)]^T\dot{\hat\mu}_n(z,\theta_0)d\varphi(z)+o_p(1).$$

Calculation shows that $E\int[\dot{\hat\mu}_n(z,\theta_0)]^T\dot{\hat\mu}_n(z,\theta_0)d\varphi(z)=O(1)$ under (H6). Hence $D_{n2}=O_p(1)$. This fact, (3.21) and the fact $n\|u_n\|^2=O_p(1)$, implied by Theorem 3.3.2, together with (3.35) imply $\widehat{D}_n(\hat\theta_n)=o_p\big((nh^{p/2})^{-1}\big)$, thereby proving (3.29)(a).

In order to prove (3.29)(b), let $U_n:=U_{n1}-U_{n2}$ and rewrite

$$\begin{aligned}K_n(\hat\theta_n) &= \int U_n(z)\Big\{\frac{1}{nN}\sum_{i=1}^n\sum_{k=1}^N K_{hi}(z)[m_{\hat\theta_n}(Z_i+\widetilde\eta_k)-m_{\theta_0}(Z_i+\widetilde\eta_k)]\Big\}d\hat\varphi(z)\\ &= \int U_n(z)\Big\{\frac{1}{nN}\sum_{i=1}^n\sum_{k=1}^N K_{hi}(z)[d_{nik}+u_n^T\dot m_{\theta_0}(Z_i+\widetilde\eta_k)]\Big\}d\hat\varphi(z)\\ &= \int U_n(z)\Big\{\frac{1}{nN}\sum_{i=1}^n\sum_{k=1}^N K_{hi}(z)d_{nik}\Big\}d\hat\varphi(z)+\int U_n(z)u_n^T\dot{\hat\mu}_n(z,\theta_0)d\hat\varphi(z)\\ &=: R_1+R_2.\end{aligned}$$

The facts (3.4), (3.31), and the Cauchy-Schwarz inequality imply that

$$
\begin{aligned}
nh^{p/2}|R_1| &\le J_n^{1/2} \times \Big\{ \int \Big(\frac{1}{nN} \sum_{i=1}^{n} \sum_{k=1}^{N} K_{hi}(z) d_{nik} \Big)^2 d\hat{\varphi}(z) \Big\}^{1/2} \\
&= nh^{p/2} o_p(\|u_n\|) O_p((nh^p)^{-1/2}) = o_p(1).
\end{aligned}
$$

Next, rewrite

$$
\begin{aligned}
R_2 &= u_n^T \int U_n(z) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) \dot{\hat{H}}_{\theta_0}(Z_i) \big] d\hat{\varphi}(z) \\
&= u_n^T \int U_n(z) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) \dot{\hat{H}}_{\hat{\theta}_n}(Z_i) \big] d\hat{\varphi}(z) \\
&\qquad - u_n^T \int U_n(z) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) (\dot{\hat{H}}_{\hat{\theta}_n}(Z_i) - \dot{\hat{H}}_{\theta_0}(Z_i)) \big] d\hat{\varphi}(z) \\
&=: R_{21} - R_{22}.
\end{aligned}
$$

The score equation (3.22) implies that

$$
\begin{aligned}
R_{21} &= u_n^T \int V_n(z, \hat{\theta}_n) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) \dot{\hat{H}}_{\hat{\theta}_n}(Z_i) \big] d\hat{\varphi}(z) \\
&= u_n^T \int V_n(z, \hat{\theta}_n) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) \dot{\hat{H}}_{\theta_0}(Z_i) \big] d\hat{\varphi}(z) \\
&\qquad + u_n^T \int V_n(z, \hat{\theta}_n) \big[n^{-1} \sum_{i=1}^{n} K_{hi}(z) (\dot{\hat{H}}_{\hat{\theta}_n}(Z_i) - \dot{\hat{H}}_{\theta_0}(Z_i)) \big] d\hat{\varphi}(z) \\
&=: R_{211} + R_{212}.
\end{aligned}
$$

Direct calculations together with (3.29)(a) and (3.8) yield

$$
nh^{p/2}|R_{211}| \le nh^{p/2} \|u_n\| [\widehat{D}_n(\hat{\theta}_n)]^{1/2} \Big(\int \|n^{-1} \sum_{i=1}^{n} K_{hi}(z) \dot{\hat{H}}_{\theta_0}(Z_i)\|^2 d\hat{\varphi}(z) \Big)^{-1/2}
$$

$$= \quad nh^{p/2}O_p(n^{-1/2})o_p((nh^{p/2})^{-1/2})O_p(1) = o_p(1).$$

Similarly, assumption (H5), $n^{1/2}\|u_n\| = O_p(1)$ and (3.29)(a) imply that $nh^{p/2}R_{212} = o_p(1)$ thereby $nh^{p/2}R_{21} = o_p(1)$.

Regarding R_{22}, the Cauchy-Schwarz inequality implies that

$$\begin{aligned} nh^{p/2}\big|R_{22}\big| &\leq \|u_n\|J_n^{1/2} \times \Big\{\int \|\frac{1}{nN}K_{hi}(z)(\dot{m}_{\hat{\theta}_n}(Z_i+\widetilde{\eta}_k) - \dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k))\|^2 d\hat{\varphi}(z)\Big\}^{1/2} \\ &= nh^{p/2}O_p(n^{-1/2})O_p((nh^p)^{-1/2})o_p(h^{p/2}) = o_p(1). \end{aligned}$$

The last equality holds because of assumption (H5) and (3.8). This completes the proof of the lemma.

Proof of Lemma 3.4.3. Recall that

$$\hat{\xi}_i = Y_i - \widehat{H}_{\hat{\theta}}(Z_i) = [Y_i - H_{\theta_0}(Z_i)] + [H_{\theta_0}(Z_i) - \widehat{H}_{\hat{\theta}_n}(Z_i)] = \xi_i + \tilde{\delta}_i.$$

Note that $\tilde{\delta}_i$ are not independent due to the common use of validation sample, we further decompose the residual as

$$\tilde{\delta}_i = H_{\theta_0}(Z_i) - \widehat{H}_{\theta_0}(Z_i) + \widehat{H}_{\theta_0}(Z_i) - \widehat{H}_{\hat{\theta}_n}(Z_i) = s_i + t_i, \quad \text{say}.$$

Proof of (3.30)(a). Let

$$\begin{aligned} &\bar{C}_n = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(z)\tilde{\delta}_i^2 d\varphi(z), \qquad B_n = \frac{1}{n^2}\sum_{i=1}^n \int K_{hi}^2(z)\xi_i\tilde{\delta}_i d\varphi(z), \\ &\phi_5(Z_i,\widetilde{\eta}_k) = \frac{1}{nN}\int K_{hi}^2(z)[m_{\theta_0}(Z_i+\widetilde{\eta}_k) - H_{\theta_0}(Z_i)]^2 d\varphi(z), \end{aligned}$$

$$\phi_6(Z_i, \widetilde{\eta}_k, \widetilde{\eta}_l) = \frac{1}{n} \int K_{hi}^2(z)[m_{\theta_0}(Z_i + \widetilde{\eta}_k) - H_{\theta_0}(Z_i)][m_{\theta_0}(Z_i + \widetilde{\eta}_l) - H_{\theta_0}(Z_i)] d\varphi(z).$$

Let C_n^* denote the $\widehat{C}_n$ with $d\hat{\varphi}$ replaced by $d\varphi$. Arguing as for (3.31), it suffices to show that (3.30)(a) holds with $\widehat{C}_n$ replaced by C_n^*. Decompose

$$C_n^* = \frac{1}{n^2} \sum_{i=1}^{n} \int K_{hi}^2(z)(\xi_i + \tilde{\delta}_i)^2 d\varphi(z) = \widetilde{C}_n + \bar{C}_n + 2B_n.$$

We claim

$$\text{(a)} \quad nh^{p/2}\bar{C}_n = o_p(1), \qquad \text{(b)} \quad nh^{p/2}B_n = o_p(1). \tag{3.36}$$

