SPECIFICATION TESTING FOR ERRORS-IN-VARIABLES MODELS

TAISUKE OTSU AND LUKE TAYLOR

Abstract. This paper considers specification testing for regression models with errors-in-variables and proposes a test statistic by comparing the distance between the parametric and nonparametric fits based on deconvolution techniques. In contrast to the existing method by Hall and Ma (2007), our test allows general nonlinear regression models. Since our test employs the smoothing approach, it complements the nonsmoothing one by Hall and Ma in terms of local power properties. We establish the asymptotic properties of the test statistic for the ordinary and super smooth measurement error densities and develop a bootstrap method to approximate the critical value. The simulation results endorse our theoretical findings: our test has advantages in detecting high frequency alternatives.

1. Introduction

As is the case with most decisions, the choice to employ nonparametric techniques over parametric ones is not always obvious, and making the wrong decision can be costly. Parametric estimation relies on specifying an exact form for the model; if this specification is incorrect, estimation will be inconsistent in general. In contrast, nonparametric estimation will be consistent under much weaker assumptions. However, the cost is paid by slower convergence rates, which can be significantly slower in high dimensional problems. It is clear that if we are able to confirm that a parametric model is correctly specified, we can gain considerably by using parametric estimators, and that if we are not fully convinced of this, we should appeal to nonparametric estimation. A popular solution to this problem involves comparing the distance between some parametric and nonparametric estimators; this has been studied in detail by Härdle and Mammen (1993). Other tests for the suitability of parametric models have been studied by Azzalini, Bowman and Härdle (1989), Eubank and Spiegelman (1990), Horowitz and Spokoiny (2001), and Fan and Huang (2001) among many others.

Measurement error is a problem that is rife in datasets from many disciplines. For example, Fuller (1987) investigated the role of nitrogen level in soil on corn yields, where measurements of nitrogen content are known to be imprecisely measured. A common example from the field of medicine involves experiments in which systolic blood pressure is measured using a sphygmomanometer; a patient’s blood pressure can differ significantly over the course of a day and can vary depending on recent physical activity (see, Carroll et al., 1984). Examples from economics, geography, and physics are also abundant (see, e.g., Meister, 2009, for a review). Determining the validity of a parametric model becomes even more important in the presence of measurement error because in this setting nonparametric estimators have even slower convergence properties whilst in many cases parametric estimators retain their $\sqrt{n}$-consistency. However, when the data

The authors gratefully acknowledge financial support from the ERC Consolidator Grant (SNP 615882) (Otsu) and the ESRC (Taylor).
are contaminated by measurement error, conventional specification tests have incorrect sizes in general and may also suffer from low power properties.

Little attention has been given to this issue despite its obvious need. Butucea (2007), Holzmann and Boysen (2006), and Holzmann, Bissantz and Munk (2007) considered a similar problem for testing probability densities. Cheng and Kukush (2004) were the first to tackle this problem for a regression function. They constructed a squared difference test statistic having a chi-squared limiting distribution under the null. Although their test enjoys considerable power, it is only able to detect departures from the null hypothesis in one direction, and only polynomial regression models are considered. Hall and Ma (2007) proposed a nonsmoothing specification test for regression models with errors-in-variables, which is able to detect local alternatives at the $\sqrt{n}$ rate. However, similar to Cheng and Kukush (2004), only polynomial models are allowed. Also, due to the unconventional construction of the test statistic, its asymptotic distribution is complex and non-standard. As a result, Hall and Ma (2007) suggested a bootstrap method to determine the critical value.

Consistent specification tests can be broadly split into those that use a nonparametric estimator (called smoothing tests) and those that do not (called nonsmoothing or integral-transform tests). In contrast to Hall and Ma (2007) which adopted the nonsmoothing approach, we propose a kernel-based smoothing test for goodness-of-fit of parametric regression models with errors-in-variables. There are two important features of our test. First, our smoothing test is not restricted to polynomial models and allows testing of general nonlinear regression models. Second, analogous to the literature of conventional specification testing, our smoothing test complements Hall and Ma’s (2007) test (if applied to polynomial models) due to its distinct power properties. Rosenblatt (1975) explained that although local power properties of nonsmoothing tests suggest they are more powerful than smoothing tests, ‘there are other types of local alternatives for which tests based on density estimates are more powerful’. Fan and Li (2000) showed that in the non-measurement error case, smoothing tests are generally more powerful for high frequency alternatives and nonsmoothing tests are more powerful for low frequency alternatives. ‘Thus, smoothing tests ‘should be viewed as complements to, rather than substitutes for, [nonsmoothing tests].’ Our simulation results suggest that this phenomenon extends to errors-in-variables models.

In contrast to the above papers and our own, Ma et al. (2011) moved away from Wald-type tests where restricted and unrestricted estimates are compared. They proposed a local test that is more analogous to the score test where only the model under the null hypothesis must be estimated. They extended this idea to an omnibus test that is able to detect departures from the null in virtually all directions using a system of different basis functions with which to test against.

To determine critical values for our smoothing test, we propose a bootstrap procedure. Measurement error can cause difficulties for applying conventional bootstrap procedures because the true regressor, regression error, and measurement error are all unobserved. Moreover, in order to estimate the distributions of test statistics, deconvolution techniques are typically required which converge at a much slower rate than $\sqrt{n}$. Hall and Ma (2007) discussed this issue and note, ‘the
bootstrap is seldom used in the context of errors-in-variables’. They outline a procedure which involves estimating the distribution of the unobservable regressor using a kernel deconvolution estimator, and obtaining bootstrap counterparts for the regression error using a wild bootstrap method. We propose a much simpler procedure involving a perturbation of each summand of our test statistic.

This paper is organized as follows. Section 2 describes the setup in detail and introduces the test statistic and its motivation. Section 3 outlines the main asymptotic properties of the test statistic and also discusses how to implement the test in the case where the distribution of the measurement error is unknown but repeated measurements on the contaminated covariates are available. Section 4 analyses the small sample properties of the test through a Monte Carlo experiment. All mathematical proofs are deferred to the Appendix.

2. Setup and test statistic

Consider the nonparametric regression model

\[ Y = m(X) + U \quad \text{with} \quad E[U|X] = 0, \]

where \( Y \in \mathbb{R} \) is a response variable, \( X \in \mathbb{R}^d \) is a vector of covariates, and \( U \in \mathbb{R} \) is the error term. In this paper, we focus on the situation where \( X \) is not directly observable due to the measurement mechanism or nature of the environment. Instead a vector of variables \( W \) is observed through

\[ W = X + \epsilon, \]

where \( \epsilon \in \mathbb{R}^d \) is a vector of measurement errors that has a known density \( f_\epsilon \) and is independent of \( (Y, X) \). The case of unknown density \( f_\epsilon \) will be discussed in Section 3.1. We are interested in specification or goodness-of-fit testing of a parametric functional form of the regression function \( m \). More precisely, for a parametric model \( m_\theta \), we wish to test the hypothesis

\[ H_0 : \quad m(x) = m_\theta(x) \quad \text{for almost every} \quad x \in \mathbb{R}^d, \]

\[ H_1 : \quad H_0 \text{ is false}, \]

based on the random sample \( \{Y_i, W_i\}_{i=1}^n \) of observables (but \( X_i \) is unobservable).

To test the null \( H_0 \), we adapt the approach of Härdle and Mammen (1993), which compares nonparametric and parametric regression fits, to the errors-in-variables model. As a nonparametric estimator of \( m \), we use the deconvolution kernel estimator (see, e.g., Fan and Truong, 1993, and Meister, 2009, for a review)

\[ \hat{m}(x) = \frac{\sum_{i=1}^n Y_i K_b(x - W_i)}{\sum_{i=1}^n K_b(x - W_i)}, \]

where

\[ K_b(a) = \frac{1}{(2\pi)^d} \int e^{-it \cdot a} \frac{K^b(tb)}{f^b(t)} dt, \]
is the so-called deconvolution kernel, \( i = \sqrt{-1} \), \( b \) is a bandwidth, and \( K^\text{ft} \) and \( f^\text{ft}_t \) are the Fourier transforms of a kernel function \( K \) and the measurement error density \( f_e \), respectively.\(^1\)

Throughout the paper we assume \( f^\text{ft}_t(t) \neq 0 \) for all \( t \in \mathbb{R}^d \) and \( K^\text{ft} \) has compact support so that the above integral is well-defined. On the other hand, if one imposes a parametric functional form \( m_\theta \) on the regression function, several methods are available to estimate \( \theta \) under certain regularity conditions. For example, based on Taupin (2001), we can estimate the parameter \( \theta \) by the (weighted) least squares regression of \( Y \) on the implied conditional mean function \( E[m_\theta(X)|W] \). In this paper, we do not specify the construction of the estimator \( \hat{\theta} \) for \( \theta \) except for a mild assumption on the convergence rate (see Section 3 for details).

