# 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 method proposed 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. The other existing method, by Song (2008), is shown to posses trivial power under certain alternatives. We establish the asymptotic properties of our test statistic for the ordinary and supersmooth measurement error densities and develop a bootstrap method to approximate the critical value. We apply the test to the specification of Engel curves in the US. Finally, some 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

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. If we are able to confirm that a parametric model is correctly specified, we can gain considerably by using parametric estimators. Meanwhile, 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. Examples from biology, economics, geography, medicine, and physics are abundant (see, e.g., Fuller, 1987, and Meister, 2009). 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 are contaminated by measurement error, conventional specification tests have incorrect size in general and may also suffer from low power properties.

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

The authors gratefully acknowledge financial support from the ERC Consolidator Grant (SNP 615882) (Otsu) and the ESRC (Taylor).

test statistic and propose a bootstrap critical value. As we discuss below, in contrast to existing methods, our test allows nonlinear regression models and possesses desirable 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 Section 2, 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 regression models with errors-in-variables, which is able to detect local alternatives at the $\sqrt{n}$-rate. We propose a smoothing specification test that complements Hall and Ma's (2007) test (see further discussion below).[1]

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. 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.

[1]Other papers that study specification testing under measurement error includes Butucea (2007), Holzmann and Boysen (2006), Holzmann, Bissantz and Munk (2007), and Ma *et al.* (2011) (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).

To determine critical values for our smoothing test, we propose a bootstrap procedure. Measurement error can cause difficulties in 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 noted, '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 obtained 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 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 and Section 5 applies the test to the specification of Engel curves. 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 3.1. 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

$$\begin{aligned} \mathrm{H}_0 &: \; m(x) = m_\theta(x) \text{ for almost every } x \in \mathbb{R}^d, \\ \mathrm{H}_1 &: \; \mathrm{H}_0 \text{ is false,} \end{aligned}$$

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

To test the null $\mathrm{H}_0$, 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)

$$\hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i \mathcal{K}_b(x - W_i)}{\sum_{i=1}^{n} \mathcal{K}_b(x - W_i)},$$

where

$$\mathcal{K}_b(a) = \frac{1}{(2\pi)^d} \int e^{-\mathrm{i}t \cdot a} \frac{K^{\mathrm{ft}}(tb)}{f_\epsilon^{\mathrm{ft}}(t)} dt,$$

is the so-called deconvolution kernel, $\mathrm{i} = \sqrt{-1}$, $b$ is a bandwidth, and $K^{\mathrm{ft}}$ and $f_\epsilon^{\mathrm{ft}}$ are the Fourier transforms of a kernel function $K$ and the measurement error density $f_\epsilon$, respectively.[2] Throughout the paper we assume $f_\epsilon^{\mathrm{ft}}(t) \neq 0$ for all $t \in \mathbb{R}^d$ and $K^{\mathrm{ft}}$ 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 Butucea and Taupin (2008), 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 $\mathrm{H}_0$, 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, $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_b(x - W_i)$ is the deconvolution kernel density estimator for $X$, $K_b(x) = \frac{1}{b^d} K\left(\frac{x}{b}\right)$, 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

$$D_n = \frac{n}{(2\pi)^d} \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} \left| \frac{1}{n}\sum_{i=1}^{n} Y_i e^{\mathrm{i}t \cdot W_i} - [m_{\hat{\theta}}\hat{f}]^{\mathrm{ft}}(t) f_\epsilon^{\mathrm{ft}}(t) \right|^2 dt.$$

Based on this expression, the distance $D_n$ can be interpreted as a contrast of the nonparametric and model-based estimators for $E[Ye^{\mathrm{i}t \cdot W}]$. Let $\zeta_i(t) = Y_i e^{\mathrm{i}t \cdot W_i} - \int e^{\mathrm{i}s \cdot W_i} m_{\hat{\theta}}^{\mathrm{ft}}(t-s) \frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t)$. To define the test statistic for $\mathrm{H}_0$, we further decompose $D_n$ as

$$\begin{aligned} D_n &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{(2\pi)^d} \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} |\zeta_i(t)|^2 dt + \frac{1}{n}\sum_{i \neq j} \frac{1}{(2\pi)^d} \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} \zeta_i(t)\overline{\zeta_j(t)} dt \\ &\equiv B_n + T_n, \end{aligned} \tag{1}$$

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

[2] 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 $\hat{m}(x) = \frac{\sum_{i=1}^{n} Y_i K_{1b}(x_1 - X_{1i})\mathcal{K}_b(x_2 - W_i)}{\sum_{i=1}^{n} K_{1b}(x_1 - X_{1i})\mathcal{K}_b(x_2 - W_i)}$, where $K_{1b}(a) = \frac{1}{b^{d_1}} K_1\left(\frac{a}{b}\right)$ and $K_1$ is a conventional kernel function for $X_1$, and analogous results can be established.