To prove (3.36)(a), by the triangular inequality, we obtain

$$\begin{aligned}
\bar{C}_n &= \frac{1}{n^2} \sum_{i=1}^{n} \int K_{hi}^2(z)[H_{\theta_0}(Z_i) - \widehat{H}_{\theta_0}(Z_i) + \widehat{H}_{\theta_0}(Z_i) - \widehat{H}_{\hat{\theta}_n}(Z_i)]^2 d\varphi(z) \\
&\leq 2\Big[\frac{1}{n^2} \sum_{i=1}^{n} \int K_{hi}^2(z)[H_{\theta_0}(Z_i) - \widehat{H}_{\theta_0}(Z_i)]^2 d\varphi(z) \\
&\qquad + \frac{1}{n^2} \sum_{i=1}^{n} \int K_{hi}^2(z)[\widehat{H}_{\theta_0}(Z_i) - \widehat{H}_{\hat{\theta}_n}(Z_i)]^2 d\varphi(z)\Big] \\
&=: 2\,\bar{C}_{n1} + 2\,\bar{C}_{n2}, \qquad \text{say.}
\end{aligned}$$

First, consider

$$\begin{aligned}
\bar{C}_{n1} &= \frac{1}{n^2N^2} \sum_{i=1}^{n} \sum_{k,l=1}^{N} \int K_{hi}^2(z)[m_{\theta_0}(Z_i + \widetilde{\eta}_k) - H_{\theta_0}(Z_i)][m_{\theta_0}(Z_i + \widetilde{\eta}_l) - H_{\theta_0}(Z_i)] d\varphi(z) \\
&= \frac{1}{nN} \sum_{i=1}^{n} \sum_{k=1}^{N} \phi_5(Z_i, \widetilde{\eta}_k) + \frac{1}{nN^2} \sum_{i=1}^{n} \sum_{k \neq l=1}^{N} \phi_6(Z_i, \widetilde{\eta}_k, \widetilde{\eta}_l) =: \bar{C}_{n11} + \bar{C}_{n12}.
\end{aligned}$$

The summand $\bar{C}_{n11}$ is a two sample U statistic with the kernel function ϕ_5. Direct calculations yield the following facts.

$$E\phi_5 = O(1/(nNh^p)), \qquad E(\phi_5|Z_i) = \frac{1}{nN}\int K_{hi}^2(z)\sigma^2(Z_i)d\varphi(z),$$
$$E(\phi_5|\widetilde{\eta}_k) = \frac{K_2}{nNh^p}\int [m_{\theta_0}(z+\widetilde{\eta}_k) - H_{\theta_0}(z)]^2 f(z)d\varphi(z) + o_p(1/(nNh^p)),$$
$$\mathrm{Var}(E(\phi_5|Z_i)) = O\Big(\frac{1}{n^2N^2h^{3p}}\Big), \qquad \mathrm{Var}(E(\phi_5|\widetilde{\eta}_k)) = O\Big(\frac{1}{n^2N^2h^{2p}}\Big).$$

Because $\lambda = \lim N/n < \infty$, by Lemma 3.6.1, we obtain

$$\sqrt{N}\bar{C}_{n11} = O_p\Big(\mathrm{Var}(E(\phi_5|Z_i)) + \lambda \mathrm{Var}(E(\phi_5|\widetilde{\eta}_k))\Big) = O_p\big(1/(nNh^{3p/2})\big).$$

Therefore, (W1) implies

$$nh^{p/2}\bar{C}_{n11} = O_p\Big(\frac{1}{Nh^p\sqrt{N}}\Big) = o_p(1).$$

Next, consider $\bar{C}_{n12}$. It is a two sample degenerated U statistic with the kernel function ϕ_6. Similar to the analysis of Q_3 in Lemma 3.3.1, we have $\mathrm{Var}(C_{n12}) = O\big(N^{-2}(nh^p)^{-2}\big)$. Hence under (W1),

$$nh^{p/2}C_{n12} = O_p\Big(\frac{nh^{p/2}}{Nnh^p}\Big) = O_p\Big(\frac{1}{\sqrt{N}\sqrt{Nh^p}}\Big) = o_p(1).$$

Therefore $nh^{p/2}\bar{C}_{n1} = o_p(1)$.

Next, consider

$$\bar{C}_{n2} \;=\; \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\Big\{\frac{1}{N}\sum_{k=1}^{N}[m_{\hat{\theta}_n}(Z_i+\widetilde{\eta}_k) - m_{\theta_0}(Z_i+\widetilde{\eta}_k)]\Big\}^2 d\varphi(z)$$

$$
\begin{aligned}
&= \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\Big\{\frac{1}{N}\sum_{k=1}^{N}\big[d_{nik}+u_n^T\dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k)\big]\Big\}^2 d\varphi(z)\\
&\le \frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\Big\{\frac{1}{N}\sum_{k=1}^{N}d_{nik}\Big\}^2 d\varphi(z)\\
&\qquad\qquad +\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\Big\{\frac{1}{N}\sum_{k=1}^{N}u_n^T\dot{m}_{\theta_0}(Z_i+\widetilde{\eta}_k)\Big\}^2 d\varphi(z)\\
&:= \bar{C}_{n21}+\bar{C}_{n22}, \qquad \text{say.}
\end{aligned}
$$

By the facts (3.4) and $n\|u_n\|^2 = O_p(1)$, we obtain

$$
\bar{C}_{n21} = o_p(\|u_n\|^2)O_p\Big(\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)d\varphi(z)\Big) = o_p(n^{-2}h^{-p}).
$$

The facts $N^{-1}\sum_{k=1}^{N}\dot{m}_{\theta_0}(z+\widetilde{\eta}_k) = \dot{H}_{\theta_0}(z)+o_p(1)$, $n\|u_n\|^2 = O_p(1)$ and (H6) yield

$$
\bar{C}_{n22} = O_p\Big(\frac{2}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\big[u_n^T\dot{H}_{\theta_0}(Z_i)\big]^2 d\varphi(z)\Big) = O_p(n^{-2}h^{-p}).
$$

Hence, by assumption (W1), we obtain $nh^{p/2}\bar{C}_{n2} = nh^{p/2}O_p\big((n^2h^p)^{-1}\big) = o_p(1)$, thereby completing the proof of (3.36)(a).

Next, consider

$$
B_n = \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\xi_i s_i d\varphi(z) + \frac{1}{n^2}\sum_{i=1}^{n}\int K_{hi}^2(z)\xi_i t_i d\varphi(z) =: B_{n1}+B_{n2}.
$$

Recall the notation in (3.32). Rewrite

$$
B_{n1} = \frac{1}{n^2N}\sum_{i=1}^{n}\sum_{k=1}^{N}\int K_{hi}^2(z)\xi_i\Delta_{ik}d\varphi(z).
$$

Algebra shows that $E(B_{n1}) = 0$ and

$$\begin{aligned}
\mathrm{Var}(B_{n1}) &= \frac{1}{n^4N^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{N} E\int K_{hi}^2(y)K_{hj}^2(z)\xi_i\xi_j\Delta_{ik}\Delta_{jl}d\varphi(y)d\varphi(z) \\
&= O\Big(\frac{1}{n^3Nh^{2p}}\Big).
\end{aligned}$$

Therefore, $nh^{p/2}B_{n1} = o_p(1)$. An argument similar to the one used in the analysis of $\bar{C}_{n2}$ yields that $nh^{p/2}B_{n2} = o_p(1)$, thereby completing the proof of (3.36)(b), and also of (3.30)(a).