In order to construct a test statistic for \( H_0 \), as in Härdle and Mammen (1993), we compare the nonparametric and parametric estimators of the regression function based on the \( L_2 \)-distance,

\[
D_n = n \int \left| \hat{m}(x) \hat{f}(x) - [K_b + m_\hat{\theta} \hat{f}](x) \right|^2 dx,
\]

where \( | \cdot | \) is the Euclidean norm, \( \hat{f}(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - W_i) \) is the deconvolution kernel density estimator for \( X \), \( K_b(x) = \frac{1}{nb} K \left( \frac{x}{b} \right) \), and \( [K_b + m_\hat{\theta} \hat{f}](x) = \int K_b(x-a)m_\hat{\theta}(a)\hat{f}(a)da \) is a convolution. The convolution by the (original) kernel function \( K_b \) plays an analogous role to the smoothing operator in Härdle and Mammen (1993). Note that the Fourier transform of a convolution is given by the product of the Fourier transforms. Thus by Parseval’s identity, the distance \( D_n \) is alternatively written as

\[
D_n = \frac{n}{(2\pi)^d} \int \left| K^\text{ft}_t(tb) \right|^2 \left| \frac{1}{f^\text{ft}_t(t)} \right|^2 \left| \frac{1}{n} \sum_{i=1}^n Y_i e^{itW_i} - \left[ m_\hat{\theta} \hat{f} \right]^\text{ft}_t(t) f^\text{ft}_t(t) \right|^2 dt.
\]

Based on this expression, the distance \( D_n \) can be interpreted as a contrast of the nonparametric and model-based estimators for \( E[Y e^{itW}] \). Let \( \zeta_i(t) = Y_i e^{itW_i} - \int e^{isW_i} m_\hat{\theta}(t-s) K^\text{ft}_t(s)ds f^\text{ft}_t(t) \). To define the test statistic for \( H_0 \), we further decompose \( D_n \) as

\[
D_n = \frac{1}{n} \sum_{i=1}^n \frac{1}{(2\pi)^d} \int \left| K^\text{ft}_t(tb) \right|^2 |\zeta_i(t)|^2 dt + \frac{1}{n} \sum_{i\neq j} \frac{1}{(2\pi)^d} \int \left| K^\text{ft}_t(tb) \right|^2 |\zeta_i(t)| |\overline{\zeta}_j(t)| dt
\]

\[
\equiv B_n + T_n,
\]

where \( \overline{\zeta}_j(t) \) is the complex conjugate of \( \zeta_j(t) \). The second term \( T_n \) plays a dominant role in the limiting behavior of \( D_n \) and the first term \( B_n \) is considered a bias term. Therefore, we neglect \( B_n \) and employ \( T_n \) as our test statistic for \( H_0 \). In the next section, we study the asymptotic behaviour of \( T_n \).

We close this section by a remark on an alternative testing approach. To test the null hypothesis \( H_0 \), one may consider testing some implication of \( H_0 \) on the conditional mean \( E[Y|W] \) of observables, i.e., consider \( H'_0 : E[Y|W] = \int m_\theta(W - \epsilon)f_e(\epsilon)d\epsilon \) and test \( H'_0 \) by a conventional method, such as Härdle and Mammen (1993). To clarify the rationale of our testing approach.

---

\(^1\)To simplify the exposition, we concentrate on the case where all elements of \( X \) are mismeasured. If \( X \) contains both correctly measured and mismeasured covariates (denoted by \( X_1 \) and \( X_2 \), respectively), then the kernel estimator is modified as \( \hat{m}(x) = \sum_{i=1}^n Y_i K_{1b}(x-X_i_1)K_1(x-X_2-W_i) \), where \( K_{1b}(a) = \frac{1}{\sqrt{b}} K_1 \left( \frac{a}{b} \right) \) and \( K_1 \) is a conventional kernel function for \( X_1 \), and analogous results can be established.
based on $T_n$ against the conventional approach for $H'_0$, let us consider the following local alternative hypothesis for the regression function

$$m_n(x) = m_\theta(x) + a_n \cos(A_n x) \frac{(\sin x)}{x}^2,$$

where $a_n \to 0$ and $A_n \to \infty$ as $n \to \infty$. In this case, $m_n$ converges to $m_\theta$ at the rate of $a_n$ under the $L^2$-norm, and the test based on $T_n$ will have non-trivial power for a certain rate of $a_n$. On the other hand, local power of the test based on the implied null $H'_0$ is determined by the $L^2$-norm of the convolution $(m_n - m_\theta) * f_\epsilon$. For example, if $f_\epsilon$ is Laplace, then we can see that this $L_2$-norm is of order $a_n/A_n^2$. By letting $A_n$ diverge at an arbitrarily fast rate, the rate $a_n/A_n^2$ becomes arbitrarily fast so that any conventional test for $H'_0$ fails to detect deviations from this null. Therefore, as far as the researcher is concerned with testing the functional form of the regression function $m$, we argue that our statistic $T_n$ tests directly the null hypothesis $H_0$ and possesses desirable local power properties compared to the conventional tests on $H'_0$.

3. ASYMPTOTIC PROPERTIES

In this section, we present asymptotic properties of the test statistic $T_n$ defined in (1). We first derive the limiting distribution of $T_n$ under the null hypothesis $H_0$. To this end, we impose the following assumptions.

**Assumption D.**

(i): $\{Y_i, X_i, \epsilon_i\}_{i=1}^n$ are i.i.d. $\epsilon$ is independent of $(Y,X)$ and has a known density $f_\epsilon$.

(ii): $f^{\alpha}, m^{\alpha}, \frac{\partial m^{\alpha}}{\partial \theta} \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$.

(iii): $K^{\alpha}(t)$ is compactly supported on $[-1,1]^d$, is symmetric around zero (i.e., $K^{\alpha}(t) = K^{\alpha}(-t)$), and is bounded.

(iv): As $n \to \infty$, it holds that $b \to 0$ and $nb^d \to \infty$.

Assumption D (i) is common in the literature of classical measurement error. Extensions to the case of unknown $f_\epsilon$ will be discussed in Section 3.1. Assumption D (ii) contains boundedness conditions on the Fourier transforms of the density $f$ of $X$, regression function $m$, and derivative of $m_\theta$ with respect to $\theta$. Assumption D (iii) and (iv) contain standard conditions on the kernel function $K$ and bandwidth $b$, respectively. A popular choice for the kernel function in the context of deconvolution methods is the sinc kernel $K(x) = \frac{\sin \pi x}{\pi x}$ whose Fourier transform is equal to $K^{\alpha}(t) = \mathbb{1}\{-1 \leq t \leq 1\}$.

For additional assumptions, we consider two cases characterized by bounds on the decay rate of the tail of the characteristic function of the measurement error, $f^{\alpha}_\epsilon$. Let $\sigma^2(x) = E[U^2|X = x]$ be the conditional variance of the error term. The first case, called the ordinary smooth measurement error case, contains the following assumptions.

**Assumption O.**

(i): $f^{\alpha}_\epsilon(t) \neq 0$ for all $t \in \mathbb{R}^d$ and there exist positive constants $c$, $C$, and $\alpha$ such that

$$c|t|^{-\alpha} \leq |f^{\alpha}_\epsilon(t)| \leq C|t|^{-\alpha},$$

5
as $|t| \to \infty$.