$B_n$ and employ $T_n$ as our test statistic for $\mathrm{H}_0$. 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 $\mathrm{H}_0$, one may consider testing some implication of $\mathrm{H}_0$ on the conditional mean $E[Y|W]$ of observables, i.e., consider $H_0' : f_W(w)E[Y\big|W = w] = \int m_\theta(w-u)f_X(w-u)f_\epsilon(u)du$ for almost every $w$, and test $\mathrm{H}_0'$ by a conventional method, such as Härdle and Mammen (1993). This approach was employed by Song (2008). To clarify the rationale of our testing approach based on $T_n$ against the conventional approach for $\mathrm{H}_0'$, let us consider the following local alternative hypothesis for the regression function

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

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 $\mathrm{H}_0'$ is determined by the $L_2$-norm of the convolution $\{(m_n - m_\theta)f_X\} * f_\epsilon$. By Parseval's identity and the Fourier shift formula, we have

$$\|\{(m_n - m_\theta)f_X\} * f_\epsilon\|^2 = a_n^2 \left\|\{q^{\mathrm{ft}}(\cdot - A_n) + q^{\mathrm{ft}}(\cdot + A_n)\}f_\epsilon^{\mathrm{ft}}\right\|^2,$$

where $q(x) = \left(\frac{\sin x}{x}\right)^2 f_X(x)$. For example, if $f_\epsilon$ is Laplace with $f_\epsilon^{\mathrm{ft}}(t) = 1/(1+t^2)$, then we can see that the $L_2$-norm $\|\{(m_n - m_\theta)f_X\} * f_\epsilon\|$ 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 $\mathrm{H}_0'$ 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 $\mathrm{H}_0$ and possesses desirable local power properties compared to the conventional tests on $\mathrm{H}_0'$.

## 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 $\mathrm{H}_0$. 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):** $f^{\mathrm{ft}}, m^{\mathrm{ft}}, \frac{\partial}{\partial\theta}(m_\theta^{\mathrm{ft}}) \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$.
- **(iii):** $K^{\mathrm{ft}}(t)$ *is compactly supported on* $[-1,1]^d$, *is symmetric around zero (i.e.,* $K^{\mathrm{ft}}(t) = K^{\mathrm{ft}}(-t)$*), 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 $X$ and the regression function $m$, as well as the derivative, with respect to $\theta$, of the Fourier transform of $m_\theta$. 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 $K(x) = \frac{\sin x}{\pi x}$ whose Fourier transform is equal to $K^{\text{ft}}(t) = \mathbb{I}\{-1 \leq t \leq 1\}$.

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

$$c|t|^{-d\alpha} \leq |f_\epsilon^{\text{ft}}(t)| \leq C|t|^{-d\alpha},$$

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

**(ii):** $\int |t|^{-2d\alpha}|f^{\text{ft}}(t)|^2 dt < \infty$, $\int |t|^{-2d\alpha}|m^{\text{ft}}(t)|^2 dt < \infty$, $\int |t|^{-2d\alpha}|[mf]^{\text{ft}}(t)|^2 dt < \infty$, $\int |t|^{-2d\alpha}|[m^2 f]^{\text{ft}}(t)|^2 dt < \infty$, *and* $\int |t|^{-2d\alpha}|[\sigma^2 f]^{\text{ft}}(t)|^2 dt < \infty$.

**(iii):** $\hat{\theta} - \theta = o_p(n^{-1/2} b^{-d(\frac{1}{4}-\alpha)})$ *under* $\mathrm{H}_0$.

Assumption O (i) requires that the Fourier transform $f_\epsilon^{\text{ft}}$ decays in some finite power. A popular example of an 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), 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 $\sqrt{n}(\hat{\theta} - \theta)$ 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):** $f_\epsilon^{\text{ft}}(t) \neq 0$ *for all* $t \in \mathbb{R}$ *and there exist positive constants* $C_\epsilon$, $\mu$, $\gamma_0$, *and* $\gamma > 1$ *such that*

$$f_\epsilon^{\text{ft}}(t) \sim C_\epsilon |t|^{\gamma_0} e^{-|t|^\gamma/\mu},$$

*as* $|t| \to \infty$. *Also, there exist constants* $A > 0$ *and* $\beta \geq 0$ *such that*

$$K^{\text{ft}}(1-t) = At^\beta + o(t^\beta),$$

*as* $t \to 0$.

**(ii):** $E[Y^4] < \infty$, $E[W^4] < \infty$, $\int |t|^{2\beta} \left| \frac{\partial}{\partial\theta} m_\theta^{\mathrm{ft}}(t) \right|^2 dt < \infty$, *and* $\int |t|^{2\beta} |m^{\mathrm{ft}}(t)|^2 dt < \infty$.
**(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 $f_\epsilon^{\mathrm{ft}}$ decays 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 $K(x) = \frac{\sin x}{\pi x}$ 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) regards 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* $\mathrm{H}_0$,

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

*where* $C_{V,b} = O(b^{-d(1+4\alpha)})$ *is defined in (3) in the Appendix.*

**(ii):** *Suppose that Assumptions D and S hold true with* $d = 1$ *and* $\epsilon \sim N(0,1)$. *Then under* $\mathrm{H}_0$,