Proof of (3.30)(b). Rewrite

$$\begin{aligned}
\widehat{\Gamma}_n &= \frac{2h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)(\xi_i+\tilde{\delta}_i)(\xi_j+\tilde{\delta}_j)d\hat{\varphi}(z)\Big)^2 \\
&= \widetilde{\Gamma}_n + \frac{2h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)(\xi_i\tilde{\delta}_j+\xi_j\tilde{\delta}_i+\tilde{\delta}_i\tilde{\delta}_j)d\hat{\varphi}(z)\Big)^2 \\
&\quad + \frac{4h^p}{n^2}\sum_{i\neq j}\int K_{hi}(z)K_{hj}(z)\xi_i\xi_j d\hat{\varphi}(z)\int K_{hi}(z)K_{hj}(z)(\xi_i\tilde{\delta}_j+\xi_j\tilde{\delta}_i+\tilde{\delta}_i\tilde{\delta}_j)d\hat{\varphi}(z) \\
&=: \widetilde{\Gamma}_n + \Gamma_{n1} + \Gamma_{n2}, \qquad \text{say.}
\end{aligned}$$

It suffices to show that $\Gamma_{n1} = o_p(1)$ and $\Gamma_{n2} = o_p(1)$. The triangular inequality implies that

$$\begin{aligned}
\Gamma_{n1} &\leq \frac{6h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\xi_i\tilde{\delta}_j d\hat{\varphi}(z)\Big)^2 + \frac{6h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\tilde{\delta}_i\xi_j d\hat{\varphi}(z)\Big)^2 \\
&\quad + \frac{6h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\tilde{\delta}_i\tilde{\delta}_j d\hat{\varphi}(z)\Big)^2 \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

Substituting $s_j + t_j$ for $\tilde{\delta}_j$ in I_1, it can be seen that

$$
\begin{aligned}
I_1 &\leq \frac{12h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\xi_i t_i d\hat{\varphi}(z)\Big)^2 + \frac{12h^p}{n^2}\sum_{i\neq j}\Big(\int K_{hi}(z)K_{hj}(z)\xi_i s_i d\hat{\varphi}(z)\Big)^2 \\
&=: I_{11} + I_{12}.
\end{aligned}
$$

Rewrite

$$
I_{11} = \frac{12h^p}{n^2}\sum_{i\neq j=1}^{n}\Big(\frac{1}{N}\sum_{k=1}^{N}\int K_{hi}(z)K_{hj}(z)\xi_i[m_{\hat{\theta}_n}(Z_j+\widetilde{\eta}_k) - m_{\theta_0}(Z_j+\widetilde{\eta}_k)]d\hat{\varphi}(z)\Big)^2.
$$

Analogous to the analysis of $\bar{C}_{n2}$, by (3.7) and the Cauchy-Schwarz inequality, for $1 \leq i \neq j \leq n$, we obtain

$$
\begin{aligned}
&\int K_{hi}(z)K_{hj}(z)\xi_i\Big\{\frac{1}{N}\sum_{k=1}^{N}\big[m_{\hat{\theta}_n}(Z_j+\widetilde{\eta}_k) - m_{\theta_0}(Z_j+\widetilde{\eta}_k)\big]\Big\}d\hat{\varphi}(z) \\
&= O_p\Big(\int \Big|K_{hi}(z)K_{hj}(z)\xi_i\frac{1}{N}\sum_{k=1}^{N}\big[m_{\hat{\theta}_n}(Z_j+\widetilde{\eta}_k) - m_{\theta_0}(Z_j+\widetilde{\eta}_k)\big]\Big|d\varphi(z)\Big) \\
&\leq O_p\Big(\Big\{\int K_{hi}^2(z)\xi_i^2 d\varphi(z) \times \int K_{hj}^2(z)\Big\{\frac{1}{N}\sum_{k=1}^{N}\big[d_{nik} + u_n^T\dot{m}_{\theta_0}(Z_j+\widetilde{\eta}_k)\big]\Big\}^2 d\varphi(z)\Big\}^{1/2}\Big) \\
&= O_p(h^{-p/2})O_p((nh^p)^{-1/2}) = O_p(n^{-1/2}h^{-p}).
\end{aligned}
$$

Hence $I_{11} = h^p O_p(n^{-1/2}h^{-p}) = o_p(1)$.

Regarding I_{12}, let I_{12}^* denote the I_{12} with $\hat{\varphi}$ replaced by φ, then it suffices to prove that $I_{12}^* = o_p(1)$ by (3.5). Rewrite

$$
I_{12}^* = \frac{h^p}{n^2N^2}\sum_{i\neq j=1}^{n}\sum_{k,l=1}^{N}\int K_{hi}(y)K_{hj}(y)K_{hi}(z)K_{hj}(z)\xi_i^2\Delta_{jk}\Delta_{jl}d\varphi(y)d\varphi(z).
$$

Define

$$\phi_7(D_i, D_j, \widetilde{\eta}_k) = \frac{1}{2N}\int K_{hi}(y)K_{hj}(y)K_{hi}(z)K_{hj}(z)[\xi_i^2\Delta_{jk}^2 + \xi_j^2\Delta_{ik}^2]d\varphi(y)d\varphi(z),$$
$$\phi_8(D_i, D_j, \widetilde{\eta}_k, \widetilde{\eta}_l) = \frac{1}{2}\int K_{hi}(y)K_{hj}(y)K_{hi}(z)K_{hj}(z)[\xi_i^2\Delta_{jk}\Delta_{jl} + \xi_j^2\Delta_{ik}\Delta_{il}]d\varphi(y)d\varphi(z).$$

Then I_{12}^* can be rewritten as

$$\begin{aligned} I_{12}^* &= \frac{h^p}{n^2N}\sum_{i\neq j=1}^{n}\sum_{k=1}^{N}\phi_7(D_i, D_j, \widetilde{\eta}_k) + \frac{h^p}{n^2N^2}\sum_{i\neq j=1}^{n}\sum_{k\neq l=1}^{N}\phi_8(D_i, D_j, \widetilde{\eta}_k, \widetilde{\eta}_l) \\ &=: L_1 + L_2, \qquad \text{say.} \end{aligned}$$

Both L_1 and L_2 are two sample U statistics. Verify that by (A2), (A5) and (W1),

$$E(I_{12}^*) = E(L_1) = \frac{(n-1)h^p}{nN}E\phi_6 = O((Nh^p)^{-1}) = o(1).$$

Furthermore, by calculating the second moments of the conditional expectations in Lemma 2.6.4, it can be shown that, under (A2), (A4), (A5) and (W1),

$$E\phi_7^2 = O\big(N^{-2}h^{-3p}\big), \quad \mathrm{Var}(E(\phi_7|D_i)) = O\big((Nh^p)^{-2}\big),$$
$$\mathrm{Var}(E(\phi_7|D_i, D_j)) = O\big(N^{-2}h^{-3p}\big), \quad \mathrm{Var}(E(\phi_7|\widetilde{\eta}_k)) = O\big((Nh^p)^{-1}\big).$$

Lemma 2.6.4 implies that $\mathrm{Var}(L_1) = o(1)$. Thereby $L_1 = o_p(1)$. Similarly, Lemma 2.6.5 yields that $L_2 = o_p(1)$. The results $I_2 = o_p(1)$ and $I_3 = o_p(1)$ are obtained in a similar manner. Details are skipped for the sake of brevity of the chapter. The fact $\Gamma_{n2} = o_p(1)$ is derived by using the fact that

$$h^p n^{-2} \sum_{i \neq j} \Big(\int K_{hi}(z) K_{hj}(z) |\xi_i| |\xi_j| d\hat{\varphi}(z) \Big)^2 = O_p(1)$$

proved in KN, the application of Cauchy-Scharwz inequality and the fact that $\Gamma_{n1} = o_p(1)$. This completes the proof of (3.30)(b) and also of Lemma 3.4.3.

We further briefly discuss the consistency of these tests. We establish that under some regularity conditions, $|\widehat{\mathcal{T}}_n| \to_p \infty$, under certain fixed alternatives, which implies the consistency of the sequences of tests based on $\widehat{\mathcal{T}}_n$.

Recall the definitions of $H(z)$ and $T(H)$ in the beginning of Section 3.3.1. Let θ_n be an consistent estimator of $T(H)$ and define

$$\xi_i = Y_i - H(Z_i), \quad \xi_{ni} = Y_i - \widehat{H}_{\theta_n}(Z_i),$$
$$C_n = \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z) \xi_{ni}^2 d\hat{\varphi}(z), \quad \Gamma_n = \frac{2h^p}{n^2} \sum_{i \neq j=1}^n \Big(\int K_{hi}(z) K_{hj}(z) \xi_{ni} \xi_{nj} d\hat{\varphi}(z) \Big)^2.$$

Let $\mathcal{T}_n := nh^{p/2} \Gamma_n^{-1/2} (\widehat{M}_n(\theta_n) - C_n)$. Then the theorem below presents the asymptotic behavior of the proposed test under certain alternative hypotheses.

