Goodness-of-fit tests in measurement error models with replications

by

Weijia Jia

M.S., Nankai University, 2010

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Abstract

In this dissertation, goodness-of-fit tests are proposed for checking the adequacy of parametric distributional forms of the regression error density functions and the error-prone predictor density function in measurement error models, when replications of the surrogates of the latent variables are available.

In the first project, we propose goodness-of-fit tests on the density function of the regression error in the errors-in-variables model. Instead of assuming that the distribution of the measurement error is known as is done in most relevant literature, we assume that replications of the surrogates of the latent variables are available. The test statistic is based upon a weighted integrated squared distance between a nonparametric estimate and a semi-parametric estimate of the density functions of certain residuals. Under the null hypothesis, the test statistic is shown to be asymptotically normal. Consistency and local power results of the proposed test under fixed alternatives and local alternatives are also established. Finite sample performance of the proposed test is evaluated via simulation studies. A real data example is also included to demonstrate the application of the proposed test.

In the second project, we propose a class of goodness-of-fit tests for checking the parametric distributional forms of the error-prone random variables in the classic additive measurement error models. We also assume that replications of the surrogates of the error-prone variables are available. The test statistic is based upon a weighted integrated squared distance between a nonparametric estimator and a semi-parametric estimator of the density functions of the averaged surrogate data. Under the null hypothesis, the minimum distance estimator of the distribution parameters and the test statistics are shown to be asymptotically normal. Consistency and local power of the proposed tests under fixed alternatives and local alternatives are also established. Finite sample performance of the proposed tests is evaluated via simulation studies.

Goodness-of-fit tests in measurement error

models with replications

by

Weijia Jia

M.S., Nankai University, 2010

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2018

Approved by:

Major Professor
Dr. Weixing Song

Copyright

Abstract

In this dissertation, goodness-of-fit tests are proposed for checking the adequacy of parametric distributional forms of the regression error density functions and the error-prone predictor density function in measurement error models, when replications of the surrogates of the latent variables are available.

In the first project, we propose goodness-of-fit tests on the density function of the regression error in the errors-in-variables model. Instead of assuming that the distribution of the measurement error is known as is done in most relevant literature, we assume that replications of the surrogates of the latent variables are available. The test statistic is based upon a weighted integrated squared distance between a nonparametric estimate and a semi-parametric estimate of the density functions of certain residuals. Under the null hypothesis, the test statistic is shown to be asymptotically normal. Consistency and local power results of the proposed test under fixed alternatives and local alternatives are also established. Finite sample performance of the proposed test is evaluated via simulation studies. A real data example is also included to demonstrate the application of the proposed test.

In the second project, we propose a class of goodness-of-fit tests for checking the parametric distributional forms of the error-prone random variables in the classic additive measurement error models. We also assume that replications of the surrogates of the error-prone variables are available. The test statistic is based upon a weighted integrated squared distance between a nonparametric estimator and a semi-parametric estimator of the density functions of the averaged surrogate data. Under the null hypothesis, the minimum distance estimator of the distribution parameters and the test statistics are shown to be asymptotically normal. Consistency and local power of the proposed tests under fixed alternatives and local alternatives are also established. Finite sample performance of the proposed tests is evaluated via simulation studies.

Table of Contents

Acknowledgments

I would like to express my appreciation to Dr. Weixing Song, my major advisor, for all his knowledge, guidance and suggestions. He was always available when I had doubts or questions. I learned a lot from him through many helpful discussions. My appreciation also goes to Dr. Christopher Pinner for his willingness to serve as the chairperson of the examining committee for my doctoral degree. I would also like to thank Dr. James Neill, Dr. Juan Du and Dr. Zhoumeng Lin for serving on my committee, and for their valuable guidance through my study process.

I would like to thank the department of statistics for offering me graduate assistantships so that I could come to the states and complete my graduate studies at Kansas State University. I would like to thank everyone in the department for their kindness, thank all the professors in the department for the their excellent courses and for their help. Finally, many thanks to my family for their endless love, support, understanding and encouragement. I would never have been able to finish my graduate studies without their support.

Chapter 1

Introduction

The relationship between a random variable Y and a random vector X is often investigated through a regression model. In the classic regression, both Y and X are assumed to be observable. However, in many experiments, it is expensive or impossible to observe X. Instead, one observes some surrogates for predictors. These models are often called errors-in-variables models or measurement errors models.

Extensive research has been done on the estimation of the underlying parameters in the measurement errors models. There is also an increase in research activity in recent years emphasizing the study of lack-of-fit testing of a parametric regression model with measurement errors in the predictors. Relatively, however, there is little published literature aiming at checking the appropriateness of the distributional assumption on regression errors and/or error-prone predictors. The focus of this dissertation is to make an attempt at partly filling this void.

1.1 The Density Function of the Regression Error

Statistical inferences could be made in regression models without knowing the distributions of the regression errors, but more efficient procedures can be developed if these distributions are known. However, misspecified distributional forms can severely undermine the reliability or validity of the conclusions. Goodness-of-fit tests for checking the suitability of specified distributions for the regression errors in the classical regression models have been extensively studied in the literature. Ranging from simple graphical tools to complicated formal analytical tests, the existing methods include histograms and density plots of various residuals, the minimum Hellinger distance test in Beran (1977), and tests based on empirical residual processes in Koul (2002), Khmaladze and Koul (2004, 2009), among others.

It is often the case that in regression models, the predictor X, possibly multidimensional, cannot be observed directly due to some uncontrollable reasons. Instead, observations on some surrogates Z of the variables of interest are available. It is commonly assumed that the surrogate variables Z are related to the latent variables X in an additive way $Z = X + U$, where U is called the measurement error. How to denoise the measurement error from the surrogate data Z and correctly modeling the relationship among true variables Y and X, is the primary objective in measurement error modeling. See the monographs of Fuller (1987), Cheng and Van Ness (1999), Buonaccorsi (2010) for a comprehensive introduction, and Carroll et al. (2006) for more advanced research directions on this field. Compared to the rich statistical explorations on testing the parameters and regression functions in measurement error models, the goodness-of-fit tests for the random components in these models are less developed.

To be specific, in Chapter 2 of this dissertation, we consider the following linear regression model with measurement error.

$$Y = \alpha + \beta^T X + \varepsilon, \quad Z = X + U, \tag{1.1}$$

where Y is a scalar response, X is a d-dimensional latent variable, and U is a d-dimensional measurement error vector. X, U, and the regression error ε are assumed to be independent, and ε has mean 0 and finite variance. Knowing the distributions of the measurement error or other random components in the measurement error models might help us construct more efficient estimates. For example, in the simple linear regression model ($d = 1$) of $Y = \alpha + \beta X + \varepsilon$, $Z = X + U$, if the variance σ_u^2 of U is known, and there is no further distributional assumptions on the latent variable X, the measurement error U, and the regression error ε other than $E(\varepsilon|X) = 0$, $\mathrm{Var}(\varepsilon) < \infty$, then the commonly used estimate is the bias-corrected estimator $\hat{\beta}_n = (S_{ZZ} - \sigma_u^2)^{-1} S_{ZY}$, $\hat{\alpha}_n = \bar{Y} - \hat{\beta}_n \bar{Z}$, where S_{ZZ} is the sample covariance of Z, S_{ZY} is the sample covariance of Z and Y, $\bar{Y}$ and $\bar{Z}$ are the sample means of Y's and Z's, respectively. This estimator is consistent and asymptotically normal even when the actual distribution of ε is misspecified but without violating the basic assumptions of expectation being zero and second moment being finite. However, in addition to the normality assumptions on X, U, if we can further assume the normality on ε, and the ratio of variances of U and ε to be 1, then simulation studies have shown that the adjusted maximum likelihood estimator

$$\tilde{\beta} = [S_{YY} - S_{ZZ} + \sqrt{(S_{YY} - S_{ZZ})^2 + 4S_{ZY}^2}]/2S_{ZY}$$

has smaller mean squared error than that of the bias-corrected estimator, in particular, when the sample size is small.

Therefore, by taking the distributional information of the random variables into account, one can construct more efficient estimators of the underlying parameters in the measurement error models.

Throughout this dissertation, for any generic random variable or vector V, its density function will be denoted as $f_V(\cdot)$. In Chapter 2, we will focus on the goodness-of-fit tests for the following hypothesis on the density function of ε.

$$H_0 : f_\varepsilon(x) = f_\varepsilon(x, \theta), \quad \theta \in \Theta, \quad x \in \mathbb{R}^d \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true.} \tag{1.2}$$

Rewrite the model (1.1) as

$$Y = \alpha + \beta^T Z + \xi, \quad \xi = \varepsilon - \beta^T U. \tag{1.3}$$

Then $f_\xi(v) = \int f_\varepsilon(v + \beta^T u) f_U(u) du$. As argued in Koul and Song (2012), when f_U is assumed to be known, the density functions of ε and ξ are uniquely determined by each other. As a result, testing for H_0 in (1.2) is equivalent to testing for

$$H_0 : f_\xi(v) = f_\xi(v, \theta), \quad \theta \in \Theta, \quad v \in \mathbb{R}^d \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true}, \tag{1.4}$$

where $f_\xi(v, \theta) = \int f_\varepsilon(v + \beta^T u, \theta) f_U(u) du$. Under the assumption that the density function of U is known, Koul and Song (2012) proposed a class of tests for the testing (1.4) based on kernel density estimators of f_ξ obtained from the residuals $Y_i - \hat{\alpha}_n - \hat{\beta}_n^T Z_i$, where $\hat{\alpha}_n, \hat{\beta}_n$ are some $n^{1/2}$-consistent estimators of α, β under H_0 based on a sample (Z_i, Y_i), $1 \leq i \leq n$, from model (1.1). According to Holzmann et al. (2007), the test developed in Koul and Song (2012) can be labeled as a direct test, due to the fact that the test is about hypothesis on the distribution of ξ rather than ε. Recently, Koul et al. (2017) developed an indirect test based on the deconvolution estimate of the density function of ε. Simulation studies show that when the variance of measurement error is small, the direct test performs better than the indirect test based on the comparison with respect to their finite sample powers, but the trend reverses when the variance of the measurement error becomes larger. See Holzmann et al. (2007) and Laurent et al. (2011) for more discussion on the direct and indirect procedures.

For the sake of model identifiability, the variance or the density function of U is often assumed to be known in the measurement error literature. This assumption plays a critical role in the tests developed in Koul and Song (2012), Koul et al. (2017). However, in real applications, the distribution of U is rarely known. To our best knowledge, no test has yet been proposed for checking the hypothesis in (1.2) or (1.4) when f_U is unknown, up to now. In Chapter 2 of this dissertation, we will try to fill out this void by assuming replications can be made on X. Under some regularity assumptions on f_U, the replications make it possible to construct a nonparametric estimate of $f_{\bar{U}}$, the density function of the average of the measurement errors in each replicated observations, which in turn can be used for constructing the test. In fact, the research on estimation problems in measurement error models using replication is abundant, see Blas et al. (2013), Dalen et al. (2009), Delaigle et al. (2008), Gimenez and Patat (2005), Huwang (1995), Lin and Cao (2013), and Xiao et al. (2010) and the references therein.

1.2 The Density Function of the Latent Variable

In the second part of this dissertation, we go on to consider the classical measurement error model. Depending on the assumption about X, measurement error models can be generally classified into two separate types (see Carroll et al. (2006)): functional model, where the X's are viewed as fixed unknown constants, and structural model, where the X's are regarded as random variables. Much recent emphasis has been on structural models and methods (see Huwang (1995), Huang et al. (2006), Thompson and Carter (2007), Lin and Cao (2013)), in that by making no assumptions about the distribution of X, likelihood functions for the functional models are either not available or can only be calculated via complex methods generally with low efficiency. Inference based on structural modelling is generally simpler than that in functional modelling.

When the distribution of X can be well identified, one can always construct better estimates. A convincing example is given by the famous Tweedie formula. Suppose the latent variable X follows a normal distribution, X and U are independent, even if the density function of U is unknown, Tweedie formula states that $E(X|Z = z) = z + \sigma_U^2 p'(z)/p(z)$, where p denotes the density function of Z. Therefore, $E(U|Z = z) = -\sigma_U^2 p'(z)/p(z)$. This amazing result can be directly used for constructing more efficient estimation and testing procedures via the regression calibration technique.

Concerns inevitably arises, in a structural modelling approach, that the estimates and inferences will depend upon the distribution of the X assumed. Misspecification of the distribution for X can result in inconsistent estimators. Many parametric or nonparametric methods are employed to identify the distribution of the latent variable X or dampen the effect of misspecification, see, for example, deconvolution-type methods, both parametric and nonparametric in Section 12.1 of Carroll et al. (2006), nonparametric modelling method in Schafer (2001), flexible parametric modelling method in Carroll et al. (1999), or Richardson et al. (2002), and latent-model robustness in Huang et al. (2006). Parametric structural modelling is generally much more favorable in practice because of its simplicity, potential efficiency, as well as in terms of drawing inferences, accuracy, power, etc.

In Chapter 3, we are interested in developing a goodness-of-fit test for the density function X in the framework of Holzmann et al. (2007), in which the density function of U is assumed to be known. The test in Holzmann et al. (2007) is based on the L_2-distance between a deconvolution density estimator of X and its expected values under the null hypothesis. Three drawbacks can be easily identified in Holzmann et al. (2007)'s procedure: (1) the measurement error is restricted to the cases of ordinary smooth, that is, the characteristic function of the measurement error decays to 0 in the tail at the algebraic rate. This excludes some important measurement errors, such as the normal error; (2) The theoretical development of the test statistic is rather complicated due the complexity of deconvolution technique; (3) The null hypothesis is simple. Although the theory might be able to be extended to composite cases, its derivation is not provided. We shall develop a test procedure by dropping the assumption of the density function of U being known, also the test applies to both ordinary and super smooth measurement errors. The test will be based on the ordinary kernel density estimator of the averaged surrogate observations, thus avoiding the cumbersome deconvolution arguments.

The hypothesis of the goodness-of-fit tests on the density function of X considered in the second project is defined as follows.

$$H_0 : f_X(x) = f_X(x, \theta), \quad \theta \in \Theta, \quad x \in \mathbb{R}^d \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true.} \tag{1.5}$$

The tests are based on certain minimized L_2 distances between a nonparametric density function of Z and the convolution of $f_X(x, \theta)$ and a nonparametric density function of U. This test is labeled as a direct test from Holzmann et al. (2007). The goodness-of-fit testing problem on the density function of X has been also studied by several authors from direct or indirect perspective. For indirect testing, we mention Holzmann and Boysen (2006) for a study of the asymptotic distribution of the integrated square error of a deconvolution kernel density estimator when the error term distribution is assumed to be supersmooth and known. When the density of the measurement error is assumed to be known, Loubes and Marteau (2014) compared the inverse problem procedure (i.e. indirect testing procedure) and direct procedure on a goodness-of-fit test of whether the density of the latent variable is equal to a benchmark density function.

1.3 Measurement Error Models with Replication

The measurement model (1.1) has the non-identifiable issue, as showed in Reiersol and Koopmans (1950), when normality of X is assumed, unless further information about the parameters can be found. The assumptions of variance of the regression error σ_ε^2 or variance of measurement error σ_U^2 being known are commonly used in the measurement error literature. However, the non-identifiability problem will not appear in the replicated measurement error model, since the error variances can be estimated through the replicated data.

The research on estimation problems in measurement error models using replication is abundant. We mention, for instance, White et al. (2001) for developing the regression calibration approach for problems with a replication study where the covariates comprise both continuously distributed and binary variables and the outcome is continuous. Devanarayan and Stefanski (2002) presented a variation of the simex algorithm, which can accommodate heteroscedastic measurement error, when the measurement error variances are unknown but replicate measurements are available. To fit the replicated measurement error data with more robust model, Lin and Cao (2013) assumed the replicated observations jointly follow scale mixtures of normal distribution, and based on this assumption, the maximum likelihood estimates are computed via an EM type algorithm method. Research on estimating the density of X with replicate data available can also be found in the literature, we mention Dalen et al. (2009) for estimating the true exposure densities in the model for a dichotomous outcome variable Y.

To our best knowledge, no test has been proposed for checking the appropriateness of distributional assumption on the regression error ε or the error-prone predictor X, using replication data. A class of goodness-of-fit tests are proposed to check the appropriateness of a specified family of density of the regression error ε or the latent variable X. We assume in this dissertation that for each $X_i, i = 1, 2, \cdots, n$, we have two replications on Z having the additive relation

$$Z_{i1} = X_i + U_{i1}, \quad Z_{i2} = X_i + U_{i2},$$

where U are independent and identically distributed. Moreover, we assume the density function of the measurement error U is symmetric about 0, which plays a crucial role in our tests construction.

Chapter 2

Goodness-of-Fit Tests on the Density Function of the Regression Error

We start with a brief introduction to Koul and Song (2012)'s direct testing procedure. Denote the true parameters of α, β, θ as $\alpha_0, \beta_0, \theta_0$, respectively. Under H_0, the density function of $\xi = \varepsilon - \beta_0^T U$ has the form of $f_\xi(u; \beta_0, \theta_0) := \int f_\varepsilon(u + \beta_0^T v, \theta_0) f_U(v) dv$. Let K be a kernel density function and b_n be a sequence of bandwidths, which are positive numbers tending to 0 as the sample size $n \to \infty$. A kernel density estimator of $f_\xi(\cdot)$ can be defined as

$$\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) = \frac{1}{n} \sum_{i=1}^{n} K_{b_n}(v - \hat{\xi}_i), \tag{2.1}$$

where $\hat{\xi}_i = Y_i - \hat{\alpha}_n - \hat{\beta}_n^T Z_i$, $i = 1, 2, \ldots, n$, $\hat{\alpha}_n$ and $\hat{\beta}_n$ are any $\sqrt{n}$-consistent estimates of α_0 and β_0, respectively, and $K_{b_n}(\cdot) := b_n^{-1} K(\cdot / b_n)$. Denote $f_{\xi b_n}(v; \beta_0, \theta_0) = E_0 \hat{f}_{\xi n}(v; \alpha_0, \beta_0) = \int K_{b_n}(v - u) f_\xi(u; \beta_0, \theta_0) du$, where $\hat{f}_{\xi n}(v; \alpha_0, \beta_0)$ is the same as $\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n)$ with $\hat{\alpha}_n$ and $\hat{\beta}_n$ being replaced by α_0 and β_0. The test proposed in Koul and Song (2012) is based upon the statistic

$$T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n) = \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b_n}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v), \tag{2.2}$$

where Π is a weight function supported on a compact subset of $\mathbb{R}$. To see the rationality of using T_n to construct the test, note that $T_n^0 = \int [\hat{f}_{\xi n}(v; \alpha_0, \beta_0) - f_{\xi b_n}(v; \beta_0, \theta_0)]^2 d\Pi(v)$ is a weighted integration of the L_2-distance between $\hat{f}_{\xi n}(v;\alpha_0,\beta_0)$ and its expectation under the null hypothesis, and T_n is an analogue of T_n^0 with $\alpha_0,\beta_0,\theta_0$ replaced by $\sqrt{n}$-consistent estimates $\hat{\alpha}_n,\hat{\beta}_n$ and $\hat{\theta}_n$. One might be thinking about using other distances, such as L^∞ or L_p to measure the discrepancy between $\hat{f}_{\xi n}(v;\alpha_0,\beta_0)$ and its expectation under the null hypothesis, however, the theoretical derivation of the corresponding asymptotic distributions will be much more complicated than using L_2-distance.

However, often times the density function f_U is unknown in real application, which renders the test procedure of Koul and Song (2012) not applicable in many cases. In the following, we shall assume that replications can be made at each X-value, and the associated measurement errors U are independent and identically distributed. Moreover, we shall assume that U is symmetric about 0 which plays a critical role in our test construction.

This chapter is organized as follows. The test statistic incorporating the replications is constructed in Section 2.1; Technical assumptions and the asymptotic distribution of the test statistic under the null hypothesis will be stated in Section 2.2; Consistency of the test under fixed alternatives, and the power of the test under some local alternatives will be discussed in Section 2.3; Finally, the finite sample performance of the proposed test will be examined through some simulation studies in Section 2.4, together with an application of the proposed test on the Framingham data set. The proofs of all theoretical results are postponed to Section 2.5.

In the sequel, all the integrations are denoted by a single integration sign, single or multiple integration can be understood from the context. Integration limits are understood from $-\infty$ to ∞ unless specified otherwise. For a vector a, $\|a\|$ denotes its L_2 norm, and for a matrix A, $\|A\|$ denotes its Frobenius norm, or $\|A\| = \sqrt{\mathrm{tr}(A^T A)}$.

2.1 Test Statistics

Suppose for each X_i, $i = 1, 2, \ldots, n$, we have two replications of Z from

$$Z_{i1} = X_i + U_{i1}, \quad Z_{i2} = X_i + U_{i2}. \tag{2.3}$$

For two generic random variables V_{i1} and V_{i2}, we use $\bar{V}_i$ to denote their average, and $\tilde{V}_i$ to denote $(V_{i1} - V_{i2})/2$. Then from (2.3),

$$\bar{Z}_i = X_i + \bar{U}_i, \quad \tilde{Z}_i = \tilde{U}_i.$$

Because of the independent and identical structure and symmetry, $\tilde{U}_i$, $\bar{U}_i$ have the same distribution. Instead of considering model (1.1), we can consider the following linear errors-in-variables regression model:

$$Y_i = \alpha + \beta^T X_i + \varepsilon_i, \quad \bar{Z}_i = X_i + \bar{U}_i.$$

Following the idea in Koul and Song (2012) we can define the same entities treating $\bar{Z}_i$ as Z_i, $\bar{U}_i$ as U_i in (2.1), (2.2). Since $\bar{U}_i$ and $\tilde{U}_i$ have the same distribution, so estimating the density function of $\bar{U}_i$ can be realized by estimating the density function of $\tilde{U}_i$, which in turn can be estimated through observations on $\tilde{Z}_i$. To be specific, let L be a d-dimensional kernel function on $\mathbb{R}^d$, and w_n be another sequence of bandwidths. In the sequel, we write b for b_n, w for w_n for the sake of simplicity. Define

$$\hat{f}_{\bar{U}n}(u) = \frac{1}{n}\sum_{i=1}^{n} L_w(u - \tilde{Z}_i), \quad L_w(u) = \frac{1}{w^d} L\left(\frac{u}{w}\right),$$

and redefine $\hat{\xi}_i = Y_i - \hat{\alpha}_n - \hat{\beta}_n^T \bar{Z}_i$,

$$\tilde{f}_{\xi n}(u; \hat{\beta}_n, \hat{\theta}_n) = \int f_\varepsilon(u + \hat{\beta}_n^T t, \hat{\theta}_n) \hat{f}_{\bar{U}n}(t) dt,$$
$$\tilde{f}_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n) = \int K_b(v - u) \tilde{f}_{\xi n}(u; \hat{\beta}_n, \hat{\theta}_n) du.$$

Then the proposed test in this dissertation is based upon

$$T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n) = \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - \tilde{f}_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v). \tag{2.4}$$

Remark 2.1: As suggested in Koul and Song (2012), one can estimate θ_0 using the minimum distance (MD) method. That is, for any preliminary estimators $\hat{\alpha}_n, \hat{\beta}_n$ of α_0 and β_0, we estimate θ_0 using

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} T_n(\hat{\alpha}_n, \hat{\beta}_n, \theta). \tag{2.5}$$

The $\sqrt{n}$-consistency and the asymptotic normality of $\hat{\theta}_n$ can be derived using the similar arguments in Koul and Ni (2004) and Koul and Song (2010).

Meanwhile, the preliminary estimates of α_0 and β_0 can be chosen as the well-known bias-corrected estimate. If no replication on Z is available, and the covariance matrix Σ_U of U is known, then the bias-corrected estimates of α_0 and β_0 are given by $\hat{\alpha}_n = \bar{Y} - \bar{Z}^T \hat{\beta}_n$ and $\hat{\beta}_n = (S_{ZZ} - \Sigma_U)^{-1} S_{ZY}$, where S_{ZZ} and S_{ZY} denote the sample covariance matrices of Z, and of Z and Y, respectively. In our current setup, Σ_U is unknown, but it can be estimated by the data from $\tilde{Z}_i$'s. Therefore, modified bias-corrected estimates of α_0, β_0 can be obtained by replacing Σ_U by the sample covariance matrix of $\tilde{Z}_i$'s, S_{ZZ} and S_{ZY} by $S_{\bar{Z}\bar{Z}}$ and $S_{\bar{Z}Y}$. One can show that such $\hat{\alpha}_n, \hat{\beta}_n$ are still $\sqrt{n}$-consistent and asymptotically normal, even if the regression error distribution is misspecified.

Remark 2.2: In the above development, we assume that the measurement error U_1 and U_2 are identically distributed and symmetric. If the question of interest is to estimate the distribution of X or U, then the assumption of identical distribution is not necessary. Indeed, based on Kotlarski's argument, see Rao (1992), the distributions of X and U can be uniquely determined by the joint distribution of the replicated observations on X, given that the characteristic functions of X and U are non-vanishing. However, if such estimators are used in the proposed test statistic, then the asymptotic distribution of the resulting test statistic might be hard to derive. For example, if Li and Vuong (1998)'s estimator is used, how the parameter in the truncation limit will affect the convergence rate of the statistic T_n, formula (2.4), is not clear. On the other hand, the estimators proposed in Li and Vuong (1998) only deal with the case of univariate X and U, it is not clear what the large sample properties look like in our current multidimensional case. The assumption of symmetry plays an important role in our current setup. The significance of the symmetry lies in the fact that

U_1+U_2 and U_1-U_2 will have the same distribution, and $Z_1-Z_2=U_1-U_2$ simply tells us that the distribution of U_1+U_2 can be estimated using Z_1-Z_2, and without using characteristic functions, or deconvolution related techniques. That said, to develop a more general test without these strong assumptions deserves further study and will be future research.

2.2 Asymptotic Null Distributions

To define a proper test statistic from T_n in (2.4), we have to investigate the asymptotic distribution of T_n under H_0. The following is a list of technical assumptions needed to derive such a result. Throughout, for any generic smooth function $\gamma(x;\eta)$, $\dot{\gamma}_\eta(\cdot)$ and $\ddot{\gamma}_\eta(\cdot)$ denote the first order and the second order derivative of γ with respect to the parameter η, while $\gamma'(x;\eta)$ and $\gamma''(x;\eta)$ denote the first order and the second order derivative of γ with respect to x, and x can be a d-dimensional vector.

The assumptions related to f_ε are

(f1). The density function f_ε and its second order derivative $\ddot{f}_\varepsilon(t)$ are continuous and bounded, $\int |\ddot{f}_\varepsilon(t)|dt<\infty$.

(f2). For any $\sqrt{n}$-consistent estimates $\hat{\beta}_n$, $\hat{\theta}_n$ of β_0, θ_0, respectively,

$$\sup_{u,t}|f_\varepsilon(u+\hat{\beta}_n^Tt,\hat{\theta}_n)-f_\varepsilon(u+\beta_0^Tt,\theta_0)-(\hat{\beta}_n-\beta_0)^T\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{f}_{\varepsilon\theta}(u+\beta_0^Tt,\theta_0)|$$

is of the order $O_p(n^{-1})$.

(f3). For all v,β,θ, $\dot{f}_{\varepsilon\beta}(v+\beta^Tt,\theta)$, $\dot{f}_{\varepsilon\theta}(v+\beta^Tt,\theta)$ are Lipschitz continuous in v. That is, there exists a $B(v+\beta^Tt,\theta)$, a continuous function of v, such that

$$\begin{aligned}&\|\dot{f}_{\varepsilon\beta}(v+bx+\beta^Tt,\theta)-\dot{f}_{\varepsilon\beta}(v+\beta^Tt,\theta)\|+\|\dot{f}_{\varepsilon\theta}(v+bx+\beta^Tt,\theta)-\dot{f}_{\varepsilon\theta}(v+\beta^Tt,\theta)\|\\ \le& b|x|B(v+\beta^Tt,\theta),\end{aligned}$$

where $\dot{f}_{\varepsilon\beta}(v + \beta^T t, \theta)$, $\dot{f}_{\varepsilon\theta}(v + \beta^T t, \theta)$ and $B(v + \beta^T, \theta)$ are integrable and square integrable with respect to t.

The assumption on the weighting function Π is

(w). The weighting function Π has a compact support $\mathscr{C}$ in $\mathbb{R}$, and its derivative $\pi(\cdot)$ is twice continuously differentiable.

For the measurement error U and the density function of $\bar{U}$, we assume that

(g1). The measurement error U is symmetric about 0.

(g2). The density function $f_{\bar{U}}$ of $\bar{U}$ is twice continuously differentiable, $\sup_t \|f''_{\bar{U}}(t)\| < \infty$, and $\int \|f''_{\bar{U}}(t)\|dt < \infty$.

(g3). $\int \sqrt{f_{\bar{U}}(t)}dt < \infty$, and there exists a positive constant $\varepsilon_0 > 0$, such that

$$\sup_{\theta\in[0,1], w\in[0,\varepsilon_0)} \int \left\| \int L(v) v^T f''_{\bar{U}}(t + \theta v w) v dv \right\|^{\frac{1}{2}} dt < \infty.$$

Condition (g3) is needed for deriving an upper bound for the integrated mean squared error of $\hat{f}_{\bar{U}n}$. It can be replaced by the boundedness of $\dot{f}_{\varepsilon\beta}(v + \beta^T t, \theta)$, $\dot{f}_{\varepsilon\theta}(v + \beta^T t, \theta)$.

For the sake of simplicity, we shall use the product kernel, with identical component, to estimate the density function of $\bar{U}$. We assume that

(kl). K and L are univariate and bounded d-dimensional product kernel density functions, respectively, such that $\int_{\mathbb{R}} vK(v)dv = 0$, $\int_{\mathbb{R}^d} vL(v)dv = 0$, and $\int_{\mathbb{R}} v^2 K(v)dv \neq 0$, $\int_{\mathbb{R}^d} vv^T L(v)dv = \mu_2(L) I_{d\times d}$ for some positive constant $\mu_2(L)$.

Here we abuse the notation $\mu_2(L)$ a little bit and it is simply a positive constant. The conditions on K are much weaker than the corresponding conditions adopted in Koul and Song (2012).

About the bandwidths b and w, we assume that

(b1). $nb \to \infty$, $nb^{1/2}w^4 \to 0$ as $n \to \infty$.

(b2). $nw^d \to \infty$.

The bandwidth assumptions $nb \to \infty$ and $nw^d \to \infty$ are commonly used in the univariate and multivariate kernel smoothing estimation procedures. The condition $nb^{1/2}w^4 \to 0$ is required to dampen the effect of estimating $f_{\bar{U}}$ by the d-dimensional kernel density estimate $\hat{f}_{\bar{U}n}$. However, $nb \to \infty$ and $nw^d \to \infty$ imply $nb^{1/2}w^{d/2} \to \infty$, combining this with the assumption $nb^{1/2}w^4 \to 0$, we must have $d < 8$. Therefore, one limitation of the proposed test is that the linear regression model under consideration cannot have more than 8 predictors.

In fact, there are two kernel smoothing procedures involved in the construction of test statistic. The kernel density estimator of f_ξ is a univariate smoothing, and the kernel density estimator of $f_{\bar{U}}$ is a d-dimensional multivariate smoothing. It is well known that the larger the dimension, the more difficult to estimate the density function. The two bandwidth sequences must be selected carefully to make sure the test statistic to have a manageable asymptotic distribution. The limitation of $d < 8$ for the proposed test procedure is another evidence of the unpleasant effect of the curse of dimensionality.

To state our main results, the following notations are needed.

$$
\begin{aligned}
\hat{C}_n =& \frac{1}{n^2}\sum_{i=1}^n \int K_b^2(v-\hat{\xi}_i)d\Pi(v), \quad \hat{\Gamma}_n = 2\int \hat{f}_{\xi n}^2(x;\hat{\alpha}_n,\hat{\beta}_n)\pi^2(x)dx \int (K_*(u))^2du, \\
\Gamma =& 2\int f_\xi^2(v;\beta_0,\theta_0)\pi^2(v)dv \int (K_*(u))^2du, \quad K_*(u) := \int K(v)K(u+v)dv, \qquad (2.6)\\
C_n =& \frac{1}{n^2}\sum_{i=1}^n \int [K_b(v-\xi_i) - EK_b(v-\xi_1)]^2 d\Pi(v)
\end{aligned}
$$

Theorem 2.2.1. *Suppose conditions (f1)-(f3), (w), (g1)-(g3), (b1)-(b2) hold. Then under* H_0*,* $\mathcal{T}_n := nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) - \hat{C}_n)$ *converges to the standard normal in distribution and denoted as* $\mathcal{T}_n \Rightarrow N(0,1)$.

Comparing with Koul and Song (2012)'s result, one can see that replacing the density function of U with a kernel density estimate does not slow down the convergence rate of $T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) - \hat{C}_n$. From the proof we can see that this is a consequence of requiring $nb^{1/2}w^4 \to 0$. Otherwise, the bias caused by replacing $f_{\bar{U}}$ by its kernel density estimator $\hat{f}_{\bar{U}n}$

would make the test statistic $\mathcal{T}_n$ not tight.

According to Theorem 2.2.1, at the significance level α, the null hypothesis will be rejected whenever $|\mathcal{T}_n| > z_{1-\alpha/2}$, where $z_{1-\alpha}$ is the upper $(1-\alpha)$100-th percentile of the standard normal distribution.

To conclude this section, we would like to point out that when applying any tests based on smoothing techniques, bandwidth selection is always a vexing issue. Assumptions (b1) and (b2) are only meaningful when the sample size is sufficiently large, which is seldom true in real applications. In general, two approaches could possibly used to select the bandwidth when implementing such tests. The first one is the naive method, simply using an estimation-based optimal bandwidth, such as a cross-validation bandwidth; The second one is to consider a set of suitable values for the bandwidth and check how sensitive the test is. Some formal discussion on this issue can be found in Gao and Gijbels (2008), and they suggest to select the bandwidth based on the consideration of size and power functions of the tests. However, such development in our current setup is very challenging and we shall investigate this possibility in a future study.

2.3 Consistency and Local Power

A desirable and also a basic requirement for any reasonable test is consistency. That is, the power of the test at any fixed alternative hypothesis should approach 1 when the sample size goes to infinity. To be specific, the alternative hypothesis we are testing is

$$H_a: \quad f_\varepsilon(x) = f_{\varepsilon a}(x), \quad f_{\varepsilon a}(x) \neq f_\varepsilon(x;\theta) \quad \text{for any } \theta, \text{ and } x \text{ a.e.}(\lambda),$$

where λ denotes the Lebesgue measure.

