This is the accepted manuscript (post-print version) of the article.
The content in the accepted manuscript version is identical to the final published version, although typography and layout may differ.

How to cite this publication
Please cite the final published version:


Publication metadata

- **Title:** Specification Testing for Errors-in-Variables Models.
- **Author(s):** Otsu, Taisuke ; Taylor, Luke Nicholas.
- **Journal:** Econometric Theory.
- **DOI/Link:** 10.1017/S0266466620000262
- **Document version:** Accepted manuscript (post-print)
- **Document license:** CC-BY-NC-ND

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.
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 comparing the distance between the parametric and nonparametric fits based on deconvolution techniques. In contrast to the methods proposed by Hall and Ma (2007) and Song (2008), our test allows general nonlinear regression models and possesses complementary local power properties. We establish the asymptotic properties of our test statistic for the ordinary and supersmooth measurement error densities. Simulation results endorse our theoretical findings: our test has advantages in detecting high-frequency alternatives and dominates the existing tests under certain specifications.

1. INTRODUCTION

In this paper, we propose a specification, or goodness-of-fit test, for (possibly nonlinear) regression models with errors-in-variables by comparing the distance between the parametric and nonparametric fits based on deconvolution techniques. We establish asymptotic properties of the test statistic and use the bootstrap procedure of Hall and Ma (2007) to obtain critical values. Compared to existing methods, our test allows nonlinear regression models and possesses complementary local power properties.

In the enormous literature on specification testing, relatively little attention has been given to the issue of measurement error despite its obvious need. Papers such as Zhu, Song and Cui (2003), Zhu and Cui (2005), and Cheng and Kukush (2004) proposed \( \chi^2 \) statistics based on moment conditions of observables implied from errors-in-variables regression models. However, as is the case without measurement error, these tests are generally inconsistent for some fixed alternatives. Song (2008) proposed a consistent specification test for linear errors-in-variables regression models by comparing nonparametric and model-based estimators on the conditional mean function of the dependent variable \( Y \) given the mismeasured observable covariates \( W \), that is \( E[Y|W] \). As we clarify at the end of the next section, this approach may not have sensible local power for the original hypothesis on \( E[Y|X] \), where \( X \) is a vector of error-free unobservable covariates. Hall and Ma (2007) proposed a nonsmoothing specification test for polynomial regression models with errors-in-variables, which is able to detect local alternatives at the \( \sqrt{n} \)-rate.

The authors would like to thank anonymous referees and a co-editor for helpful comments, and acknowledge financial support from the ERC Consolidator Grant (SNP 615882) (Otsu) and the ESRC (Taylor).
We propose a smoothing specification test that complements Hall and Ma’s (2007) test (see further discussion below). Other papers that study specification testing under measurement error include Butucea (2007), Holzmann and Boysen (2006) and Holzmann, Bissantz and Munk (2007) (for testing probability densities), Koul and Song (2009, 2010) (for Berkson measurement error models), and Song (2009) and Xu and Zhu (2015) (for errors-in-variables models with validation data).

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 the 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 on 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, then they extended this to an omnibus test using a system of different basis functions. However, the test does not allow the number of basis functions to increase with the sample size and so is not strictly a nonparametric test. They also discuss that as a result of the way the test is constructed, it has low power against high-frequency alternatives - similar to the tests of Song (2008) and Hall and Ma (2007).

The rest of 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, outlines the bootstrap procedure used, and discusses how to implement the
test, including in the case where the distribution of the measurement error is unknown but repeated measurements are available. Finally, Section 4 analyses the small sample properties of the test through a Monte Carlo simulation. All mathematical proofs are deferred to the Appendix.

2. Setup and test statistic

Consider the nonparametric regression model

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

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 3.2. 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

\[ H_0 : \text{there exists some } \theta \in \mathbb{R}^k \text{ such that } m(x) = m_\theta(x) \text{ for almost every } x \in \mathbb{R}^d, \]

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

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

\[ K_b(a) = \frac{1}{(2\pi)^d} \int e^{-it \cdot a} \frac{K^R(t b)}{f^R_\epsilon(t)} dt, \]

is the so-called deconvolution kernel, \( i = \sqrt{-1}, b \) is a bandwidth, and \( K^R \) and \( f^R_\epsilon \) are the Fourier transforms of a kernel function \( K \) and the measurement error density \( f_\epsilon \), respectively.
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

$$\hat{m}(x) = \sum_{i=1}^{n} Y_i K(b(\frac{x}{X_1 - X_{1i}})) K(b(\frac{x}{W_i})), \quad K(b(a)) = \frac{1}{b^d} K(\frac{a}{b})$$

where $K$ is a conventional kernel function for $X_1$, and analogous results can be established.

Throughout the paper we assume $f_t(\tau) \neq 0$ for all $\tau \in \mathbb{R}$ and $K$ has compact support so that the above integral is well-defined.

For the parametric functional form, $m_\theta$, several methods are available to estimate $\theta$ under certain regularity conditions. For example, based on Butucea and Taupin (2008), we can estimate $\theta$ 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 assuming consistency.

In order to construct a test statistic for $H_0$, we compare the nonparametric and parametric estimators of the regression function based on the $L_2$-distance,

$$D_n = n \int \left| \hat{m}(x)\hat{f}(x) - m_\hat{\theta}(x)\hat{f}(x) \right|^2 dx \quad (1)$$

where $|\cdot|$ is the Euclidean norm, and $\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_b(x - W_i)$ is the deconvolution kernel density estimator for $X$.

We close this section with a remark on an alternative testing approach. To test the null hypothesis $H_0$, one may consider testing an alternative formulation of $H_0$ based on the conditional mean $E[Y|W]$ of observables, i.e., consider $H'_0 : f_W(w)E[Y|W = w] = \int m_\theta(w - u)f_X(w - u)f_\epsilon(u)du$ for almost every $w$, and test $H'_0$ by a conventional method, such as Härdle and Mammen (1993). This approach was employed by Song (2008), and we argue that our testing approach can be a useful complement to this method since it may have non-trivial local power against local alternatives which may not be detected by Song’s (2008) approach.

To illustrate this point, consider the following local alternative hypothesis for the regression function

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

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 $D_n$ 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 $\{(m_n - m_\theta)f_X\} \ast f_\epsilon$. By Parseval’s identity and the Fourier shift
formula, we have
\[
\|\{(m_n - m_{\theta})f_X\} \ast f_\epsilon\|^2 = a_n^2 \|q^\ast(\cdot - A_n) + q^\ast(\cdot + A_n)\|_{L^2}^2,
\]
where \(q(x) = \left(\frac{\sin x}{x}\right)f_X(x)\). For example, if \(f_\epsilon\) is Laplace with \(f_{\epsilon}^0(t) = 1/(1 + t^2)\), then we have that the \(L_2\)-norm \(\|\{(m_n - m_{\theta})f_X\} \ast f_\epsilon\|\) 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 \(H_0^f\) fails to detect deviations from this null.