Theorem 3.4.2. *Suppose (A1), (A2), (A4), (A5),(F1), (F2), (H3), (K), (W1) and (W2) hold and the alternative hypothesis* $H_1 : \mu(x) = m(x)$, $x \in \mathcal{C}$ *satisfies that* $\inf_\theta \rho(H, H_\theta) > 0$ *and* $T(H)$ *is unique. Then* $|\mathcal{T}_n| \to_p \infty$ *for any consistent estimator* θ_n *of* $T(H)$.

By Lemma 3.3.2, $\hat{\theta}_n$ is consistent for $T(H)$, therefore the above theorem implies that $|\widehat{\mathcal{T}}_n| \to \infty$ in probability under the same regularity conditions, and the test based on $\widehat{\mathcal{T}}_n$ is consistent against the alternative m for which $\inf_\theta \rho(H, H_\theta) > 0$. The proof of Theorem 3.4.2 is similar to that of Theorem 5.1 in KS with slight modifications. The techniques used for analyzing $W_n(\theta)$ in Lemma 3.3.1 and $\widehat{D}_n(\theta)$ in the proof of Theorem 3.3.2 are enough to produce the conclusions. Details are skipped for the sake of brevity.

3.5 Simulation

In this section, we present the results of a Monte Carlo study of the proposed estimation and testing procedures for $p = 1, 2$. For $p = 1$, both linear and nonlinear functions are chosen as the underlying true regression to generate the primary and validation data. For $p = 2$, a linear regression is assumed. Various values of the ratio N/n are selected to demonstrate its role on the performance of these inference procedures. Throughout the simulation, the kernel function K is chosen as $K(u) = 0.75(1 - u^2)I_{(|u|\le 1)}$ for $p = 1$ while $K(u) = 0.75^2(1 - u_1^2)(1 - u_2^2)I_{(|u_1|\le 1, |u_2|\le 1)}$ for $p = 2$. All of the results obtained are based on 1000 replications.

We need to determine the two bandwidths for the implementation of the above estimation and testing procedures. As mentioned in the beginning of Section 3.4, one bandwidth used for estimating f_Z is $w = c(\log n/n)^{1/(p+4)}$, $c > 0$. We propose to obtain c by minimizing, w.r.t. c, the unbiased cross-validation criterion $UCV(w)$ developed in Wolfgang, Marron and Wand (1990), where

$$UCV(w) = \frac{(R(K))^p}{nw^p} + \frac{1}{n(n-1)w^p} \sum_{i\neq j=1}^{n} (K * K - K)\Big(\frac{Z_i - Z_j}{w}\Big),$$

with $R(K) = \int K^2(x)dx$ and $K * K(x) = \int K(y)K(x - y)dy$. We apply a grid search to choose the optimal coefficient c starting from 0.1 with step 0.02, i.e.,

$$c_n^* := \text{argmin}_{0.1\le c\le 10}\, UCV\Big(c(\log n/n)^{1/(p+4)}\Big), \qquad w_{opt} = c_n^*(\log n/n)^{1/(p+4)}.$$

For the bandwidth h, in order to satisfy (W2), we choose to use $h = \hat{\sigma}_Z\, n^{-1/3}$ for $p = 1$ recommended by Sepanski and Carroll (1993) and $h = n^{-1/4.5}$ for $p = 2$ as used in KS.

In order to interpret the performance of the proposed estimator $\hat{\theta}_n$, we also present the performance of the KS estimator $\tilde{\theta}_{KS}$. Recall that in KS, the measurement error density f_η is assumed to be known.

Both the means and square root of mean square error (RMSE) of the two estimators are reported. For the proposed estimator $\hat{\theta}_n$, N/n are chosen as 1, 1/2 and 1/10 to illustrate how N/n affects the estimator performance. In both linear and nonlinear cases, the bias and RMSE decrease as the sample sizes increase. In the linear case, as shown in Example 3.3.1, the asymptotic variance of $\hat{\theta}_n$ is the same as that of $\tilde{\theta}_{KS}$. This is also reflected in this finite sample study as the RMSEs of $\hat{\theta}_n$ and $\tilde{\theta}_{KS}$ in Table 3.1 are very similar for all of the three choices of N/n. In the nonlinear case, Table 3.2 shows that the obtained RMSE of $\hat{\theta}_n$ is larger than $\tilde{\theta}_{KS}$ and it decreases as N/n increases from 1/10 to 1.

In the testing procedure, with nominal level 0.05, the empirical level and power are obtained by computing $\#\{|\widehat{\mathcal{T}}_n| \ge 1.96\}/1000$. The sample size ratios are chosen as $N/n = 4, 1$ and $1/4$. Both the linear and nonlinear regressions are used as the null for $p = 1$ while the linear regression is chosen as the null for $p = 2$. The empirical power is obtained under various choices of alternative models.

3.5.1 Finite sample performance of $\hat{\theta}_n$

In this subsection we report the findings of a finite sample performance of the estimator $\hat{\theta}_n$ in linear and nonlinear cases.

The linear case with $q = 1 = p$. In this case, we generated the data from (3.1) with

$$m_\theta(x) = \theta x, \quad \theta_0 = 1, \tag{3.37}$$

where $\varepsilon \sim \mathcal{N}_1(0, 0.2^2), \eta \sim \mathcal{N}_1(0, 0.1^2), Z \sim U[-1, 1]$. Then

$$H_\theta(z) = \theta z, \quad \widehat{H}_\theta(z) = \frac{1}{N}\sum_{k=1}^{N} \theta(z + \widetilde{\eta}_k) = \theta(z + \bar{\eta}).$$

The two bandwidths are chosen as described above. Throughout the simulation, $\mathcal{C} = [-1, 1]$, and G is the uniform measure on $[-1, 1]$. Hence, as noted in Example 3.3.1, here $\Sigma_2 = 0$ and the asymptotic variances of $\hat{\theta}_n$ is equivalent to that of $\tilde{\theta}_{KS}$. This fact is also reflected in this finite sample study by observing that the RMSE of $\hat{\theta}_n$ remains the same for different choices of N/n as seen in the Table 3.1.

The nonlinear case with $q = 1 = p$. In this section, the regression function is

$$m_\theta(x) = e^{\theta x}, \quad \theta_0 = -1, \tag{3.38}$$

and all other setup is the same as in the above simulation for the linear case. Then the regression function, given z, is

$$H_\theta(z) = e^{\theta^2\sigma_\eta^2/2}\, e^{\theta z}, \quad \widehat{H}_\theta(z) = \frac{1}{N}\sum_{k=1}^{N} e^{\theta(z+\widetilde{\eta}_k)} = e^{\theta z}\frac{1}{N}\sum_{k=1}^{N} e^{\theta\widetilde{\eta}_k}.$$

$N/n = 1$	(n, N)	(100,100)	(200,200)	(400,400)	(600,600)
	mean($\hat{\theta}_n$)	1.0016	0.9994	1.0005	0.9996
	RMSE($\hat{\theta}_n$)	0.0389	0.0298	0.0194	0.0163
$N/n = 1/2$	(n, N)	(100,50)	(200,100)	(400,200)	(600,300)
	mean($\hat{\theta}_n$)	0.9979	0.9985	1.0005	1.0005
	RMSE($\hat{\theta}_n$)	0.0381	0.0295	0.0194	0.0165
$N/n = 1/10$	(n, N)	(100,10)	(200,20)	(400,40)	(600,60)
	mean($\hat{\theta}_n$)	0.9974	0.9984	1.0001	0.9999
	RMSE($\hat{\theta}_n$)	0.0399	0.0299	0.0195	0.0170
KS	n	100	200	400	600
	mean($\tilde{\theta}_{KS}$)	0.9999	0.9996	1.0006	0.9995
	RMSE($\tilde{\theta}_{KS}$)	0.0393	0.0298	0.0194	0.0170

Table 3.1: Performance of $\hat{\theta}_n, \tilde{\theta}_n$ in the linear case (3.37), $p = 1$.

