GOODNESS-OF-FIT TESTING OF ERROR DISTRIBUTION IN NONPARAMETRIC ARCH(1) MODELS AND LINEAR MEASUREMENT ERROR MODELS

By

Xiaoqing Zhu

A DISSERTATION

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

Statistics—Doctor of Philosophy

2015

ABSTRACT

GOODNESS-OF-FIT TESTING OF ERROR DISTRIBUTION IN NONPARAMETRIC ARCH(1) MODELS AND LINEAR MEASUREMENT ERROR MODELS

By

Xiaoqing Zhu

This thesis discusses the goodness-of-fit testing of an error distribution in a nonparametric autoregressive conditionally heteroscedastic model of order one and in the linear measurement error model.

For the nonparametric autoregressive conditionally heteroscedastic model of order one, the test is based on a weighted empirical distribution function of the residuals, where the residuals are obtained from a local linear fit for the autoregressive and heteroscedasticity functions, and the weights are chosen to adjust for the undesirable behavior of these nonparametric estimators in the tails of their domains. An asymptotically distribution free test is obtained via Khmaladze martingale transformation. A simulation study is included to assess the finite sample level and power behavior of this test. It exhibits some superiority of this test compared to the classical Kolmogorov-Smirnov and Cramér-von Mises tests in terms of the finite sample level and power.

For the linear measurement error model, a class of test statistics are based on the integrated square difference between the deconvolution kernel density estimators of the regression model error density and a smoothed version of the null error density, an analog of the so called Bickel and Rosenblatt test statistics. The asymptotic null distributions of the proposed test statistics are derived for both the ordinary smooth and super smooth cases. The asymptotic powers of the proposed tests against a fixed alternative and a class of local nonparametric alternatives for both cases are also described. A finite sample simulation study shows some superiority of the proposed test compared to some other tests.

To my beloved parents, Shijun Zhu and Guangxia Lv, my brother, Yingming Zhu, and my boyfriend, Silong Zhang.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude and appreciation to my advisor Dr. Hira L. Koul for his continuous support for my Ph.D. study and research, for his patient guidance, encouragement, valuable criticism, and immense knowledge. His enlightening ideas, constructive comments, numerous advises and uncountable time he spent make it possible for me to finish this dissertation. His enthusiasm in research, and optimistic and energetic attitude will help me in my future life.

Besides of my advisor, I would also like to thank Dr. Lyudmila Sakhanenko, Dr. David Todem, and Dr. Yimin Xiao for serving as members of my doctoral committee and their invaluable suggestions. My sincere thanks also go to Dr. Shiyuan Zhong from the Department of Geography at MSU, for offering me collaboration opportunities in her group and providing me chances to work in diverse exciting projects. I would also like to thank Dr. Weixing Song from the Department of Statistics at Kansas State University for his interesting course on measurement error models. I would like to express my deepest appreciation to Dr. Zhiying Wen from Tsinghua University for his continuous support and invaluable encouragement all these years.

Thanks also go to the entire faculty and staff members in the Department of Statistics and Probability who have taught me and helped me during my study at MSU. My special thanks go to Dr. Yimin Xiao for his interesting courses, valuable advise and encouragement.

Thanks to the graduate school, the College of Natural Science and the Department of Statistics and Probability who provided me the Dissertation Continuation Fellowship and traveling fellowships for working on my thesis and attending conferences. This dissertation is also supported in part by the grant NSF DMS 1205271, P.I. Hira L. Koul.

I would also like to thank my academic family members: Liqian Cai, Bin Gao, Tao He, Abhishek Kaul, Lisi Pei, Dr. Xin Qi, Honglang Wang, Yuzhen Zhou and all other students from the Department of Statistics and Probability for the numerous discussions and all the fun we had in the last four years.

Last but not least, I would express my profound gratitude to my beloved parents, Shijun Zhu and Guangxia Lv, my older brother, Yingming Zhu and my boyfriend, Silong Zhang for their love, endless support and encouragements all these years.

TABLE OF CONTENTS

LIST OF TABLES

KEY TO ABBREVIATIONS

- ARCH(1): autoregressive conditionally heteroscedastic model of order 1.
- Gof: Goodness-of-fit.
- KS: Kolmogorov-Smirnov.
- CvM: Cramér-von Mises.
- KK: Khmaldaze and Koul (2009).
- MSW: Müller, Schick, and Wefelmeyer (2012).
- d.f.: distribution function.
- i.i.d: independent and identically distributed

Chapter 1

Introduction

One of the classical problems of statistical inference is to test if a given random sample comes from a given continuous distribution. This is the so called goodness-of-fit testing problem. A well known test for this problem is Kolmogorov's tests based on empirical distribution function, which is asymptotically distribution free. This is desirable because it makes this test implementable for large or moderate sample sizes. This property is lost as soon as there is a nuisance parameter present in the testing problem as happens to be case when, for example, one is fitting a given distribution up to an unknown location parameter or up to unknown location and scale parameters.

Similarly, analogous tests based on the residual empirical process in regression or in autoregressive conditionally heteroscedastic time series models are not asymptotically distribution free for fitting a known distribution function (d.f.) to the error d.f. One way to obtain asymptotically distribution free tests from residual empirical process in these models is to base tests on its Khmaladze (1981) martingale transform. This has been successfully done in parametric and non-parametric regression models in Khmaladze and Koul (2004, 2009). Müller, Schick and Wefelmeyer (2012) developed analogous transformation test based on certain weighted residual empirical process for fitting a known error d.f. in nonparametric autoregressive time series models of order 1. Chapter 2 of this thesis pertains to developing and analyzing analogously transformed process for fitting a known error d.f. to the error d.f. in nonparametric autoregressive conditionally heteroscedastic time series models of order

1. The supremum test based on this transform is asymptotically distribution free. A finite sample study shows accuracy of the asymptotic null distribution of this test, and that its empirical power dominates that of the Kolmogorov test based on weighted residual empirical process at all chosen alternatives, levels, and sample sizes.

Another way to obtain asymptotically distribution free tests in these problems is to assume densities exist and use nonparametric estimators of densities to fit a given density. For fitting a known density in the one sample set up, Bickel and Rosenblatt (1973) were the first to investigate the asymptotic null distribution of a test based on a L_2-distance between a kernel type density estimator and its null expected value. The asymptotic null distribution of a suitably standardized version of this statistics was shown to be standard Gaussian. Since then numerous papers have appeared proposing tests based on analogs of this statistics in various models having some nuisance parameters. A desirable property of this statistics is that its asymptotic null distribution is not affected by not knowing the nuisance parameters in the one sample location-scale models. Lee and Na (2002), Bachmann and Dette (2005), Horvath and Zitikis (2006) and Koul and Mimoto (2012) observed that this fact continues to hold for the analog of this statistics when fitting an error density based on residuals in parametric autoregressive and generalized autoregressive conditionally heteroscedastic time series models. A similar fact has been observed to hold by Ducharme and Lafaye de Micheaux (2004) in parametric autoregressive moving average models, by Cheng and Sun (2008) in parametric nonlinear autoregressive time series models, by Bercu and Portier (2008) for multivariate ARMAX models in adaptive tracking, and by Na (2009) for infinite-order autoregressive models.

The regression models where covariates are not directly observable are abound in real world applications as is evidenced by the three monographs of Fuller (1987), Carroll, Ruppert and Stefanski (1995), and Cheng and Van Ness (1999). In these models one observes a surrogate for the covariates with some error. These are known as measurement error regression models or errors-in-variables regression models. Statistical inference in these models is highly sensitive to the knowledge of the error distributions. Knowing the regression model error distribution can help to develop efficient inference for the underlying parameter in these models. It is thus of interest to develop goodness-of-fit tests for fitting a known error density to the regression model error density in the presence of measurement error in the covariates. Chapter 3 of this thesis pertains to developing a goodness-of-fit tests for this testing problem in linear measurement error regression models. The test statistics are of the above L_2 type distance based on a class of deconvoluted error density estimators and the smoothed version of null error density. Two types of tail properties of the measurement error distribution are considered, which are the ordinary smooth case and super smooth case. For each case, a comprehensive theoretical analysis of the asymptotic distributions of these statistics under null hypothesis, under a fixed alternative and under a sequence of local nonparametric alternatives is presented. A member of this class of tests is compared via a finite sample simulation with some other tests. It dominates several of these tests in terms of the power at the chosen alternatives when the measurement error is large.

Chapter 2

Nonparametric ARCH(1) Models

2.1 Introduction

In recent years, there has been a considerable focus for providing asymptotically distribution free tests for fitting a known error distribution in regression and autoregressive and moving average models. Boldin (1982, 1990), Koul (1991, 2002), Khmaldaze and Koul (2004), Koul and Ling (2006), among others, focus on tests based on residual empirical distribution function (d.f.) in parametric cases. Khmaldaze and Koul (2009) provide martingale transform tests based on residual empirical d.f. for nonparametric regression models, and Müller, Schick, and Wefelmeyer (2012) provide similar tests fitting an error distribution in semiparametric partially linear regression models.

The focus of the present chapter is to analyze an analog of the above tests for fitting an error distribution in nonparametric autoregressive conditionally heteroscedastic models of order 1 (ARCH(1)). One of the main problems faced here is the construction of the nonparametric residuals so that the corresponding residual empirical d.f. obeys uniform asymptotic linearity expansion up to the first order. Müller et al. (2009) obtained this type of a result for nonparametric homoscedastic autoregressive time series models of order 1. In this chapter we extend this result to a class of ARCH(1) models.

The chapter is organized as follows. In section 2, we introduce the local linear estimators of autoregressive and variance functions and state their uniform strong consistency. The asymptotic uniform linear expansion of a suitably standardized weighted residual empirical process based on the corresponding residuals, and the asymptotic distributions of the test based on the martingale transform of these weighted residual empirical processes are established in section 3. Several examples of error d.f.'s where the results of this chapter are applicable are also discussed in section 3. A simulation study of section 4 shows that the finite sample power of the martingale transform test is uniformly higher than that of the Kolmogorov-Smirnov test based on a weighted residual empirical process at all chosen alternatives. This finding is consistent with that reported in Khmaladze and Koul (2009) (KK) when dealing with nonparametric regression models. The same simulation study also shows some superiority of the proposed test over the Cramér-von Mises based on a weighted residual empirical process in terms of the finite sample level and power at the chosen alternatives.

The proofs of some technical results pertaining to nonparametric estimators of autoregressive and heteroscedasticity functions and those of the asymptotic uniform linearity of the weighted residual empirical process are deferred to the last section of this chapter, section 2.5.

One of the novelties of this chapter is in the implementation of the Khmaladze martingale transform test in ARCH(1) models even when the incomplete Fisher information matrix is singular. In the location set up alone this matrix is known to be singular for double exponential error distribution. In this chapter we note that this matrix is singular also for a class of t-distributions in the present location-scale context, which is unlike in the location set up where it is nonsingular as was noted in KK.

2.2 Autoregressive and Variance Functions Estimation

Consider the nonparametric ARCH model of order 1

$$X_i = m(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i, \quad i \in \mathbb{Z} := \{0, \pm 1, \cdots\}, \tag{2.2.1}$$

where $\varepsilon_i, i \in \mathbb{Z}$ are independent copies of a standardized random variable (r.v.) ε, and ε_i is independent of X_{i-1}, for all $i \in \mathbb{Z}$. Note that then $m(x) = E(X_i|X_{i-1} = x)$, and $\sigma^2(x) = E\{(X_i - m(X_{i-1})^2|X_{i-1} = x\}$, $x \in \mathbb{R}$, $i \in \mathbb{Z}$.

Let F be a known d.f. We are interested in testing the hypothesis that the d.f. of ε is F. Any test of such a hypothesis has to be based on the estimated residuals, which in turn needs suitable estimators of the nonparametric functions m and σ.

Several researchers have investigated numerous nonparametric estimators of m and σ in regression and autoregressive models. In order to use these estimators in the above testing problem, one needs their uniform consistency. For homoscedastic regression models with bounded dependent variable, Ojeda (2008) established the Hölder continuity properties of the local polynomial estimators of the regression function for the one dimensional covariate case. For heteroscedastic regression models, Neumeyer and Van Keilegom (2010) established the uniform consistency of the local polynomial estimators for the regression and variance functions in the case of multidimensional covariates. To estimate the variance function, they use the estimators of the type $\hat{a} - \hat{m}^2$ (see also Yao and Tong (1994), where $\hat{a}(x)$ and $\hat{m}(x)$ are estimators of $E(Y^2|X = x)$ and $m(x)$, respectively. For homoscedastic autoregressive models, Masry (1996) proved the uniform consistency over compact sets of multivariate local polynomial estimators of the autoregressive function, provided the time series is $\alpha-$mixing. For stationary and ergodic auto-regressive time series of order 1, MSW proved the uniform consistency over a sequence of compact intervals increasing to $\mathbb{R}$ of the local linear estimators of the autoregressive function. For the one dimensional $\alpha-$mixing time series model, Neumeyer and Selk (2013) proved the uniform consistency over a sequence of compact intervals increasing to $\mathbb{R}$ of the Nadaraya-Waston estimators for autoregressive and variance functions. Fan and Yao (1998) provided the asymptotic properties for an efficient fully-adaptive estimator for the variance function, i.e. the local linear estimator of $E((Y-\hat{m}(X))^2|X=x)$ in the one dimensional $\beta-$mixing case. Different from the mixing condition, Wu, Huang and Huang (2010) gave a moment contracting condition for the dependence properties of a general autoregressive model, and established the uniform consistency for the Nadaraya-Waston type estimators of the autoregressive function over a bounded compact set. Based on the moment contracting condition for one dimensional stationary autoregressive model, Borkowski and Mielniczuk (2012) established the asymptotic distributional properties of the efficient fully-adaptive local linear estimator of the variance function $E((Y-\hat{m}(X))^2|X=x)$.

To proceed further, we now define the estimators of interest here. Let K and W be density kernel functions and h_1 and h_2 be the bandwidths. Define

$$(\hat{a}_0(x), \hat{b}_0(x)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \Big\{X_i - \alpha - \beta(X_{i-1}-x)\Big\}^2 K\Big(\frac{X_{i-1}-x}{h_1}\Big), \quad x \in \mathbb{R}. \quad (2.2.2)$$

Note that $\hat{a}_0(x)$ and $\hat{b}_0(x)$ are the local linear estimators of $m(x)$ and the first derivative $\dot{m}(x)$ of $m(x)$, respectively. Henceforth, $\hat{m}(x) = \hat{a}_0(x)$. To estimate $\sigma^2(x)$, we shall consider the following two methods. The first one is based on Yao and Tong (1994), where $\hat{\sigma}^2(x) \equiv \hat{\sigma}_1^2(x) = \hat{a}_1(x) - \hat{m}^2(x)$, and

$$(\hat{a}_1(x), \hat{b}_1(x)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \Big\{X_i^2 - \alpha - \beta(X_{i-1}-x)\Big\}^2 W\Big(\frac{X_{i-1}-x}{h_2}\Big). \quad (2.2.3)$$

The second estimator is based on the work of Fan and Yao (1998), who suggested an efficient fully-adaptive procedure, $\hat{\sigma}^2(x) \equiv \hat{\sigma}_2^2(x) = \hat{a}_2(x)$, where

$$(\hat{a}_2(x), \hat{b}_2(x)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} \Big\{\hat{r}_i - \alpha - \beta(X_{i-1} - x)\Big\}^2 W\Big(\frac{X_{i-1} - x}{h_2}\Big). \tag{2.2.4}$$

Here $\hat{r}_i = [X_i - \hat{m}(X_{i-1})]^2$. We shall show that both of these estimators of $\sigma^2(x)$ yield the same asymptotic result for the proposed goodness-of-fit tests under similar conditions. Here we shall present some consistency results about these estimators. In order to do so, we need some assumptions as follows.

In the sequel, for any twice differentiable function g, $\dot{g}$ and $\ddot{g}$ represent the first and second derivatives of g, respectively. All limits are taken as $n \to \infty$, unless specified otherwise.

Assumptions:

(E) There exists some $b > 1 + \sqrt{3}$ such that $E[|X_0|^{2b}] < \infty$ and $E[|\varepsilon_1|^{2b}] < \infty$.

(F) The innovation ε_j, $j \in \mathbb{Z}$, are i.i.d. F. The density f of F is continuously differentiable and $\sup_{x\in\mathbb{R}} |xf(x)| < \infty$ as well as $\sup_{x\in\mathbb{R}} |x^2 \dot{f}(x)| < \infty$.

(H) The sequence of bandwidths $h_i = \alpha_i c_n$, $i = 1, 2$, $\alpha_i > 0$, $c_n \to 0$ and

$$(\log n)^\eta/(nc_n^{2+\sqrt{3}}) \to 0, \quad nc_n^4(\log n)^\eta \to 0, \quad \forall\, \eta > 0. \tag{2.2.5}$$

If $\hat{\sigma}_2^2(x)$ is used, c_n also satisfies,

$$(\log n)^\eta/(nc_n^{3.8}) \to 0, \quad \forall\, \eta > 0. \tag{2.2.6}$$

(I) The two sequences of real numbers a_n, b_n satisfy the following conditions: $a_n < 0 < b_n$,

$-a_n$ and b_n tend to infinity such that for an $0 \leq r_1 < \infty$, $(b_n - a_n) = O((\log n)^{r_1})$, and $P(X_0 \leq a_n + \lambda) + P(X_0 \geq b_n - \lambda) = o((\log n)^{-1})$, for any $\lambda > 0$.

(KZ) For the $\alpha-$mixing process, the kernel density K is supported on $[-1, 1]$, symmetric around 0 and three times differentiable, with all three derivatives bounded. Moreover $K(1) = \dot{K}(1) = 0$. The kernel W satisfies the same conditions.

(KZ′) For the geometric moment contracting process, the kernel density K is supported on $[-1, 1]$, symmetric around 0 and three times continuously differentiable. The kernel W satisfies the same conditions.

(M) The functions m and σ are four times differentiable and there exist constants $0 < d_1 < d_2 < \infty$, $0 \leq r_q, r_s < \infty$, and sequences $q_n, q_{n,\sigma}$ such that for all sufficiently large n, $d_1 < q_n < d_2(\log n)^{r_q}$, $d_1 < q_{n,\sigma} < d_2(\log n)^{r_s}$, $\sup_{x\in[a_n-h_1,\, b_n+h_1]} |m^{(k)}(x)| = O(q_n)$, and $\sup_{x\in[a_n-h_2,\, b_n+h_2]} |\sigma^{(k)}(x)| = O(q_n)$, $k = 0, 1, 2, 3, 4$, and $(\inf_{x\in I_n} |\sigma(x)|)^{-1} = O(q_{n,\sigma})$, where h_1, h_2 are as in (H) above.

(X) The observations X_j, $j \in \mathbb{Z}$ have a common marginal density g, which is bounded and four times differentiable with bounded derivatives. The density is also bounded away from zero on compact intervals. There exists some $0 \leq r_g < \infty$ such that $q_{n,g} = (\inf_{x\in I_n} g(x))^{-1} = O((\log n)^{r_g})$, where $I_n := [a_n, b_n]$, with a_n, b_n as in Assumption (I).

(Z) The process $(X_j)_{j\in\mathbb{Z}}$ is α-mixing with mixing-coefficient $\alpha(n) = O(n^{-\kappa})$, for some

$$\kappa > \max\Big(2\frac{(3+\sqrt{3})b + 2 + \sqrt{3}}{(1+\sqrt{3})b - 2(2+\sqrt{3})}, 7\Big).$$

Moreover, $\sup_{x\in\mathbb{R}}((|m(x)|+|\sigma(x)|)^{2k})g(x)<\infty$, and there exists a $j^*\geq 1$ such that

$$\sup_{x,x'\in\mathbb{R}}\left((|m(x)|+|\sigma(x)|)^k(|m(x')|+|\sigma(x')|)^k g_{X_0,X_{j-1}}(x,x')\right)<\infty,\quad \forall\, j>j^*+1,$$

for $k=1,2$, where $g_{U,V}$ denotes joint density of any two r.v's (U,V).

(Z′) $X_n=\mathcal{J}(\cdots,\varepsilon_{n-1},\varepsilon_n)$, which is a $\sigma-$field generated by $\cdots,\varepsilon_{n-1},\varepsilon_n$. Also $(X_t)_{t\in\mathbb{Z}}$ is geometric moment contracting, i.e. let $\|Y\|_p=(E|Y|^p)^{1/p}$, for $n>0$, some $q>1$ and $0<r<1$, $\|X_n-X_n^*\|_q=O(r^n)$, where $X_n^*=\mathcal{J}(\cdots,\varepsilon_{-1},\varepsilon_0^*,\cdots,\varepsilon_{n-1},\varepsilon_n)$ and ε_0^* is an independent copy of ε_0.

The above assumptions (E), (F), (H), (I), (KZ), (M), (X) and (Z) are similar to the conditions in Neumeyer and Selk (2013) for the mixing processes. Assumption (Z') is similar as in Borkowski and Mielniczuk (2012) when the process satisfies the moment contracting condition, and the kernel conditions (KZ′) are similar to those in Müller, Schick, and Wefelmeyer (2009) (MSW). The relation (2.2.6) in assumption (H) is needed only for the analysis of $\hat{\sigma}_2^2(x)$.

We are now ready to state a uniform consistency result for the above estimators of m and σ^2. Its proof is deferred to the last section. Throughout the chapter, $I_n:=[a_n,b_n]$, with a_n,b_n as in Assumption (I).

Lemma 2.2.1 *Suppose (2.2.1), (F), (H), (I), (KZ) or (KZ′), (X), (Z) or (Z′), and (M) hold. Then*

$$\sup_{x\in I_n}\left|\frac{\hat{m}(x)-m(x)}{\sigma(x)}\right|=O_p\Big(\Big(c_n^{-1/2}n^{-1/2}(\log n)^{1/2}\Big)Q_n\Big), \tag{2.2.7}$$

$$\sup_{x\in I_n}\left|\frac{\hat{\sigma}_i(x)-\sigma(x)}{\sigma(x)}\right|=O_p\Big(\Big(c_n^{-1/2}n^{-1/2}(\log n)^{1/2}\Big)Q_n^2\Big),\quad i=1,2, \tag{2.2.8}$$

where $Q_n = q_n q_{n,g} q_{n,\sigma}$.

The next section describes the proposed weighted empirical d.f. $\hat{\mathbb{F}}$, the Khmaladze martingale transform test based on $\hat{\mathbb{F}}$, its asymptotic distribution under null hypothesis and computation of the test statistics for several distributions.

2.3 Goodness-of-fit Tests

2.3.1 Asymptotic expansion for the weighted empirical distribution function

To begin with we need to introduce the weighted residual empirical d.f. Unlike in the regression case, MSW noted that the dependency and unboundedness of the observations create some technical difficulties in autoregressive time series models because of the poor performances of the estimator $\hat{m}(x)$ for large values of x. They used only those residuals $\hat{\varepsilon}_j = X_j - \hat{m}(X_{j-1})$, for which X_{j-1} falls in the interval $I_n = [a_n, b_n]$. Analogously, we use the following weighted residual empirical process.

Fix a $\lambda > 0$. Let $\omega_n(x) \in (0, 1)$ be a sequence of functions arbitrarily defined for x in the intervals $[a_n, a_n + \lambda)$ and $(b_n - \lambda, b_n]$. In addition, assume that $\omega_n(x)$ is three times differentiable in x with uniformly three bounded derivatives, i.e., $\sup_{n \in \mathbb{N}} \sup_{x \in \mathbb{R}} |\omega_n^{(j)}(x)| < \infty$, $j = 1, 2, 3$, and satisfies

$$\omega_n(x) = \begin{cases} 1, & x \in [a_n + \lambda, b_n - \lambda], \\ 0, & x \notin [a_n, b_n]. \end{cases} \tag{2.3.1}$$

Let $\omega_{nj} = \omega_n(X_{j-1})$ and

$$\bar{\omega}_j = \frac{\omega_{nj}}{\sum_{i=1}^n \omega_{ni}}, \quad j = 1, \cdots, n.$$

Let $\hat{\varepsilon}_j := (X_j - \hat{m}(X_{j-1}))/\hat{\sigma}(X_{j-1})$, where $\hat{\mu}, \hat{\sigma}$ are as in the previous section. Then the weighted residual empirical d.f. of interest is

$$\hat{\mathbb{F}}(x) = \sum_{j=1}^n \bar{\omega}_j I(\hat{\varepsilon}_j \le x), \quad x \in \mathbb{R}. \tag{2.3.2}$$

We also need the empirical d.f. based on the true errors

$$F_n(x) = \frac{1}{n}\sum_{j=1}^n I[\varepsilon_j \le x], \quad x \in \mathbb{R}.$$

For the one dimension autoregressive homoscedastic regression model, where $\hat{\varepsilon}_j = X_j - \hat{m}(X_{j-1})$, MSW established, under the null hypothesis and under some conditions, that

$$\sup_{x\in\mathbb{R}} |\hat{\mathbb{F}}(x) - F_n(x) - f(x)\frac{1}{n}\sum_{j=1}^n \varepsilon_j| = o_p(n^{-1/2}).$$

Neumeyer and Selk (2013) obtained an analogous result for the ARCH(1) model (2.2.1) by using nonparametric residuals based on Nadaraya-Waston type estimators of autoregressive and variance functions. Under some conditions, they proved that

$$\sup_{x\in\mathbb{R}} \left|\hat{\mathbb{F}}(x) - F_n(x) - f(x)\frac{1}{n}\sum_{j=1}^n [\varepsilon_j + \frac{x}{2}(\varepsilon_j^2 - 1)]\right| = o_p(n^{-1/2}). \tag{2.3.3}$$

Theorem 2.3.1 below shows that this result continues to hold when residuals are based on the local linear fitting of $m(x)$ and $\sigma^2(x)$ as defined in (2.2.2)–(2.2.4).