We emphasize that our test should be treated as a complement to the existing approach by Song (2008). One may construct examples where Song’s (2008) approach yields better power properties. This point is illustrated in our simulation study in Section 4.

3. Asymptotic Properties

In this section, we present asymptotic properties of the test statistic \(D_n\). We first derive the limiting distribution 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 \(f_\epsilon\). Elements of \(\epsilon\) are mutually independent.

(ii): \(m_\theta\) is twice differentiable in a neighborhood of \(\theta\). Each of the following are in \(L_1\) and \(L_2\): \(m^2f_X, m_2f_X, \left(\frac{dm_2}{d\theta}\right)f_X\), and \(\frac{d^2m_\theta}{d\theta^2}f_X\), where \(m_2(x) = E[Y^2|X = x]\).

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

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

(v): \(\hat{\theta} - \theta = o_p(1)\) under \(H_0\).

Assumption D (i) is common in the literature of classical measurement error. Extensions to the case of unknown \(f_\epsilon\) are discussed in Section 3.2. Assumption D (ii) contains boundedness conditions on \(f_X\), the regression functions \(m\) and \(m_\theta\), and \(m_2\). 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 \(K(x) = \frac{\sin x}{\pi x}\) whose Fourier transform is equal to \(K^\theta(t) = \mathbb{1}\{-1 \leq t \leq 1\}\). Assumption (v) stipulates that we only require some consistent estimator of \(\theta\). When the regression model under the null hypothesis is linear (i.e., \(m_\theta(x) = x^T\theta\)), we can employ the methods in, for example, Gleser (1981), Bickel and
Ritov (1987), or van der Vaart (1988). For nonlinear regression, we may choose the estimators by e.g., Taupin (2001) or 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 \((\hat{\theta} - \theta)\) does not influence the first-order asymptotic properties of the test statistic \(D_n\). This is due to the measurement error slowing down the convergence rate of the dominant term of \(D_n\).

For simplicity, we use product kernels in the following way. Following Masry (1993), let \(\tilde{K}(\cdot)\) be a univariate kernel and \(\tilde{K}_b(\cdot)\) denote its Fourier transform. Define the univariate deconvolution kernel as

\[
\tilde{K}_b(x_j) = \frac{1}{2\pi b} \int e^{-ita} \tilde{K}_b(t) \tilde{f}_\epsilon(t/b) \, dt.
\]

where \(\tilde{f}_\epsilon(t)\) is the Fourier transform of \(\epsilon_j\). Finally, set \(K(x) = \prod_{j=1}^{\dim(x)} \tilde{K}(x_j)\) and \(K_b(x) = \prod_{j=1}^{\dim(x)} \tilde{K}_b(x_j)\). Since we assume that \(\epsilon\) is vector valued with independent elements, we can write \(f(t) = \prod_{j=1}^{\dim(t)} \tilde{f}_\epsilon(t_j)\). Combining these facts, we have \(K_b(x) = \prod_{j=1}^{\dim(x)} \tilde{K}_b(x_j)\).

We impose additional assumptions based on the bounds of the rate of decay of the tail of the characteristic function of the measurement error, \(f_\epsilon(t)\). Let \(\sigma^2(x) = E[U^2 | X = x]\) be the conditional variance of the error term. The first case, known as ordinary smooth measurement error, is characterised as follows.

**Assumption O.**

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

\[c|s|^{-\alpha} \leq |\tilde{f}_\epsilon(s)| \leq C|s|^{-\alpha},\]

for all \(1 \leq j \leq d\) as \(|s| \to \infty|s|\to \infty\).

Assumption O requires that the Fourier transform \(\tilde{f}_\epsilon(s)\) decays in some finite power. A popular example of an ordinary smooth density is the Laplace density. Although it is beyond the scope of this paper, isolated zeros in \(f_\epsilon(t)\) may be allowed by introducing the spectral method (Carrasco and Florens, 2011) or ridge approach (Hall and Meister, 2007).

For the second case, known as supersmooth measurement error, we impose the following assumptions.

**Assumption S.**

(i): \(f_\epsilon(t) \neq 0\) for all \(t \in \mathbb{R}^d\) and there exist positive constants \(C_\epsilon, \mu, \gamma_0,\) and \(\gamma > 1\) such that

\[\tilde{f}_\epsilon(s) \sim C_\epsilon|s|^{\gamma_0}e^{-|s|}/\mu,\]
as $|s| \to \infty$.

(ii): There exist constants $A > 0$ and $\beta \geq 0$ such that
\[ \hat{K}^\text{ft}(1 - s) = A s^\beta + o(s^\beta), \]

as $s \to 0$.


Assumption S (i) is adopted from Holzmann and Boysen (2006). This assumption requires that the Fourier transform $\hat{f}_\epsilon$ decays at an exponential rate. An example of a 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), the condition $\gamma > 1$ is imposed to make a bias term negligible.

Assumption S (ii) contains an additional condition on the kernel function. The sinc kernel, $K(x) = \frac{\sin \pi x}{\pi x}$, for example, satisfies this assumption with $A = 1$ and $\beta = 0$. Assumption S (iii) requires bounded fourth moments for our observable random variables.

Under these assumptions, the null distribution of $D_n$ is given as follows.

**Theorem 1.**

(i): Suppose that Assumptions D and O hold true. Then under $H_0$,
\[ C_{V,b}^{-1/2} D_n \xrightarrow{d} N \left( 0, \frac{2}{(2\pi)^{2d}} \right), \]

where $C_{V,b} = O(b^{-d(1+4\alpha)})$ is defined in (2) in the Appendix.

(ii): Suppose that Assumptions D and S hold true and $\epsilon$ is multivariate normal. Then under $H_0$,
\[ \varphi(b) D_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1), \]

where $\varphi(b) = \frac{(2\pi)^d}{b^{(1+4\beta)d/6^\gamma A^2 d^\Gamma(1+2d\beta)}}$ with the gamma function $\Gamma$, $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ is defined in (6) in the Appendix.

(iii): Suppose that Assumptions D and S hold true. Then under $H_0$,
\[ \phi(b) D_n \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1), \]

where $\phi(b) = \frac{(2\pi)^d \gamma_1 + 2d^{\beta_5} C_{2d} \Gamma(1+2d\beta)}{b^{1+2d\beta} (\gamma - 1+2\beta) \mu^2 (\mu^{\beta_7}) \Gamma^2(1+2d\beta)}$. 

7
Theorem 1 (i) says that for the ordinary smooth case, the test statistic $D_n$ is asymptotically normal. The normalizing term $C_{V,b}$ comes from the variance of the U-statistic of the leading term in $D_n$. Note that the convergence rate $C_{V,b}^{-1/2} = O(t^{d(\frac{1}{2}+2\alpha)})$ of the statistic $D_n$ is slower than the rate $O(t^{d/2})$ 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 $f_{\epsilon}$ increases, the convergence rate of $D_n$ becomes slower. This theorem can also be used to show the ‘rule-of-thumb’ rate for the bandwidth is of the order $b = n^{-\left(\frac{1}{4+d+2\alpha}\right)}$.