In this case, the second term Σ_2 in the asymptotic variance is calculated as

$$\sigma_{\theta_0}(x, y) = e^{\sigma_\eta^2}(e^{\sigma_\eta^2} - 1)e^{\theta_0(x+y)}, \quad \Sigma_2 = e^{2\sigma_\eta^2}(e^{\sigma_\eta^2} - 1)\Big[\int (x + \sigma_\eta^2\theta_0)e^{\theta_0 x}dG(x)\Big]^2 > 0.$$

. Table 3.2 shows the consistency of $\hat{\theta}_n$ as the bias is very little and the RMSE decreases as the samples sizes increase. The RMSEs of $\hat{\theta}_n$ are larger than those of $\tilde{\theta}_{KS}$, for all chosen values of N/n. Furthermore, the RMSE of $\hat{\theta}_n$ decreases as N/n increases.

The linear case with $q = 2 = p$. We further consider the case $m_\theta(x) = \theta_1 x_1 + \theta_2 x_2$, where $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$ and $x = (x_1, x_2)^T \in \mathbb{R}^2$. The true parameter $\theta_0 = (1, 1)$ is used to generate the data. Denote $Z_i = (Z_{i1}, Z_{i2})^T$ and $\eta_i = (\eta_{i1}, \eta_{i2})^T$ for $1 \le i \le n$. Both Z_{i1} and Z_{i2} are generated independently from $U[-1, 1]$ while η_{i1} and η_{i2} are generated from $\mathcal{N}_1(0, 0.1^2)$ and $\mathcal{N}_1(0, 0.2^2)$, respectively. Then $X_i = (X_{i1}, X_{i2})$ is obtained as the sum of Z_i and η_i. The primary data $\{(Y_i, Z_i), 1 \le i \le n\}$ are obtained with the above regression function and the error ε following $\mathcal{N}_1(0, 0.2^2)$. The validation data $\{\widetilde{\eta}_k, 1 \le k \le N\}$ are independently simulated from η_i. The bandwidth w is obtained based on the UCV criterion

$N/n = 1$	(n, N)	(100,100)	(200,200)	(400,400)	(600,600)
	mean($\hat{\theta}_n$)	-0.9999	-1.0004	-0.9999	-1.0002
	RMSE($\hat{\theta}_n$)	0.0360	0.0249	0.0172	0.0141
$N/n = 1/2$	(n, N)	(100,50)	(200,100)	(400,200)	(600,300)
	mean($\hat{\theta}_n$)	-1.0032	-1.0004	-1.0000	-1.0001
	RMSE($\hat{\theta}_n$)	0.0392	0.0267	0.0181	0.0149
$N/n = 1/10$	(n, N)	(100,10)	(200,20)	(400,40)	(600,60)
	mean($\hat{\theta}_n$)	-1.0023	-1.0009	-1.0004	-1.0003
	RMSE($\hat{\theta}_n$)	0.0498	0.0358	0.0245	0.0200
KS	n	100	200	400	600
	mean($\tilde{\theta}_{KS}$)	-1.0005	-0.9998	-0.9998	-1.0002
	RMSE($\tilde{\theta}_{KS}$)	0.0321	0.0233	0.0162	0.0132

Table 3.2: Performance of $\hat{\theta}_n, \tilde{\theta}_n$ in the nonlinear case (3.38), $p = 1$.

with $p = 2$ while h is taken as $h = n^{-1/4.5}$ as used in KS. In this case, $\mathcal{C} = [-1, 1]^2$ and G is the uniform measure on $[-1, 1]^2$. The choices of N/n are the same as the previous cases. Both means and RMSE of the estimator $\hat{\theta}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2})^T$ and $\tilde{\theta}_{KS} = (\tilde{\theta}_{KS,1}, \tilde{\theta}_{KS,2})^T$ are presented in Table 3.3. It shows small estimation bias and reduced RMSE for increased sample sizes.

3.5.2 Test performance

Here we present the test performance of the proposed test associated with $\widehat{M}_n(\hat{\theta}_n))$ in terms of empirical level and power for different alternative hypotheses and various sample size ratio choices.

The case $q = 1 = p$. The finite sample performance of the $\widehat{\mathcal{T}}_n$ test is assessed for both the above linear (3.37) and nonlinear (3.38) regression models as the null. For each case, the three different alternatives are chosen to obtain the empirical power of a member of the class of the proposed tests.

$N/n=1$	(n,N)	(100,100)	(200,200)	(300,300)	(400,400)
	mean($\hat{\theta}_{n,1}$)	0.9953	1.0004	0.9999	1.0013
	RMSE($\hat{\theta}_{n,1}$)	0.0728	0.0397	0.0332	0.0275
	mean($\hat{\theta}_{n,2}$)	0.9989	1.0032	1.0013	0.9999
	RMSE($\hat{\theta}_{n,2}$)	0.0634	0.0398	0.0307	0.0271
$N/n=1/2$	(n,N)	(100,50)	(200,100)	(300,150)	(400,200)
	mean($\hat{\theta}_{n,1}$)	0.9943	1.0000	0.9998	1.0007
	RMSE($\hat{\theta}_{n,1}$)	0.0779	0.0399	0.0332	0.0269
	mean($\hat{\theta}_{n,2}$)	0.9975	1.0011	1.0009	0.9983
	RMSE($\hat{\theta}_{n,2}$)	0.0644	0.0395	0.0308	0.0261
$N/n=1/10$	(n,N)	(100,10)	(200,20)	(300,30)	(400,40)
	mean($\hat{\theta}_{n,1}$)	0.9928	0.9990	0.9990	1.0006
	RMSE($\hat{\theta}_{n,1}$)	0.0813	0.0400	0.0333	0.0275
	mean($\hat{\theta}_{n,2}$)	0.9892	0.9965	0.9980	0.9971
	RMSE($\hat{\theta}_{n,2}$)	0.0679	0.0399	0.0311	0.0273
KS	n	100	200	300	400
	mean($\tilde{\theta}_{KS,1}$)	0.9957	1.0006	0.9999	1.0013
	RMSE($\tilde{\theta}_{KS,1}$)	0.0732	0.0397	0.0334	0.0275
	mean($\tilde{\theta}_{KS,2}$)	0.9999	1.0037	1.0017	1.0002
	RMSE($\tilde{\theta}_{KS,2}$)	0.0633	0.0399	0.0306	0.0271

Table 3.3: Performance of $\hat{\theta}_n, \tilde{\theta}_n$ in the linear case with $p=2$

Model 0: $Y = X + \varepsilon$ $\qquad$ $Y = e^{-X} + \varepsilon$

Model 1: $Y = X + 0.2X^2 + \varepsilon$ $\qquad$ $Y = e^{-X} + 0.2X^2 + \varepsilon$

Model 2: $Y = X + 0.5\sin(2X) + \varepsilon$ $\qquad$ $Y = e^{-X} + 0.5\sin(2X) + \varepsilon$

Model 3: $Y = XI_{(X\leq 0.5)} + 0.5I_{(X>0.5)} + \varepsilon$ $\qquad$ $Y = e^{-X}I_{(X\leq 0.5)} + e^{-0.5}I_{(X>0.5)} + \varepsilon$

The entities G, K, f_Z, U and ε are as in the $q = 1 = p$ cases in Section 3.5.1.

The empirical levels under Model 0 and the empirical power under Models 1, 2, 3, are shown in Table 3.4 with increasing sample sizes. The left and right panels of Table 3.4 correspond to the left and right panels of the models above, respectively. With nominal level 0.05, the empirical level is well controlled in the linear case while it is slightly conservative under the exponential null model with larger sample sizes. The proposed test rejects the

$N/n=4$				$N/n=4$		
(n,N)	(100,400)	(200,800)	(500,2000)	(100,400)	(200,800)	(500,2000)
Model 0	0.031	0.041	0.046	0.032	0.022	0.039
Model 1	0.686	0.982	1.000	0.326	0.748	1.000
Model 2	0.392	0.781	1.000	0.996	1.000	1.000
Model 3	0.913	1.000	1.000	0.321	0.740	1.000
$N/n=1$				$N/n=1$		
(n,N)	(100,100)	(200,200)	(500,500)	(100,100)	(200,200)	(500,500)
Model 0	0.029	0.037	0.041	0.029	0.032	0.040
Model 1	0.671	0.984	1.000	0.345	0.746	1.000
Model 2	0.409	0.790	1.000	0.996	1.000	1.000
Model 3	0.921	1.000	1.000	0.327	0.730	1.000
$N/n=1/4$				$N/n=1/4$		
(n,N)	(100,25)	(200,50)	(500,125)	(100,25)	(200,50)	(500,125)
Model 0	0.056	0.048	0.052	0.037	0.033	0.038
Model 1	0.679	0.967	1.000	0.344	0.744	1.000
Model 2	0.483	0.814	0.996	0.995	1.000	1.000
Model 3	0.902	1.000	1.000	0.339	0.733	0.996

Table 3.4: Empirical level and power under linear null model (left panel) and nonlinear null model (right panel) for $p=1$

null hypotheses with high power for moderate and large sample sizes, for all the three chosen alternatives. Moreover, it is observed that for the same primary sample size n, the empirical power changes little when the validation sample size N increases. This finding also is somewhat consistent with the theoretical result that the sample size ratio N/n does not play a critical role in the asymptotic behavior of the proposed test statistic.