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

*where* $\varphi(b) = \frac{(2\pi)2^{2\beta}}{b^{1+4\beta} e^{1/b^2} A^2 \Gamma(1+2\beta)}$ *with the gamma function* $\Gamma$, $\{Z_k\}$ *is an independent sequence of standard normal random variables and* $\{\lambda_k\}$ *is defined in (13) in the Appendix.*

**(iii):** *Suppose that Assumptions D and S hold true with* $d = 1$. *Then under* $\mathrm{H}_0$,

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

*where* $\phi(b) = \frac{(2\pi)2^{2\beta}\gamma^{1+2\beta} C_\epsilon^2}{\mu^{1+2\beta} b^{\gamma-1+2\gamma\beta+2\gamma_0} e^{2/(\mu b^\gamma)} A^2 \Gamma(1+2\beta)}$ *with the gamma function* $\Gamma$, $\{Z_k\}$ *is an independent sequence of standard normal random variables and* $\{\lambda_k\}$ *is defined in (13) 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 $C_{V,b}^{-1/2} = O(b^{d(\frac{1}{2}+2\alpha)})$ 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 $f_\epsilon^{\mathrm{ft}}$ increases, the convergence rate of $T_n$ becomes slower.

Theorem 1 (ii) focuses on the case of 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 $\beta = 0$), 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 $f_\epsilon^{\text{ft}}(t)$ 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 extended to the vector case. 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 $\mathrm{H}_0$ based on $T_n$. Alternatively, we can compute the critical values by bootstrap methods. A bootstrap counterpart of $T_n$ is given by perturbing each summand in $T_n$ as follows

$$T_n^* = \frac{1}{n}\sum_{i \neq j} \nu_i^* \nu_j^* \frac{1}{(2\pi)^d} \int \frac{|K^{\text{ft}}(tb)|^2}{|f_\epsilon^{\text{ft}}(t)|^2} \zeta_i(t)\overline{\zeta_j(t)}dt, \tag{2}$$

where $\{\nu_i^*\}_{i=1}^n$ 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 to Delgado, Dominguez and Lavergne (2006, Theorem 6).

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

$$\mathrm{H}_{1n} : m(x) = m_\theta(x) + c_n\Delta(x), \text{ for almost every } x \in \mathbb{R}^d$$

where $c_n \to 0$ and $\Delta(x)$ is a non-zero function such that the limits $\lim_{n\to\infty} \Delta_n$ and $\lim_{n\to\infty} \Upsilon_n$ defined in (15) and (16), 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* $\mathrm{H}_{1n}$ *with* $c_n = n^{-1/2} b^{-d(\frac{1}{4}+\alpha)}$,

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

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

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

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

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

Theorem 2 (i) says that under the ordinary smooth case, our test has non-trivial power against local alternatives drifting with the rate of $c_n = n^{-1/2} b^{-d(\frac{1}{4}+\alpha)}$. 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 $f_\epsilon^{\mathrm{ft}}$ 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 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 $T_n$ by inserting the estimated Fourier transform of the measurement error density, $\hat{f}_\epsilon^{\mathrm{ft}}$.

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

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

Delaigle, Hall and Meister (2008) studied the asymptotic properties of the deconvolution density and regression estimators using $\hat{f}_\epsilon^{\mathrm{ft}}$ 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 $T_n$ obtained above remain unchanged when we replace $f_\epsilon^{\mathrm{ft}}$ with $\hat{f}_\epsilon^{\mathrm{ft}}$. If the researcher wishes to remove the assumption that $f_\epsilon^{\mathrm{ft}}$ 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 we consider the same data generating process as Hall and Ma (2007) for ease of comparison. We also compare our test to Song (2008). Recall that although Song's (2008) and Hall and Ma's (2007) test are 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 supersmooth 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 $\{0.01, 0.05, 0.1, 0.5, 1, 1.5\}$ so 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). For the test of Song (2008) we use the same kernel as for our test and choose bandwidths by cross-validation (there was little dependence on the bandwidths so we report only for these cross-validated values). All results are based on 1000 Monte Carlo replications.

Table 1 takes $C = 0$ so as to asses 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 for all tests 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.

The columns labeled 'HM' correspond to the power of the test proposed in Hall and Ma (2007) and the columns labeled 'S' correspond to the power of the test proposed by Song (2008).