Theorem 1 (ii) focuses on the case of normal measurement error and shows that the test statistic converges to a 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 $\beta = 0$), the normalization becomes $\varphi(b) = \frac{(2\pi)^d}{\sqrt{e^{(1/2)^2}}} \Gamma(d)$. In this supersmooth case, the non-normal limiting distribution emerges because the leading term of the statistic $D_n$ is characterized by a 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 error. In this case, after normalization by $\varphi(b)$, 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 $\varphi(b)$ is characterized by the shapes of the kernel function and Fourier transform $f_{\epsilon}^R(t)$ of the measurement error specified in Assumption S (i).

In order to investigate the power properties of the test based on $T_n$, we consider a local alternative hypothesis of the form

$$H_{1n} : m(x) = m_0(x) + c_n \Delta(x), \text{ for almost every } x \in \mathbb{R}^d$$

where $c_n \to 0$ and $\Delta(x)$ is a non-zero function satisfying $\Delta(\cdot)^2 f_X(\cdot) \in L_2$ and such that the limit $\lim_{n \to \infty} \Delta_n$ defined in (7) in the Appendix exists. The local power properties are obtained as follows.

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

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

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

$$\varphi(b) D_n \xrightarrow{d} \lim_{n \to \infty} \Delta_n + \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1).$$

(iii): Suppose that Assumptions D and S hold true. Then under $H_{1n}$ with $c_n = b^{d((\lambda - 1)/2 + \lambda \beta + \lambda_0)} e^{d/\mu(b)}$,

$$\varphi(b) D_n \xrightarrow{d} \lim_{n \to \infty} \Delta_n + \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),$$

Theorem 2 (i) says that in the ordinary smooth case, our test has non-trivial power against local alternatives drifting with the rate of $c_n = n^{-1/2} b^{-d(1/4 + \alpha)}$. This is a nonparametric rate, with the test becoming less powerful as the dimension $d$ of $X$ or the decay rate $\alpha$ of $f_t$ increases. For the no measurement error case, Härdle and Mammen’s (1993) statistic has non-trivial power for local alternatives with the rate $n^{-1/2} b^{-d/4}$. Therefore, as expected, the test becomes less powerful in the presence of 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 $b \sim \left(\frac{d}{\log(n)}\right)^{1/2}$, then the rate for the local alternative will be $c_n \sim \left(\frac{d}{\log(n)}\right)^{d(1/4 + \beta)}$. So the rate at which we can detect local alternatives is typically a log rate in the supersmooth measurement error case.

3.1. Bootstrap and implementation. As is typically the case in testing problems which involve nonparametric components, the asymptotic distribution is relatively complex and can be difficult to estimate in practice. As such, we suggest a bootstrap procedure to calculate critical values for this test. We follow the approach of Hall and Ma (2007) who propose a type of moment matching bootstrap that is adapted to the setting of mismeasured regressors. Measurement error poses many problems for the implementation of bootstrap procedures, resulting in a relative scarcity of bootstraps in the literature. We do not have access to the true regressors, so cannot compute any residual based bootstrap, we also do not have access to the measurement errors which provides further issues. Hall and Ma (2007) suggest the following approach in answer to
these issues. Compute $\hat{m}$, $\hat{f}$, $m_\theta$, and $\hat{F}_X$ from the original data using standard deconvolution techniques and a parametric estimator for $\theta$. Estimate $\omega_2 = E[U^2]$ (Hall and Ma, 2007, provide a method to do this in their polynomial setting, for more general models, one can use the approach of Delaigle, Hall and Jamshidi, 2015) and choose a distribution, $G_r(\cdot|\hat{\omega}_2)$, which has first and second moments equal to 0 and $\hat{\omega}_2$, respectively (in Section 4 we use Mammen’s 2-point distribution for $G_r$). Now, draw data $\{X_i\}^n_{i=1}$ from $\hat{F}_X$, $\{e_i\}^n_{i=1}$ from $F_\epsilon(\cdot)$, and $\{U_i^2\}^n_{i=1}$ from $G_r(\cdot|\hat{\omega}_2)$. Construct $Y_i^* = m_\theta(X_i^*) + U_i^*$ and $W_i^* = X_i^* + e_i^*$. Using the pairs $(Y_i^*, W_i^*)$ compute $D_n^*$ in the same manner as $D_n$. The distribution of $D_n^*$ can be approximated by repeatedly carrying out this procedure.

In order to implement our test, we also need to choose a suitable bandwidth. Unfortunately, to the best of our knowledge, there is currently no theory on the optimal choice of bandwidth in a measurement error testing situation. It is beyond the scope of this paper to provide such a choice, however, we recommend the estimated plug-in approach of Delaigle and Gijbels (2004) as a good starting point; this is the approach we use in our Monte Carlo study. Although this procedure is designed to select the optimal bandwidth in a regression context - rather than for a testing problem - it provides good results in terms of power and size control (as shown in Section 4) and tends to produce greater power than other bandwidth choices around the selected value.

In summary, we suggest the following procedure for implementation of our specification test.

1. Based on the deconvolution regression and density estimators $\hat{m}$ and $\hat{f}$, and the parameter estimator $\hat{\theta}$, compute the test statistic $D_n$ in (1). The bandwidths $b$ in $\hat{m}$ and $\hat{f}$ are chosen as outlined in the 'two-stage selection plug-in' method of Delaigle and Gijbels (2004). That is, start with a normal-reference assumption to calculate $\int f_X^{(4)}(x)^2dx$ (where $f_X^{(4)}(x)$ denotes the 4th order derivative) which can be used to select an 'optimal' plug-in bandwidth for estimating $\int f_X^{(3)}(x)^2dx$ (required for the estimation of the asymptotic mean integrated squared error (AMISE) for $f_X^{(1)}$), in turn this bandwidth can be used to estimate $\int f_X^{(2)}(x)^2dx$ which is plugged into the AMISE for $f_X$. The final bandwidth is given as the minimiser of this estimated AMISE.

2. Compute the cdf $\hat{F}_X$ by integrating $\hat{f}$ obtained in Step 1, and compute $\hat{\omega}_2$ using the method of Delaigle, Hall and Jamshidi (2015), that is, $\hat{\omega}_2 = \max\{\hat{\mu}_2 - \hat{\xi}_2, 0\}$, where $\hat{\mu}_2 = \frac{1}{n} \sum^n_{i=1} Y_i^2$ and $\hat{\xi}_2 = \int \hat{m}(x)^2 \hat{f}(x)dx$.

3. Draw $\{X_i^*\}^n_{i=1}$ from $\hat{F}_X$, $\{e_i^*\}^n_{i=1}$ from $F_\epsilon$, and $\{U_i^*\}^n_{i=1}$ from $G_r(\cdot|\hat{\omega}_2)$ (some distribution with variance $\hat{\omega}_2$, such as Mammen’s 2 point distribution). Construct $Y_i^* = m_\theta(X_i^*) + U_i^*$.
and \( W_i^* = X_i^* + \epsilon_i^* \) for \( i = 1, \ldots, n \). Using the pairs \( \{Y_i^*, W_i^*\}_{i=1}^n \) compute the bootstrap counterpart \( D_n^* \) in the same manner as \( D_n \) in (1).