(ii): $\int |t|^{-2\alpha}|f^\mathfrak{R}(t)|^2 dt < \infty$, $\int |t|^{-2\alpha}|m^\mathfrak{R}(t)|^2 dt < \infty$, $\int |t|^{-2\alpha}||m^\mathfrak{R}f^\mathfrak{R}(t)|^2 dt < \infty$, $\int |t|^{-2\alpha}||m^2f^\mathfrak{R}(t)|^2 dt < \infty$, and $\int |t|^{-2\alpha}|\sigma^2 f^\mathfrak{R}(t)|^2 dt < \infty$.

(iii): $\hat{\theta} - \theta = o_p(n^{-1/2}b^{-d/4-\alpha})$ under $H_0$.

Assumption O (i) requires that the Fourier transform $f^\mathfrak{R}_\epsilon$ decay in some finite power. A popular example of the ordinary smooth density is the Laplace density. Assumption O (ii) contains boundedness conditions on the Fourier transforms of the density $f$ of $X$, regression function $m$, and conditional error variance $\sigma^2$. Assumption O (iii) is on the convergence rate of the estimator $\hat{\theta}$ for $\theta$ when the parametric model is correctly specified. Note that this assumption is satisfied if $\hat{\theta}$ is $\sqrt{n}$-consistent for $\theta$. When the regression model under the null hypothesis is linear (i.e., $m_\theta(x) = x^\theta$), we can employ the methods in, for example Gleser (1981), Bickel and Ritov (1987), and van der Vaart (1988). For nonlinear regression, we may choose the estimators by e.g., Taupin (2001) and Butucea and Taupin (2008) under certain regularity conditions. It is interesting to note that in contrast to the no measurement error case as in Härdle and Mammen (1993), the limiting distribution of the estimation error $\sqrt{n}(\hat{\theta} - \theta)$ does not influence the first-order asymptotic properties of the test statistic $T_n$. This is due to the fact that the measurement error slows down the convergence rate of the dominant term of $T_n$.

For the second case, known as the supersmooth measurement error case, we concentrate on the case of scalar $X$ (i.e., $d = 1$), and impose the following assumptions.

**Assumption S.** Suppose $d = 1$.

(i): $f^\mathfrak{R}_\epsilon(t) \neq 0$ for all $t \in \mathbb{R}$ and there exist positive constants $C_\epsilon$, $\mu$, $\gamma_0$, and $\gamma > 1$ such that
\[
f^\mathfrak{R}_\epsilon(t) \sim C_\epsilon|t|^\gamma_0 e^{-|t|^\gamma/\mu},\]
as $|t| \to \infty$. Also, there exist constants $A > 0$ and $\beta \geq 0$ such that
\[
K^\mathfrak{R}(1 - t) = At^\beta + o(t^\beta),
\]as $t \to 0$.

(ii): $E[Y^4] < \infty$, $E[W^4] < \infty$, $\int |t|^{2\beta} \left| \frac{\partial m_0^\mathfrak{R}(t)}{\partial \theta} \right|^2 dt < \infty$, and $\int |t|^{2\beta}|m^\mathfrak{R}(t)|^2 dt < \infty$.

(iii): $\hat{\theta} - \theta = o_p(n^{-1/2}b^{(\gamma-1)/2+\gamma/\beta+\gamma_0}e^{1/(\mu b^{\gamma})})$.

Assumption S (i) is adopted from Holzmann and Boysen (2006). This assumption requires that the Fourier transform $f^\mathfrak{R}_\epsilon$ decay at an exponential rate. An example of the supersmooth density satisfying this assumption is the normal density, where $C_\epsilon = 1$, $\gamma_0 = 0$, $\gamma = 2$, and $\mu = 2$. However, due to the requirement $\gamma > 1$, the Cauchy density is excluded. As is clarified in the proof of Theorem 1 (iii) below, the condition $\gamma > 1$ is imposed to make a bias term negligible. Assumption S (i) also contains an additional condition on the kernel function. For example, the sinc kernel $K(x) = \frac{\sin x}{x}$ satisfies this assumption with $A = 1$ and $\beta = 0$. Similarly to the ordinary smooth case, Assumption S (ii) contains boundedness conditions on the Fourier transforms, and Assumption S (iii) is on the convergence rate of the estimator $\hat{\theta}$. Again, the $\sqrt{n}$-consistency of $\hat{\theta}$ is sufficient.
Under these assumptions, the null distribution of $T_n$ is obtained as follows.

**Theorem 1.**

(i): Suppose that Assumptions D and O hold true. Then under $H_0$,

$$C_{V,b}^{-1/2}T_n \overset{d}{\to} N \left(0, \frac{2}{(2\pi)^{2d}} \right),$$

where $C_{V,b} = O(b^{-d-4\alpha})$ is defined in (2) in the Appendix.

(ii): Suppose that Assumptions D and S hold true with $d = 1$ and $\epsilon \sim N(0,1)$. Then under $H_0$,

$$\varphi(b)T_n \overset{d}{\to} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\varphi(b) = \frac{(2\pi)^{2\beta}}{b^{1+4\beta}/2^\beta A^2 \Gamma(1+2\beta)}$ with the gamma function $\Gamma$, $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ is defined in (12) in the Appendix.

(iii): Suppose that Assumptions D and S hold true with $d = 1$. Then under $H_0$,

$$\phi(b)T_n \overset{d}{\to} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

where $\phi(b) = \frac{(2\pi)^{2\beta} \mu^{1+2\beta} b^{1+4\beta}/2^\beta e^{2}/(\mu b^2)^{1/2}}{2 \mu^{1+2\beta} b^{1+4\beta}/2^\beta A^2 \Gamma(1+2\beta)}$ with the gamma function $\Gamma$, $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ is defined in (12) in the Appendix.

Theorem 1 (i) says that for the ordinary smooth case, the test statistic $T_n$ is asymptotically normal. The normalizing term $C_{V,b}$ comes from the variance of the U-statistic of the leading term in $T_n$. Note that the convergence rate $C_{V,b}^{-1/2} = O(b^{d/2+2\alpha})$ of the statistic $T_n$ is slower than the rate $O(b^{d/2})$ of Härdle and Mammen’s (1993) statistic for the no measurement error case. As the dimension $d$ of $X$ or the decay rate $\alpha$ of $f_{ft}$ increases, the convergence rate of $T_n$ becomes slower.

Theorem 1 (ii) focuses on the case of the normal measurement error, and shows that the test statistic converges to the weighted sum of chi-squared random variables. The normalizing term $\varphi(b)$ is characterized by the shape of the kernel function specified in Assumption S (i). For example, if we employ the sinc kernel (i.e., $A = 1$ and $\beta = 0$), then the normalization becomes $\varphi(b) = \frac{2\pi}{be^{1/b^2} \Gamma(1)}$. In this supersmooth case, the non-normal limiting distribution emerges because the leading term of the statistic $T_n$ is characterized by the degenerate U-statistic with a fixed kernel (see, e.g., Serfling, 1980, Theorem 5.5.2). In contrast, for the ordinary smooth case in Part (i) of this theorem, the leading term is characterized by a U-statistic with a varying kernel so that the central limit theorem in Hall (1984) applies. An analogous result is obtained in Holzmann and Boysen (2006) for the integrated squared error of the deconvolution density estimator.

Theorem 1 (iii) presents the limiting null distribution of the test statistic for the case of general supersmooth measurement errors. In this case, after normalization by $\phi(b)$, the test statistic obeys the same limiting distribution as the normal case in Part (ii) of this theorem. Thus, similar comments to Part (ii) apply. The normalization term $\phi(b)$ is characterized by the
shapes of the kernel function and Fourier transform $f^R_\epsilon(t)$ of the measurement error specified in Assumption S (i).

Although Theorem 1 (ii) and (iii) focus on the case of scalar $\epsilon$, our technical argument may be generalized. For example, if we assume that the elements of the $d$-dimensional vector $\epsilon$ are mutually independent, then the Fourier transform $f_\epsilon$ becomes the product of the Fourier transforms of the marginals. We may impose Assumption S (i) for each marginal density. To keep things simple we can choose the multivariate kernel function to be a product kernel. With these assumptions in place, the deconvolution kernel analogously becomes a product deconvolution kernel. The proofs of the theorem remain very similar using inner products and terms defined as products over the $d$ dimensions.