To show that the proposed test is consistent, we have to assume that under the fixed alternative, $\hat{\theta}_n \to \theta_a$, $\hat{\beta}_n \to \beta_a$, $\hat{\alpha}_n \to \alpha_a$ for some θ_a, β_a and α_a. This assumption is by no means a strict one, many estimation procedures can generate such estimates. In fact, as we mentioned in the previous section, the bias-corrected estimate $\hat{\alpha}_n$ and $\hat{\beta}_n$ are consistent and asymptotically normal even the density function of the regression error is misspecified. Using the bias-corrected estimates, one can show that the minimum distance estimate $\hat{\theta}_n$ defined in (2.5) converges to some constant θ_a and is asymptotically normal. The theoretical justification for this consistency and asymptotic normality is similar to the classic regression setup. See Jennrich (1969) and White (1981, 1982) for more details. In the following, we simply assume without justifying rigorously that

(c1). Under the alternative H_a, for some α_a, β_a and θ_a,

$$\sqrt{n}(\hat{\alpha}_n - \alpha_a) = O_p(1), \quad \sqrt{n}(\hat{\beta}_n - \beta_a) = O_p(1), \quad \sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1).$$

Define

$$f_{\xi a}(v;\beta) = \int f_{\varepsilon a}(v + \beta^T t) f_{\bar{U}}(t)dt.$$

We further assume that

(c2). $\int [f_{\xi a}(v;\beta_a) - f_\xi(v;\beta_a,\theta_a)]^2 d\Pi(v) > 0.$

Note that if the bias-corrected estimates are used in the test statistic, then $\beta_a = \beta_0$.

The following theorem states that the proposed test is consistent.

Theorem 2.3.1. *In addition to the conditions (f1),(f3), (w), (g1)-(g3), (b1)-(b2), (c1), (c2), we further assume that (f2) holds for β_a and θ_a. Then under H_a, $|\mathcal{T}_n| \to \infty$ in probability, as $n \to \infty$.*

Next, we shall show that the proposed test possesses nontrivial power for certain local alternatives which converges to the null hypothesis at the rate of $1/\sqrt{nb^{1/2}}$. For this purpose, let φ be a known continuous density on $\mathbb{R}$ with mean 0 and positive variance σ_φ^2, and we consider the following local alternative hypothesis

$$H_{\text{loc}} : f_\varepsilon(x) = (1-\delta_n) f_\varepsilon(x, \theta_0) + \delta_n \varphi(x)$$

with $\delta_n = 1/\sqrt{nb^{1/2}}$. Similar to the fixed alternative case, to show the local power result, we have to assume that the preselected estimate $\hat{\alpha}_n$, $\hat{\beta}_n$ and $\hat{\theta}_n$ are $\sqrt{n}$-consistent. The legitimacy of this assumption is guaranteed by many well-documented arguments in the literature, such as Koul and Song (2010), hence omitted here for the sake of brevity.

Theorem 2.3.2. *Assume all the conditions in Theorem 2.2.1 hold. If the density function $\varphi(\cdot)$ is twice continuously differentiable and the second derivative is bounded, then under H_{loc},*

$$\mathcal{T}_n \Longrightarrow N(\mu_T, 1)$$

as $n \to \infty$, where $\mu_T = \Gamma^{-\frac{1}{2}} \int \left[\int [f_\varepsilon(v+\beta_0^T t, \theta_0) - \varphi(v+\beta_0^T t)] f_{\bar{U}}(t)dt\right]^2 d\Pi(v)$.

Similar to the discussion in Koul and Song (2012), the optimal weight function Π which maximizes the asymptotic local power of the proposed test is the one to maximize the mean of the asymptotic normal distribution, or

$$\Psi(\pi) := \Gamma^{-\frac{1}{2}} \int \left[\int [f_\varepsilon(v+\beta_0^T t, \theta_0) - \varphi(v+\beta_0^T t)] f_{\bar{U}}(t)dt\right]^2 \pi(v)dv.$$

By the Cauchy-Schwarz inequality, and recalling the definition of Γ in (2.6), we have

$$\Psi(\pi) \le \frac{1}{(2\int K_*^2(v)dv)^{1/2}} \cdot \left(\int \frac{\left[\int [f_\varepsilon(v+\beta_0^T t, \theta_0) - \varphi(v+\beta_0^T t)] f_{\bar{U}}(t)dt\right]^4}{f_\xi^2(v;\theta_0,\beta_0)} dv\right)^{1/2}$$

with equality if, and only if,

$$\pi(v) \propto \left[\int [f_\varepsilon(v+\beta_0^T t, \theta_0) - \varphi(v+\beta_0^T t)] f_{\bar{U}}(t)dt\right]^2 \Big/ f_\xi^2(v;\beta_0,\theta_0)$$

for all v. Since the functional Ψ is scale-invariant, that is $\Psi(a\pi) = \Psi(\pi)$ for all positive constant $a > 0$, we may simply take the optimal $\pi(\cdot)$ to be

$$\pi(v) = \left(\frac{\int [f_\varepsilon(v+\beta_0^T u, \theta_0) - \varphi(v+\beta_0^T u)] f_{\bar{U}}(u)du}{\int f_\varepsilon(v+\beta_0^T u, \theta_0) f_{\bar{U}}(u)du}\right)^2.$$

Clearly the optimal weight $\pi(\cdot)$ is practically useless because β_0, θ_0, and the density function $f_{\bar{U}}$ are unknown. Estimators of these unknown parameters and function should be found in order to use the optimal weight in practice.

2.4 Simulation Studies and Application

To evaluate the finite sample performance of the proposed test, we conducted some numerical simulations in this section, together with an illustrative application of the proposed test on the Framingham data set.

2.4.1 Simulation Study

The simulated data are generated from the simple linear regression model $Y = \alpha + \beta X + \varepsilon$. The null hypothesis H_0 we want to test is $\varepsilon \sim N(0, \sigma_\varepsilon^2)$, so the unknown parameter θ in the distribution of ε is σ_ε^2. The latent variable X follows $N(0,1)$, and $U \sim N(0, \sigma_U^2)$. The true values of both α and β are chosen to be 1, σ_ε^2 is chosen to be 0.5^2, and σ_U^2 to be 0.5^2 and 0.8^2. At each X-value, double measurements on Z are obtained. In the simulation study, the sample size n is chosen to be 200 and 500 and 800.

To evaluate the power of the proposed test, nine non-normal distributions will be used to serve as the alternative hypotheses.

- Double exponential distribution with mean 0 and variance 1 (DE(0,1));
- Cauchy distribution with location parameter 0 and scale parameter 1;
- Logistic distribution with location parameter 0 and scale parameter 1;
- t-distribution with degrees of freedom 3, 5 and 10;
- Two-component normal mixture models $0.5N(c, \sigma_\varepsilon^2) + 0.5N(-c, \sigma_\varepsilon^2)$ with $c = 0.5, 0.75$ and 1.

The above chosen alternative distributions deviate from the normal distributions from different directions, some have heavier tails, such as the double exponential, Cauchy and t distributions with small degrees of freedom (3, or 5); some have more than one modes, like the two-component normal mixture distributions. Logistic distribution and t-distribution with degrees of freedom 10 are closer to normal. For the sake of brevity, the two-component normal mixture models $0.5N(c, \sigma_\varepsilon^2) + 0.5N(-c, \sigma_\varepsilon^2)$ will be denoted by $0.5N(\pm c, \sigma_\varepsilon^2)$.

In the simulation, the weighting function Π is taken as a uniform distribution on the closed interval $[-6, 6]$ so that computationally the integration over this interval is nearly same as the integration over the whole real line. The kernel functions K and L are chosen to be standard normal density function, and the bandwidths are chosen to be $b = n^{-1/5}$, $w = n^{-1/4}$ based on the assumptions (b1) and (b2). For each scenario, we repeat the test procedure 500 times, and the empirical level and power are calculated from $\#\{|\mathcal{T}_n| \geq z_{1-\alpha/2}\}/500$. Here, $\hat{\alpha}_n$, $\hat{\beta}_n$ are chosen to be the bias-corrected estimates, $\hat{\theta}_n = \hat{\sigma}_\varepsilon^2 = \hat{S}_\xi^2 - \hat{\beta}_n^2 \hat{\sigma}_{\tilde{U}}^2$, with $\hat{S}_\xi^2$ is the sample variance of $\hat{\xi}_i = Y_i - \hat{\alpha}_n - \hat{\beta}_n \bar{Z}_i$, where $\bar{Z}_i = (Z_{i1} + Z_{i2})/2$, and $\hat{\sigma}_{\tilde{U}}^2$ is the sample variance of $\tilde{U}_i = \tilde{Z}_i = (Z_{i1} - Z_{i2})/2$, $i = 1, 2, \cdots, n$. In the simulation, the significance level α is 0.05.

Table 2.1: Simulation results of the proposed test

	$\sigma_U^2 = 0.5^2$			$\sigma_U^2 = 0.8^2$		
	$n = 200$	$n = 500$	$n = 800$	$n = 200$	$n = 500$	$n = 800$
$N(0, \sigma_\varepsilon^2)$	0.000	0.002	0.002	0.002	0.002	0.004
Logistic(0,1)	0.090	0.200	0.364	0.030	0.128	0.280
Cauchy(0,1)	0.974	0.996	1.000	0.992	0.990	0.990
DE(0,1)	0.640	0.994	1.000	0.354	0.892	0.998
$t(3)$	0.762	0.998	1.000	0.662	0.972	1.000
$t(5)$	0.236	0.634	0.884	0.096	0.390	0.678
$t(10)$	0.022	0.048	0.108	0.016	0.034	0.044
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.000	0.004	0.012	0.002	0.004	0.004
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.098	0.744	0.988	0.012	0.150	0.340
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.930	1.000	1.000	0.396	0.952	0.998

The simulation results in Table 2.1 show that proposed test is more conservative, even for large sample sizes as $n = 800$, as evidenced by the small empirical levels and the small powers against the close-to-normal distributions such as the Logistic, $t(10)$ and the two-component normal mixtures with means ± 0.5. However, for other non-normal distributions, the empirical powers are greatly improved as n gets bigger. It is well known that tests based on smoothing techniques are generally conservative, see Koul and Song (2012) and the references therein for more discussion on this phenomenon. To alleviate the conservativeness, one may resort to some possible resampling techniques. As a preliminary attempt, we have designed the following bootstrap procedure to implement the proposed test in which the same kernel functions and bandwidths are used.

A Bootstrap Test

1. Calculate $\hat{\alpha}_n, \hat{\beta}_n, \hat{\sigma}_\varepsilon^2$ based on the full data set $(Y_i, Z_{i1}, Z_{i2}), i = 1, 2, \ldots, n$;

2. Calculate $\tilde{Z}_i = (Z_{i1} - Z_{i2})/2$ and $\bar{Z}_i = (Z_{i1} + Z_{i2})/2, i = 1, 2, \ldots, n$;

3. Generate a parametric bootstrap sample from $N(0, \hat{\sigma}_\varepsilon^2)$, denoted by $\varepsilon_i^*, i = 1, 2, \ldots, n$;

4. Draw a sample of size n with replacement from $\tilde{Z}_i$ and denote them as $\tilde{Z}_i^*, i = 1, 2, \ldots, n$;

5. Draw a samples $\bar{U}_i^*, i = 1, 2, \ldots, n$ from the kernel density $\hat{f}_{\bar{U}}$ with normal kernel in which the mean is $\tilde{Z}_i^*$, and the standard deviation is the bandwidth w;

6. Compute $Y_i^* = \hat{\alpha}_n + \hat{\beta}_n \bar{Z}_i + \varepsilon_i^* - \hat{\beta}_n \bar{U}_i^*$;

7. Use the bootstrap sample (Y_i^*, Z_{i1}, Z_{i2}) to calculate $\mathcal{T}_n^* = nb^{1/2}(\hat{\Gamma}_n^{-1/2}(T_n(\hat{\alpha}, \hat{\beta}_n, \hat{\theta}_n) - \hat{C}_n))$, and $\hat{\Gamma}_n, T_n(\hat{\alpha}, \hat{\beta}_n, \hat{\theta}_n)$ and $\hat{C}_n$ are all calculated using (Y_i^*, Z_{i1}, Z_{i2});

8. Repeat (3)–(7) B times to obtain B $\mathcal{T}_n^*$-values, denoted as $\mathcal{T}_{n,j}^*, j = 1, 2, \ldots, B$. Then sort these $\mathcal{T}_n^*$-values in ascending order and find $\mathcal{T}_{n[0.025n]}^*, \mathcal{T}_{n[0.975n]}^*$, the 2.5-th and 97.5-th percentiles of $\mathcal{T}_{n,j}^*, j = 1, 2, \ldots, B$.

9. Reject the null hypothesis whenever $\mathcal{T}_n \leq \mathcal{T}_{n[0.025n]}^*$ or $\mathcal{T}_n \geq \mathcal{T}_{n[0.975n]}^*$, where $\mathcal{T}_n$ is obtained from the original data; otherwise, accept the null hypothesis.

The simulation results based on the above Bootstrap algorithm for $n = 100, 200$ are shown in Table 2.2. Clearly, the conservativeness of the proposed test is alleviated using the Bootstrap procedure, and the powers are improved significantly as well.

Table 2.2: Simulation results based on bootstrap

	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 1$	
	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
$N(0, \sigma_\varepsilon^2)$	0.010	0.012	0.018	0.050	0.016	0.044
Logistic(0,1)	0.084	0.174	0.074	0.102	0.064	0.102
Cauchy(0,1)	0.972	0.990	0.980	0.988	0.960	0.988
DE(0,1)	0.542	0.882	0.318	0.620	0.168	0.434
$t(3)$	0.634	0.906	0.528	0.810	0.386	0.678
$t(5)$	0.238	0.460	0.142	0.256	0.114	0.182
$t(10)$	0.098	0.102	0.044	0.086	0.054	0.048
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.016	0.040	0.038	0.044	0.026	0.050
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.120	0.426	0.058	0.116	0.044	0.074
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.774	0.988	0.294	0.658	0.166	0.354

2.4.2 A Real Data Example: Farminham Heart Study

In this subsection, we apply the proposed test procedure to a data set in the Framinham Heart Study. The data set includes 1615 observations from men aged between 31 and 65 years old in several health exams taken two years apart. The variables we are interested in the study include the CHD (the indicator of the first evidence of coronary heart disease within an 8-year period following the second exam), the age at Exam 2, systolic blood pressures (SBP) at Exam 2 and Exam 3, smoking status, and serum cholesterol levels (SCL) at Exam 2 and Exam 3. For each individual, SBP are measured twice by independent examiners at each exam. To check the consistency of the blood pressure measurements between the two exams, we fit a simple linear regression model with the average of log(SBP-50) from Exam 3 being the response variable, and the log(SBP-50) from Exam 2 being the predictor. This log transformation of the SBP is also used in Eckert et al. (1997). Since the true SBP cannot be obtained directly, we treat the two measurements in the Exam 2 as replicates. The statistical hypothesis is to see if the regression error follows a normal distribution with mean 0.

To apply our proposed test, the kernel functions K and L are chosen to be standard normal density function, and the bandwidths b and w are chosen to be $n^{-1/5}$ and $n^{-1/4}$, respectively, where $n = 1615$ is the sample size. Calculation shows that $|\mathscr{T}_n| = 11.0395$ which far exceeds the 95-th percentile 1.96 of standard normal. So the normality of regression error is rejected.

2.5 Proofs

This section contains all the proofs of the main theorems stated in Section 2.2 and 2.3. Since the main idea of the proofs are similar to those in Koul and Song (2012), only differences are presented here for the sake of brevity. In particular, we will focus the discussion on the statistic $T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n)$, which will be decomposed into two parts, one part can be dealt with directly using Koul and Song (2012)'s argument, and another part involving all terms related to the kernel density estimator $\hat{f}_{\bar{U}n}$ has to be investigated separately. The discussions on the normalizing constants $\hat{C}_n$ and $\hat{\Gamma}_n$ are similar to Koul and Song (2012)'s argument, hence omitted for the sake of brevity.

The proof of Theorem 2.2.1: Note that

$$\begin{aligned}\tilde{f}_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n) &= f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n) + \iint K_b(v-u) f_\varepsilon(u + \hat{\beta}_n^T t, \hat{\theta}_n)(\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t))du dt \\ &:= f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n) + R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n),\end{aligned}$$

then the statistics in (2.4) can be written as

$$\begin{aligned}T_n(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n) &= \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n) - R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) \\ &= \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) + \int [R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) \\ &\quad - 2\int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)] R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v). \end{aligned} \tag{2.7}$$

To proceed, we consider the term R_{bw} first. Adding and subtracting $f_\varepsilon(u + \beta_0^T t, \theta_0)$ from

$f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n)$, $E\hat f_{\bar U n}(t)$ from $\hat f_{\bar U n}(t)$, R_{bw} can be written as the sum of the following four terms:

$$\begin{aligned}
R_{bw1} &= \iint K_b(v-u)[f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n)-f_\varepsilon(u+\beta_0^T t,\theta_0)][\hat f_{\bar U n}(t)-E\hat f_{\bar U n}(t)]dudt,\\
R_{bw2} &= \iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)(\hat f_{\bar U n}(t)-E\hat f_{\bar U n}(t))dudt,\\
R_{bw3} &= \iint K_b(v-u)[f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n)-f_\varepsilon(u+\beta_0^T t,\theta_0)][E\hat f_{\bar U n}(t)-f_{\bar U}(t)]dudt,\\
R_{bw4} &= \iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)[E\hat f_{\bar U n}(t)-f_{\bar U}(t)]dudt.
\end{aligned}$$

It is well known that $E\hat f_{\bar U n}(t)=f_{\bar U}(t)+w^2\mu_2(L)\mathrm{tr}(f''_{\bar U}(t))/2+o(w^2)$, then from (f1), we can show that $R_{bw4}=2^{-1}w^2\mu_2(L)\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)\mathrm{tr}(f''_{\bar U}(t))dudt+o(w^2)$. This, together with (g2), one can easily show that $|R_{bw4}(v)|=O(w^2)$ uniformly on v, this in turn implies that $\int R^2_{bw4}(v)d\Pi(v)=O(w^4)$. Hence, by assumption (b1),

$$nb^{\frac{1}{2}}\int R^2_{bw4}(v)d\Pi(v)=O(nb^{1/2}w^4)=o(1). \tag{2.8}$$

Now consider R_{bw3}. For the sake of brevity, denote

$$f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n)-f_\varepsilon(u+\beta_0^T t,\theta_0)-(\hat\beta_n-\beta_0)^T\dot f_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)-(\hat\theta_n-\theta_0)^T\dot f_{\varepsilon\theta}(u+\beta_0^T t,\theta_0)$$

by $\Delta f_\varepsilon(t,u;\hat\beta_n,\hat\theta_n)$. First we can write R_{bw3} as the sum of the following two terms,

$$R_{bw31}=\iint K_b(v-u)\Delta f_\varepsilon(t,u;\hat\beta_n,\hat\theta_n)[E\hat f_{\bar U n}(t)-f_{\bar U}(t)]dudt,$$

and

$$\begin{aligned}
R_{bw32} \;=\; & \iint K_b(v-u)[(\hat\beta_n-\beta_0)^T\dot f_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)+(\hat\theta_n-\theta_0)^T\dot f_{\varepsilon\theta}(u+\beta_0^T t,\theta_0)]\cdot\\
& [E\hat f_{\bar U n}(t)-f_{\bar U}(t)]dudt.
\end{aligned}$$

By (f2), R_{bw31} is bounded above by

$$\sup_{u,t}|\Delta f(t,u;\hat{\beta}_n,\hat{\theta}_n)|\cdot\left[\frac{1}{2}w^2\mu_2(L)\iint K_b(v-u)|\mathrm{tr}(f''_{\bar{U}}(t))|dudt+o(w^2)\right]=O_p(w^2/n),$$

and R_{bw32} is bounded above by

$$\|\hat{\beta}_n-\beta_0\|\cdot\iint K_b(v-u)\|\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)\|\cdot|E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t)|dudt$$
$$+\|\hat{\theta}_n-\theta_0\|\cdot\iint K_b(v-u)\|\dot{f}_{\varepsilon\theta}(u+\beta_0^Tt,\theta_0)\|\cdot|E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t)|dudt.$$

Note that $\iint K_b(v-u)\|\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)\||E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t)|dudt\le O(w^2)\iint K_b(v-u)\|\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)\|dudt+o(w^2)=O(w^2)$, and by changing variables, $u=v+bx$, from (f3),

$$\iint K_b(v-u)\|\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)\|dudt=\iint K(x)\|\dot{f}_{\varepsilon\beta}(v+bx+\beta_0^Tt,\theta_0)\|dxdt$$
$$\le\iint K(x)\|\dot{f}_{\varepsilon\beta}(v+\beta_0^Tt,\theta_0)\|dxdt+b\iint|x|K(x)B(v+\beta_0^Tt,\theta_0)dxdt$$
$$=\int\|\dot{f}_{\varepsilon\beta}(v+\beta_0^Tt,\theta_0)\|dt+b\int|x|K(x)dx\cdot\int B(v+\beta_0^Tt,\theta_0)dt.$$

The $\sqrt{n}$-consistency of $\hat{\beta}_n$ and $\hat{\theta}_n$, and the integrability of $\dot{f}_{\varepsilon\beta}(v+\beta_0^Tt,\theta_0)$ and $B(v+\beta_0^Tt,\theta_0)$ with respect to t imply $\int|R_{bw3}|^2d\Pi(v)=2\left[O_p\left(n^{-2}w^4\right)+O_p\left(n^{-1}w^4\right)\right]$. Thus

$$nb^{1/2}\cdot\int|R_{bw3}|^2d\Pi(v)=nb^{1/2}O_p\left(\frac{w^4}{n^2}\right)+nb^{1/2}O_p\left(\frac{w^4}{n}\right)=o_p(1). \tag{2.9}$$

Next, we consider R_{bw1}. Adding and subtracting $(\hat{\beta}_n-\beta_0)^T\dot{f}_{\varepsilon\beta}(u+\beta_0^Tt,\theta_0)+(\hat{\theta}_n-\theta_0)^T\dot{f}_{\varepsilon\theta}(u+\beta_0^Tt,\theta_0)$ from $f_\varepsilon(u+\hat{\beta}_n^Tt,\hat{\theta}_n)-f_\varepsilon(u+\beta_0^Tt,\theta_0)$, we can rewrite R_{bw1} as the summation of three terms

$$\begin{aligned}R_{bw11} = &\iint K_b(v-u)[f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n) - f_\varepsilon(u+\beta_0^T t,\theta_0) - (\hat\beta_n-\beta_0)^T\dot f_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)\\ &- (\hat\theta_n-\theta_0)^T\dot f_{\varepsilon\theta}(u+\beta_0^T t,\theta_0)][\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)]dudt,\\ R_{bw12} =&(\hat\beta_n-\beta_0)^T\iint K_b(v-u)\dot f_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)[\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)]dudt,\\ R_{bw13} =&(\hat\theta_n-\theta_0)^T\iint K_b(v-u)\dot f_{\varepsilon\theta}(u+\beta_0^T t,\theta_0)[\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)]dudt.\end{aligned}$$

From condition (f2),

$$\begin{aligned}|R_{bw11}| \le&\sup_{t,u}|f_\varepsilon(u+\hat\beta_n^T t,\hat\theta_n) - f_\varepsilon(u+\beta_0^T t,\theta_0) - (\hat\beta_n-\beta_0)^T\dot f_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)\\ &- (\hat\theta_n-\theta_0)^T\dot f_{\varepsilon\theta}(u+\beta_0^T t,\theta_0)|\iint K_b(v-u)|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dudt\\ =&O_p\left(n^{-1}\right)\int K_b(v-u)du\int|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dt = O_p(n^{-1}).\end{aligned}$$

To consider R_{bw12} and R_{bw13}, we need an upper bound for $E\int|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dt$. By the Cauchy-Schwarz inequality, we have $E\int|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dt \le \int(E|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|^2)^{\frac{1}{2}}dt$. Note that $E[\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)]^2$ equals

$$\begin{aligned}&\frac{1}{n}\left\{\frac{f_{\bar U}(t)}{w^d}\int L^2(v)dv + \frac{1}{2w^{d-2}}\int L^2(v)v^T f''_{\bar U}(\tilde t_1)vdv - \left[f_{\bar U}(t) + \frac{w^2}{2}\int L(v)v^T f''_{\bar U}(\tilde t_2)vdv\right]^2\right\}\\ =&\frac{f_{\bar U}(t)}{nw^d}\int L^2(v)dv + \frac{1}{2nw^{d-2}}\int L^2(v)v^T f''_{\bar U}(\tilde t_1)vdv - \frac{1}{n}f^2_{\bar U}(t)\\ &-\frac{w^4}{4n}\left(\int L(v)v^T f''_{\bar U}(\tilde t_2)vdv\right)^2 - \frac{w^2 f_{\bar U}(t)}{n}\int L(v)v^T f''_{\bar U}(\tilde t_2)vdv.\end{aligned}$$

where $\tilde t_1$ and $\tilde t_2$ are between t and $t+vw$. Then by condition (g2) and (g3), we have $E\int|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dt = O((nw^d)^{-1/2})$. Hence $\int|\hat f_{\bar U n}(t) - E\hat f_{\bar U n}(t)|dt = O_p\left((nw^d)^{-1/2}\right)$.

For R_{bw12}, we have

$$\begin{aligned}
&\left\| \iint K_b(v-u)\dot{f}_{\varepsilon\beta}(u+\beta_0^T t,\theta_0)[\hat{f}_{\bar{U}n}(t)-E\hat{f}_{\bar{U}n}(t)]dudt \right\| \\
\leq &\iint K(x)\|\dot{f}_{\varepsilon\beta}(v+bx+\beta_0^T t,\theta_0)\|\cdot|\hat{f}_{\bar{U}n}(t)-E\hat{f}_{\bar{U}n}(t)|dxdt \\
\leq &\iint K(x)[\|\dot{f}_{\varepsilon\beta}(v+\beta_0^T t,\theta_0)\|+b|x|B(v+\beta_0^T t,\theta_0)]\cdot|\hat{f}_{\bar{U}n}(t)-E\hat{f}_{\bar{U}n}(t)|dxdt.
\end{aligned}$$

which has the order of $O_p(1/\sqrt{nw^d})$ by condition (f3). This, together with the $\sqrt{n}$-consistency of $\hat{\beta}_n$, implies that $R_{bw12}(v) = O_p((nw^{d/2})^{-1})$ uniformly for v. Similarly, we also have $R_{bw13}(v) = O_p((nw^{d/2})^{-1})$ uniformly for v as well. Therefore,

$$nb^{1/2}\int (R_{bw1}(v))^2 d\Pi(v) = nb^{1/2}\cdot O_p\left(\frac{1}{n^2w^d}\right) = O_p\left(\frac{b^{1/2}}{nw^d}\right) = o_p(1) \tag{2.10}$$

from assumption (b2). Next we consider R_{bw2}. Note that

$$\begin{aligned}
R_{bw2}(v) &= \frac{1}{nw^d}\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)\left[\sum_{i=1}^n L\left(\frac{t-\tilde{Z}_i}{w}\right) - \sum_{i=1}^n EL\left(\frac{t-\tilde{Z}_i}{w}\right)\right]dudt \\
&= \frac{1}{n}\sum_{i=1}^n \iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)[L_w(t-\tilde{Z}_i)-EL_w(t-\tilde{Z}_i)]dudt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
E(R_{bw2}(v))^2 &= \frac{1}{n}E\left[\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)[L_w(t-\tilde{Z})-EL_w(t-\tilde{Z})]dtdu\right]^2 \\
&= \frac{1}{n}\int\left[\iint K(x)f_\varepsilon(v+bx+\beta_0^T t,\theta_0)\left[\frac{1}{w^d}L\left(\frac{t-z}{w}\right)-f_{\bar{U}}(t)+O(w^2)\right]dtdx\right]^2 f_{\tilde{Z}}(z)dz,
\end{aligned}$$

which is of the order $O(n^{-1})$, implying that $nb^{1/2}\int(R_{bw2}(v))^2 d\Pi(v) = nb^{1/2}O_p(n^{-1}) = o_p(1)$. Therefore, by the compact support of Π, and the Cauchy-Schwarz inequality, we eventually show that

$$nb^{1/2}\int [R_{bw}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 d\Pi(v) = o_p(1). \tag{2.11}$$

Now let's consider the cross term in $T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)$. Using the decomposition of R_{bw}, we can see that

$$\int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)] R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v) = \sum_{j=1}^{4} Q_{nj}, \tag{2.12}$$

where $Q_{nj} = \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)] R_{bwj}(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v)$. We know from Koul and Song (2012) that

$$nb^{\frac{1}{2}} \left[\int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) - \hat{C}_n \right] \Longrightarrow N(0, \Gamma), \quad nb^{\frac{1}{2}} \hat{C}_n = O_p(b^{-1/2}), \tag{2.13}$$

where Γ is defined in (2.6). By the Cauchy-Schwarz inequality, we can see that $nb^{1/2}|Q_{n1}|$ is bounded above by

$$\begin{aligned} & \left\{ nb^{\frac{1}{2}} \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) \right\}^{\frac{1}{2}} \left\{ nb^{\frac{1}{2}} \int [R_{bw1}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) \right\}^{\frac{1}{2}} \\ = & \left\{ nb^{\frac{1}{2}} \left[\int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) - \hat{C}_n \right] + nb^{\frac{1}{2}} \hat{C}_n \right\}^{\frac{1}{2}} \\ & \cdot \left\{ nb^{\frac{1}{2}} \int [R_{bw1}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v) \right\}^{\frac{1}{2}}, \end{aligned}$$

this, together with (2.10), implies that we can conclude

$$nb^{1/2}|Q_{n1}| = O_p(b^{-1/4}) \cdot O_p\left(\frac{b^{1/4}}{\sqrt{nw^d}} \right) = o_p(1). \tag{2.14}$$

Similarly, from (2.9), we can show that

$$nb^{1/2}|Q_{n3}| = O_p(b^{-1/4}) \cdot O_p\left(b^{1/4} w^2 \right) = o_p(1). \tag{2.15}$$

Now we shall show that $nb^{1/2} Q_{nj} = o_p(1)$ holds for $j = 2, 4$. Recall the definitions of

$\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n)$, $f_{\xi b}(v; \hat{\beta}_n, \hat{\theta}_n)$, we see that $nb^{1/2}Q_{n2}$ can be written as

$$
\begin{aligned}
nb^{1/2}Q_{n2} =& \int \Big[\frac{1}{nb}\sum_{i=1}^n K\Big(\frac{v-\hat{\xi}_i}{b}\Big) - \int K_b(v-u) f_\xi(u; \hat{\beta}_n, \hat{\theta}_n) du\Big] \\
&\cdot \Big[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)(\hat{f}_{\bar{U}n}(t) - E\hat{f}_{\bar{U}n}(t)) du dt\Big] d\Pi(v) \\
=& \int \Big[\frac{1}{nb}\sum_{i=1}^n K\Big(\frac{v - Y_i + \hat{\alpha}_n + \hat{\beta}_n^T \bar{Z}_i}{b}\Big) \mp \frac{1}{nb}\sum_{i=1}^n K\Big(\frac{v - Y_i + \alpha_0 + \beta_0^T \bar{Z}_i}{b}\Big) \\
&\mp \int K_b(v-u) f_\xi(u; \beta_0, \theta_0) du - \int K_b(v-u) f_\xi(u; \hat{\beta}_n, \hat{\theta}_n) du\Big] \\
&\cdot \Big[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)[\hat{f}_{\bar{U}n}(t) - E\hat{f}_{\bar{U}n}(t)] dt du\Big] d\Pi(v) \\
=& \int \Big[\frac{1}{nb}\sum_{i=1}^n K\Big(\frac{v - Y_i + \alpha_0 + \beta_0^T \bar{Z}_i}{b}\Big) - \int K_b(v-u) f_\xi(u; \beta_0, \theta_0) du\Big] \\
&\Big[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)(\hat{f}_{\bar{U}n}(t) - E\hat{f}_{\bar{U}n}(t)) du dt\Big] d\Pi(v) + R_n,
\end{aligned}
$$

where $\mp$ stands for first minus then plus the term after the sign, and the remainder term R_n converges to 0 faster than the first term. So, it is sufficient to consider the first term only. By the definition of $\hat{f}_{\bar{U}n}(t)$, we can rewrite the first term as S_n,

$$
\begin{aligned}
S_n =& \frac{1}{n^2}\sum_{i=1}^n \sum_{j=1}^n \int \Big[\frac{1}{b}K\Big(\frac{v-\xi_i}{b}\Big) - E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big] \\
&\cdot \Big[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)\Big[\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}_j}{w}\Big) - E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)\Big] du dt\Big] d\Pi(v).
\end{aligned}
$$