(4) Repeat Step 3 many times and use the \((1 - \alpha)\)-th quantile of the distribution of \( D_n^* \) as the critical value.

3.2. Case of unknown \( f_\epsilon \). In practical applications, it is unrealistic to assume that the density of the measurement error, \( f_\epsilon \), is known to the researcher. In the literature on nonparametric deconvolution, 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 \( D_n \) by inserting the estimated Fourier transform of the measurement error density, \( \hat{f}_\epsilon \).

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 \( f_\epsilon^{ft} \) is real-valued (i.e. \( f_\epsilon \) is symmetric around zero), then we can employ the estimator proposed by Delaigle, Hall and Meister (2008)

\[
\hat{f}_\epsilon^{ft}(t) = \left| \frac{1}{n} \sum_{i=1}^{n} \cos\{t(W_i - W_r^i)\} \right|^{1/2}.
\]

Delaigle, Hall and Meister (2008) studied the asymptotic properties of the deconvolution density estimator and the regression estimator using \( \hat{f}_\epsilon^{ft} \) and found conditions to guarantee that the differences between the estimators with known \( f_\epsilon \) and those with unknown \( f_\epsilon \) are asymptotically negligible. Under similar conditions, we can expect that the asymptotic distributions of the test statistic \( D_n \) obtained above remain unchanged when we replace \( f_\epsilon^{ft} \) with \( \hat{f}_\epsilon^{ft} \). If the researcher wishes to remove the assumption that \( f_\epsilon^{ft} \) is real-valued and \( \epsilon \) and \( \epsilon^r \) are identically distributed, 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. We also compare our test to Hall and Ma (2007) and Song (2008). Recall that although Hall and Ma’s (2007) test is confined to a polynomial regression model and Song’s (2008) test places restrictions on the type of model that can be tested, our test allows any nonlinear model. We take the true unobservable regressor \( \{X_i\}_{i=1}^n \) to be distributed as \( U[-1, 1] \) and \( Y_i = X_i + 0.5 \sin(\delta X_i) + U_i \), where \( U_i \sim N(0, 1/4) \) and \( \delta \) 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 \( 1/12 \). For the supersmooth case, we use \( N(0, 1/12) \). So the signal-to-noise ratio is \( \sigma_X^2/\sigma_\epsilon^2 = 4 \) in both cases. We use the following kernel for all of 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 small (250) and a medium (500) sample size. As discussed in Section 3.1, we use the estimated plug-in approach of Delaigle and Gijbels (2004) to select our bandwidth. For the moment matching bootstrap, we use Mammen’s 2-point distribution for \( G(\cdot|\omega_2) \).

For the parametric estimator, we use the polynomial estimator of degree 1 proposed by Cheng and Schneeweiss (1998) so as to remain consistent with the experiment conducted by Hall and Ma (2007). For the test of Song (2008) we use the same kernel as for our test and also choose bandwidths by cross-validation. All results are based on 1000 Monte Carlo replications.

Table 1 shows the results for the power and the size of three tests. The column labelled \( D_n \) refers to the test proposed in this paper, the column labelled 'S' corresponds to the test put forward in Song (2008), and the column labelled ‘HM’ corresponds to the test proposed by Hall and Ma (2007). When \( \delta = 0 \) we have the null model. We also consider \( \delta = \{1, 2, 3\} \) corresponding to departures from the null of varying frequencies which we term low, medium, and high frequency alternatives, respectively.

The results are encouraging and seem to be consistent with the theory. The first column \( (\delta = 0) \) indicates that all three tests track the nominal level closely. This indicates that the bootstrap procedure discussed in Section 3.1 works well in practice.

As we expected, across all parameter settings, our test performs more poorly for supersmooth measurement error in comparison to ordinary smooth error. This reflects the slower convergence rates derived in Section 3. Somewhat surprisingly, the test of Hall and Ma (2007) lags behind the other two tests in all cases, although it has the benefit of not requiring a tuning parameter.

As conjectured, in the low frequency setting \( (\delta = 1) \) our test is generally less powerful than the test of Song (2008). Although, in the ordinary smooth case for a sample size of 250, our test displays higher power. This reflects that the test of Song (2008) is able to detect local alternatives at the rate \( \sqrt{nb^{d/2}} \) for both ordinary and supersmooth measurement error distributions. However, our test achieves a slightly slower polynomial rate in the ordinary smooth case and only a \( \log(n) \)-rate in the super smooth case. Thus it is not surprising to see our test underperform when faced with Gaussian measurement error.
On the other hand, as previously mentioned, we suspect that our test is better suited to detecting high-frequency alternatives than Song (2008). This is confirmed in the final column of Table 1 ($\delta = 3$). We find that the test of Song (2008) and Hall and Ma (2007) have trivial power in this setting for both sample sizes, while our test still enjoys reasonable power which increases with the sample size.

The dominance over the test of Hall and Ma (2007) in this high-frequency setting 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 allow for the superior detection of high-frequency alternatives over nonsmoothing tests. Furthermore, as discussed at the end of Section 2, the test of Song (2008) will have poor power properties for some high-frequency alternatives due to testing the hypothesis based on $E[Y|W]$ rather than $E[Y|X]$.

Table 1: $Y = X + 0.5 \sin(\delta X) + U$

<table>
<thead>
<tr>
<th>Ordinary Smooth</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>Level</td>
</tr>
<tr>
<td>250</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>500</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>250</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>500</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>10%</td>
</tr>
</tbody>
</table>
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).

We start by decomposing \( D_n \) as

\[
D_n = \frac{1}{n(2\pi)^d} \sum_{i=1}^{n} |\zeta_i(x)|^2 dx + \frac{1}{n(2\pi)^d} \sum_{i \neq j} |\zeta_i(x)\overline{\zeta_j(x)}| dx = B_n + T_n,
\]

where \( \zeta_i(x) = \int \frac{K_n(ab)}{f^D(t)} \{Y_i - m_\theta(x)\} e^{-it(x-W_i)} dt \) and \( \overline{\zeta_j(x)} \) is the complex conjugate of \( \zeta_j(x) \).

We will show that \( B_n \) is of smaller asymptotic order than \( T_n \), as such, \( T_n \) will characterise the asymptotic distribution of the test.