Table 1: $Y = 1 + 1.5X + U$

| **Ordinary Smooth** | | **Bandwidth** | | | | | |
|---|---|---|---|---|---|---|---|
| **n** | **Level** | **0.01** | **0.05** | **0.1** | **0.5** | **1** | **1.5** |
| **50** | **1%** | 3.0% | 2.6% | 2.3% | 3.1% | 2.4% | 0.6% |
| | **5%** | 7.6% | 7.1% | 6.3% | 6.5% | 6.4% | 2.4% |
| | **10%** | 12.7% | 11.9% | 10.7% | 9.8% | 10.1% | 5.7% |
| **100** | **1%** | 2.2% | 2.0% | 2.8% | 2.3% | 2.2% | 0.6% |
| | **5%** | 4.9% | 5.6% | 6.8% | 6.3% | 6.9% | 3.4% |
| | **10%** | 11.6% | 10.5% | 11.9% | 12.4% | 11.1% | 7.3% |
| **Super Smooth** | | | | | | | |
| **50** | **1%** | 2.1% | 1.9% | 1.3% | 1.4% | 1.9% | 1.2% |
| | **5%** | 5.3% | 5.2% | 4.9% | 5.4% | 5.7% | 4.3% |
| | **10%** | 11.1% | 9.3% | 9.6% | 10.8% | 10.4% | 7.7% |
| **100** | **1%** | 2.9% | 2.4% | 2.7% | 2.0% | 1.7% | 1.9% |
| | **5%** | 6.9% | 6.5% | 6.0% | 6.6% | 5.3% | 5.8% |
| | **10%** | 12.3% | 10.7% | 10.3% | 10.6% | 10.8% | 10.5% |

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

| **Ordinary Smooth** | | **Bandwidth** | | | | | | HM | S |
|---|---|---|---|---|---|---|---|---|---|
| **n** | **Level** | **0.01** | **0.05** | **0.1** | **0.5** | **1** | **1.5** | | |
| **50** | **1%** | 43.2% | 33.7% | 26.5% | 40.6% | 81.8% | 63.3% | 71.0% | 86.1% |
| | **5%** | 63.1% | 52.7% | 45.5% | 61.6% | 92.1% | 78.3% | 76.9% | 92.8% |
| | **10%** | 73.5% | 63.1% | 57.8% | 72.2% | 94.8% | 86.2% | 85.3% | 95.3% |
| **100** | **1%** | 67.4% | 50.1% | 38.5% | 66.5% | 99.3% | 97.6% | 95.8% | 88.6% |
| | **5%** | 84.2% | 71.3% | 61.4% | 85.1% | 99.8% | 99.2% | 97.1% | 99.1% |
| | **10%** | 91.0% | 80.1% | 73.0% | 92.3% | 99.9% | 99.9% | 99.1% | 99.9% |
| **Super Smooth** | | | | | | | | | |
| **50** | **1%** | 28.1% | 22.3% | 19.7% | 12.2% | 24.5% | 36.1% | 66.2% | 67.2% |
| | **5%** | 46.6% | 38.7% | 33.5% | 24.1% | 44.5% | 55.2% | 72.0% | 84.2% |
| | **10%** | 57.4% | 48.4% | 43.2% | 34.9% | 56.9% | 67.3% | 80.4% | 88.6% |
| **100** | **1%** | 54.0% | 35.9% | 23.5% | 15.4% | 47.2% | 69.0% | 94.3% | 89.3% |
| | **5%** | 70.7% | 52.5% | 41.5% | 28.8% | 63.8% | 84.6% | 95.9% | 94.5% |
| | **10%** | 78.1% | 64.5% | 53.9% | 42.1% | 73.7% | 89.8% | 97.7\|% | 97.8% |

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

| **Ordinary Smooth** | | **Bandwidth** | | | | | | HM | S |
|---|---|---|---|---|---|---|---|---|---|
| **n** | **Level** | **0.01** | **0.05** | **0.1** | **0.5** | **1** | **1.5** | | |
| **100** | **1%** | 21.3% | 15.9% | 13.7% | 12.9% | 6.6% | 2.4% | 13.2% | 8.9% |
| | **5%** | 40.3% | 30.3% | 26.5% | 27.5% | 17.6% | 6.8% | 27.5% | 19.6% |
| | **10%** | 51.4% | 45.0% | 38.2% | 38.0% | 26.5% | 13.6% | 38.7% | 26.8% |
| **200** | **1%** | 35.6% | 19.7% | 12.7% | 19.4% | 11.0% | 2.6% | 20.3% | 14.4% |
| | **5%** | 54.5% | 37.2% | 27.9% | 36.2% | 22.2% | 8.7% | 37.4% | 24.4% |
| | **10%** | 66.4% | 50.0% | 39.8% | 49.0% | 32.2% | 17.0% | 50.0% | 39.0% |
| **Super Smooth** | | | | | | | | | |
| **100** | **1%** | 15.3% | 12.4% | 7.3% | 4.3% | 4.0% | 2.6% | 4.7% | 11.2% |
| | **5%** | 29.0% | 22.3% | 18.1% | 11.5% | 9.6% | 7.1% | 11.0% | 17.4% |
| | **10%** | 38.7% | 31.7% | 27.3% | 19.5% | 17.1% | 14.8% | 20.3% | 30.3% |
| **200** | **1%** | 23.8% | 16.7% | 10.6% | 4.3% | 5.3% | 3.0% | 5.7% | 14.7% |
| | **5%** | 40.9% | 29.1% | 22.0% | 13.7% | 11.0% | 8.2% | 14.2% | 26.0% |
| | **10%** | 52.9% | 38.4% | 32.3% | 21.3% | 18.6% | 12.7% | 24.3% | 34.9% |