Recall the notation $\xi = Y - \alpha_0 - \beta_0^T \bar{Z} = \varepsilon - \beta_0^T \Big(U_1 + U_2\Big)/2$, $\tilde{Z} = (U_1 - U_2)/2$. We have

$$
\begin{aligned}
ES_n =& \frac{1}{n}E\int \frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big) dt du\Big] d\Pi(v) \\
&- \frac{1}{n}\int E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0) E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big) dt du d\Pi(v)
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{n}\int\Bigg[\iiint\frac{1}{b}K\Big(\frac{v-\varepsilon+\beta_0^T(u_1+u_2)/2}{b}\Big)\Big[\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)\\
&\frac{1}{w^d}L\Big(\frac{t-(u_1-u_2)/2}{w}\Big)dtdu\Big]f(\varepsilon)f_U(u_1)f_U(u_2)\ \mathrm{d}\varepsilon\ \mathrm{d}u_1\ \mathrm{d}u_2\Bigg]d\Pi(v)\\
&-\frac{1}{n}\int E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)dtdud\Pi(v)
\end{aligned}
$$

which is of order $O(n^{-1})$. We also have

$$
\begin{aligned}
ES_n^2=&E\Bigg[\frac{1}{n^2}\sum_{i,j}\int\Big[\frac{1}{b}K\Big(\frac{v-\xi_i}{b}\Big)-E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big]\\
&\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)\Bigg[\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}_j}{w}\Big)-E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)\Bigg]dudtd\Pi(v)\Bigg]^2\\
=&\frac{1}{n^4}\sum_{i,j}E\Bigg[\int\Big[\frac{1}{b}K\Big(\frac{v-\xi_i}{b}\Big)-E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big]\\
&\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)\Big[\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}_j}{w}\Big)-E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)\Big]dudtd\Pi(v)\Bigg]^2\\
&+\frac{n(n-1)}{n^4}E\Bigg[\int\Big[\frac{1}{b}K\Big(\frac{v-\xi_1}{b}\Big)-E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big]\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)\\
&\Big[\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}_2}{w}\Big)-E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)\Big]dudtd\Pi(v)\int\Big[\frac{1}{b}K\Big(\frac{v-\xi_2}{b}\Big)-E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Big]\\
&\iint K_b(v-u)f_\varepsilon(u+\beta_0^Tt,\theta_0)\Big[\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}_1}{w}\Big)-E\frac{1}{w^d}L\Big(\frac{t-\tilde{Z}}{w}\Big)\Big]dudtd\Pi(v)\Bigg],
\end{aligned}
$$

which is the order of $O(n^{-2})$. The expectation and variance arguments imply that $S_n = O_p(1/n)$. Hence

$$
nb^{1/2}Q_{n2}=o_p(1). \tag{2.16}
$$

Finally, we are going to prove $nb^{\frac{1}{2}}Q_{n4}=o_p(1)$. First note that $nb^{\frac{1}{2}}Q_{n4}$ can be written as the sum of $nb^{1/2}S_{nj}$, $j = 1, 2, 3$, where

$$\begin{aligned}
nb^{\frac{1}{2}}S_{n1} =& \int \Bigg[\frac{1}{nb}\sum_{i=1}^{n} K\Big(\frac{v - Y_i + \alpha_0 + \beta_0^T \bar{Z}_i}{b}\Big) - \int K_b(v-u) f_\xi(u;\beta_0,\theta_0)du\Bigg] \\
& \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)(E\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t))dudt\Bigg] d\Pi(v), \\
nb^{\frac{1}{2}}S_{n2} =& \int \Bigg[\frac{1}{nb}\sum_{i=1}^{n} K\Big(\frac{v - Y_i + \hat{\alpha}_n + \hat{\beta}_n^T \bar{Z}_i}{b}\Big) - \frac{1}{nb}\sum_{i=1}^{n} K\Big(\frac{v - Y_i + \alpha_0 + \beta_0^T \bar{Z}_i}{b}\Big)\Bigg] \\
& \cdot \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)[E\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t)]dtdu\Bigg] d\Pi(v), \\
nb^{\frac{1}{2}}S_{n3} =& \int \Bigg[\int K_b(v-u) f_\xi(u;\beta_0,\theta_0)du - \int K_b(v-u) f_\xi(u;\hat{\beta}_n,\hat{\theta}_n)du\Bigg] \\
& \cdot \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)[E\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t)]dtdu\Bigg] d\Pi(v).
\end{aligned}$$

We can easily see that $ES_{n1} = 0$. From the boundedness of $f''_{\bar{U}}(t)$, we further have ES_{n1}^2 is bounded above by

$$\begin{aligned}
& \frac{1}{n}E\Bigg\{\int \Bigg|\frac{1}{b}K\Big(\frac{v-\xi_1}{b}\Big) - E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Bigg| \\
& \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)\frac{w^2}{2}\int L(z)|z^T f''_{\bar{U}}(\tilde{t})z|dzdudt\Bigg] d\Pi(v)\Bigg\}^2 \\
\leq & \frac{B^2 w^4}{4n} E\Bigg\{\int \Bigg|\frac{1}{b}K\Big(\frac{v-\xi_1}{b}\Big) - E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Bigg| \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)dudt\Bigg] d\Pi(v)\Bigg\}^2
\end{aligned}$$

for some finite positive constant B. Note that

$$E\Bigg\{\int \Bigg|\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big) - E\frac{1}{b}K\Big(\frac{v-\xi}{b}\Big)\Bigg| \Bigg[\iint K_b(v-u) f_\varepsilon(u+\beta_0^T t, \theta_0)dudt\Bigg] d\Pi(v)\Bigg\}^2$$

is the order of $O(1)$, so, we have $S_{n1} = O_p(w^2/\sqrt{n})$. Thus

$$nb^{\frac{1}{2}}S_{n1} = O_p\left(\sqrt{n}b^{\frac{1}{2}}w^2\right) = O_p(\sqrt{nbw^4}) = o_p(1).$$

Using the Cauchy-Schwarz inequality and from the proof of Theorem 3.1 in Koul and Song

(2012), we have

$$
\begin{aligned}
nb^{1/2}|S_{n2}| =& nb^{\frac{1}{2}}\left|\int\left[\frac{1}{n}\sum_{i=1}^{n}(K_b(v-\hat{\xi}_i)-K_b(v-\xi_i))\right]\right.\\
&\left.\cdot\left[\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)(E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t))dudt\right]d\Pi(v)\right|\\
\leq&\left\{nb^{1/2}\int\left[\frac{1}{n}\sum_{i=1}^{n}(K_b(v-\hat{\xi}_i)-K_b(v-\xi_i))\right]^2 d\Pi(v)\right\}^{1/2}\\
&\cdot\left\{nb^{1/2}\int\left[\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)(E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t))dudt\right]^2 d\Pi(v)\right\}^{1/2}\\
\leq& o_p(1)O(\sqrt{nb^{1/2}w^4})=o_p(1),
\end{aligned}
$$

and

$$
\begin{aligned}
nb^{1/2}S_{n3} =& nb^{\frac{1}{2}}\left|\int[f_{\xi b}(v;\beta_0,\theta_0)-f_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]\right.\\
&\left.\cdot\left[\iint K_b(v-u)f_\varepsilon(u+\beta_0^T t,\theta_0)(E\hat{f}_{\bar{U}n}(t)-f_{\bar{U}}(t))dudt\right]d\Pi(v)\right|\\
\leq& nb^{\frac{1}{2}}\int[(\hat{\theta}_n-\theta_0)^T\dot{f}_{\xi b\theta}(v;\beta_0,\theta_0)+(\hat{\beta}_n-\beta_0)^T\dot{f}_{\xi b\beta}(v;\beta_0,\theta_0)+O_p(1/n)]\\
&\cdot\left[\iint K_b(v-u)f_\varepsilon(u+\beta_0 t,\theta_0)\frac{w^2}{2}\int L(z)|z^T f''_{\bar{U}}(\tilde{t})z|dzdudt\right]d\Pi(v)\\
\leq& O_p(nb^{\frac{1}{2}}w^2/\sqrt{n})=o_p(1).
\end{aligned}
$$

Therefore, we have

$$
nb^{1/2}Q_{n4}=o_p(1). \tag{2.17}
$$

Note that $nb^{1/2}(T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n)-\hat{C}_n)$ can be rewritten as

$$
\begin{aligned}
&nb^{\frac{1}{2}}\left[\int[\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n)-f_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 d\Pi(v)-\hat{C}_n\right]+nb^{1/2}\int[R_{bw}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 d\Pi(v)\\
&-2nb^{1/2}\int[\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n)-f_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]R_{bw}(v;\hat{\beta}_n,\hat{\theta}_n)d\Pi(v).
\end{aligned}
$$

Combining (2.11)–(2.17), we can show that $\mathcal{T}_n \Longrightarrow N(0,\Gamma)$. This, together with $\hat{\Gamma}_n\to\Gamma$

in probability, which can be easily shown based on the consistency of $\hat{\alpha}_n, \hat{\beta}_n$ and the kernel density estimator $\hat{f}_{\xi n}$, completes the proof of Theorem 2.2.1.

Proof of Theorem 2.3.1: Define

$$\check{f}_{\xi b}(v;\beta) = \int K_b(v-u)\tilde{f}_{\xi a}(u;\beta)du, \quad \tilde{f}_{\xi a}(u;\beta) = \int f_{\varepsilon a}(u+\beta^T t)\hat{f}_{\bar{U}n}(t)dt,$$

By adding and subtracting $\check{f}_{\xi b}(v;\hat{\beta}_n)$ from $\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)$, we can rewrite $T_n = T_{1n} - 2T_{2n} + T_{3n}$, where

$$\begin{aligned}
T_{1n} &= \int [\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n) - \check{f}_{\xi b}(v;\hat{\beta}_n)]^2 d\Pi(v),\\
T_{2n} &= \int [\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n) - \check{f}_{\xi b}(v;\hat{\beta}_n)][\check{f}_{\xi b}(v;\hat{\beta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]d\Pi(v),\\
T_{3n} &= \int [\check{f}_{\xi b}(v;\hat{\beta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 d\Pi(v).
\end{aligned}$$

Therefore,

$$\mathcal{T}_n = nb^{\frac{1}{2}}\hat{\Gamma}_n^{-\frac{1}{2}}(T_{1n} - \hat{C}_n) - 2nb^{\frac{1}{2}}\hat{\Gamma}_n^{-\frac{1}{2}}T_{2n} + nb^{\frac{1}{2}}\hat{\Gamma}_n^{-\frac{1}{2}}T_{3n}.$$

One can show that $nb^{\frac{1}{2}}\hat{\Gamma}_n^{-\frac{1}{2}}(T_{1n}-\hat{C}_n) \Rightarrow N(0,1)$. The proof is similar to that of Theorem 2.2.1. Note that

$$\hat{\Gamma}_n \to 2\int f_{\xi a}^2(v;\beta_a)\pi^2(v)dv \int K_*^2(u)du := \tilde{\Gamma} > 0,$$

where $K_*(u) = \int K(t)K(u+t)dt$, and

$$\begin{aligned}
T_{3n} &= \int \left[\iint K_b(v-u)[f_{\varepsilon a}(u+\hat{\beta}_n^T t) - f_\varepsilon(u+\hat{\beta}_n^T t,\hat{\theta}_n)]\hat{f}_{\bar{U}n}(t)dtdu\right]^2 d\Pi(v)\\
&= \int \left[\iint K(x)[f_{\varepsilon a}(v+bx+\hat{\beta}_n^T t) - f_\varepsilon(v+bx+\hat{\beta}_n^T t,\hat{\theta}_n)]\hat{f}_{\bar{U}n}(t)dtdx\right]^2 d\Pi(v)\\
&\to \int \left[\int f_{\varepsilon a}(v+\beta_a^T t)f_{\bar{U}}(t)dt - \int f_\varepsilon(v+\beta_a^T t,\theta_a)f_{\bar{U}}(t)dt\right]^2 d\Pi(v)\\
&= \int [f_{\xi a}(v;\beta_a) - f_\xi(v;\beta_a,\theta_a)]^2 d\Pi(v) > 0
\end{aligned}$$

we have $nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{3n} = nb^{1/2}\tilde{\Gamma}^{-1/2}\int[f_{\xi a}(v;\beta_a) - f_\xi(v;\beta_a,\theta_a)]^2 d\Pi(v) + o_p(nb^{1/2})$ as $n \to \infty$.

By the Cauchy-Schwarz inequality, and using the fact $\hat{C}_n = O_p(1/(nb))$ from Koul and Song (2012), $nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}|T_{2n}|$ is bounded above by

$$
\begin{aligned}
&[nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}T_{1n}]^{\frac{1}{2}}[nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}T_{3n}]^{\frac{1}{2}} = [nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}(T_{1n} - \hat{C}_n + \hat{C}_n)]^{\frac{1}{2}}[nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}T_{3n}]^{\frac{1}{2}} \\
\le &[nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}|T_{1n} - \hat{C}_n| + nb^{\frac{1}{2}}\hat{\Gamma}_n^{-1/2}\hat{C}_n]^{\frac{1}{2}}O_p(\sqrt{nb^{1/2}}) \\
= &[O_p(1) + O_p(b^{-1/2})]^{\frac{1}{2}}O_p(\sqrt{nb^{1/2}}) = o_p(nb^{1/2})
\end{aligned}
$$

from $nb \to \infty$ guaranteed by the assumption (b1). Therefore, $\mathcal{T}_n = nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{1n} - \hat{C}_n) + nb^{1/2}\tilde{\Gamma}^{-1/2}\int[f_{\xi a}(v;\beta_a) - f_\xi(v;\beta_a,\theta_a)]^2 d\Pi(v) + o_p(nb^{1/2})$. Clearly, the right hand side of the above expression tends to ∞ as $n \to \infty$, implying that the proposed test is consistent. □

Proof of Theorem 2.3.2: Denote

$$
\begin{aligned}
&\tilde{f}_\xi^{\text{loc}}(v;\beta_0,\theta_0) = \int \left[(1-\delta_n)f_\varepsilon(v+\beta_0^T u,\theta_0) + \delta_n\varphi(v+\beta_0^T u)\right] f_{\bar{U}}(u)du \\
= &\int f_\varepsilon(v+\beta_0^T u,\theta_0)f_{\bar{U}}(u)du - \delta_n\int \left[f_\varepsilon(v+\beta_0^T u,\theta_0) - \varphi(v+\beta_0^T u)\right] f_{\bar{U}}(u)du. \\
&\tilde{f}_{\xi b}^{\text{loc}}(v;\hat{\beta}_n,\hat{\theta}_n) = \int K_b(v-u)\tilde{f}_\xi^{\text{loc}}(u;\hat{\beta}_n,\hat{\theta}_n)du.
\end{aligned}
$$

Adding and subtracting $\tilde{f}_{\xi b}^{\text{loc}}(v;\hat{\beta}_n,\hat{\theta}_n)$ from $\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)$, we can rewrite the test statistic as

$$
T_n(\hat{\alpha}_n,\hat{\beta}_n,\hat{\theta}_n) = \int[\hat{f}_{\xi n}(v;\hat{\alpha}_n,\hat{\beta}_n) - \tilde{f}_{\xi b}^{\text{loc}}(v;\hat{\beta}_n,\hat{\theta}_n) + \tilde{f}_{\xi b}^{\text{loc}}(v;\hat{\beta}_n,\hat{\theta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n)]^2 d\Pi(v).
$$

Note that

$$
\begin{aligned}
&\tilde{f}_{\xi b}^{\text{loc}}(v;\hat{\beta}_n,\hat{\theta}_n) - \tilde{f}_{\xi b}(v;\hat{\beta}_n,\hat{\theta}_n) = \int K_b(v-u)[\tilde{f}_\xi^{\text{loc}}(u;\hat{\beta}_n,\hat{\theta}_n) - \tilde{f}_\xi(u;\hat{\beta}_n,\hat{\theta}_n)]du \\
= &-\int K_b(v-u)\cdot\delta_n\int[f_\varepsilon(u+\hat{\beta}_n^T t,\hat{\theta}_n) - \varphi(u+\hat{\beta}_n^T t)]\hat{f}_{\bar{U}n}(t)dtdu \\
= &-\delta_n\iint K_b(v-u)[f_\varepsilon(u+\hat{\beta}_n^T t,\hat{\theta}_n) - \varphi(u+\hat{\beta}_n^T t)]\hat{f}_{\bar{U}n}(t)dtdu := -\delta_n D_n(v;\hat{\beta}_n,\hat{\theta}_n).
\end{aligned}
$$

We can rewrite T_n as the sum of the following three terms

$$
\begin{aligned}
T_{n1} =& \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - \tilde{f}_{\xi b}^{\text{loc}}(v; \hat{\beta}_n, \hat{\theta}_n)]^2 d\Pi(v), \\
T_{n2} =& - 2\delta_n \int [\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - \tilde{f}_{\xi b}^{\text{loc}}(v; \hat{\beta}_n, \hat{\theta}_n)] D_n(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v), \\
T_{n3} =& \delta_n^2 \int D_n^2(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v).
\end{aligned}
$$

For the sake of simplicity, denote $f_n(u,t) := (1-\delta_n) f_\varepsilon(u + \hat{\beta}_n^T t, \hat{\theta}_n) + \delta_n \varphi(u + \hat{\beta}_n^T t)$. Adding and subtracting $f_{\bar{U}}(t)$ from $\hat{f}_{\bar{U}n}(t)$, T_{n1} can be further written as the sum of the following three terms,

$$
\begin{aligned}
T_{n11} =& \int \left[\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - \iint K_b(v-u) f_n(u,t) f_{\bar{U}}(t) dt du \right]^2 d\Pi(v), \\
T_{n12} =& \int \left[\iint K_b(v-u) f_n(u,t) [\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t)] dt du \right]^2 d\Pi(v), \\
T_{n13} =& \int \left[\hat{f}_{\xi n}(v; \hat{\alpha}_n, \hat{\beta}_n) - \iint K_b(v-u) f_n(u,t) f_{\bar{U}}(t) dt du \right] \cdot \\
& \quad \left[\iint K_b(v-u) f_n(u,t) [\hat{f}_{\bar{U}n}(t) - f_{\bar{U}}(t)] dt du \right] d\Pi(v).
\end{aligned}
$$

Similar to the discussion as in $R_{bw}(v; \hat{\beta}_n, \hat{\theta}_n)$, one can show that $nb^{\frac{1}{2}} T_{n12} = o_p(1)$. Follow the proof of Theorem 2.2.1 in Koul and Song (2012), we can show that $nb^{\frac{1}{2}}[T_{n11} - \hat{C}_n] \Rightarrow N(0, \Gamma)$ and using the Cauchy-Schwarz inequality, we also have $nb^{\frac{1}{2}} T_{n13} = o_p(1)$. Therefore, we have $nb^{\frac{1}{2}}[T_{n1} - \hat{C}_n] = nb^{\frac{1}{2}}[T_{n11} - \hat{C}_n] + o_p(1)$.

By the boundedness of $f''(t)$ and $\varphi''(t)$, then we have

$$
\begin{aligned}
& nb^{\frac{1}{2}} \hat{\Gamma}_n^{-\frac{1}{2}} T_{n3} = nb^{\frac{1}{2}} \hat{\Gamma}_n^{-\frac{1}{2}} \delta_n^2 \int D_n^2(v; \hat{\beta}_n, \hat{\theta}_n) d\Pi(v) = \hat{\Gamma}_n^{-\frac{1}{2}} \int D_n^2(v, \hat{\beta}_n, \hat{\theta}_n) d\Pi(v) \\
=& \hat{\Gamma}_n^{-\frac{1}{2}} \int \left[\iint K_b(v-u) [f_\varepsilon(u + \hat{\beta}_n^T t, \hat{\theta}_n) - \varphi(u + \hat{\beta}_n^T t)] \hat{f}_{\bar{U}n}(t) dt du \right]^2 d\Pi(v) \\
=& \Gamma^{-\frac{1}{2}} \int \left[\int [f_\varepsilon(v + \beta_0^T t, \theta_0) - \varphi(v + \beta_0^T t)] f_{\bar{U}}(t) dt \right]^2 d\Pi(v) + o_p(1).
\end{aligned}
$$

Similarly, we can obtain

$$
\begin{aligned}
nb^{\frac{1}{2}}T_{n2} =& \sqrt{nb^{\frac{1}{2}}} \int \left[\frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{v - Y_i + \alpha_0 + \beta_0^T \bar{Z}_i}{b} \right) - \int K_b(v-u) \tilde{f}_{\xi}^{\text{loc}}(u; \beta_0, \theta_0) du \right] \\
& \cdot \left[\iint K_b(v-u)[f_{\varepsilon}(u + \beta_0^T t, \theta_0) - \varphi(u + \beta_0^T t)] f_{\bar{U}}(t) du dt \right] d\Pi(v) + o_p(1) \\
=& O_p(b^{\frac{1}{4}}) = o_p(1).
\end{aligned}
$$

Summarizing the above results, we can conclude the proof of Theorem 2.3.2. □

Chapter 3

Goodness-of-Fit Test on the Density Function of the Latent Variable

This Chapter is organized as follows. Two minimum distance estimators of the parameter under the null hypothesis, and the statistics based on which the test being built will be defined in Section 3.1; A list of technical assumptions needed for the main results will be given in Section 3.2, as well as some notations used in the later sections; The large sample properties of the minimum distance estimators of the distribution parameters will be stated in Section 3.3, including the weak consistency and asymptotic normality; Asymptotic distributions of the test statistic under null hypothesis will be discussed in Section 3.4, together with its power performance under fixed and local alternatives in Section 3.5; Simulation and comparison studies will be conducted in Section 3.6.

3.1 Minimum Distance Estimators and Test

Recall that in the measurement error model $Z = X + U$, the hypothesis to be tested is

$$H_0 : f_X(x) = f_X(x, \theta) \quad \text{for some } \theta \in \Theta, \Theta \subset \mathbb{R}^q \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true.}$$

We assume that X and U are independent one-dimensional random variables. We also assume that the parameter space Θ is compact subset in $\mathbb{R}^q$, $q \geq 1$. Moreover, the measurement error is assumed to be symmetric around 0. Suppose at each value of X, two measurements of Z can be obtained. That is, we can observe

$$Z_{i1} = X_i + U_{i1}, \quad Z_{i2} = X_i + U_{i2}, \tag{3.1}$$

$i = 1, 2, \ldots, n$, U_{i1} and U_{i2} are independent and identically distributed. By (3.1) and simple algebra, we have

$$\frac{Z_{i1} - Z_{i2}}{2} = \frac{U_{i1} - U_{i2}}{2} := \tilde{U}_i, \quad \frac{Z_{i1} + Z_{i2}}{2} = X_i + \frac{U_{i1} + U_{i2}}{2} := X_i + \bar{U}_i. \tag{3.2}$$

Denote $\tilde{Z}_i = (Z_{i1} - Z_{i2})/2$, and $\bar{Z}_i = (Z_{i1} + Z_{i2})/2$. Then from the second equality in (3.2), $\bar{Z}$ is the convolution of X and $\bar{U}$. Therefore, $f_{\bar{Z}}(z) = \int f_X(z-u) f_{\bar{U}}(u)du$, and under H_0, $f_{\bar{Z}}(z, \theta) = \int f_X(z-u, \theta) f_{\bar{U}}(u)du$. Due to the fact that $f_{\bar{U}}(u)$ is unknown, this expression cannot be used directly. However, from the first equality in (3.2), it can be estimated by the classic kernel estimator defined by $\hat{f}_{\bar{U}}(u) = n^{-1} \sum_{i=1}^n K_h(u - \tilde{U}_i)$, where $K_h(\cdot) = h^{-1}K(\cdot/h)$, K is a kernel function and h is a bandwidth depending on the sample size n. We use h, instead of h_n, for simplicity. Therefore, $f_{\bar{Z}}(z, \theta)$ can be estimated by

$$\hat{f}_{\bar{Z}}(z, \theta) = \int f_X(z-u, \theta) \hat{f}_{\bar{U}}(u)du. \tag{3.3}$$

Since $\bar{Z}_i$'s are available, so $f_{\bar{Z}}(\cdot)$ can also be estimated by the following kernel estimator

$$\hat{f}_{\bar{Z}}(z) = \frac{1}{n} \sum_{i=1}^n L_b(z - \bar{Z}_i), \tag{3.4}$$

where $L_b(\cdot) = b^{-1}L(\cdot/b)$, L is a kernel function and b is a sequence of bandwidth depending on n. In the sequel, we use b other than b_n for simplicity.

Intuitively, if H_0 holds, then the semi-parametric estimator $\hat{f}_{\bar{Z}}(z, \theta)$ defined in (3.3) should be close to the kernel density estimator defined in (3.4). This motivates us to define the following weighted L_2-distance between these two estimators

$$T_n^*(\theta) = \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z,\theta)]^2 d\Pi(z), \quad \theta \in \Theta. \tag{3.5}$$

Since θ is unknown, we can estimate θ by $\theta_n^* = \underset{\theta}{\text{minimizer}}\ T_n^*(\theta)$, and a potential test can be built upon the statistic $T_n^*(\theta_n^*)$.

The second potential test is based on the centralization idea from Bickle and Rosenblatt (1973). Note that under the null hypothesis,

$$E\hat{f}_{\bar{Z}}(z) = EL_b(z - \bar{Z}) = \int L_b(z-x)\left[\int f_X(x-u,\theta) f_{\bar{U}}(u)du\right] dx,$$

thus $E\hat{f}_{\bar{Z}}(z)$ can be estimated by $\int L_b(z-x)\left[\int f_X(x-u,\theta)\hat{f}_{\bar{U}}(u)du\right] dx$ by replacing $f_{\bar{U}}(u)$ with $\hat{f}_{\bar{U}}(u)$. Define

$$\begin{aligned} T_n(\theta) &= \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta)dx\right]^2 d\Pi(z) \\ &= \int \left\{\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\left[\int f_X(x-u,\theta)\hat{f}_{\bar{U}}(u)du\right] dx\right\}^2 d\Pi(z), \quad \theta \in \Theta, \end{aligned} \tag{3.6}$$

which is a weighted L_2-distance between $\hat{f}_{\bar{Z}}(z)$ and its estimated expectation. θ thus can be estimated by $\hat{\theta}_n = \underset{\theta}{\text{minimizer}}\ T_n(\theta)$. Then we can develop a test procedure based on $T_n(\hat{\theta}_n)$.

The main difference between $T_n^*(\theta)$ and $T_n(\theta)$ is the semiparametric part in their definitions. $T_n^*(\theta)$ uses the density estimator of $\bar{Z}$ under the null hypothesis, while $T_n(\theta)$ uses the expectation of the nonparametric estimator of $\bar{Z}$ under the null hypothesis. Because of the centralization in $T_n(\theta)$, no under smoothing is needed for the kernel density estimator of $\bar{Z}$, thus avoid the potential non-tightness of the test statistic caused by the nonnegligible bias. A similar phenomenon can be found in the regression setup, see Koul and Ni (2004) for detail.

3.2 Assumptions

This section include a list of technical assumptions needed for the theoretical results which will be presented later. Also, some notations will be also introduced here for the sake of convenience in stating various lemmas and theorems in the following sections.

As for the density function of $\bar{Z}$, we assume that

(z1) For each θ, $f_{\bar{Z}}(z, \theta) = \int f_X(z-u, \theta) f_{\bar{U}}(u) du$ is integrable, twice differentiable in z w.r.t. Π;

(z2) $f_{\bar{Z}}(z, \theta)$ is identifiable. That is, $\int [f_{\bar{Z}}(z, \theta_0) - f_{\bar{Z}}(z, \theta)]^2 d\Pi(z) = 0$ implies $\theta = \theta_0$;

(z3) For some positive continuous function l on I, with $l(z)$ bounded, and for some $r > 0$, $|f_{\bar{Z}}(z, \theta_1) - f_{\bar{Z}}(z, \theta_2)| \leq \|\theta_1 - \theta_2\|^r l(z), \quad \forall \theta_1, \theta_2 \in \Theta, z \in I$;

(z4) $f''_{\bar{Z}}(z, \theta_0)$ is bounded. That is, $|f''_{\bar{Z}}(z, \theta_0)| \leq c$ for some positive number c.

For the density function f_X of X, we have

(x1) f_X is bounded, twice continuously differentiable w.r.t. θ.

(x2) For every x, $f_X(x, \theta)$ is differentiable in θ in a neighborhood of θ_0 with the vector of derivatives $\dot{f}_X(x, \theta)$ such that if $\theta_n \to \theta_0$ in probability, then

$$\sup_x \frac{|f_X(x, \theta_n) - f_X(x, \theta_0) - (\theta_n - \theta_0)^T \dot{f}_X(x, \theta_0)|}{\|\theta_n - \theta_0\|} = o_p(1).$$

(x3) The vector function $x \mapsto \dot{f}_X(x, \theta_0)$ is continuous in $x \in I$ and for every $\varepsilon > 0$, there is an $N_\varepsilon < \infty$ such that for every $0 < k < \infty$,

$$P\left(\sup_{x \in I, (nb_n)^{1/2} \|\theta_n - \theta_0\| \leq k} b_n^{-1/2} \|\dot{f}_X(x, \theta_n) - \dot{f}_X(x, \theta_0)\| \geq \varepsilon \right) < \varepsilon, \quad \forall n > N_\varepsilon.$$

For the weight function Π, we assume that

(π1) The weight function Π is supported on a close interval I, and its derivative π is continuous and bounded.

For the density function of $\bar{U}$, we assume that

(u1) $f_{\bar{U}}(u)$ is twice differentiable in u, and $f''_{\bar{U}}(u)$ is bounded, integrable and square integrable.

About the kernel function K and L, we shall assume

(kl) The kernel K and L are bounded, symmetric, continues density functions.

About the bandwidth b, we shall make the following assumption

(b1) $b \to 0$ as $n \to \infty$.

(b2) $nb^2 \to \infty$ as $n \to \infty$.

(b3) $nb^4 \to 0$ as $n \to \infty$.

About the bandwidth h, we shall make the following assumption

(h1) $h \to 0$ as $n \to \infty$.

(h2) $nh \to \infty$ as $n \to \infty$.

(h3) $nh^4 \to 0$ as $n \to \infty$.

Assumptions $b \to 0, h \to 0, nb^2 \to \infty, nh \to \infty$ are commonly used in univariate kernel smoothing estimation procedures. Under the null hypothesis, the assumptions $nb^4 \to 0$ and $nh^4 \to 0$ are both required in the proof of the asymptotic distribution of the minimum distance estimator of the distribution parameter or the asymptotic distribution of the test statistic based on $T_n^*(\theta)$ defined in (3.5), while only $nh^4 \to 0$ is needed for those results based on $T_n(\theta)$ defined in (3.6).

In addition to the assumptions listed above, the following notations are also needed.

$$
\begin{aligned}
\Sigma_0 =& \int \dot f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}^T(z,\theta_0)d\Pi(z),\\
\Sigma_1 =& \int f_{\bar Z}(z,\theta_0)[\dot f_{\bar Z}(z,\theta_0)][\dot f_{\bar Z}(z,\theta_0)]^T\pi^2(z)dz - \left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]\\
&\cdot\left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]^T,\\
\Sigma_2 =& \int\left[\int f_X(z-u,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]\left[\int f_X(z-u,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]^T f_{\bar U}(u)du\\
&-\left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]\left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]^T,\\
\Sigma_3 =& 2\left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]\left[\int f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}(z,\theta_0)d\Pi(z)\right]^T\\
&-4\iint\left(\int f_X(y-u,\theta_0)f_U(z+u)f_U(z-u)du\right)\dot f_{\bar Z}(z,\theta_0)\dot f_{\bar Z}^T(y,\theta_0)d\Pi(z)d\Pi(y).\\
\hat C_n(\theta) =& \frac{1}{n^2}\sum_{i=1}^n\int\left(L_b(z-\bar Z_i)-\int L_b(z-x)\int f_X(x-u,\theta)\hat f_{\bar U}(u)dudx\right)^2 d\Pi(z),\\
\Gamma =& 2\int[f_{\bar Z}(y,\theta_0)]^2\pi^2(y)dy\int\left(\int L(t)L(z+t)dt\right)^2dz.
\end{aligned}
\tag{3.7}
$$

3.3 Consistency and Asymptotic Normality of the MD Estimators

This section states the large sample properties of the minimum distance estimators θ_n^* and $\hat\theta_n$, including the consistency and asymptotic normality.

3.3.1 Consistency

We begin with the consistency of θ_n^*, which is the minimizer of $T_n^*(\theta)$ defined in (3.5).

Theorem 3.3.1. *Suppose H_0, (z1)–(z4), (x1), (b1)–(b3) and (π1) hold. Then $\theta_n^* \to \theta_0$ in probability.*

Proof. Define $T_n^{**}(\theta) = \int[\hat f_{\bar Z}(z) - f_{\bar Z}(z,\theta)]^2 d\Pi(z)$, and $\theta_n^{**} = \underset{\theta\in\Theta}{\text{minimizer}}\, T_n^{**}(\theta)$. According to

Lemma 3.1 in Koul and Ni (2004), we only need to show $T_n^{**}(\theta_0) \xrightarrow{p} 0 \quad \text{as } n \to \infty$. In fact, by the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$,

$$\begin{aligned} T_n^{**}(\theta_0) =& \int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z) + E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) \\ \leq& 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) + 2\int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z). \end{aligned}$$

By Fubini's Theorem and (z4),

$$\begin{aligned} & E\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) = \int E\left[\frac{1}{n}\sum_{i=1}^n L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z})\right]^2 d\Pi(z) \\ \leq& \int \frac{1}{nb^2} EL^2\left(\frac{z-\bar{Z}}{b}\right) d\Pi(z) = \int \frac{1}{nb^2}\int L^2\left(\frac{z-v}{b}\right) f_{\bar{Z}}(v,\theta_0) dv d\Pi(z) \\ =& \frac{1}{nb}\int\left[\int L^2(v)[f_{\bar{Z}}(z,\theta_0) + bvf'_{\bar{Z}}(z,\theta_0) + \frac{1}{2}b^2v^2 f''_{\bar{Z}}(\tilde{z};\theta_0)]dv\right] d\Pi(z) = O\left(\frac{1}{nb}\right), \end{aligned}$$

where $\tilde{z}$ is between z and $z+bv$. So

$$\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) = O_p\left(\frac{1}{nb}\right) = o_p(1). \tag{3.8}$$

Also, it is easy to show by the well known result for the bias term in kernel density estimation that

$$\int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) = \int [EL_b(z-\bar{Z}) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) = O(b^4), \tag{3.9}$$

which is of order $o(1)$. From this, together with the assumptions (b2) and (b3), we have

$$T_n^{**}(\theta_0) = O_p\left(\frac{1}{nb}\right) = o_p(1) \tag{3.10}$$

and thus $\theta_n^{**} \to \theta_0$ in probability by Lemma 3.1 from Koul and Ni (2004).

Now, let's show that $\theta_n^* \to \theta_0$. It is sufficient to show that

$$\sup_{\theta\in\Theta} |T_n^*(\theta) - T(\theta)| = o_p(1), \qquad \sup_{\theta\in\Theta} |T_n^{**}(\theta) - T(\theta)| = o_p(1), \tag{3.11}$$

where $T(\theta) = \int [f_{\bar{Z}}(z,\theta_0) - f_{\bar{Z}}(z,\theta)]^2 d\Pi(z)$.

In fact, (3.11) implies

$$\sup_{\theta\in\Theta} |T_n^*(\theta) - T_n^{**}(\theta)| = o_p(1). \tag{3.12}$$

If $\theta_n^{**} - \theta_n^* \nrightarrow 0$, then, using the fact that Θ is compact, there must be a sub-sequence $\{n_k\}$ such that $\theta_{n_k}^* \to \theta_1,\ \theta_{n_k}^{**} \to \theta_0,\ \text{and } \theta_0 \neq \theta_1$.