First, we define the normalization term \( C_{V,b} \) and characterize its asymptotic order. Let

\[
\xi_i(x) = \int \frac{K_n}{f^D(t)} \{Y_i - m_\theta(x)\} e^{-it(x-W_i)} dt, \quad H_{i,j} = \int \zeta_i(x)\overline{\zeta_j(x)} dx.
\]

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

\[
C_{V,b} = E[H_{1,2}^2] = \int \int |E[\xi_i(x)\xi_j(z)]|^2 dx dz = Q_1 + Q_2 - 2Q_3,
\]

where

\[
Q_1 = \int \int \int \int E \left[ Y_i^2 e^{-it(x-W_i)} e^{-is(z-W_i)} \right] \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds dx dz,
Q_2 = \int \int \int \int E \left[ e^{-it(x-W_i)} e^{-is(z-W_i)} m_\theta(x)m_\theta(z) \right] \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds dx dz,
Q_3 = \int \int \int \int E \left[ Y_i e^{-it(x-W_i)} e^{-is(z-W_i)} \right] m_\theta(z) \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds dx dz.
\]

For \( Q_1 \), we have

\[
Q_1 = \int \int \int \int f^D(t+s) m_2(v)f_X(v) e^{-it(x-v)} e^{-is(z-v)} dv \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds dx dz
= \int \int f^D(t+s) m_2(v)f_X(v)e^{it(s+v)} dv \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds
= \int \int [m_2f_X]^D(t+s)f^D(t+s) \frac{K_n}{f^D(t)} \frac{K_n}{f^D(s)} dt ds
= b^{-d} \int \int [m_2f_X]^D(a)f^D(a) \frac{K_n}{f^D(a-r/b)} \frac{K_n}{f^D(r/b)} dr da
= O(b^{-d(1+4a)}),
\]
where we have used Parseval’s identity in the second equality, the change of variables \((t + s = a, s = r/b)\) in the penultimate equality, and for the final equality we have used Assumption D (ii) \((f_Xm_2 \in L_2)\) which implies \(\sup_a |[m_2 f_X]|^2 < \infty\), Assumption D (iii) (compactness of \(K^\theta\)), Assumption O (i) (rate of decay of \(f^\theta\)), and that a characteristic function is bounded by 1.

For \(Q_2\), we can use similar arguments as for \(Q_1\) to write

\[
Q_2 \leq \int \int \int \int f^\theta(t + s) e^{-it(x-v)} e^{-is(z-v)} f_X(v)m_\theta(x)m_\theta(z) \frac{K^\theta(tb)}{f^\theta(t)} \frac{K^\theta(sb)}{f^\theta(s)} dt ds dv dx dz
\]

\[
= b^{2d} \int \int \int \int f^\theta(t + s) e^{-it ub} e^{-is rb} f_X(v)m_\theta(v + ub)m_\theta(v + rb) \frac{K^\theta(tb)}{f^\theta(t)} \frac{K^\theta(sb)}{f^\theta(s)} dt ds dv dx dz
\]

\[
= b^{-2d} \int \int \int \int f^\theta(t + s) e^{-it u} e^{-is r} f_X(v)m_\theta(v + ub)m_\theta(v + rb) \frac{K^\theta(t)}{f^\theta(t / b)} \frac{K^\theta(s)}{f^\theta(s / b)} dt ds dv dx dz
\]

\[
\approx b^{-2d} \int \int \int \int f^\theta(t + s) e^{-it u} e^{-is r} f_X(v)m_\theta(v)m_\theta(v) \frac{K^\theta(t)}{f^\theta(t / b)} \frac{K^\theta(s)}{f^\theta(s / b)} dt ds dv dx dz
\]

\[
= b^{-2d} \int |f_X(v)m_\theta(v)m_\theta(v)|^2 dv \int \int |f^\theta(t + s) - f^\theta(t / b) - f^\theta(s / b)|^2 dt ds
\]

\[
= b^{-2d} \int ||f_Xm_\theta^2(\zeta)|^2 d\zeta \int \int |f^\theta(t + s) - f^\theta(t / b) - f^\theta(s / b)|^2 dt ds
\]

\[
= O(b^{-d(1+4\alpha)}),
\]

where we have used the change of variables \((x = ub + v, z = rb + v)\) in the first equality, the wave relation follows from two Taylor expansions in each \(m_\theta\) around \(v\), Parseval’s theorem is used in the third and fourth equalities, and the same arguments as used for bounding \(Q_1\) are used in the final step.

We can similarly show \(Q_3 = O(b^{-d(1+4\alpha)})\) and conclude \(C_{V,b} = O(b^{-d(1+4\alpha)})\).

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

\[
\rho_i(x) = \int \frac{K^\theta(tb)}{f^\theta(t)} \{m_\theta(x) - m_\theta(x)\} e^{-it(x-W_i)} dt.
\]

Then \(T_n\) is written as

\[
T_n = \frac{1}{n} \sum_{i \neq j} \int \xi_i(x) \xi_j(x) dx + \frac{1}{n} \sum_{i \neq j} \int \rho_i(x) \rho_j(x) dx
\]

\[
+ \frac{1}{n} \sum_{i \neq j} \int \rho_i(x) \xi_j(x) dx + \frac{1}{n} \sum_{i \neq j} \int \xi_i(x) \rho_j(x) dx
\]

\[
\equiv \bar{T}_n + R_1n + R_2n + R_3n.
\]
By an expansion around $\hat{\theta} = \theta$ and Assumption O (iii) ($\hat{\theta} - \theta = o_p(1)$), the term $R_{1n}$ satisfies

$$R_{1n} = o_p(1) \left| \frac{1}{n} \sum_{i \neq j} \rho_{1i}(x)\rho_{1j}(x) dx \right|,$$

where $\rho_{1i}(x) = \int \frac{\partial m_{0}(x)}{\partial \theta} \frac{K_{\theta}(tb)}{f_{X}(t)} e^{-it(x-W_{i})} dt$. Using similar arguments as those used to bound $Q_2$ together with Assumption D (ii) ($\frac{\partial m_{0}}{\partial \theta} f_{X} \in L_{2}$), we have

$$E \left[ \int \rho_{1i}(x)\rho_{1j}(x) dx \right] \leq \int \int \left| \int \frac{\partial m_{0}(v + ub)}{\partial \theta} f_{X}(v)K_{\theta}(t) e^{-it-u} dt \right|^2 dvdu = b^{-d} \int \int \left| \int \frac{\partial m_{0}(v)}{\partial \theta} f_{X}(v)K_{\theta}(t) e^{-it} dt \right|^2 dvdu \sim b^{-d} \int \int \left| \frac{\partial m_{0}}{\partial \theta} f_{X}(t) \right|^2 dt dvdu = b^{-d} \left[ \frac{\partial m_{0}}{\partial \theta} f_{X} \right]^2 \int d\zeta \left| K_{\theta}(t) \right|^2 dt = O(1).$$

Also, by applying the same argument as used to derive the bound of $C_{V,b}$, we have

$$E \left[ \left( \int \rho_{1i}(x)\rho_{1j}(x) dx \right)^2 \right] = O(b^{-d(1+4\alpha)}).$$

Combining these two results, along with $\hat{\theta} - \theta = o_p(1)$ and $C_{V,b} = O(b^{-d(1+4\alpha)})$, 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 $C_{V,b}^{-1/2} R_{3n} = o_p(1)$ and thus $C_{V,b}^{-1/2} T_{n} = C_{V,b}^{-1/2} \tilde{T}_{n} + o_p(1)$.