Theorem 2.3.1 *Under the assumptions (2.2.1), (E), (F), (H), (I), (KZ) or (KZ′), (M), (X) and (Z) or (Z′), (2.3.3) continues to hold.*

2.3.2 Khmaladze martingale transformation

The classical tests for the goodness-of-fit testing of an error distribution are the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CvM) tests. Using the the asymptotic expansion (2.3.3) we readily obtain the following

Corollary 2.3.1 *Under the conditions of Theorem 2.3.1,*

$$\begin{aligned}
\mathrm{KS} &= n^{1/2} \sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}(x) - F(x)| \to_d \sup_{x \in \mathbb{R}} |R(x)|, \\
\mathrm{CvM} &= n \int (\hat{\mathbb{F}}(x) - F(x))^2 d\hat{\mathbb{F}}(x) \to_d \int R^2(x) dF(x),
\end{aligned}$$

where $R(x)$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned}
\mathrm{Cov}(R(x_1), R(x_2)) &= E\Big\{\big[I(\varepsilon \le x_1) - F(x_1) + f(x_1)(\varepsilon + \frac{x_1}{2}(\varepsilon^2 - 1))\big] \\
&\qquad \times \big[I(\varepsilon \le x_2) - F(x_2) + f(x_2)(\varepsilon + \frac{x_2}{2}(\varepsilon^2 - 1)\big]\Big\}.
\end{aligned}$$

Clearly these limiting null distributions depend on F in a complicated fashion and to date no theoretical results about their quantiles are available, which makes it impractical to implement these tests in practice, even for large samples. Instead, we propose to use the Khmaladze martingale transformation of $\hat{\mathbb{F}}$ to obtain asymptotically distribution free tests.

To proceed further, as in KK, assume F has an absolutely continuous density f with almost derivative $\dot{f}$. Let $\psi_f(x) = -\dot{f}(x)/f(x)$. We assume

$$I(f) = \int \psi_f^2(x) dF(x) = \int \left(\frac{\dot{f}}{f}\right)^2 dF < \infty. \tag{2.3.4}$$

Note that $E\varepsilon^2 < \infty$ and (2.3.4) imply

$$\int [x\psi_f(x) - 1]^2 dF(x) < \infty. \tag{2.3.5}$$

Thus (2.3.4) and (2.3.5) guarantee the finiteness of the Fisher information for location and scale parameters.

Consider the extended score function vector $h(x) = (1, \psi_f(x), x\psi_f(x) - 1)^T$, for location-scale family $F((y-\theta)/\sigma)$ with respect to both θ and σ, at $\theta = 0$ and $\sigma = 1$. Define the incomplete information matrix

$$\begin{aligned}\Gamma_{F(x)} &= \int_x^\infty h(y)h^T(y)dF(y) \\ &= \begin{pmatrix} 1 - F(x) & f(x) & xf(x) \\ f(x) & \int_x^\infty (\dot{f}^2(y)/f(y)dy & \int_x^\infty (f(y) + y\dot{f}(y))\dot{f}(y)/f(y)dy \\ xf(x) & \int_x^\infty (f(y) + y\dot{f}(y))\dot{f}(y)/f(y)dy & \int_x^\infty (f(y) + y\dot{f}(y))^2/f(y)dy \end{pmatrix}.\end{aligned}$$

Suppose $\Gamma_{F(x)}$ is nonsingular, for all $x \in \mathbb{R}$, and define, as in KK, for a signed measure v,

$$K(x, v) = \int_{-\infty}^x h^T(y)\Gamma_{F(y)}^{-1} \int_y^\infty h(z)dv(z)dF(y), \quad x \in \mathbb{R}. \tag{2.3.6}$$

If we define a vector function

$$H(x) = \int_{-\infty}^{x} h dF = (1 - F(x), -f(x), -xf(x))^T,$$

then analogous to (2.4) of KK, we obtain

$$H^T(x) - K(x, H^T) = 0, \quad \forall\, x \in \mathbb{R}. \tag{2.3.7}$$

Let

$$\hat{v}_n(x) = \sqrt{n}[\hat{\mathbb{F}}(x) - F(x)], \qquad v_n(x) = \sqrt{n}[F_n(x) - F(x)], \qquad x \in \mathbb{R}.$$

The Khmaladze martingale transformed processes $\widehat{\mathcal{U}}_n$ and $\mathcal{U}_n$ are defined as

$$\widehat{\mathcal{U}}_n(x) = \sqrt{n}[\hat{\mathbb{F}}(x) - K(x, \hat{\mathbb{F}})] = \hat{v}_n(x) - K(x, \hat{v}_n), \tag{2.3.8}$$

$$\mathcal{U}_n(x) = \sqrt{n}[F_n(x) - K(x, F_n)] = v_n(x) - K(x, v_n).$$

Based on the asymptotic expansion (2.3.3), we can rewrite

$$\widehat{\mathcal{U}}_n(x) = \mathcal{U}_n(x) + \eta_n(x), \qquad \eta_n(x) = \xi_n(x) - K(x, \xi_n),$$

$$\xi_n(x) = \hat{v}_n(x) - v_n(x) - f(x)\frac{1}{\sqrt{n}}\sum_{j=1}^{n}[\varepsilon_j + \frac{x}{2}(\varepsilon_j^2 - 1)],$$

$$\sup_x |\xi_n(x)| = o_p(1).$$

If the matrix $\Gamma_{F(x)}$ is singular, then $\Gamma_{F(x)}^{-1}$ cannot be uniquely defined. But, the above transformation is still well defined as is evidenced in the following lemma. This lemma is an extension of Lemma 2.1 of KK, suitable for the location-scale set up. As mentioned in KK, it is an adaptation and simplification of a more general argument presented in Nikabadze (1987) and Tsigroshvili (1998).

Lemma 2.3.1 *Suppose, for some x_0, such that $0 < F(x_0) < 1$, the matrix $\Gamma_{F(x)}$, for $x > x_0$ degenerates to the form*

$$\Gamma_{F(x)} = (1 - F(x)) \begin{pmatrix} 1 & 1 & x \\ 1 & 1 & x \\ x & x & x^2 + 1 \end{pmatrix}, \quad \forall\, x > x_0, \tag{2.3.9}$$

or

$$\Gamma_{F(x)} = (1 - F(x)) \begin{pmatrix} 1 & \frac{k}{x} & k \\ \frac{k}{x} & \frac{(k+1)^2 k}{(k+2)x^2} & \frac{k^2}{x} \\ k & \frac{k^2}{x} & k^2 \end{pmatrix}, \quad \forall\, x > x_0, \quad \textit{some} \quad k > 0. \tag{2.3.10}$$

Then in both cases, the equalities (2.3.7) and, hence, (2.3.8) are still valid. Besides, for (2.3.9),

$$h^T(x)\Gamma^{-1}_{F(x)} \int_x^\infty h(y)dv_n(y) = -\frac{2v_n(x) - \int_x^\infty v_n(y)dy}{1 - F(x)}, \quad x \in \mathbb{R}; \tag{2.3.11}$$

for (2.3.10),

$$h^T(x)\Gamma^{-1}_{F(x)} \int_x^\infty h(y)dv_n(y) = -\frac{(k+1)}{k} \frac{2v_n(x) + (k+2)x \int_x^\infty \frac{v_n(y)}{y^2} dy}{1 - F(x)}, \quad x \in \mathbb{R}. \tag{2.3.12}$$

The conclusions (2.3.11) and (2.3.12) continue to hold with v_n replaced by $\hat{v}_n$.

Proof. The proof of this lemma is similar to that of Lemma 2.1 of KK, which was proved for the location model only where the analog of Γ is 2×2. In the present set up Γ is 3×3 matrix, which creates some complexity. For the sake of self containment and completeness, we give details here to deal with this situation.

When $\Gamma_{F(x)}$ is degenerate of the form (2.3.9), $h(x) = (1, 1, x-1)^T$. The image of the linear operator in $\mathbb{R}^3$ of $\Gamma_{F(x)}$ is

$$\begin{aligned} \mathcal{I}(\Gamma_{F(x)}) &= \{b : b = \Gamma_{F(x)} a, \text{ for some } a \in \mathbb{R}^3\} \\ &= \{b : b = (1 - F(x))(\beta, \beta, \beta x + \gamma), \beta, \gamma \in \mathbb{R}\}, \end{aligned}$$

and the kernel of this operator is

$$\begin{aligned} \mathcal{K}(\Gamma_{F(x)}) &= \{a : \Gamma_{F(x)} a = 0\} \\ &= \{a : \ a = \alpha(1, -1, 0), \ \alpha \in \mathbb{R}\}, \end{aligned}$$

To prove the equalities (2.3.7), it suffices to show that for any $b \in \mathcal{I}(\Gamma_{F(x)})$, $a \in \mathcal{K}(\Gamma_{F(x)})$,

$$h(x)^T \Gamma_{F(x)}^{-1} \Gamma_{F(x)} (b + a) = h(x)^T (b + a).$$

Note that for any $b \in \mathcal{I}(\Gamma_{F(x)})$, $a \in \mathcal{K}(\Gamma_{F(x)})$,

$$\Gamma_{F(x)}(b + a) = \Gamma_{F(x)} b = (2\beta + \beta x^2 + \gamma x, 2\beta + \beta x^2 + \gamma x, 3\beta x + \beta x^3 + \gamma x^2 + \gamma)^T.$$

For any $g = (\lambda, \lambda, \lambda x + \eta) \in \mathcal{I}(\Gamma_{F(x)})$, if $\Gamma_{F(x)} g = \Gamma_{F(x)} b$, then

$$2\lambda + \lambda x^2 + \eta x = 2\beta + \beta x^2 + \gamma x, \qquad 3\lambda x + \lambda x^3 + \eta x^2 + \eta = 3\beta x + \beta x^3 + \gamma x^2 + \gamma.$$

From these two equations we obtain $\lambda = \beta$ and $\eta = \gamma$. Then $\Gamma^{-1}_{F(x)}$ is any linear operator on $\mathcal{I}(\Gamma_{F(x)})$ such that

$$\Gamma^{-1}_{F(x)} \Gamma_{F(x)} b = b + a_1, \quad a_1 \in \mathcal{K}(\Gamma_{F(x)}).$$

From this fact we obtain that for any $a \in \mathcal{K}(\Gamma_{F(x)})$, $h^T a = 0$,

$$h(x)^T \Gamma^{-1}_{F(x)} \Gamma_{F(x)} (b + a) = h(x)^T \Gamma^{-1}_{F(x)} \Gamma_{F(x)} b = h(x)^T (b + a_1) = h(x)^T (b + a).$$

Similarly, one proves (2.3.7) in the case $\Gamma_{F(x)}$ is degenerate of the form (2.3.10). This completes the proof of (2.3.7), which in turn yields the claims (2.3.11) and (2.3.12) for v_n and $\hat{v}_n$, in an obvious way.

Sometimes it is convenient to use the time transformation $t = F(x)$, $u_n = v_n(F^{-1}(t))$, $\hat{u} = \hat{v}_n(F^{-1}(t))$, $\gamma(t) = h(F^{-1}(t))$, and $\Gamma_t = \int_t^1 \gamma(s)\gamma(s)^T ds$, $0 \le t \le 1$. Now consider a function parametric version of the u- and u_n-processes and their transforms:

$$\begin{aligned}
&u(\varphi) = \int_0^1 \varphi(s) du(s), \qquad u_n(\varphi) = \int_0^1 \varphi(s) du_n(s), \\
&K(\varphi) = K(\varphi, u) = \int_0^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(s) du(s) dt, \\
&K_n(\varphi) = K(\varphi, u_n) = \int_0^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(s) du_n(s) dt, \\
&b(\varphi) = u(\varphi) - K_(\varphi), \qquad b_n(\varphi) = u_n(\varphi) - K_n(\varphi), \quad \varphi \in L_2[0,1].
\end{aligned}$$

Write $b(t)$ and $b_n(t)$ for $b(\varphi)$ and $b_n(\varphi(\cdot))$, respectively, when $\varphi(\cdot) = I(\cdot \leq t)$. Then

$$b(t) = u(t) - \int_0^t \gamma^T(z)\Gamma_z^{-1} \int_z^1 \gamma(s)du(s)dz, \quad t \in [0,1], \tag{2.3.13}$$
$$b_n(t) = u_n(t) - \int_0^t \gamma^T(z)\Gamma_z^{-1} \int_z^1 \gamma(s)du_n(s)dz, \quad t \in [0,1].$$

If $\Phi \subset L_2[0,1]$ is a subset of square integrable functions such that the sequence $u_n(\varphi), n \geq 1$, is uniformly in n equicontinuous on Φ, then $u_n \to_d u$ in $\ell^\infty(\Phi)$, where u is standard Brownian bridge, and $\ell^\infty(\Phi)$ is the set of all uniformly bounded real valued functions on Φ (see van der Vaart and Wellner (1996)).

The following theorem describes the weak convergence of the process $K_n(\varphi), \varphi \in \Phi$. It is an extension of Theorem 2.1 of KK, which is valid for the location model only, to the location-scale model.

Theorem 2.3.2 *(i) Let $L_{2,\varepsilon} \subset L_2[0,1]$ be the subspace of all square integrable functions which are equal to* 0 *on the interval* $(1-\varepsilon, 1]$. *Then, $K_n \to_d K$, on $L_{2,\varepsilon}$, for any* $0 < \varepsilon < 1$.

(ii) Let, for any arbitrarily small but fixed $\varepsilon > 0, C < \infty$, and $\alpha < 1/2, \Phi_\varepsilon \subset L_2[0,1]$ be a class of all square integrable functions satisfying the following right tail condition:

$$|\varphi(s)| \leq C[\gamma^T(s)\Gamma_s^{-1}\gamma(s)]^{-1/2}(1-s)^{-1/2-\alpha}, \quad \forall\, s > 1-\varepsilon. \tag{2.3.14}$$

Then, $K_n \to_d K$, on Φ_ε.

The following theorem describes the weak limit of the b_n process and is an extension of Theorem 2.2 of KK to the location-scale set up. Recall that as in van der Vaart and Wellner (1996), the family of Gaussian random variable $b(\phi)$, $\phi \in \Phi$, $\Phi \subset L_2[0,1]$, is continuous on Φ, with covariance function $Eb(\phi)b(\phi') = \int_0^1 \phi(t)\phi'(t)dt$ is called Browian motion on Φ.

Theorem 2.3.3 *(i) Let Φ be a Donsker class, that is, let $u_n \to_d u$ in $l^\infty(\Phi)$. Then, for every $\varepsilon > 0$,*

$$b_n \to_d b \quad in \quad \ell^\infty(\Phi \cap \Phi_\varepsilon),$$

where $\{b(\varphi), \varphi \in \Phi\}$ is standard Brownian motion.

(ii) If the envelop function $\Psi(t)$ of (2.3.14) tends to positive (finite or infinite) limit at $t = 1$, then for the process (2.3.13) we have

$$b_n \to_d b \quad on \quad [0, 1].$$

2.3.3 Examples

Here, we shall assess the behavior of $\gamma^T(s)\Gamma_s^{-1}\gamma(s)$, as $s \to 1$, for some well known distributions. This is needed to understand the behavior of the bound in (2.3.14), which in turn sheds some light on the class of functions φ one can use in this testing problem. Many technical details are similar to those appearing in KK when dealing with the location model only where Γ_s is 2×2 matrix. In the current set up we are dealing with the 3×3 matrix, which makes the details of derivations a bit more involved.

First, let F be standard normal d.f. Then $h(x) = (1, x, x^2 - 1)^T$. With $\zeta \equiv \zeta(x) =$

$f(x)/(1-F(x))$, we obtain

$$\Gamma_{F(x)} = (1-F(x)) \begin{pmatrix} 1 & \zeta & x\zeta \\ \zeta & x\zeta+1 & (1+x^2)\zeta \\ x\zeta & (1+x^2) & 2\zeta+(x+x^3)\zeta \end{pmatrix},$$

$$\Gamma_{F(x)}^{-1} = \frac{(1-F(x))^{-1}}{2-3\zeta^2+3x\zeta+x\zeta^3-2x^2\zeta^2+x^3\zeta} \times$$
$$\begin{pmatrix} 2-\zeta^2+3x\zeta-x^2\zeta^2+x^3\zeta & -2\zeta & \zeta^2-x\zeta \\ -2\zeta & 2+x\zeta-x^2\zeta^2+x^3\zeta & -\zeta+x\zeta^2-x^2\zeta \\ \zeta^2-x\zeta & -\zeta+x\zeta^2-x^2\zeta & 1-\zeta^2+x\zeta \end{pmatrix},$$

and

$$h^T(x)\Gamma_{F(x)}^{-1}h(x) = \frac{1}{(1-F(x))}\frac{3-4\zeta^2+4x\zeta+x^2\zeta^2+x^4-2x^3\zeta}{2-3\zeta^2+3x\zeta+x\zeta^3-2x^2\zeta^2+x^3\zeta}.$$

Using the asymptotic expansion for the tail of the normal d.f. for $\zeta(x)$ we obtain, as in KK,

$$\zeta(x) = \frac{x}{1-S(x)}, \quad \text{where} \quad S(x) = \sum_{i=1}^{n} \frac{(-1)^{i-1}(2i-1)!!}{x^{2i}} = \frac{1}{x^2} - \frac{3}{x^4} + \frac{15}{x^6} - \cdots .$$

From this one can derive that

$$\frac{3-4\zeta^2+4x\zeta+x^2\zeta^2+x^4-2x^3\zeta}{2-3\zeta^2+3x\zeta+x\zeta^3-2x^2\zeta^2+x^3\zeta} \to 9/5, \quad x\to\infty,$$

and hence $h(x)^T\Gamma_{F(x)}^{-1}h(x) \sim 9(1-F(x))^{-1}/5$, $x\to\infty$, equivalently,

$$\gamma^T(s)\Gamma_s^{-1}\gamma(s) \sim 9(1-s)^{-1}/5, \quad s\to 1.$$

This result is similar to the one obtained in KK for the location model only, where 9/5 is replaced by 2.

Next, consider logistic d.f. $F(x)$ with scale parameter 1, or equivalently $\psi_f(x) = 2F(x) - 1$. Then $h(x) = (1, 2F(x)-1, x(2F(x)-1)-1)^T$ or in terms of $s = F(x)$, $\gamma(s) = h(F^{-1}(s)) = (1, 2s-1, F^{-1}(s)(2s-1)-1)^T$, and when s is close to 1,

$$\Gamma_s \sim (1-s) \begin{pmatrix} 1 & s & xs \\ s & \dfrac{1-2s+4s^2}{3} & \dfrac{s+x(1-s)^2}{3} + xs \\ xs & \dfrac{s+x(1-s)^2}{3} + xs & 3 - 9x^2 - 6(x-3x^2)s + \dfrac{\pi^2+12x^2}{9(1-s)} \end{pmatrix}_{s=F(x)}.$$

From this formula, one can verify that $\gamma^T(s)\Gamma_s^{-1}\gamma(s) \sim (1-s)^{-1}$, for $s \to 1$. This result is different from the one reported in KK, where analogous γ and Γ satisfy $\gamma^T(s)\Gamma_s^{-1}\gamma(s) = 4(1-s)^{-1}$, for all $0 \le s < 1$.

Next, consider the double exponential d.f. with density $f(x) = e^{-|x|}/2$. For $x > 0$, we get $h(x) = (1, 1, x-1)^T$, and $\Gamma_{F(x)}$ is degenerate and equals to (2.3.9). An argument similar to the proof of Lemma 2.3.1 yields $h(x)^T\Gamma_{F(x)}^{-1}h(x) = 2(1-F(x))^{-1}$, for all $x > 0$ with $F(x) < 1$.

Finally, consider student t_k-distribution with degrees of freedom k. In this case,

$$f(x) = \frac{1}{\sqrt{\pi k}} \frac{\Gamma((k+1/2))}{\Gamma(k/2)} \frac{1}{(1+x^2/k)^{(k+1)/2}}.$$

As shown in KK, using the results of Soms (1976), for every $k \geq 1$,

$$1 - F(x) \sim \frac{1 + x^2/k}{x} f(x) \sim \frac{d_k}{k} \frac{1}{x^k}, \quad d_k = \frac{1}{\sqrt{\pi}} \frac{\Gamma((k+1/2))}{\Gamma(k/2)} k^{k/2},$$
$$f(x) \sim \frac{d_k}{x^{k+1}}, \quad \psi_f(x) = \frac{k+1}{k} \frac{x}{1 + (x^2/k)} \sim \frac{k+1}{x}, \qquad x \to \infty.$$

Hence, $h(x) = (1, \psi_f(x), x\psi_f(x) - 1)^T \sim (1, (k+1)/x, k)^T$, and $\Gamma_{F(x)}$ degenerates and has the form as in (2.3.10). This is unlike the location model case where KK observed that the analog of $\Gamma_{F(x)}$ is non-degenerate. Nevertheless, one still continues to have the same right tail behavior for the quadratic from $\gamma(s)^T \Gamma_s^{-1} \gamma(s)$ as in the location model case, viz, $\gamma(s)^T \Gamma_s^{-1} \gamma(s) \sim \{2(k+1)/k\}(1-s)^{-1}, \ s \to 1$.

2.3.4 Limiting process

In this section we discuss the weak convergence of the $\widehat{\mathcal{U}}_n$ process. Towards this goal we assume the same tail conditions for $\hat{v}_n$ as in KK, which is that for some $0 < \beta < 1/2$,

$$\sup_{y > x} \frac{|\hat{v}_n(y)|}{(1 - F(y))^\beta} = o_p(1), \quad \text{as} \quad x \to \infty, \tag{2.3.15}$$

uniformly in n. To simplify the notation, we let

$$\psi_1(x) = -\dot{f}(x)/f(x), \quad \psi_2(x) = -x\dot{f}(x)/f(x) - 1,$$

and denote the right tail mean of ψ_1 and ψ_2 by

$$E_x\psi_i = E[\psi_i(e_1)|e_1 > x], \quad \psi_{i0} = \psi_i - E_x\psi_i,$$

$$\mathrm{Var}_x(\psi_i) = \mathrm{Var}[\psi_i(e_1)|e_1 > x], \quad \mathrm{Cov}_x(\psi_1, \psi_2) = \mathrm{Cov}[\psi_1(e_1), \psi_2(e_1)|e_1 > x] \quad i = 1, 2.$$

Now we formulate three more conditions on F:

(a) For any $\varepsilon > 0$, the function $\psi_i(F^{-1})$, $i = 1, 2$, is monotone on $[1-\varepsilon, 1]$.

(b) For some $\delta > 0$, $\varepsilon > 0$ and some $C < \infty$, and for all x, such that $F(x) > 1 - \varepsilon$,

$$\begin{aligned}
\left|h^T(x)\Gamma_{F(x)}^{-1}(0, \psi_{10}(x), 0)^T\right| &= \frac{|\psi_{10}^2 \mathrm{Var}_x(\psi_2) - \psi_{10}\psi_{20}\mathrm{Cov}_x(\psi_1, \psi_2)|}{\mathrm{Var}_x(\psi_1)\mathrm{Var}_x(\psi_2) - \mathrm{Cov}_x(\psi_1, \psi_2)} \\
&\leq C(1 - F(x))^{-2\delta}, \\
\left|h^T(x)\Gamma_{F(x)}^{-1}(0, 0, \psi_{20}(x))^T\right| &= \frac{|\psi_{20}^2 \mathrm{Var}_x(\psi_1) - \psi_{10}\psi_{20}\mathrm{Cov}_x(\psi_1, \psi_2)|}{\mathrm{Var}_x(\psi_1)\mathrm{Var}_x(\psi_2) - \mathrm{Cov}_x(\psi_1, \psi_2)} \\
&\leq C(1 - F(x))^{-2\delta}.
\end{aligned}$$

Note that in terms of the above notation, with $t = F(x)$,

$$\begin{aligned}
&\gamma^T(t)\Gamma_t^{-1}\gamma(t) \\
&= \frac{1}{(1 - F(x))} \times \left[1 + \frac{\psi_{10}^2 \mathrm{Var}_x(\psi_2) + \psi_{20}^2 \mathrm{Var}_x(\psi_1) - 2\psi_{10}\psi_{20}\mathrm{Cov}_x(\psi_1, \psi_2)}{\mathrm{Var}_x(\psi_1)\mathrm{Var}_x(\psi_2) - \mathrm{Cov}_x(\psi_1, \psi_2)}\right].
\end{aligned}$$

Hence, condition (b) implies

$$\gamma^T(t)\Gamma_t^{-1}\gamma(t) \leq C(1 - t)^{-1-2\delta}, \quad \forall\, t > 1 - \varepsilon. \tag{2.3.16}$$

(c) For some $0 < C < \infty$ and $\beta > 0$ as in (2.3.15),

$$\left|\int_x^\infty [1 - F(y)]^\beta d\psi_i(y)\right| \leq C|\psi_{i0}(x)|, \quad i = 1, 2.$$

Remark 2.3.1 *As mentioned in KK, (2.3.15) also holds for v_n for any $0 < \beta < 1/2$. Conditions (a), (b) and (c) are easy to check for all the examples in Section 2.3.3 by following similar* *procedures even with $\delta = 0$ in condition (b), so we omit the details here.*

Now we consider the asymptotic behaviors for the $K(\psi, \xi_n)$, which is

$$K(\psi, \xi_n) = \int_0^1 \psi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(s)\xi_n(F^{-1}(ds))dt,$$

and for a given indexing class Φ of functions from $L_2[0,1]$. Let $\Phi \circ F = \{\varphi(F(\cdot)), \varphi \in \Phi\}$. We can prove the similar limiting process for $\widehat{\mathcal{U}}_n$ as Theorem 4.1 in KK.