From (3.12), we have

$$T_n^*(\theta_{n_k}^{**}) - T_n^{**}(\theta_{n_k}^{**}) = o_p(1), \quad T_n^*(\theta_{n_k}^{*}) - T_n^{**}(\theta_{n_k}^{*}) = o_p(1),$$

this immediately implies

$$T_n^*(\theta_{n_k}^{**}) - T_n^*(\theta_{n_k}^{*}) = T_n^{**}(\theta_{n_k}^{**}) - T_n^{**}(\theta_{n_k}^{*}) + o_p(1). \tag{3.13}$$

By the definition of $\theta_{n_k}^*$ and $\theta_{n_k}^{**}$, for every n, the left-hand side of (3.13) is nonnegative, while the right-hand side is nonpositive. This implies $T_n^*(\theta_{n_k}^{**}) - T_n^*(\theta_{n_k}^{*}) = o_p(1)$, $T_n^{**}(\theta_{n_k}^{**}) - T_n^{**}(\theta_{n_k}^{*}) = o_p(1)$, and therefore $|T(\theta_{n_k}^*) - T(\theta_{n_k}^{**})|$ is bounded above by

$$|T(\theta_{n_k}^*) - T_n^*(\theta_{n_k}^*)| + |T_n^*(\theta_{n_k}^*) - T_n^*(\theta_{n_k}^{**})| + |T_n^*(\theta_{n_k}^{**}) - T_n^{**}(\theta_{n_k}^{**})| + |T_n^{**}(\theta_{n_k}^{**}) - T(\theta_{n_k}^{**})|$$

which is the order of $o_p(1)$. By the continuity of $T(\theta)$, we have $|T(\theta_1) - T(\theta_0)| = 0$, which contradicts the uniqueness of the minimizer of $T(\theta)$ as implied by the identifiability condition (z2).

First we show the second equality in (3.11). Adding and subtracting $f_{\bar{Z}}(z,\theta_0)$ from

$\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta)$, we can rewrite $T_n^{**}(\theta)$ as the sum of $T(\theta)$ and the following two terms

$$
\begin{aligned}
A_{n1} &= \int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z), \\
A_{n2} &= 2\int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)][f_{\bar{Z}}(z,\theta_0) - f_{\bar{Z}}(z,\theta)] d\Pi(z).
\end{aligned}
$$

Thus $T_n^{**}(\theta) - T(\theta) = A_{n1} + A_{n2}(\theta)$, and

$$
\sup_{\theta\in\Theta} |T_n^*(\theta) - T(\theta)| \leq A_{n1} + \sup_{\theta\in\Theta} |A_{n2}(\theta)| \leq A_{n1} + 2A_{n1}^{\frac{1}{2}} \sup_{\theta\in\Theta} T^{\frac{1}{2}}(\theta).
$$

$A_{n1} = o_p(1)$ indeed is implied by (3.10), and $\sup_{\theta\in\Theta} T(\theta) < \infty$ can be shown by noting that the parameter space Θ is compact, and from (z3), $T(\theta) \leq \|\theta - \theta_0\|^{2r} \int l^2(z) d\Pi(z)$. Therefore, the second equality (3.11) holds.

Next, let's show the first equality in (3.11). Adding and subtracting $f_{\bar{Z}}(z,\theta)$ from $\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z,\theta)$, we can rewrite $T_n^*(\theta)$ as the sum of $T_n^{**}(\theta)$ and the following two terms

$$
\begin{aligned}
B_{n1}(\theta) =& 2\int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta)][f_{\bar{Z}}(z,\theta) - \hat{f}_{\bar{Z}}(z,\theta)] d\Pi(z), \\
B_{n2}(\theta) =& \int [f_{\bar{Z}}(z,\theta) - \hat{f}_{\bar{Z}}(z,\theta)]^2 d\Pi(z).
\end{aligned}
$$

Therefore

$$
T_n^*(\theta) - T(\theta) = T_n^{**}(\theta) - T(\theta) + B_{n1}(\theta) + B_{n2}(\theta).
$$

From (x1), we have

$$
B_{n2}(\theta) = \int \left[\int f_X(z-u,\theta)(\hat{f}_{\bar{U}}(u) - f_{\bar{U}}(u)) du\right]^2 d\Pi(z) \leq c\left[\int |\hat{f}_{\bar{U}}(u) - f_{\bar{U}}(u)| du\right]^2.
$$

On the other hand, by using Scheffe's Lemma, and the fact that

$$
\int \hat{f}_{\bar{U}}(u) du = \int f_{\bar{U}}(u) du = 1, \quad \hat{f}_{\bar{U}}(u) \to f_{\bar{U}}(u),
$$

we have

$$\int |\hat{f}_{\bar{U}}(u) - f_{\bar{U}}(u)|du = o_p(1). \tag{3.14}$$

So, $\sup_{\theta\in\Theta} B_{n2}(\theta) = o_p(1)$. Therefore, by using the elementary inequality $\sqrt{a+c} \le \sqrt{a}+\sqrt{c}$ for $a \ge 0, c \ge 0$, we can show that

$$\sup_\theta |B_{n1}(\theta)| \le 2\sup_\theta |T_n^{**}(\theta)|^{\frac{1}{2}} \cdot \sup_\theta |B_{n2}|^{\frac{1}{2}} = 2\sup_\theta |T_n^{**}(\theta) - T(\theta) + T(\theta)|^{\frac{1}{2}} \cdot \sup_\theta |B_{n2}(\theta)|^{\frac{1}{2}}$$

is the order of $o_p(1)$. This completes the proof of (3.11) and hence Theorem 3.3.1. □

The next theorem states the consistency of $\hat{\theta}_n$, the minimizer of $T_n(\theta)$ defined in (3.6).

Theorem 3.3.2. *Assume H_0, (z1)–(z4), (b1)–(b2), (x1) and (π1) hold, then $\hat{\theta}_n \to \theta_0$ in probability.*

Proof. Recall that

$$T_n(\theta) = \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta)dx\right]^2 d\Pi(z), \quad \hat{\theta}_n = \underset{\theta}{\text{minimizer}}\ T_n(\theta).$$

Define

$$\mathscr{T}_n^*(\theta) = \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)f_{\bar{Z}}(x,\theta)dx\right]^2 d\Pi(z), \quad \hat{\theta}_n^* = \underset{\theta}{\text{minimizer}}\ \mathscr{T}_n^*(\theta),$$

and

$$\mathscr{T}(\theta) = \int \left\{\int L_b(z-x)[f_{\bar{Z}}(x,\theta) - f_{\bar{Z}}(x,\theta_0)]dx\right\}^2 d\Pi(z).$$

The minimizer of $\mathscr{T}(\theta)$ is unique, as can be easily derived from the identifiable condition (z2). By Lemma 3.1(c) in Koul and Ni (2004), and from (3.8), one can easily check that

$$E\mathscr{T}_n^*(\theta_0) = E\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) = O\left(\frac{1}{nb}\right). \tag{3.15}$$

Therefore,

$$\mathscr{T}_n^*(\theta_0) = O_p\left(\frac{1}{nb}\right) = o_p(1). \tag{3.16}$$

Using Lemma 3.1 in Koul and Ni (2004), we have $\hat{\theta}_n^* \to \theta_0$ in probability. Next, we will show that if

$$\sup_{\theta\in\Theta} |T_n(\theta) - \mathscr{T}(\theta)| = o_p(1), \qquad \sup_{\theta\in\Theta} |\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta)| = o_p(1), \tag{3.17}$$

then $\hat{\theta}_n^* - \hat{\theta}_n \to 0$, which can be proved by contradiction.

If $\hat{\theta}_n^* - \hat{\theta}_n \nrightarrow 0$, there must be a sub-sequence $\{n_k\}$, such that $\hat{\theta}_{n_k}^* \to \theta_0$, $\hat{\theta}_{n_k} \to \theta_1$, and $\theta_1 \neq \theta_0$.

From (3.17), we have

$$\sup_{\theta\in\Theta} |T_n(\theta) - \mathscr{T}_n^*(\theta)| = o_p(1). \tag{3.18}$$

Thus

$$T_n(\hat{\theta}_n) - \mathscr{T}_n^*(\hat{\theta}_n) = o_p(1), \quad T_n(\hat{\theta}_n^*) - \mathscr{T}_n^*(\hat{\theta}_n^*) = o_p(1).$$

Therefore,

$$T_n(\hat{\theta}_n) - T_n(\hat{\theta}_n^*) = \mathscr{T}_n^*(\hat{\theta}_n) - \mathscr{T}_n^*(\hat{\theta}_n^*) + o_p(1).$$

By the definition of $\hat{\theta}_n$ and $\hat{\theta}_n^*$, for every n, the left-hand side of the equation above is nonpositive, while the first term on the right-hand side is nonnegative. Hence $T_n(\hat{\theta}_n) - \mathscr{T}_n^*(\hat{\theta}_n^*) = o_p(1)$. Then

$$|\mathscr{T}(\hat{\theta}_{n_k}^*) - \mathscr{T}(\hat{\theta}_{n_k})| \leq |\mathscr{T}(\hat{\theta}_{n_k}^*) - \mathscr{T}_n^*(\hat{\theta}_{n_k}^*)| + |\mathscr{T}_n^*(\hat{\theta}_{n_k}^*) - T_n(\hat{\theta}_{n_k})| + |T_n(\hat{\theta}_{n_k}) - \mathscr{T}(\hat{\theta}_{n_k})| = o_p(1).$$

However, by the continuity of $\mathscr{T}$, we have $\mathscr{T}(\hat{\theta}_{n_k}^*) \to \mathscr{T}(\theta_0) \neq \mathscr{T}(\theta_1) \leftarrow \mathscr{T}(\hat{\theta}_{n_k})$. This contradiction implies that we must have $\hat{\theta}_n^* - \hat{\theta}_n \to 0$.

Next, we are going to show the second equality in (3.17). Adding and subtracting $\int L_b(z-x) f_{\bar{Z}}(x, \theta_0)dx]^2 d\Pi(z)$ from $\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) f_{\bar{Z}}(x, \theta)dx$, $\mathscr{T}_n^*(\theta)$ can be written as the sum of $\mathscr{T}(\theta)$ and the following two terms

$$C_{n1} = \int [\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) f_{\bar{Z}}(x,\theta_0)dx]^2 d\Pi(z)$$
$$C_{n2} = 2\int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) f_{\bar{Z}}(x,\theta_0)dx\right] \left[\int L_b(z-x)[f_{\bar{Z}}(x,\theta_0) - f_{\bar{Z}}(x,\theta)]dx\right] d\Pi(z).$$

Then $\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta) = C_{n1} + C_{n2}(\theta)$ and

$$\sup_{\theta\in\Theta} |\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta)| \le C_{n1} + \sup_{\theta\in\Theta} C_{n2}(\theta) \le C_{n1} + 2C_{n1}^{\frac{1}{2}} \sup_{\theta\in\Theta} \mathscr{T}^{\frac{1}{2}}(\theta).$$

It thus suffices to show $C_{n1} = o_p(1)$, which is already shown in (3.16), and $\sup_{\theta\in\Theta} \mathscr{T}(\theta) = O_p(1)$.

From (z3), by changing the variables, we have

$$\begin{aligned}
&\mathscr{T}(\theta) = \int \left[\int L_b(z-x)[f_{\bar{Z}}(x,\theta) - f_{\bar{Z}}(x,\theta_0)]dx\right]^2 d\Pi(z) \\
&\le \int \left[\int L_b(z-x)\|\theta-\theta_0\|^r l(x)dx\right]^2 d\Pi(z) = \|\theta-\theta_0\|^{2r} \int \left[\int L(u)l(z+ub)du\right]^2 d\Pi(z) \\
&\le c\|\theta-\theta_0\|^{2r}
\end{aligned}$$

for some positive constant c. Since Θ is compact, we obtain $\sup_{\theta\in\Theta} \mathscr{T}(\theta) < \infty$. Therefore $\sup_{\theta\in\Theta} |\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta)| \le o_p(1)$.

Next, let's show the first equality in (3.17). Similarly, we add and subtract $\int L_b(z-x) f_{\bar{Z}}(x,\theta)$ from $\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta)$, and $T_n(\theta)$ can be written as the sum of $\mathscr{T}_n^*(\theta)$ and the following two terms

$$D_{n1} = 2\int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) f_{\bar{Z}}(x,\theta)dx\right] \left[\int L_b(z-x)(f_{\bar{Z}}(x,\theta) - \hat{f}_{\bar{Z}}(x,\theta))dx\right] d\Pi(z)$$
$$D_{n2}(\theta) = \int \left[\int L_b(z-x)(f_{\bar{Z}}(x,\theta) - \hat{f}_{\bar{Z}}(x,\theta))dx\right]^2 d\Pi(z).$$

Then $T_n(\theta) - \mathscr{T}(\theta) = \mathscr{T}_n^*(\theta) - \mathscr{T}(\theta) + D_{n1}(\theta) + D_{n2}(\theta)$.

For $D_{n2}(\theta)$, from (x1) and (3.14),

$$\begin{aligned} D_{n2}(\theta) =& \int \left[\int L_b(z-x)(f_{\bar{Z}}(x,\theta) - \hat{f}_{\bar{Z}}(x,\theta))dx \right]^2 d\Pi(z) \\ \leq& \int \left[\int L_b(z-x) \int f_X(x-u,\theta)|f_{\bar{U}}(u) - \hat{f}_{\bar{U}}(u)|dudx \right]^2 d\Pi(z) \\ \leq& c \int \left[\int L_b(z-x)dx \right]^2 d\Pi(z) \left[\int |f_{\bar{U}}(u) - \hat{f}_{\bar{U}}(u)|du \right]^2, \end{aligned}$$

which is of order $o_p(1)$. Then we obtain

$$\begin{aligned} \sup_{\theta\in\Theta} |D_{n1}(\theta)| \leq& \sup_\theta |\mathscr{T}_n^*(\theta)|^{\frac{1}{2}} \sup_\theta |D_{n2}(\theta)|^{\frac{1}{2}} = \sup_\theta |\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta) + \mathscr{T}(\theta)|^{\frac{1}{2}} \sup_\theta |D_{n2}(\theta)|^{\frac{1}{2}} \\ \leq& \left[\sup_\theta |\mathscr{T}_n^*(\theta) - \mathscr{T}(\theta)|^{\frac{1}{2}} + \sup_\theta |\mathscr{T}(\theta)|^{\frac{1}{2}} \right] \sup_\theta |D_{n2}(\theta)|^{\frac{1}{2}} \leq o_p(1). \end{aligned}$$

This concludes the proof of (3.17). □

Remark 3.3.1. *One can replace condition (b2) with $nb \to \infty$, and Theorem 3.3.2 still holds.*

3.3.2 Asymptotic Normality of the Non-Centered MD Estimator

In this section, we shall report the asymptotic normality of θ_n^*. Let

$$\begin{aligned} U_n^*(z,\theta) =& \hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z,\theta), \quad U_n^*(z) = U_n^*(z,\theta_0), \\ Z_n^*(z,\theta) =& U_n^*(z) - U_n^*(z,\theta) = \hat{f}_{\bar{Z}}(z,\theta) - \hat{f}_{\bar{Z}}(z,\theta_0), \\ d_n(x,\theta,\theta_0) =& \hat{f}_{\bar{Z}}(x,\theta) - \hat{f}_{\bar{Z}}(x,\theta_0) - (\theta-\theta_0)^T \dot{\hat{f}}_{\bar{Z}}(x,\theta_0). \end{aligned} \tag{3.19}$$

Taking the derivative with respect to θ,

$$\dot{T}_n^*(\theta) = -2 \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z,\theta)] \dot{\hat{f}}_{\bar{Z}}(z,\theta) d\Pi(z) = -2 \int U_n^*(z,\theta) \dot{\hat{f}}_{\bar{Z}}(z,\theta) d\Pi(z).$$

Note that θ_n^* is the minimizer of $T_n^*(\theta)$ and also θ_0 is an interior point of Θ, so by the consistency, for sufficiently large n, θ_n^* will be an interior point of Θ, therefore $\dot{T}_n^*(\theta_n^*) = 0$.

Thus

$$\int U_n^*(z,\theta_0)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z) - \int U_n^*(z,\theta_n^*)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z) = \int U_n^*(z,\theta_0)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z).$$

Then we obtain

$$\int Z_n^*(z,\theta_n^*)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z) = \int U_n^*(z)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z),$$

which can be written as the sum of the following three terms:

$$\begin{aligned} S_n &= \int U_n^*(z)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z), \\ g_{n1} &= \int U_n^*(z)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)]d\Pi(z), \\ g_{n2} &= \int U_n^*(z)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]d\Pi(z). \end{aligned} \tag{3.20}$$

To proceed, the following lemmas are needed.

Lemma 3.3.1. *Suppose H_0, (z1)–(z4), (x1), (x2), (π1) and (b1)–(b3) holds. Then*

$$nb\|\theta_n^* - \theta_0\|^2 = O_p(1).$$

Proof. From (3.10) and (h3), one can easily verify

$$\int (U_n^*(z))^2 d\Pi(z) = T_n^*(\theta_0) = O_p\left(\frac{1}{nb}\right). \tag{3.21}$$

Let $D_n(\theta) = \int [Z_n^*(z,\theta)]^2 d\Pi(z)$. We are going to show $nbD_n(\theta_n^*) = O_p(1)$. To see this, observe that $T_n^*(\theta_n^*) \le T_n^*(\theta_0) = O_p(\frac{1}{nb})$. Thus $nbT_n^*(\theta_n^*) = O_p(1)$ and

$$nbD_n(\theta_n^*) = nb\int [U_n^*(z) - U_n^*(z,\theta_n^*)]^2 d\Pi(z) \le 2nb[T_n^*(\theta_0) + T_n^*(\theta_n^*)] = O_p(1).$$

Next, we shall show $\frac{D_n(\theta_n^*)}{\|\theta_n^*-\theta_0\|^2} \ge B$ with arbitrarily large probability, where B is an arbitrary positive number.

$$O_p(1) = nbD_n(\theta_n^*) = nb\|\theta_n^* - \theta_0\|^2 \frac{D_n(\theta_n^*)}{\|\theta_n^* - \theta_0\|^2}.$$

Recalling the definition of $d_n(z,\theta,\theta_0)$ and $Z_n^*(z,\theta)$ from (3.19), we have

$$\begin{aligned}
\frac{D_n(\theta_n^*)}{\|\theta_n^* - \theta_0\|^2} =& \frac{\int [\hat{f}_{\bar{Z}}(z,\theta_n^*) - \hat{f}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z)}{\|\theta_n^* - \theta_0\|^2} \\
=& \frac{\int [d_n(z,\theta_n^*,\theta_0) + (\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z)}{\|\theta_n^* - \theta_0\|^2} \\
\geq & \int \frac{d_n^2(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|^2} d\Pi(z) + \int \left[\frac{(\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)}{\|\theta_n^* - \theta_0\|}\right]^2 d\Pi(z) \\
& - 2\left[\int \frac{d_n^2(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|^2} d\Pi(z)\right]^{\frac{1}{2}} \left[\int \left[\frac{(\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)}{\|\theta_n^* - \theta_0\|}\right]^2 d\Pi(z)\right]^{\frac{1}{2}}.
\end{aligned}$$

By assumption (x2) and the consistency of θ_n^*, we can show that

$$\begin{aligned}
& \int \frac{d_n^2(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|^2} d\Pi(z) \\
=& \int \left[\int \frac{f_X(z-u,\theta_n^*) - f_X(z-u,\theta_0) - (\theta_n^* - \theta_0)^T \dot{f}_X(z-u,\theta_0)}{\|\theta_n^* - \theta_0\|} \hat{f}_{\bar{U}}(u) du\right]^2 d\Pi(z)
\end{aligned}$$

is of order $o_p(1)$.

For the second term, we notice that $\int \left[\frac{(\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)}{\|\theta_n^* - \theta_0\|}\right]^2 d\Pi(z) \geq \inf_{\|s\|=1} \Sigma_n(s)$, where $\Sigma_n(s) = \int [s^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z)$. By the usual calculations, one sees for each $s \in \mathbb{R}^q$,

$$\begin{aligned}
\Sigma_n(s) =& s^T \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) \dot{\hat{f}}_{\bar{Z}}^T(z,\theta_0) d\Pi(z)\ s \\
=& s^T \int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0) + \dot{f}_{\bar{Z}}(z,\theta_0)][\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0) + \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)\ s \\
=& s^T \int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)][\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)\ s \\
& + 2s^T \int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)][\dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)\ s + s^T \int \dot{f}_{\bar{Z}}(z,\theta_0) \dot{f}_{\bar{Z}}^T(z,\theta_0) d\Pi(z)\ s \\
\to & s^T \int \dot{f}_{\bar{Z}}(z,\theta_0) \dot{f}_{\bar{Z}}^T(z,\theta_0) d\Pi(z)\ s = s^T \Sigma_0 s,
\end{aligned}$$

where Σ_0 is as in (3.7). Thus, by the Cauchy–Schwarz inequality, the cross term

$$\left[\int \frac{d_n^2(z,\theta_n^*,\theta_0)}{\|\theta_n^*-\theta_0\|^2} d\Pi(z)\right]^{\frac{1}{2}} \left[\int \left(\frac{(\theta_n^*-\theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)}{\|\theta_n^*-\theta_0\|}\right)^2 d\Pi(z)\right]^{\frac{1}{2}} = o_p(1).$$

For any $\delta > 0$, and any two unit vectors $s, s_1 \in \mathbb{R}^q$, $\|s-s_1\| < \delta$, we have $|\Sigma_n(s)-\Sigma_n(s_1)| \leq \delta(\delta+2)\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\|^2 d\Pi(z)$. By observing

$$\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\|^2 d\Pi(z) = O_p(1) \tag{3.22}$$

and the compactness of the set $\{s \in \mathbb{R}^q; \|s\| = 1\}$, we have $\sup_{\|s\|=1} |\Sigma_n(s) - s^T\Sigma_0 s| = o_p(1)$. In sum, $\frac{D_n(\theta_n^*)}{\|\theta_n^*-\theta_0\|^2} \geq \inf_{\|s\|=1} s^T\Sigma_0 s$ with arbitrarily large probability.

This concludes the proof of Lemma 3.3.1. □

Lemma 3.3.2. *Suppose H_0, (z1)–(z4), (x1)–(x3), (b1)–(b3), (h1)–(h3), (u1),(π1) and (kl) hold. Then $n^{\frac{1}{2}}S_n \Rightarrow N_q(0,\Sigma)$, $n^{\frac{1}{2}}g_{n1} = o_p(1)$, $n^{\frac{1}{2}}g_{n2} = o_p(1)$, where S_n, g_{n1}, g_{n2} are as defined in (3.20), $\Sigma = \Sigma_1+\Sigma_2+\Sigma_3$ with $\Sigma_i, i = 1,2,3$ as in (3.7) .*

Proof. For convenience, we shall give the proof here only for the case $d = 1$, i.e., when $\dot{f}_{\bar{Z}}(z,\theta_0)$ is one dimensional. For multidimensional case, the result can be proved by using linear combination of its components instead of $\dot{f}_{\bar{Z}}(z,\theta_0)$, and applying the same argument.

Add and subtract $E\hat{f}_{\bar{Z}}(z), f_{\bar{Z}}(z,\theta_0)$ and $E\hat{f}_{\bar{Z}}(z,\theta_0)$ inside the parenthesis of $U_n^*(z)$, then

$$\begin{aligned}
S_n &= \int U_n^*(z)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\
&= \int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z) + \int [E\hat{f}_{\bar{Z}}(z,\theta_0) - \hat{f}_{\bar{Z}}(z,\theta_0)]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\
&\quad + \int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z) + \int [f_{\bar{Z}}(z,\theta_0) - E\hat{f}_{\bar{Z}}(z,\theta_0)]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\
&=: S_{n1} + S_{n2} + S_{n3} + S_{n4}.
\end{aligned}$$

In order to show $n^{\frac{1}{2}}S_n \Rightarrow N_q(0,\Sigma)$, we need only to show $n^{\frac{1}{2}}(S_{n1}+S_{n2}) \Rightarrow N_q(0,\Sigma)$, and $n^{\frac{1}{2}}S_{n3} = o(1)$, $n^{\frac{1}{2}}S_{n4} = o(1)$.

Consider S_{n3} and S_{n4} first. Since $E\hat{f}_{\bar{Z}}(z) = f_{\bar{Z}}(z,\theta_0) + \frac{b^2}{2}\int L(v)v^2 f''_{\bar{Z}}(z+\tau_1 vb,\theta_0)dv$, where $0<\tau_1<1$, from (z4) and (b3), one can verify

$$\begin{aligned}
n^{\frac{1}{2}}S_{n3} =& n^{\frac{1}{2}}\frac{b^2}{2}\iint L(v)v^2 f''_{\bar{Z}}(z+\tau_1 vb,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)dvd\Pi(z)\\
\leq& \frac{C}{2}n^{\frac{1}{2}}b^2\int L(v)v^2dv\int \dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z) = O(n^{\frac{1}{2}}b^2) = o(1).
\end{aligned}$$

Similarly, there is a $0<\tau_2<1$ such that

$$\begin{aligned}
E\hat{f}_{\bar{Z}}(z,\theta_0) =& \int f_X(z-u,\theta_0)\left[f_{\bar{U}}(u) + \frac{h^2}{2}\int K(v)v^2 f''_{\bar{U}}(u+\tau_2 vh)dv\right]du\\
=& f_{\bar{Z}}(z,\theta_0) + \frac{h^2}{2}\int f_X(z-u,\theta_0)\left[\int K(v)v^2 f''_{\bar{U}}(u+\tau_2 vh)dv\right]du
\end{aligned}$$

and from (u1) and (h3), one obtains $n^{\frac{1}{2}}S_{n4} = O(n^{\frac{1}{2}}h^2) = o(1)$.

Next, we consider $n^{\frac{1}{2}}(S_{n1}+S_{n2})$. Let $G_{\tilde{U}_i}(z,\theta_0) = \int f_X(z-u,\theta_0)K_h(u-\tilde{U}_i)du$. and rewrite $\hat{f}_{\bar{Z}}(z,\theta_0) = \frac{1}{n}\sum_{i=1}^n G_{\tilde{U}_i}(z,\theta_0)$. Then

$$\begin{aligned}
& n^{\frac{1}{2}}(S_{n1}+S_{n2})\\
=& n^{\frac{1}{2}}\int [(\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)) + (E\hat{f}_{\bar{Z}}(z,\theta_0) - \hat{f}_{\bar{Z}}(z,\theta_0))]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\
=& n^{\frac{1}{2}}\int \frac{1}{n}\sum_{i=1}^n[(L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z}_1)) + (EG_{\tilde{U}_1}(z,\theta_0) - G_{\tilde{U}_i}(z,\theta_0))]\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\
=& \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\int (L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z}_1))\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right.\\
& \left. + \int (EG_{\tilde{U}_1}(z,\theta_0) - G_{\tilde{U}_i}(z,\theta_0))\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right]\\
=:& \frac{1}{\sqrt{n}}\sum_{i=1}^n(s_{ni1}+s_{ni2}) =: \frac{1}{\sqrt{n}}\sum_{i=1}^n s_{ni}.
\end{aligned}$$

Note that $s_{ni}, 1\leq i\leq n$ are i.i.d centered r.v.'s for each n. By the Lindeberg-Feller

Central Limit Theorem, it is sufficient to show that as $n \to \infty$,

$$Es_{ni}^2 \to \Sigma, \tag{3.23}$$

$$E\{s_{ni}^2 I(|s_{ni}| > n^{\frac{1}{2}}\eta)\} \to 0, \quad \forall \eta > 0. \tag{3.24}$$

By Fubini,

$$\begin{aligned} Es_{ni1}^2 =& E\left[\int (L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z}_1))\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right]^2 \\ =& \iint \left[EL_b(z-\bar{Z}_1)L_b(y-\bar{Z}_1) - EL_b(z-\bar{Z}_1)EL_b(y-\bar{Z}_1)\right] \\ & \cdot \dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y), \end{aligned}$$

where

$$\begin{aligned} &EL_b(z-\bar{Z}_i)L_b(y-\bar{Z}_i) = \int L_b(z-t)L_b(y-t)f_{\bar{Z}}(t,\theta_0)dt, \\ &EL_b(z-\bar{Z}_1) = \int L(s)f_{\bar{Z}}(z-bs,\theta_0)ds, \quad EL_b(y-\bar{Z}_1) = \int L(s)f_{\bar{Z}}(y-bs,\theta_0)ds. \end{aligned} \tag{3.25}$$

In the sequel, for the sake of brevity, we shall often write $dxdydz = \mathrm{d}(x,y,z)$. By the transformation $z - t = bz_1$ and $y - t = by_1$ and continuity of $f_{\bar{Z}}$ and π, we obtain

$$\begin{aligned} Es_{ni1}^2 =& \int L_b(z-t)L_b(y-t)f_{\bar{Z}}(t,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)\,\mathrm{d}(t,z,y) \\ & - \iint \left(\int L(s)f_{\bar{Z}}(z-bs,\theta_0)ds\right)\left(\int L(s)f_{\bar{Z}}(y-bs,\theta_0)ds\right)\dot{f}_{\bar{Z}}(z,\theta_0) \\ & \cdot \dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)dzdy \\ =& \int L(z_1)L(y_1)f_{\bar{Z}}(t,\theta_0)\dot{f}_{\bar{Z}}(t+bz_1,\theta_0)\dot{f}_{\bar{Z}}(t+by_1,\theta_0)\pi(t+bz_1) \\ & \cdot \pi(t+by_1)\,\mathrm{d}(t,z_1,y_1) - \iint \left(\int L(s)f_{\bar{Z}}(z-bs,\theta_0)ds\right) \\ & \cdot \left(\int L(s)f_{\bar{Z}}(y-bs,\theta_0)ds\right)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)dzdy \end{aligned}$$

which converges to Σ_1 as $b \to 0$. Next, consider Es_{ni2}^2. By Fubini, we have

$$\begin{aligned}
Es_{ni2}^2 =& E\left[\int (EG_{\tilde{U}_1}(z,\theta_0) - G_{\tilde{U}_i}(z,\theta_0))\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right]^2 \\
=& E\iint \left(G_{\tilde{U}_i}(z,\theta_0) - EG_{\tilde{U}_1}(z,\theta_0)\right)\left(G_{\tilde{U}_i}(y,\theta_0) - EG_{\tilde{U}_1}(y,\theta_0)\right) \\
& \cdot \dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y) \\
=& \iint \left(EG_{\tilde{U}_1}(z,\theta_0)G_{\tilde{U}_1}(y,\theta_0) - EG_{\tilde{U}_1}(z,\theta_0)EG_{\tilde{U}_1}(y,\theta_0)\right) \\
& \cdot \dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y),
\end{aligned}$$

where

$$\begin{aligned}
EG_{\tilde{U}_1}(z,\theta_0)G_{\tilde{U}_1}(y,\theta_0) &= \int f_X(z-u,\theta_0)K_h(u-t)f_X(y-v,\theta_0)K_h(v-t)f_{\bar{U}}(t)\,\mathrm{d}(u,v,t), \\
EG_{\tilde{U}_1}(z,\theta_0) &= \iint f_X(z-u,\theta_0)K(s)f_{\bar{U}}(u-hs)duds, \\
EG_{\tilde{U}_1}(y,\theta_0) &= \iint f_X(y-u,\theta_0)K(s)f_{\bar{U}}(u-hs)duds. \qquad (3.26)
\end{aligned}$$

Therefore, by the transformation $u - t = hu_1, v - t = hv_1$, taking the limit, and using the continuity of f_X and $f_{\bar{U}}$, one obtains

$$\begin{aligned}
Es_{ni2}^2 =& \int f_X(z-u,\theta_0)K_h(u-t)f_X(y-v,\theta_0)K_h(v-t)f_{\bar{U}}(t)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0) \\
& \cdot \pi(z)\pi(y)\,\mathrm{d}(u,v,t,z,y) - \iint \left(\iint f_X(z-u,\theta_0)K(s)f_{\bar{U}}(u-hs)duds\right) \\
& \cdot \left(\iint f_X(y-u,\theta_0)K(s)f_{\bar{U}}(u-hs)duds\right)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)dzdy \\
\xrightarrow{h\to 0}& \int f_X(z-t,\theta_0)K(u_1)f_X(y-t,\theta_0)K(v_1)f_{\bar{U}}(t)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0) \\
& \cdot \pi(z)\pi(y)\,\mathrm{d}(u_1,v_1,t,z,y) - \iint \left(\iint f_X(z-u,\theta_0)K(s)f_{\bar{U}}(u)duds\right) \\
& \cdot \left(\iint f_X(y-u,\theta_0)K(s)f_{\bar{U}}(u)duds\right)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)d\Pi(z)dy
\end{aligned}$$

The last term is indeed Σ_2 by simple algebra.

Next, we consider $Es_{ni1}s_{ni2}$,

$$\begin{aligned}Es_{ni1}s_{ni2} =&E\int \left(L_b(z-\bar{Z}_i)-EL_b(z-\bar{Z}_1)\right)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\\&\cdot\int \left(EG_{\tilde{U}_1}(y,\theta_0)-G_{\tilde{U}_i}(y,\theta_0)\right)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(y)\\=&\iint \left(EL_b(z-\bar{Z}_1)EG_{\tilde{U}_1}(y,\theta_0)-EL_b(z-\bar{Z}_1)G_{\tilde{U}_1}(y,\theta_0)\right)\\&\cdot\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)dzdy.\end{aligned}$$

By the transformation $2(z-s)-p-q=2bv$ and $2u-p+q=2ht$, we have

$$\begin{aligned}&EL_b(z-\bar{Z}_i)G_{\tilde{U}_i}(y,\theta_0)=E\left[L_b(z-\bar{Z}_i)\int f_X(y-u,\theta_0)K_h(u-\tilde{U}_i)du\right]\\=&E\left[L_b\left(z-X_i-\frac{U_1+U_2}{2}\right)\int f_X(y-u,\theta_0)K_h\left(u-\frac{U_1-U_2}{2}\right)du\right]\\=&\int L_b\left(\frac{2(z-s)-p-q}{2}\right)f_X(y-u,\theta_0)K_h\left(\frac{2u-p+q}{2}\right)f_X(s,\theta_0)f_U(p)f_U(q)\,\mathrm{d}(u,s,p,q)\\=&2\int L(v)f_X(y-u,\theta_0)K(t)f_X(s,\theta_0)f_U(z-s-bv+u-ht)\\&\cdot f_U(z-s-bv-u+ht)\,\mathrm{d}(u,s,v,t).\end{aligned}\tag{3.27}$$

Combining (3.25), (3.26) and (3.27), using the assumed continuity of $f_{\bar{Z}}, f_X, f_{\bar{U}}, f_U$, we obtain $2Es_{ni1}s_{ni2}$ converges to

$$\begin{aligned}&2\left[\int f_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right]^2-4\iint\left(\iint f_X(y-u,\theta_0)f_U(z-s+u)\right.\\&\left.\cdot f_U(z-s-u)f_X(s,\theta_0)dsdu\right)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y)=\Sigma_3.\end{aligned}$$

Therefore, $Es_{ni}^2\to\Sigma=\Sigma_1+\Sigma_2+\Sigma_3$. Hence (3.23) is proved.