Thirdly, we derive the limiting distribution of $C_{V,b}^{-1/2} \tilde{T}_{n}$. Note that $\tilde{T}_{n}$ is written as $\tilde{T}_{n} = \frac{2}{n(2\pi)^{d}} \sum_{i<j} H_{i,j}$ and is a U-statistic with zero mean under $H_{0}$. To prove the asymptotic normality of $\tilde{T}_{n}$, we apply the central limit theorem in Hall (1984, Theorem 1). To this end, it is enough to show

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

where $G_{i,j} = E[H_{i,j}H_{i,j}|Y_{1},W_{1}]$. Recall that $C_{V,b} = E[H_{i,j}^{2}]$ satisfies $C_{V,b} = O(b^{-d(1+4\alpha)})$. By a similar arguments used to bound $E[H_{i,j}^{2}]$, we can show

$$E[H_{i,j}^{4}] = E \left[ \prod_{k=1}^{4} \xi_{1}(x_{k})\xi_{2}(x_{k})dx_{1} \cdots dx_{4} \right] = O(b^{-3d(1+8\alpha)}).$$
For $E[C_{1,2}^2]$, we can equivalently look at (see, p. 5 of Hall, 1984)

$$
E[H_{1,3}H_{1,4}H_{2,3}H_{2,4}]
= \int \cdots \int \prod_{k=1}^{4} \xi_1(x_1)\xi_3(x_1)\xi_1(x_2)\xi_3(x_2)\xi_2(x_3)\xi_2(x_4)\xi_4(x_4)dx_1 \cdots dx_4
= O(b^{-d(1+8\alpha)}).
$$

These results combined with Assumption D (iv) guarantee the conditions in (4). Thus, Hall (1984, Theorem 1) implies

$$
C_{V,b}^{-1/2}T_n \overset{d}{\to} N\left(0, \frac{4}{(2\pi)^{2d}} \right).
$$

Finally, the $B_n$ term is simply a sample mean of independent and identically distributed random variables. The second moment can be calculated in the same manner as $T_n$ to give $C_{V,b}^{-1/2}B_n = O_p(n^{-1})$. Combining these results, the conclusion follows.

A.1.2. Proof of (ii). A similar argument to the proof of Part (i) guarantees $\varphi(b)T_n = \varphi(b)\tilde{T}_n + o_p(1)$, where $\varphi(b) = \frac{(2\pi)^d}{\gamma^{(2+4\beta)}e^{d/b^2}A^{2d/2+2d\beta}}$. Thus, we hereafter derive the limiting distribution of $\tilde{T}_n$. Decompose $\tilde{T}_n = \tilde{T}_n + r_{1n} - 2r_{2n}$, where

$$
\tilde{T}_n = \frac{1}{n} \sum_{i \neq j} \int \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} Y_i Y_j e^{-it(x-W_i)}e^{-is(x-W_j)} dt ds dx
$$

$$
r_{1n} = \frac{1}{n} \sum_{i \neq j} \int \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} \frac{K^{\text{f}(s)}(sb)}{f^{\text{f}(s)}(s)} m_\theta(x)^2 e^{-it(x-W_i)}e^{-is(x-W_j)} dt ds dx
$$

$$
r_{2n} = \frac{1}{n} \sum_{i \neq j} \int \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} \frac{K^{\text{f}(s)}(sb)}{f^{\text{f}(s)}(s)} Y_i m_\theta(x) e^{-it(x-W_i)}e^{-is(x-W_j)} dt ds dx
$$

First, we derive the limiting distribution of $\tilde{T}_n$. Observe that

$$
\tilde{T}_n = \frac{1}{n} \sum_{i \neq j} \left( \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} Y_i e^{-itW_i} dt \right) \left( \int \frac{K^{\text{f}(s)}(sb)}{f^{\text{f}(s)}(s)} Y_j e^{-isW_j} ds \right)
$$

$$
= \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^{2d}} \int \left| \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} \right| \left| \frac{K^{\text{f}(s)}(sb)}{f^{\text{f}(s)}(s)} \right| Y_i Y_j \{ \cos(tW_i) \cos(sW_j) + \sin(tW_i) \sin(sW_j) \} dt ds
$$

$$
= \frac{1}{(2\pi)^{2d}} \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} dt \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\}
$$

$$
+ O_p \left( b^{d(2+4\beta)}e^{d/b^2} \right)
$$

$$
= \left( \frac{1}{(2\pi)^{2d}} \int \frac{K^{\text{f}(t)}(tb)}{f^{\text{f}(t)}(t)} dt \right)^2 \tilde{T}_n + O_p \left( b^{d(2+4\beta)}e^{d/b^2} \right)
$$
where the first equality follows from Plancheral’s theorem, the second equality follows from $f^b_t(t) = e^{-b^2 t/2}$ and $e^{itW} = \cos(tW) + is(tW)$, the fourth equality follows from the change of variables $(t = v/b, s = t + r)$ and from a simple multivariate extension of Holzmann and Boysen (2006, Theorem 1) based on Assumption S (ii). Note that
\[
\frac{\dot{V}_n}{n} = \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{X_i}{b} \right) \cos \left( \frac{Y_i}{b} \right) - \sin \left( \frac{X_i}{b} \right) \sin \left( \frac{Y_i}{b} \right) \right\} \left\{ \cos \left( \frac{X_j}{b} \right) \cos \left( \frac{Y_j}{b} \right) - \sin \left( \frac{X_j}{b} \right) \sin \left( \frac{Y_j}{b} \right) \right\}.
\]

From van Es and Uh (2005, proof of Lemma 6), it holds $\left( \frac{X_i}{b} \mod 2\pi \right) \xrightarrow{d} V_i^X \sim U[0,2\pi]$ and $\left( \frac{Y_i}{b} \mod 2\pi \right) \xrightarrow{d} V_i^Y \sim U[0,2\pi]$ as $b \to 0$ for each $i$, where $V_i^X$ is independent from $(Y_i, V_i^X)$. Thus by applying Holzmann and Boysen (2006, Lemma 1), $\frac{\dot{V}_n}{n}$ has the same limiting distribution as $\frac{\dot{V}_n}{n} = \frac{1}{n} \sum_{i \neq j} h(Q_i, Q_j)$, where $Q_i = (Y_i, V_i^X, V_i^Y)$ and
\[
h(Q_i, Q_j) = Y_i Y_j \left\{ \cos(V_i^X) \cos(V_j^Y) - \sin(V_i^X) \sin(V_j^Y) \right\} \left\{ \cos(V_j^X) \cos(V_i^Y) - \sin(V_j^X) \sin(V_i^Y) \right\}.
\]