Theorem 2.3.4 *(i) Suppose conditions (2.3.15) and (a)-(c) are satisfied with $\beta > \delta$. Then, on the class Φ_ε as in Theorem 2.3.2, with $\alpha < \beta - \delta$, we have*

$$\sup_{\varphi \in \Phi_\varepsilon} |K(\varphi, \xi_n)| = o_p(1), \quad n \to \infty.$$

Therefore, if Φ is a Donsker class, then, for every $\varepsilon > 0$,

$$\widehat{\mathcal{U}}_n \to_d b \quad in \quad \ell^\infty(\Phi \cap \Phi_\varepsilon \circ F),$$

where $\{b(\varphi), \varphi \in \Phi\}$ is standard Brownian motion.

(ii) If, in addition, $\delta < \alpha$, then for the time transformed process $\widehat{\mathcal{U}}_n(F^{-1}(\cdot))$ of (2.3.8),

$$\widehat{\mathcal{U}}_n(F^{-1}(\cdot)) \to_d b(\cdot) \quad in \quad D[0,1].$$

Proof. The proof below is similar to that of Theorem 4.1 in KK.

Note that

$$\begin{aligned}
\gamma^T(t)\Gamma_t^{-1}(0,a_1,0)^T &= \frac{[\psi_{10}\mathrm{Var}_x(\psi_2)-\psi_{20}\mathrm{Cov}_x(\psi_1,\psi_2)]a_1}{\mathrm{Var}_x(\psi_1)\mathrm{Var}_x(\psi_2)-\mathrm{Cov}_x(\psi_1,\psi_2)};\\
\gamma^T(t)\Gamma_t^{-1}(0,0,a_2)^T &= \frac{[\psi_{20}\mathrm{Var}_x(\psi_1)-\psi_{10}\mathrm{Cov}_x(\psi_1,\psi_2)]a_2}{(1-F(x))[\mathrm{Var}_x(\psi_1)\mathrm{Var}_x(\psi_2)-\mathrm{Cov}_x(\psi_1,\psi_2)]}.
\end{aligned}$$

The above equalities used with $a_i=\int_t^1(1-s)^\beta d\psi_i(F^{-1}(s))$, $i=1,2$, combined with conditions (b) and (c), yield

$$|\gamma^T(t)\Gamma_t^{-1}(0,a_1,a_2)^T|\le C(1-t)^{-1-2\delta},\quad \forall\, 1-\epsilon<t<1. \tag{2.3.17}$$

Now we prove the first claim.

(i) Denote $\xi_n'(t)=\xi_n(x)$ with $t=F(x)$. Because of the singularities at $t=0$ and $t=1$ in both integrals in $K(\varphi,\xi_n)$, we will isolate the neighborhood of $t=1$. The neighborhood of $t=0$ can be treated more easily. First assume $\Gamma_t>0$ for all $t<1$. Then,

$$\begin{aligned}
\int_0^1\varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_t^1\gamma(t)\xi_n'(ds)dt &= \int_0^{1-\varepsilon}\varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_t^{1-\varepsilon}\gamma(t)\xi_n'(ds)dt\\
&\quad+\int_0^{1-\varepsilon}\varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_{1-\varepsilon}^1\gamma(t)\xi_n'(ds)dt\\
&\quad+\int_{1-\varepsilon}^1\varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_t^1\gamma(t)\xi_n'(ds)dt.
\end{aligned}$$

We shall show that each of these three terms are $o_p(1)$.

First consider the third summand on the right-hand side. By definition,

$$\xi_n'(t)=\hat{u}_n(t)-u_n(t)-f(F^{-1}(t))n^{-1/2}\sum_{i=1}^n[\varepsilon_i+\frac{F^{-1}(t)}{2}(\varepsilon_i^2-1)].$$

The third summand is then sum of the two terms, one corresponding to the difference $\hat{u}_n - u_n$ and the other corresponding to the remaining term. Now, since $df(F^{-1}(s)) = \psi_f(x)f(x)dx$ and $dF^{-1}(s)f(F^{-1}(s)) = [1 + x\psi_f(x)]f(x)dx$, $F(x) = s$, then

$$\int_{1-\varepsilon}^{1} \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(t)(df(F^{-1}(s)) + dF^{-1}(s)f(F^{-1}(s)))dt$$

is the sum of the second and the third coordinate of $\int_{1-\varepsilon}^{1} \varphi(t)\gamma(t)dt$, and is small for small ε anyway. Assumption (a) guarantees the monotonicity of $\psi_f(F^{-1})$ and $dF^{-1}(s)f(F^{-1}(s))$, so the integration by parts is justified, and we obtain

$$\begin{aligned}
&\int_{1-\varepsilon}^{1} \varphi(t)\gamma^T(t)\Gamma_t^{-1} \int_t^1 \gamma(t)\hat{u}_n(ds)dt \\
&= \int_{1-\varepsilon}^{1} \varphi(t)\gamma^T(t)\Gamma_t^{-1} \Big[-\gamma(t)\hat{u}_n(t) - \int_t^1 \hat{u}_n(s)d\gamma(s)\Big] dt.
\end{aligned}$$

Using assumption (2.3.14) on φ and (2.3.16), we obtain

$$\begin{aligned}
&\Big| \int_{1-\varepsilon}^{1} \varphi(t)\gamma^T(t)\Gamma_t^{-1}\gamma(t)\hat{u}_n(t)dt \Big| \\
&\leq C \int_{1-\varepsilon}^{1} [\gamma^T(t)\Gamma_t^{-1}\gamma(t)]^{1/2} \frac{1}{(1-t)^{1/2+\alpha-\beta}} dt \sup_{t>1-\varepsilon} \frac{|\hat{u}_n(t)|}{(1-t)^{\beta}} \\
&\leq C \int_{1-\varepsilon}^{1} \frac{1}{(1-t)^{1+\alpha+\delta-\beta}} dt \sup_{t>1-\varepsilon} \frac{|\hat{u}_n(t)|}{(1-t)^{\beta}},
\end{aligned}$$

which is small for small ϵ as soon as $\alpha < \beta - \delta$.

Note that $\int_t^1 \hat{u}_n(s)d\Gamma(s) = \Big(0, \int_t^1 \hat{u}_n(s)d\psi_f(F^{-1}(s)), \int_t^1 \hat{u}_n(s)d(F^{-1}(s)\psi_f(F^{-1}(s))\Big)^T$. Using monotonicity of $\psi_f(F^{-1}(s))$ and $F^{-1}(s)\psi_f(F^{-1}(s))$ for small enough ε, we obtain, for all $t > 1-\varepsilon$,

$$\Big|\int_t^1 \hat{u}_n(s)d\psi_f(F^{-1}(s))\Big| < C\Big|\int_t^1 (1-s)^\beta d\psi_f(F^{-1}(s))\Big| \sup_{t>1-\varepsilon} \frac{|\hat{u}_n(t)|}{(1-t)^\beta}; \tag{2.3.18}$$

$$\Big|\int_t^1 \hat{u}_n(s)d(F^{-1}(s)\psi_f(F^{-1}(s)))\Big| < C\Big|\int_t^1 (1-s)^\beta d(F^{-1}(s)\psi_f(F^{-1}(s)))\Big| \sup_{t>1-\varepsilon} \frac{|\hat{u}_n(t)|}{(1-t)^\beta}.$$

Therefore, using (2.3.17), for the double integral

$$\Big|\int_{1-\varepsilon}^1 \varphi(t)\gamma^T(t)\Gamma_t^{-1}\int_t^1 \hat{u}_n(s)d\gamma(s)dt\Big| \le C\int_{1-\varepsilon}^1 (1-t)^{-1-2\delta}dt \sup_{t>1-\varepsilon} \frac{|\hat{u}_n(t)|}{(1-t)^\beta},$$

which is small as soon as $\alpha < \beta - \delta$. The same conclusion is true for $\hat{u}_n$ replaced by u_n. Since (2.3.18) implies the smallness of

$$\int_{1-\varepsilon}^1 \hat{u}_n(s)d\psi_f(F^{-1}(s)) \quad \text{and} \quad \int_{1-\varepsilon}^1 u_n(s)d\psi_f(F^{-1}(s));$$
$$\int_{1-\varepsilon}^1 \hat{u}_n(s)d(F^{-1}(s)\psi_f(F^{-1}(s))) \quad \text{and} \quad \int_{1-\varepsilon}^1 u_n(s)d(F^{-1}(s)\psi_f(F^{-1}(s))),$$

to prove that the middle summand on the right-hand side is small one needs only finiteness of $\psi_1(x), \psi_2(x)$ in each x with $0 < F(x) < 1$, which follows from (a). This and uniform in x smallness of ξ_n proves smallness of the first summand as well.

The smallness of integrals

$$\int_0^\varepsilon \varphi(t)\gamma^T(t)\Gamma_t^{-1}\gamma(t)\int_t^1 \gamma(s)\xi_n'(ds)dt,$$

follows from $\Gamma_t^{-1} \sim \Gamma_0^{-1}$ for small t, and square integrability of φ and Γ.

If Γ_t is degenerate of the form (2.3.9) for any $t > t_0$, we get

$$\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)\xi_n'(ds)dt = -\frac{2\xi_n'(t) - \int_t^1 \xi_n'(t)dt}{1-t}.$$

If Γ_t is degenerate of the type (2.3.10) for any $t > t_0$, we get

$$\gamma^T(t)\Gamma_t^{-1}\int_t^1 \gamma(s)\xi_n'(ds)dt = -\frac{(k+1)}{k}\frac{2\xi_n'(t) + (k+2)F^{-1}(t)\int_t^1 \xi_n'(t)/F^{-1}(t)^2 dt}{1-t}.$$

The smallness of all tail integrals easily follows by the tail condition (2.3.15) for our choice of the indexing functions φ.

(ii) Since for $\delta < \alpha$ the envelope function $\Psi(t)$ of (2.3.14) satisfies inequality

$$\Psi(t) \geq (1-t)^{\delta-\alpha}.$$

It has positive finite or infinite lower limit at $t = 1$. We can choose an indexing class of indicator functions $\varphi(t) = I[\tau \leq t]$ and the claim follows.

2.4 Simulations

In this section we report the findings of a simulation study. To examine the performance of the proposed test, we consider the following autoregressive and conditional variance functions

$$m(x) = \sqrt{(1/2 + x^2/2)} - 1/2, \qquad \sigma^2(x) = 3/4 + x^2/4, \quad x \in \mathbb{R}.$$

In the null hypothesis, F is the d.f. of a standardized normal r.v., as in Section 2.3.3, then $h(x) = (1, x, x^2 - 1)^T$, and $\Gamma^{-1}_{F(y)}$ is as in (2.3.15). The interval $I_n := [-\log(n), \log(n)]$. For the purpose of computation, we use the following representation of

$$\widehat{\mathcal{U}}_n(x) = n^{1/2} \sum_{i=1}^{n} \bar{\omega}_i [I(\hat{e}_i \le x) - h(\hat{e}_i)^T \mathcal{G}(x \wedge \hat{e}_i)], \quad x \in \mathbb{R},$$
$$\hat{e}_i := \hat{\varepsilon}_i I(-\log n \le X_{i-1} \le \log(n)), \quad \hat{\varepsilon}_i := (X_i - \hat{m}(X_{i-1}))/\hat{\sigma}(X_{i-1}),$$

where $\mathcal{G}(x) = \int_{y \le x} \Gamma^{-1}_{F(y)} h(y) dF(y)$. Let $\hat{e}_{(j)}, 1 \le j \le n$ denote the ordered residuals $\hat{e}_i, 1 \le i \le n$. Then $U_n := \sup_{x \in \mathbb{R}} |\widehat{\mathcal{U}}_n(x)| = \max\{\max_{1 \le j \le n} |\widehat{\mathcal{U}}_n(\hat{e}_{(j)})|, \sup_{x < \hat{e}_{(1)}} |\widehat{\mathcal{U}}_n(x)|\}$.

The asymptotic critical values of the U_n-test are the critical values of the distribution of $\sup_{0 \le t \le 1} |b(t)|$. From Khmaladze and Koul (2004) these critical values at the levels 5%, 2.5% and 1%, respectively, are 2.24241, 2.49771 and 2.80705. To compare the effect of the two estimators $\hat{\sigma}_1^2(x)$ and $\hat{\sigma}_2^2(x)$ of $\sigma^2(x)$ given at (2.2.2) and (2.2.4) on the finite sample behavior of the test, we first compared the type I error for different sample sizes obtained by computing the number of times U_n exceeded the given asymptotic critical value, divided by the number of repetitions, based on the sample sizes n = 300, 500, each repeated 1000 times. The results are displayed in Table 2.2. One sees that $\hat{\sigma}_2^2$ is more effective than $\hat{\sigma}_1^2$ in preserving the nominal level of this test.

Then we used the adaptive estimator $\hat{\sigma}_2^2$ to examine the finite sample power of the proposed Khmaladze martingale transform U_n test. The alternatives chosen are the mixture distributions of standard normal and standardized t-distribution with degree of freedom 4, i.e $(1-p)N(0,1) + pt_4/\sqrt{2}$, for $p \in [0,1]$.

We compared the U_n test with the two classical tests, KS and CvM tests. The critical values for the latter two tests are simulated by Monte Carlo method. We choose $n = 500$

and 1000 repetitions for each test. The critical values thus obtained are given in Table 2.1.

Level	KS	CvM
0.01	1.03159	0.21080
0.025	0.93812	0.17630
0.05	0.86067	0.15036

Table 2.1: Monte carlo critical values of the KS and CvM tests.

The empirical powers, i.e., the relative rejection frequencies under the chosen alternatives, for all three tests based on the sample sizes $n = 300$ and $n = 500$ with 1000 repetitions and 5%, 2.5% and 1% levels are displayed in Table 2.3. As in KK, the martingale transform test U_n again has larger empirical power than the KS test, uniformly at all chosen levels and for all values of p. Its empirical powers are also higher than those of the CvM test, at all chosen levels and for all values of p, except for $p = .8$ and $p = 1$.

In this simulation study the time series X_i was generated as follows. For each simulation, $900 + n$ observation of X_i were generated, and only the last n observations were used in the test, to ensure stationarity. The local linear estimators for $\hat{m}$ and $\hat{\sigma}^2$ were calculated using the biweight kernel function $K(x) \equiv W(x) \equiv 15(1 - x^2)^2 I(|x| \leq 1)/16$. Both the bandwidths were chosen according to the assumption by a rule of thumb as $h_1 = h_2 = 1.06 * min(sd(\hat{e}), IQR(\hat{e})/1.34) * h^{-2/(6+1.9)}$, where $\hat{e}$ is the vector of all residuals with $X_{i-1} \in I_n = [-\log n, \log n], i = 1, \cdots, n$, and IQR means the interquartile range. Let $s = (\log n - |x|)/0.1$, $x \in \mathbb{R}$. The weight function used was

$$w_n(x) = \begin{cases} 0, & x \notin [-\log n, \log n]; \\ 1, & x \in [-\log n + 0.1, \log n - 0.1]; \\ -20s^7 + 70s^6 - 84s^5 + 35s^4, & \text{otherwise.} \end{cases}$$

Table 2.2: Empirical levels of U_n test

	$\hat{\sigma}_1^2$			$\hat{\sigma}_2^2$		
Level	$n = 300$	$n = 500$	$n = 600$	$n = 300$	$n = 500$	$n = 600$
0.05	0.014	0.021	0.031	0.031	0.047	0.051
0.025	0.005	0.010	0.014	0.009	0.017	0.027
0.01	0.005	0.006	0.006	0.004	0.008	0.010

Table 2.3: Empirical powers of tests based on $\hat{\sigma}_2^2$.

		$n = 300$			$n = 500$		
p	Level	U_n	KS	CvM	U_n	KS	CvM
0	0.05	0.030	0.053	0.049	0.049	0.049	0.052
	0.025	0.018	0.022	0.018	0.020	0.021	0.029
	0.01	0.007	0.006	0.007	0.007	0.014	0.013
0.2	0.050	0.073	0.038	0.041	0.134	0.046	0.052
	0.025	0.057	0.015	0.024	0.118	0.025	0.025
	0.010	0.045	0.006	0.012	0.106	0.013	0.014
0.4	0.050	0.148	0.071	0.099	0.303	0.089	0.169
	0.025	0.129	0.049	0.066	0.263	0.052	0.117
	0.010	0.110	0.024	0.037	0.229	0.030	0.075
0.6	0.050	0.261	0.109	0.182	0.494	0.241	0.411
	0.025	0.223	0.066	0.131	0.445	0.172	0.336
	0.010	0.188	0.032	0.076	0.398	0.101	0.253
0.8	0.050	0.404	0.216	0.408	0.673	0.422	0.716
	0.025	0.342	0.141	0.311	0.612	0.326	0.627
	0.010	0.300	0.087	0.217	0.563	0.209	0.516
1	0.050	0.556	0.331	0.575	0.812	0.587	0.873
	0.025	0.499	0.235	0.478	0.760	0.447	0.816
	0.010	0.437	0.153	0.368	0.710	0.356	0.738

2.5 Proofs

In this section we give the proof of Theorem 2.3.1. To this end, we list some useful lemmas. For $\alpha-$mixing processes, we can follow the same proof as in Selk and Neumeyer (2013), and for the moment contracting stationary processes, the proofs are similar to those of Wu et al. (2010). Many details that follow lemma will be brief.

Let t_1, t_2, $\cdots$ be measurable functions which are bounded by the same constant B. Let

$$T_n(x) = \frac{1}{nh} \sum_{j=1}^{n} t_n(X_j) K\Big(\frac{X_j - x}{h_1}\Big), \quad x \in \mathbb{R}. \tag{2.5.1}$$

We have

Lemma 2.5.1 *Under the conditions of Theorem 2.3.1,*

$$\sup_{x \in I_n} |T_n(x) - E(T_n(x))| = O_p\Big(\Big(\frac{\log n}{nc_n}\Big)^{1/2}\Big).$$

Proof. (i) Under condition (Z) for $\alpha-$mixing processes, the proof is similar to that of Lemma B.1 in Selk and Neumeyer (2010) with $k = 0$ in their proof.

(ii) For the moment contracting processes, since the t_1, t_2, $\cdots$ are bounded on I_n, the claim follows from Proposition 2 and Lemma 4 of Wu et al. (2010).

Next, consider

$$U_{n,l}(x) = \frac{1}{nh} \sum_{j=1}^{n} \varepsilon_j \sigma(X_{j-1}) K^{(l)}\Big(\frac{X_{j-1} - x}{h_1}\Big), \quad x \in I_n, l = 0, 1, 2, \tag{2.5.2}$$

where $K^{(l)}$ is the l-th derivative of K. We have

Lemma 2.5.2 *Under the conditions of Theorem 2.3.1,*

$$\sup_{x \in I_n, l=0,1,2} |U_{n,l}(x)| = O_p\Big(\Big(c_n^{-1/2-l} n^{-1/2} (\log n)^{1/2} + c_n^2\Big) q_n\Big).$$

Proof. i) Under the condition (Z) for $\alpha-$mixing processes, it follows from Lemma B.1 and Lemma B.2 of Selk and Neumeyer (2010) applied with $k = 1$.

(ii) Under the condition (Z$'$) for the moment contracting processes, because of the stationarity, it follows from Lemma 4 of Müller et al. (2009).

Proof of Lemma 2.2.1. The general idea of the proof this lemma and Theorem 2.3.1 is similar to that of Theorem 1 in Müller et al. (2009), so we use similar notation as in their paper and shall be brief whenever possible. Let $K_i(u) = u^i K(u)$, $i \geq 0$, Let $K_i(u) = u^i K(u)$, $i \geq 0$,

$$\hat{p}_i(x) = \frac{1}{nh_1} \sum_{j=1}^{n} K_i\Big(\frac{X_{j-1} - x}{h_1}\Big), \qquad \hat{q}_i(x) = \frac{1}{nh_1} \sum_{j=1}^{n} X_j K_i\Big(\frac{X_{j-1} - x}{h_1}\Big), \quad x \in \mathbb{R}.$$

On the event, $\hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x) > 0$,

$$\hat{m}(x) = \frac{\hat{p}_2(x)\hat{q}_0(x) - \hat{p}_1(x)\hat{q}_1(x)}{\hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x)}.$$

Assumption (F), (H), (K), and Lemmas 2.5.1 imply

$$\sup_{x \in I_n} |\hat{p}_i(x) - E[\hat{p}_i(x)]| = O_p(h_1), \quad i = 0, 1, 2, \cdots. \tag{2.5.3}$$

Let $\bar{p}_i(x) = E[\hat{p}_i(x)]$ and $\lambda_i = \int K_i(u)du = \int u^i K(u)du$. Note that $\bar{p}_i(x) = \int g(x - h_1 u)u^i K(u)du$, and $\lambda_0 = 1$, $\lambda_1 = 0$, $\lambda_2 > 0$. By (2.5.3),

$$\|\hat{p}_i/g - \lambda_i\|_{I_n} + \|\bar{p}_i/g - \lambda_i\|_{I_n} = O_p(h_1), \quad i = 0, 1, 2, \cdots. \tag{2.5.4}$$

Hence

$$\|\hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x) - \lambda_2 g^2\|_{I_n} = O_p(h_1).$$

With $(\inf_{x\in I_n} g(x))^{-1} = q_{n,g}$ in assumption (X), there exists an $\eta > 0$ such that

$$P\Big(q_{n,g}^2 \inf_{x\in I_n} |\hat{p}_2(x)\hat{p}_0(x) - \hat{p}_1^2(x)| > \eta\Big) \to 1. \tag{2.5.5}$$

Write $\hat{q}_i = A_i + B_i$, for $i = 0, 1$, where

$$A_i(x) = \frac{1}{nh_1}\sum_{j=1}^{n} \sigma(X_{j-1})\varepsilon_j K_i\Big(\frac{X_{j-1}-x}{h_1}\Big),$$
$$B_i(x) = \frac{1}{nh_1}\sum_{j=1}^{n} m(X_{j-1}) K_i\Big(\frac{X_{j-1}-x}{h_1}\Big), \quad x \in \mathbb{R}.$$

Since the second derivative $\ddot{m}$ of m is bounded, a Taylor expansion shows that

$$\|(B_i - m\hat{p}_i - \dot{m}h_1\hat{p}_{i+1} - \frac{1}{2}\ddot{m}h_1^2\hat{p}_{i+2})/g\|_{I_n} = O_p(h_1^3), \tag{2.5.6}$$

where $\|\cdot\|_{I_n}$ denotes the super norm over I_n.

Note that the proof of the properties of $\hat{\sigma}_1^2$ is similar to one for $\hat{m}$, so we give the details for $\hat{m}$ and $\hat{\sigma}_2^2$ only. By $g_n = u_p(h_n)$, we mean that there exists constant $C > 0$, such that $P(\|g_n\|_{I_n} \le C\|h_n\|_{I_n}) \to 1$. Based on the analysis above, we obtain the following expansions, which are similar to those appearing in Yao and Tong (1994). With $\hat{r}_j \equiv (X_j - \hat{m}(X_{j-1}))^2$,

$$\hat{m}(x) - m(x) = \frac{1}{nh_1 g(x)}\sum_{j=1}^{n} \sigma(X_{j-1})\varepsilon_j K\Big(\frac{X_{j-1}-x}{h_1}\Big) + \frac{h_1^2\lambda_2}{2}\ddot{m}(x) + u_p(R_{n,1(x)}), \tag{2.5.7}$$

$$\hat{\sigma}_2^2(x) - \sigma_2^2(x) \tag{2.5.8}$$
$$= \frac{1}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1}-x}{h_2}\Big)\{\hat{r}_j - \sigma^2(x) - \dot{\sigma}^2(x)(X_{j-1}-x)\} + u_p\{R_{n,2}(x)\},$$

where

$$\begin{aligned} R_{n,1}(x) &= \frac{1}{ng(x)}\Big[\Big|\sum_{j=1}^{n}\sigma(X_{j-1})\varepsilon_j K\Big(\frac{X_{j-1}-x}{h_1}\Big)\Big| \\ &\quad + \Big|\sum_{j=1}^{n}\frac{X_{j-1}-x}{h_1}\sigma(X_{j-})\varepsilon_j K\Big(\frac{X_{j-1}-x}{h_1}\Big)\Big|\Big] + O(q_{n,g}^2 q_n h_1^3); \\ R_{n,2}(x) &= \frac{1}{ng(x)}\Big[\Big|\sum_{j=1}^{n} W\Big(\frac{X_{j-1}-x}{h_2}\Big)\{\hat{r}_j - \sigma^2 - \dot{\sigma}^2(x)(X_{j-1}-x)\}\Big| \\ &\quad + \Big|\sum_{j=1}^{n}\frac{X_{j-1}-x}{h_2} W\Big(\frac{X_{j-1}-x}{h_2}\Big)\{\hat{r}_j - \sigma^2 - \dot{\sigma}^2(x)(X_{j-1}-x)\}\Big|\Big] \\ &\quad + O(q_{n,g}^2 q_n^2 h_2^3). \end{aligned}$$

From Lemma 2.5.2, we have

$$\sup_{x\in I_n}\Big|\frac{1}{nh_1 g(x)}\sum_{j=1}^{n}\sigma(X_{j-1})\varepsilon_j K_i\Big(\frac{X_{j-1}-x}{h_1}\Big)\Big| = O_p\Big(q_n q_{n,g}\Big(\frac{\log n}{nh_1}\Big)^{1/2}\Big).$$

From (2.5.7) and the above bounds we readily obtain

$$\sup_{x\in I_n}|\hat{m}(x) - m(x)| = O_p\Big(\Big(c_n^{-1/2} n^{-1/2}(\log n)^{1/2}\Big) q_n q_{n,g}\Big). \tag{2.5.9}$$

Combining this fact with condition (M) completes the proof of (2.2.7).