Theorem 1 can be applied to obtain critical values for testing the null $H_0$ based on $T_n$. Alternatively, we can compute the critical values by some bootstrap method. A bootstrap counterpart of $T_n$ is given by perturbing each summand in $T_n$ as follows

$$T^*_n = \frac{1}{n} \sum_{i \neq j} \nu^i_n \nu^j_n \frac{1}{(2\pi)^d} \int \frac{|K^R(\lambda t)|^2}{|f^R_\epsilon(t)|^2} \zeta_i(t) \zeta_j(t) dt,$$

where $\{\nu^i_n\}_{i=1}^n$ is an i.i.d. sequence which is mean zero, unit variance, and independent of $\{Y_i, W_i\}_{i=1}^n$. The asymptotic validity of this bootstrap procedure follows by a similar argument as in Delgado, Dominguez and Lavergne (2006, Theorem 6).

In order to investigate the power properties of the test based on $T_n$, we consider the local alternative hypothesis in the form of $H_1^n$: $m(x) = m_\theta(x) + c_n \Delta(x)$, for almost every $x \in \mathbb{R}^d$ where $c_n \to 0$ and $\Delta(x)$ is a non-zero function such that the limits $\lim_{n \to \infty} \Delta_n$ and $\lim_{n \to \infty} \Upsilon_n$ defined in (14) and (15) respectively in the Appendix exist. The local power properties are obtained as follows.

**Theorem 2.**

(i): Suppose that Assumptions D and O hold true. Then under $H_{1n}$ with $c_n = n^{-1/2}b^{-d/4-\alpha}$,

$$C_{V,b}^{-1/2} T_n \xrightarrow{d} N \left( \lim_{n \to \infty} \Delta_n, \frac{2}{(2\pi)^{2d}} \right).$$

(ii): Suppose that Assumptions D and S hold true with $d = 1$ and $\epsilon \sim N(0, 1)$. Then under $H_{1n}$ with $c_n = n^{-1/2}b^{1/2+2\beta}e^{1/(2b^2)}$,

$$\varphi(b) T_n \xrightarrow{d} \lim_{n \to \infty} \Upsilon_n + \sum_{k=1}^\infty \lambda_k (Z^2_k - 1).$$

(iii): Suppose that Assumptions D and S hold true with $d = 1$. Then under $H_{1n}$ with $c_n = b^{(1-1)/2+\lambda} \lambda e^{1/(ub^2)}$,

$$\phi(b) T_n \xrightarrow{d} \lim_{n \to \infty} \Upsilon_n + \sum_{k=1}^\infty \lambda_k (Z^2_k - 1),$$
Theorem 2 (i) says that under the ordinary smooth case, our test has non-trivial power against local alternatives drifting with the rate of $c_n = n^{-1/2}b^{-d/4-\alpha}$. This is a nonparametric rate, and the test based on $T_n$ becomes less powerful as the dimension $d$ of $X$ or the decay rate $\alpha$ of $f_{\epsilon}^{th}$ increases. For the no measurement error case, Härdle and Mammen’s (1993) statistic has non-trivial power for local alternatives with the rate of $n^{-1/2}b^{-d/4}$. Therefore, the test becomes less powerful due to the measurement error. Theorem 2 (ii) and (iii) present local power properties of our test for the normal and general supersmooth measurement error cases, respectively. Except for the normalizing constants, the test statistic has the same limiting distribution. Also, for $c_n \to 0$, the bandwidth $b$ should decay at a log rate. As an example, consider the case of $\epsilon \sim N(0, 1)$. In this case, if we choose $b \sim (\log n)^{-1/2}$, then the rate for the local alternative will be $c_n \sim (\log n)^{-1/4-\beta}$. Therefore, for the supersmooth case, the rate for the local alternative is typically a log rate.

3.1. Case of unknown $f_{\epsilon}$. In practical applications, it is sometimes unrealistic to assume that the density $f_{\epsilon}$ of the measurement error is known to the researcher. In the literature of nonparametric deconvolution methods several estimation methods for $f_{\epsilon}$ are available, these are typically based on additional data (see, e.g., Section 2.6 of Meister (2009) for a review). Although the analysis of the asymptotic properties is different, we can modify the test statistic $T_n$ by inserting the estimated Fourier transform $\hat{f}_{\epsilon}^{th}$ of the measurement error density.

For example, suppose the researcher has access to repeated measurements on $X$ in the form of $W = X + \epsilon$ and $W^r = X + \epsilon^r$, where $\epsilon$ and $\epsilon^r$ are identically distributed and $(X, \epsilon, \epsilon^r)$ are mutually independent, see Delaigle, Hall and Meister (2008) for a list of examples of such repeated measurements. If we further assume that the Fourier transform $f_{\epsilon}^{th}$ is real-valued (or the density $f_{\epsilon}$ is symmetric around zero), then we can employ the estimator proposed by Delaigle, Hall and Meister (2008)

$$\hat{f}_{\epsilon}^{th}(t) = \left| \frac{1}{n} \sum_{i=1}^{n} \cos\{t(W_i - W^r_i)\} \right|^{1/2}.$$

Delaigle, Hall and Meister (2008) studied the asymptotic properties of the deconvolution density and regression estimators using $f_{\epsilon}^{th}$ and found conditions to guarantee that the differences between the estimators with known $f_{\epsilon}$ and the ones with unknown $f_{\epsilon}$ are asymptotically negligible. Under similar conditions, we can expect that the asymptotic distributions of the test statistic $T_n$ obtained above remain unchanged even if we replace $f_{\epsilon}^{th}$ with $\hat{f}_{\epsilon}^{th}$. If the researcher wants to remove the assumption that $f_{\epsilon}^{th}$ is real-valued, it may be possible to employ the estimator by Li and Vuong (1998) based on Kotlarski’s identity.

4. Simulation

We evaluate the small sample performance of our test through a Monte Carlo experiment. To begin with we consider the same data generating process as Hall and Ma (2007) for ease of comparison. We recall that although Hall and Ma’s (2007) test is confined to polynomial regression models, our test allows nonlinear models. Specifically we take the true unobservable regressor $\{X_i\}_{i=1}^{n}$ to be distributed as $U[-3, 4]$ and $Y_i = 1 + 1.5X_i + C\cos(X_i) + U_i$, where $U_i \sim$
$N(0,1)$ and $C$ is a constant to be varied. The contaminated regressor is given by $W_i = X_i + \epsilon_i$. We consider two distributions for $\epsilon_i$ to be drawn from. For the ordinary smooth case, we use the Laplace distribution with variance of 0.5. For the supersmooth case, we use $N(0,1)$. We use the following kernel for our simulations (Fan, 1992)

$$K(x) = \frac{48 \cos(x)}{\pi x^4} \left( 1 - \frac{15}{x^2} \right) - \frac{144 \sin(x)}{\pi x^5} \left( 2 - \frac{5}{x^2} \right).$$

We report results for a range of sample sizes, bandwidths, and nominal levels of the test. Specifically, for the ordinary and super smooth cases, we choose the bandwidths according to the rules of thumb $b = c \left( \frac{5\sigma^4}{n} \right)^{1/9}$ and $b = c \left( \frac{2\sigma^2}{\log(n)} \right)^{1/2}$, respectively, where $\sigma$ is the standard deviation of the measurement error and $c$ varies in the grid $\{0.01, 0.05, 0.1, 0.5, 1.0, 1.5\}$ so that we can analyse the sensitivity of our test to the bandwidth. For the parametric estimator we use the polynomial estimator of degree 2 proposed by Cheng and Schneeweiss (1998) so as to remain consistent with the experiment conducted by Hall and Ma (2007). All results are based on 1000 Monte Carlo replications.

Table 1 takes $C = 0$ so as to assess the level accuracy of our test. To study the power properties of the test, we take $C = 1.5$ in Table 2. The critical values are based on 99 replications of the bootstrap procedure (results were very similar for 199 replications and hence are not reported). The perturbation random variable $\nu^*$ for the bootstrap is drawn from the Rademacher distribution.