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

| **Ordinary Smooth** | | **Bandwidth** | | | | | | HM | S |
|---|---|---|---|---|---|---|---|---|---|
| **n** | **Level** | **0.01** | **0.05** | **0.1** | **0.5** | **1** | **1.5** | | |
| **100** | **1%** | 20.9% | 16.7% | 13.8% | 6.9% | 5.6% | 1.8% | 9.2% | 7.3% |
| | **5%** | 38.7% | 31.5% | 28.0% | 17.1% | 12.5% | 5.6% | 20.6% | 13.6% |
| | **10%** | 49.3% | 43.2% | 37.8% | 25.4% | 19.5% | 9.9% | 29.7% | 23.7% |
| **200** | **1%** | 35.9% | 20.8% | 15.3% | 7.8% | 4.8% | 1.3% | 9.3% | 13.6% |
| | **5%** | 55.9% | 37.4% | 28.9% | 18.4% | 11.2% | 4.6% | 21.7% | 23.8% |
| | **10%** | 66.8% | 49.8% | 40.4% | 28.6% | 17.6% | 10.1% | 31.4% | 31.9% |
| **Super Smooth** | | | | | | | | | |
| **100** | **1%** | 16.1% | 11.2% | 9.1% | 5.3% | 4.0% | 3.1% | 5.2% | 7.6% |
| | **5%** | 30.4% | 22.4% | 17.9% | 12.6% | 11.0% | 7.3% | 11.6% | 17.3% |
| | **10%** | 41.3% | 35.0% | 26.1% | 20.6% | 17.4% | 13.7% | 19.3% | 27.1% |
| **200** | **1%** | 23.6% | 13.4% | 9.4% | 5.1% | 4.7% | 3.7% | 5.3% | 8.9% |
| | **5%** | 39.4% | 25.0% | 20.0% | 11.8% | 12.0% | 8.4% | 13.1% | 25.0% |
| | **10%** | 50.8% | 35.8% | 30.5% | 20.5% | 19.4% | 13.9% | 20.3% | 32.2% |

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

| **Ordinary Smooth** | | **Bandwidth** | | | | | | HM | S |
|---|---|---|---|---|---|---|---|---|---|
| **n** | **Level** | **0.01** | **0.05** | **0.1** | **0.5** | **1** | **1.5** | | |
| **100** | **1%** | 22.8% | 17.0% | 11.8% | 7.9% | 5.4% | 1.8% | 9.4% | 7.3% |
| | **5%** | 39.3% | 32.6% | 28.2% | 17.7% | 12.9% | 5.9% | 20.8% | 19.0% |
| | **10%** | 50.6% | 46.1% | 38.1% | 27.2% | 19.8% | 10.4% | 31.8% | 27.1% |
| **200** | **1%** | 36.4% | 20.1% | 14.0% | 8.1% | 4.2% | 1.9% | 9.9% | 7.9% |
| | **5%** | 54.1% | 36.3% | 29.2% | 19.4% | 10.7% | 5.0% | 21.8% | 21.8% |
| | **10%** | 67.1% | 48.9% | 40.0% | 27.6% | 18.0% | 10.3% | 32.0% | 27.7% |
| **Super Smooth** | | | | | | | | | |
| **100** | **1%** | 17.4% | 10.9% | 7.5% | 4.7% | 4.6% | 3.3% | 5.6% | 8.5% |
| | **5%** | 31.4% | 22.8% | 19.6% | 12.8% | 10.1% | 9.1% | 12.0% | 17.4% |
| | **10%** | 42.0% | 33.2% | 28.3% | 20.8% | 16.9% | 14.3% | 21.0% | 26.1% |
| **200** | **1%** | 21.5% | 15.6% | 9.3% | 5.5% | 4.6% | 2.9% | 4.8% | 14.0% |
| | **5%** | 39.1% | 29.0% | 21.6% | 12.3% | 11.6% | 8.1% | 14.1% | 22.5% |
| | **10%** | 50.9% | 37.5% | 30.2% | 21.1% | 17.9% | 13.7% | 22.6% | 27.8% |

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 an under-rejection of the null hypothesis.

Table 2 gives a direct comparison to the tests proposed in Hall and Ma (2007) and Song (2008). As we expected, in this low frequency alternative setting, our test is generally slightly less powerful than the other tests. Having said this, in the ordinary smooth case for several choices of bandwidth our test does display the highest power. Hall and Ma's (2007) test is able to detect local alternatives at the $\sqrt{n}$-rate for both ordinary and supersmooth measurement error distributions, and the test of Song (2008) is able to detect local alternatives at the rate $\sqrt{nb^{d/2}}$ in both cases. However, our test achieves a 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. However, the test is still able to enjoy considerable power in this case especially for larger sample sizes.

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 $\delta$. 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 better able to pick up on these rapid changes.

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 $E[Y|X]$. This fact is also reflected in the Monte Carlo simulations where the power falls as we move to higher frequency alternatives and is inferior to the test we propose. Interestingly the test of Song (2008) appears to dominate the test of Hall and Ma (2007) for the supersmooth case but not for the ordinary smooth case.