To prove (3.24), note that by the C_r inequality, $E\{s_{ni}^2I(|s_{ni}|>n^{\frac{1}{2}}\eta)\}$ has upper bound

$$\eta^{-\delta}n^{-\delta/2}E|s_{ni}|^{2+\delta}\le\eta^{-\delta}n^{-\delta/2}2^{1+\delta}E\left(|s_{ni1}|^{2+\delta}+|s_{ni2}|^{2+\delta}\right)$$

By the Hölder's inequality, $E(|s_{ni1}|^{2+\delta})$ is bounded above by

$$\begin{aligned}
&E\left|\left[\int \left(L_b(z-\bar{Z}_i)-EL_b(z-\bar{Z}_1)\right)^{\frac{2+\delta}{2}}d\Pi(z)\right]^{\frac{2}{2+\delta}}\left[\int \left(\dot{f}_{\bar{Z}}(z,\theta_0)\right)^{\frac{2+\delta}{\delta}}d\Pi(z)\right]^{\frac{\delta}{2+\delta}}\right|^{2+\delta}\\
=&E\left|\int \left(L_b(z-\bar{Z}_i)-EL_b(z-\bar{Z}_1)\right)^{\frac{2+\delta}{2}}d\Pi(z)\right|^{2}\left[\int \left(\dot{f}_{\bar{Z}}(z,\theta_0)\right)^{\frac{2+\delta}{\delta}}d\Pi(z)\right]^{\delta}=O(b^{-\delta}),
\end{aligned}$$

and

$$\begin{aligned}
E\left(|s_{ni2}|^{2+\delta}\right)=&E\left|\int \left(EG_{\tilde{U}_1}(z,\theta_0)-G_{\tilde{U}_i}(z,\theta_0)\right)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right|^{2+\delta}\\
=&E\left|\int G_{\tilde{U}_i}(z,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)-\int EG_{\tilde{U}_1}(z,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right|^{2+\delta}\\
=&\int\Big|\iint f_X(z-u,\theta_0)K_h(u-s)\dot{f}_{\bar{Z}}(z,\theta_0)dud\Pi(z)\\
&-\int EG_{\tilde{U}_1}(z,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\Big|^{2+\delta}f_{\bar{U}}(s)ds=O(1).
\end{aligned}$$

Therefore, from (b2),

$$E\left\{s_{ni}^2 I(|s_{ni}|>n^{\frac{1}{2}}\eta)\right\}\le n^{-\delta/2}\left(O(b^{-\delta})+O(1)\right)=o(1),\quad \forall\eta>0.$$

By L-F C.L.T, we have $n^{\frac{1}{2}}S_n\Rightarrow N(0,\Sigma)$, where $\Sigma=\Sigma_1+\Sigma_2+\Sigma_3$.

To finish proving Lemma 3.3.2, we need only to show $n^{\frac{1}{2}}g_{n1}=o_p(1)$ and $n^{\frac{1}{2}}g_{n2}=o_p(1)$. In fact, from (b2), (h3), (u1), (kl) and (3.21), by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
|n^{\frac{1}{2}}g_{n1}|=&\left|n^{\frac{1}{2}}\int U_n^*(z)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)-\dot{f}_{\bar{Z}}(z,\theta_0)]d\Pi(z)\right|\\
\le& n^{\frac{1}{2}}\left(\int [U_n^*(z)]^2d\Pi(z)\right)^{\frac{1}{2}}\left(\int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)-\dot{f}_{\bar{Z}}(z,\theta_0)]^2d\Pi(z)\right)^{\frac{1}{2}}\\
=&n^{\frac{1}{2}}O_p\left(\frac{1}{\sqrt{nb}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=O_p\left(\frac{1}{\sqrt{nb}}\right)=o_p(1).
\end{aligned}\tag{3.28}$$

Similarly, from (x3), (3.21), and the result from Lemma 3.3.1, we obtain

$$\begin{aligned}|n^{\frac{1}{2}}g_{n2}| &= \left| n^{\frac{1}{2}} \int U_n^*(z)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]d\Pi(z) \right| \\ &\leq \left[n^{\frac{1}{2}} \int (U_n^*(z))^2 d\Pi(z) \right]^{\frac{1}{2}} \left[n^{\frac{1}{2}} \int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) \right]^{\frac{1}{2}} \\ &= O_p\left(\frac{1}{\sqrt{n^{\frac{1}{2}}b}} \right) o_p\left(\sqrt{n^{\frac{1}{2}}b} \right) O_p(1) = o_p(1).\end{aligned}$$

This completes the proof of Lemma 3.3.2. □

Lemma 3.3.3. *Suppose H_0, (z1)–(z4), (x1), (x2), (b1)–(b3), and (π1). Then*

$$\sqrt{n} \int Z_n^*(z,\theta_n^*)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z) = R_n\sqrt{n}(\theta_n^* - \theta_0)$$

with $R_n = \Sigma_0 + o_p(1)$, where $Z_n^*(z,\theta)$ is as defined in (3.19), and Σ_0 is as in (3.7).

Proof. Recalling $d_n(z,\theta,\theta_0)$ defined in (3.19), $n^{\frac{1}{2}} \int Z_n^*(z,\theta_n^*)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z)$ can be rewritten as

$$\begin{aligned}&n^{\frac{1}{2}} \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)(d_n(z,\theta_n^*,\theta_0) + (\dot{f}_{\bar{Z}}(z,\theta_0))^T(\theta_n^* - \theta_0))d\Pi(z) \\ =& \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) \left[\frac{d_n(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|} \frac{(\theta_n^* - \theta_0)^T}{\|\theta_n^* - \theta_0\|} + (\dot{\hat{f}}_{\bar{Z}}(z,\theta_0))^T \right] d\Pi(z) \cdot [n^{\frac{1}{2}}(\theta_n^* - \theta_0)]\end{aligned}$$

Therefore, we only need to show that

$$\left\| \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) \frac{d_n(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|} \frac{(\theta_n^* - \theta_0)^T}{\|\theta_n^* - \theta_0\|} d\Pi(z) \right\| = o_p(1) \tag{3.29}$$

and

$$\int \dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) = \Sigma_0 + o_p(1). \tag{3.30}$$

To prove (3.29), from (x2) and the consistency of θ_n^*, the L.H.S. of (3.29) is bounded above by

$$
\begin{aligned}
&\sup_x \frac{|f_X(x,\theta_n^*) - f_X(x,\theta_0) - (\theta_n^* - \theta_0)^T \dot{f}_X(x,\theta_0)|}{\|\theta_n^* - \theta_0\|} \int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)\| d\Pi(z) \int \hat{f}_{\bar{U}}(u) du \\
=&o_p(1)O_p(1) = o_p(1),
\end{aligned}
$$

by observing the fact that $\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)\| d\Pi(z)$ is bounded above by

$$
\begin{aligned}
&\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\| d\Pi(z) + \int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\| d\Pi(z) \\
\le& \int \|\dot{f}_X(z-u,\theta_n^*) - \dot{f}_X(z-u,\theta_0)\| \hat{f}_{\bar{U}}(u) du d\Pi(z) + \int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\| d\Pi(z) \\
=&o_p(b^{1/2})O_p(1) + O_p(1) = O_p(1),
\end{aligned}
$$

from Lemma 3.3.1, (x3) and (3.22). Next, we will prove (3.30). By the Cauchy-Schwarz inequality, from (b1) and (x3), one sees that

$$
\begin{aligned}
&\left\| \int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)] (\dot{\hat{f}}_{\bar{Z}}(z,\theta_0))^T d\Pi(z) \right\| \\
\le& \left[\int \left\| \dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) \right\|^2 d\Pi(z) \right]^{1/2} \left[\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)\|^2 d\Pi(z) \right]^{1/2} \\
\le& o_p(b^{1/2})O_p(1) = o_p(1).
\end{aligned}
$$

Therefore, the L.H.S. of (3.30) can be written as

$$
\begin{aligned}
&\int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*) - \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)] [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) + \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) \\
\le& o_p(1) + \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) = \Sigma_0 + o_p(1),
\end{aligned}
$$

where the last step is due to the fact $\int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) = \Sigma_0 + o_p(1)$, which is verified below.

In fact,

$$
\begin{aligned}
&\int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) \\
=&\int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0) + \dot{f}_{\bar{Z}}(z,\theta_0)][\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0) + \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) \\
=&\int [\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)][\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) \\
&+ 2\int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) + \int \dot{f}_{\bar{Z}}(z,\theta_0)[\dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)
\end{aligned}
$$

where $\int \|\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)\| d\Pi(z) = o_p(1)$ and by using the Cauchy–Schwarz inequality, we know

$$
\left\| \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0) - \dot{f}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z) \right\| \le O_p(1)o_p(1) = o_p(1).
$$

$\square$

Theorem 3.3.3. *Under H_0, when (x1)–(x3), (z1)–(z4), (h1)–(h3), (b1)–(b3), (π1), (kl) and (u1) hold, we have $n^{\frac{1}{2}}(\theta_n^* - \theta_0) \to N_q(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, where $\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_3$ and Σ_i, $i = 0,1,2,3$ are defined in (3.7).*

Proof. Based on the discussion at the beginning of this section, it is sufficient to show that $\sqrt{n}\int U_n^*(z)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z)$ converges in distribution to the normal distribution $N(0,\Sigma)$, while $\sqrt{n}\int Z_n^*(z,\theta_n^*)\dot{\hat{f}}_{\bar{Z}}(z,\theta_n^*)d\Pi(z) = \sqrt{n}\Sigma_0(\theta_n^* - \theta_0) + o_p(1)$. Then the theorem can be proved by combining the results from Lemma 3.3.1, 3.3.2 and 3.3.3. $\square$

3.3.3 Asymptotic Normality of the Centered MD Estimator

We will derive the asymptotic normality of $\hat{\theta}_n$, the minimizer of $T_n(\theta)$ in (3.6).

The following fact shall be used constantly in the subsequent proofs. Although the proof is not complicated, it is reproduced here, with a short justification, for the sake of completeness.

Suppose $g(z,x)$ is a bivariate function such that and for each z in the support of Π,

$Eg(z, X)$ and $Eg^2(z, X)$ are continuous function of z. Then we have

$$\begin{aligned}
&\int E\left[\frac{1}{n}\sum_{i=1}^{n} g(z, X_i)\right]^2 d\Pi(z) = \int E\left[\frac{1}{n^2}\sum_{i=1}^{n} g^2(z, X_i) + \frac{1}{n^2}\sum_{i\neq j} g(z, X_i)g(z, X_j)\right] d\Pi(z)\\
=&\frac{1}{n}\int Eg^2(z, X_1)d\Pi(z) + \frac{n(n-1)}{n^2}\int [Eg(z, X_1)]^2 d\Pi(z) = O_p(1). \qquad (3.31)
\end{aligned}$$

Since the derivation of the asymptotic normality of $\hat{\theta}_n$ is tedious, a series of lemmas will be introduced first to help us better understand the whole process.

Define

$$\begin{aligned}
\dot{\mu}_n(z, \theta) =& \int L_b(z-x)\dot{\hat{f}}_{\bar{Z}}(x, \theta)dx, \quad \dot{\mu}_h(z) = E\dot{\mu}_n(z, \theta_0),\\
U_n(z, \theta) =& \hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x, \theta)dx, \quad U_n(z) = U_n(z, \theta_0),\\
Z_n(z, \theta) =& U_n(z) - U_n(z, \theta) = \int L_b(z-x)[\hat{f}_{\bar{Z}}(x, \theta) - \hat{f}_{\bar{Z}}(x, \theta_0)]dx, \qquad (3.32)
\end{aligned}$$

Lemma 3.3.4. *Suppose (x1), then* $L_n := \int \dot{\mu}_n(z, \theta_0)(\dot{\mu}_n(z, \theta_0))^T d\Pi(z) = \Sigma_0 + o_p(1)$, *where* Σ_0 *is as in* (3.7).

Proof. Recall $\dot{\mu}_h(z) = E\dot{\mu}_n(z, \theta_0)$. Then

$$\begin{aligned}
&\left\|L_n - \int \dot{\mu}_h(z)(\dot{\mu}_h(z))^T d\Pi(z)\right\|\\
=&\left\|\int (\dot{\mu}_n(z, \theta_0) \pm \dot{\mu}_h(z))(\dot{\mu}_n(z, \theta_0) \pm \dot{\mu}_h(z))^T d\Pi(z) - \int \dot{\mu}_h(z)(\dot{\mu}_h(z))^T d\Pi(z)\right\|\\
\leq&\int \|\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)\|^2 d\Pi(z) + 2\int \|\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)\|\ \|\dot{\mu}_h(z)\| d\Pi(z) \qquad (3.33)
\end{aligned}$$

where $\pm$ stands for minus and plus the term afterwards.

For $\int \|\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)\|^2 d\Pi(z)$, first note that $\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)$ is an average of centered iid r.v.'s. Using Fubini Theorem, doing the transformation $z-x = bx_1,\ z-y = by_1,\ u-s = hu_1,\ v-s = hv_1$, by the fact that variance is bounded above by the second moment, and the assumed continuity of f_X, we obtain that $E\int \|\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)\|^2 d\Pi(z)$ is bounded above by

$$
\begin{aligned}
&\frac{1}{n}\int E\left\|\int \frac{1}{b}L\left(\frac{z-x}{b}\right)\int \dot{f}_X(x-u,\theta_0)\frac{1}{h}K\left(\frac{u-\tilde{U}_1}{h}\right)dudx\right\|^2 d\Pi(z)\\
=&\frac{1}{n}\int L(x_1)\dot{f}_X(z-bx_1-s-hu_1,\theta_0)K(u_1)L(y_1)\dot{f}_X^T(z-by_1-s-hv_1,\theta_0)\\
&\cdot K(v_1)f_{\bar{U}}(s)\,\mathrm{d}(u_1,x_1,v_1,y_1,s,\Pi(z))=O\left(\frac{1}{n}\right)=o(1).
\end{aligned}
$$

Therefore,

$$
\int \|\dot{\mu}_n(z,\theta_0)-\dot{\mu}_h(z)\|^2 d\Pi(z)=O_p\left(\frac{1}{n}\right)=o_p(1). \tag{3.34}
$$

For the second term in (3.33), from (3.31), one can observe that

$$
\int \|\dot{\mu}_h(z)\|^2 d\Pi(z)=\int \|E\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z)=O(1). \tag{3.35}
$$

By the Cauchy-Schwarz inequality, $\int \|\dot{\mu}_n(z,\theta_0)-\dot{\mu}_h(z)\|\;\|\dot{\mu}_h(z)\|d\Pi(z)$ is of order $o_p(1)$. Therefore $\|L_n-\int \dot{\mu}_h(z)(\dot{\mu}_h(z))^T d\Pi(z)\|=o(1)$. To finish this proof, we only need to show $\int \dot{\mu}_h(z)(\dot{\mu}_h(z))^T d\Pi(z)=\Sigma_0+o(1)$.

By changing variables, we have

$$
\begin{aligned}
\dot{\mu}_h(z)=&E\dot{\mu}_n(z,\theta_0)\\
=&\frac{1}{n}\sum_{i=1}^n E\left[\int L_b(z-x)\int \dot{f}_X(x-u,\theta_0)K_h(u-\tilde{U}_i)dudx\right]\\
=&\int L_b(z-x)\dot{f}_X(x-u,\theta_0)K_h(u-s)f_{\bar{U}}(s)\,\mathrm{d}(s,u,x)\\
=&\int L(x)\dot{f}_X(z-bx-s-hu,\theta_0)K(u)f_{\bar{U}}(s)\,\mathrm{d}(s,u,x).
\end{aligned} \tag{3.36}
$$

Therefore, by using the continuity assumption of f_X, one obtains

$$
\begin{aligned}
&\int \dot{\mu}_h(z)(\dot{\mu}_h(z))^T d\Pi(z) \\
=&\int L(x)\dot{f}_X(z-bx-s-hu,\theta_0)K(u)f_{\bar{U}}(s) \\
&\cdot L(y)\dot{f}_X^T(z-by-t-hv,\theta_0)K(v)f_{\bar{U}}(t)\,\mathrm{d}(s,u,x,t,v,y,\Pi(z)) \\
\xrightarrow[h\to 0]{b\to 0}&\int L(x)\dot{f}_X(z-s,\theta_0)K(u)f_{\bar{U}}(s)L(y)\dot{f}_X^T(z-t,\theta_0)K(v)f_{\bar{U}}(t)\,\mathrm{d}(s,u,x,t,v,y,\Pi(z)) \\
=&\int \dot{f}_X(z-s,\theta_0)f_{\bar{U}}(s)\dot{f}_X^T(z-t,\theta_0)f_{\bar{U}}(t)\,\mathrm{d}(s,t,\Pi(z))
\end{aligned}
$$

which is $\int \dot{f}_{\bar{Z}}(z,\theta_0)(\dot{f}_{\bar{Z}}(z,\theta_0))^T d\Pi(z)$ by simple algebra. ☐

Lemma 3.3.5. *Suppose H_0, (z1)–(z4), (b1)–(b2), (x1)–(x2), (π1) hold, then $nb\|\hat{\theta}_n-\theta_0\|^2 = O_p(1)$.*

Proof. Note that $\int U_n^2(z)d\Pi(z) = \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_0)dx\right]^2 d\Pi(z) = T_n(\theta_0)$. From (3.15) and (h3), one can verify that

$$\int U_n^2(z)d\Pi(z) = T_n(\theta_0) = O_p\left(\frac{1}{nb}\right). \tag{3.37}$$

Recall $Z_n(z,\theta) = U_n(z) - U_n(z,\theta)$ and let $D_n(\theta) = \int Z_n^2(z,\theta)d\Pi(z)$. We are going to show $nbD_n(\hat{\theta}_n) = O_p(1)$. To see this, observe that $nbT_n(\theta_0) = O_p(1)$ as shown above and $\hat{\theta}_n$ is the minimizer of $T_n(\theta)$. From $T_n(\hat{\theta}_n) \le T_n(\theta_0) = O_p(\frac{1}{nb})$, we know $nbT_n(\hat{\theta}_n) = O_p(1)$ and

$$
\begin{aligned}
nbD_n(\hat{\theta}_n) =& nb\int [U_n(z) - U_n(z,\hat{\theta}_n)]^2 d\Pi(z) \le 2\left[nb\int U_n^2(z)d\Pi(z) + nb\int U_n^2(z,\hat{\theta}_n)d\Pi(z)\right] \\
=& 2[nbT_n(\theta_0) + nbT_n(\hat{\theta}_n)] = O_p(1).
\end{aligned}
$$

Next, we shall show $\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n-\theta_0\|^2} \ge B$ with arbitrarily large probability, where B is an arbitrary positive number.

$$O_p(1) = nbD_n(\hat{\theta}_n) = nb\|\hat{\theta}_n-\theta_0\|^2 \frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n-\theta_0\|^2}.$$

Recall the definition of $d_n(z,\theta,\theta_0)$ in (3.19) and $Z_n(z,\hat{\theta}_n)$ in (3.32), we have

$$
\begin{aligned}
\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n-\theta_0\|^2} =& \frac{\int\left[\int L_b(z-x)(\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)-\hat{f}_{\bar{Z}}(x,\theta_0))dx\right]^2 d\Pi(z)}{\|\hat{\theta}_n-\theta_0\|^2}\\
=& \frac{\int\left[\int L_b(z-x)d_n(x,\hat{\theta}_n,\theta_0)dx+\int L_b(z-x)(\hat{\theta}_n-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)dx\right]^2 d\Pi(z)}{\|\hat{\theta}_n-\theta_0\|^2}\\
\geq& B_{n1}+B_{n2}-2B_{n1}^{1/2}B_{n2}^{1/2},
\end{aligned}
$$

where

$$
\begin{aligned}
B_{n1} &= \int\left[\int L_b(z-x)\frac{d_n(x,\hat{\theta}_n,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dx\right]^2 d\Pi(z),\\
B_{n2} &= \int\left[\int L_b(z-x)\frac{(\hat{\theta}_n-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dx\right]^2 d\Pi(z).
\end{aligned}
$$

We can verify $B_{n1}=o_p(1)$. In fact, from (x2) and the consistency of $\hat{\theta}_n$,

$$
\begin{aligned}
EB_{n1} =& E\int\left[\int L_b(z-x)\frac{\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)-\hat{f}_{\bar{Z}}(x,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dx\right]^2 d\Pi(z)\\
=& E\int\Big[\int L_b(z-x)\int\frac{f_X(x-u,\hat{\theta}_n)-f_X(x-u,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{f}_X(x-u,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}\\
&\cdot\hat{f}_{\bar{U}}(u)dudx\Big]^2 d\Pi(z)\\
\leq& \left[\sup_x\frac{|f_X(x,\hat{\theta}_n)-f_X(x,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{f}_X(x-u,\theta_0)|}{\|\hat{\theta}_n-\theta_0\|}\right]^2\\
&\cdot E\int\left[\int L_b(z-x)\int\hat{f}_{\bar{U}}(u)dudx\right]^2 d\Pi(z)
\end{aligned}
$$

is of order $o_p(1)$ by observing that $E\int[\int L_b(z-x)\int\hat{f}_{\bar{U}}(u)dudx]^2 d\Pi(z)=O(1)$.

For B_{n2}, we notice that $B_{n2}\geq\inf_{\|s\|=1}\Sigma_n(s)$, where

$$
\Sigma_n(s)=\int\left[\int L_b(z-x)s^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)dx\right]^2 d\Pi(z)=\int[s^T\dot{\mu}_n(z,\theta_0)]^2 d\Pi(z).
$$

By Lemma 3.3.4, we have $\Sigma_n(s) = \int s^T \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T s d\Pi(z) = s^T\Sigma_0 s + o_p(1)$.

Also note that for any $\delta > 0$, and any two unit vector $s, s_1 \in \mathbb{R}^q$, $\|s - s_1\| < \delta$, we have

$$
\begin{aligned}
|\Sigma_n(s) - \Sigma_n(s_1)| = &\left| \int s^T \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T s d\Pi(z) - \int s_1^T \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T s_1 d\Pi(z) \right| \\
=&\Big| (s - s_1)^T \int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T d\Pi(z)(s - s_1) \\
&+ 2s_1^T \int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T d\Pi(z)(s - s_1)\Big| \\
\le&\delta(\delta + 2) \int \|\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z).
\end{aligned}
$$

By Lemma 3.3.4, we have $\int \|\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z) = O_p(1)$. This fact together with the compactness of the set $\{s \in \mathbb{R}^q; \|s\| = 1\}$ imply $\sup_{\|s\|=1} |\Sigma_n(s) - s^T\Sigma_0 s| = o_p(1)$. Therefore, $B_{n2} \ge \inf_{\|s\|=1} s^T\Sigma_0 s + o_p(1)$. We also have $\frac{1}{2}B_{n1}^{1/2}B_{n2}^{1/2} = o_p(1)$ by the Cauchy-Schwarz inequality. These facts imply

$$
\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n - \theta_0\|} \ge \inf_{\|s\|=1} s^T\Sigma_0 s \quad \text{with arbitrarily large probability.}
$$

This concludes the proof of Lemma 3.3.5. □

Lemma 3.3.6. *Suppose H_0, (z1)–(z4), (b1)–(b2), (x1), (x3) hold. Then*

$$
\int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\hat{\theta}_n))^T d\Pi(z) = \Sigma_0 + o_p(1).
$$

Proof. Note that

$$
\begin{aligned}
&\int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\hat{\theta}_n))^T d\Pi(z) \\
=&\int \dot{\mu}_n(z,\theta_0)[\dot{\mu}_n(z,\hat{\theta}_n) - \dot{\mu}_n(z,\theta_0)]^T d\Pi(z) + \int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T d\Pi(z).
\end{aligned}
$$

In view of Lemma 3.3.4, we only need to show the upper bound of $\int \dot{\mu}_n(z,\theta_0)[\dot{\mu}_n(z,\hat{\theta}_n) -$

$\dot{\mu}_n(z,\theta_0)]^T d\Pi(z)$ is $o_p(1)$. In fact, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\left\|\int \dot{\mu}_n(z,\theta_0)[\dot{\mu}_n(z,\hat{\theta}_n)-\dot{\mu}_n(z,\theta_0)]^T d\Pi(z)\right\|^2\\ \leq &\int \|\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z)\cdot \int \|\dot{\mu}_n(z,\hat{\theta}_n)-\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z).\end{aligned}$$

Note that $\int \|\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z) = O_p(1)$. Moreover, from (b1), (x3), and by the consistency of $\hat{\theta}_n$,

$$\begin{aligned}&\int \|\dot{\mu}_n(z,\hat{\theta}_n)-\dot{\mu}_n(z,\theta_0)\|^2 d\Pi(z)\\ =&\int \left\|\int L_b(z-x)\int (\dot{f}_X(x-u,\hat{\theta}_n)-\dot{f}_X(x-u,\theta_0))\hat{f}_{\bar{U}}(u)dudx\right\|^2 d\Pi(z)\\ \leq &\sup_x \|\dot{f}_X(x,\hat{\theta}_n)-\dot{f}_X(x,\theta_0)\|^2 \int \left[\int L_b(z-x)\int \hat{f}_{\bar{U}}(u)dudx\right]^2 d\Pi(z)\\ =&o_p(b)\int \left[\int \hat{f}_{\bar{U}}(u)du\right]^2 d\Pi(z) = o_p(b)O_p(1) = o_p(1). \qquad (3.38)\end{aligned}$$

Therefore,

$$\int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\hat{\theta}_n))^T d\Pi(z) = \int \dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T d\Pi(z) + o_p(1) = \Sigma_0 + o_p(1).$$

□

Lemma 3.3.7. *Suppose H_0, (b1)–(b2), (h3), (π1), (z1)–(z4), (x1), (x3), (kl), and (u1) hold, then*

$$\sqrt{n}\int U_n(z,\theta_0)\dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z) \to N_q(0,\Sigma),$$

where $\Sigma = \Sigma_1+\Sigma_2+\Sigma_3$ and $\Sigma_i, i = 1,2,3$ are as defined in (3.7).

Proof. For convenience, we shall give the proof here only for the case $d = 1$, i.e., when $\dot{\mu}_n(z,\theta)$ is one dimensional. For multidimensional case, the result can be proved by using linear combination of its components instead of $\dot{\mu}_n(z,\theta)$.

Note that $\sqrt{n}\int U_n(z)\dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z)$ can be written as the sum of the three terms:

$$
\begin{aligned}
g_{n1} =& \sqrt{n}\int U_n(z)[\dot{\mu}_n(z,\hat{\theta}_n)-\dot{\mu}_n(z,\theta_0)]d\Pi(z) \\
g_{n2} =& \sqrt{n}\int U_n(z)[\dot{\mu}_n(z,\theta_0)-\dot{\mu}_h(z)]d\Pi(z) \\
G_n =& \sqrt{n}\int U_n(z)\dot{\mu}_h(z)d\Pi(z)
\end{aligned}
$$

We are going to show the first two terms are $o_p(1)$ and the last one converges to $N(0,\Sigma)$.

We add and subtract $\int L_b(z-x)f_{\bar{Z}}(x,\theta_0)dx$ and $\int L_b(z-x)E\hat{f}_{\bar{Z}}(x,\theta_0)dx$ from $U_n(z)$, then G_n can be written as the sum of G_{n1}, G_{n2} and G_{n3}, where

$$
\begin{aligned}
G_{n1} =& \sqrt{n}\int\left[\hat{f}_{\bar{Z}}(z)-\int L_b(z-x)f_{\bar{Z}}(x,\theta_0)dx\right]\dot{\mu}_h(z)d\Pi(z) \\
G_{n2} =& \sqrt{n}\int\left[\int L_b(z-x)\int f_X(x-u,\theta_0)(f_{\bar{U}}(u)-E\hat{f}_{\bar{U}}(u))dudx\right]\dot{\mu}_h(z)d\Pi(z) \\
G_{n3} =& \sqrt{n}\int\left[\int L_b(z-x)\int f_X(x-u,\theta_0)(E\hat{f}_{\bar{U}}(u)-\hat{f}_{\bar{U}}(u))dudx\right]\dot{\mu}_h(z)d\Pi(z)
\end{aligned}
$$

It suffices to show $G_{n1}+G_{n3}\to N(0,\Sigma), G_{n2}=o_p(1)$. Let $s_{ni}=s_{ni1}+s_{ni2}$, where

$$
\begin{aligned}
s_{ni1} =& \int\left[L_b(z-\bar{Z}_i)-\int L_b(z-x)f_{\bar{Z}}(x,\theta_0)dx\right]\dot{\mu}_h(z)d\Pi(z), \\
s_{ni2} =& \int\left[\int L_b(z-x)\int f_X(x-u,\theta_0)(EK_h(u-\tilde{U}_1)-K_h(u-\tilde{U}_i))dudx\right]\dot{\mu}_h(z)d\Pi(z).
\end{aligned}
$$

Then We can rewrite $G_{n1}+G_{n3}=\frac{1}{\sqrt{n}}\sum_{i=1}^n(s_{ni1}+s_{ni2})=\frac{1}{\sqrt{n}}\sum_{i=1}^n s_{ni}$.

Note that for each n, s_{ni} are iid centered r.v.'s. To prove $G_{n1}+G_{n3}\to N(0,\Sigma)$, we only need to show

$$
\begin{aligned}
& Es_{ni1}^2\to\Sigma_1,\quad Es_{ni2}^2\to\Sigma_2,\quad 2Es_{ni1}s_{ni2}\to\Sigma_3, \\
& E\{s_{ni}^2 I(|s_{ni}|>n^{\frac{1}{2}}\eta)\}\to 0,\quad \forall\eta>0. \qquad (3.39)
\end{aligned}
$$

By Fubini,

$$
\begin{aligned}
Es_{ni1}^2 =& \iint [EL_b(z-\bar{Z})L_b(y-\bar{Z}) - EL_b(z-\bar{Z})EL_b(y-\bar{Z})]\dot{\mu}_h(z)\dot{\mu}_h(y)d\Pi(z)d\Pi(y) \\
=& \int L_b(z-s)L_b(y-s)f_{\bar{Z}}(s,\theta_0)\dot{\mu}_h(z)\dot{\mu}_h(y)\pi(z)\pi(y)\,\mathrm{d}(s,z,y) \\
& - \left[\iint L_b(z-s)f_{\bar{Z}}(s,\theta_0)\dot{\mu}_h(z)\pi(z)dsdz\right]^2.
\end{aligned}
$$

By using the result in (3.36), and the assumed continuity of L and π, one obtains

$$
\begin{aligned}
\lim_{b\to 0} Es_{ni1}^2 =& \lim_{b\to 0}\int L(z)L(y)f_{\bar{Z}}(s,\theta_0)\dot{\mu}_h(s+bz)\dot{\mu}_h(s+by)\pi(s+bz)\pi(s+by)\,\mathrm{d}(s,z,y) \\
& - \left[\iint L(z)f_{\bar{Z}}(s,\theta_0)\dot{\mu}_h(s+bz)\pi(s+bz)dsdz\right]^2 \\
=& \int L(z)L(y)f_{\bar{Z}}(s,\theta_0)\left(\int \dot{f}_X(s-u,\theta_0)f_{\bar{U}}(u)du\right)^2\pi^2(s)\,\mathrm{d}(s,z,y) \\
& - \left[\iint L(z)f_{\bar{Z}}(s,\theta_0)\left(\int \dot{f}_X(s-u,\theta_0)f_{\bar{U}}(u)du\right)\pi(s)dsdz\right]^2.
\end{aligned}
$$

By simple algebra, we see this is indeed Σ_1. Denote

$$
H_{\tilde{U}_i}(z,\theta_0) = \int L_b(z-x)\int f_X(x-u,\theta_0)K_h(u-\tilde{U}_i)dudx. \tag{3.40}
$$

Then

$$
\begin{aligned}
Es_{ni2}^2 =& E\left[\int [EH_{\tilde{U}}(z,\theta_0) - H_{\tilde{U}_i}(z,\theta_0)]\dot{\mu}_h(z)d\Pi(z)\right]^2 \\
=& E\left[E\left(\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)\right) - \int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)\right]^2
\end{aligned}
$$

which equals the variance of $\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)$, and can be written as the difference between $E[\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)]^2$ and $[E\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)]^2$. We then calculate these two terms one by one. By Fubini, one obtains that

$$
\begin{aligned}
& E\left[\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z)\right]^2 \\
=& E\iint H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)H_{\tilde{U}}(y,\theta_0)\dot{\mu}_h(y)d\Pi(z)d\Pi(y) \\
=& \iint E(H_{\tilde{U}}(z,\theta_0)H_{\tilde{U}}(y,\theta_0))\dot{\mu}_h(z)\dot{\mu}_h(y)d\Pi(z)d\Pi(y) \\
=& \iint \Big[\int L_b(z-x)f_X(x-u,\theta_0)K_h(u-s)L_b(y-t)f_X(t-v,\theta_0) \\
& \cdot K_h(v-s)f_{\bar{U}}(s)\,\mathrm{d}(u,x,v,t,s)\Big]\dot{\mu}_h(z)\dot{\mu}_h(y)\pi(z)\pi(y)dzdy
\end{aligned}
$$

which converges to $\int\left[\int f_X(x-s,\theta_0)\dot{f}_{\bar{Z}}(x,\theta_0)\pi(x)dx\right]^2 f_{\bar{U}}(s)ds$ as $b\to 0, h\to 0$ by simple algebra. And

$$
\begin{aligned}
& E\int H_{\tilde{U}}(z,\theta_0)\dot{\mu}_h(z)d\Pi(z) \\
=& \int\left[\int L_b(z-x)f_X(x-u,\theta_0)K_h(u-s)f_{\bar{U}}(s)\,\mathrm{d}(u,x,s)\right]\dot{\mu}_h(z)d\Pi(z) \\
=& \int L(z_1)f_X(x-s-hu_1,\theta_0)K_h(u_1)f_{\bar{U}}(s)\dot{\mu}_h(x+bz_1)\pi(x+bz_1)\,\mathrm{d}(u_1,x,s,z_1)
\end{aligned}
$$

converges to $\int\left[\int f_X(x-s,\theta_0)\dot{f}_{\bar{Z}}(x,\theta_0)\pi(x)dx\right]f_{\bar{U}}(s)ds$. Therefore, Es_{ni2}^2 converges to

$$
\int\left[\int f_X(x-s,\theta_0)\dot{f}_{\bar{Z}}(x,\theta_0)d\Pi(x)\right]^2 f_{\bar{U}}(s)ds-\left[\int f_{\bar{Z}}(x,\theta_0)\dot{f}_{\bar{Z}}(x,\theta_0)d\Pi(x)\right]^2=\Sigma_2.
$$