Observe that Cov $\left(h(Q_1, Q_2), h(Q_1, Q_3)\right) = 0$ because $E[\cos(V_i^X)] = E[\sin(V_i^Y)] = 0$. Therefore, by applying the limit theorem for degenerate U-statistics with a fixed kernel $h$ (Serfling, 1980, Theorem 5.5.2), we obtain
\[
\frac{\dot{V}_n}{n} \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1),
\]
where $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ are the eigenvalues of the integral operator
\[
(\Lambda g)(Q_1) = \lambda g(Q_1).
\]

where $\langle g \rangle(Q_1) = E[h(Q_1, Q_2)g(Q_2)|Q_1]$. Also, we have
\[
\left| \int \frac{K^t(b \phi)}{\frac{f^b(t)}{f^b(t/b)}} dt \right|^2 \leq \int \left| \frac{K^t(b \phi)}{\frac{f^b(t)}{f^b(t/b)}} \right|^2 dt = \frac{1}{b^2} \int \left| \frac{K^t(t)}{f^b(t/b)} \right|^2 dt = \frac{1}{b^2} \int \left| K^t(t) e^{b \phi(t)} \right|^2 dt \sim \frac{1}{\varphi(b)},
\]
where the inequality follows from Cauchy’s inequality, the second step follows from a simple change of variables, the third step follows from the definition of the characteristic function of the Gaussian distribution, and the final wave relation comes from van Es and Uh (2005, Lemma 5).

Combining these results gives
\[
\varphi(b) \frac{\dot{V}_n}{n} \xrightarrow{d} \sum_{k=1}^{\infty} \lambda_k (Z_k^2 - 1).
\]
Next, we show that \( r_{1n} \) is negligible. Notice that \( E[r_{1n}^2] \) is equal to \( Q_2 \) from the proof of Theorem 1 (i), hence we can write

\[
E[r_{1n}^2] = b^{-2d} \int \left| |f_X m_0^{\gamma_1}(\zeta)|^2 \right| d\zeta \int \left| f^{\gamma_0}_b(t) \frac{K^\theta(t)}{f^{\gamma_0}_b(t/b)} \right|^2 dt ds
\]

\[
= b^{-d} \int \left| |f_X m_0^{\gamma_1}(\zeta)|^2 \right| d\zeta \int \left| f^{\gamma_0}_b(v) \frac{K^\theta(vb-s)}{f^{\gamma_0}_b(v-s/b)} \right|^2 dv ds
\]

\[
\sim b^{-d} \int \left| |f_X m_0^{\gamma_1}(\zeta)|^2 \right| d\zeta \int \left| f^{\gamma_0}_b(v) \right|^2 dv \int \left| \frac{K^\theta(s)}{f^{\gamma_0}_b(s/b)} \right|^2 ds
\]

\[
= O(b^{-d}) \int \left| \frac{K^\theta(s)}{f^{\gamma_0}_b(s/b)} \right|^4 ds \sim \frac{1}{\varphi(2b)},
\]

where the second equality follows from a change of variables \( t = vb - s \), the penultimate equality uses Assumption D (ii), and the final relation uses van Es and Uh (2005, Lemma 5) again. The first moment of \( r_{1n} \) is bounded in a similar manner.

By the same argument used to show the asymptotic behaviour of \( \bar{T}_n \), it holds

\[
\frac{1}{n} \sum_{i \neq j} \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\} = O_p(1).
\]

Combining these results, we have

\[
\varphi(b)r_{1n} = O_p(t_d^{(1+2\beta)}),
\]

and thus \( r_{1n} \) is negligible. Similar arguments imply that the terms \( r_{2n} \) and \( r_{3n} \) are also asymptotically negligible, and 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 \( \bar{T}_n \) defined in (5) satisfies

\[
\bar{T}_n \sim \frac{b^{2d(\gamma_0 - 1)}}{(2\pi)^d C_\epsilon^{2d}} \int |K^\theta(t)||K^\theta(s)||t|^{-\gamma_0}|s|^{-\gamma_0} e^{\frac{d}{\varphi(\gamma)}(||s|| + |s|^\gamma)}
\]

\[
\times \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{tW_i}{b} \right) \cos \left( \frac{sW_j}{b} \right) + \sin \left( \frac{tW_i}{b} \right) \sin \left( \frac{sW_j}{b} \right) \right\} dt ds
\]
We now show that

\[
D_{\cos} = \frac{b^{4d}\gamma_{0}^{-1}}{(2\pi)^d C_{d}^{2d}} \int |K^f(t)||K^f(s)||t|^{-d_{0}|s|^{-d_{0}} e^{\frac{d}{\mu_{b}+\gamma} (|t|+|s|^{\gamma})} \\
\times \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{t W_i}{b} \right) \cos \left( \frac{s W_j}{b} \right) - \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) \right\} dt ds
\]

is asymptotically negligible, as well as the correspondingly defined \(D_{\sin}\). We have seen in the proof of part (ii) that each term is zero mean. Following the proof of Holzmann and Boysen (2006, Theorem 1), we obtain

\[
\left| \cos \left( \frac{t W_i}{b} \right) \cos \left( \frac{s W_j}{b} \right) - \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) \right| \leq \frac{(1-t)|W_i|}{b} + \frac{(1-s)|W_k|}{b}.
\]

Similar arguments to van Es and Uh (2005, Lemmas 1 and 5) and Assumption S (ii), imply

\[
\text{Var}(D_{\cos}) = O(n^{-2}b^{4d_{0}-6d}) \sum_{i \neq j} E \left[ |Y_i||Y_j| \int |K^f(t)||K^f(s)||t|^{-d_{0}|s|^{-d_{0}} e^{\frac{d}{\mu_{b}+\gamma} (|t|+|s|^{\gamma})} \\
\times \{(1-t)|W_i| + (1-s)|W_k|\} dt ds \right]^2
\]

\[
= O(b^{4d_{0}-6d}) \left[ \int (1-t)|K^f(t)||K^f(s)||t|^{-d_{0}|s|^{-d_{0}} e^{\frac{d}{\mu_{b}+\gamma} (|t|+|s|^{\gamma})} dt ds \right]^2
\]

\[
= O(b^{4d_{0}-1}) \left[ \int (1-t)|K^f(t)||K^f(t + vb)||t|^{-d_{0}|t + vb|^{-d_{0}} e^{\frac{d}{\mu_{b}+\gamma} (|t|+|t + vb|^{\gamma})} dt dv \right]^2
\]

\[
\sim O(b^{4d_{0}-1}) \left[ \int (1-t)|K^f(t)|^2|t|^{-2d_{0}} e^{2d_{0} \frac{\gamma}{\mu_{b}+\gamma}} dt \right]^2
\]

\[
= O(b^{4d_{0}-1}) \left( b^{2}\gamma \int |K^f(1 - b^\gamma v)|^2 |1 - b^\gamma v|^{-2d_{0}} e^{2\frac{\gamma^2}{\mu_{b}+\gamma}} dv \right)^2
\]

\[
= O \left( b^{4d_{0}-1} \left( b^{2}\gamma^{(2+2\beta)} e^{\frac{2\gamma}{\mu_{b}+\gamma}} \right)^2 \right),
\]