To deal with $\hat{\sigma}_2^2$, a similar analysis as in Fan and Yao (2002) can be followed, where

$$\begin{aligned}
\hat{r}_j &= \{X_j - \hat{m}(X_{j-1})\}^2 = \{\sigma(X_{j-1})\varepsilon_j + m(X_{j-1}) - \hat{m}(X_{j-1})\}^2 \\
&= \sigma^2(X_{j-1})\varepsilon_j^2 + 2\sigma(X_{j-1})\varepsilon_j\{m(X_{j-1}) - \hat{m}(X_{j-1})\} \\
&\quad +\{m(X_{j-1}) - \hat{m}(X_{j-1})\}^2.
\end{aligned}$$

Then

$$\begin{aligned}
&\hat{\sigma}_2^2(x) - \sigma^2(x) \\
= \;& J_1 + J_2 - J_3 + J_4 + O_p(h_2)(|J_1 + J_2 - J_3 + J_4| + |J_1^* + J_2^* - J_3^* + J_4^*|),
\end{aligned}$$

where

$$\begin{aligned}
J_1 &= \frac{1}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1} - x}{h_2}\Big)\{\sigma^2(X_{j-1}) - \sigma^2(x) - \dot{\sigma}^2(x)(X_{j-1} - x)\}, \\
J_2 &= \frac{1}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1} - x}{h_2}\Big)\sigma^2(X_{j-1})(\varepsilon_j^2 - 1), \\
J_3 &= \frac{2}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1} - x}{h_2}\Big)\sigma(X_{j-1})\varepsilon_j\{\hat{m}(X_{j-1}) - m(X_{j-1})\}, \\
J_4 &= \frac{1}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1} - x}{h_2}\Big)\{\hat{m}(X_{j-1}) - m(X_{j-1})\}^2,
\end{aligned}$$

and J_i^* is defined in the same way as J_i with one more factor $h_2^{-1}(X_{j-1} - x)$ in the jth summand, for $j = 1, \cdots, n$ and $i = 1, \cdots, 4$. Condition (M) implies

$$\|J_1\|_{I_n} = O_p(q_n q_{n,g} h_2^2),$$

and from Lemma 2.5.2, we obtain

$$\|J_2\|_{I_n} = O_p\Big(q_n^2 q_{n,g}\Big(\frac{\log n}{nh_2}\Big)^{1/2}\Big).$$

Based on (2.5.9),

$$\|J_4\|_{I_n} = O_p\Big(q_{n,g}^3 q_n^2 \frac{\log n}{nh_1h_2}\Big).$$

To deal with J_3, rewrite $J_3 = J_{31} + J_{32} + J_{33}$, where

$$\begin{aligned}
J_{31} &= \frac{1}{n^2h_1h_2g(x)}\sum_{i,j=1}^{n} K\Big(\frac{X_{i-1}-X_{j-1}}{h_1}\Big)\sigma(X_{i-1})\sigma(X_{j-1})\varepsilon_i\varepsilon_j \\
&\quad \Big\{g^{-1}(X_{i-1})W\Big(\frac{X_{i-1}-x}{h_2}\Big) + g^{-1}(X_{j-1})W\Big(\frac{X_{j-1}-x}{h_2}\Big)\Big\} \\
&= \frac{1}{n^2h_1h_2g(x)}\sum_{1\le i,j\le n}\phi_{ij},
\end{aligned}$$

$$\begin{aligned}
J_{32} &= \frac{h_1^2\lambda_2}{nh_2g(x)}\sum_{i=1}^{n} W\Big(\frac{X_{i-1}-x}{h_2}\Big)\sigma(X_{i-1})\varepsilon_i\ddot{m}(X_{i-1}), \\
|J_{33}| &\le \frac{O_p(1)}{n^2h_2}\sum_{i,j=1}^{n}\Big|W\Big(\frac{X_{i-1}-x}{h_2}\Big)K\Big(\frac{X_{i-1}-X_{j-1}}{h_1}\Big)\sigma(X_{i-1})\sigma(X_{j-1})|\varepsilon_i|\varepsilon_j/g(X_{i-1})\Big|,
\end{aligned}$$

where

$$\begin{aligned}
\phi_{ij} &= K\Big(\frac{X_{i-1}-X_{j-1}}{h_1}\Big)\sigma(X_{i-1})\sigma(X_{j-1})\varepsilon_i\varepsilon_j\Big\{g^{-1}(X_{i-1})W\Big(\frac{X_{i-1}-x}{h_2}\Big) \\
&\quad + g^{-1}(X_{j-1})W\Big(\frac{X_{j-1}-x}{h_2}\Big)\Big\}.
\end{aligned}$$

Argue as in Borkowski and Mielniczuk (2012), to obtain

$$E\Big\{\sum_{1\leq i,j\leq n}\phi_{ij}\Big\}^2 = O_p(n^2c_n^2q_n^4q_{n,g}^2).$$

To obtain the uniform bound, we consider the equal-length cover I_{nk} and with center x_{nk},

$k=1,\cdots,L(n)$, for I_n, where

$$L(n) = O((\log n)^{r_1}/(c_n^3(nc_n)^{1/2}q_nq_{n,g})).$$

Then

$$\sup_{x\in I_n}|J_{31}(x)| \leq \max_{1\leq k\leq L(n)}\sup_{x\in I_n\cap I_{nk}}|J_{31}(x)-J_{31}(x_{nk})| + \max_{1\leq k\leq L(n)}|J_{31}(x_{nk})| = R_1+R_2.$$

Note that

$$R_1 \leq \frac{C(\log n)^{r_1}q_n^3q_{n,g}^2}{L(n)h_2(nh_1)^{1/2}} = O_p\Big(q_n^2q_{n,g}c_n^2\Big).$$

For any $\epsilon>0$, by the relation (2.2.6) in assumption (H), for a constant $C<\infty$,

$$\begin{aligned}
&P\Big(\Big|q_n^{-2}q_{n,g}^{-1}c_n^{-2}R_2\Big|>\epsilon\Big)\\
&\leq L(n)P\Big(\Big|q_n^{-2}q_{n,g}^{-1}c_n^{-2}\Big(\frac{1}{n^2h_1h_2g(x)}\sum_{1\leq i\leq j\leq n}\phi_{ij}\Big)\Big|>\varepsilon\Big)\\
&\leq \frac{C(\log n)^{r_1}}{c_n^3(nc_n)^{1/2}q_nq_{n,g}}\frac{1}{\epsilon^2n^4c_n^4h_1^2h_2^2q_n^4q_{n,g}^2}E\Big\{\sum_{i<j}\phi_{ij}\Big\}^2\\
&= \frac{C(\log n)^{r_1}}{\epsilon^2n^{5/2}c_n^{19/2}q_nq_{n,g}}\to 0.
\end{aligned}$$

So $\|J_{31}\|_{I_n} = O_p(q_n^2 q_{n,g} c_n^2)$. Similarly, we obtain $\|J_{33}\|_{I_n} = O_p(q_n^2 q_{n,g} c_n^2)$. Also, it is obvious that $\|J_{32}\|_{I_n} = o_p(q_n^2 q_{n,g} h_2^2)$.

Based on the above analysis, we obtain

$$\hat{\sigma}_2^2 - \sigma_2^2 \;=\; \frac{1}{nh_2 g(x)} \sum_{j=1}^{n} W\Big(\frac{X_{j-1} - x}{h_2}\Big)\sigma^2(X_{j-1})(\varepsilon_j^2 - 1) + O_p(q_n^2 q_{n,g} c_n^2).$$

This relation, the fact $(\hat{\sigma}_2 - \sigma)/\sigma = (\hat{\sigma}_2^2 - \sigma^2)/2\sigma^2 - (\hat{\sigma}_2 - \sigma)^2/2\sigma^2$, Lemma (2.5.2) and the condition (M) together imply (2.2.8) in routine fashion. This also completes the proof of Lemma 2.2.1.

Proof of Theorem 2.3.1. We denote $\hat{S} = (\hat{m} - m)/\sigma$, $\hat{T} = (\hat{\sigma} - \sigma)/\sigma$. Let $\mathbb{F}_\omega$ denote the weighted empirical distribution function based on the unobserved innovations, which is

$$\mathbb{F}_\omega(t) = \sum_{j=1}^{n} \bar{\omega}_j I[\varepsilon_j \le t], \quad t \in \mathbb{R}.$$

Similarly as in Lemma B.5 of Selk and Neumeyer (2013), we obtain

$$\sup_{t\in\mathbb{R}} |\mathbb{F}_\omega(t) - \mathbb{F}(t)| = o_p(n^{-1/2}), \qquad \overline{W} = \frac{1}{n}\sum_{j=1}^{n} \omega_{nj} = 1 + o_p(1).$$

Next, define

$$B(t, \hat{S}, \hat{T}) = \frac{1}{n}\sum_{j=1}^{n} \omega_{nj}\big\{F\big(t + \hat{S}(X_{j-1}) + \hat{T}(X_{j-1})t\big) - F(t)\big\},$$

and

$$H(t,\hat{S},\hat{T}) = \frac{1}{n}\sum_{j=1}^{n}\omega_{nj}\big\{I\big(\varepsilon \le t+\hat{S}(X_{j-1})+\hat{T}(X_{j-1})t\big) - F\big(t+\hat{S}(X_{j-1})+\hat{T}(X_{j-1})t\big)\big\},$$

for t in $\mathbb{R}$ and S, T in $C(\mathbb{R})$, the set of continuous functions from $\mathbb{R}$ to $\mathbb{R}$. Then we can rewrite

$$\overline{W}(\hat{\mathbb{F}}(t) - \mathbb{F}_\omega(t)) = H(t,\hat{S},\hat{T}) - H(t,0,0) + B(t,\hat{S},\hat{T}).$$

It follows from Lemma 2.5.3 below that

$$\begin{aligned} \frac{1}{n}\sum_{j=1}^{n}\omega_{nj}\frac{\hat{m}(X_{j-1})-m(X_{j-1})}{\sigma(X_{j-1})} &= \frac{1}{n}\sum_{j=1}^{n}\varepsilon_j + o_p(n^{-1/2}), \\ \frac{1}{n}\sum_{j=1}^{n}\omega_{nj}\frac{\hat{\sigma}(X_{j-1})-\sigma(X_{j-1})}{\sigma(X_{j-1})} &= \frac{1}{2n}\sum_{j=1}^{n}(\varepsilon_j^2-1) + o_p(n^{-1/2}). \end{aligned}$$

As $\dot{f}$ exists, we derive

$$\sup_{t\in\mathbb{R}}\Big|B(t,\hat{S},\hat{T}) - f(t)\frac{1}{n}\sum_{j=1}^{n}\Big(\varepsilon_j+\frac{1}{2}t(\varepsilon_j^2-1)\Big)\Big| \le \frac{1}{n}\sum_{j=1}^{n}\omega_{nj}\dot{f}(\xi_{t,X_{j-1}})[\hat{S}(X_{j-1})+\hat{T}(X_{j-1})t]^2,$$

for some $\xi_{t,X_{j-1}}$ between $\hat{S}(X_{j-1})+\hat{T}(X_{j-1})t$ and t. The relation $\sup_{t\in\mathbb{R}}|t^2\dot{f}(t)| < \infty$ yields

$$\sup_{t\in\mathbb{R}}\Big|B(t,\hat{S},\hat{T}) - f(t)\frac{1}{n}\sum_{j=1}^{n}\Big(\varepsilon_j+\frac{1}{2}t(\varepsilon_j^2-1)\Big)\Big| = o_p(n^{-1/2}).$$

Thus to complete the proof of claim (2.3.3), it remains to show that

$$\sup_{t\in\mathbb{R}} \Big| H(t,S,T) - H(t,0,0) \Big| = o_p(n^{-1/2}). \tag{2.5.10}$$

Based on condition (E), we have

$$\max_{1\le j\le n} |\varepsilon_j| = o_p(n^{1/2}).$$

Since $\|S\|_{I_n} = o_p(1)$, $\|T\|_{I_n} = o_p(n^{-1/4})$, the probability of the event

$$A_n := \{\max_{1\le j\le n} |\varepsilon_j| \le 2n^{1/2} - 1\} \cap \{\|\hat{S}\|_{I_n} = o_p(1), \quad \|\hat{T}\|_{I_n} = o_p(n^{-1/4})\},$$

tends to one. On the event A_n

$$\begin{aligned} \sup_{|t|>n^{1/2}} |H(t,\hat{S},\hat{T} - H(t,0,0)| &= \sup_{|t|>n^{1/2}, 1\le i\le n} B(t,\hat{S},\hat{T}) \\ &\le 2(1 - F(\sqrt{n}/2)) + 2F(1 - \sqrt{n}/2). \end{aligned}$$

Since F has a finite second moment, we have $F(t) = o(t^{-2})$, as $t \to -\infty$ and $1 - F(t) = o(t^{-2})$, as $t \to \infty$. This implies that

$$\sup_{|t|>n^{1/2}} |H(t,\hat{S},\hat{T} - H(t,0,0)| = o_p(n^{-1/2}).$$

So we are left to show

$$\sup_{|t|\le n^{1/2}} |H(t,\hat{S},\hat{T} - H(t,0,0)| = o_p(n^{-1/2}). \tag{2.5.11}$$

Now let $\delta = 1/(1+\sqrt{3})$. For any interval I, let $C_1^{1+\delta}(I)$ be the set of differentiable functions h on $\mathbb{R}$ that satisfy $\|h\|_{I,\delta} \le 1$, where

$$\|h\|_{I,\delta} = \|h\|_I + \|\dot{h}\|_I + \sup_{x,y\in I, x\neq y} \frac{|\dot{h}(x)-\dot{h}(y)|}{|x-y|^\delta}.$$

Now let $\mathcal{D}_n = \{u+\nu : u \in \mathcal{U}_n, \nu \in \mathcal{V}_n\}$, where

$$\begin{aligned}
\mathcal{U}_n &= \{h \in C(\mathbb{R}) : \|h\|_{I_n} \le n^{-1/2}\}, \\
\mathcal{V}_n &= \{h \in C_1^{1+\delta}(\mathbb{R}) : \|h\|_{I_n} \le n^{-1/2} c_n^{-1/2} \log n Q_n^2\},
\end{aligned}$$

with $Q_n = q_n q_{n,g} q_{n,\sigma}$. Let $\hat{u}(x) := \hat{m}(x) - m(x) - \hat{v}(x)$, and $\hat{u}_\sigma(x) := \hat{\sigma}^2(x) - \sigma^2(x) - \hat{v}_\sigma(x)$, where

$$\begin{aligned}
\hat{v}(x) &:= \frac{1}{nh_1 g(x)} \sum_{j=1}^n \sigma(X_{j-1})\varepsilon_j K\Big(\frac{X_{j-1}-x}{h_1}\Big) + O_p(q_n c_n^2), \\
\hat{v}_\sigma(x) &:= \frac{1}{nh_2 g(x)} \sum_{j=1}^n W\Big(\frac{X_{j-1}-x}{h_2}\Big)\sigma^2(X_{j-1})(\varepsilon_j^2 - 1).
\end{aligned}$$

It follows from Lemma 2.5.2 and similar argument as in Selk and Neumeyer (2013), $\hat{S}$ and $\hat{T}$ belong to $\mathcal{D}_n$ with probability tending to one. So (2.5.11) will be followed if we prove

$$\sup_{|t|\le n^{1/2}, S,T\in\mathcal{D}_n} |H(t,S,T - H(t,0,0)| = o_p(n^{-1/2}).$$

To this end, set $\eta_n = n^{-1/2}$. Let $t_1, \cdots, t_{M_n}$ be η_n-net of $[-n^{1/2}, n^{1/2}]$, and set

$\nu_1, \cdots, \nu_{N_n}$ for $\mathcal{V}_n$. We can choose the former net such that

$$M_n \leq 2 + n, \tag{2.5.12}$$

the second net is

$$N_n \leq \exp(K_*(2 + b_n - a_n)n^{1/(2+2\delta)}), \tag{2.5.13}$$

where K_* is some positive constant, see also (Van der Vaart and Wellner (1996)). Note that $\nu_1, \cdots, \nu_{N_n}$ is an $2\eta_n$-net for $\mathcal{D}_n$. We have

$$\begin{aligned}
&\sup_{|t| \leq n^{1/t}, S, T \in \mathcal{D}_n} |H(t, S, T) - H(t, 0, 0)| \\
\leq\ & \max_{i,l,m} |H_n(t_i, \nu_l, \nu_m) - H_n(t_i, 0, 0)| + \max_{i,l,m} D_{i,l,m},
\end{aligned}$$

where

$$\begin{aligned}
D_{i,l,m} = & \sup_{|t-t_i| \leq \eta_n, \|S-\nu_l\|_I \leq 2\eta_n, \|T-\nu_m\|_I \leq 2\eta_n} \big(|H(t_i, S, T) - H(t, \nu_l, \nu_m)| \\
& + |H(t_i, 0, T) - H(t, 0, \nu_m)| + |H(t_i, S, 0) - H(t, \nu_l, 0)|\big) + |H_n(t_i, 0, 0) - H_n(t, 0, 0)|.
\end{aligned}$$

For $|t - t_i| \leq \eta_n$, $\|S - \nu_l\|_I \leq 2\eta_n$, $\|T - \nu_m\|_I \leq 2\eta_n$, we have

$$\begin{aligned}
I\big(y \leq t_i + \nu_l(x) + \nu_m(x)t_i - \eta_n(A + 3)\big) &\leq I\big(y \leq t + S(x) + T(x)t\big) \\
&\leq I\big(y \leq t_i + \nu_l(x) + \nu_m(x)t_i - \eta_n(A + 3)\big),
\end{aligned}$$

and

$$F\big(t_i + \nu_l(x) + \nu_m(x)t_i - \eta_n(A+3)\big)$$
$$\leq F\big(t + S(x) + T(x)t\big) \leq F\big(t_i + \nu_l(x) + \nu_m(x)t_i + \eta_n(A+3)\big),$$

for all $y \in \mathbb{R}$ and $x \in I_n$, where $A = |T| + 2|t_i| + 2\eta_n$. Hence

$$|H(t_i, S, T) - H(t, \nu_l, \nu_m)|$$
$$\leq \quad |H\big(t_i + \eta_n(A+3), \nu_l(x), \nu_m(x)\big) - H\big(t_i - \eta_n(A+3), \nu_l(x), \nu_m(x)\big)| + 2R_{i,l,m},$$

with

$$\begin{aligned} &R_{i,l,m} \\ &= \sum_{j=1}^{n} \frac{\omega_{nj}}{n} \{F\big(t_i + \nu_l(x) + \nu_m(x)t_i + \eta_n(A+3)\big) - F\big(t_i + \nu_l(x) + \nu_m(x)t_i - \eta_n(A+3)\big)\} \\ &\leq 2\eta_n(\sup_t |Af(\xi)| + 3\|f\|_\infty), \quad \text{say.} \end{aligned}$$

for some ξ is between $t_i + \nu_l(x) + \nu_m(x)t_i - \eta_n(A+3)$ and $t_i + \nu_l(x) + \nu_m(x)t_i + \eta_n(A+3)$. By assumption (F), there exists some L, such that $|Af(\xi)| < L < \infty$. Similarly, we derive the bound for the following terms,

$$\begin{aligned} |H(t_i, 0, T) - H(t, 0, \nu_m)| &\leq |H\big(t_i + \eta_n(A+1), 0, \nu_m(x)\big) - H\big(t_i - \eta_n(A+1), 0, \nu_m(x)\big)| \\ &\leq 4\eta_n L + 4\|f\|_\infty, \\ |H(t_i, S, 0) - H(t, \nu_l, 0)|) &\leq |H\big(t_i + 3\eta_n, \nu_l(x), 0\big) - H\big(t_i - 3\eta_n, \nu_l(x), 0\big)| \leq \eta_n 12\|f\|_\infty, \\ |H_n(t_i, 0, 0) - H_n(t, 0, 0)| &\leq |H(t_i + \eta_n, 0, 0) - H(t_i - \eta_n, 0, 0)| \leq \eta_n 4\|f\|_\infty. \end{aligned}$$

So

$$\sup_{|t|\le n^{1/2}, S,T\in\mathcal{D}_n} |H(t,\hat{S},\hat{T}-H(t,0,0)| = T_1+T_2+T_3+T_4+T_5+\eta_n(8L+32\|f\|_\infty),$$

where

$$\begin{aligned}
T_1 &= \max_{i,l,m}|H(t_i,\nu_l,\nu_m)-H(t_i,0,0)|,\\
T_2 &= \max_{i,l,m}|H\big(t_i+\eta_n(A+3),\nu_l(x),\nu_m(x)\big)-H(t_i-\eta_n(A+3),\nu_l(x),\nu_m(x))|,\\
T_3 &= \max_{i,l,m}|H\big(t_i+\eta_n(A+1),0,\nu_m(x)\big)-H\big(t_i-\eta_n(A+1),0,\nu_m(x)\big)|,\\
T_4 &= \max_{i,l,m}|H\big(t_i+3\eta_n,\nu_l(x),0\big)-H\big(t_i-3\eta_n,\nu_l(x),0\big)|,\\
T_5 &= \max_{i,l,m}|H(t_i+\eta_n,0,0)-H(t_i-\eta_n,0,0)|.
\end{aligned}$$

To continue, for any υ_i and τ_i, $i=1,2$, let

$$\begin{aligned}
Y_j &= \omega_{nj}\Big\{I\big(\varepsilon_j\le s+\upsilon_1(X_{j-1})+\tau_1(X_{j-1})s\big)-I\big(\varepsilon_j\le t+\upsilon_2(X_{j-1})+\tau_2(X_{j-1})t\big)\\
&- F\big(s+\upsilon_1(X_{j-1})+\tau_1(X_{j-1})s\big)+F\big(t+\upsilon_2(X_{j-1})+\tau_2(X_{j-1})t\big)\Big\}.
\end{aligned}$$

We have $|Y_j|\le 2$, $E(Y_j|X_0,\cdots,X_{j-1})=0$, and

$$\begin{aligned}
V_n &= \sum_{j=1}^n \mathrm{E}\big(Y_j^2|X_0,\cdots,X_{j-1}\big)\\
&\le \sum_{j=1}^n \Big|F\big(s+\upsilon_1(X_{j-1})+\tau_1(X_{j-1})s\big)-bF\big(t+\upsilon_2(X_{j-1})+\tau_2(X_{j-1})t\big)\Big|\\
&\le n\Big|f(\xi)\Big\{\big(s+\upsilon_1(X_{j-1})+\tau_1(X_{j-1})s\big)-\big(t+\upsilon_2(X_{j-1})+\tau_2(X_{j-1})t\big)\Big\}\Big|,
\end{aligned}$$

where ξ is between $s+\upsilon_1(X_{j-1})+\tau_1(X_{j-1})s$ and

$t + \upsilon_2(X_{j-1}) + \tau_2(X_{j-1})t$. Since $\sup_t |tf(t)| < \infty$, there exists some constant L, such that

$$V_n \leq n\{\|f\|_\infty(|s-t|(1+\|\sigma\|_{I_n}) + \|\upsilon_1 - \upsilon_2\|_{I_n}) + L\|\tau_1 - \tau_2\|_{I_n}\} = n\|f\|_\infty B.$$

Then by martingale inequality in Freedman (1975),

$$\begin{aligned} P(|H(s,\upsilon_1,\tau_1)| - H(t,\upsilon_2,\tau_2)| > \beta n^{1/2}) &= P\Big(\Big|\sum_{j=1}^n Y_j\Big| > \beta n^{1/2}, V_n \leq n\|f\|_\infty B\Big), \\ &\leq 2\exp(-\frac{\beta^2 n}{4\beta n^{1/2} + 2n\|f\|_\infty B}). \end{aligned}$$