Finally, to highlight the power advantages of our test under high frequency alternatives we consider the slightly altered data generating process $Y_i = 1 + 1.5X_i + \cos(\pi\delta X_i) + U_i$, where $\delta$ is a constant to be varied; larger values corresponding to higher frequency alternatives. All other parameter settings remain unchanged. Results for these experiments are shown in Tables 3-5, where the column labeled ‘HM’ corresponds to the power of the test proposed in Hall and Ma (2007).
Table 1: $Y = 1 + 1.5X + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>50</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>100</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>Super Smooth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
</tbody>
</table>

Table 2: $Y = 1 + 1.5X + 1.5 \cos(X) + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>Bandwidth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
</tr>
<tr>
<td>50</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>100</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>Super Smooth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
</tbody>
</table>
Table 3: $Y = 1 + 1.5X + \cos(\pi X) + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>Bandwidth</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>21.3%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>40.3%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>51.4%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>35.6%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>54.5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>66.4%</td>
</tr>
<tr>
<td>Super Smooth</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>15.3%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>29.0%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>38.7%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>23.8%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>40.9%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>52.9%</td>
</tr>
</tbody>
</table>

Table 4: $Y = 1 + 1.5X + \cos(2\pi X) + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>Bandwidth</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>20.9%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>38.7%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>49.3%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>35.9%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>55.9%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>66.8%</td>
</tr>
<tr>
<td>Super Smooth</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>16.1%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>30.4%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>41.3%</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1%</td>
<td>23.6%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>39.4%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>50.8%</td>
</tr>
</tbody>
</table>

12
Table 5: $Y = 1 + 1.5X + \cos(3\pi X) + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>Bandwidth</th>
<th>HM</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>Level</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>1%</td>
<td>22.8%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>39.3%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>50.6%</td>
</tr>
<tr>
<td>200</td>
<td>1%</td>
<td>36.4%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>54.1%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>67.1%</td>
</tr>
<tr>
<td>Super Smooth</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>1%</td>
<td>17.4%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>31.4%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>42.0%</td>
</tr>
<tr>
<td>200</td>
<td>1%</td>
<td>21.5%</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>39.1%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>50.9%</td>
</tr>
</tbody>
</table>

The results are encouraging and seem to be consistent with the theory. Table 1 indicates that our test tracks the nominal level relatively closely. There does appear to be some dependence on the bandwidth; smaller bandwidths tending to lead to an over-rejection and larger bandwidths leading to under-rejection of the null hypothesis. Table 2 gives a direct comparison to the results presented in Hall and Ma (2007). As we expected, in this low frequency alternative setting, our test is generally slightly less powerful. Having said this, in the ordinary smooth case for several choices of bandwidth our test outperforms that of Hall and Ma (2007). Hall and Ma’s (2007) test achieves a $\sqrt{n}$-rate for both ordinary and super smooth measurement error distributions, whereas our test achieves a slower polynomial rate in the ordinary smooth case but only a $\log(n)$-rate in the super smooth case. Thus it is not surprising to see our test underperform when faced with the Gaussian measurement error. The test is still able to enjoy considerable power in this case especially for larger sample sizes. However we can learn from these simulations that for reasonably small samples with supersmooth measurement errors, perhaps Hall and Ma’s (2007) test would be a wiser choice if one suspects deviations from the null of a low frequency type.

On the other hand, as mentioned earlier, we suspect that our test is better suited to detecting high frequency alternatives than Hall and Ma (2007). This is confirmed in Tables 3-5. We find that for smaller bandwidths our test is more powerful across the range of $c$. Unfortunately, the power of our test shows considerable variation across the bandwidth choices. For smaller bandwidths the power is generally much higher. This is intuitive and is explained in Fan and Li (2000). Nonsmoothing tests can be thought of as smoothing tests but with a fixed bandwidth. Thus it is the asymptotically vanishing nature of the bandwidth in smoothing tests that allows for the superior detection of high frequency alternatives. When smaller bandwidths are employed, the test is able to pick up on these rapid changes more readily.
Hereafter, \( f(b) \sim g(b) \) means \( f(b)/g(b) \to 1 \) as \( b \to 0 \).

### A.1. Proof of Theorem 1.

#### A.1.1. Proof of (i). First, we define the normalization term \( C_{V,b} \) and characterize its asymptotic order. Let

\[
\xi_i(t) = Y_i e^{itW_i} - \int e^{isW_i} m_i^f(t-s) \frac{K^f(sb)}{f^f(s)} ds f^f(t),
\]

\[
H_{ij} = \int \frac{|K^f(tb)|^2}{|f^f(t)|^2} \xi_i(t) \xi_j(t) dt.
\]

Then \( C_{V,b} \) is defined as

\[
C_{V,b} = E[H_{12}^2] = \int \int \frac{|K^f(t_1b)|^2}{|f^f(t_1)|^2} \frac{|K^f(t_2b)|^2}{|f^f(t_2)|^2} \left\{ \left[ m^2 f^f(t_1 + t_2) + |\sigma^2 f^f(t_1 + t_2)| \right] f^f_i(t_1 + t_2) \right. \\
+ \int f^f_W(s_1 + s_2) m^f_i(t_1 - s_1) m^f(t_2 - s_2) K^f(s_1 b) K^f(s_2 b) f^f_i(s_1) f^f_i(s_2) ds_1 ds_2 f^f_i(t_1) f^f_i(t_2) \\
- \int [m^f_i(t_2 + s_1) f^f_i(t_2 + s_1) m^f_i(t_1 - s_1) K^f(s_1 b) f^f_i(s_1)] ds_1 f^f_i(t_1) \\
- \int [m^f_i(t_1 + s_1) f^f_i(t_1 + s_1) m^f_i(t_2 - s_1) K^f(s_1 b) f^f_i(s_1)] ds_1 f^f_i(t_2) \left. \right|^2 dt_1 dt_2.
\]

To find the order of \( C_{V,b} \), we consider the square of each of these four terms and all of their cross products. For example,

\[
\int \int \frac{|K^f(t_1b)|^2}{|f^f(t_1)|^2} \frac{|K^f(t_2b)|^2}{|f^f(t_2)|^2} \left\{ \left[ m^2 f^f(t_1 + t_2) + |\sigma^2 f^f(t_1 + t_2)| \right] f^f_i(t_1 + t_2) \right. \\
\sim b^{-2d-4a} \int \int |K^f(a_1)|^2 |K^f(a_2)|^2 |a_1|^{2a_1} |a_2|^{2a_2} |(a_1 + a_2)/b|^{-2a} \left| (m^2 + \sigma^2 f^f((a_1 + a_2)/b)) \right|^2 da_1 da_2 \\
\sim b^{-d-4a} \int |a|^{-2a} \left| (m^2 + \sigma^2 f^f(a)) \right|^2 da \int |K^f(a_2)|^4 |a_2|^{4a} da_2 \\
= O(b^{-d-4a}),
\]

where the first wave relation follows from the change of variables \( (a_1, a_2) = (t_1 b, t_2 b) \) and Assumption O (i), the second wave relation follows from the change of variables \( a = (a_1 + a_2)/b \), and the equality follows from Assumption D (iii) and O (ii). Since all other squared and cross terms can be bounded in the same manner, we obtain \( C_{V,b} = O(b^{-d-4a}) \).