where the first two steps follow from standard arguments, the third step follows from the change of variable \(s = vb + t\), the fourth step results from the definition of the wave relation, the fifth step uses the change of variable \(t = 1 - b^\gamma v\), and the final step uses van Es and Uh (2005, Lemma 5). Using similar arguments as in the proof of Theorem 1 (ii), we obtain \(D_{\cos} = \)
The same argument applies to $D_{\sin}$. Note that

$$\bar{T}_n = \frac{b^{d(\gamma-1)}}{(2\pi)^d C_{2d}^2} \int |K_{\mathbf{t}}(t)||K_{\mathbf{s}}(s)||t|^{-d\gamma_0} |s|^{-d\gamma_0} e^{\frac{2d}{\sqrt{\pi}}(|t|^{\gamma} + |s|^{\gamma})}$$

$$\times \frac{1}{n} \sum_{i \neq j} Y_i Y_j \left\{ \cos \left( \frac{W_i}{b} \right) \cos \left( \frac{W_j}{b} \right) + \sin \left( \frac{W_i}{b} \right) \sin \left( \frac{W_j}{b} \right) \right\} \, dt \, ds$$

$$+ O \left( b^{d(2(\gamma-1)+2\gamma\beta+2\gamma_0)} e^{\frac{2d}{\sqrt{\pi}}} \right)$$

$$= \frac{A^{2d} \mu_1^{1+2d} \mu_2^{d(\gamma-1)+2\gamma\beta+2\gamma_0} e^{\frac{2d}{\sqrt{\pi}}}}{\lambda^{1+2d} (2\pi)^d C_{2d}^2} \bar{T}_n + O \left( b^{d(2(\gamma-1)+2\gamma\beta+2\gamma_0)} e^{\frac{2d}{\sqrt{\pi}}} \right)$$

$$\equiv \frac{\bar{T}_n}{\varphi(b)} + O \left( b^{d(2(\gamma-1)+2\gamma\beta+2\gamma_0)} e^{\frac{2d}{\sqrt{\pi}}} \right),$$

where the second equality follows from the definition of $\bar{T}_n$ in (3) and a modification of van Es and Uh (2005, Lemma 5). Therefore, we obtain

$$\varphi(b) T_n = \bar{T}_n + O(b^{d(\gamma-1)}).$$

The limiting distribution of $\bar{T}_n$ - and hence $D_n$ - is obtained in the proof of Theorem 1 (ii). The remainder term becomes negligible if we impose $\gamma > 1$. 

---

21
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

\[
T_n = \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \xi_i(x)\xi_j(x)dx + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \eta_i(x)\eta_j(x)dx
\]

\[
+ \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \xi_i(x)\eta_j(x)dx + \frac{1}{n} \sum_{i \neq j} \frac{1}{(2\pi)^d} \int \eta_i(x)\xi_j(x)dx + o_p(C_{V,b}^{1/2})
\]

\[
\equiv T_n + R_1^* + R_2^* + R_3^* + o_p(C_{V,b}^{1/2}),
\]

where

\[
\eta_i(x) = \int \frac{K^\hat{\imath}(tb)}{f^\hat{\imath}(t)} \{m(x) - m_\theta(x)\} e^{-it(x-W_i)} dt
\]

\[
= c_n \int \frac{K^\hat{\imath}(tb)}{f^\hat{\imath}(t)} \Delta(x)e^{-it(x-W_i)} dt,
\]

under \( H_{in} \). By Theorem 1 (i), we’ve seen \( C_{V,b}^{-1/2} \tilde{T}_n \Rightarrow N\left(0, \frac{2}{(2\pi)^d}\right) \). For \( R_1^* \), observe that \( \eta_i(x) \) takes the same form as \( \rho_{1i}(x) \) in the proof of 1 (i), but with \( \frac{\partial m(x)}{\partial t} \) replaced by \( \Delta(x) \) and multiplied by \( c_n \). As such, we can reuse the same arguments as used in that proof. First,

\[
E[C_{V,b}^{-1/2} R_1^*] = E \left[ \frac{c_n^2}{(2\pi)^d C_{V,b}^{2/3}} \int \int \frac{K^\hat{\imath}(tb)}{f^\hat{\imath}(t)} \Delta(x)e^{-it(x-W_i)} dt \int \frac{K^\hat{\imath}(sb)}{f^\hat{\imath}(s)} \Delta(x)e^{-is(x-W_j)} ds dx \right]
\]

\[
= E \left[ \int \frac{c_n^2}{(2\pi)^d C_{V,b}^{2/3}} \int \frac{K^\hat{\imath}(tb)}{f^\hat{\imath}(t)} e^{-itW_i} dt \int \frac{K^\hat{\imath}(sb)}{f^\hat{\imath}(s)} e^{-isW_j} ds \right]
\]

\[
= \int \Delta^2(x) dx \frac{(n-1)c_n^2}{(2\pi)^d C_{V,b}^{2/3}} \int K^\hat{\imath}(tb) f^\hat{\imath}(t) dt \right|^2 \equiv \Delta_n. \tag{7}
\]

By the definition of \( c_n \), the fact that \( C_{V,b} = O(b^{-d(1+4\alpha)}) \), and Assumption D (ii), it holds that \( E[C_{V,b}^{-1/2} R_1^*] = O(1) \) and the limit of \( \Delta_n \) exists. Again, using a similar argument used to bound the variance of \( C_{V,b} \), we obtain

\[
E[R_1^*] = O(b^{-d(1+4\alpha)})c_n^4 = O(n^{-2}b^{-2d(1+4\alpha)})
\]

Therefore, \( \text{Var}(C_{V,b}^{-1/2} R_1^*) \rightarrow 0 \) and we obtain \( C_{V,b}^{-1/2} R_1^* \overset{p}{\rightarrow} \lim_{n \rightarrow \infty} \Delta_n \). Finally, using similar arguments, combined with \( E[\xi_i(t)] = 0 \), we can show that \( C_{V,b}^{-1/2} R_2^* \overset{p}{\rightarrow} 0 \) and \( C_{V,b}^{-1/2} R_3^* \overset{p}{\rightarrow} 0 \).

Taken together, the conclusion follows.
A.2.2. **Proof of (ii).** Similar to the proof of Theorem 2 part (i), we can decompose

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

Theorem 1 (ii) gives the limiting distribution of \( \varphi(b)\hat{T}_n \). For \( R_{1n}^* \), we have the same result as for 2 part (i) and the limit of \( \Delta_n \) exists from the definition of \( c_n \) and the bound of \( C_{V,b} \) in the supersmooth case. The conclusion follows.

A.2.3. **Proof of (iii).** The proof is identical to that of Part (ii) after setting \( c_n = b^{d\{\lambda - 1/2 + \lambda \beta + \lambda_0 \}} e^{\mu b \beta}. \)

Each of the objects in the asymptotic distribution are defined analogously to 2 part (ii).
REFERENCES


24


DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A 2AE, UK.

E-mail address: t.otsumlse.ac.uk

DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS, FUGLESTANGS ALLE 4 BUILDING 2631, 12 8210 AARHUS V, DENMARK

E-mail address: lintaylor@econ.au.dk