The case $q=2=p$. In this case, the setup is the same as in the estimation subsection 3.5.1 for $p=2$. We investigate the empirical level of the proposed test under Model $\emptyset$ and power under alternative Models I, II and III below.

$$\text{Model } \emptyset\text{:} \quad Y=\theta_0^T X+\varepsilon, \quad \theta_0=(1,1)^T, \quad X=(X_1,X_2)^T$$

$$\text{Model I:} \quad Y=\theta_0^T X+0.2X_1X_2+\varepsilon$$

$$\text{Model II:} \quad Y=\theta_0^T X+0.5\sin(2X_1X_2)+\varepsilon$$

$$\text{Model III:} \quad Y=\theta_0^T X I_{(\theta_0^T X\le 0.5)}+0.5I_{(\theta_0^T X>0.5)}+\varepsilon$$

$N/n = 4$				
(n, N)	(40,160)	(100,400)	(200,800)	(400,1600)
Model $\emptyset$	0.041	0.034	0.048	0.047
Model I	0.052	0.103	0.278	0.596
Model II	0.581	0.900	0.953	0.998
Model III	0.654	0.910	0.961	0.999
$N/n = 1$				
(n, N)	(40,40)	(100,100)	(200,200)	(400,400)
Model $\emptyset$	0.040	0.035	0.049	0.046
Model I	0.056	0.121	0.295	0.608
Model II	0.585	0.900	0.958	0.998
Model III	0.636	0.908	0.963	0.997
$N/n = 1/4$				
(n, N)	(40,10)	(100,25)	(200,50)	(400,100)
Model $\emptyset$	0.092	0.081	0.077	0.069
Model I	0.105	0.188	0.340	0.658
Model II	0.604	0.901	0.961	0.999
Model III	0.638	0.912	0.964	0.999

Table 3.5: Empirical level and power under linear null model for $p = 2$

The numerical findings are summarized in Table 3.5. It is observed that the empirical levels preserve the nominal size 0.05 for larger sample sizes when $N/n = 1$ and 4. The empirical levels for $N/n = 1/4$ are slightly inflated due to limited validation sample and it decreases towards the nominal size 0.05 as sample sizes increase. The empirical power under all chosen alternatives increases as the sample size increases.

3.6 Proofs

In this section, we provide the detailed proofs of Lemmas 3.3.1 and 3.3.2. To proceed, we state Theorem B.1 in Sepanski and Lee (1995) and recall Lemmas 2.6.4 and 2.6.5 pertaining to some two sample U statistics.

Lemma 3.6.1. *Let $\{x_i\}, i = 1, ..., n$ be an i.i.d. sample, and $\{v_j\}, j = 1, ..., m$ be another i.i.d. sample which is independent with $\{x_i\}$. The functions $\psi_n(v, x, h)$ is a sequence of ran-*

dom functions with a bandwidth h. In addition, suppose the following hold.

(1) There exists square integrable functions $q_1(v)$ and $q_2(x)$ such that $|E\{\psi_n(v,x,h)|v\}| \leq q_1(v)$ and $|E\{\psi_n(v,x,h)|x\}| \leq q_2(x)$,

(2) $\lim_{n\to\infty} E\{\psi_n(v,x,h)|v\} \to p_1(v)$, a.e., and $\lim_{n\to\infty} E\{\psi_n(v,x,h)|x\} \to p_2(x)$, a.e. for some measurable functions $p_1(v)$ and $p_2(x)$, and

(3) $\lim_{n\to\infty} \sqrt{n}E\{\psi_n(v,x,h)\} \to 0$.

Then

$$\frac{1}{m\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m}\psi_n(v_j,x_i,h) \to_d \mathcal{N}_1(0, \lambda Var\{p_1(v)\} + Var\{p_2(x)\}),$$

where $\lambda = \lim_{n\wedge m\to\infty}(n/m)$, assumed to be finite.

Proof of Lemma 3.3.1. Let $W_n^*(\theta)$ be the $W_n(\theta)$ with $d\hat{\varphi}$ replaced by $d\varphi$. To proceed, for $1 \leq i \neq j \leq n, 1 \leq k \neq l \leq N$, define

$$\begin{aligned}
&\phi_1(Z_i,\widetilde{\eta}_k) = \int K_{hi}^2(z)[m_\theta(Z_i+\widetilde{\eta}_k) - H_\theta(Z_i)]^2 d\varphi(z),\\
&\phi_2(Z_i,Z_j,\widetilde{\eta}_k,\widetilde{\eta}_l) = \int K_{hi}(z)K_{hj}(z)[m_\theta(Z_i+\widetilde{\eta}_k) - H_\theta(Z_i)][m_\theta(Z_j+\widetilde{\eta}_l) - H_\theta(Z_j)]d\varphi(z),\\
&\phi_3(Z_i,Z_j,\widetilde{\eta}_k) = \int K_{hi}(z)K_{hj}(z)[m_\theta(Z_i+\widetilde{\eta}_k) - H_\theta(Z_i)][m_\theta(Z_j+\widetilde{\eta}_k) - H_\theta(Z_j)]d\varphi(z),\\
&\phi_4(Z_i,\widetilde{\eta}_k,\widetilde{\eta}_l) = \int K_{hi}(z)K_{hj}(z)[m_\theta(Z_i+\widetilde{\eta}_k) - H_\theta(Z_i)][m_\theta(Z_i+\widetilde{\eta}_l) - H_\theta(Z_i)]d\varphi(z).
\end{aligned}$$

Rewrite

$$\begin{aligned}
&W_n^*(\theta)\\
= &\frac{1}{n^2}\int\sum_{i,j=1}^{n}K_{hi}(z)K_{hj}(z)[\widehat{H}_\theta(Z_i) - H_\theta(Z_i)][\widehat{H}_\theta(Z_j) - H_\theta(Z_j)]d\varphi(z)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2N^2}\sum_{i,j=1}^{n}\sum_{k,l=1}^{N}\int K_{hi}(z)K_{hj}(z)[m_\theta(Z_i+\widetilde{\eta}_k)-H_\theta(Z_i)][m_\theta(Z_j+\widetilde{\eta}_l)-H_\theta(Z_j)]d\varphi(z)\\
&= \frac{1}{n^2N^2}\Big\{\sum_{i=j}\sum_{k=l}\phi_1(Z_i,\widetilde{\eta}_k)+\sum_{i\neq j}\sum_{k\neq l}\phi_2(Z_i,Z_j,\widetilde{\eta}_k,\widetilde{\eta}_l)\\
&\qquad +\sum_{i\neq j}\sum_{k=l}\phi_3(Z_i,Z_j,\widetilde{\eta}_k)+\sum_{i=j}\sum_{k\neq l}\phi_4(Z_i,\widetilde{\eta}_k,\widetilde{\eta}_l)\Big\}\\
&=: Q_1+Q_2+Q_3+Q_4, \qquad \text{say.}
\end{aligned}
$$

Because $E[m_\theta(Z+\widetilde{\eta})|Z=z]=H_\theta(z)$, then $EQ_2=EQ_4=0$. Therefore

$$
\begin{aligned}
& E(W_n^*(\theta)) = EQ_1+EQ_3=\frac{1}{nN}E\phi_1(Z,\widetilde{\eta})+\frac{1}{N}E\phi_3(Z_1,Z_2,\widetilde{\eta})\\
&= \frac{1}{nN}E\int K_h^2(z-Z_1)\sigma_\theta^2(Z_1)d\varphi(z)+\frac{1}{N}E\int K_h(z-Z_1)K_h(z-Z_2)\sigma_\theta(Z_1,Z_2)d\varphi(z)\\
&= O\Big(\frac{1}{N}\int\sigma_\theta^2(z)dG(z)\Big)=O\Big(A_N(\theta)\Big).
\end{aligned}
$$

This fact and (3.5) imply that $W_n(\theta)=W_n^*(\theta)+o_p(W_n^*(\theta))=W_n^*(\theta)+o_p(1/N)$. Therefore, it suffices to prove that (3.10) holds with $W_n(\theta)$ replaced by $W_n^*(\theta)$.