Second, we show that the estimation error of \( \theta \) is negligible for the limiting distribution of \( T_n \). Decompose \( \zeta_i(t) = \xi_i(t) + \rho_i(t) \), where

\[
\rho_i(t) = \int e^{isW_i} \left( m^f_\theta(t-s) - m^f_\theta(t-s) \right) \frac{K^f(sb)}{f^f(s)} ds f^f_i(t).
\]
Then the test statistic $T_n$ is written as

$$T_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \left| K^{f^i(t)}(t) \right|^2 \xi_i(t)\xi_j(t)dt + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \left| K^{f^j(t)}(t) \right|^2 \rho_i(t)\rho_j(t)dt$$

$$+ \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \left| K^{f^i(t)}(t) \right|^2 \rho_i(t)\xi_j(t)dt + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \left| K^{f^j(t)}(t) \right|^2 \xi_i(t)\rho_j(t)dt$$

$$\equiv \hat{T}_n + R_{1n} + R_{2n} + R_{3n}.$$ 

By an expansion around $\theta = \theta$ and Assumption O (iii), the term $R_{1n}$ satisfies

$$R_{1n} = o_p(b^{-d/2-2\alpha}) \left| \frac{1}{n^2} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \left| K^{f^i(t)}(t) \right|^2 \rho_i(t)\rho_{ij}(t)dt \right|,$$  

where $\rho_{ij}(t) = \int e^{i\theta W^i} \frac{\partial m^{(i)}(s)}{\partial \theta} K^{f^i(s)} ds f^i(t)$. By the Cauchy-Schwarz inequality and Assumption D (ii),

$$E \left[ \left| \int \left| K^{f^i(t)}(t) \right|^2 \rho_i(t)\rho_{ij}(t)dt \right| \right] = \int \left| K^{f^i(t)}(t) \right|^2 \left| \int f^i(s) \frac{\partial m^{(i)}(t-s)}{\partial \theta} K^{f^i(s)} ds \right|^2 dt$$

$$= O(1).$$

Also, by applying the same argument to (3) under Assumption O (ii), we have

$$E \left[ \left( \int \left| K^{f^i(t)}(t) \right|^2 \rho_i(t)\rho_{ij}(t)dt \right)^2 \right] = O(b^{-d-4\alpha}).$$

Combining (4)-(6) and $C_{V,b} = O(b^{-d-4\alpha})$, we obtain $C_{V,b}^{-1/2}R_{1n} = o_p(1)$. In the same manner we can show $C_{V,b}^{-1/2}R_{2n} = o_p(1)$ and $C_{V,b}^{-1/2}R_{3n} = o_p(1)$ under Assumption O (ii)-(iii) and thus $C_{V,b}^{-1/2}T_n = C_{V,b}^{-1/2}\hat{T}_n + o_p(1)$.

Second, we derive the limiting distribution of $C_{V,b}^{-1/2}\hat{T}_n$. Note that $\hat{T}_n$ is written as $\hat{T}_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} H_{i,j}$ and is a U-statistic with zero mean (because $E[Y \exp(itW)] = [m_0] f^i(t)$ under $H_0$). To prove the asymptotic normality of $\hat{T}_n$, we apply the central limit theorem in Hall (1984, Theorem 1). To this end, it is enough to show

$$\frac{E[H_{1,2}^4]}{n(E[H_{1,2}^2])^2} \rightarrow 0, \quad \text{and} \quad \frac{E[G_{1,2}^2]}{(E[H_{1,2}^2])^2} \rightarrow 0,$$

where $G_{i,j} = E[H_{1,i}H_{1,j}|Y_1,W_1]$. Recall that $C_{V,b} = E[H_{1,2}^2]$ defined in (2) satisfies $C_{V,b} = O(b^{-d-4\alpha})$. By a similar argument to bound $E[H_{1,2}^2]$ in (3), we can show

$$E[H_{1,2}^2] = E \left[ \int \cdots \int \prod_{k=1}^4 \left| K^{f^i(t_k)}(t_k) \right|^2 \xi_1(t_k)\xi_2(t_k)dt_1 \cdots dt_4 \right] = O(b^{-3d-8\alpha}).$$
For $E[C_{1,2}^2]$, we can equivalently look at
\[
E[H_{1,3}H_{1,4}H_{2,3}H_{2,4}]
= \int \cdots \int \prod_{k=1}^{4} \frac{|K^f(t_k b)|^2}{|f^f(t_k)|^2} \xi_1(t_1)\xi_3(t_3)\xi_2(t_2)\xi_4(t_4)dt_1 \cdots dt_4
= O(b^{d-8\alpha}).
\]
These results combined with Assumption D (iv) guarantee the conditions in (7). Thus, Hall (1984, Theorem 1) implies
\[
C_{V,b}^{-1/2} \tilde{T}_n \xrightarrow{d} N \left(0, \frac{2}{(2\pi)^{2d}} \right),
\]
and the conclusion follows.

A.1.2. Proof of (ii). A similar argument to the proof of Part (i) guarantees $\varphi(b)T_n = \varphi(b)\tilde{T}_n + o_p(1)$. Thus we hereafter derive the limiting distribution of $\tilde{T}_n$. Decompose $\tilde{T}_n = \tilde{T}_n + r_{1n} + r_{2n} + r_{3n}$, where

\[
\tilde{T}_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{2\pi} \int \frac{|K^f(t b)|^2}{|f^f(t)|^2} Y_i e^{itW_i} Y_j e^{itW_j} dt,
\]
\[
r_{1n} = \frac{1}{n} \sum_{i \neq j} \frac{1}{2\pi} \int \frac{|K^f(t b)|^2}{|f^f(t)|^2} \left( \int e^{iW_i t} \frac{K^\ell(s)}{f^\ell(s)} dt \right) \frac{K^f(s)}{f^f(s)} ds \frac{f^f(t)}{f^\ell(s)} dt,
\]
\[
r_{2n} = \frac{1}{n} \sum_{i \neq j} \frac{1}{2\pi} \int \frac{|K^f(t b)|^2}{|f^f(t)|^2} Y_i e^{itW_i} \left( \int e^{iW_j (t-s)} \frac{K^\ell(s)}{f^\ell(s)} ds \frac{f^f(t)}{f^\ell(s)} dt \right) dt,
\]
\[
r_{3n} = \frac{1}{n} \sum_{i \neq j} \frac{1}{2\pi} \int \frac{|K^f(t b)|^2}{|f^f(t)|^2} \left( \int e^{iW_i (t-s)} \frac{K^\ell(s)}{f^\ell(s)} ds \frac{f^f(t)}{f^\ell(s)} dt \right) Y_j e^{itW_j} dt.
\]

First, we derive the limiting distribution of $\tilde{T}_n$. Observe that
\[
\tilde{T}_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{2\pi} \int |K^f(t b)|^2 e^{itW_i} Y_i Y_j \left( \cos(tW_i) \cos(tW_j) + \sin(tW_i) \sin(tW_j) \right) dt
\]
\[
= \frac{1}{nb} \sum_{i \neq j} \frac{1}{2\pi} \int |K^f(t b)|^2 e^{itW_i/b} Y_i Y_j \left( \cos \left( \frac{tW_i}{b} \right) \cos \left( \frac{tW_j}{b} \right) + \sin \left( \frac{tW_i}{b} \right) \sin \left( \frac{tW_j}{b} \right) \right) dt
\]
\[
= \left( \frac{1}{b^2} \right) \int |K^f(t b)|^2 e^{itW_i/b} dt \sum_{i \neq j} Y_i Y_j \left( \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right)
+ O_p(b^{2+4\beta} e^{1/b^2})
\]
\[
= \left( \frac{1}{b^2} \right) \int |K^f(t b)|^2 e^{itW_i/b} dt \tilde{T}_n + O_p(b^{2+4\beta} e^{1/b^2}),
\]
where the first equality follows from $f^f(t) = e^{-t^2/2}$ and $e^{itW_i} = \cos(tW_i) + i \sin(tW_i)$, the second equality follows from a change of variables, and the third equality follows from Holzmann and
Boysen (2006, Theorem 1) based on Assumption S (ii). Note that
\[
\tilde{T}_n = \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left[ \left\{ \cos \left( \frac{x_i}{b} \right) \cos \left( \frac{\pi}{b} \right) - \sin \left( \frac{x_i}{b} \right) \sin \left( \frac{\pi}{b} \right) \right\} \left\{ \cos \left( \frac{x_j}{b} \right) \cos \left( \frac{\pi}{b} \right) - \sin \left( \frac{x_j}{b} \right) \sin \left( \frac{\pi}{b} \right) \right\} + \left\{ \sin \left( \frac{x_i}{b} \right) \cos \left( \frac{\pi}{b} \right) + \cos \left( \frac{x_i}{b} \right) \sin \left( \frac{\pi}{b} \right) \right\} \left\{ \sin \left( \frac{x_j}{b} \right) \cos \left( \frac{\pi}{b} \right) + \cos \left( \frac{x_j}{b} \right) \sin \left( \frac{\pi}{b} \right) \right\} \right].
\] (10)