Also $\|\nu_l\|_{I_n} \leq n^{-1/2}c_n^{-1/2}\log nQ_n^2 + \eta_n$. Thus we obtain that

$$\begin{aligned} &P(T_1 > \beta n^{-1/2}) \\ &\leq \sum_{i,l,m} P(|H(t_i,\nu_l,\nu_m)| - H(t_i,0,0)| > \beta n^{1/2}) \\ &\leq 2M_nN_n^2 \exp\Big(-\frac{\beta^2 n}{4\beta n^{1/2} + 4n(n^{-1/2}c_n^{-1/2}\log nQ_n^2(L+1) + \eta_n)\|f\|_\infty}\Big). \end{aligned}$$

Similarly, there exists some constant L_2 and L_3, such that

$$\begin{aligned} P(T_2 > \beta n^{-1/2}) &\leq 2M_nN_n^2 \exp\Big(-\frac{\beta^2 n}{4\beta n^{1/2} + n\eta_n(L_2 + 12\|f\|_\infty)}\Big), \\ P(T_3 > \beta n^{-1/2}) &\leq 2M_nN_n^2 \exp\Big(-\frac{\beta^2 n}{4\beta n^{1/2} + n\eta_n(L_3 + 4\|f\|_\infty)}\Big), \\ P(T_4 > \beta n^{-1/2}) &\leq 2M_nN_n^2 \exp\Big(-\frac{\beta^2 n}{4\beta n^{1/2} + 12n\eta_n\|f\|_\infty}\Big), \\ P(T_5 > \beta n^{-1/2}) &\leq 2M_nN_n^2 \exp\Big(-\frac{\beta^2 n}{4\beta n^{1/2} + 4n\eta_n\|f\|_\infty}\Big). \end{aligned}$$

As $\delta = 1/(1+\sqrt{3})$ and relation (2.2.5) in condition (H), together with relations (2.5.12) and (2.5.13), we obtain that

$$P(T_i > \beta n^{-1/2}) \to 0, \quad i = 1, 2, \cdots, 5, \quad \beta > 0.$$

This completes the proof of (2.5.10) and hence the proof of Theorem 2.3.1

Lemma 2.5.3 *Under the conditions of Theorem 2.3.1,*

$$\frac{1}{n}\sum_{j=1}^{n} \omega_n(X_{j-1}) \frac{\hat{m}(X_{j-1}) - m(X_{j-1})}{\sigma_(X_{j-1})} = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}),$$

and for $i = 1, 2,$

$$\frac{1}{n}\sum_{j=1}^{n} \omega_n(X_{j-1}) \frac{\hat{\sigma}_i(X_{j-1}) - \sigma(X_{j-1})}{\sigma_(X_{j-1})} = \frac{1}{2n}\sum_{j=1}^{n} (\varepsilon_j^2 - 1) + o_p(n^{-1/2}).$$

Proof. To prove the first equation, from the proof of Lemma 2.2.1, we have

$$\hat{m}(x) - m(x) = \frac{1}{nh_1 g(x)} \sum_{j=1}^{n} \sigma(X_j)\varepsilon_j K\Big(\frac{X_j - x}{h_1}\Big) + o_p(n^{-1/2}).$$

Then we only need to prove

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\omega_n(X_i)}{nh_1 g(X_i)\sigma(X_i)} \sum_{j=1}^{n} \sigma(X_j)\varepsilon_j K\Big(\frac{X_j - X_i}{h_1}\Big) = \frac{1}{n}\sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}).$$

Denote

$$\hat{d}(x) = \sum_{i=1}^{n} \frac{\omega_n(X_i)\sigma(x)}{nh_1 g(X_i)\sigma(X_i)} K\Big(\frac{x - X_i}{h_1}\Big).$$

Let $\bar{d}(x) = E(\hat{d}(x))$, we have

$$\bar{d}(x) = \int \frac{\omega_n(u)\sigma(x)}{h_1\sigma(u)} K\Big(\frac{x-u}{h_1}\Big) du.$$

Then we have $E[(\bar{d}(X) - 1)^2] \to 0$. Therefore

$$\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j(\bar{d}(X_j) - 1) = o_p(n^{-1/2}).$$

Thus we only need to prove that

$$\frac{1}{n}\sum_{j=1}^{n} \varepsilon_j \tilde{d}(X_j) = o_p(n^{-1/2}), \tag{2.5.14}$$

where $\tilde{d}(x) = \hat{d}(x) - \bar{d}(x)$. But the proof of (2.5.14) is similar to that of Lemma B.3 of Selk and Neumeyer (2013) under the mixing condition (Z), and as that appearing in section 5 of Müller et al. (2009) under the moment contracting condition (Z′). The second equation can be followed by similar proof.

Chapter 3

Linear Measurement Error Models

3.1 Introduction

The problem of fitting an error distribution in regression models has been well studied when covariates are fully observed, see, e.g., Loynes (1980), Koul (2002), Khamalze and Koul (2004, 2009) and the references therein. However, in practice there are numerous examples of real world applications where covariates are not observable. Instead, one observes some surrogates for covariates. The monographs of Cheng and Van Ness (1999), Fuller (2009) and Carroll, Rupert, Stefanski, and Crainiceanu (2012) are full of such important applications. These models are often called errors-in-variables models or measurement errors models. Relatively little is known about fitting an error distribution to the regression model in these models. In this chapter we investigate a class of tests for this testing problem based on deconvoluted density estimators of the error density.

Let $p \geq 1$ be a given dimension of the covariate vector X. In a multiple linear regression model with measurement error in X one observes the response variable Y and a surrogate p-vector Z obeying the model

$$Y = \alpha + \beta' X + \varepsilon, \qquad Z = X + u, \tag{3.1.1}$$

for some $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^p$, where the p-vector u is the measurement error in X. Here b'

denotes the transpose of any vector $b \in \mathbb{R}^p$. The variables ε, u and X are assumed to be mutually independent, with $E\varepsilon = 0$ and $Eu = 0$. And for the model identifiability reasons, we assume the density g of the measurement error u to be known.

Let f denote density of ε, and f_0 be a known density with zero mean. Consider the problem of testing the hypothesis

$$H_0 : f = f_0 \quad v.s \quad H_1 : f \neq f_0, \tag{3.1.2}$$

based on a random sample (Y_i, Z_i), $1 \leq i \leq n$ from the joint distribution of (Y, Z) obeying the model (3.1.1).

Note that if in (3.1.1), $\beta = 0$, then Y bears no relation with X and hence whether X is observable or not is irrelevant for making inference about f. In particular any goodness-of-fit test based on $Y_i, 1 \leq i \leq n$, useful for fitting a density up to an unknown location parameter may be used to test the above hypotheses. Thus, from now onwards we shall assume $\|\beta\| \neq 0$ in this chapter.

Since we observe Z instead of X, we shall rewrite the model (3.1.1) as

$$Y = \alpha + \beta' Z + e, \quad e = \varepsilon - \beta' u.$$

Because u and ε are independent, the density of e is $h(v) = \int f(v + \beta' u) g(u) \mathrm{d}u$, $v \in \mathbb{R}$. Let $h_0(v) = \int f_0(v + \beta' u) g(u) \mathrm{d}u$, $v \in \mathbb{R}$. As argued in Koul and Song (2012), there is a one-to-one map between the densities of ε and e. Hence, testing for H_0 is equivalent to testing for

$$\mathcal{H}_0 : h = h_0, \quad \text{vs.} \quad \mathcal{H}_1 : h \neq h_0. \tag{3.1.3}$$

In the one sample i.i.d. set up, Bickel and Rosenblatt (1973) goodness-of-fit test for fitting a known density is based on an L_2 distance between a kernel density estimator and its null expected value. This test is adapted to fitting an error density up to an unknown location parameter, where the density estimator would be based on the estimated residuals. This statistics has the property that its asymptotic null distribution is not affected by not knowing the location parameter. In other words, not knowing the nuisance location parameter has no effect on asymptotic level of the test based on the analog of this statistics. What is remarkable is that this property continues to hold in several more complicated additive models. Lee and Na (2002), Bachmann and Dette (2005), and Koul and Mimoto (2012) observed that this fact continues to hold for the analog of this statistics when fitting an error density based on residuals in autoregressive and generalized autoregressive conditionally heteroscedastic time series models. This type of property makes these L_2-distance type tests more desirable, compared to the tests based on residual empirical processes, because the asymptotic null distribution of the standardized residual empirical process depends on the estimators of the underlying nuisance parameters in these models in a complicated fashion. In all of these works all data are completely observable.

In the above measurement error model, Koul and Song (2012) proposed analogous class of tests for the testing problem (3.1.3) based on kernel density estimators of h obtained from the residuals $Y_i - \hat{\alpha} - \hat{\beta}' Z_i$, $1 \leq i \leq n$, directly, where $\hat{\alpha}, \hat{\beta}$ are some $n^{1/2}$-consistent estimators of α, β, under H_0.

Alternately, because f is involved in the convolution h, it is natural to construct tests of H_0 based on a deconvolution density estimators. In this chapter we develop an analogs of the above tests for testing H_0 based on deconvolution density estimators.

There is a vast literature on the deconvolution estimators of the density of X in the measurement error model (3.1.1), as is evidenced in the papers of Carroll and Hall (1988), Stefanski and Carroll (1990), Fan (1991), van Es and Uh (2004), and Delaigle and Hall (2006) among others. The goodness-of-fit testing problem pertaining to the density function of X has been studied by several authors including Butucea (2004), Holzman and Boysen (2006), Holzman, Bissantz and Munk (2007), and Loubes and Marteau (2014). All of these authors use analogs of the above L_2-distance type tests based either on the deconvoluted estimator of density of X or on a density estimator of Z density. None of them address the above problem of testing (3.1.2) or (3.1.3) pertaining to the error density in the above measurement error model (3.1.1).

Consider the model (3.1.1) and assume for the time being α, β are known. Since we observe Y and Z, we can construct a kernel density estimator of density h of $e := Y - \alpha - \beta' Z = \varepsilon - \beta' u$, which is also an estimator of the convolution of the density f of ε with the known density of $\beta' u$. From this we obtain a deconvolution density estimator of f, which we shall use to construct tests of H_0.

Let Φ_γ denote the characteristic function of a density γ. Proceeding a bit more precisely, by the independence of ε and u, $\Phi_h(t) = \Phi_f(t)\Phi_g(-\beta t)$. Assuming $\Phi_g(t) \neq 0$, for all $t \in \mathbb{R}$, the characteristic function of ε is $\Phi_f(t) = \Phi_h(t)/\Phi_g(-\beta t)$. Using the data Y_i, $Z_i, 1 \leq i \leq n$, an estimate of Φ_h is provided by the empirical characteristic function $\Psi_n(t) := n^{-1} \sum_{j=1}^n e^{ite_j}$ of $e_j := Y_j - \alpha - \beta' Z_j$, $1 \leq j \leq n$. A kernel density estimator of h is

$$h_n(x, \alpha, \beta) = \frac{1}{nb} \sum_{j=1}^n K\Big(\frac{x - e_j}{b}\Big),$$

where K is a kernel function with its characteristic function Φ_K compactly support and $b > 0$ is a bandwidth sequence. Then the characteristic function of h_n is $\Phi_K(bt)\Psi_n(t)$. Since

Φ_g is known, a kernel estimate of $\Phi_f(t)$ is $\Phi_K(bt)\Psi_n(t)/\Phi_g(-\beta t)$. By the inversion formula,

$$f_n(x,\alpha,\beta) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\Phi_K(bt)\frac{\Psi_n(t)}{\Phi_g(-\beta t)}\,\mathrm{d}t,$$

is a deconvolution estimate of f when α and β are known. But, in practice α, β are seldom known. Let $\hat{\alpha}, \hat{\beta}$ be estimators of α, β, respectively. Then the corresponding deconvolution estimator of f is $\hat{f}_n(x) := f_n(x, \hat{\alpha}, \hat{\beta})$ obtained from f_n after replacing α, β by $\hat{\alpha}, \hat{\beta}$, respectively. The proposed class of tests, one for each K and b, of H_0 is to be based on

$$\hat{T}_n = \int_{\mathbb{R}} \left(\hat{f}_n(x) - K_b * f_0(x)\right)^2 \mathrm{d}x,$$

where for any function γ, $K_b * \gamma(x) := b^{-1}\int K((x-y)/b)\gamma(y)\mathrm{d}y$.

It is well known that the convergence rate of the deconvolution density estimators depends sensitively on the tail behaviour of the characteristic function of the underlying measurement error, which in the present set up is Φ_g. There are two general cases: one is the ordinary smooth case, where $|\Phi_g(t)|$ is of polynomial order $|t|^{-\kappa}$, for some $\kappa > 0$, as $|t| \to \infty$; the other is the super smooth case, where $|\Phi_g(t)|$ is of the order $|t|^{\lambda_0} e^{-|t|^{\lambda}/\nu}$, for some $\lambda_0 \in \mathbb{R}$, $\lambda > 0$ and $\nu > 0$, as $|t| \to \infty$. In this chapter, we obtain asymptotic distributions of $\hat{T}_n$ under H_0 in both the ordinary smooth and super smooth cases in section 2. The consistency against a fixed alternative, the asymptotic power against a class of local nonparametric alternatives and against a fixed alternative for both cases is described in section 3.

The findings of a finite sample simulation that compares the empirical power of a member of the proposed class of tests with that of the Kolmogorov–Smirnov, Cramér–von Mises tests based on the empirical d.f. of $\{\hat{e}_j := Y_j - \hat{\alpha} - \hat{\beta}' Z_j, 1 \le j \le n\}$, and a Koul and Song (2012) test based on $h_n(\cdot, \hat{\alpha}, \hat{\beta})$ are presented in section 4. The comparison is made for the three choices of the measurement error variance σ_u^2. In the ordinary smooth case, the proposed test dominates the Koul-Song test at almost all chosen alternatives for all three choices of σ_u^2. It also dominates the other two tests for the larger values of σ_u^2 at most of the chosen alternatives and for a larger sample size. The findings in the super smooth case are similar. In general the proposed test has better empirical power at the chosen alternatives compared to some of these other tests for larger values of σ_u^2, while Cramér–von Mises test dominates in terms of the empirical power for smaller values of σ_u^2. See section 3.4 for more on this finite sample comparison.

Throughout this chapter, $\mathcal{N}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, all limits are taken as $n \to \infty$, $\to_d$ and $\to_p$ denoted the convergence in distribution and probability, respectively, and the range of integration in all the integrals is $\mathbb{R}$, unless specified otherwise.

3.2 Asymptotic Null Distribution

This section discusses the asymptotic null distribution $\hat{T}_n$ for the ordinary smooth and super smooth cases.

3.2.1 Ordinary smooth case

Here we shall first derive the limiting null distribution of $\hat{T}_n$ for the ordinary smooth case. To begin with we state the needed assumptions.

(A): The characteristic function Φ_g of the error vector u satisfies $\Phi_g(t) \neq 0$, for all $t \in \mathbb{R}^p$, and $|\Phi_g(t)| \approx \|t\|^{-\kappa}$, for a $\kappa > 0$, i.e. there are c, $C > 0$ such that $c\|t\|^{-\kappa} \leq |\Phi_g(t)| \leq$

$C\|t\|^{-\kappa}$, for all sufficiently large $\|t\|$.

(B): The characteristic function Φ_f of the density f of ε satisfies $|\Phi_f(t)| = O(|t|^{-r})$, for some $r > 1$, as $|t| \to \infty$.

(C): The characteristic function Φ_K of the kernel function K is symmetric around 0 and compactly supported on $[-1, 1]$.

(D): $E\{\|X\|^4 + |\varepsilon|^4 + \|u\|^4\} < \infty$.

Next, define $\psi(\beta, s, t) := \Phi_g(\beta t + \beta s)\Phi_f(t + s)$, and let

$$
\begin{aligned}
T_n(\alpha, \beta) &:= \int \big(f_n(x, \alpha, \beta) - K_b * f_0(x)\big)^2 \, \mathrm{d}x, \quad C_{M,b} := \int \frac{|\Phi_K(tb)|^2}{|\Phi_g(\beta t)|^2} \, \mathrm{d}t, \\
C_{V,b} &:= \int\int \frac{|\Phi_K(tb)|^2 |\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2 |\Phi_g(\beta s)|^2} |\psi(\beta, s, t)|^2 \, \mathrm{d}t.
\end{aligned}
$$

Using Theorem 1 of Holzman et al. (2007) one can derive the following result. Suppose H_0 and the assumptions (A)–(C) hold and $b \to 0$, $nb \to \infty$. Then

$$
C_{M,b} \approx b^{-(2\kappa+1)}, \qquad C_{V,b} \approx b^{-(4\kappa+1)}, \tag{3.2.1}
$$

$$
nC_{V,b}^{-1/2}\Big(T_n(\alpha, \beta) - C_{M,b}/(2\pi n)\Big) \to_d \mathcal{N}\big(0, 1/2\pi^2\big). \tag{3.2.2}
$$

Note that $\hat{T}_n = T_n(\hat{\alpha}, \hat{\beta})$. Thus we need the above results to hold with α, β replaced by $\hat{\alpha}$ and $\hat{\beta}$, respectively. Accordingly, write $\hat{C}_{M,b}$, $\hat{C}_{V,b}$ and $\hat{\Psi}_n(t)$ for $C_{M,b}$, $C_{V,b}$ and $\Psi_n(t)$, when α, β are replaced by $\hat{\alpha}, \hat{\beta}$, respectively. We are now ready to state the following theorem, which provides yet another example where the asymptotic null distributions of these L_2-distance statistics are not affected by not knowing the nuisance parameters α, β.

Theorem 3.2.1 *Suppose H_0 holds, assumptions (A), (B) with $r > 3/2$, (C) and (D) hold, and that*

$$n^{1/2}\{|\hat{\alpha}-\alpha| + \|\hat{\beta}-\beta\|\} = O_p(1). \tag{3.2.3}$$

In addition, suppose $b \to 0$, and $nb^{\max\{2\kappa+3, 3.5\}} \to \infty$, with κ as in (A). Then $\hat{C}_{M,b} \approx b^{-(2\kappa+1)}$, $\hat{C}_{V,b} \approx b^{-(4\kappa+1)}$, and

$$n\hat{C}_{V,b}^{-1/2}\Big(\hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n}\Big) \to_d \mathcal{N}\big(0, \frac{1}{2\pi^2}\big). \tag{3.2.4}$$

The proof of this theorem is given in the last section. Let z_a be $(1-a)$100th percentile of the $\mathcal{N}(0,1)$ distribution. An immediate consequence of (3.2.4) is that for any $0 < a < 1$, the test that rejects H_0 whenever

$$\mathcal{T}_n := \sqrt{2}\pi n\hat{C}_{V,b}^{-1/2}\big|\hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n}\big| > z_{a/2}$$

has the asymptotic size a.

Examples of g that satisfy assumption (A) include uniform distribution with $\kappa = 1$, gamma distributions with scale γ where $\kappa = \gamma$, exponential where $\kappa = 1$, and Laplace distribution with location 0 and scale 1 where $\kappa = 2$. The class of the regression error densities f that satisfy assumption (B) includes Laplace where $r = 2$, normal and Cauchy for any $r > 0$.

3.2.2 Super smooth case

Now we consider the problem of obtaining the limiting distribution of $\hat{T}_n$ in the super smooth case. Here we need the following assumptions.

(A′): The characteristic function Φ_g of the error variable u satisfies $\Phi_g(t) \neq 0$, for any $t \in \mathbb{R}^p$. For any $\beta \in \mathbb{R}^p$, $\beta_k \neq 0$, for $k = 1, \cdots, p$, $|\Phi_g(\beta t)| \sim C(\beta)|t|^{\lambda_0} e^{-\nu(\beta)|t|^{\lambda}}$, as $|t| \to \infty$, for a $\lambda > 1$, $C(\beta) > 0$, $\nu(\beta) > 0$, and $\lambda_0 \in \mathbb{R}$. Also, $C(\beta)$, $\nu(\beta)$, exist bounded first derivatives.

(B′): The density f is square-integrable, and $E\varepsilon^2 < \infty$.

(C′): The characteristic function Φ_K of the kernel function K is symmetric around 0 and compactly supported on $[-1, 1]$. Moreover $\Phi_K(0) = 1$, and there exist $A > 0$, $\omega \geq 0$ such that

$$\Phi_K(1-t) = At^{\omega} + o(t^{\omega}), \quad \text{as} \quad t \to 0.$$

From Holzmann and Boysen (2006) we can deduce that under the conditions (A′)–(C′), as $n \to \infty$ and $b \to 0$,

$$\frac{(2\lambda)^{1+2\omega} \pi C^2(\beta) n}{A^2 \nu^{1+2\omega}(\beta) b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp\left(2\nu(\beta)/b^{\lambda}\right) \Gamma(2\omega+1)} T_n(\alpha, \beta) \to_d \chi_2^2/2, \quad (3.2.5)$$

where χ_2^2 is a r.v. having chi-square distribution with 2 degree of freedom, and $\Gamma(\cdot)$ is the Gamma function.

In order to derive a similar result for $\hat{T}_n$, we need the following additional condition. Let $\dot{q}$ be the first derivative of q for any function q.

(D′): There exists some $\lambda_1 > 1$, $\Phi_f(t)| = O(|t|^{-\lambda_1})$ as $|t| \to \infty$.

Theorem 3.2.2 *Suppose H_0 and the assumptions (A′), (B′), (C′), (D′), (D) hold, $b \to 0$, and*

$$nb^{-\eta} \exp\big(- 2\nu(\beta)/b^{\lambda}\big) \to \infty, \quad \text{for any} \quad \eta > 0. \tag{3.2.6}$$

Then

$$\mathcal{T}_{n,s} := \frac{(2\lambda)^{1+2\omega}\pi C(\hat{\beta})^2 n}{A^2 \nu(\hat{\beta})^{1+2\omega} b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp\big(2\nu(\hat{\beta})/b^{\lambda}\big)\Gamma(2\omega+1)} \hat{T}_n \to_d \chi_2^2/2. \tag{3.2.7}$$

Note that the factor multiplying $\hat{T}_n$ here is all known. Again, the proof of this theorem appears in the last section. The corresponding test is to rejects H_0 with asymptotic size a, for $0 < a < 1$, whenever $\mathcal{T}_{n,s} > \mathcal{X}_a/2$, where $\mathcal{X}_a$ is $(1-a)$100th percentile of the χ_2^2 distribution.

Examples satisfying assumption (A′) include normal densities. If g is a standard normal density then $C_g = 1$, $\lambda_0 = 0$, $\lambda = 2$ and $\nu = 2$. For kernel functions satisfying assumption (C′), Holzmann and Boysen (2006) used the sinc kernel $K(x) = \sin(x)/(\pi x)$, with $A = 1$ and $\omega = 0$, and Fan (1992) used $\Phi_K(t) = (1-t^2)^3$ with $A = 8$ and $\omega = 3$. Other suitable kernel functions can also be found in Delaigle and Hall (2006).

3.3 Consistency and Asymptotic Power

In this section we shall discuss the consistency and asymptotic power for fixed and local nonparametric alternatives of the above tests for both ordinary and super smooth cases.

Consistency. Let f_1 be another fixed density of ε such that

$$\|f_1 - f_0\| := \Big(\int \big[f_1(x) - f_0(x)\big]^2 dx\Big)^{1/2} > 0. \tag{3.3.1}$$

Consider the fixed alternatives, $H_1 : f(x) = f_1(x)$, for all $x \in \mathbb{R}$.

The following two theorems yield the consistency of the above $\mathcal{T}_n$ and $\mathcal{T}_{n,s}$ tests against H_1 for the ordinary and super smooth cases, respectively.

Theorem 3.3.1 *Suppose assumptions (A) and (C) hold, f_0 and f_1 satisfy (B) with $r > 3/2$, and have finite fourth moment, and (3.2.3) holds under H_1. Furthermore, suppose (D) holds, $b \to 0$, and $nb^{\max\{2\kappa+3,3.5\}} \to \infty$. Then*

$$\sqrt{2}\pi n \hat{C}_{V,b}^{-1/2} \big|\hat{T}_n - \frac{\hat{C}_{M,b}}{2\pi n}\big| \to_p \infty. \tag{3.3.2}$$

Theorem 3.3.2 *Assume (3.2.3) holds under H_1, and that the assumptions of Theorem 3.2.2 hold. Then*

$$\frac{n}{b^{\lambda-1+2\lambda\omega+2\lambda_0} \exp\big(2\nu(\hat{\beta})/b^{\lambda}\big)} \hat{T}_n \to_p \infty.$$

Asymptotic local power. First we consider the ordinary smooth case. We shall describe the asymptotic distribution of $\hat{T}_n$ under a sequence of the local nonparametric alternatives

$$f_{1n}(x) = f_0(x) + \delta_{1n}\ell(x), \quad x \in \mathbb{R},$$

with $\delta_{1n} = (C_{V,b}/2)^{1/4}/(n\pi)^{1/2}$, and f_{1n} a nonnegative function, $\ell \in L_2(\mathbb{R})$, and $\int \ell(x)dx = 0$. We obtain

Theorem 3.3.3 *Suppose the assumptions of Theorem 3.2.1 hold and that under* H_{1n} : $f(x) = f_{1n}(x)$, *(3.2.3) holds. Then, under under* H_{1n},

$$\sqrt{2}\pi n \hat{C}_{V,b}^{-1/2}(\hat{T}_n - \hat{C}_{M,b}/(2\pi n)) \to_d \mathcal{N}(\|\ell\|^2, 1).$$

Similarly for the super smooth case, consider a sequence of the local nonparametric alternatives

$$\begin{aligned} f_{2n}(x) &= f_0(x) + \delta_{2n}\ell(x), \quad x \in \mathbb{R}, \\ \delta_{2n} &= \Big(\frac{(2\lambda)^{1+2\omega}\pi C(\beta)^2 n}{A^2\nu(\beta)^{1+2\omega}b^{\lambda-1+2\lambda\omega+2\lambda_0}\exp\big(2\nu(\beta)/b^{\lambda}\big)\Gamma(2\omega+1)}\Big)^{-1/2}, \end{aligned}$$

with f_{2n} a nonnegative function, $\ell \in L_2(\mathbb{R})$, and $\int \ell(x)dx = 0$. We obtain

Theorem 3.3.4 *Suppose the assumptions of Theorem 3.2.2 hold and (3.2.3) holds under* $H_{2n} : f(x) = f_{2n}(x)$. *Then, under* H_{2n},

$$\frac{(2\lambda)^{1+2\omega}\pi C(\hat{\beta})^2 n}{A^2\nu(\hat{\beta})^{1+2\omega}b^{\lambda-1+2\lambda\omega+2\lambda_0}\exp\big(2\nu(\hat{\beta})/b^{\lambda}\big)\Gamma(2\omega+1)}\hat{T}_n - \|\ell\|^2 \to_d \chi_2^2/2.$$

The above two theorems show that the proposed tests can detect alternatives which converge to f_0 at a rate slower than $n^{-1/2}$.