We can learn from these simulations that for reasonably small samples with supersmooth measurement error, perhaps the tests proposed by Hall and Ma (2007) or Song (2008) would be a wiser choice if one suspects deviations from the null of a low frequency type, otherwise our test appears to be superior. However, we suggest that to avoid any risk of very low power the test proposed in this paper may be the best option.

In order to account for the dependence of our test on the bandwidth we may look to employ the ideas of Horowitz and Spokoiny (2001) to construct a test that is adaptive to the smoothness of the regression function. In order to do this we could construct a test statistic of the form

$$T_{A,n} = \max_{b_n \in \mathbb{B}_n} T_n$$

where $\mathbb{B}_n$ is a finite set of bandwidths. To obtain valid critical values we can use a bootstrap procedure similar to the one proposed in Section 3. Specifically we construct a bootstrap counterpart as

$$T^*_{A,n} = \max_{b_n \in \mathbb{B}_n} T^*_n$$

where $T^*_n$ is defined in (2). It is beyond the scope of this paper to determine the asymptotic properties of such a test but this could prove to be a fruitful area for future research.

## 5. Empirical Example

We apply our test to the specification of Engel curves for food, clothing and transport. An Engel curve describes the relationship between an individual's purchases of a particular good and their total resources and hence provides an estimate of a good's income elasticity. Much work has been carried out on the estimation of Engel curves and in particular the correct functional form which has been shown to significantly affect estimates of income elasticity (see for example Leser, 1963). Hausmann, Newey and Powell (1995) highlighted the problem that measurement error plays in the estimation of Engel curves. To the best of our knowledge no previous work has tested the parametric specification of Engel curves whilst accounting for the inherent measurement error in the data.

We concentrate on the Working-Leser specification put forward by Leser (1963)

$$Y_{gi} \quad = \quad a_0 + a_1\big(X_i \log(X_i)\big) + a_2 X_i + U_i$$

where $Y_{gi}$ is the expenditure on good $g$ of consumer $i$ and $X_i$ is the *true* total expenditure of consumer $i$. It is commonly believed that the measurement error in total expenditure is multiplicative, hence we take $\tilde{X}_i = \log(X_i)$ as our true regressor and adjust the specification accordingly, as in Schennach (2004). We use data from the Consumer Expenditure Survey where we take the third quarter of 2014 as our sample, giving 4312 observations. To account for the measurement error we make use of repeated measurements of $X$. Specifically, we use total expenditure from the current quarter as one measurement and total expenditure from the previous quarter as the other. To estimate the parametric form we employ the estimator proposed by Schennach (2004). For the nonparametric estimator we use Delaigle, Hall and Meister (2008) and select the bandwidth using the cross-validation approach also proposed in that paper. To analyse the sensitivity of our test to the choice of bandwidth we report results for various bandwidths around the cross-validated choice $b \approx 0.15$. We use the same kernel and bootstrap procedure as implemented in the Monte Carlo simulations.

Table 6 reports the p-value for our specification test on food, clothing and transport for a range of bandwidths.

Table 6: Engel curve p-values

| **Good** | **Bandwidth** | | | | | |
|---|---|---|---|---|---|---|
| | **0.05** | **0.10** | **0.15** | **0.25** | **0.35** | **0.5** |
| **Food** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.07 |
| **Clothing** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 |
| **Transport** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

We can see that the test is fairly robust to the choice of bandwidth. The parametric specification is rejected for all bandwidths apart from 0.5 where we fail to reject the null hypothesis at the 5% level for food and at the 10% level for clothing. Interestingly Härdle and Mammen (1993) obtained similar findings in the case of transport but tended to fail to reject the Working-Leser specification for food. Thus, it appears that accounting for measurement error is indeed very important to draw the correct conclusions and must not simply be ignored.

## Appendix A. Mathematical Appendix

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

$$\begin{aligned}\xi_i(t) &= Y_i e^{\mathrm{i}t\cdot W_i} - \int e^{\mathrm{i}s\cdot W_i} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds f_\epsilon^{\mathrm{ft}}(t),\\ H_{i,j} &= \int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\xi_j(t)}dt.\end{aligned}$$