We investigate each quantity in the decomposition of $W_n^*(\theta)$. First, Q_1 is a two sample U statistic with kernel function ϕ_1. In order to apply Lemma 3.6.1, it is necessary to calculate the projections of ϕ_1, i.e.,

$$
E(\phi_1|Z_i)=h^p\int K_{hi}^2(z)\sigma^2(Z_i)d\varphi(z), \quad E(\phi_1|\widetilde{\eta}_k)=O_p\Big(K_1\int[m_\theta(z+\widetilde{\eta}_k)-H_\theta(z)]^2dG(z)\Big).
$$

It can be verified that $E\phi_1$, $\mathrm{Var}(E(\phi_1|Z_i))$ and $\mathrm{Var}(E(\phi_1|\widetilde{\eta}_k))$ are all finite. Therefore Lemma 3.6.1 implies that for finite $0<\lambda<\infty$,

$$
Z_n := \sqrt{N}\times\frac{1}{nN}\sum_{i=1}^{n}\sum_{k=1}^{N}[\phi_1(Z_i,\widetilde{\eta}_k)-E\phi_1]
$$

is asymptotically normally distributed. Hence

$$N\,Q_1 = \frac{1}{nN^{1/2}} Z_n + \frac{1}{nN^{1/2}} E(\phi_1) = o_p(1). \tag{3.39}$$

Similarly, Q_2 is a two sample statistic with the kernel function ϕ_2. Note that $E[\phi_2|Z_i] = E[\phi_2|Z_i, Z_j] = E[\phi_2|Z_i, Z_j, \widetilde{\eta}_k] = E[\phi_2|\widetilde{\eta}_k] = E[\phi_2|Z_i, \widetilde{\eta}_k] = 0$ for $1 \le i \ne j \le n$ and $1 \le k \le N$.

To proceed further, define

$$\begin{aligned}
\tilde{\phi}_2(\widetilde{\eta}_k, \widetilde{\eta}_l) &= \int [m_\theta(z + \widetilde{\eta}_k) - H_\theta(z)][m_\theta(z + \widetilde{\eta}_l) - H_\theta(z)] f^2(z) d\varphi(z), \\
\widetilde{Q}_2 &= \frac{1}{N(N-1)} \sum_{k \ne l = 1}^{N} \tilde{\phi}_2(\widetilde{\eta}_k, \widetilde{\eta}_l).
\end{aligned}$$

Calculation shows that,

$$\mathrm{Var}(\phi_2) = O\Big(\frac{1}{h^{2p}}\Big), \quad E(\phi_2|\widetilde{\eta}_k, \widetilde{\eta}_l) = O_p\Big(\tilde{\phi}_2(\widetilde{\eta}_k, \widetilde{\eta}_l)\Big), \quad \mathrm{Var}\big(\tilde{\phi}_2(\widetilde{\eta}_k, \widetilde{\eta}_l)\big) = \Sigma_\theta.$$

Then Lemma 2.6.5 implies that

$$\mathrm{Var}\big(N(Q_2 - \widetilde{Q}_2)\big) = O\big((nh^p)^{-2}\big) = o(1).$$

Thereby $N(Q_2 - \widetilde{Q}_2) = o_p(1)$. Furthermore, with $\widetilde{Q}_2$ being a degenerated U statistic, applying Theorem 1 in Hall (1984) to $\widetilde{Q}_2$ yields $N\widetilde{Q}_2 \to \mathcal{N}_1(0, 2\gamma(\theta))$, which in turn implies

$$NQ_2 \to \mathcal{N}_1(0, 2\gamma(\theta)). \tag{3.40}$$

Next, consider Q_3, which is defined with kernel function ϕ_3. Algebra shows

$$\begin{aligned}
E(\phi_3|Z_i,Z_j) &= O_p\Big(\frac{1}{N}\int K_{hi}(z)K_{hj}(z)\sigma_\theta(Z_i,Z_j)d\varphi(z)\Big),\\
E(\phi_3|Z_i) &= O_p\Big(\frac{1}{N}\int K_{hi}(z)\sigma_\theta(Z_i,z)f(z)d\varphi(z)\Big),\ E(\phi_3)=O_p\Big(A_N(\theta)\Big),\\
E(\phi_3|\widetilde{\eta}_k) &= O_p\Big(\frac{1}{N}\int[m_\theta(z+\widetilde{\eta}_k)-H_\theta(z)]^2f_Z^2(z)d\varphi(z)\Big),\\
E(\phi_3|Z_i,\widetilde{\eta}_k) &= O_p\Big(\frac{1}{N}\int K_{hi}(z)[m_\theta(Z_i+\widetilde{\eta}_k)-H_\theta(Z_i)][m_\theta(z+\widetilde{\eta}_k)-H_\theta(z)]f_Z(z)d\varphi(z)\Big).
\end{aligned}$$

Furthermore, the second moments of the above projections can be derived

$$E\phi_3^2=O(N^{-2}h^{-2p}),\quad E[E(\phi_3|Z_i,Z_j)]^2=O(N^{-2}h^{-2p}),\quad E[E(\phi_3|Z_i)]^2=O(N^{-2}h^{-p}),$$
$$E[E(\phi_3|\widetilde{\eta}_k)]^2=O(N^{-2}),\quad E[E(\phi_3|Z_i,\widetilde{\eta}_k)]^2=O(N^{-2}h^{-p}).$$

Then Lemma 2.6.4 yields that

$$\mathrm{Var}(Q_3)=O\Big(\frac{4}{n}\mathrm{Var}(E(\phi_3|Z_1)+\frac{1}{N}\mathrm{Var}(E(\phi_3|\widetilde{\eta}_1))\Big)=O\Big(\frac{4}{nh^pN^2}+\frac{1}{N^3}\Big).$$

Hence $N\,(Q_3-E\phi_3)=o_p(1)$ for sufficient large N and $E\phi_3=O(1/N)$. Therefore

$$Q_3=Q_3-E\phi_3+E\phi_3 \;=\; E\phi_3+o_p(1/N)=A_N(\theta)+o_p(1/N). \tag{3.41}$$

The same routine argument and Lemma 2.6.4 applying to Q_4 lead to

$$\mathrm{Var}(Q_4)=O\Big(\frac{1}{(nh^p)^2N^2}\Big)\text{ and }EQ_4=0. \tag{3.42}$$

Combining all the results of components of W_n (3.39)–(3.42), one can see that Q_2 dominates the convergence rate of W_n and only Q_3 contributes to the mean of W_n asymptotically, which in turn yields (3.10).

Proof of Lemma 3.3.2. KS has shown that $\tilde{\theta}_n = T(H) + o_p(1)$ by proving that

$$\sup_{\theta \in \Theta} |M_n(\theta) - \rho(H, H_\theta)| = o_p(1). \tag{3.43}$$

In the current setup, if we show

$$\sup_{\theta \in \Theta} |\widehat{M}_n(\theta) - M_n(\theta)| = o_p(1), \tag{3.44}$$

then, by (3.43),

$$\sup_{\theta \in \Theta} |\widehat{M}_n(\theta) - \rho(H, H_\theta)| = o_p(1).$$

Then arguing as in KS will yield the lemma.