Asymptotic power against a fixed alternative. Now we describe the asymptotic power for the ordinary smooth case against a fixed alternative f_1 such that $\|f_1 - f_0\| > 0$. To proceed further we state the following result, which follows from Theorem 2 of Holzmann et al. (2007). Assume $f_1 \neq f_0$ satisfies (3.3.1), assumptions (A) and (C) hold, f_1 and f_0 satisfy assumption (B) for some $r > \kappa + 1$, and have bounded second derivatives $b \to 0$, and (3.2.6) holds. Then, under H_1,

$$n^{1/2}\big(T_n(\alpha,\beta) - \|K_b * (f_1 - f_0)\|^2\big) \to_d \mathcal{N}(0, \tau_0^2), \tag{3.3.3}$$

where

$$\tau_0^2 = \frac{1}{2\pi^3}\mathrm{Var}\Big(\int e^{-it\varepsilon}\frac{\Phi_{f_1}(t) - \Phi_{f_0}(t)}{\Phi_g(\beta t)}\,\mathrm{d}t\Big).$$

We shall use this result to analyze the asymptotic distribution of $\hat{T}_n$ under the fixed alternative H_1. To proceed further, let $\mu_Z := EZ$, and suppose the first derivatives $\dot{f}_1$ and $\dot{f}_0$ exist. Define

$$A_f = 2\int (f_1 - f_0)\dot{f}_0(x)\,\mathrm{d}x, \quad B_f = 2\mu_Z\int (f_1 - f_0)\dot{f}_1(x)\,\mathrm{d}x.$$

Theorem 3.3.5 *Assume that (A), (C) and (D) hold, f_1 and f_0 satisfy assumption (B) with $r > \kappa + 1$, $r > 3/2$, and κ as in (A) and have bounded second derivatives. Also, assume (3.3.1) and (3.2.3) hold under H_1. Furthermore, if $b \to 0$, $nb^{\max\{4\kappa+2, 2\kappa+3\}} \to \infty$, then*

$$n^{1/2}\Big(\hat{T}_n - \|K_b * (f_1 - f_0)\|^2 - (\hat{\alpha} - \alpha)A_f - (\hat{\beta} - \beta)'B_f\Big) \to_d \mathcal{N}(0, \tau_0^2). \tag{3.3.4}$$

Note that the effect of estimating α and β introduces another bias term $n^{1/2}((\hat{\alpha} - \alpha)A_f + (\hat{\beta} - \beta)'B_f)$ in the asymptotic distribution of the statistics $\hat{T}_n$. This bias will vanish if to begin with there is no intercept parameter in the model and $\mu_Z = 0$. It also vanishes under the following linearity condition on the estimators.

Furthermore, suppose under H_1, the estimators $\hat{\alpha}$ and $\hat{\beta}$ satisfy the following expansion:

$$\hat{\alpha} - \alpha = \frac{1}{n}\sum_{j=1}^{n} \eta_j + o_p(n^{-1/2}), \tag{3.3.5}$$

$$\hat{\beta}_k - \beta_k = \frac{1}{n}\sum_{j=1}^{n} \zeta_{jk} + o_p(n^{-1/2}), \quad k = 1, \cdots, p, \tag{3.3.6}$$

where η_j are i.i.d. with $\mathrm{E}\eta = 0$, $\mathrm{Var}(\eta) > 0$, $\mathrm{E}|\eta|^{2+\vartheta} < \infty$, for some $\vartheta > 0$. Moreover, the same conditions are satisfied by ζ_{jk}'s, and also for $i \neq j \neq k$ η_i, ζ_j and e_k are mutually independent.

Examples of the estimators of $\hat{\alpha}, \hat{\beta}$ that satisfy these two conditions include the naive least square estimators, maximum likelihood estimators (see Huăková and Meintanis (2007)), and the bias-corrected estimators (see Fuller (1987)). Using the above expansion, we obtain

Theorem 3.3.6 *Assume the conditions of Theorem 3.3.5 and (3.3.5)-(3.3.6) for $\hat{\alpha}$ and $\hat{\beta}$ hold. Then, for some $\tau > 0$,*

$$n^{1/2}\big(\hat{T}_n - \|K_b * (f_1 - f_0)\|^2\big) \to_d \mathcal{N}(0, \tau^2). \tag{3.3.7}$$

The form of τ is described in the proof of this theorem in the last section, see (3.5.26). Although τ is complicated to calculate in practice, the bootstrap simulation methods can be used to estimate τ.

For the super smooth case, in order to obtain a similar result as above, we need to make the following stronger assumptions on f_1 and f_0:

(B*) The characteristic function Φ_f of the density f of ε satisfies $|\Phi_f(t)| = O(|t|^{\xi_0} e^{-|t|^{\xi}/\zeta})$ for some $\xi_0 \in \mathbb{R}$, $\zeta > 0$ and $\xi > \lambda$.

Assumption (B^*) implies (D'), and assures $\int \left|\Phi_f(t)/\Phi_g(\beta t)\right| \, dt < \infty$. An example of f and g satisfying the above condition is where f is a normal density with variance smaller than 1, and g is standard normal density.

A result analogous to (3.3.3) can be obtained in the super smooth case also by following the proof of Theorem 2 in Holzmann et al. (2007) with known α and β. To be clear, assume f_1, f_0 satisfying (3.3.1), assumptions (A') and (C') hold, and f_1 and f_0 satisfy assumption (B^*). Assume $b \to 0$, and

$$nb^{-\eta} \exp\big(- 4\nu(\beta)/b^{\lambda}\big) \to \infty, \quad \text{for any} \quad \eta > 0. \tag{3.3.8}$$

Then (3.3.3) holds. In the case of unknown α and β, we obtain the following theorem.

Theorem 3.3.7 *Suppose assumptions (A'), (C'), and (B^*) hold, f_1, f_0 satisfy (3.3.1), and have bounded second derivatives. If, in addition, $b \to 0$, and (3.2.6) holds, then we have (3.3.4).*

Furthermore, if $\hat{\alpha}$ and $\hat{\beta}$ satisfy (3.3.5)-(3.3.6), then (3.3.7) holds for some $\tau > 0$.

3.4 Simulations

In this section we report the findings of some extensive simulations, which assess some finite sample level and power behavior of a member of the above class of tests. The results are presented in the two subsections for ordinary and super smooth cases.

3.4.1 Ordinary smooth case

Consider the measurement error model

$$Y = 1 + X + \varepsilon, \quad Z = X + u, \tag{3.4.1}$$

where $X \sim \mathcal{N}(0,1)$ and $\Phi_g(t) = 16/(4+\sigma_u^2 t^2)^2$. This Φ_g satisfies assumption (A) of the ordinary smooth case with $\kappa = 4$. We wish to test the hypothesis that $\varepsilon \sim \mathcal{N}(0, 0.25)$, i.e., f_0 in H_0 is the density of normal distribution with mean zero and variance 0.25. As in Koul and Song (2012), we use the bias-corrected estimators $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{Z}$ and $\hat{\beta} = S_{ZY}/(S_{ZZ} - \sigma_u^2)$, where $\bar{Y}$ and $\bar{Z}$ denote the sample mean of Y and Z, and S_{ZY} and S_{ZZ} denote the sample covariance of Z and Y and the sample variance of Z, respectively. In the deconvolution estimator of f, we used the sinc kernel $K(x) = \sin x/(\pi x)$. The proposed test based on $\hat{T}_n$ rejects H_0 for the large values of $\hat{\mathcal{T}}_n := n\hat{C}_{V,b}^{-1/2}|\hat{T}_n - \hat{C}_{M,b}/(2\pi n)|$.

We shall compare this test with with the Kolmogorov-Smirnov (T_{KS}), the Cramér-von Mises (T_{CvM}) tests and the W_n test proposed by Koul and Song (2012), all based directly on residuals $\hat{e}_i := Y_i - \hat{\alpha} - \hat{\beta}Z_i$, $1 \le i \le n$. The first two statistics are defined as

$$T_{KS} := \sup_{x \in \mathbb{R}} n^{1/2}|\hat{F}_n(x) - F_0(x)|, \quad T_{CvM} := n \int \left(\hat{F}_n(x) - F_0(x)\right)^2 \mathrm{d}F_0(x),$$

where $\hat{F}_n(x) := n^{-1}\sum_{i=1}^n I(\hat{e}_i \le x)$. To define W_n, let φ be a density kernel on $\mathbb{R}$, $\varphi_2(u) := \int \varphi(v)\varphi(u+v)\,\mathrm{d}v$, $c \equiv c_n$ be another window width, w be a compactly supported density on $\mathbb{R}$, and let

$$\tilde{h}_n(x) := \frac{1}{nc}\sum_{j=1}^{n}\varphi\Big(\frac{x-\hat{e}_j}{c}\Big), \quad \hat{C}_n := \frac{1}{n^2c^2}\sum_{i=1}^{n}\int \varphi^2\Big(\frac{v-\hat{e}_i}{c}\Big)w(v)\,\mathrm{d}v,$$
$$\hat{\Gamma}_n := 2\int \tilde{h}_n^2(x)w^2(x)dx\int \big[\varphi_2(u)\big]^2\mathrm{d}u.$$

Then, with $h_0(x,\hat{\beta}) := \int f_0(x+\hat{\beta}u)dx$,

$$W_n := nb^{1/2}\hat{\Gamma}_n^{-1}\Big|\int \big(\tilde{h}_n(x)-h_0(x,\hat{\beta})\big)^2 w(x)\,\mathrm{d}x - \hat{C}_n\Big|.$$

In this simulation study, we chose the kernel function φ to be the standard normal density and the bandwidth $c = n^{-0.27}$, and $w(\cdot)$ was chosen to be the uniform density on the closed interval $[-6, 6]$. All these three tests reject H_0 for large values of their corresponding statistics.

To assess the effect of the measurement error on the finite sample level and power of these tests, we conducted simulations for the three values of $\sigma_u^2 = 0.25$, $\sigma_u^2 = 0.5$, $\sigma_u^2 = 1$ with bandwidth $b = 0.5n^{-1/12}$,$b = 0.65n^{-1/12}$, and $b = 0.8n^{-1/12}$, respectively .

It is well known that the approximation of the distributions of the test statistics based on density estimators by their asymptotic distributions is generally slow. For that reason in this simulation study we use the Monte Carlo simulation method to obtain the critical values for all tests considered. At level 0.05, the critical values of all four tests are simulated by the Monte Carlo method, based on sample size 300 and 500, repeating 1000 times. The 95% quantiles are calculated for 1000 repetitions and the mean values of the 1000 quantiles are chosen as the critical values given in Table 3.1 for different values of σ_u^2. The Monte Carlo level of the three tests $\hat{\mathcal{T}}_n$, T_{KS} and T_{CvM} is relatively more robust against the variation in the measurement error, compared to that of the W_n test.

n	σ_u^2	$\hat{\mathcal{T}}_n$	T_{KS}	T_{CvM}	W_n
300	0.25	1.18118	0.86714	0.16988	1.34799
	0.5	1.14531	0.87355	0.17395	1.40213
	1	1.15600	0.88576	0.18056	1.49112
500	0.25	1.19852	0.86223	0.16929	1.40258
	0.5	1.20847	0.86674	0.17357	1.44800
	1	1.20484	0.87829	0.18073	1.52883

Table 3.1: Monte Carlo critical values of all the tests, ordinary smooth case.

The alternatives considered here are t-distributions with k degrees of freedom, denoted by t_k, for $k = 4, 6, 8, 10, 15, 20$, double exponential (DE) and logistic (L) distributions, all having zero mean and standard deviation 0.5. The sample sizes chosen are 300 and 500 and the level is .05. From Table 3.2, we see that in terms of the empirical power, the $\hat{\mathcal{T}}_n$ dominates the W_n test uniformly across the chosen alternatives, sample sizes and the values of σ_u^2, while for $n = 500$, it also dominates the T_{KS} test for at almost all chosen alternatives, when $\sigma_u^2 = .5, 1$, while the T_{CvM} test dominates all other tests for the smallest value of σ_u^2.

We also considered the following normal and logistic mixture alternatives.

$$\begin{aligned} f_1 &= 0.5\mathcal{N}(-\mu, 0.25) + 0.5\mathcal{N}(\mu, 0.25), \quad \mu > 0, \\ f_2 &= 0.5\ell(-\lambda, 1.5/\pi) + 0.5\ell(\lambda, 1.5/\pi), \quad \lambda \geq 0, \end{aligned}$$

where $\ell(a, b)$ is the density of the logistic d.f. $1/(1 + e^{-\frac{x-a}{b}})$. The empirical powers for normal and logistic mixture alternatives are given in Table 3.3. In both cases the sample sizes are 300 and 500 and the level is 0.05. From Table 3.3 one observes the following. First, as σ_u^2 increases, the empirical powers decrease generally. Secondly, for the alternatives f_1 and $\sigma_u^2 = 1$, the proposed test $\hat{\mathcal{T}}_n$ based on deconvolution density estimator has larger empirical powers than the T_{KS} and W_n tests in most of the cases, while the T_{CvM} test dominates all other three tests in terms of the empirical power. For $\sigma_u^2 = 0.25$, the W_n test dominates all three tests $\hat{\mathcal{T}}_n$, T_{KS} and T_{CvM} at all normal mixture alternatives. For $\sigma_u^2 = 0.5$, the $\hat{\mathcal{T}}_n$ test has larger empirical powers than T_{KS}, but smaller empirical powers than T_{CvM} and W_n. Similar phenomena can be found from Table 3.3 for the alternatives f_2.

n	σ_u^2	Test	t_4	t_6	t_8	t_{10}	t_{15}	t_{20}	DE	L
300	0.25	$\hat{\mathcal{T}}_n$	0.240	0.093	0.072	0.052	0.047	0.060	0.183	0.059
		T_{KS}	0.191	0.084	0.062	0.058	0.041	0.053	0.150	0.064
		T_{CvM}	0.268	0.103	0.065	0.053	0.053	0.062	0.226	0.068
		W_n	0.035	0.023	0.023	0.027	0.027	0.039	0.033	0.036
	0.5	$\hat{\mathcal{T}}_n$	0.093	0.072	0.038	0.051	0.052	0.047	0.050	0.050
		T_{KS}	0.064	0.057	0.055	0.054	0.060	0.053	0.051	0.049
		T_{CvM}	0.100	0.077	0.065	0.052	0.056	0.051	0.071	0.049
		W_n	0.018	0.037	0.038	0.043	0.040	0.046	0.032	0.034
	1	$\hat{\mathcal{T}}_n$	0.046	0.050	0.043	0.040	0.049	0.049	0.046	0.048
		T_{KS}	0.052	0.046	0.046	0.050	0.055	0.047	0.047	0.040
		T_{CvM}	0.042	0.052	0.047	0.050	0.057	0.060	0.054	0.047
		W_n	0.037	0.041	0.050	0.039	0.046	0.038	0.043	0.038
500	0.25	$\hat{\mathcal{T}}_n$	0.397	0.158	0.092	0.078	0.068	0.060	0.344	0.097
		T_{KS}	0.244	0.122	0.082	0.077	0.050	0.044	0.223	0.076
		T_{CvM}	0.398	0.159	0.101	0.083	0.054	0.053	0.315	0.090
		W_n	0.051	0.037	0.027	0.032	0.034	0.035	0.059	0.025
	0.5	$\hat{\mathcal{T}}_n$	0.131	0.052	0.049	0.054	0.043	0.049	0.107	0.044
		T_{KS}	0.113	0.070	0.055	0.062	0.052	0.064	0.106	0.053
		T_{CvM}	0.162	0.059	0.057	0.062	0.045	0.069	0.129	0.052
		W_n	0.059	0.043	0.034	0.041	0.048	0.049	0.043	0.037
	1	$\hat{\mathcal{T}}_n$	0.069	0.050	0.045	0.052	0.061	0.047	0.060	0.063
		T_{KS}	0.059	0.049	0.043	0.044	0.053	0.050	0.049	0.050
		T_{CvM}	0.058	0.046	0.042	0.056	0.057	0.049	0.068	0.059
		W_n	0.042	0.034	0.054	0.047	0.048	0.041	0.034	0.048

Table 3.2: Empirical powers against chosen alternatives, ordinary smooth case.

			μ				λ			
n	σ_u^2	Test	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8
300	0.25	$\hat{\mathcal{T}}_n$	0.084	0.672	0.999	1.000	0.252	0.523	0.767	0.929
		T_{KS}	0.078	0.577	0.998	1.000	0.236	0.460	0.703	0.889
		T_{CvM}	0.101	0.787	1.000	1.000	0.367	0.651	0.862	0.966
		W_n	0.137	0.818	1.000	1.000	0.419	0.687	0.874	0.968
	0.5	$\hat{\mathcal{T}}_n$	0.071	0.457	0.973	1.000	0.162	0.363	0.551	0.742
		T_{KS}	0.078	0.350	0.920	0.998	0.163	0.278	0.431	0.645
		T_{CvM}	0.082	0.514	0.985	1.000	0.226	0.430	0.631	0.824
		W_n	0.095	0.473	0.974	1.000	0.206	0.403	0.584	0.772
	1	$\hat{\mathcal{T}}_n$	0.079	0.262	0.729	0.964	0.147	0.209	0.291	0.450
		T_{KS}	0.059	0.199	0.593	0.924	0.117	0.164	0.228	0.347
		T_{CvM}	0.079	0.265	0.760	0.974	0.153	0.233	0.328	0.496
		W_n	0.075	0.212	0.655	0.949	0.132	0.181	0.255	0.404
500	0.25	$\hat{\mathcal{T}}_n$	0.113	0.893	1.000	1.000	0.390	0.747	0.947	0.992
		T_{KS}	0.095	0.835	1.000	1.000	0.393	0.697	0.913	0.987
		T_{CvM}	0.143	0.955	1.000	1.000	0.586	0.879	0.985	1.000
		W_n	0.172	0.955	1.000	1.000	0.585	0.872	0.982	1.000
	0.5	$\hat{\mathcal{T}}_n$	0.063	0.646	0.998	1.000	0.300	0.520	0.776	0.922
		T_{KS}	0.091	0.529	0.990	1.000	0.247	0.444	0.668	0.858
		T_{CvM}	0.084	0.709	0.999	1.000	0.373	0.619	0.845	0.961
		W_n	0.094	0.639	0.999	1.000	0.322	0.531	0.790	0.934
	1	$\hat{\mathcal{T}}_n$	0.089	0.394	0.898	0.996	0.182	0.304	0.446	0.627
		T_{KS}	0.065	0.305	0.801	0.988	0.156	0.218	0.379	0.528
		T_{CvM}	0.084	0.415	0.930	1.000	0.197	0.342	0.484	0.681
		W_n	0.073	0.320	0.855	0.995	0.153	0.265	0.373	0.550

Table 3.3: Empirical powers against mixture normal (left panel) and logistic alternatives, ordinary smooth case.

3.4.2 Super smooth case

Now consider the measurement error model (3.4.1), where again $X \sim \mathcal{N}(0,1)$, $\varepsilon \sim \mathcal{N}(0, 0.25)$ but $u \sim \mathcal{N}(0, \sigma_u^2)$. The bias-corrected estimators are also used to estimate α and β. The sinc kernel $K(x) = \sin x/(\pi x)$ is consider for the deconvolution kernel estimator, with the bandwidth $b = 0.55(\log n)^{-0.5}$, $b = (\sqrt{0.5} + 0.05)(\log n)^{-0.5}$ and $b = 1.15(\log n)^{-0.4}$ when $\sigma_u^2 = 0.25$, $\sigma_u^2 = 0.5$ and $\sigma_u^2 = 1$, respectively. Thus, $C_g = 1$, $\nu = 2/\sigma_u^2$, $\lambda_0 = 0$, $\lambda = 2$, $A = 1$, and $\omega = 0$ in equation (3.2.4). Then the left side of (3.2.7) can be written as

$$\hat{\mathcal{T}}_{n,s} := \frac{2\pi n \sigma_u^2 \hat{\beta}^2 \hat{T}_n}{b \exp(|\hat{\beta}\sigma_u|^2/b^2)}.$$

The Monte Carlo distribution of $\hat{\mathcal{T}}_{n,s}$ for the sample size 1000 based on 1000 repetitions is very close to $\chi_2^2/2$. Hence the critical values of this test are obtained from $\chi_2^2/2$ distribution. To examine the power, we compared our test with the same three direct tests as in the previous section. We generated the critical values for T_{KS}, T_{CvM} and W_n defined as above by Monte Carlo methods, based on 500 and 1000 sample size, repeated 1000 times. The 95% quantiles are calculated for 1000 repetitions and the mean values of 1000 these quantiles are chosen as the critical values. These critical values are listed in Table 3.4.

n	σ_u^2	T_{KS}	T_{CvM}	W_n
500	0.25	0.85655	0.16540	1.39447
	0.5	0.85670	0.16500	1.45467
	1	0.85545	0.16446	1.53780
1000	0.25	0.85038	0.16482	1.46119
	0.5	0.85183	0.16535	1.51706
	1	0.85210	0.16492	1.59195

Table 3.4: Monte Carlo critical values of the T_{KS} , T_{CvM}, and W_n, super smooth case.

We consider the same alternative as in the ordinary smooth case of the subsection 3.4.1.