Next, we consider $Es_{ni1}s_{ni2}$. Note that $Es_{ni1}s_{ni2}$ can be rewritten as

$$
\begin{aligned}
& Es_{ni1}s_{ni2} \\
=& E\int[L_b(z-\bar{Z}_i)-EL_b(z-\bar{Z}_1)]\dot{\mu}_h(z)d\Pi(z)\left[\int[EH_{\tilde{U}}(z,\theta_0)-H_{\tilde{U}_i}(z,\theta_0)]\dot{\mu}_h(z)d\Pi(z)\right] \\
=& E\iint(L_b(z-\bar{Z}_i)-EL_b(z-\bar{Z}_1))\dot{\mu}_h(z)(EH_{\tilde{U}_1}(y,\theta_0)-H_{\tilde{U}_i}(y,\theta_0))\dot{\mu}_h(y)d\Pi(z)d\Pi(y) \\
=& \iint[EL_b(z-\bar{Z}_1)EH_{\tilde{U}_1}(y,\theta_0)-EL_b(z-\bar{Z}_i)H_{\tilde{U}_i}(y,\theta_0)]\dot{\mu}_h(z)\dot{\mu}_h(y)d\Pi(z)d\Pi(y).
\end{aligned}
$$

Note that

$$
\begin{aligned}
EL_b(z-\bar{Z}_1) &= \int L(t) f_{\bar{Z}}(z-bt,\theta_0)dt,\\
EH_{\tilde{U}_1}(y,\theta_0) &= \int L(x) f_X(y-bx-u,\theta_0)K(s)f_{\bar{U}}(u-hs)\,\mathrm{d}(u,x,s).
\end{aligned}
$$

By the transformation $z-s-(p+q)/2=bv$ and $u-(p-q)/2=ht$, we obtain

$$
\begin{aligned}
& EL_b(z-\bar{Z}_i)H_{\tilde{U}_i}(y,\theta_0)\\
=& E\left[L_b\left(z-X_i-\frac{U_1+U_2}{2}\right)\int L_b(y-x)\int f_X(x-u,\theta_0)K_h\left(u-\frac{U_1-U_2}{2}\right)dudx\right]\\
=&\int L_b\left(z-s-\frac{p+q}{2}\right)L_b(y-x)f_X(x-u,\theta_0)K_h\left(u-\frac{p-q}{2}\right)f_X(s,\theta_0)f_U(p)f_U(q)\,\mathrm{d}(u,x,s,p,q)\\
=&2\int L(v)L(x)f_X(y-bx-u,\theta_0)K(t)f_U(z-s-bv+u-ht)\\
&\cdot f_U(z-s-bv-u+ht)f_X(s,\theta_0)\,\mathrm{d}(u,x,s,v,t)
\end{aligned}
$$

Therefore, when $b\to 0, h\to 0$, by using the continuity of $f_{\bar{Z}}, f_X, f_{\bar{U}}$ and f_U, we have $2Es_{ni1}s_{ni2}$ converges to

$$
\begin{aligned}
&\left(\int L(t)f_{\bar{Z}}(z,\theta_0)dt\right)\left(\int L(x)f_X(y-u,\theta_0)K(s)f_{\bar{U}}(u)\,\mathrm{d}(u,x,s)\right)\\
&\cdot \dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y)-2\int L(v)L(x)f_X(y-u,\theta_0)K(t)f_X(s,\theta_0)\\
&\cdot f_U(z-s+u)f_U(z-s-u)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)\,\mathrm{d}(u,x,s,v,t,z,y)\\
=&2\iint f_{\bar{Z}}(z,\theta_0)f_{\bar{Z}}(y,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)d\Pi(z)d\Pi(y)-2\int f_X(y-u,\theta_0)\\
&\cdot f_U(z-s+u)f_U(z-s-u)f_X(s,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)\pi(z)\pi(y)\,\mathrm{d}(u,s,z,y)\\
=&2\left[\int f_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)d\Pi(z)\right]^2-2\iiint\Big(\int f_X(y-u,\theta_0)f_U(z-s+u)\\
&\cdot f_U(z-s-u)du\Big)f_X(s,\theta_0)\dot{f}_{\bar{Z}}(z,\theta_0)\dot{f}_{\bar{Z}}(y,\theta_0)dsd\Pi(z)d\Pi(y)
\end{aligned}
$$

which is indeed Σ_3 by simple algebra. Next, we are going to prove (3.39). Note that

$$E\{s_{ni}^2 I(|s_{ni}| > n^{\frac{1}{2}}\eta)\} = E\left\{s_{ni}^2 I\left(\frac{|s_{ni}|}{n^{\frac{1}{2}}\eta} > 1\right)\right\} \leq E\left\{\frac{s_{ni}^2\ |s_{ni}|^\delta}{(n^{\frac{1}{2}}\eta)^\delta}\right\} = \eta^{-\delta} n^{-\delta/2} E|s_{ni}|^{2+\delta}.$$

By using C_r inequality, we have $E|s_{ni}|^{2+\delta} \leq 2^{1+\delta} E(|s_{ni1}|^{2+\delta} + |s_{ni2}|^{2+\delta})$. Using Hölder's inequality, $E(|s_{ni1}|^{2+\delta})$ is bounded above by

$$\begin{aligned}
&E\left|\left\{\int [L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z})]^{\frac{2+\delta}{2}} d\Pi(z)\right\}^{\frac{2}{2+\delta}} \cdot \left\{\int (\dot{\mu}_h(z))^{\frac{2+\delta}{\delta}} d\Pi(z)\right\}^{\frac{\delta}{2+\delta}}\right|^{2+\delta} \\
=&E\left[\int [L_b(z-\bar{Z}_i) - EL_b(z-\bar{Z})]^{\frac{2+\delta}{2}} d\Pi(z)\right]^2 \cdot \left[\int (\dot{\mu}_h(z))^{\frac{2+\delta}{\delta}} d\Pi(z)\right]^\delta = O(b^{-\delta}),
\end{aligned}$$

and it is not hard to see

$$\begin{aligned}
E|s_{ni2}|^{2+\delta} =&E\left|\int (EH_{\tilde{U}_i}(z,\theta_0) - H_{\tilde{U}_i}(z,\theta_0))\dot{\mu}_h(z) d\Pi(z)\right|^{2+\delta} \\
=&E\left|\int H_{\tilde{U}_i}(z,\theta_0)\dot{\mu}_h(z) d\Pi(z) - \int EH_{\tilde{U}_i}(z,\theta_0)\dot{\mu}_h(z) d\Pi(z)\right|^{2+\delta} = O(1).
\end{aligned}$$

Then

$$\eta^{-\delta} n^{-\delta/2} E|s_{ni}|^{2+\delta} \leq \eta^{-\delta} 2^{1+\delta} n^{-\delta/2} E(|s_{ni1}|^{2+\delta} + |s_{ni2}|^{2+\delta}) \leq O(n^{-\delta/2} b^{-\delta}) + O(n^{-\delta/2})$$

is of the order $o_p(1)$. By the L-F C.L.T., we have $G_{n1} + G_{n3} \to N(0, \Sigma)$, where $\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_3$.

To finish the proof of Lemma 3.3.7, it suffices to show $g_{n1} = o_p(1)$, $g_{n2} = o_p(1)$ and $G_{n2} = o_p(1)$. In fact, from (3.37) and (3.38), by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
g_{n1} =&\sqrt{n}\int U_n(z)[\dot{\mu}_n(z,\hat{\theta}_n) - \dot{\mu}_n(z,\theta_0)] d\Pi(z) \\
\leq&\sqrt{n}\left[\int U_n^2(z) d\Pi(z)\right]^{\frac{1}{2}} \left\{\int [\dot{\mu}_n(z,\hat{\theta}_n) - \dot{\mu}_n(z,\theta_0)]^2 d\Pi(z)\right\}^{\frac{1}{2}} = \sqrt{n} O_p\left(\frac{1}{\sqrt{nb}}\right) o_p(b^{1/2})
\end{aligned} \tag{3.41}$$

which is of order $o_p(1)$. From (b2), (3.34) and (3.37), we obtain

$$\begin{aligned}
g_{n2} =& \sqrt{n} \int U_n(z)[\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)] d\Pi(z) \\
\leq & \sqrt{n} \left[\int U_n^2(z) d\Pi(z) \right]^{\frac{1}{2}} \left[\int [\dot{\mu}_n(z, \theta_0) - \dot{\mu}_h(z)]^2 d\Pi(z) \right]^{\frac{1}{2}} \\
=& \sqrt{n} O_p\left(\frac{1}{\sqrt{nb}} \right) O_p\left(\frac{1}{\sqrt{n}} \right) = O_p\left(\frac{1}{\sqrt{nb}} \right) = o_p(1).
\end{aligned}$$

From (u1) and (h3), we have

$$\begin{aligned}
G_{n2} =& \sqrt{n} \int \left[\int L_b(z-x) \int f_X(x-u, \theta_0)(f_{\bar{U}}(u) - E\hat{f}_{\bar{U}}(u)) du dx \right] \dot{\mu}_h(z) d\Pi(z) \\
=& \sqrt{n} \int \left[\int L_b(z-x) \int f_X(x-u, \theta_0) \frac{h^2}{2} \int K(v) v^2 f''_{\bar{U}}(u + \theta v h) dv du dx \right] \dot{\mu}_h(z) d\Pi(z) \\
=& O(\sqrt{n} h^2) = o_p(1),
\end{aligned}$$

where $0 < \theta < 1$. This completes the proof of Lemma 3.3.7. □

Now we are ready to prove the asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Let

$$G_n = \int U_n(z) \dot{\mu}_h(z) d\Pi(z).$$

Theorem 3.3.4. *Assume H_0, (x1)–(x3), (z1)–(z4), (u1), (b1)–(b2), (h1), (h3), and (π1) hold. Then $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) = \Sigma_0^{-1} n^{\frac{1}{2}} G_n + o_p(1)$. Consequently, $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0) \Rightarrow N_q(0, \Sigma_0^{-1} \Sigma \Sigma_0^{-1})$, where $\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_3$, and $\Sigma_i, i = 0, 1, 2, 3$ are as in (3.7).*

Proof. Recall that $\hat{\theta}_n$ is the minimizer of $T_n(\theta)$. By the consistency of $\hat{\theta}_n$, for sufficiently large n, $\hat{\theta}_n$ will be in the interior of Θ and $\dot{T}_n(\hat{\theta}_n) = 0$. Recall definition $\dot{\mu}_n(z, \theta)$ in (3.32), $\dot{T}_n(\hat{\theta}_n)$ can be written as

$$\begin{aligned}
\dot{T}_n(\hat{\theta}_n) =& -2 \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) \hat{f}_{\bar{Z}}(x, \hat{\theta}_n) dx \right] \left(\int L_b(z-x) \dot{\hat{f}}_{\bar{Z}}(x, \hat{\theta}_n) dx \right) d\Pi(z) \\
=& -2 \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x) \hat{f}_{\bar{Z}}(x, \hat{\theta}_n) dx \right] \dot{\mu}_n(z, \hat{\theta}_n) d\Pi(z).
\end{aligned}$$

Therefore,

$$\int \hat{f}_{\bar{Z}}(z)\ \dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z) = \int \left[\int L_b(z-x)\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)dx\ \dot{\mu}_n(z,\hat{\theta}_n)\right] d\Pi(z). \tag{3.42}$$

Adding and subtracting $\int\int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_0)dx\ \dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z)$ from the R.H.S. of (3.42), recalling the definition of $U_n(z,\theta)$ in (3.32), one obtains

$$\begin{aligned}&\int\int L_b(z-x)(\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)-\hat{f}_{\bar{Z}}(x,\theta_0))dx\ \dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z)\\ =&\int\left[\hat{f}_{\bar{Z}}(z)-\int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_0)dx\right]\ \dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z) =: \int U_n(z,\theta_0)\ \dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z).\end{aligned}$$

Recall the definition of $d_n(x,\theta,\theta_0) = \hat{f}_{\bar{Z}}(x,\theta) - \hat{f}_{\bar{Z}}(x,\theta_0) - (\theta-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)$, the term $\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)-\hat{f}_{\bar{Z}}(x,\theta_0)$ can be written as $d_n(x,\hat{\theta}_n,\theta_0)+(\hat{\theta}_n-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)$. Thus we have

$$\begin{aligned}&\Big\{\int \dot{\mu}_n(z,\hat{\theta}_n)\int L_b(z-x)\frac{d_n(x,\hat{\theta}_n,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dxd\Pi(z)\frac{(\hat{\theta}_n-\theta_0)^T}{\|\hat{\theta}_n-\theta_0\|}\\ &+\int \dot{\mu}_n(z,\hat{\theta}_n)(\dot{\mu}_n(z,\theta_0))^Td\Pi(z)\Big\}\sqrt{n}(\hat{\theta}_n-\theta_0)\\ =&\sqrt{n}\Big\{\int U_n(z,\theta_0)\dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z)\Big\}.\end{aligned}$$

From (x2) and the consistency of $\hat{\theta}_n$, $\|\int \dot{\mu}_n(z,\hat{\theta}_n)[\int L_b(z-x)\frac{d_n(x,\hat{\theta}_n,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dx]d\Pi(z)\ \frac{(\hat{\theta}_n-\theta_0)^T}{\|\hat{\theta}_n-\theta_0\|}\|$ is bounded above by

$$\begin{aligned}&\sup_x \frac{|f_X(x,\hat{\theta}_n)-f_X(x,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{f}_X(x,\theta_0)|}{\|\hat{\theta}_n-\theta_0\|}\int\|\dot{\mu}_n(z,\hat{\theta}_n)\|\\ &\cdot\left[\int L_b(z-x)\int f_{\bar{U}}(u)dudx\right]d\Pi(z) = o_p(1),\end{aligned}$$

which is due to

$$\int\|\dot{\mu}_n(z,\hat{\theta}_n)\|d\Pi(z)\le\int\|\dot{\mu}_n(z,\hat{\theta}_n)-\dot{\mu}_n(z,\theta_0)\|d\Pi(z)+\int\|\dot{\mu}_n(z,\theta_0)\|d\Pi(z) = O_p(1).$$

Then the result of Theorem 3.3.4 is a consequence of Lemma 3.3.4 to 3.3.7. $\square$

Remark 3.3.2. *For the bandwidth assumptions, Theorem 3.3.3 requires both $nb^4 \to 0$ and $nh^4 \to 0$, while Theorem 3.3.4 only requires $nh^4 \to 0$. The condition $nb^4 \to 0$ is needed to deal with the asymptotic bias $E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)$, while the condition $nh^4 \to 0$ is required to dampen the effect of estimating $f_{\bar{U}}$ by its kernel density estimate $\hat{f}_{\bar{U}}(u)$.*

3.4 Asymptotic Distribution of the MD Test Statistic

This section contains the proofs of the asymptotic normality of the minimized distance $T_n^*(\theta_n^*)$ and $T_n(\hat{\theta}_n)$. We begin this section with a lemma, which will be used in the subsequent proofs.

Lemma 3.4.1. *Let $X_i, i = 1, 2, \ldots, n$ be a sequence of i.i.d. random variables, $f(z, \cdot)$ and $g(z, \cdot)$ be two measurable functions. Suppose $\int Ef(z, X)d\Pi(z) < \infty$, $\int Eg(z, X)d\Pi(z) < \infty$ and $\int Ef(z, X)g(z, X)d\Pi(z) < \infty$, then*

$$\int \left[\frac{1}{n}\sum_{i=1}^{n}(f(z, X_i) - Ef(z, X))\right]\left[\frac{1}{n}\sum_{i=1}^{n}(g(z, X_i) - Eg(z, X))\right] d\Pi(z) = O_p\left(\frac{1}{n}\right).$$

Proof. First, for $i \neq j$, $E\int[f(z, X_i) - Ef(z, X)][g(z, X_j) - Eg(z, X)]d\Pi(z) = 0$ due to the independence of X_i and X_j. Therefore,

$$\begin{aligned}
&E\int \left[\frac{1}{n}\sum_{i=1}^{n}(f(z, X_i) - Ef(z, X))\right]\left[\frac{1}{n}\sum_{i=1}^{n}(g(z, X_i) - Eg(z, X))\right] d\Pi(z)\\
=&\frac{1}{n^2}\sum_{i=1}^{n} E\int [f(z, X_i) - Ef(z, X)][g(z, X_i) - Eg(z, X)]d\Pi(z)\\
=&\frac{1}{n}\int [Ef(z, X)g(z, X) - Ef(z, X)Eg(z, X)]d\Pi(z) = O\left(\frac{1}{n}\right).
\end{aligned}$$

Moreover,

$$
\begin{aligned}
&E\left\{\int\left[\frac{1}{n}\sum_{i=1}^{n}(f(z,X_i)-Ef(z,X))\right]\left[\frac{1}{n}\sum_{i=1}^{n}(g(z,X_i)-Eg(z,X))\right]d\Pi(z)\right\}^2\\
=&\frac{1}{n^4}\sum_{i,j}E\left[\int(f(z,X_i)-Ef(z,X))(g(z,X_i)-Eg(z,X))d\Pi(z)\right]^2\\
&+\frac{n(n-1)}{n^4}E\Big[\int(f(z,X_1)-Ef(z,X))(g(z,X_1)-Eg(z,X))d\Pi(z)\\
&\cdot\int(f(z,X_2)-Ef(z,X))(g(z,X_2)-Eg(z,Z))d\Pi(z)\Big]\\
&+\frac{n(n-1)}{n^4}E\Big[\int(f(z,X_1)-Ef(z,X))(g(z,X_2)-Eg(z,X))d\Pi(z)\\
&\cdot\int(f(z,X_2)-Ef(z,X))(g(z,X_1)-Eg(z,Z))d\Pi(z)\Big]=O\left(\frac{1}{n^2}\right).
\end{aligned}
$$

Hence the desired result. □

The following theorem states the asymptotic distribution of the minimum distance test statistic based on $T_n^*(\theta_n^*)$.

Theorem 3.4.1. *Suppose H_0, (x1)–(x3), (z1)–(z4), (b1)–(b3), (h1)–(h3), (π1), (kl), and (u1). Then*

$$nb^{1/2}(T_n^*(\theta_n^*)-\hat{C}_n(\theta_n^*))\Rightarrow N(0,\Gamma),$$

where $\hat{C}_n(\theta_n^*)$ and Γ are as defined in (3.7) and

$$\hat{\Gamma}_n=\frac{2b}{n^2}\sum_{i\neq j}\left(\int\big[L_b(z-\bar{Z}_i)-\frac{1}{n}\sum_{k=1}^{n}L_b(z-\bar{Z}_k)\big]\big[L_b(z-\bar{Z}_j)-\frac{1}{n}\sum_{k=1}^{n}L_b(z-\bar{Z}_k)\big]d\Pi(z)\right)^2$$

Moreover, $|\hat{\Gamma}_n\Gamma^{-1}-1|=o_p(1)$.

Define

$$\mathcal{T}_n^*(\theta_n^*)=\hat{\Gamma}_n^{-1/2}nb^{1/2}(T_n^*(\theta_n^*)-\hat{C}_n(\theta_n^*)).\tag{3.43}$$

Consequently, H_0 will be rejected whenever $|\mathcal{T}_n^*(\theta_n^*)|>Z_{\alpha/2}$, where α is the asymptotic size and Z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.

The proof of this theorem is facilitated by the following three lemmas. Define

$$\tilde{T}_n(\theta_0) = \int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) \tag{3.44}$$

Lemma 3.4.2. *Suppose (b1)–(b2), (z1) and (π1) hold. Then* $nb^{1/2}(\tilde{T}_n(\theta_0) - C_n) \Rightarrow N(0, \Gamma)$.

To prove Lemma 3.4.2, we need the following Theorem 3.4.2, which is Theorem 1 of Hall (1984) and reproduced here for the sake of completeness.

Theorem 3.4.2. *Let $\tilde{X}_i, 1 \le i \le n$, be i.i.d. random vectors, and let*

$$U_n = \sum_{1 \le i < j \le n} H_n(\tilde{X}_i, \tilde{X}_j), \quad G_n(x, y) = EH_n(\tilde{X}_1, x)H_n(\tilde{X}_1, y),$$

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

$$EH_n(\tilde{X}_1, \tilde{X}_2 | \tilde{X}_1) = 0, \text{ almost surely, and } EH_n^2(\tilde{X}_1, \tilde{X}_2) < \infty, \text{ for each } n \ge 1.$$

If

$$[EG_n^2(\tilde{X}_1, \tilde{X}_2) + n^{-1}EH_n^4(\tilde{X}_1, \tilde{X}_2)]/[EH_n^2(\tilde{X}_1, \tilde{X}_2)]^2 \to 0,$$

then U_n is asymptotically normally distributed with mean zero and variance $n^2 EH_n^2(\tilde{X}_1, \tilde{X}_2)/2$.

Now let's prove Lemma 3.4.2.

Proof. Expanding the square, $\tilde{T}_n(\theta_0)$ can be written as the sum of the following two terms:

$$\begin{aligned}
C_n =& \frac{1}{n^2} \sum_{i=1}^n \int \left(L_b(z - \bar{Z}_i) - \int L_b(z - x) f_{\bar{Z}}(x, \theta_0) dx \right)^2 d\Pi(z), \\
M_n =& \frac{1}{n^2} \sum_{i \ne j} \int \left(L_b(z - \bar{Z}_i) - \int L_b(z - x) f_{\bar{Z}}(x, \theta_0) dx \right) \\
& \cdot \left(L_b(z - \bar{Z}_j) - \int L_b(z - x) f_{\bar{Z}}(x, \theta_0) dx \right) d\Pi(z).
\end{aligned}$$

Let

$$H_n(\bar{Z}_i, \bar{Z}_j) = \frac{1}{n} b^{1/2} \int [L_b(z - \bar{Z}_i) - EL_b(z - \bar{Z}_i)][L_b(z - \bar{Z}_j) - EL_b(z - \bar{Z}_j)] d\Pi(z), \quad (3.45)$$

$$G_n(p, q) = EH_n(\bar{Z}_1, p) H_n(\bar{Z}_1, q). \quad (3.46)$$

One can easily show that the relation between $H_n(\bar{Z}_i, \bar{Z}_j)$ and M_n can be built as

$$\sum_{1 \le i \le j \le n} H_n(\bar{Z}_i, \bar{Z}_j) = \frac{1}{2} n b^{1/2} M_n.$$

Observe that $H_n(\bar{Z}_1, \bar{Z}_2)$ is symmetric, we have $E(H_n(\bar{Z}_1, \bar{Z}_2)|\bar{Z}_1) = 0$. Applying Theorem 3.4.2, in order to show $nb^{1/2} M_n \to N(0, \Gamma)$, we need to further prove the following two results:

$$EH_n^2(\bar{Z}_1, \bar{Z}_2) < \infty, \quad \text{for any } n. \quad (3.47)$$

$$\frac{EG_n^2(\bar{Z}_1, \bar{Z}_2) + n^{-1} EH_n^4(\bar{Z}_1, \bar{Z}_2)}{[EH_n^2(\bar{Z}_1, \bar{Z}_2)]^2} \to 0. \quad (3.48)$$

To prove (3.47), observe that for each $n \ge 1$,

$$\begin{aligned}
&EH_n^2(\bar{Z}_1, \bar{Z}_2) \\
=& n^{-2} b E\Big[\int (L_b(z - \bar{Z}_1) - EL_b(z - \bar{Z}_1))(L_b(z - \bar{Z}_2) - EL_b(z - \bar{Z}_2)) d\Pi(z) \\
& \cdot \int (L_b(y - \bar{Z}_1) - EL_b(y - \bar{Z}_1))(L_b(y - \bar{Z}_2) - EL_b(y - \bar{Z}_2)) d\Pi(y) \Big] \\
=& n^{-2} b \iint \Big\{ E[L_b(z - \bar{Z}_1) L_b(y - \bar{Z}_1)] - EL_b(z - \bar{Z}_1) EL_b(y - \bar{Z}_1) \Big\}^2 d\Pi(z) d\Pi(y).
\end{aligned}$$

By changing variable, $\frac{z-s}{b} = t$, we have

$$E[L_b(z - \bar{Z}_1) L_b(y - \bar{Z}_1)] = \frac{1}{b} \int L(t) L\left(\frac{y - z}{b} + t\right) f_{\bar{Z}}(z - bt, \theta_0) dt,$$

$$EL_b(z - \bar{Z}_1) = \int L(t) f_{\bar{Z}}(z - bt, \theta_0) dt, \quad EL_b(y - \bar{Z}_1) = \int L(t) f_{\bar{Z}}(y - bt, \theta_0) dt.$$

Usual calculation shows that $EH_n^2(\bar{Z}_1, \bar{Z}_2)$ equals

$$n^{-2}\iint\Big[\int L(t)L(z_1+t)f_{\bar{Z}}(y-bz_1-bt,\theta_0)dt - b\int L(t)f_{\bar{Z}}(y-bz_1-bt,\theta_0)dt$$
$$\int L(t)f_{\bar{Z}}(y-bt,\theta_0)dt\Big]^2\cdot\pi(y-bz_1)\pi(y)\mathrm{d}z_1dy =: n^{-2}\kappa_n, \tag{3.49}$$

where $\kappa_n \to \Gamma/2$ as $n \to \infty$. Next, consider (3.48). Similar to the argument above, one obtains

$$\begin{aligned}
&EH_n^4(\bar{Z}_1, \bar{Z}_2)\\
=&n^{-4}b^2E\left[\int(L_b(z-\bar{Z}_1)-EL_b(z-\bar{Z}_1))(L_b(z-\bar{Z}_2)-EL_b(z-\bar{Z}_2))d\Pi(z)\right]^4\\
=&n^{-4}b^2\int[E(L_b(z-\bar{Z}_1)-EL_b(z-\bar{Z}_1))(L_b(y-\bar{Z}_1)-EL_b(y-\bar{Z}_1))\\
&\cdot(L_b(s-\bar{Z}_1)-EL_b(s-\bar{Z}_1))(L_b(t-\bar{Z}_1)-EL_b(t-\bar{Z}_1))]\,\mathrm{d}(\Pi(z),\Pi(y),\Pi(s),\Pi(t))\\
=&n^{-4}b^2\int\Big[\frac{1}{b^3}\int L(v)L\left(\frac{y-z}{b}+v\right)L\left(\frac{s-z}{b}+v\right)L\left(\frac{t-z}{b}+v\right)f_{\bar{Z}}(z-vb,\theta_0)\\
&+o(1/b^3)\Big]^2\mathrm{d}(\Pi(z),\Pi(y),\Pi(s),\Pi(t)) = O(n^{-4}b^2(1/b^3)^2b^3) = O(n^{-4}b^{-1}).
\end{aligned}\tag{3.50}$$

Recall $G_n(p,q)$ defined in (3.46). For $p,q\in\mathbb{R}$,

$$\begin{aligned}
G_n(p,q) =&EH_n(\bar{Z}_1,p)H_n(\bar{Z}_1,q)\\
=&n^{-2}b\iint[L_b(z-p)-EL_b(z-p)][L_b(y-q)-EL_b(y-q)]\\
&\cdot E[L_b(z-\bar{Z}_1)-EL_b(z-\bar{Z}_1)][L_b(y-\bar{Z}_1)-EL_b(y-\bar{Z}_1)]d\Pi(z)d\Pi(y).
\end{aligned}$$

Let

$$\begin{aligned}
B_b(y-z) =&E[(L_b(z-\bar{Z}_1)-EL_b(z-\bar{Z}_1))(L_b(y-\bar{Z}_1)-EL_b(y-\bar{Z}_1))]\\
=&\frac{1}{b}\int L(t)L(\frac{y-z}{b}+t)f_{\bar{Z}}(z-bt,\theta_0)dt\\
&-\int L(t)f_{\bar{Z}}(z-bt,\theta_0)dt\int L(t)f_{\bar{Z}}(y-bt,\theta_0)dt.
\end{aligned}$$

Then by expanding the square of integrals and changing the variables, we obtain

$$
\begin{aligned}
EG_n^2(\bar{Z}_1, \bar{Z}_2) =& n^{-4}b^2 \int B_b(v-z)B_b(w-y)B_b(y-z)B_b(w-v) \\
& \cdot \mathrm{d}(\Pi(z), \Pi(y), \Pi(v), \Pi(w)) \\
=& O(n^{-4}b^2(1/b)^4b^3) = O(n^{-4}b).
\end{aligned}
$$

By observing the facts below,

$$
\begin{aligned}
\frac{EG_n^2(\bar{Z}_1, \bar{Z}_2)}{[EH_n^2(\bar{Z}_1, \bar{Z}_2)]^2} =& \frac{O(n^{-4}b)}{n^{-4}\kappa_n^2} = o(1), \\
\frac{n^{-1}EH_n^4(\bar{Z}_1, \bar{Z}_2)}{[EH_n^2(\bar{Z}_1, \bar{Z}_2)]^2} =& \frac{O(n^{-5}b^{-1})}{n^{-4}\kappa_n^2} = O\left(\frac{1}{nb}\right) = o(1).
\end{aligned}
$$

we have shown (3.48) holds.

Define

$$
\begin{aligned}
\Gamma_n =& 2b\frac{n-1}{n} \iint \left\{E[L_b(z-\bar{Z}_1)L_b(y-\bar{Z}_1)] - EL_b(z-\bar{Z}_1)EL_b(y-\bar{Z}_1)\right\}^2 d\Pi(z)d\Pi(y) \\
=& 2n(n-1)EH_n^2(\bar{Z}_1, \bar{Z}_2).
\end{aligned}
$$

From (3.49), and by using the continuity of $f_{\bar{Z}}$, π, one obtains

$$
\begin{aligned}
\frac{\Gamma_n}{4} =& \frac{n(n-1)}{2}EH_n^2(\bar{Z}_1, \bar{Z}_2) \\
=& \frac{1}{2}n(n-1)n^{-2}\iiint \Big[\int L(t)L(z+t)f_{\bar{Z}}(y-bz-bt, \theta_0)dt \\
& - b\int L(t)f_{\bar{Z}}(y-bz-bt, \theta_0)dt \int L(t)f_{\bar{Z}}(y-bt, \theta_0)dt\Big]^2 \pi(y-bz)\pi(y)dzdy \\
\stackrel{b\to 0}{\to}& \frac{1}{2}\iint \left[\int L(t)L(z+t)f_{\bar{Z}}(y, \theta_0)dt\right]^2 \pi^2(y)dzdy \\
=& \frac{1}{2}\iint \left[f_{\bar{Z}}(y, \theta_0)\int L(t)L(z+t)dt\right]^2 \pi^2(y)dzdy \\
=& \frac{1}{2}\iint f_{\bar{Z}}^2(y, \theta_0)\pi^2(y)dy\left(\int L(t)L(z+t)dt\right)^2 dz =: \frac{\Gamma}{4}.
\end{aligned}
$$

In sum, $nb^{1/2}(\tilde{T}_n(\theta_0) - C_n) \Rightarrow N(0, \Gamma)$. $\square$

Lemma 3.4.3. *Under H_0, when (x1)–(x3), (z1)–(z4), (h1)–(h3), (b1)–(b3), (π1), (kl), and (u1) hold, we have $nb^{1/2}(T_n^*(\theta_n^*) - T_n^*(\theta_0)) = o_p(1)$.*

Proof. Recall the definition of $U_n^*(z)$ and $Z_n^*(z, \theta)$ from (3.19),

$$T_n^*(\theta_0) - T_n^*(\theta_n^*) = 2\int U_n^*(z) Z_n^*(z, \theta_n^*) d\Pi(z) - \int [Z_n^*(z, \theta_n^*)]^2 d\Pi(z) =: 2Q_1 - Q_2.$$

Thus, it suffices to show that $nb^{1/2}Q_1 = o_p(1)$ and $nb^{1/2}Q_2 = o_p(1)$.

Recalling $d_n(z, \theta, \theta_0)$ as defined in (3.19), we have

$$\begin{aligned} Q_1 &= \int U_n^*(z) Z_n^*(z, \theta_n^*) d\Pi(z) = \int U_n^*(z)[\hat{f}_{\bar{Z}}(z, \theta_n^*) - \hat{f}_{\bar{Z}}(z, \theta_0)] d\Pi(z) \\ &= \int U_n^*(z) d_n(z, \theta_n^*, \theta_0) d\Pi(z) + (\theta_n^* - \theta_0)^T \int U_n^*(z) \dot{\hat{f}}_{\bar{Z}}(z, \theta_0) d\Pi(z) \\ &=: Q_{11} + Q_{12}. \end{aligned}$$

For Q_{11}, by using the Cauchy-Schwarz inequality, and from Theorem 3.3.3, (x2), and (3.21), $nb^{1/2}|Q_{11}|$ is bounded above by

$$\begin{aligned} & nb^{1/2}\|\theta_n^* - \theta_0\| \left[\int (U_n^*(z))^2 d\Pi(z)\right]^{\frac{1}{2}} \left[\int \frac{d_n^2(z, \theta_n^*, \theta_0)}{\|\theta_n^* - \theta_0\|^2} d\Pi(z)\right]^{\frac{1}{2}} \\ =& nb^{1/2} O_p\left(\frac{1}{\sqrt{n}}\right) O_p\left(\frac{1}{\sqrt{nb}}\right) o_p(1) = o_p(1). \end{aligned} \tag{3.51}$$

Consider Q_{12} and notice that

$$\begin{aligned} & \int U_n^*(z) \dot{\hat{f}}_{\bar{Z}}(z, \theta_0) d\Pi(z) \\ =& \int U_n^*(z)[\dot{\hat{f}}_{\bar{Z}}(z, \theta_0) - \dot{f}_{\bar{Z}}(z, \theta_0)] d\Pi(z) + \int U_n^*(z) \dot{f}_{\bar{Z}}(z, \theta_0) d\Pi(z) =: g_{n1} + S_n. \end{aligned}$$

Therefore, from Lemma 3.3.2 and (3.28), one can easily verify

$$nb^{1/2}|Q_{12}| = |nb^{1/2}(\theta_n^* - \theta_0)^T(g_{n1} + S_n)| \le nb^{1/2} O_p\left(\frac{1}{\sqrt{n}}\right)\left[O_p\left(\frac{1}{nb^{1/2}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right)\right]$$

which is of order $o_p(1)$. Next, we show $nb^{1/2}Q_2 = o_p(1)$. In fact,

$$\begin{aligned}
&\int [Z_n^*(z,\theta_n^*)]^2 d\Pi(z) = \int [d_n(z,\theta_n^*,\theta_0) + (\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) \\
=& \int d_n^2(z,\theta_n^*,\theta_0) d\Pi(z) + (\theta_n^* - \theta_0)^T \int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)(\theta_n^* - \theta_0) \\
&+ 2\int d_n(z,\theta_n^*,\theta_0)(\theta_n^* - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(z,\theta_0) d\Pi(z) \\
=:& Q_{21} + Q_{22} + Q_{23}.
\end{aligned}$$

Consider Q_{21}. From Theorem 3.3.3,

$$\begin{aligned}
nb^{1/2}Q_{21} =& nb^{1/2}\int d_n^2(z,\theta_n^*,\theta_0) d\Pi(z) = nb^{1/2}\|\theta_n^* - \theta_0\|^2 \int \frac{d_n^2(z,\theta_n^*,\theta_0)}{\|\theta_n^* - \theta_0\|^2} d\Pi(z) \\
=& nb^{1/2} O_p(1/n) o_p(1) \int \left[\int f_{\bar{U}}(u) du\right]^2 d\Pi(z) = o_p(1).
\end{aligned}$$

The proof of $nb^{1/2}Q_{22} = o_p(1)$ is similar.

$$\begin{aligned}
\|nb^{1/2}Q_{22}\| =& nb^{1/2}\|\theta_n^* - \theta_0\|^2 \left\|\int \dot{\hat{f}}_{\bar{Z}}(z,\theta_0)[\dot{\hat{f}}_{\bar{Z}}(z,\theta_0)]^T d\Pi(z)\right\| \\
=& O_p(nb^{1/2}) O_p\left(\frac{1}{n}\right) O_p(1) = o_p(1).
\end{aligned}$$

For Q_{23}, by using the Cauchy-Schwarz inequality, we have $nb^{1/2}Q_{23} = o_p(1)$.