Then $C_{V,b}$ is defined as

$$\begin{aligned}C_{V,b} &= E[H_{1,2}^2] \qquad (3)\\ &= \int\int \frac{|K^{\mathrm{ft}}(t_1b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_1)|^2}\frac{|K^{\mathrm{ft}}(t_2b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_2)|^2}\Bigg|\left\{[m^2f]^{\mathrm{ft}}(t_1+t_2)+[\sigma^2 f]^{\mathrm{ft}}(t_1+t_2)\right\}f_\epsilon^{\mathrm{ft}}(t_1+t_2)\\ &\quad + \int\int f_W^{\mathrm{ft}}(s_1+s_2)m^{\mathrm{ft}}(t_1-s_1)m^{\mathrm{ft}}(t_2-s_2)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}\frac{K^{\mathrm{ft}}(s_2b)}{f_\epsilon^{\mathrm{ft}}(s_2)}ds_1ds_2f_\epsilon^{\mathrm{ft}}(t_1)f_\epsilon^{\mathrm{ft}}(t_2)\\ &\quad - \int [mf]^{\mathrm{ft}}(t_2+s_1)f_\epsilon^{\mathrm{ft}}(t_2+s_1)m^{\mathrm{ft}}(t_1-s_1)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}ds_1f_\epsilon^{\mathrm{ft}}(t_1)\\ &\quad - \int [mf]^{\mathrm{ft}}(t_1+s_1)f_\epsilon^{\mathrm{ft}}(t_1+s_1)m^{\mathrm{ft}}(t_2-s_1)\frac{K^{\mathrm{ft}}(s_1b)}{f_\epsilon^{\mathrm{ft}}(s_1)}ds_1f_\epsilon^{\mathrm{ft}}(t_2)\Bigg|^2dt_1dt_2.\end{aligned}$$

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,

$$\begin{aligned}&\int\int \frac{|K^{\mathrm{ft}}(t_1b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_1)|^2}\frac{|K^{\mathrm{ft}}(t_2b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_2)|^2}\left|\left\{[m^2f]^{\mathrm{ft}}(t_1+t_2)+[\sigma^2 f]^{\mathrm{ft}}(t_1+t_2)\right\}f_\epsilon^{\mathrm{ft}}(t_1+t_2)\right|^2dt_1dt_2\\ \sim\ & b^{-2d-4d\alpha}\int\int |K^{\mathrm{ft}}(a_1)|^2|K^{\mathrm{ft}}(a_2)|^2|a_1|^{2d\alpha}|a_2|^{2d\alpha}|(a_1+a_2)/b|^{-2d\alpha}\left|[(m^2+\sigma^2)f]^{\mathrm{ft}}((a_1+a_2)/b)\right|^2da_1da_2\\ \sim\ & b^{-d-4d\alpha}\int |a|^{-2d\alpha}\left|[(m^2+\sigma^2)f]^{\mathrm{ft}}(a)\right|^2da\int |K^{\mathrm{ft}}(a_2)|^4|a_2|^{4d\alpha}da_2\\ =\ & O(b^{-d(1+4\alpha)}), \qquad (4)\end{aligned}$$

where the first wave relation follows from the change of variables $(a_1,a_2)=(t_1b,t_2b)$ 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 $C_{V,b}=O(b^{-d(1+4\alpha)})$.

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

$$\rho_i(t) = \int e^{\mathrm{i}s\cdot W_i}\{m_\theta^{\mathrm{ft}}(t-s)-m_{\hat\theta}^{\mathrm{ft}}(t-s)\}\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds f_\epsilon^{\mathrm{ft}}(t).$$

Then the test statistic $T_n$ is written as

$$\begin{aligned} T_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\xi_j(t)}dt+\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_i(t)\overline{\rho_j(t)}dt \\ &\quad +\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_i(t)\overline{\xi_j(t)}dt+\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\rho_j(t)}dt \\ &\equiv \tilde{T}_n+R_{1n}+R_{2n}+R_{3n}. \end{aligned}$$

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

$$R_{1n}=o_p(b^{-d/2-2\alpha})\left|\frac{1}{n^2}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right|, \tag{5}$$

where $\rho_{1i}(t)=\int e^{\mathrm{i}s\cdot W_i}\frac{\partial m_\theta^{\mathrm{ft}}(t-s)}{\partial\theta}\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}dsf_\epsilon^{\mathrm{ft}}(t)$. By the Cauchy-Schwarz inequality and Assumption D (ii),

$$\begin{aligned} E\left[\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right] &= \int|K^{\mathrm{ft}}(tb)|^2\left|\int f^{\mathrm{ft}}(s)\frac{\partial}{\partial\theta}\big(m_\theta^{\mathrm{ft}}(t-s)\big)K^{\mathrm{ft}}(sb)ds\right|^2dt \\ &= O(1). \end{aligned} \tag{6}$$

Also, by applying the same argument to (4) under Assumption O (ii), we have

$$E\left[\left(\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\rho_{1i}(t)\overline{\rho_{1j}(t)}dt\right)^2\right]=O(b^{-d(1+4\alpha)}). \tag{7}$$

Combining (5)-(7) 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 $C_{V,b}^{-1/2}R_{3n}=o_p(1)$ under Assumption O (ii)-(iii) and thus $C_{V,b}^{-1/2}T_n=C_{V,b}^{-1/2}\tilde{T}_n+o_p(1)$.