The empirical powers against t, double exponential and logistics distributions are given in Table 3.5. From this table one sees that the proposed deconvolution test provides the largest empirical powers in all the cases compared to the other three testing methods when $\sigma_u^2 = 1$, while it dominates the W_n test for smaller values of σ_u^2. The empirical powers against normal and logistics mixture alternatives are given in Table 3.6, for sample size 500 and 1000. From this table we see that for both normal and logistic mixture alternatives, the W_n test dominates the $\hat{\mathcal{T}}_{n,s}$ and T_{KS} tests for all chosen sample sizes and for all values of σ_u^2, while the T_{CvM} test dominates all other tests uniformly.

n	σ_u^2	Test	t_3	t_4	t_5	t_6	t_8	t_{10}	DE	L
500	0.25	$\hat{\mathcal{T}}_{n,s}$	0.207	0.118	0.111	0.087	0.072	0.061	0.121	0.075
		T_{KS}	0.451	0.184	0.119	0.090	0.071	0.062	0.130	0.071
		T_{CvM}	0.694	0.300	0.160	0.120	0.082	0.066	0.206	0.082
		W_n	0.146	0.037	0.037	0.035	0.029	0.040	0.034	0.035
	0.5	$\hat{\mathcal{T}}_{n,s}$	0.143	0.105	0.087	0.080	0.065	0.069	0.105	0.072
		T_{KS}	0.152	0.086	0.073	0.066	0.053	0.045	0.073	0.051
		T_{CvM}	0.250	0.104	0.082	0.059	0.055	0.058	0.088	0.045
		W_n	0.043	0.047	0.051	0.039	0.046	0.051	0.045	0.034
	1	$\hat{\mathcal{T}}_{n,s}$	0.155	0.093	0.088	0.071	0.074	0.058	0.070	0.075
		T_{KS}	0.084	0.050	0.046	0.044	0.054	0.047	0.054	0.051
		T_{CvM}	0.087	0.059	0.052	0.047	0.053	0.049	0.057	0.050
		W_n	0.043	0.047	0.044	0.045	0.049	0.050	0.044	0.054
1000	0.25	$\hat{\mathcal{T}}_{n,s}$	0.262	0.117	0.096	0.067	0.066	0.062	0.165	0.062
		T_{KS}	0.714	0.364	0.193	0.123	0.090	0.067	0.274	0.080
		T_{CvM}	0.960	0.551	0.304	0.171	0.114	0.088	0.403	0.101
		W_n	0.499	0.098	0.051	0.036	0.038	0.037	0.095	0.044
	0.5	$\hat{\mathcal{T}}_{n,s}$	0.184	0.085	0.079	0.066	0.046	0.060	0.081	0.075
		T_{KS}	0.273	0.102	0.090	0.072	0.056	0.048	0.100	0.046
		T_{CvM}	0.468	0.164	0.120	0.090	0.060	0.053	0.095	0.048
		W_n	0.092	0.041	0.041	0.044	0.041	0.049	0.046	0.050
	1	$\hat{\mathcal{T}}_{n,s}$	0.199	0.116	0.088	0.078	0.074	0.063	0.081	0.073
		T_{KS}	0.094	0.062	0.051	0.054	0.048	0.044	0.049	0.047
		T_{CvM}	0.135	0.072	0.057	0.053	0.063	0.047	0.050	0.039
		W_n	0.066	0.035	0.047	0.058	0.043	0.055	0.050	0.042

Table 3.5: Empirical powers against alternative distributions, super smooth case.

			m				λ			
n	σ_u^2	Test	0.2	0.4	0.6	0.8	1.2	1.4	1.6	1.8
500	0.025	$\hat{\mathcal{T}}_{n,s}$	0.048	0.140	0.547	0.890	0.092	0.134	0.189	0.305
		T_{KS}	0.079	0.800	1.000	1.000	0.389	0.645	0.853	0.974
		T_{CvM}	0.136	0.930	1.000	1.000	0.565	0.844	0.961	0.997
		W_n	0.143	0.908	1.000	1.000	0.502	0.810	0.942	0.997
	0.5	$\hat{\mathcal{T}}_{n,s}$	0.048	0.065	0.207	0.599	0.079	0.105	0.165	0.240
		T_{KS}	0.042	0.171	0.803	0.997	0.199	0.326	0.549	0.765
		T_{CvM}	0.049	0.263	0.938	1.000	0.297	0.512	0.732	0.897
		W_n	0.045	0.191	0.835	0.999	0.195	0.362	0.558	0.783
	1	$\hat{\mathcal{T}}_{n,s}$	0.040	0.042	0.352	0.925	0.042	0.071	0.159	0.291
		T_{KS}	0.049	0.099	0.452	0.865	0.115	0.164	0.286	0.342
		T_{CvM}	0.059	0.128	0.590	0.950	0.145	0.237	0.412	0.523
		W_n	0.054	0.085	0.359	0.828	0.081	0.127	0.227	0.279
1000	0.025	$\hat{\mathcal{T}}_{n,s}$	0.063	0.150	0.684	0.979	0.076	0.110	0.208	0.356
		T_{KS}	0.154	0.983	1.000	1.000	0.632	0.929	0.996	0.998
		T_{CvM}	0.225	1.000	1.000	1.000	0.824	0.984	0.997	0.998
		W_n	0.187	0.994	1.000	1.000	0.755	0.966	0.997	0.998
	0.5	$\hat{\mathcal{T}}_{n,s}$	0.070	0.153	0.553	0.892	0.070	0.108	0.169	0.271
		T_{KS}	0.096	0.742	1.000	1.000	0.361	0.634	0.879	0.984
		T_{CvM}	0.147	0.898	1.000	1.000	0.557	0.803	0.964	0.996
		W_n	0.044	0.631	1.000	1.000	0.342	0.610	0.862	0.973
	1	$\hat{\mathcal{T}}_{n,s}$	0.034	0.275	0.968	1.000	0.094	0.231	0.365	0.625
		T_{KS}	0.067	0.368	0.933	1.000	0.174	0.295	0.484	0.660
		T_{CvM}	0.083	0.511	0.977	1.000	0.252	0.420	0.649	0.837
		W_n	0.052	0.238	0.879	1.000	0.126	0.199	0.358	0.508

Table 3.6: Empirical powers against mixture normal (left panel) and logistic distributions, super smooth case.

3.5 Proofs

Here we present proof of Theorems 3.2.1–3.3.7. We write $T_n := T_n(\alpha, \beta)$ and $f_n(x) := f_n(x, \alpha, \beta)$ with known α, β and $\hat{f}_n(x) := f_n(x, \hat{\alpha}, \hat{\beta})$ for expressions simplicity.

Since $C_{V,b} \approx b^{-(4\kappa+1)}$, we first show

$$nb^{2\kappa} \int (\hat{f}_n - f_n)^2(x)\, \mathrm{d}x = o_p(1). \tag{3.5.1}$$

Using Parseval's equation, we have

$$\begin{aligned}
&\int (\hat{f}_n - f_n)^2(x)\, \mathrm{d}x \qquad (3.5.2)\\
&= \frac{1}{4\pi^2} \int \Big(\int e^{-itx} \Phi_K(ht) \big(\frac{\hat{\Psi}_n(t)}{\Phi_g(-\hat{\beta}t)} - \frac{\Psi_n(t)}{\Phi_g(-\beta t)} \big)\, \mathrm{d}t \Big)^2\, \mathrm{d}x \\
&= \frac{1}{2\pi} \int |\Phi_K(ht)|^2 \Big| \frac{\hat{\Psi}_n(t)}{\Phi_g(-\hat{\beta}t)} - \frac{\Psi_n(t)}{\Phi_g(-\beta t)} \Big|^2\, \mathrm{d}t \\
&\leq \frac{1}{2\pi} \int |\Phi_K(ht)|^2 \frac{|\hat{\Psi}_n(t) - \Psi_n(t))|^2}{|\Phi_g(-\hat{\beta}t)|^2} dt \\
&\qquad + \frac{1}{2\pi} \int |\Phi_K(bt)\Psi_n(t)|^2 \frac{|\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)|^2}{|\Phi_g(-\hat{\beta}t)\Phi_g(-\beta t)|^2}\, \mathrm{d}t \\
&= \frac{1}{2\pi} S_1 + \frac{1}{2\pi} S_2, \quad \text{say.}
\end{aligned}$$

Since Φ_K is supported on $[-1, 1]$, $\Phi_K(bt) = 0$, for $|t| > 1/b$. Thus in the above two integrals, $t \in [-1/b, 1/b]$. Since $\mu_g := \int |x| g(x) dx < \infty$, $\dot{\Phi}_g$ exists and is uniformly bounded above by μ_g. This fact together with (3.2.3) and assumption (A) imply,

$$
\begin{aligned}
|\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)| &\leq \mu_g\,|t|\,\|\hat{\beta} - \beta\|, && (3.5.3)\\
\max_{|t|\leq 1/b} \Big|\frac{\Phi_g(-\hat{\beta}t)}{\Phi_g(-\beta t)} - 1\Big| &= \max_{|t|\leq 1/b} \Big|\frac{\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)}{\Phi_g(-\beta t)}\Big| && (3.5.4)\\
&= O_p(n^{-1/2}b^{-\kappa-1}).
\end{aligned}
$$

Let $A_n := \{|\Phi_g(-\hat{\beta}t)| \geq |\Phi_g(-\beta t)|/2,\, t \in [-1/b, 1/b]\}$. Since $nb^{2\kappa+3} \to \infty$, (3.5.4) implies $P(A_n) \to 1$. Thus we need only to restrict our attention to A_n.

Consider S_2. Conditions (A) and (B) imply that there exists a M, c_β, C_β and C_f, such that for all $|t| > M$, $c_\beta|t|^{-\kappa} \leq |\Phi_g(\beta t)| \leq C_\beta|t|^{-\kappa}$ and $\Phi_f(t) \leq C_f|t|^{-r}$. Take n large enough so that $M < 1/b$. Split the integral in S_2 into two ranges, one with $|t| \leq M$ and the other with $|t| > M$. Then by (3.2.3) and (3.5.3) we obtain that on the event A_n, S_2 is bounded from the above by

$$
\begin{aligned}
&4\mu_g^2\|\hat{\beta} - \beta\|^2 \int_{1/b\geq|t|>M} \frac{|t\Phi_K(bt)\Psi_n(t)|^2}{|\Phi_g(-\beta t)|^4}\,\mathrm{d}t + O_p(n^{-1})\\
&\leq 8\mu_g^2\|\hat{\beta} - \beta\|^2\Big\{\int_{1/b\geq|t|>M}\Big[\frac{|t\Phi_K(bt)|^2|\Psi_n(t) - \Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4}\\
&\qquad + \frac{|t\Phi_K(bt)\Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4}\Big]\mathrm{d}t\Big\} + O_p(n^{-1}).
\end{aligned}
$$

By the Parseval's identity

$$
T_n(\alpha, \beta) = \frac{1}{2\pi}\int \frac{|\Phi_K(bt)|^2|\Psi_n(t) - \Phi_h(t)|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t = O_p(n^{-1}b^{-2\kappa-1}), \qquad (3.5.5)
$$

because of (3.2.1) and (3.2.2). Because $|\Phi_g(\beta t)|^{-2} \leq c_\beta^2|t|^{2\kappa}$, the first term within the curly brackets in the above bound is bounded above by $b^{-2\kappa-2}T_n(\alpha, \beta) = O_p(n^{-1}b^{-4\kappa-3})$.

Similarly, assumptions (A) and (B) imply

$$\int_{|t|>M} \frac{|t\Phi_K(bt)\Phi_h(t)|^2}{|\Phi_g(-\beta t)|^4}\,\mathrm{d}t = \int_{|t|>M} \frac{|t\Phi_K(bt)\Phi_f(t)|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t = O(b^{\min(2r-2\kappa-3,0)}).$$

Hence, in view of (3.2.3),

$$S_2 = O_p(n^{-2}b^{-4\kappa-3}) + O_p(n^{-1}b^{\min(2r-2\kappa-3,0)}) = o_p(n^{-1}b^{-2\kappa}). \tag{3.5.6}$$

Next, to analyze S_1. Let

$$\begin{aligned}
S_{11} &:= \frac{1}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n t(\hat\beta-\beta)'Z_j e^{it(Y_j-\alpha-\beta'Z_j)}|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t,\\
S_{12} &:= \frac{1}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n t e^{it(Y_j-\alpha-\beta'Z_j)}|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t,\\
S_{13} &:= \frac{1}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n t((\hat\beta-\beta)'Z_j)^2 e^{it(Y_j-\alpha-\beta'Z_j)}|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t.
\end{aligned}$$

Using the fact $Y_j - \alpha - \beta' Z_j = \varepsilon_j - \beta' u_j$, we obtain on the event A_n,

$$
\begin{aligned}
& S_1 \qquad\qquad (3.5.7) \\
\leq\ & 4 \int |\Phi_K(bt)|^2 \frac{|\hat{\Psi}_n(t) - \Psi_n(t)|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t \\
\leq\ & \frac{16}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t(\hat{\beta} - \beta)' Z_j e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t \\
& + \frac{16(\hat{\alpha} - \alpha)^2}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t \\
& + \frac{16}{n^2 b^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t((\hat{\beta} - \beta)' Z_j)^2 e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t \\
& + \frac{16(\hat{\alpha} - \alpha)^4}{n^2 b^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t + O_p\big(n^{-3} b^{-2\kappa - 7}\big) \\
=\ & 16[S_{11} + (\hat{\alpha} - \alpha)^2 S_{12} + b^{-2} S_{13} + (\hat{\alpha} - \alpha)^4 b^{-2} S_{12}] + O_p\big(n^{-3} b^{-2\kappa - 7}\big),
\end{aligned}
$$

by (3.2.3), assumption (A), and the fact that $\sum_{j=1}^n |Z_j|^3 = O_p(n)$.

Now, consider S_{11}.

$$
S_{11} \leq p \sum_{k=1}^{p} \frac{(\hat{\beta}_k - \beta_k)^2}{n^2} \int \frac{|\Phi_K(bt)|^2 |\sum_{j=1}^n t Z_{kj} e^{it(\varepsilon_j - \beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\, \mathrm{d}t
$$

Since X, u and ε are mutually independent, for any $k = 1, \cdots, p$,

$$
\mathrm{E} Z_k e^{it(\varepsilon_j - \beta' u_j)} = \mathrm{E} X_k \Phi_h(t) + \Phi_f(t) \mathrm{E} u_k e^{-it\beta' u}.
$$

We use this to obtain

$$
\begin{aligned}
&\frac{1}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n tZ_{kj}e^{it(\varepsilon_j-\beta' u_j)}|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t \\
\le\ &\frac{3}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n [Z_{kj}e^{it(\varepsilon_j-\beta' u_j)} - \mathrm{E}Z_{kj}e^{it(\varepsilon_j-\beta' u_j)}]|^2}{b^2|\Phi_g(-\beta t)|^2}\,\mathrm{d}t \\
&+\frac{3}{n^2}\int \frac{|\Phi_K(bt)|^2\left|\sum_{j=1}^n \mathrm{E}X_k t\Phi_h(t)]\right|^2}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t \\
&+\frac{3}{n^2}\int \frac{|\Phi_K(bt)|^2|\sum_{j=1}^n t\Phi_f(t)\mathrm{E}u_k e^{-it\beta' u}|}{|\Phi_g(-\beta t)|^2}\,\mathrm{d}t.
\end{aligned}
\tag{3.5.8}
$$

An argument similar to the one used in the proof of Theorem 1 in Holzmann et al. (2007) implies that the first summand in the upper bound of (3.5.8) is $O_p(n^{-1}b^{-2\kappa-3})$. The second summand is $O_p(1)$, by assumption (B), and $\Phi_h(t)/\Phi_g(-\beta t) = \Phi_f(t)$. To analyze the third summand in the upper bound of (3.5.8), decompose the integral into two ranges, $|t| > M$ and $|t| \le M$, and use the conditions (A)-(B) to show that the term with integration over $|t| \le M$ is $O_p(1)$, while the term with $|t| > M$ is of the order $O_p(b^{\min(2r-2\kappa-3,0)})$, thereby showing that the third summand in (3.5.8) is of the order $O_p(1) + O_p(b^{2r-2\kappa-3})$. Thus

$$
S_{11} = O_p(n^{-1}b^{-2\kappa-3}) + O_p(1) + O_p(b^{2r-2\kappa-3}). \tag{3.5.9}
$$

Similarly one obtains that S_{12} and S_{13} are of the same order as S_{11}. Then (3.5.7), (3.5.9), $nb^{2\kappa+3} \to \infty$, $nb^{7/2} \to \infty$ imply

$$
nb^{2\kappa}S_1 = O_p(n^{-1}b^{-3}) + O_p(b^{2\kappa}) + O_p(b^{2r-3}) + O_p(n^{-2}b^{-7}) = o_p(1).
$$

This together with (3.5.6) completes the proof of (3.5.1).

From (3.5.1) and (3.5.5) we obtain

$$\begin{aligned}\hat{T}_n - T_n &= \int (f_n - \hat{f}_n)^2(x)\,\mathrm{d}x + 2\int (f_n - \hat{f}_n)(f_n - K_b * f_0)(x)\,\mathrm{d}x \\ &= o_p(n^{-1}b^{-2\kappa-1/2}),\end{aligned}$$

by (3.2.1) and (3.2.2). Hence, in view of (3.2.2),

$$n/C_{V,b}^{1/2}\Big(\hat{T}_n - C_{M,b}/\big((2\pi)n\big)\Big) \to_d \mathcal{N}\big(0, 1/2\pi^2\big). \tag{3.5.10}$$

To complete the proof of (3.2.4), it suffices to show that

$$(a)\ \Big|1 - \frac{\hat{C}_{V,b}^{1/2}}{C_{V,b}^{1/2}}\Big| = o_p(b^{1/2}), \quad (b)\ \Big|\frac{\hat{C}_{M,b}}{\hat{C}_{V,b}^{1/2}} - \frac{C_{M,b}}{C_{V,b}^{1/2}}\Big| = o_p(1). \tag{3.5.11}$$

To show (3.5.11)(a), recall $\psi(\beta, s, t) := \Phi_g(\beta t + \beta s)\Phi_f(t+s)$. Then

$$\begin{aligned}&|C_{V,b} - \hat{C}_{V,b}| \\ =\ & \Big|\int\int \frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2|\Phi_g(\beta s)|^2}|\psi(\beta,s,t)|^2\,\mathrm{d}s\,\mathrm{d}t \\ & - \int\int \frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2}{|\Phi_g(\hat{\beta} t)|^2|\Phi_g(\hat{\beta} s)|^2}|\psi(\hat{\beta},s,t)|^2\,\mathrm{d}s\,\mathrm{d}t\Big| \\ \le\ & \int\int \frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2\big||\Phi_g(\beta t)|^2 - |\Phi_g(\hat{\beta} t)|^2\big|}{|\Phi_g(\beta t)|^2|\Phi_g(\beta s)|^2|\Phi_g(\hat{\beta} t)|^2}|\psi(\beta,s,t)|^2\,\mathrm{d}s\,\mathrm{d}t \\ & + \int\int \frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2\big||\Phi_g(\beta s)|^2 - |\Phi_g(\hat{\beta} s)|^2\big|}{|\Phi_g(\hat{\beta} t)|^2|\Phi_g(\beta s)|^2|\Phi_g(\hat{\beta} s)|^2}|\psi(\beta,s,t)|^2\,\mathrm{d}s\,\mathrm{d}t \\ & + \int\int \frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2}{|\Phi_g(\hat{\beta} t)|^2|\Phi_g(\hat{\beta} s)|^2}\big||\psi(\beta,s,t)|^2 - |\psi(\hat{\beta},s,t)|^2\big|\,\mathrm{d}s\,\mathrm{d}t\end{aligned}$$

In view of (3.5.4), the first term in the above bound is bounded from the above by

$$\max_{|t|\le 1/b}\Big|1-\frac{|\Phi_g(\beta t)|^2}{|\Phi_g(\hat\beta t)|^2}\Big|\int\int\frac{|\Phi_K(tb)|^2|\Phi_K(sb)|^2}{|\Phi_g(\beta t)|^2|\Phi_g(\beta s)|^2}|\psi(\beta,s,t)|^2\,\mathrm{d}s\,\mathrm{d}t$$
$$= O_p(n^{-1/2}b^{-1-\kappa}C_{V,b}).$$

The other two terms in the above bounds are bounded similarly. Together with (3.2.1) and $nb^{2\kappa+1}\to\infty$, we obtain

$$|1-\hat C_{V,b}/C_{V,b}| = O_p(n^{-1/2}b^{-1-\kappa}) = o_p(b^{-1/2}),$$

which implies (3.5.11)(a).

Next, consider (3.5.11)(b). Applying (3.2.1), (3.5.11)(a) and $nb^{2\kappa+1}\to\infty$,

$$\begin{aligned}
&|\hat C_{M,b}/\hat C_{V,b}^{1/2} - C_{M,b}/C_{V,b}^{1/2}|\\
&\le\ |\hat C_{M,b}-C_{M,b}||\hat C_{V,b}^{-1/2}| + C_{M,b}C_{V,b}^{-1/2}\Big|1-\hat C_{V,b}^{1/2}/C_{V,b}^{1/2}\Big|\\
&\le\ \max_{|t|\le 1/b}\Big|1-\frac{|\Phi_g(\beta t)|^2}{|\Phi_g(\hat\beta t)|^2}\Big|C_{M,b}\hat C_{V,b}^{1/2} + o_p(1)\\
&=\ O_p(n^{-1/2}b^{-3/2-\kappa}) = o_p(1).
\end{aligned}$$

This completes the proof of (3.5.11), which combined with (3.5.10) also prove (3.2.4), thereby completing the proof of Theorem 3.2.1.

Proof of Theorem 3.2.2. Let $\zeta_\beta(b) := \exp(2\nu(\beta)/b^\lambda)$, $\beta\in\mathbb{R}$. We shall first show that

$$\frac{n}{b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)}\int(\hat f_n - f_n)^2(x)dx = o_p(1). \qquad (3.5.12)$$

The proof is similar as in the ordinary smooth case. We only list some main differences.

First, arguing as for (3.5.4), for the super smooth case, (A′) implies

$$\begin{aligned}\max_{|t|\leq 1/b}\Big|\frac{\Phi_g(-\hat{\beta}t)}{\Phi_g(-\beta t)}-1\Big| &= \max_{|t|\leq 1/b}\Big|\frac{\Phi_g(-\hat{\beta}t)-\Phi_g(-\beta t)}{\Phi_g(-\beta t)}\Big| \\ &= O_p(n^{-1/2}b^{-1+2\lambda_0}\zeta_\beta^{1/2}(b)).\end{aligned} \tag{3.5.13}$$

By (3.2.6), hence $P(A_n)\to 1$, with $A_n := \{|\Phi_g(-\hat{\beta}t)| \geq |\Phi_g(-\beta t)|/2,\ t \in [-1/b, 1/b]\}$.

Assumptions (B′) and (D′) imply that there exist constants M, c_β, $C_\beta<\infty$, such that for $|t|>M$, $c_\beta|t|^{\lambda_0}e^{-\nu(\beta)|t|^\lambda} \leq |\Phi_g(t)| \leq C_\beta|t|^{\lambda_0}e^{-\nu(\beta)|t|^\lambda}$ and $|\Phi_f(t)| \leq C_{g1}|t|^{-\lambda_1}$. Also, on the event A_n, there exists some $\tilde{\beta}$ between $\hat{\beta}$ and β, such that S_2 is bounded from the above by

$$\frac{2\mu_g\|\hat{\beta}-\beta\|^2}{b^2}\int_{|t|\geq M}\frac{|\Phi_K(bt)|^2\big(|\Psi_n(t)-\Phi_h(t)|^2+|\Phi_h(t)|^2\big)}{|\Phi_g(-\beta t)|^4}\,\mathrm{d}t$$
$$+O_p(n^{-1}).$$

Based on (3.2.5),

$$\int_{|t|\geq M}\frac{|\Phi_K(bt)|^2|\Psi_n(t)-\Phi_g(t)|^2}{|\Phi_g(-\beta t)|^4}\,\mathrm{d}t = O_p(n^{-1}b^{\lambda-1+2\lambda\omega+4\lambda_0}\zeta_\beta^2(b)).$$

From Lemma 5 in van Es and Uh (2005), it follows that

$$\int_{|t|\geq M}\frac{|\Phi_K(bt)\Phi_h(-\beta t)|^2}{|\Phi_g(-\beta t)|^4}\,\mathrm{d}t = O_p(b^{2\lambda_0+2\lambda_1+\lambda(1+2\omega)}\zeta_\beta(b)).$$

So when $\lambda_1 > 1$ and n, b satisfy (3.2.6), we have

$$S_2 = o_p(n^{-1} b^{\lambda-1+2\lambda\omega+2\lambda_0} \zeta_\beta(b)).$$

Now we consider S_1. Follow the same arguments as (3.5.7), using assumptions (A′) and (B′) to obtain

$$S_1 \le 8(\hat\beta - \beta)^2/b^2 S_{11} + 8(\hat\alpha - \alpha)^2/b^2 S_{12} + O_p\Big(\frac{b^{2\lambda_0-1}\zeta_\beta(b)}{n^2 b^4}\Big). \tag{3.5.14}$$

Consider S_{11} first. Similar as (3.5.8), together with assumptions (A′)-(D′), we obtain

$$S_{11} = O_p(n^{-1} b^{-1+2\lambda_0}\zeta_\beta(b)) + O_p(1) + O_p(b^{2\lambda_0-1+2\lambda_1+\lambda(1+2\omega)}\zeta_\beta(b)). \tag{3.5.15}$$

S_{12} can be considered the same way. Thus the above arguments(3.2.6), (3.5.14) and (3.5.15) imply

$$\begin{aligned}\frac{nS_1}{b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)} &= O_p(n^{-1}b^{-\lambda-2\lambda\omega}) + O_p(n^{-1}b^{-\lambda-2\lambda\omega-1-2\lambda_0}) \\ &\quad + O(b^{2\lambda_1-2}) + O_p(n^{-1}b^{-\lambda-4-2\lambda\omega}) = o_p(1).\end{aligned}$$

This completes the proof of (3.5.12). Combining this with (3.2.5), we obtain

$$\begin{aligned} &\hat T_n - T_n \qquad (3.5.16)\\ = &\int (f_n - \hat f_n)^2(x)\,\mathrm{d}x + 2\int (f_n - \hat f_n)(f_n - K_b * f_0)(x)\,\mathrm{d}x \\ = &\; o_p(n^{-1}b^{\lambda-1+2\lambda\omega+2\lambda_0}\zeta_\beta(b)).\end{aligned}$$

Also, since $\|\hat{\beta} - \beta\| = O_p(n^{-1/2})$, the first derivatives of $\nu(\beta)$ and $C(\beta)$ exist,

$$|1 - \exp\big(- 2(\nu(\hat{\beta}) - \nu(\beta))/b^{\lambda}\big)| = o_p(1). \tag{3.5.17}$$

Then (3.5.16) and (3.5.17) yield to (3.2.7), thus we complete the proof of Theorem 3.2.2.

Proof of Theorem 3.3.1. Define

$$\tilde{T}_n = \int \big(\hat{f}_n(x) - K_b * f_1(x)\big)^2 \,\mathrm{d}x,$$

Argue as in the proof of Theorem 3.2.1 to obtain

$$n\tilde{C}_{V,b}^{-1/2}\Big(\tilde{T}_n - \hat{C}_{M,b}/\big(2\pi n\big)\Big) \to_d \mathcal{N}\big(0, 1/2\pi^2\big), \tag{3.5.18}$$

where $\tilde{C}_{V,b}$ is same as $\hat{C}_{V,b}$ with f replaced by f_1. Hence, $\tilde{C}_{V,b} \approx b^{-(4\kappa+1)}$.