From van Es and Uh (2005, proof of Lemma 6), it holds \( \tilde{X}_i \sim U[0, 2\pi] \) for \( b \to 0 \) for each \( i \), where \( \tilde{X}_i \) is independent from \( (Y_i, V_i^X) \). Thus by applying Holzmann and Boysen (2006, Lemma 1), \( \tilde{T}_n \) has the same limiting distribution with \( \tilde{T}^V_n = \frac{1}{n} \sum_{i \neq j} h(Q_i, Q_j) \), where \( Q_i = (Y_i, V_i^X, V_i^Y) \) and
\[
h(Q_i, Q_j) = Y_i Y_j \left[ \left\{ \cos(V_i^X) \cos(V_j^X) - \sin(V_i^X) \sin(V_j^Y) \right\} \left\{ \cos(V_j^X) \cos(V_j^Y) - \sin(V_j^X) \sin(V_j^Y) \right\} + \left\{ \sin(V_i^X) \cos(V_j^X) + \cos(V_i^X) \sin(V_j^Y) \right\} \left\{ \sin(V_j^X) \cos(V_j^Y) + \cos(V_j^X) \sin(V_j^Y) \right\} \right].
\]
Observe that \( \text{Cov}(h(Q_1, Q_2), h(Q_1, Q_3)) = 0 \) because \( E[\cos(V_i^Y)] = 0 \). Therefore, by applying the limit theorem for degenerate U-statistics with a fixed kernel \( h \) (Serfling, 1980, Theorem 5.5.2), we obtain
\[
\tilde{T}^V_n \overset{d}{\to} \sum_{k=1}^{\infty} \lambda_k(Z_k^2 - 1),
\] (11)
where \( \{Z_k\} \) is an independent sequence of standard normal random variables and \( \{\lambda_k\} \) are the eigenvalues of the integral operator
\[
(Ag)(Q_1) = \lambda g(Q_1),
\] (12)
where \( (Ag)(Q_1) = E[h(Q_1, Q_2)g(Q_2)|Q_1] \). Also, van Es and Uh (2005, Lemma 5) gives
\[
\frac{1}{2\pi} \int |K^t(t)|^2 e^{(t/b)^2} dt \sim \frac{b}{\varphi(b)},
\] (13)
where \( \Gamma(\cdot) \) is the gamma function. Combining (9)-(13),
\[
\varphi(b) \tilde{T}_n \overset{d}{\to} \sum_{k=1}^{\infty} \lambda_i(Z_i^2 - 1).
\]

Next, we show that \( r_{1n} \) is negligible. Observe that
\[
r_{1n} = \frac{1}{nb^3} \sum_{i \neq j} \frac{1}{2\pi} \int |K^t(t)|^2 \left( \int e^{isW_i/b} m^n \left( \frac{t - s}{b} \right) K^s(s) f^t_b(s/b) ds \right) \left( \int e^{isW_j/b} m^n \left( \frac{t - s}{b} \right) K^s(s) f^t_b(s/b) ds \right) dt
\]
\[
= \frac{1}{nb^3} \sum_{i \neq j} \frac{1}{2\pi} \int |K^t(t)|^2 \left[ \left\{ \int \cos \left( \frac{s_1 W_i}{b} \right) m^n \left( \frac{t-s_1}{b} \right) \frac{K^s(s_1)}{f^t_b(s_1/b)} ds_1 \right\} \left\{ \int \cos \left( \frac{s_2 W_i}{b} \right) m^n \left( \frac{t-s_2}{b} \right) \frac{K^s(s_2)}{f^t_b(s_2/b)} ds_2 \right\} \right. \\
\left. + \left\{ \int \sin \left( \frac{s_1 W_i}{b} \right) m^n \left( \frac{t-s_1}{b} \right) \frac{K^s(s_1)}{f^t_b(s_1/b)} ds_1 \right\} \left\{ \int \sin \left( \frac{s_2 W_i}{b} \right) m^n \left( \frac{t-s_2}{b} \right) \frac{K^s(s_2)}{f^t_b(s_2/b)} ds_2 \right\} \right] dt
\]
\[
= \left( \frac{1}{2\pi} \int \int |K^t(t)|^2 \frac{K^s(s_1)}{f^t_b(s_1/b)} f^t_b(-s_2/b) m^n \left( \frac{t-s_1}{b} \right) \frac{m^n \left( \frac{s_2 - t}{b} \right) ds_1 ds_2 dt} \right) \times \frac{1}{nb^3} \sum_{i \neq j} \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\} + O_p(b^{2+4\beta} e^{1/b^2}),
\]
where the first equality follows from a change of variables, the second equality follows from a direct calculation using \( e^{i\theta W_i} = \cos(s W_i) + i\sin(s W_i) \), the third equality follows from Holzmann and Boysen (2006, Theorem 1) based on Assumption S (ii). By a similar argument to show (11), it holds
\[
\frac{1}{n} \sum_{i \neq j} \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\} = O_p(1).
\]

Also, we obtain
\[
\frac{1}{2\pi} \int \int \left| K^{\beta}_{\gamma} (t) \right|^2 K^{\beta}_{\gamma} (s_1) K^{\beta}_{\gamma} (s_2) m^\alpha \left( \frac{t - s_1}{b} \right) m^\alpha \left( \frac{s_2 - t}{b} \right) ds_1 ds_2 dt
\]
\[
= \frac{b^4 e^{1/b^2}}{2\pi} \int \int \left[ e^\left( \frac{1-b^2 v_1}{2b^2} \right)^2 \frac{K^{\beta}_{\gamma}(t)^2 K^{\beta}_{\gamma}(1-b^2 v_1) K^{\beta}_{\gamma}(1-b^2 v_2)}{e^\left( \frac{1-b^2 v_2}{2b^2} \right)} m^\alpha \left( \frac{t - 1}{b} \right) m^\alpha \left( \frac{1 - t}{b} \right) dt \right] dv_1 dv_2 dt
\]
\[
= \frac{A^2 b^{4+2\beta} e^{1/b^2}}{2\pi} \int |t|^{2\beta} |m^\beta(t)|^2 dt
\]
\[
= O(b^{5+2\beta} e^{1/b^2}),
\]
where the first equality follows from changes of variables \( s_1 = 1 - b^2 v_1 \) and \( s_2 = 1 - b^2 v_2 \), the wave relations follow from Assumption S (i), and the last equality follows from Assumption S (ii). Combining these results,
\[
\varphi(b) r_{1n} = O_p(b^{1+2\beta}),
\]
and thus \( r_{1n} \) is negligible. Similar arguments imply that the terms \( r_{2n} \) and \( r_{3n} \) are also asymptotically negligible. Therefore, the conclusion follows.

A.1.3. Proof of (iii). The proof for the general supersmooth case follows the same steps as in the proof of Part (ii) for the normal case. As the proof is similar, we omit the most part. Hereafter we show why the condition \( \gamma > 1 \) is imposed in this case. The dominant term \( \tilde{T}_n \) defined in (8) satisfies
\[
\tilde{T}_n \sim \frac{1}{nb} \sum_{i \neq j} \frac{1}{2\pi C^2_e} \int |K^{\beta}_{\gamma}(t)|^2 \left| \frac{t}{b} \right|^{2\gamma_0} e^{2i\gamma_7 Y_i Y_j} \left\{ \cos \left( \frac{t W_i}{b} \right) \cos \left( \frac{t W_j}{b} \right) + \sin \left( \frac{t W_i}{b} \right) \sin \left( \frac{t W_j}{b} \right) \right\} dt.
\]