This concludes the proof of Lemma 3.4.3.

□

Lemma 3.4.4. *Suppose (z1), (π1), (b1)–(b2) hold. Then* $\hat{\Gamma}_n - \Gamma = o_p(1)$. *Consequently,* $\Gamma > 0$ *implies that* $|\hat{\Gamma}_n \Gamma^{-1} - 1| = o_p(1)$.

Proof. Let $\tilde{\Gamma}_n := \frac{2b}{n^2}\sum_{i\neq j}\left(\int [L_b(z - \bar{Z}_i) - EL_b(z - \bar{Z}_i)][L_b(z - \bar{Z}_j) - EL_b(z - \bar{Z}_j)] d\Pi(z)\right)^2$.

By using LLN, it is easy to show $|\hat{\Gamma}_n - \tilde{\Gamma}_n| = o_p(1)$. We only need to show $\hat{\Gamma}_n - \Gamma_n = o_p(1)$. The claim of this lemma follows from this result and the fact that $\Gamma_n \to \Gamma$.

Note that $\Gamma_n = E\hat{\Gamma}_n$. Hence

$$
\begin{aligned}
E(\Gamma_n - \hat{\Gamma}_n)^2 =& 4E\left[\sum_{i\neq j}(H_n^2(\bar{Z}_i, \bar{Z}_j) - EH_n^2(\bar{Z}_i, \bar{Z}_j))\right]^2 \\
\leq& 4\sum_{i\neq j} EH_n^4(\bar{Z}_i, \bar{Z}_j) + 4\sum_{i\neq j\neq k} EH_n^2(\bar{Z}_i, \bar{Z}_j)H_n^2(\bar{Z}_j, \bar{Z}_k) \\
\leq& 4(n^2 + n^3)EH_n^4(\bar{Z}_i, \bar{Z}_j).
\end{aligned}
$$

By using (3.50) and (b2), this upper bound is $O((nb)^{-1}) = o(1)$. $\square$

Combing the results of Lemma 3.4.2–3.4.4, we can prove Theorem 3.4.1 as follows:

Proof of Theorem 3.4.1. Recall $\tilde{T}(\theta_0)$ defined in (3.44). Adding and subtracting $E\hat{f}_{\bar{Z}}(z)$, $f_{\bar{Z}}(z, \theta_0)$ and $E\hat{f}_{\bar{Z}}(z, \theta_0)$ from $\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z, \theta_0)$, we can rewrite $T_n^*(\theta_0)$ as the sum of the following ten terms:

$$
\begin{aligned}
\tilde{T}_n(\theta_0) =& \int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z), \quad t_{n1}^* = \int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)]^2 d\Pi(z) \\
t_{n2}^* =& \int [f_{\bar{Z}}(z, \theta_0) - E\hat{f}_{\bar{Z}}(z, \theta_0)]^2 d\Pi(z), \quad t_{n3}^* = \int [E\hat{f}_{\bar{Z}}(z, \theta_0) - \hat{f}_{\bar{Z}}(z, \theta_0)]^2 d\Pi(z), \\
t_{n4}^* =& 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)][E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)]d\Pi(z), \\
t_{n5}^* =& 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)[f_{\bar{Z}}(z, \theta_0) - E\hat{f}_{\bar{Z}}(z, \theta_0)]d\Pi(z), \\
t_{n6}^* =& 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)][E\hat{f}_{\bar{Z}}(z, \theta_0) - \hat{f}_{\bar{Z}}(z, \theta_0)]d\Pi(z), \\
t_{n7}^* =& 2\int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)][f_{\bar{Z}}(z, \theta_0) - E\hat{f}_{\bar{Z}}(z, \theta_0)]d\Pi(z), \\
t_{n8}^* =& 2\int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)][E\hat{f}_{\bar{Z}}(z, \theta_0) - \hat{f}_{\bar{Z}}(z, \theta_0)]d\Pi(z), \\
t_{n9}^* =& 2\int [f_{\bar{Z}}(z, \theta_0) - E\hat{f}_{\bar{Z}}(z, \theta_0)][E\hat{f}_{\bar{Z}}(z, \theta_0) - \hat{f}_{\bar{Z}}(z, \theta_0)]d\Pi(z).
\end{aligned}
$$

We are going to show $nb^{1/2}t_{ni}^* = o_p(1), i = 1, \cdots, 9$.

From (z4) and (b3), and $E\hat{f}_{\bar{Z}}(z) = f_{\bar{Z}}(z,\theta_0) + \frac{b^2}{2}\int L(v)v^2 f''_{\bar{Z}}(z+\tau_1 vb,\theta_0)dv$, where $0 < \tau_1 < 1$, we have

$$\begin{aligned}
nb^{1/2}t^*_{n1} =& nb^{1/2}\int [E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) \\
=& nb^{1/2}\int \left[\frac{b^2}{2}\int L(v)v^2 f''_{\bar{Z}}(z+\tau_1 vb,\theta_0)dv\right]^2 d\Pi(z) \\
\leq& nb^{1/2}\frac{b^4}{4}c^2\int\left[\int L(v)v^2 dv\right]^2 d\Pi(z) = O(nb^{9/2}) = o(1).
\end{aligned} \tag{3.52}$$

From (u1), (b1), and (h3), one obtains

$$\begin{aligned}
nb^{1/2}t^*_{n2} =& nb^{1/2}\int [f_{\bar{Z}}(z,\theta_0) - E\hat{f}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) \\
=& nb^{1/2}\int\left[\int f_X(z-u,\theta_0)(f_{\bar{U}}(u) - E\hat{f}_{\bar{U}}(u))du\right]^2 d\Pi(z) \\
=& nb^{1/2}\int\left[\int f_X(z-u,\theta_0)\left(\frac{h^2}{2}\int K(v)v^2 f''_{\bar{U}}(u+\tau_2 vh)dv\right)du\right]^2 d\Pi(z) \\
=& O(nb^{1/2}h^4) = o(1),
\end{aligned}$$

where $0 < \tau_2 < 1$. For t^*_{n3}, we have

$$\begin{aligned}
nb^{1/2}Et^*_{n3} =& nb^{1/2}E\int [E\hat{f}_{\bar{Z}}(z,\theta_0) - \hat{f}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) = nb^{1/2}\int var(\hat{f}_{\bar{Z}}(z,\theta_0))d\Pi(z) \\
=& nb^{1/2}\int \frac{1}{n}var(G_{\tilde{U}}(z,\theta_0))d\Pi(z) \leq b^{1/2}\int EG^2_{\tilde{U}}(z,\theta_0)d\Pi(z).
\end{aligned}$$

Since

$$\begin{aligned}
EG^2_{\tilde{U}}(z,\theta_0) =& E\left[\int f_X(z-u,\theta_0)K_h(u-\tilde{U})du\right]^2 \\
=& \int f_X(z-u,\theta_0)K_h(u-s)f_X(z-v,\theta_0)K_h(v-s)f_{\bar{U}}(s)\,\mathrm{d}(u,v,s) \\
=& \int f_X(z-s-hu,\theta_0)K(u)f_X(z-s-hv,\theta_0)K(v)f_{\bar{U}}(s)\,\mathrm{d}(u,v,s),
\end{aligned}$$

we get $nb^{1/2}t^*_{n3} \leq O_p(b^{1/2}) = o_p(1)$. By the Cauchy-Schwarz inequality, we also have

$nb^{1/2}t^*_{ni} = o_p(1),\ i = 7, 8, 9$. We can show $C_n = O_p(1/(nb))$. In fact,

$$EC_n = E\left[\frac{1}{n^2}\sum_{i=1}^{n}\int\left(L_b(z-\bar{Z}_i) - \int L_b(z-x)f_{\bar{Z}}(x,\theta_0)dx\right)^2 d\Pi(z)\right]$$
$$= \frac{1}{n}\iint\left(L_b(z-s) - \int L(x)f_{\bar{Z}}(z-bx,\theta_0)dx\right)^2 d\Pi(z)f_{\bar{Z}}(s,\theta_0)ds = O\left(\frac{1}{nb}\right). \quad (3.53)$$

Therefore, from (3.52) and (3.53), by using the elementary inequality $(a+c)^{1/2} \leq a^{1/2} + c^{1/2}$ for $a \geq 0$, $c \geq 0$, we have

$$nb^{1/2}t^*_{n4} \leq nb^{1/2}(\tilde{T}_n(\theta_0) - C_n + C_n)^{\frac{1}{2}}(t^*_{n1})^{\frac{1}{2}}$$
$$\leq nb^{1/2}(\tilde{T}_n(\theta_0) - C_n)^{\frac{1}{2}}(t^*_{n1})^{\frac{1}{2}} + nb^{1/2}(C_n)^{\frac{1}{2}}(t^*_{n1})^{\frac{1}{2}}$$
$$= o_p(1) + nb^{1/2}O_p\left(\frac{1}{\sqrt{nb}}\right)O_p(b^2) = o_p(1).$$

Similarly, $nb^{1/2}t^*_{n5} = o_p(1)$. By using Lemma 3.4.1, it is not hard to see that $t^*_{n6} = o_p(\frac{1}{n})$. Therefore, $nb^{1/2}t^*_{n6} = O_p(b^{1/2}) = o_p(1)$.

Moreover, we can show $nb^{1/2}(\hat{C}_n(\theta^*_n) - C_n) = o_p(1)$. By adding and subtracting $\int L_b(z-x)f_{\bar{Z}}(x,\theta_0)dx$, $\int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_0)dx$, we can see that $\hat{C}_n(\theta^*_n)$ is the sum of C_n, D_{n1}, $D_{n2}(\theta^*_n)$ and the cross products, where

$$D_{n1} = \frac{1}{n}\int\left(\int L_b(z-x)(f_{\bar{Z}}(x,\theta_0) - \hat{f}_{\bar{Z}}(x,\theta_0))dx\right)^2 d\Pi(z),$$
$$D_{n2}(\theta^*_n) = \frac{1}{n}\int\left(\int L_b(z-x)(\hat{f}_{\bar{Z}}(x,\theta_0) - \hat{f}_{\bar{Z}}(x,\theta^*_n))dx\right)^2 d\Pi(z).$$

We only need to show $nb^{1/2}D_{n1} = o_p(1)$ and $nb^{1/2}D_{n2}(\theta^*_n) = o_p(1)$. The cross products are of order $o_p(1)$ from these two facts and $C_n = O_p(1)$ by using the Cauchy-Schwarz inequality.

Actually, for D_{n1}, we know $\int(\int L_b(z-x)(f_{\bar{Z}}(x,\theta_0) - \hat{f}_{\bar{Z}}(x,\theta_0))dx)^2 d\Pi(z)$ is bounded above by the sum of

$$2\int\left[\int L_b(z-x)(f_{\bar{Z}}(x,\theta_0) - E\hat{f}_{\bar{Z}}(x,\theta_0))dx\right]^2 d\Pi(z) = O(h^4)$$

and

$$2\int\left[\int L_b(z-x)(E\hat{f}_{\bar{Z}}(x,\theta_0)-\hat{f}_{\bar{Z}}(x,\theta_0))dx\right]^2 d\Pi(z)=O(\frac{1}{n}).$$

Then $nb^{1/2}D_{n1}=o_p(1)$. For $D_{n2}(\theta_n^*)$, from (x2), we can show

$$\int\left(\int L_b(z-x)(\hat{f}_{\bar{Z}}(x,\theta_0)-\hat{f}_{\bar{Z}}(x,\theta_n^*))dx\right)^2 d\Pi(z)=O(\frac{1}{n}).$$

Thus $nb^{1/2}D_{n2}(\theta_n^*)=o_p(1)$.

The facts shown above together with the results from Lemma 3.4.2 – Lemma 3.4.4 complete the proof of Theorem 3.4.1.

□

Remark 3.4.1. *The conclusion of Theorem 3.4.1 still holds if θ_n^* is replaced by any $\sqrt{n}$-consistent estimator of θ_0. One can also check the proof of Theorem 3.4.3 below and see that the conclusion of Theorem 3.4.3 is still valid when any $\sqrt{n}$-consistent estimator of θ_0, say the method of moment estimate, is used, in the place of $\hat{\theta}_n$.*

The following theorem presents the asymptotic distribution of the minimum distance test based on $T_n(\hat{\theta}_n)$.

Theorem 3.4.3. *Suppose H_0, (b1)–(b2), (h1)–(h3), (z1)–(z4), (x1)–(x3), (π1), (kl), and (u1) hold. Then $nb^{1/2}(T_n(\hat{\theta}_n)-\hat{C}_n(\hat{\theta}_n))\Rightarrow N(0,\Gamma)$.*

Define

$$\mathcal{T}_n(\hat{\theta}_n)=\hat{\Gamma}_n^{-1/2}nb^{1/2}(T_n(\hat{\theta}_n)-\hat{C}_n(\hat{\theta}_n)). \tag{3.54}$$

Then for the proposed test, we reject H_0 whenever $|\mathcal{T}_n(\hat{\theta}_n)|>Z_{\alpha/2}$, where Z_α is the $100(1-\alpha)\%$ percentile of the standard normal distribution.

Recall $H_{\tilde{U}_i}(z,\theta_0)$ as defined in (3.40). Then $T_n(\theta_0)$ can be written as

$$\begin{aligned}&\int\left\{\hat{f}_{\bar{Z}}(z)-\int L_b(z-x)\left[\int f_X(x-u,\theta_0)\hat{f}_{\bar{U}}(u)du\right]dx\right\}^2 d\Pi(z)\\ =&\int\left\{\hat{f}_{\bar{Z}}(z)-\frac{1}{n}\sum_{i=1}^n H_{\tilde{U}_i}(z,\theta_0)\right\}^2 d\Pi(z).\end{aligned}$$

Again, we will introduce Lemma 3.4.5 to facilitate our proof of Theorem 3.4.3.

Lemma 3.4.5. *Suppose H_0, (z1)–(z4), (b1)–(b2),(h1)–(h3), (x1)–(x3), (π1), (kl), and (u1). Then $nb^{1/2}(T_n(\hat{\theta}_n) - T_n(\theta_0)) = o_p(1)$.*

Proof. Recall $U_n(z)$ and $Z_n(x,\theta)$ defined in (3.32). We have

$$T_n(\hat{\theta}_n) - T_n(\theta_0) = 2\int U_n(z) Z_n(z,\hat{\theta}_n) d\Pi(z) - \int [Z_n(z,\hat{\theta}_n)]^2 d\Pi(z) = 2Q_1 - Q_2.$$

It suffices to show $nb^{1/2}Q_i = o_p(1), i = 1,2$. Recall $d_n(x,\hat{\theta}_n,\theta_0) = \hat{f}_{\bar{Z}}(x,\hat{\theta}_n) - \hat{f}_{\bar{Z}}(x,\theta_0) - (\hat{\theta}_n - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(x,\theta_0)$. Q_1 can be written as the sum of the following two terms:

$$\begin{aligned}
Q_{11} &= \int U_n(z) \int L_b(z-x) d_n(x,\hat{\theta}_n,\theta_0) dx d\Pi(z),\\
Q_{12} &= \int U_n(z) \int L_b(z-x)(\hat{\theta}_n - \theta_0)^T \dot{\hat{f}}_{\bar{Z}}(x,\theta_0) dx d\Pi(z).
\end{aligned}$$

For Q_{11}, from Theorem 3.3.4, (x2), and (3.37), by the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
nb^{1/2}Q_{11} \le & nb^{1/2}\left[\int U_n^2(z) d\Pi(z)\right]^{1/2} \left(\int \left[\int L_b(z-x) d_n(x,\hat{\theta}_n,\theta_0) dx\right)^2 d\Pi(z)\right]^{1/2}\\
\le & nb^{1/2}\left[\int U_n^2(z) d\Pi(z)\right]^{1/2} \sup_x \frac{f_X(x,\hat{\theta}_n) - f_X(x,\theta_0) - (\hat{\theta}_n - \theta_0)^T \dot{f}_X(x,\theta_0)}{\|\hat{\theta}_n - \theta_0\|}\\
& \cdot \|\hat{\theta}_n - \theta_0\| \left\{\int \left[\int L_b(z-x) \int \hat{f}_{\bar{U}}(u) du dx\right]^2 d\Pi(z)\right\}^{1/2}\\
= & O_p\left(nb^{1/2}\frac{1}{\sqrt{nb}}\frac{1}{\sqrt{n}}\right) o_p(1) O_p(1) = o_p(1). \qquad (3.55)
\end{aligned}$$

For Q_{12}, recall $\dot{\mu}_n(z,\theta_0) = \int L_b(z-x)\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)dx$ as defined in (3.32). From (3.41) and by Lemma 3.3.7, we have

$$\begin{aligned}
nb^{1/2}Q_{12} =& nb^{1/2}\int U_n(z)\int L_b(z-x)(\hat{\theta}_n-\theta_0)^T\dot{\hat{f}}_{\bar{Z}}(x,\theta_0)dxd\Pi(z) \\
=& nb^{1/2}(\hat{\theta}_n-\theta_0)^T\int U_n(z)\dot{\mu}_n(z,\theta_0)d\Pi(z) \\
=& nb^{1/2}(\hat{\theta}_n-\theta_0)^T\int U_n(z)(\dot{\mu}_n(z,\theta_0)-\dot{\mu}_n(z,\hat{\theta}_n))d\Pi(z) \\
& + nb^{1/2}(\hat{\theta}_n-\theta_0)^T\int U_n(z)\dot{\mu}_n(z,\hat{\theta}_n)d\Pi(z) \\
\leq& nb^{1/2}O_p\left(\frac{1}{\sqrt{n}}\right)o_p\left(\frac{b^{1/2}}{\sqrt{nb}}\right)+nb^{1/2}O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{\sqrt{n}}\right)=o_p(1).
\end{aligned}$$

Next, we will show $nb^{1/2}Q_2=o_p(1)$. Recall $d_n(x,\theta,\theta_0)$ as defined in (3.19), Q_2 can be written as the sum of the following three terms:

$$\begin{aligned}
Q_{21} =& \int\left[\int L_b(z-x)d_n(x,\hat{\theta}_n,\theta_0)dx\right]^2 d\Pi(z), \\
Q_{22} =& \int[(\hat{\theta}_n-\theta_0)^T\dot{\mu}_n(z,\theta_0)]^2 d\Pi(z), \\
Q_{23} =& 2\int\left(\int L_b(z-x)d_n(x,\hat{\theta}_n,\theta_0)dx\right)(\hat{\theta}_n-\theta_0)^T\dot{\mu}_n(z,\theta_0)d\Pi(z).
\end{aligned}$$

We will show $nb^{1/2}Q_{2i}=o_p(1), i=1,2,3$. For Q_{21}, from (x3), by the consistency of $\hat{\theta}_n$,

$$\begin{aligned}
nb^{1/2}Q_{21} =& nb^{1/2}\|\hat{\theta}_n-\theta_0\|^2\Big[\int L_b(z-x)\int\frac{d_n(x,\hat{\theta}_n,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}dx\Big]^2 d\Pi(z) \\
\leq& nb^{1/2}\|\hat{\theta}_n-\theta_0\|^2\sup_x\left[\frac{f_X(x,\hat{\theta}_n)-f_X(x,\theta_0)-(\hat{\theta}_n-\theta_0)^T\dot{f}_X(x,\theta_0)}{\|\hat{\theta}_n-\theta_0\|}\right]^2 \\
& \cdot\int\left[\int L_b(z-x)\int\hat{f}_{\bar{U}}(u)dudx\right]^2 d\Pi(z) \\
=& O_p\left(nb^{1/2}\frac{1}{n}\right)o_p(1)=o_p(1).
\end{aligned}$$

For Q_{22}, by using Theorem 3.3.4 and Lemma 3.3.4, one obtains

$$nb^{1/2}Q_{22}=nb^{1/2}(\hat{\theta}_n-\theta_0)^T\int\dot{\mu}_n(z,\theta_0)(\dot{\mu}_n(z,\theta_0))^T d\Pi(z)(\hat{\theta}_n-\theta_0)=O_p\left(nb^{1/2}\frac{1}{n}\right)=o_p(1).$$

And the Cauchy-Schwarz inequality implies $nb^{1/2}Q_{23} = o_p(1)$. $\square$

Next, let's prove Theorem 3.4.3.

Proof of Theorem 3.4.3. Based on Lemma 3.4.5, it suffices to show $nb^{1/2}(T_n(\theta_0) - C_n) \Rightarrow N(0, \Gamma)$ and $nb^{1/2}(\hat{C}_n(\hat{\theta}_n) - C_n) = o_p(1)$. Recall $\tilde{T}_n(\theta_0)$ defined in (3.44), $T_n(\theta_0)$ can be written as the sum of the following six terms:

$$
\begin{aligned}
\tilde{T}_n(\theta_0) &= \int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)]^2 d\Pi(z) = \int \left\{ \hat{f}_{\bar{Z}}(z) - \int L_b(z-x) f_{\bar{Z}}(x, \theta_0) dx \right\}^2 d\Pi(z), \\
t_{n1} &= \int [E\hat{f}_{\bar{Z}}(z) - EH_{\tilde{U}}(z, \theta_0)]^2 d\Pi(z), \\
t_{n2} &= \int \left[EH_{\tilde{U}}(z, \theta_0) - \frac{1}{n}\sum_{i=1}^{n} H_{\tilde{U}_i}(z, \theta_0) \right]^2 d\Pi(z), \\
t_{n3} &= 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)][E\hat{f}_{\bar{Z}}(z) - EH_{\tilde{U}}(z, \theta_0)] d\Pi(z), \\
t_{n4} &= 2\int [\hat{f}_{\bar{Z}}(z) - E\hat{f}_{\bar{Z}}(z)] \left[EH_{\tilde{U}}(z, \theta_0) - \frac{1}{n}\sum_{i=1}^{n} H_{\tilde{U}_i}(z, \theta_0) \right] d\Pi(z), \\
t_{n5} &= 2\int [E\hat{f}_{\bar{Z}}(z) - EH_{\tilde{U}}(z, \theta_0)] \left[EH_{\tilde{U}}(z, \theta_0) - \frac{1}{n}\sum_{i=1}^{n} H_{\tilde{U}_i}(z, \theta_0) \right] d\Pi(z).
\end{aligned}
$$

From Lemma 3.4.2, we know $nb^{1/2}(\tilde{T}_n(\theta_0) - C_n) \Rightarrow N(0, \Gamma)$. Next, we need only to show $nb^{1/2}t_{ni} = o_p(1), i = 1, 2, 3, 4, 5$.

If we can show $nb^{1/2}t_{n1} = o_p(1), nb^{1/2}t_{n2} = o_p(1)$, by the Cauchy–Schwarz inequality, we obtain $nb^{1/2}t_{n5} \le [nb^{1/2}t_{n1}]^{\frac{1}{2}}[nb^{1/2}t_{n2}]^{\frac{1}{2}} = o_p(1)$.

By the transformation $u - s = ht$, we have

$$
\begin{aligned}
& EH_{\tilde{U}}(z, \theta_0) \\
&= \int L_b(z-x) f_X(x-u, \theta_0) K_h(u-s) f_{\tilde{U}}(s)\, \mathrm{d}(s, u, x) \\
&= \int L_b(z-x) f_X(x-u, \theta_0) K(t) [f_{\tilde{U}}(u) + f'_{\tilde{U}}(u) ht + \frac{(ht)^2}{2} f''_{\tilde{U}}(u + \tau ht)]\, \mathrm{d}(t, u, x) \\
&= \int L_b(z-x) f_{\bar{Z}}(x, \theta_0) dx + \frac{h^2}{2} \int L_b(z-x) f_X(x-u, \theta_0) K(t) t^2 f''_{\tilde{U}}(u + \tau ht)\, \mathrm{d}(t, u, x),
\end{aligned}
$$

where $0 < \tau < 1$. From (h3) and (z4), we have

$$
\begin{aligned}
nb^{1/2}t_{n1} =& nb^{1/2}\int \Big[\int L_b(z-x)f_{\tilde{Z}}(x,\theta_0)dx - EH_{\tilde{U}}(z,\theta_0)\Big]^2 d\Pi(z) \\
=& nb^{1/2}\int \left[\frac{h^2}{2}\int L_b(z-x)f_X(x-u,\theta_0)K(t)t^2 f''_{\tilde{U}}(u+\theta ht)\,\mathrm{d}(t,u,x)\right]^2 d\Pi(z) \\
=& O(nb^{1/2}h^4) = o(1).
\end{aligned}
$$

What's more,

$$
\begin{aligned}
Et_{n2} =& E\int \left[\frac{1}{n}\sum_{i=1}^n H_{\tilde{U}_i}(z,\theta_0) - E\Big(\frac{1}{n}\sum_{i=1}^n H_{\tilde{U}_i}(z,\theta_0)\Big)\right]^2 d\Pi(z) \\
=& \int var\left(\frac{1}{n}\sum_{i=1}^n H_{\tilde{U}_i}(z,\theta_0)\right) d\Pi(z) = \frac{1}{n}\int var(H_{\tilde{U}}(z,\theta_0))d\Pi(z) \\
=& \frac{1}{n}\int \left[E[H_{\tilde{U}}(z,\theta_0)]^2 - [EH_{\tilde{U}}(z,\theta_0)]^2\right] d\Pi(z).
\end{aligned}
$$

Since

$$
E[H_{\tilde{U}}(z,\theta_0)]^2 = \int \left\{\int L_b(z-x)\left[\int f_X(x-u,\theta_0)K_h(u-s)du\right]dx\right\}^2 f_{\tilde{U}}(s)ds = O(1),
$$

we know $nb^{1/2}Et_{n2} \leq nb^{1/2}O(1/n) = O(b^{1/2}) = o(1)$.

Next, we are going to show $nb^{1/2}t_{n3} = o_p(1)$ and $nb^{1/2}t_{n4} = o_p(1)$.

From (3.53), we have $C_n = O_p(\frac{1}{nb})$. Therefore, using the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
nb^{1/2}t_{n3} \leq& nb^{1/2}[\tilde{T}(\theta_0)]^{1/2}[t_{n1}]^{1/2} = nb^{1/2}[\tilde{T}(\theta_0) - C_n + C_n]^{1/2}(t_{n1})^{1/2} \\
\leq& nb^{1/2}(\tilde{T}(\theta_0) - C_n)^{1/2}(t_{n1})^{1/2} + nb^{1/2}C_n^{1/2}(t_{n1})^{1/2} \\
=& O_p(1)o_p(1) + nb^{1/2}O_p\left(\frac{1}{\sqrt{nb}}\right)O(h^2) = o_p(1).
\end{aligned}
$$

Moreover, applying Lemma 3.4.1, we have $nb^{1/2}t_{n4} = nb^{1/2}O_p(\frac{1}{n}) = o_p(1)$.

By similar procedure as in the proof of Theorem 3.4.1, we can show $nb^{1/2}(\hat{C}_n(\hat{\theta}_n) - C_n) =$

$o_p(1)$. The proof is omitted here for the sake of simplicity.

In sum, $nb^{1/2}(T_n(\hat{\theta}_n) - \hat{C}_n(\hat{\theta}_n)) \Rightarrow N(0, \Gamma)$. □

Remark 3.4.2. *For the same reason as stated in Remark 3.3.2, the non-centered test in Theorem 3.4.1 requires both $nb^4 \to 0$ and $nh^4 \to 0$, while the centered test in Theorem 3.3.4 only requires $nh^4 \to 0$.*

3.5 Consistency and Local Power of MD Test Statistic

Consistency is a basic requirement of any reasonable test. It requires that the test should have a power tending to 1 for any fixed alternative hypothesis when the sample size n goes to ∞. In this section, we shall show that, under some regularity conditions, the tests in Section 3.4 are consistent against certain fixed alternatives.

Let $f_{X,a}$ be a density on $\mathbb{R}$ and consider the alternative $H_a : f_X(x) = f_{X,a}(x)$, for all $x \in \mathbb{R}$. Under H_a, density of $\bar{Z}$ is $f_{\bar{Z},a}(z) = \int f_{X,a}(z-u) f_{\bar{U}}(u) du$, which can be estimated by $\hat{f}_{\bar{Z},a}(z) = \int f_{X,a}(z-u) \hat{f}_{\bar{U}}(u) d\Pi(z)$. We shall assume that $\hat{\theta}_n$ converges to a value θ_a in probability under H_a. In fact, if

$$\theta_a = \underset{\theta}{\text{minimizer}} \int [f_{\bar{Z},a}(z) - f_{\bar{Z}}(z, \theta)]^2 d\Pi(z) \tag{3.56}$$

is well defined, then one can show that the minimum distance estimator θ_a^* or $\hat{\theta}_a$ converges to θ_a in probability. The proof is omitted for the sake of brevity.

The following theorem states the consistency of the test $\mathcal{T}_n^*(\theta_n^*)$ defined in (3.43).

Theorem 3.5.1. *Suppose (x1)–(x3), (z1)–(z4), (h1)–(h3), (b1)–(b3), (π1), (kl) and (u1) hold. Under H_a, assume that θ_a in (3.56) is well defined, and the additional assumption that $f_{X,a}(z)$ is bounded and $\int [f_{\bar{Z},a}(z) - f_{\bar{Z}}(z, \theta_a)]^2 d\Pi(z) > 0$, we have $\mathcal{T}_n^*(\theta_n^*) = nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n^*(\theta_n^*) - \hat{C}_n) \to_p \infty$. Consequently, the above test is consistent against H_a.*

Proof. Adding and subtracting $f_{\bar{Z},a}(z)$ inside the quadratic term of the integrand, one obtains

$$\begin{aligned}
T_n^*(\theta_n^*) &= \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z}}(z, \theta_n^*)]^2 d\Pi(z) = \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z},a}(z) + \hat{f}_{\bar{Z},a}(z) - \hat{f}_{\bar{Z}}(z, \theta_n^*)]^2 d\Pi(z) \\
&= \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z},a}(z)]^2 d\Pi(z) + \int [\hat{f}_{\bar{Z},a}(z) - \hat{f}_{\bar{Z}}(z, \theta_n^*)]^2 d\Pi(z) \\
&\quad + 2 \int [\hat{f}_{\bar{Z}}(z) - \hat{f}_{\bar{Z},a}(z)][\hat{f}_{\bar{Z},a}(z) - \hat{f}_{\bar{Z}}(z, \theta_n^*)] d\Pi(z) \\
&=: T_{n1}^* + T_{n2}^* + T_{n3}^*.
\end{aligned}$$

One can show that $nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{n1}^* - \hat{C}_n) \Rightarrow N(0,1)$. The proof is similar to that of Theorem 3.4.1. Note that now

$$\hat{\Gamma}_n \to 2 \int f_{\bar{Z},a}^2(v)\pi^2(y)dy \int \left(\int L(t)L(z+t)dt \right)^2 dz =: \tilde{\Gamma} \quad \text{in probability.}$$

What's more, adding and subtracting $f_{\bar{Z},a}(z), f_{\bar{Z}}(z,\theta_a), \hat{f}_{\bar{Z}}(z,\theta_a)$ in the quadratic term of the integrand in T_{n2}^*, expanding the term, using the fact $\int |\hat{f}_{\bar{U}}(u) - f_{\bar{U}}(u)| du = o_p(1)$ and from (x3), one verifies $T_{n2}^* = \int [f_{\bar{Z},a}(z) - f_{\bar{Z}}(z,\theta_a)]^2 d\Pi(z) + o_p(1)$. Therefore,

$$nb^{1/2}\hat{\Gamma}_n^{-1/2} T_{n2}^* = nb^{1/2}\tilde{\Gamma}^{-1/2} \int [f_{\bar{Z},a}(z) - f_{\bar{Z}}(z,\theta_a)]^2 d\Pi(z) + o_p(nb^{1/2}).$$

By the Cauchy–Schwarz inequality, the elementary inequality $(a+c)^{\frac{1}{2}} \le a^{\frac{1}{2}} + c^{\frac{1}{2}}$ for $a \ge 0, c \ge 0$, and from (3.53), one can show that

$$\begin{aligned}
nb^{1/2}\hat{\Gamma}_n^{-1/2}|T_{n3}|^* \le& 2nb^{1/2}\hat{\Gamma}_n^{-1/2}|T_{n1}^* - \hat{C}_n|^{1/2}(T_{n2}^*)^{1/2} + 2(nb^{1/2}\hat{\Gamma}_n^{-1/2}\hat{C}_n)^{1/2}(nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n2}^*)^{1/2} \\
=& O_p(1)O_p(\sqrt{nb^{1/2}}) + O_p\left(\sqrt{nb^{1/2}}\frac{1}{\sqrt{nb}}\right) O_p(\sqrt{nb^{1/2}}) = o_p(nb^{1/2})
\end{aligned}$$

from $nb \to \infty$ guaranteed by the assumption (b2). Thus

$$\begin{aligned}
& nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n^*(\theta_n^*) - \hat{C}_n) \\
=& nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{n1}^* - \hat{C}_n) + nb^{1/2}\tilde{\Gamma}^{-1/2} \int [f_{\bar{Z},a}(a) - f_{\bar{Z}}(z,\theta_a)]^2 d\Pi(z) + o_p(nb^{1/2}).
\end{aligned}$$

Clearly, the right hand side of the above expression tends to ∞ as $n \to \infty$, implying that the proposed test is consistent. $\square$

Next, we consider the consistency of the test $\mathcal{T}_n(\hat{\theta}_n)$ in (3.54).