Next, consider

$$\begin{aligned}
& nb^{2\kappa+1/2}(\hat{T}_n - \tilde{T}_n) \\
= \; & nb^{2\kappa+1/2} \int \big(K_b * f_0(x) - K_b * f_1(x)\big)^2 \,\mathrm{d}x \\
& +2nb^{2\kappa+1/2} \int \big(\hat{f}_n(x) - K_b * f_1(x)\big)\big(K_b * f_1(x) - K_b * f_0(x)\big) \,\mathrm{d}x.
\end{aligned}$$

Because $\int \big(K_b * f_0(x) - K_b * f_1(x)\big)^2 \mathrm{d}x \to \|f_1 - f_0\|^2 > 0$ and $nb^{2\kappa+3} \to \infty$, the first term in the right hand side above is of the order $O(nb^{2\kappa+1/2}) \to \infty$, while by (3.5.18) and the Cauchy-Schwarz inequality, the second term is of the order $o_p(nb^{2\kappa+1/2})$. This completes the proof of Theorem 3.3.1.

The proofs of Theorems 3.3.2, 3.3.3 and 3.3.4 are similar to those of Theorems 3.3.1 and

3.2.1, and hence no details are given.

Proof of Theorem 3.3.5. For the sake of completeness of this chapter, we first provide a brief proof of (3.3.3). For $j = 1, \cdots, n$, let

$$D_j = \frac{1}{\pi} \int |\Phi_K(tb)|^2 \Big(\frac{e^{it(\varepsilon_j - \beta' u_j)}}{\Phi_g(-\beta t)} - \Phi_{f_1}(t) \Big) \overline{\big(\Phi_{f_1}(t) - \Phi_{f_0}(t)\big)} \, \mathrm{d}t.$$

Note that since K is symmetric, D_j is real. Rewrite

$$\begin{aligned} & T_n - \|K_b * (f_1 - f_0)\|^2 \\ & \quad = \int (f_n - K_b * f_1)^2 \, \mathrm{d}x + 2 \int (f_n - K_b * f_1)(K_b * (f_1 - f_0)) \, \mathrm{d}x. \end{aligned}$$

Recall (3.2.2) and that $nb^{4\kappa+2} \to \infty$. Hence, the first term on the right hand side above is $O_p(n^{-1}b^{2\kappa+1}) = o_p(n^{-1/2})$. Using Parseval's equation, the second term can be written as $n^{-2} \sum_{j=1}^{n} D_j$. Note that D_j's are independent arrays identically distributed r.v.'s, with $ED_1 = 0$, and $\mathrm{Var}(D_1)$ converging to

$$\begin{aligned} \tau_0^2 : \ &= \ \frac{1}{\pi^2} \int \int \Phi_h(t-s) \frac{(\Phi_{f_1}(s) - \Phi_{f_0}(s))\overline{(\Phi_{f_1}(t) - \Phi_{f_0}(t))}}{\Phi_g(\beta s)\Phi_g(-\beta t)} \, \mathrm{d}s \, \mathrm{d}t \\ & \quad - \frac{1}{\pi^2} \Big(\int \Phi_{f_1}(-t)(\Phi_{f_1}(t) - \Phi_{f_0}(t)) \, \mathrm{d}t \Big)^2 \\ &= \ \frac{1}{2\pi^3} \mathrm{Var} \Big(\int e^{-it\varepsilon} \frac{\Phi_{f_1}(t) - \Phi_{f_0}(t)}{\Phi_g(\beta t)} \, \mathrm{d}t \Big). \end{aligned}$$

Moreover,

$$\mathrm{E}|D_1|^4 \leq \frac{1}{\pi^4} \Big(\int \big(\frac{1}{|\Phi_g(-\beta t)|} + |\Phi_{f_1}(t)| \big) \big(|\Phi_{f_1}(t)| + |\Phi_{f_0}(t)| \big) \, \mathrm{d}t \Big)^4 = O(1),$$

by the assumption (B) with $r > \kappa + 1$. Hence one obtains (3.3.3), by the Lindeberg-Feller CLT.

To complete the proof of Theorem 3.3.5, first, consider the case where α is known, so that $\hat{f}_n$ is based on the residuals $Y_i - \alpha - \hat{\beta}' Z_i$'s only. Without loss of generality, assume $\alpha = 0$. Under the alternative H_1,

$$\begin{aligned} & n^{1/2}(\hat{T}_n - T_n) \\ = \; & n^{1/2} \int (\hat{f}_n - f_n)^2(x)\,\mathrm{d}x + 2n^{1/2} \int (\hat{f}_n - f_n)(f_n - K_b * f_1)(x)\,\mathrm{d}x \\ & +2n^{1/2} \int (\hat{f}_n - f_n)(K_b * f_1 - K_b * f_0)(x)\,\mathrm{d}x. \end{aligned}$$

The same proof as that of (3.5.1) and $nb^{4\kappa+2} \to \infty$ imply

$$n^{1/2} \int (\hat{f}_n - f_n)^2(x)\,\mathrm{d}x = o_p(n^{-1/2} b^{-2\kappa}) = o_p(1). \tag{3.5.19}$$

This fact together with (3.2.4) and the Cauchy-Schwarz inequality implies

$$2n^{1/2} \int (\hat{f}_n - f_n)(f_n - K_b * f_1)(x)\,\mathrm{d}x = o_p(n^{-1} b^{-3\kappa-1}) = o_p(1). \tag{3.5.20}$$

To deal with the the remaining part, let $\Delta_f(x) := (K_b * f_1 - K_b * f_0)(x)$. Rewrite $\hat{f}_n - f_n$ as the sum of the following two terms:

$$\mathcal{D}_1 := \int\int e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n (e^{it(\varepsilon_j - \hat{\beta}' u_j)} - e^{it(\varepsilon_j - \beta' u_j)})}{2\pi n \Phi_g(-\hat{\beta} t)} \Delta_f(x)\,\mathrm{d}t\,\mathrm{d}x,$$

$$\mathcal{D}_2 := \int\int e^{-itx} \Phi_K(bt) \frac{\sum_{j=1}^n e^{it(\varepsilon_j - \beta' u_j)}}{2\pi n} \Big(\frac{1}{\Phi_g(-\hat{\beta} t)} - \frac{1}{\Phi_g(-\beta t)} \Big) \Delta_f(x)\mathrm{d}t\mathrm{d}x.$$

Consider $\mathcal{D}_1$ first. Since $nb^{4\kappa+2} \to \infty$, and $\kappa > 1$, then uniformly in $|t| \le 1/b$,

$$\frac{1}{n}\sum_{j=1}^{n}\Big(e^{it(\varepsilon_j-\hat{\beta}'u_j)} - e^{it(\varepsilon_j-\beta'u_j)}\Big) = \frac{\sum_{j=1}^{n} it(\beta-\hat{\beta})'Z_j e^{it(\varepsilon_j-\hat{\beta}'u_j)}}{n} + o_p(n^{-1/2}).$$

Let

$$\mathcal{C}_0 := \int\int te^{-itx}\Phi_K(bt)\frac{\sum_{j=1}^{n}[Z_{jk}e^{it(\varepsilon_j-\beta'u_j)} - \mathrm{E}Z_{jk}e^{it(\varepsilon_j-\beta'u_j)}]}{2\pi n\Phi_g(-\beta t)}\,\mathrm{d}t\Delta_f(x)\,\mathrm{d}x.$$

Then $\mathrm{E}\mathcal{C}_0 = 0$ and

$$\begin{aligned}
&\mathrm{E}\mathcal{C}_0^2 \qquad (3.5.21)\\
&\le \mathrm{E}\Big(\sum_{j=1}^{n}\int\int te^{-itx}\Phi_K(bt)\frac{Z_{jk}e^{it(\varepsilon_j-\beta'u_j)} - \mathrm{E}Z_{jk}e^{it(\varepsilon_j-\beta'u_j)}}{2\pi n\Phi_g(-\beta t)}\,\mathrm{d}t\Delta_f(x)\,\mathrm{d}x\Big)\\
&\le \frac{\mathrm{E}|Z_k|^2}{n}\Big(\int\int te^{-itx}\frac{\Phi_K(bt)}{2\pi\Phi_g(-\beta t)}\,\mathrm{d}t\Delta_f(x)\,\mathrm{d}x\Big)^2\\
&\le \frac{\mathrm{E}|Z_k|^2}{n}\Big(\int\int te^{-itx}\Phi_K(bt)\frac{\Phi_K(bt)}{2\pi\Phi_g(-\beta t)}\,\mathrm{d}t\Delta_f(x)\,\mathrm{d}x\Big)^2\\
&\le \frac{\mathrm{E}|Z_k|^2}{2\pi n}\Big(\int\frac{|\Phi_K(bt)|}{|\Phi_g(-\beta t)|}\,\mathrm{d}t\int|\Delta_{\dot{f}}(x)|\,\mathrm{d}x\Big)^2 = O(n^{-1}b^{-2\kappa-2}) = o(1).
\end{aligned}$$

Hence, $\mathcal{C}_0 = o_p(1)$. Since

$$\mathrm{E}Z_k e^{it(\varepsilon-\beta'u)} = \mu_Z\Phi_h(t) + \Phi_{f_1}(t)\mathrm{E}u_k e^{-i\beta' ut},$$

assumption (B) with $r > \kappa+1$ and the relation $\Phi_h(t) = \Phi_g(t)\Phi_{f_1}(t)$ imply

$$\int\int te^{-itx}\Phi_K(bt)\frac{\mathrm{E}Z_k e^{it(\varepsilon-\beta'u)}}{\Phi_g(-\beta t)}\mathrm{d}t\Delta_f(x)\mathrm{d}x = O(1).$$

Together with (3.5.4), and $nb^{4\kappa+2} \to \infty$, the above analysis yields

$$\begin{aligned} &\mathcal{D}_1 + \frac{i(\hat{\beta}-\beta)'}{2\pi n} \int\int te^{-itx}\Phi_K(bt)\frac{\mathrm{E}Ze^{it(\varepsilon-\beta' u)}}{\Phi_g(-\beta t)}\mathrm{d}t\Delta_f(x)\mathrm{d}x \\ &\quad = O_p(n^{-1}b^{-\kappa-1}) = o_p(n^{-1/2}). \end{aligned} \tag{3.5.22}$$

Next consider $\mathcal{D}_2$. Uniformly in $|t| \le 1/b$,

$$\begin{aligned} &\frac{\Phi_g(-\hat{\beta}t) - \Phi_g(-\beta t)}{\Phi_g^2(-\beta t)} \\ &= \sum_{k=1}^{p}\Big\{\frac{it(\beta_k - \hat{\beta}_k)\mathrm{E}u_k e^{-i\beta' ut}}{\Phi_g^2(-\beta t)} - \frac{(\beta_k - \hat{\beta}_k)^2 t^2 \mathrm{E}u^2 e^{-i\beta' ut}}{\Phi_g^2(-\beta t)}\Big\} + O_p(n^{-3/2}b^{-3-2\kappa}). \end{aligned}$$

Let

$$\mathcal{C}_1 := \int\int te^{-itx}\Phi_K(bt)\frac{\sum_{j=1}^n (e^{it(\varepsilon-\beta' u)} - \Phi_h(t))\mathrm{E}u_k e^{-i\beta' ut}}{\Phi_g^2(-\beta t)}\Delta_f(x)\mathrm{d}t\mathrm{d}x,$$

$$\mathcal{C}_2 := \int\int t^2 e^{-itx}\Phi_K(bt)\frac{\sum_{j=1}^n (e^{it(\varepsilon-\beta' u)} - \Phi_h(t))\mathrm{E}u_k^2 e^{-i\beta' ut}}{\Phi_g^2(-\beta t)}\Delta_f(x)\mathrm{d}t\mathrm{d}x.$$

Note that $\mathrm{E}\mathcal{C}_i = 0$, $i = 1, 2$, and same arguments as (3.5.21) yield

$$\mathrm{E}\mathcal{C}_1^2 = O(n^{-1}b^{-4\kappa-2}) = o(1), \quad \mathrm{E}\mathcal{C}_2^2 = O(n^{-1}b^{-4\kappa-4}) = o(b^{-2}).$$

Hence, $\mathcal{C}_1 = o_p(1)$ and $\mathcal{C}_2 = o_p(b^{-1})$. Since $\Phi_h(t) = \Phi_{f_1}(t)\Phi_g(-\beta t)$, $nb^{4\kappa+2} \to \infty$, by (3.5.4) and assumption (B) with $r > \kappa + 1$, we obtain

$$\begin{aligned} &\mathcal{D}_2 - \frac{i(\hat{\beta}-\beta)'}{2\pi n} \int\int te^{-itx}\Phi_K(bt)\frac{\Phi_{f_1}(t)\mathrm{E}ue^{-i\beta' ut}}{\Phi_g(-\beta t)}\,\mathrm{d}t\Delta_f(x)\,\mathrm{d}x \\ &\quad = o_p(n^{-1/2}). \end{aligned} \tag{3.5.23}$$

Also,

$$-\frac{1}{2\pi}\int ite^{-itx}\Phi_K(bt)\Phi_{f_1}(t)dt = K_b * \dot{f}_1(x).$$

Combine this with (3.5.22) and (3.5.23) to obtain

$$2n^{1/2}\int(\hat{f}_n - f_n)(f_n - K_b * f_0)(x)\,\mathrm{d}x = (\hat{\beta}-\beta)'B_f + o_p(n^{-1/2}).$$

Recall (3.5.19) and (3.5.20), immediately

$$n^{1/2}(\hat{T}_n - T_n - (\hat{\beta}-\beta)'B_f) = o_p(1). \tag{3.5.24}$$

Next, consider the case when the intercept parameter α is unknown. Let $a = \alpha - \hat{\alpha}$. Then

$$\begin{aligned}\hat{T}_n &= \int \big(f_n(x,\alpha,\hat{\beta}) - K_b * f_0(x+a)\big)^2\,\mathrm{d}x \\ &= \int \big(f_n(x,\alpha,\hat{\beta}) - K_b * f_0(x)\big)^2\,\mathrm{d}x \\ &\quad + \int \big(K_b * f_0(x+a) - K_b * f_0(x)\big)^2\,\mathrm{d}x \\ &\quad -2\int \big(f_n(x,\alpha,\hat{\beta}) - K_b * f_0(x)\big)\big(K_b * f_0(x+a) - K_b * f_0(x)\big)\,\mathrm{d}x.\end{aligned}$$

The first term on the right side above is $T_n(\alpha,\hat{\beta})$, and from (3.5.24) we have

$$n^{1/2}(T_n(\alpha,\hat{\beta}) - T_n - (\hat{\beta}-\beta)'B_f) = o_p(1).$$

Because $\dot{f}_0$ exists, and is finite, and $a = O_p(n^{-1/2})$, the second term is $O_p(n^{-1})$. Then to deal with the third term, rewrite the factor multiplying -2 as the sum of the following three terms:

$$\int \big(f_n(x,\alpha,\hat{\beta}) - f_n(x)\big)\big(K_b * f_0(x+a) - K_b * f_0(x)\big)\,\mathrm{d}x,$$
$$\int \big(f_n(x) - K_b * f_1(x)\big)\big(K_b * f_0(x+a) - K_b * f_0(x)\big)\,\mathrm{d}x,$$
$$\int \big(K_b * f_1(x) - K_b * f_0(x)\big)\big(K_b * f_0(x+a) - K_b * f_0(x)\big)\,\mathrm{d}x.$$

By using the Cauchy-Schwarz inequality, together with $a = O_p(n^{-1/2})$, (3.5.18) and (3.5.24), verify that each of the first two terms above are $o_p(n^{-1/2})$. The finiteness of $\ddot{f}_0$ implies that the third term is equal to

$$a\int \big(K_b * f_1(x) - K_b * f_0(x)\big)\big(K_b * \dot{f}_0(x)\big)\,\mathrm{d}x + o_p(n^{-1/2}).$$

The above analysis and (3.5.24) imply

$$n^{1/2}\big(\hat{T}_n - T_n - (\hat{\beta} - \beta)' B_f - (\hat{\alpha} - \alpha) A_f\big) = o_p(1). \tag{3.5.25}$$

This fact together with (3.3.3) completes the proof of Theorem 3.3.5.

Proof of Theorem 3.3.6. For $\hat{T}_n$, recall (3.3.5), (3.3.6) and (3.5.25). Using the details in the proof of Theorem 3.3.5, we obtain,

$$\hat{T}_n - \|K_b * (f_1 - f_0)\|^2 = \frac{1}{n}\sum_{j=1}^{n}(D_j + \eta_j A_f + \zeta_j' B_f) + o_p(n^{1/2}).$$

We write

$$\tau^2 := \mathrm{Var}(D_1 + \eta_1 A_f + \zeta_1' B_f). \tag{3.5.26}$$

Since $D_j + \eta_j A_f + \zeta_j' B_f$, for $j = 1, \cdots, n$ are arrays of i.i.d. zero mean r.v.'s and $\mathrm{E}|D_1|^4 = O(1)$, $\mathrm{E}|\eta|^{2+\vartheta} < \infty$ and $\mathrm{E}\|\zeta\|^{2+\vartheta} < \infty$ for some $\vartheta > 0$. Thus the claim (3.3.7) follows by the Lindeberg-Feller CLT, thereby completing the proof.

The proof of Theorem 3.3.7 is similar as the arguments in the proof of Theorem 3.3.5 and 3.3.6. Thus we omit the details of the proof.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Bachmann, D. and Dette, H. (2005). A note on the Bickel–Rosenblatt test in autoregressive time series. *Statistics and Probability Letters*, **74(3)**, 221–234.

[2] Bercu, B., and Portier, B. (2008). Kernel density estimation and goodness-of-fit test in adaptive tracking. *SIAM Journal on Control and Optimization*, **47(5)**, 2440–2457.

[3] Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, **1**, 1071–1095.

[4] Boldin, M.V. (1982). An estimate of the distribution of the noise in an autoregressive scheme. *Theory of Probability and Its Applications*, **27(4)**, 866–871.

[5] Boldin, M.V. (1990). On testing hypotheses in the sliding average scheme by the kolmogorovsmirnov and ω^2 tests. *Theory of Probability and Its Applications*, **34(4)**, 699–704.

[6] Borkowski, P. and Mielniczuk, J. (2012). Performance of variance function estimators for autoregressive time series of order one: asymptotic normality and numerical study. *Control and Cybernetics*, **41**, 415-441.

[7] Butucea, C. (2004). Asymptotic normality of the integrated square error of a density estimator in the convolution model. *Sort*, **28(1)**, 9–26.

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

[9] Cheng, F., and Sun, S. (2008). A goodness-of-fit test of the errors in nonlinear autoregressive time series models. *Statistics & Probability Letters*, **78(1)**, 50–59.

[10] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, **83(404)**, 1184–1186.

[11] Carroll, R. J., Ruppert, D., Stefanski, L.A., and Crainiceanu, C.M. (2012). *Measurement error in nonlinear models: a modern perspective.* CRC Press.

[12] Delaigle, A., and Hall, P. (2006). On optimal kernel choice for deconvolution. *Statistics and Probability Letters*, **76(15)**, 1594–1602.

[13] Ducharme, G. R., and Lafaye de Micheaux, P. (2004). Goodness-of-fit tests of normality for the innovations in ARMA models. *Journal of Time Series Analysis*, **25(3)**, 373–395.

[14] Durbin, J. (1973). *Distribution Theory for Tests Based on the Sample Distribution Function.* CBMS Regional Conference Series in Applied Mathematics **9**. SIAM, Philadelphia.

[15] Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, **53(1)**, 97–110.

[16] Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, **85(3)**, 645–660.

[17] Fan, J., and Yao, Q. (2002). *Nonlinear time series: Nonparametric and Parametric Methods.* Springer Series in Statistics, Springer-Verlag New York, Inc.

[18] Freedman, D.A. (1975). On tail probabilities for martingales. *The Annals of Probability* **3(1)**, 100–118.

[19] Fuller, W. A. (2009). *Measurement error models*. John Wiley & Sons.

[20] Horváth, L., and Zitikis, R. (2006). Testing goodness of fit based on densities of GARCH innovations. *Econometric theory*, **22(03)**, 457–482.

[21] Jiang, J. (2001). Goodness-of-fit tests for mixed model diagnostics. *Annals of statistics*, 1137–1164.

[22] Khmaladze, E. V. (1982). Martingale Approach in the Theory of Goodness-of-fit Tests. *Theory of Probability & Its Applications*, **26(2)**, 240–257.

[23] Khmaladze, E. V. (1993). Goodness of fit Problems and Scanning Innovation Martingales. *Annals of Statistics*, **21 (2)**, 798-829.

[24] Khmaladze, E. V. and Koul, H. L. (2004). Martingale transforms goodness-of-fit tests in regression models. *The Annals of Statistics*, **32**, 995–1034.

[25] Khmaladze, E. V. and Koul, H. L. (2009). Goodness-of-fit problem for errors in nonparametric regression. *The Annals of Statistics*, **37**, 3165–3185.

[26] Koul, H.L. Asymptotic behavior of Wilcoxon type confidence regions in multiple linear regression. *Annals of Mathematical Statistics*, **40 (1969)**,1950-1979

[27] H.L. Koul Some convergence theorems for ranks and weighted empirical cumulatives *Annals of Mathematical Statistics*, **41 (1970)**, 1768-1773

[28] Koul, H. L. (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference*, **29(3)**, 291–308.

[29] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models. Second Edition.* **Lecture Notes in Statistics**, **166**. Springer-Verlag New York, Inc.

[30] Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics*, **34(2)**, 994–1012.

[31] Koul, H. L. and Mimoto, N. (2012). A goodness-of-fit test for GARCH innovation density. *Metrika*, **75(1)**, 127–149.

[32] Koul, H. L. and Song, W. (2012). A class of goodness-of-fit tests in linear errors-in-variables model. *Journal de la Société Française de Statistique*, **153(1)**, 52–70.

[33] Lee, S. and Na, S. (2002). On the BickelRosenblatt test for first-order autoregressive models. *Statistics and Probability Letter* **56**, 23–35.

[34] Loynes, R. M. (1980). The empirical d.f. of residuals from generalized regression. *The Annals of Statistics*, **8**, 285–298.

[35] Loubes, J. M. and Marteau, C. (2004). Goodness-of-fit testing strategies from indirect observations. *Journal of Nonparametric Statistics*, **26(1)**, 85–99.

[36] Masry, E. (1996). Multivariate local polynomial regression for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis*, **17(6)**, 571–599.

[37] Müller, U. U., Schick, A., and Wefelmeyer, W. (2007). Estimating the error distribution function in semiparametric regression. *Statistics and Decisions-International Journal Stochastic Methods and Models*, **25(1)**, 1-18.

[38] Müller, U.U., Schick, A. and Wefelmeyer, W. (2009a). Estimating the innovation distribution in nonparametric autoregression. *Probability Theory and Related Fields*, **144(1-2)**, 53–77.

[39] Müller, U. U., Schick, A., and Wefelmeyer, W. (2009b). Estimating the error distribution function in nonparametric regression with multivariate covariates. *Statistics & Probability Letters*, **79(7)**, 957-964.

[40] Müller, U. U., Schick, A., and Wefelmeyer, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference*, **142(2)**, 552–566.

[41] Na, S. (2009). Goodness-of-fit test using residuals in infinite-order autoregressive models. *Journal of the Korean Statistical Society*, **38(3)**, 287–295.

[42] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, **101(5)**, 1067–1078.

[43] Neumeyer, N. and Selk, L. (2013). A note on non-parametric testing for gaussian innovations in ararch models. *Journal of Time Series Analysis*, **34(3)**, 362–367.

[44] Nikabadze, A. M. (1988). On a Method for Constructing Goodness-of-Fit Tests for Parametric Hypotheses in $\mathbb{R}^m$. *Theory of Probability and Its Applications*, **32(3)**, 539–544

[45] Ojeda, J.L. (2008). Hölder continuity properties of the local polynomial estimator. *Prepublicaciones del Seminario Matemático*, **4**, 1–21.

[46] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, **27(3)**, 832–837. Chicago

[47] Selk, L. and Neumeyer, N. (2013). Testing for a Change of the Innovation Distribution in Nonparametric Autoregression: The Sequential Empirical Process Approach, *Scandinavian Journal of Statistics*, **40(4)**, 770–788.

[48] Soms, A. P. (1976). An asymptotic expansion for the tail area of the t -distribution. *Journal of the American Statistical Association*, **71**, 728–730.

[49] Stefanski, L. A. and Carroll, R. J. (1990).Deconvolving kernel density estimators. *Statistics*, **21(2)**, 169–184.

[50] Tsigroshvili, Z. (1998). Some notes on goodness-of-fit tests and innovation martingales. *Proceedings of A. Razmadze Mathematical Institute*, **117**, 89–102.

[51] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence.* Springer, New York.

[52] Van Es, A. and Uh, H. -W. (2004). Asymptotic normality of nonparametric kernel type deconvolution density estimators: crossing the cauchy boundary. *Nonparametric Statistics*, **16(1-2)**, 261–277.

[53] Wu, W.B., Huang, Y. and Huang, Y. (2010). Kernel estimation for time series: An asymptotic theory. *Stochastic Processes and their Applications*, **120(12)**, 2412–2431.

[54] Yao, Q. and Tong, H. (1994). Quantifying the influence of initial values on non-linear prediction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **56(4)**, 701–725.