Proof of (3.44). By the Cauchy-Schwarz inequality,

$$|\widehat{M}_n(\theta) - M_n(\theta)| \leq W_n(\theta) + 2[W_n(\theta) M_n(\theta)]^{1/2}.$$

It suffices to show that $\sup_\theta |W_n(\theta)| = o_p(1)$ and $\sup_\theta |M_n(\theta)| = O_p(1)$. The the compactness of Θ and $H_\theta \in L_2(G)$ imply that $\sup_\theta |\rho(H, H_\theta)|$ is finite. Furthermore, (3.43) shows that $\sup_\theta |M_n(\theta)| = O_p(1)$.

Now we study $W_n(\theta)$. By Lemma 3.3.1, $W_n(\theta) = o_p(1)$, for every $\theta \in \Theta$. Moreover, for any θ_1, $\theta_2 \in \Theta$,

$$
\begin{aligned}
&|W_n(\theta_1) - W_n(\theta_2)| \\
&\le \Big(\int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[\widehat{H}_{\theta_1}(Z_i) - H_{\theta_1}(Z_i) + \widehat{H}_{\theta_2}(Z_i) - H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) \\
&\qquad \times \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[\widehat{H}_{\theta_1}(Z_i) - H_{\theta_1}(Z_i) - \widehat{H}_{\theta_2}(Z_i) + H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) \Big)^{1/2}.
\end{aligned}
$$

The first term on the right hand side above is $O_p(1)$ due to the boundedness of H_θ and the compactness of Θ. Similar to the proof on p143 of KS, the second term is bounded above by

$$
\begin{aligned}
2 \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[\widehat{H}_{\theta_1}(Z_i) - \widehat{H}_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) \\
+2 \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[H_{\theta_1}(Z_i) - H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z).
\end{aligned} \tag{3.45}
$$

The first term can be rewritten as 2 times the factor

$$
\begin{aligned}
&\int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[\widehat{H}_{\theta_1}(Z_i) - H_{\theta_1}(Z_i) + H_{\theta_1}(Z_i) - H_{\theta_2}(Z_i) + H_{\theta_2}(Z_i) - \widehat{H}_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) \\
&\le 3 \Big[W_n(\theta_1) + \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[H_{\theta_1}(Z_i) - H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) + W_n(\theta_2) \Big] \\
&= 3 \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[H_{\theta_1}(Z_i) - H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) + o_p(1).
\end{aligned}
$$

The last claim holds because $NA_N(\theta) = O(1)$ and hence by Lemma 3.3.1, $W_N(\theta) = o_p(1)$, for all $\theta \in \Theta$. Then, by (H3), the bound (3.45) is further bounded from the above by

$$
8 \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z)[H_{\theta_1}(Z_i) - H_{\theta_2}(Z_i)] \Big\}^2 d\hat{\varphi}(z) + o_p(1)
$$

$$\leq \; 8\|\theta_1 - \theta_2\|^{2\beta} \int \Big\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(z) r(Z_i) \Big\}^2 d\hat{\varphi}(z) + o_p(1) = \|\theta_1 - \theta_2\|^{2\beta} O_p(1),$$

by (3.8) applied with $\alpha = r$. The above result and the compactness of Θ implies that $\sup_{\theta \in \Theta} |W_n(\theta)| = o_p(1)$, by a routine argument.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abarin T. and Wang L. (2009). Second-order least squares estimation of censored regression models. *Journal of Stat. Plann. and Infer.* **139**, 125–135.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics* **24**, 3–61.

Berkson, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45**, 164–180.

Bhattacharya, P. K., Chernoff, H. and Yang, S. S. (1983). Nonparametric estimation of the slope of a truncated regression. *Ann. Statist.* **11**, 505–514.

Bierens, H. (1987). Kernel Estimators of Regression Functions. *Advances in Econometrics*, Fifth Worldcongress, Vol 1, Bewley (ed.), 99–144.

Bosq, D. (1998). Nonparametric statistics for stochastic processes, 2nd Edition, Springer, Berlin.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. (2006). Measurement Error in Nonlinear Models: A Modern Perspective. Second edition. Chapman and Hall, London.

Cheng, C. L., and Van Ness, J. W. (1999). Statistical regression with measurement error. John Wiley & Sons.

Delaigle, A., Hall, P. and Qiu, P. (2006). Nonparametric methods for solving the Berkson errors-in-variables problem. *J. R. Statist. Soc. B*, **68**(2), 201–220.

Du, L., Zou, C. and Wang, Z. (2011). Nonparametric regression function estimation for errors-in-variables models with validation data. *Statistica Sinica*, **21**, 1093–1113.

Fuller, W. A. (1987). Measurement error models. John Wiley & Sons.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of Goodness-of-Fit tests for regression models. *Test*, **22**, 361–411.

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

Huwang, L. and Huang Y.H. (2000). On errors-in-variables in polynomial regression–Berkson case. *Statistica Sinica* **10**, 923–936.

Jones, M. C. and Signorini, D. F. (1997). A Comparison of Higher-Order Bias Kernel Density Estimators *J. Amer. Statist. Assoc.* **439**, 1063–1073.

Kim, K. H., Chao, S. and Härdle, W. K. (2016), Simultaneous Inference for the Partially Linear Model with a Multivariate Unknown Function when the Covariates are Measured with Errors. SFB 649 Discussion Paper 2016–024.

Koul, L. H. and Ni, P. (2004). Minimum distance regression model checking. *Journal of Stat. Plann. and Infer.* **119**, 109–141.

Koul, H. L. and Song, W. (2009) Minimum distance regression model checking with Berkson measurement errors. *Annals of Statistics*, **37**(1), 132–156.

Koul, H. L., Song, W. and Liu, S. (2014). Model checking in Tobit regression via nonparametric smoothing. *J. Multivariate Anal.* **125**, 36–49.

Lee, L. F. and Sepanski, J. H. (1995). Estimation of Linear and Nonlinear Errors-in-Variables Models Using Validation Data, *Journal of the American Statistical Association*, **90**(429), 130–140.

Mack, Y.P. and Silverman, B.W. (1982). Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. Gebiete*, **61**, 405–415.

Schennach, S.M. (2013). Regressions with Berkson errors in covariates–a nonparametric approach. *The Annals of Statistics*, **41**(3), 1642–1668.

Sepanski, J. H. and Carroll, R. J. (1993). Semiparametric quasilikelihood and variance function estimation in measurement error models. *J. Econometrics*, **58**, 223–256.

Sepanski, J.H. and Lee, L. (1995). Semiparametric estimation of nonlinear errors-in-variables models with validation study. *Journal of Nonparametric Statistics.* **4(4)**, 365–394.

Serfling R. J. (1981). Approximation theorems of mathematical statistics. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ.

Song, W. (2008). Model checking in errors-in-variables regression. *J. Multivariate Anal.* **99**, 2406–2443.

Song, W. (2009). Lack-of-fit testing in errors-in-variables regression model with validation data. *Statist. Probab. Lett.* **79**, 765–773.

Song, W. (2011). Distribution-free test in Tobit mean regression model. *Journal of Stat. Plann. and Infer.* **141**, 2891–2901.

Song, W. and Yao, W. (2011). A lack-of-fit test in Tobit errors-in-variables regression models. *Statist. Probab. Lett.* **81**, 1792–1801.

Stute, W., Thies, S. and Zhu, L. (1998). Model checks for regression: an innovation process approach. *Ann. Statist.*, **26**, 1916–1934.

Stute, W., Xue, L. and Zhu, L. (2007). Empirical likelihood inference in nonlinear errors-in-covariables models with validation data. *J. Amer. Statist. Assoc.*,**102**, 332-346.

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

Wang, L. (1998). Estimation of censored linear errors-in-variables models. *J. Econometrics* **84**, 383–400.

Wang, L. (2004). Estimation of nonlinear models with Berkson measurement errors. *Annals of Statistics*,**32**, 2559–2579.

Wang, L. (2007). A simple nonparametric test for diagnosing nonlinearity in Tobit median regression model. *Statist. Probab. Lett.* **77**, 1034–1042.

Wang, Q. and Rao, J. N. K. (2002). Empirical Likelihood-Based Inference in Linear Errors-in-Covariables Models with Validation Data. *Biometrika*, **89**, 2, 345–358.

Wolfgang, H., Marron, J. S. and Wand, M. P. (1990). Bandwidth choice for density derivatives. *J. R. Statist. Soc. B*, **52**(1), 223–232.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **175**(2), 263–289.