Theorem 3.5.2. *Suppose (b1)–(b2), (h1)–(h3), (z1)–(z4), (x1)–(x3), (π1), (kl), and (u1) hold. Under H_a, the additional assumption that θ_a in (3.56) is well defined, and $f_{X,a}(z)$ is bounded, $\int[f_{\bar{Z},a}(z) - f_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z) > 0$, we have $\mathcal{T}_n(\hat{\theta}_n) = nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\theta}_n) - \hat{C}_n) \to_p \infty$. Consequently, the above test is consistent against H_a.*

Proof. Add and subtract $\int L_b(z-x)\hat{f}_{\bar{Z},a}(x)dx$ inside the quadratic term of the integrand. Then $T_n(\hat{\theta}_n)$ can be written as sum of the following three terms:

$$
\begin{aligned}
T_{n1} =& \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z},a}(x)dx \right]^2 d\Pi(z), \\
T_{n2} =& \int \left[\int L_b(z-x)(\hat{f}_{\bar{Z},a}(x) - \hat{f}_{\bar{Z}}(x,\hat{\theta}_n))dx \right]^2 d\Pi(z), \\
T_{n3} =& 2\int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z},a}(x)dx \right] \left[\int L_b(z-x)(\hat{f}_{\bar{Z},a}(x) - \hat{f}_{\bar{Z}}(x,\hat{\theta}_n))dx \right] d\Pi(z).
\end{aligned}
$$

One can show that $nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_{n1} - \hat{C}_n) \Rightarrow N(0,1)$. The proof is similar to that of Theorem 3.4.3. Note that now $\hat{\Gamma}_n \to \tilde{\Gamma}$ in probability.

What's more,

$$
\begin{aligned}
T_{n2} =& \int \left[\int L_b(z-x)(\hat{f}_{\bar{Z},a}(x) - \hat{f}_{\bar{Z}}(x,\hat{\theta}_n))dx \right]^2 d\Pi(z) \\
=& \int \Bigg[\int L_b(z-x)\hat{f}_{\bar{Z},a}(x)dx \mp \int L_b(z-x)f_{\bar{Z},a}(x)dx \mp \int L_b(z-x)f_{\bar{Z}}(x,\theta_a)dx \\
& \mp \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_a)dx - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\hat{\theta}_n)dx \Bigg]^2 d\Pi(z)
\end{aligned}
$$

where $\mp$ stands for first minus then plus the term after the sign. Expanding the quadratic term and using change of variables formula, one verifies $T_{n2} = \int[f_{\bar{Z},a}(z) - f_{\bar{Z}}(z,\theta_a)]^2 d\Pi(z) +$

$o_p(1)$. Therefore,

$$nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n2} = nb^{1/2}\tilde{\Gamma}^{-1/2}\int [f_{\bar{Z},a}(z) - f_{\bar{Z}}(z,\theta_a)]^2 d\Pi(z) + o_p(nb^{1/2}).$$

By similar argument as in Theorem 3.5.1, we know the proposed test is consistent. □

Next, we shall show that the proposed tests possesses nontrivial power for certain local alternatives which converges to the null hypothesis at the rate of $1/\sqrt{nb^{1/2}}$. For this purpose, let φ be a known continuous density on $\mathbb{R}$ with mean 0 and positive variance σ_φ^2, and we consider the following local alternative hypothesis

$$H_{\text{loc}} : f(x) = (1-\delta_n)f_X(x,\theta_0) + \delta_n\varphi(x)$$

with $\delta_n = 1/\sqrt{nb^{1/2}}$. Similar to the fixed alternative case, to show the local power result, we need to show the $\sqrt{n}$ consistency of θ_n^* and $\hat{\theta}_n$, which is similar and omitted here for the sake of brevity.

Theorem 3.5.3. *Assume all the conditions in Theorem 3.4.1 hold. Under H_{loc}, if the density function $\varphi(\cdot)$ is twice continuously differentiable and the second derivative is bounded, then*

$$\mathcal{T}_n^*(\theta_n^*) = nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n^*(\theta_n^*) - \hat{C}_n) \Rightarrow N(\mu_t, 1),$$

as $n \to \infty$, where $\mu_t = \Gamma^{-1/2}\int\left[\int(f_X(z-u,\theta_0) - \varphi(z-u))f_{\bar{U}}(u)du\right]^2 d\Pi(z)$.

Proof. Denote

$$\begin{aligned}
f_{\bar{Z}}^{\text{loc}}(z,\theta_0) &= \int[(1-\delta_n)f_X(z-u,\theta_0) + \delta_n\varphi(z-u)]f_{\bar{U}}(u)du,\\
\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0) &= \int[(1-\delta_n)f_X(z-u,\theta_0) + \delta_n\varphi(z-u)]\hat{f}_{\bar{U}}(u)du\\
&=(1-\delta_n)\hat{f}_{\bar{Z}}(z,\theta_0) + \delta_n\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du.
\end{aligned}$$

From Lemma 3.4.3, we have $T_n^*(\theta_0) - T_n^*(\theta_n^*) = o_p(1)$. Therefore, we only need to show $nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n^*(\theta_0)-\hat{C}_n) \Rightarrow N(\mu_t,1)$. Adding and subtracting $\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)$ from $\hat{f}_{\bar{Z}}(z)-\hat{f}_{\bar{Z}}(z,\theta_0)$, we can rewrite the statistic as

$$T_n^*(\theta_0)=\int [\hat{f}_{\bar{Z}}(z)-\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)+\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)-\hat{f}_{\bar{Z}}(z,\theta_0)]^2 d\Pi(z).$$

Note that $\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)-\hat{f}_{\bar{Z}}(z,\theta_0)=-\delta_n\left[\hat{f}_{\bar{Z}}(z,\theta_0)-\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du\right]$.

Expanding the quadratic term, we can rewrite $T_n^*(\theta_n^*)$ as the sum of the following three terms

$$\begin{aligned}
T_{n1}^{*\text{loc}} =&\int [\hat{f}_{\bar{Z}}(z)-\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)]^2 d\Pi(z),\\
T_{n2}^{*\text{loc}} =&\delta_n^2\int\left[\hat{f}_{\bar{Z}}(z,\theta_0)-\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du\right]^2 d\Pi(z),\\
T_{n3}^{*\text{loc}} =&-2\delta_n\int [\hat{f}_{\bar{Z}}(z)-\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)]\left[\hat{f}_{\bar{Z}}(z,\theta_0)-\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du\right] d\Pi(z).
\end{aligned}$$

Similar to the proof of Theorem 3.4.1, one can verify that $nb^{1/2}(T_{n1}^{*\text{loc}}-\hat{C}_n)\Rightarrow N(0,\Gamma)$. For $T_{n2}^{*\text{loc}}$, it's not hard to see

$$\begin{aligned}
nb^{1/2}\hat{\Gamma}^{-1/2}T_{n2}^{*\text{loc}} =&\hat{\Gamma}_n^{-1/2}\int\left[\int f_X(z-u,\theta_0)\hat{f}_{\bar{U}}(u)du-\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du\right]^2 d\Pi(z)\\
=&\hat{\Gamma}_n^{-1/2}\int\left[\int (f_X(z-u,\theta_0)-\varphi(z-u))\hat{f}_{\bar{U}}(u)du\right]^2 d\Pi(z)\\
=&\Gamma^{-1/2}\int\left[\int (f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du\right]^2 d\Pi(z)+o_p(1)
\end{aligned}$$

Similarly, we can obtain

$$\begin{aligned}
&nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n3}^{*\text{loc}}\\
=&-\sqrt{nb^{1/2}}\hat{\Gamma}_n^{-1/2}\int [\hat{f}_{\bar{Z}}(z)-\hat{f}_{\bar{Z}}^{\text{loc}}(z,\theta_0)]\left[\hat{f}_{\bar{Z}}(z,\theta_0)-\int\varphi(z-u)\hat{f}_{\bar{U}}(u)du\right] d\Pi(z)\\
=&-\sqrt{nb^{1/2}}\Gamma^{-1/2}\int [\hat{f}_{\bar{Z}}(z)-f_{\bar{Z}}^{\text{loc}}(z,\theta_0)]\\
&\cdot\left[\int (f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du\right] d\Pi(z)+o_p(1)
\end{aligned}$$

is of order $o_p(b^{1/4}) = o_p(1)$ from

$$E\left(\int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}^{\mathrm{loc}}(z,\theta_0)] \left[\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du\right] d\Pi(z)\right) = 0$$

and

$$\begin{aligned}
&var\left(\int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}^{\mathrm{loc}}(z,\theta_0)] \left[\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du\right] d\Pi(z)\right) \\
\leq& E\left(\int [\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}^{\mathrm{loc}}(z,\theta_0)] \left[\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du\right] d\Pi(z)\right)^2 \\
=& \frac{1}{n} E\left(\int [L_b(z-\bar{Z}) - f_{\bar{Z}}^{\mathrm{loc}}(z,\theta_0)] \left[\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du\right] d\Pi(z)\right)^2 \\
=& O\left(\frac{1}{n}\right).
\end{aligned}$$

□

Theorem 3.5.4. *Assume all the conditions in Theorem 3.4.3 hold. Under H_{loc}, if the density function $\varphi(\cdot)$ is twice continuously differentiable and the second derivative is bounded, then*

$$\mathcal{T}_n(\hat{\theta}_n) = nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\theta}_n) - \hat{C}_n) \Rightarrow N(\mu_t, 1),$$

as $n \to \infty$, where $\mu_t = \Gamma^{-1/2}\int \left[\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du\right]^2 d\Pi(z)$.

Proof. From Lemma 3.4.5, we only need to show $nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\theta_0) - \hat{C}_n) \Rightarrow N(\mu_t, 1)$. Adding and subtracting $\iint L_b(z-x)\hat{f}_{\bar{Z}}^{\mathrm{loc}}(x,\theta_0)dxd\Pi(z)$ from $\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}(x,\theta_0)dx$ in $T_n(\theta_0)$, then $T_n(\theta_0)$ can be written as the sum of the following three terms.

$$\begin{aligned}
T_{n1} &= \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}^{\mathrm{loc}}(x,\theta_0)dx\right]^2 d\Pi(z) \\
T_{n2} &= \int \left[\int L_b(z-x)(\hat{f}_{\bar{Z}}^{\mathrm{loc}}(x,\theta_0) - \hat{f}_{\bar{Z}}(x,\theta_0))dx\right]^2 d\Pi(z) \\
T_{n3} &= \int \left[\hat{f}_{\bar{Z}}(z) - \int L_b(z-x)\hat{f}_{\bar{Z}}^{\mathrm{loc}}(x,\theta_0)dx\right] \left[\int L_b(z-x)(\hat{f}_{\bar{Z}}^{\mathrm{loc}}(x,\theta_0) - \hat{f}_{\bar{Z}}(x,\theta_0))dx\right] d\Pi(z)
\end{aligned}$$

Similar to the proof of Theorem 3.4.3, we can show $nb^{1/2}(T_{n1} - \hat{C}_n) \Rightarrow N(0, \Gamma)$.

For T_{n2},

$$\begin{aligned}
nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n2} =& \hat{\Gamma}_n^{-1/2}\int\left[\int L_b(z-x)\int(f_X(x-u,\theta_0)-\varphi(x-u))\hat{f}_{\bar{U}}(u)dudx\right]^2 d\Pi(z) \\
=& \Gamma^{-1/2}\int\left[\int L(v)\int(f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)dudv\right]^2 d\Pi(z)+o_p(1) \\
=& \mu_t+o_p(1).
\end{aligned}$$

For T_{n3},

$$\begin{aligned}
nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n3} =& -\sqrt{nb^{1/2}}\hat{\Gamma}_n^{-1/2}\int\left[\hat{f}_{\bar{Z}}(z)-\int L_b(z-x)\hat{f}_{\bar{Z}}^{\text{loc}}(x,\theta_0)dx\right] \\
& \cdot\left[\int L_b(z-x)\int(f_X(x-u,\theta_0)-\varphi(x-u))\hat{f}_{\bar{U}}(u)dudx\right]d\Pi(z) \\
=& -\sqrt{nb^{1/2}}\Gamma^{-1/2}\int[\hat{f}_{\bar{Z}}(z)-f_{\bar{Z}}^{\text{loc}}(z,\theta_0)] \\
& \cdot\left[\int(f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du\right]d\Pi(z)+o_p(1).
\end{aligned}$$

By using similar argument as in the proof of Theorem 3.5.3, we obtain $nb^{1/2}\hat{\Gamma}_n^{-1/2}T_{n3} = O_p(b^{1/4}) = o_p(1)$. In sum, we have $nb^{1/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\theta}_n)-\hat{C}_n) \Rightarrow N(\mu_t, 1),\ n\to\infty$. □

Next, we shall discuss the optimal weight function Π which maximizes the asymptotic local power of the proposed tests, which in turn maximizes the mean of the asymptotic normal distribution, or $\Psi(\pi) := \Gamma^{-1/2}\int\left[\int(f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du\right]^2\pi(v)dv$. By the Cauchy-Schwarz inequality, we have

$$\Psi(\pi) \le \frac{1}{(2\int(\int L(t)L(z+t)dt)^2dz)^{1/2}}\left(\int\frac{(\int(f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du)^4}{f_{\bar{Z}}^2(z,\theta_0)}dz\right)^{1/2}$$

with equality holds if and only if

$$\pi(z) \propto \left(\int(f_X(z-u,\theta_0)-\varphi(z-u))f_{\bar{U}}(u)du\right)^2 / f_{\bar{Z}}^2(z,\theta_0)$$

for all z. Since the functional Ψ is scale-invariant, that is $\Psi(aw) = \Psi(\pi)$ for all positive constant $a > 0$, we take the optimal $\pi(\cdot)$ to be

$$\pi(z) = \left(\frac{\int (f_X(z-u,\theta_0) - \varphi(z-u)) f_{\bar{U}}(u)du}{\int f_X(z-u,\theta_0) f_{\bar{U}}(u)du} \right)^2 .$$

One can estimate the optimal weight $\pi(\cdot)$ by $w_n(z)$ where the unknown density function or parameter $f_{\bar{U}}(u), \theta_0$ are replaced by the $\sqrt{n}$ estimates $\hat{f}_{\bar{U}}(u), \theta_n^*$ or $\hat{\theta}_n$.

3.6 Simulation

To evaluate the finite sample performance of the proposed tests, a simulation study is conducted in this section. The null hypothesis H_0 to be tested is $X \sim N(0, \sigma_X^2)$, so the unknown parameter θ is the variance σ_X^2 of X. The measurement error $U \sim (0, \sigma_U^2)$, where σ_U^2 is chosen to be 0.5^2 and 0.8^2. At each X-value, double measurements on Z are obtained. The sample size n is chosen to be 100, 200, and the weight function Π is taken to be the uniform distribution over the closed interval $[-6, 6]$ so that computationally the integration over this interval is nearly same as the integration over the whole real line. The kernel functions K and L are chosen to be standard normal density function. We repeat the test procedure 500 times for each scenario.

To study the empirical size and power of the test, the following nine non-normal distributions are used. For the sake of computational efficiency, we use the method of moment estimate $\tilde{\theta}_n$ in the simulation. The empirical levels and powers are calculated as the relative frequencies of the number of times of $|\mathcal{T}_n^*(\tilde{\theta}_n)|$ or $|\mathcal{T}_n(\tilde{\theta}_n)|$, which are defined in (3.43) and (3.54) respectively, exceed the critical value $Z_{\alpha/2}$, the $100(1-\alpha)/2$-th upper percentile of standard normal distribution. The significance level α is 0.05 in all cases.

Nine non-normal distributions as the alternative hypotheses:

- Logistic distribution with location parameter 0 and scale parameter 1;
- Cauchy distribution with location parameter 0 and scale parameter 1;
- Double exponential distribution with mean 0 and variance 1 (DE(0,1));

- t-distribution with degrees of freedom 3, 5 and 10;

- Two-component normal mixture models $0.5N(c, \sigma_\varepsilon^2) + 0.5N(-c, \sigma_\varepsilon^2)$ with $c = 0.5, 0.75$ and 1.

We also test the sensitivity of the proposed tests by conducting simulation studies with different measurement error densities and bandwidths levels.

Case I: $U \sim N(0, \sigma_U^2), b = n^{-1/3}, h = n^{-1/3}$.

Case II: $U \sim \text{Laplace}(0, \sigma_U/\sqrt{2}), b = n^{-1/3}, h = n^{-1/3}$.

Case III: $U \sim N(0, \sigma_U^2), b = 0.8 * n^{-1/3}, h = 0.8 * n^{-1/3}$.

Case IV: $U \sim N(0, \sigma_U^2), b = 1.2 * n^{-1/3}, h = 1.2 * n^{-1/3}$.

Case V: $U \sim N(0, \sigma_U^2), b = n^{-1/5}, h = n^{-1/3}$.

Note that Case I and Case II differs in the density function of the measurement error term U, with one being the standard normal density and the other being Laplace density with the same level of σ_U. Case III, I, IV have the same order of bandwidth but different coefficient, $0.8, 1, 1.2$, respectively. Case I and Case V are different in the bandwidth for b. Obviously, the assumption (b3) is violated. From the theorems in Section 3.4, we know the assumption (b3) is needed for the asymptotic normality of $\mathcal{T}_n^*(\tilde{\theta}_n)$, but is not required for that of $\mathcal{T}_n(\tilde{\theta}_n)$.

The simulation results below show that the proposed tests have reasonable empirical level for both $\mathcal{T}_n^*(\tilde{\theta}_n)$ and $\mathcal{T}_n(\tilde{\theta}_n)$ when the assumptions (b3) and (h3) are satisfied. For the case that assumption (b3) is violated, the empirical level become unreasonably large for $\mathcal{T}_n^*(\tilde{\theta}_n)$, while the centered version $\mathcal{T}_n(\tilde{\theta}_n)$ still holds valid empirical levels. When sample size n increases, the power generally increases as well. Another general trend is the power decreases when σ_U gets large. When comparing the non-centered test $\mathcal{T}_n^*(\tilde{\theta}_n)$ with centered test $\mathcal{T}_n(\tilde{\theta}_n)$, power for the centered test $\mathcal{T}_n(\tilde{\theta}_n)$ is higher than the non-centered test $\mathcal{T}_n^*(\tilde{\theta}_n)$ for all the uni-modal distributions. The trend reverses for the bi-modal distributions. There is no clear change tendency with different coefficient when both b and h are chosen with order $n^{-1/3}$. Different measurement error distribution doesn't cause significant change in power either.

Table 3.1: Simulation results of the proposed test
(Case I: $U \sim N(0, \sigma_U^2)$, $b, h = n^{-1/3}$)

	$\mathcal{T}_n^*(\tilde{\theta}_n)$				$\mathcal{T}_n(\tilde{\theta}_n)$			
	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$	
	100	200	100	200	100	200	100	200
$N(0, \sigma_\varepsilon^2)$	0.046	0.050	0.054	0.066	0.048	0.050	0.050	0.066
Logistic(0,1)	0.056	0.114	0.054	0.058	0.056	0.118	0.056	0.060
Cauchy(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DE(0,1)	0.432	0.728	0.238	0.486	0.452	0.738	0.246	0.510
$t(3)$	0.528	0.796	0.376	0.698	0.536	0.802	0.390	0.702
$t(5)$	0.176	0.292	0.124	0.170	0.194	0.318	0.134	0.192
$t(10)$	0.064	0.058	0.048	0.044	0.070	0.064	0.054	0.052
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.076	0.078	0.046	0.054	0.064	0.068	0.044	0.048
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.274	0.574	0.114	0.158	0.236	0.526	0.102	0.142
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.892	1.000	0.404	0.690	0.874	0.998	0.360	0.668

Table 3.2: Simulation results of the proposed test
(Case II: $U \sim$ Laplace$(0, \sigma_U/\sqrt{2})$)

	$\mathcal{T}_n^*(\tilde{\theta}_n)$				$\mathcal{T}_n(\tilde{\theta}_n)$			
	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$	
	100	200	100	200	100	200	100	200
$N(0, \sigma_\varepsilon^2)$	0.048	0.046	0.060	0.052	0.054	0.056	0.054	0.060
Logistic(0,1)	0.094	0.102	0.072	0.086	0.094	0.106	0.072	0.092
Cauchy(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DE(0,1)	0.392	0.744	0.300	0.498	0.420	0.778	0.316	0.518
$t(3)$	0.486	0.800	0.394	0.688	0.508	0.810	0.412	0.706
$t(5)$	0.158	0.272	0.106	0.180	0.172	0.296	0.120	0.194
$t(10)$	0.056	0.068	0.048	0.048	0.060	0.076	0.052	0.048
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.058	0.074	0.058	0.058	0.042	0.050	0.050	0.054
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.314	0.618	0.118	0.206	0.276	0.574	0.098	0.184
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.898	1.000	0.542	0.868	0.880	1.000	0.522	0.846

Table 3.3: Simulation results of the proposed test
(Case III: $U \sim N(0, \sigma_U^2), b, h = 0.8n^{-\frac{1}{3}}$)

	$\mathcal{T}_n^*(\tilde{\theta}_n)$				$\mathcal{T}_n(\tilde{\theta}_n)$			
	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$	
	100	200	100	200	100	200	100	200
$N(0, \sigma_\varepsilon^2)$	0.048	0.048	0.060	0.042	0.048	0.054	0.056	0.044
Logistic(0,1)	0.072	0.104	0.052	0.086	0.074	0.110	0.052	0.086
Cauchy(0,1)	1.000	1.000	0.998	1.000	1.000	1.000	0.998	1.000
DE(0,1)	0.376	0.674	0.200	0.398	0.380	0.684	0.212	0.404
$t(3)$	0.496	0.802	0.356	0.616	0.508	0.806	0.366	0.630
$t(5)$	0.144	0.270	0.112	0.174	0.154	0.294	0.118	0.182
$t(10)$	0.058	0.062	0.044	0.054	0.066	0.064	0.046	0.056
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.058	0.066	0.050	0.052	0.052	0.052	0.042	0.048
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.252	0.502	0.090	0.148	0.236	0.468	0.080	0.134
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.880	0.994	0.378	0.676	0.862	0.992	0.360	0.648

Table 3.4: Simulation results of the proposed test
(Case IV: $U \sim N(0, \sigma_U^2), b, h = 1.2n^{-\frac{1}{3}}$)

	$\mathcal{T}_n^*(\tilde{\theta}_n)$				$\mathcal{T}_n(\tilde{\theta}_n)$			
	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$	
	100	200	100	200	100	200	100	200
$N(0, \sigma_\varepsilon^2)$	0.070	0.046	0.068	0.062	0.078	0.066	0.074	0.056
Logistic(0,1)	0.078	0.110	0.046	0.100	0.086	0.124	0.046	0.106
Cauchy(0,1)	1.000	1.000	0.996	1.000	1.000	1.000	0.996	1.000
DE(0,1)	0.458	0.768	0.262	0.542	0.502	0.792	0.288	0.588
$t(3)$	0.544	0.810	0.412	0.698	0.586	0.828	0.430	0.720
$t(5)$	0.150	0.272	0.114	0.198	0.178	0.304	0.132	0.226
$t(10)$	0.036	0.076	0.042	0.066	0.050	0.088	0.046	0.070
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.054	0.072	0.064	0.062	0.038	0.052	0.064	0.062
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.254	0.610	0.088	0.176	0.186	0.532	0.074	0.152
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.856	0.998	0.414	0.742	0.826	0.998	0.376	0.702

Table 3.5: Simulation results of the proposed test
(Case V: $U \sim N(0, \sigma_U^2), b = n^{-\frac{1}{5}}, h = n^{-\frac{1}{3}}$)

	$\mathcal{T}_n^*(\tilde{\theta}_n)$				$\mathcal{T}_n(\tilde{\theta}_n)$			
	$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$		$\sigma_U^2 = 0.5^2$		$\sigma_U^2 = 0.8^2$	
	100	200	100	200	100	200	100	200
$N(0, \sigma_\varepsilon^2)$	0.326	0.414	0.166	0.168	0.076	0.054	0.078	0.062
Logistic(0,1)	0.072	0.096	0.072	0.084	0.084	0.120	0.076	0.100
Cauchy(0,1)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DE(0,1)	0.426	0.764	0.272	0.542	0.532	0.832	0.342	0.636
$t(3)$	0.524	0.852	0.438	0.762	0.616	0.900	0.500	0.804
$t(5)$	0.148	0.316	0.124	0.190	0.232	0.432	0.182	0.250
$t(10)$	0.056	0.066	0.034	0.064	0.068	0.094	0.040	0.078
$0.5N(\pm 0.5, \sigma_\varepsilon^2)$	0.162	0.276	0.104	0.152	0.036	0.056	0.062	0.062
$0.5N(\pm 0.75, \sigma_\varepsilon^2)$	0.408	0.762	0.170	0.316	0.206	0.470	0.056	0.154
$0.5N(\pm 1, \sigma_\varepsilon^2)$	0.916	1.000	0.500	0.874	0.794	0.998	0.374	0.748

Chapter 4

Conclusion

In this dissertation, goodness-of-fit tests are proposed for checking the adequacy of parametric distributional forms of the regression error density functions and the error-prone predictor density function in measurement error models, when replications of the surrogates of the latent variables are available.

In Chapter 2, we proposed a goodness-of-fit test for checking the adequacy of parametric forms of the regression error density functions in linear errors-in-variables regression models. Instead of assuming the distribution of the measurement error being known, we assume that replication of the surrogates of the latent variables are available. The test statistic is based upon a weighted integration of the L_2 distance between a nonparametric estimator and a semiparametric estimator of the density functions of the residuals.

Under the null hypothesis, the test statistic was shown to be asymptotically normal (Theorem 2.2.1). Consistency (Theorem 2.3.1) and local power results (Theorem 2.3.2) of the proposed test under fixed alternatives and local alternatives were also established. Comparing these results with Koul and Song (2012)'s, in which the density function of measurement error was assumed to be known, one can see that replacing the density function of U with a kernel density estimate did not slow down the convergence rate of the test statistic. One can check the proof of Theorem 2.2.1 and find out that this is a consequence of requiring $nb^{1/2}w^4 \to 0$.

Actually, the condition $nb^{1/2}w^4 \to 0$ is required to dampen the effect of estimating $f_{\bar{U}}$ by its d-dimensional kernel estimate. Together with the bandwidth assumptions $nb \to \infty$ and $nw^d \to 0$, which are commonly used in the univariate and multivariate kernel estimation, we must have $d < 8$. Therefore, one limitation of the proposed test in Chapter 2 is that the linear errors-in-variables regression model under consideration cannot have more that 8 predictors. In the future work, we will figure out ways to alleviate this constraint. One possible methods might be considering higher order terms in the Taylor series when deriving the bias between $\hat{f}_{\bar{U}}$ and $f_{\bar{U}}$.

In Chapter 3, we proposed a class of goodness-of-fit tests for checking the parametric distribution forms of the error-prone random variables in the classic additive measurement error models. By giving up the commonly adopted assumptions of the distribution of the measurement error being known, we assumed replications of the surrogates of the error-prone variables are available. Two types of test statistics were defined based upon a weighted integrated squared distance between a nonparametric estimator and (centered or non-centered) semi-parametric estimator of the density functions of the averaged surrogate data. Under the null hypothesis, the minimum distance estimator (Theorem 3.3.3 and Theorem 3.3.4) of the distribution parameters and the test statistics (Theorem 3.4.1 and Theorem 3.4.3) are shown to be asymptotically normal. Consistency (Theorem 3.5.1 and Theorem 3.5.2) and local power (Theorem 3.5.3 and Theorem 3.5.4) of the proposed tests under fixed alternatives and local alternatives are also established.

Theorems in Chapter 3 show that the two different types of tests proposed share similar properties on asymptotic normality, consistency and local power, but under different requirement on the bandwidths. In addition to the assumptions $nb^2 \to \infty$ and $nh \to \infty$, which are commonly used in the univariate kernel smoothing estimation procedures, the non-centered test requires $nb^4 \to 0$ as well as $nh^4 \to 0$, while the centered test only requires $nh^4 \to 0$. The requirement $nb^4 \to 0$ is the consequence of considering the asymptotic bias $E\hat{f}_{\bar{Z}}(z) - f_{\bar{Z}}(z, \theta_0)$ in the non-centered test. The centered version avoided analyzing the asymptotic bias, but still require $nh^4 \to 0$ since $\hat{f}_{\bar{U}}(u)$ is used to replace $f_{\bar{U}}$ in the statistic $T_n(\theta)$. For considering the multivariate case X being d-dimensional, in the non-centered test, one would require

$nb^{2d} \to \infty$, combining this with $nb^4 \to 0$, we must have $d = 1$. The centered test has better potential of being generalized to higher dimensional case, which will be our next step of research.

Throughout this dissertation, we require the density function of the measurement error term U to be symmetric about 0. The significance of this symmetry assumption lies in the fact that U_1+U_2 and U_1-U_2 will have the same distribution. Therefore, one can estimate the distribution of U_1+U_2 by using U_1-U_2, which in turn can be estimated through Z_1-Z_2, on which we have observations. In the future research, it is worthwhile to consider developing more general tests by relaxing the symmetric assumption.

Bibliography

Beran, R., 1977. Minimum Hellinger distance estimates for parameteric models. Annals of Statistics 5 (3), 445–463.

Bickle, P., Rosenblatt, M., 1973. On some global measures of the deviations of density function estimates. Annals of Statistics 1 (6), 1071–1095.

Blas, B., Bolfarine, H., Lachos, V., 2013. Statistical analysis of controlled calibration model with replicates. Journal of Statistical Computation and Simulation 83, 941–961.

Buonaccorsi, J., 2010. Measurement Error: Models, Methods and Applications. Chapman & Hall/CRC Interdisciplinary Statistics.

Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., 2006. Measurement Error in Nonlinear Models. Chapman & Hall/CRC, Boca Raton.

Carroll, R. J., Roeder, K., Wasserman, L., 1999. Flexible parametric measurement error models. Biometrics 55 (1), 44–54.

Cheng, C., Van Ness, J. W., 1999. Statistical Regression with Measurement Error. John Wiley & Sons, Ltd.

Dalen, I., Buonaccorsi, J., Sexton, J., Laake, P., Thoresen, M., 2009. Correction for misclassification of a categorized exposure in binary regression using replication data. Statistics in Medicine 28, 3386–3410.

Delaigle, A., Hall, P., Meister, A., 2008. On deconvolution with repeated measurements. Annals of Statistics 36, 665–685.

Devanarayan, V., Stefanski, L. A., 2002. Empirical simulation extrapolation for measurement error models with replicate measurements. Statistics & Probability Letters 59 (3), 219–225.

Eckert, R., Carroll, R. J., Wang, N., 1997. Transformations to additivity in measurement error models. Biometrics 53, 262–272.

Fuller, W., 1987. Measurement Error Models. Wiley, New York.

Gao, J., Gijbels, I., 2008. Bandwidth selection in nonparametric kernel testing. Journal of the American Statistical Association 103, 1584–1594.

Gimenez, P., Patat, M., 2005. Estimation in comparative calibration models with replicate measurement. Statistics and Probability Letters 71, 155–164.

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

Holzmann, H., Bissantz, N., Munk, A., 2007. Density testing in a contaminated sample. Journal of Multivariate Analysis 98, 57–75.

Holzmann, H., Boysen, L., 2006. Integrated square error asymptotics for supersmooth deconvolution. Scandinavian Journal of Statistics 33 (4), 849–860.

Huang, X., Stefanski, L. A., Davidian, M., 2006. Latent-model robustness in structural measurement error models. Biometrika 93 (1), 53–64.

Huwang, L., 1995. Interval estimation in structural errors-in-variables model with partial replication. Journal of Multivariate Analysis 55, 230–245.

Jennrich, R., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.

Khmaladze, E., Koul, H., 2004. Martingale transforms goodness-of-fit tests in regression models. Annals of Statistics 32, 995–1034.

Khmaladze, E., Koul, H., 2009. Goodness-of-fit problems for errors in nonparametric regression. Annals of Statistics 37, 3165–3185.

Koul, H., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models. Lecture notes in Statistics, 166. Springer-Verlag, New York.

Koul, H., Ni, P., 2004. Minimum distance regression model checking. Journal of Statistical Planning and Inference 119, 109–141.

Koul, H., Song, W., 2010. Model checking in partial linear regression models with berkson measurement error. Statistica Sinica 20, 1551–1579.

Koul, H., Song, W., 2012. Bickel-Rosenblatt type goodness-of-fit tests in errors-in-variables model. Journal of French Statistics Society 153, 52–70.

Koul, H., Song, W., Zhu, X., 2017. Goodness-of-fit testing of error distribution in linear measurement error models. Accepted by *Annals of Statistics*.

Laurent, B., Loubes, J., Marteau, C., 2011. Testing inverse problems: A direct or an indirect problem? Journal of Statistical Planning and Inference 141, 1849–1861.

Li, T., Vuong, Q., 1998. Nonparametric estimation of the measurement error model using multiple indicators. Journal of Multivariate Analysis 65, 139–165.

Lin, J., Cao, C., 2013. On estimation of measurement error models with replication under heavy-tailed distributions. Computational Statistics 28, 809–829.

Loubes, J., Marteau, C., 2014. Goodness-of-fit testing strategies from indirect observations. Journal of Nonparametric Statistics 26 (1), 85–99.

Rao, B. P., 1992. Identifiability in Stochastic Models. Academic Press, San Diego.

Reiersol, O., Koopmans, T. C., 1950. The identification of structural characteristics. The Annals of Mathematical Statistics 21 (2), 165–181.

Richardson, S., Leblond, L., Jaussent, I., Green, P. J., 2002. Mixture models in measurement error problems, with reference to epidemiological studies. Journal of the Royal Statistical Society: Series A (Statistics in Society) 165 (3), 549–566.

Schafer, D. W., 2001. Semiparametric maximum likelihood for measurement error model regression. Biometrics 57 (1), 53–61.

Thompson, J. R., Carter, R. L., 2007. An overview of normal theory structural measurement error models. International Statistical Review 75 (2), 183–198.

White, H., 1981. Consequences and detection of misspecified nonlinear regression models. Journal of American Statistical Association 76, 419–433.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

White, I., Frost, C., Tokunaga, S., 2001. Correcting for measurement error in binary and continuous variables using replicates. Statistics in Medicine 20 (22), 3441–3457.

Xiao, Z., Shao, J., Palta, M., 2010. Instrumental variable and gmm estimation for panel data with measurement error. Statistica Sinica 20, 1725–1747.