We now show that
\[
D_{cos} = \frac{1}{nb} \sum_{i \neq j} \frac{1}{2\pi C^2_e} \int |K^{\beta}_{\gamma}(t)|^2 \left| \frac{t}{b} \right|^{2\gamma_0} e^{2i\gamma_7 Y_i Y_j} \left\{ \cos \left( \frac{t W_i}{b} \right) \cos \left( \frac{t W_j}{b} \right) \right\} dt
\]
\[
- \left( \frac{1}{2\pi C^2_e} \int |K^{\beta}_{\gamma}(t)|^2 \left| \frac{t}{b} \right|^{2\gamma_0} dt \right) \frac{1}{nb} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) \right\}
\]
is asymptotically negligible, as well as the correspondingly defined \( D_{sin} \). We have seen that each term is zero mean. Following the proof of Holzmann and Boysen (2006, Theorem 1), we obtain
\[
\left| \cos \left( \frac{t W_i}{b} \right) \cos \left( \frac{t W_j}{b} \right) - \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) \right| \leq (1 - t) \frac{||W_i|| + |W_k||}{b}.
\]
Thus, similar arguments to van Es and Uh (2005, Lemmas 1 and 5) using Assumption S (ii) imply

\[
\text{Var}(D_{\cos}) \leq O(n^{-2}b^{4\gamma_0-4}) \left( \int (1-t)|K^f(t)|^2|t|^{-2\gamma_0} e^{2(t)^{\gamma}} dt \right)^2 \sum_{i \neq j} E \left[ |Y_i|^2 |Y_j|^2 (|W_i| + |W_k|)^2 \right] = O \left( b^{4\gamma_0-4} \left( b^{2+2\beta} e^{2(t)^{\gamma}} \right)^2 \right),
\]
and we obtain \( D_{\cos} = O_p \left( b^{2(\gamma-1)+2\gamma_3+2\gamma_0} e^{2(t)^{\gamma}} \right) \). The same argument applies to \( D_{\sin} \). Note that

\[
\hat{T}_n = \left( \frac{1}{b} \frac{1}{2\pi C^2} \int |K^f(t)|^2 \left| t \right|^{-2\gamma_0} e^{2(t)^{\gamma}} dt \right) \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\} + O \left( b^{2(\gamma-1)+2\gamma_3+2\gamma_0} e^{2(t)^{\gamma}} \right)
\]

\[
\hat{T}_n = A^2 \mu^2 b^{4\gamma_0} \frac{(\gamma_2+1)}{\lambda^{1+2\beta}} \hat{T}_n + O \left( b^{2(\gamma-1)+2\gamma_3+2\gamma_0} e^{2(t)^{\gamma}} \right)
\]

where the second equality follows from the definition of \( \hat{T}_n \) in (10) and a modification of van Es and Uh (2005, Lemma 5). Therefore, we obtain

\[
\phi(b)T_n = \hat{T}_n + O(b^{7-1}).
\]

The limiting distribution of \( \hat{T}_n \) is obtained in the proof of Part (ii). The remainder term becomes negligible if we impose \( \gamma > 1 \).

A.2. Proof of Theorem 2.

A.2.1. Proof of (i). By a similar argument to the proof of Theorem 1 (i), the estimation error \( \hat{\theta} - \theta \) is negligible for the asymptotic properties of \( T_n \) and thus it is written as

\[
T_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^f(tb)|^2}{|f^f(t)|^2} \xi_i(t) \xi_j(t) dt + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^f(tb)|^2}{|f^f(t)|^2} \eta_i(t) \eta_j(t) dt
\]

\[
+ \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^h(tb)|^2}{|f^h(t)|^2} \xi_i(t) \eta_j(t) dt + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^h(tb)|^2}{|f^h(t)|^2} \eta_i(t) \xi_j(t) dt + o_p(C_{V,b}^{1/2})
\]

\[
\equiv \hat{T}_n + R_{1n}^* + R_{2n}^* + R_{3n}^* + o_p(C_{V,b}^{1/2}),
\]

where

\[
\eta_i(t) = \int e^{isW_i} \left\{ m^{\hat{f}}(t-s) - m^{\hat{f}}(t-s) \right\} \frac{K^h(sb)}{f^h(s)} ds f^{\hat{f}}(t)
\]

\[
= c_n \int e^{isW_i} \Delta^{\hat{f}}(t-s) \frac{K^h(sb)}{f^h(s)} ds f^{\hat{f}}(t),
\]

19
under $H_{1n}$. By Theorem 1 (i), it holds $C_{V,b}^{-1/2}\tilde{T}_n \overset{d}{\to} N\left(0, \frac{2}{(2\pi)^2}\right)$. For $R_{1n}^*$, observe that

$$E[C_{V,b}^{-1/2}R_{1n}^*] = \frac{(n-1)c_n^2}{(2\pi)^{d}C_{V,b}} \int \int |K^h(tb)|^2K^h(s_1b)K^h(s_2b)\Delta^\ast(t-s_1)\Delta^\ast(s_2-t)f^\ast(s_1)f^\ast(-s_2)ds_1ds_2dt$$

$$\equiv \Delta_n. \quad (14)$$

By the definition of $c_n$, $C_{V,b} = O(b^{-d-4\alpha})$ (obtained in the proof of Theorem 1 (i)), and Assumption D (ii), it holds $E[C_{V,b}^{-1/2}R_{1n}^*] = O(1)$ and the limit of $\Delta_n$ exists. Also, a similar argument to (3) yields

$$E[R_{1n}^2] = c_n^4 \int \int |K^h(t_1b)|^2|K^h(t_2b)|^2K^h(s_1b)K^h(s_2b)K^h(s_3b)K^h(s_4b)\frac{f_0^\ast(s_1)f_0^\ast(-s_2)f_0^\ast(s_3)f_0^\ast(-s_4)}{f_W^\ast(s_1+s_3)f_W^\ast(-s_2-s_4)}$$

$$\times \Delta^\ast(t_1-s_1)\Delta^\ast(s_2-t_1)\Delta^\ast(t_2-s_3)\Delta^\ast(s_4-t_2)ds_1 \cdots ds_4dt_1dt_2$$

$$\equiv O(b^{-d-4\alpha}).$$

Therefore, $\text{Var}(C_{V,b}^{-1/2}R_{1n}^*) \to 0$ and we obtain $C_{V,b}^{-1/2}R_{1n}^* \overset{p}{\to} \lim_{n \to \infty} \Delta_n$. Finally, using similar arguments combined with $E[\xi_1(t)] = 0$, we can show that $C_{V,b}^{-1/2}R_{2n}^* \overset{p}{\to} 0$ and $C_{V,b}^{-1/2}R_{3n}^* \overset{p}{\to} 0$. Combining these results, the conclusion follows.

A.2.2. Proof of (ii). Similarly to the proof of Part (i), we can decompose

$$T_n = \tilde{T}_n + R_{1n}^* + R_{2n}^* + R_{3n}^* + o_p(\varphi(b)^{-1}).$$

Theorem 1 (ii) implies the limiting distribution of $\varphi(b)\tilde{T}_n$. For $R_{1n}^*$, note that

$$E[\varphi(b)R_{1n}^*] = \varphi(b)(n-1)c_n^2 \int \int |K^h(tb)|^2K^h(s_1b)K^h(s_2b)$$

$$\times \Delta^\ast(t-s_1)\Delta^\ast(s_2-t)f^\ast(s_1)f^\ast(-s_2)ds_1ds_2dt$$

$$\equiv T_n. \quad (15)$$

and the limit of $T_n$ exists from the definition of $c_n$. Also, by similar treatment to $r_{1n}$ in the proof of Theorem 1 (ii), we can show $\text{Var}(\varphi(b)R_{1n}^*) \to 0$ and thus $\varphi(b)R_{1n}^* \overset{p}{\to} \lim_{n \to \infty} T_n$. Using similar arguments combined with $E[\xi_1(t)] = 0$, we can again show that $R_{2n}^*$ and $R_{3n}^*$ are asymptotically negligible. Therefore, the conclusion follows.

A.2.3. Proof of (iii). The proof is identical to that of Part (ii) with $\varphi(b)$ replaced by $\phi(b)$ and setting $c_n = b^{(\lambda-1)/2+\lambda\beta+\lambda_0e^{\lambda \beta}}$. 

20
REFERENCES


DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A 2AE, UK.
*E-mail address: t.otsu@lse.ac.uk*

DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A 2AE, UK.
*E-mail address: l.n.taylor@lse.ac.uk*