Second, we derive the limiting distribution of $C_{V,b}^{-1/2}\tilde{T}_n$. Note that $\tilde{T}_n$ is written as $\tilde{T}_n=\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}H_{i,j}$ and is a U-statistic with zero mean (because $E[Y\exp(\mathrm{i}t\cdot W)]=[m_\theta f]^{\mathrm{ft}}(t)f_\epsilon^{\mathrm{ft}}(t)$ under $\mathrm{H}_0$). 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

$$\frac{E[H_{1,2}^4]}{n(E[H_{1,2}^2])^2}\to 0,\quad\text{and}\quad\frac{E[G_{1,2}^2]}{(E[H_{1,2}^2])^2}\to 0, \tag{8}$$

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 (3) satisfies $C_{V,b}=O(b^{-d-4\alpha})$. By a similar argument to bound $E[H_{1,2}^2]$ in (4), we can show

$$E[H_{1,2}^4]=E\left[\int\cdots\int\prod_{k=1}^4\frac{|K^{\mathrm{ft}}(t_kb)|^2}{|f_\epsilon^{\mathrm{ft}}(t_k)|^2}\xi_1(t_k)\overline{\xi_2(t_k)}dt_1\cdots dt_4\right]=O(b^{-3d(1+8\alpha)}).$$

For $E[G_{1,2}^2]$, we can equivalently look at

$$\begin{aligned}
& E[H_{1,3}H_{1,4}H_{2,3}H_{2,4}] \\
= & \int \cdots \int \prod_{k=1}^{4} \frac{|K^{\mathrm{ft}}(t_k b)|^2}{|f_\epsilon^{\mathrm{ft}}(t_k)|^2} \xi_1(t_1)\overline{\xi_3(t_1)}\xi_1(t_2)\overline{\xi_4(t_2)}\xi_2(t_3)\overline{\xi_3(t_3)}\xi_2(t_4)\overline{\xi_4(t_4)} dt_1 \cdots dt_4 \\
= & O(b^{-d(1+8\alpha)}).
\end{aligned}$$

These results combined with Assumption D (iv) guarantee the conditions in (8). 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 $\tilde{T}_n = \bar{T}_n + r_{1n} + r_{2n} + r_{3n}$, where

$$\bar{T}_n = \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} Y_i e^{\mathrm{i}tW_i}\overline{Y_j e^{\mathrm{i}tW_j}} dt, \tag{9}$$

$$r_{1n} = \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\left(\int e^{\mathrm{i}sW_i} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t)\right)\overline{\left(\int e^{\mathrm{i}sW_j} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t)\right)} dt,$$

$$r_{2n} = \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2} Y_i e^{\mathrm{i}tW_i}\overline{\left(\int e^{\mathrm{i}sW_j} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t)\right)} dt,$$

$$r_{3n} = \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int \frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\left(\int e^{\mathrm{i}sW_i} m^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)} ds f_\epsilon^{\mathrm{ft}}(t)\right)\overline{Y_j e^{\mathrm{i}tW_j}} dt.$$

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

$$\begin{aligned}
\bar{T}_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(tb)|^2 e^{t^2} Y_i Y_j \{\cos(tW_i)\cos(tW_j) + \sin(tW_i)\sin(tW_j)\} dt \\
&= \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 e^{(t/b)^2} Y_i Y_j \left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right) + \sin\left(\frac{tW_i}{b}\right)\sin\left(\frac{tW_j}{b}\right)\right\} dt \\
&= \left(\frac{1}{b}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 e^{(t/b)^2} dt\right)\frac{1}{n}\sum_{i\neq j} Y_i Y_j \left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right) + \sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} \\
&\quad + O_p(b^{2+4\beta}e^{1/b^2}) \\
&\equiv \left(\frac{1}{b}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 e^{(t/b)^2} dt\right)\tilde{T}_n + O_p(b^{2+4\beta}e^{1/b^2}),
\end{aligned} \tag{10}$$

where the first equality follows from $f_\epsilon^{\mathrm{ft}}(t) = e^{-t^2/2}$ and $e^{\mathrm{i}tW_i} = \cos(tW_i) + \mathrm{i}\sin(tW_i)$, 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

$$\tilde{T}_n = \frac{1}{n}\sum_{i\neq j} Y_iY_j \left[\begin{array}{c} \left\{\cos\left(\frac{X_i}{b}\right)\cos\left(\frac{\epsilon_i}{b}\right) - \sin\left(\frac{X_i}{b}\right)\sin\left(\frac{\epsilon_i}{b}\right)\right\}\left\{\cos\left(\frac{X_j}{b}\right)\cos\left(\frac{\epsilon_j}{b}\right) - \sin\left(\frac{X_j}{b}\right)\sin\left(\frac{\epsilon_j}{b}\right)\right\} \\ +\left\{\sin\left(\frac{X_i}{b}\right)\cos\left(\frac{\epsilon_i}{b}\right) + \cos\left(\frac{X_i}{b}\right)\sin\left(\frac{\epsilon_i}{b}\right)\right\}\left\{\sin\left(\frac{X_j}{b}\right)\cos\left(\frac{\epsilon_j}{b}\right) + \cos\left(\frac{X_j}{b}\right)\sin\left(\frac{\epsilon_j}{b}\right)\right\} \end{array}\right]. \tag{11}$$

From van Es and Uh (2005, proof of Lemma 6), it holds $\left(\frac{X_i}{b} \bmod 2\pi\right) \xrightarrow{d} V_i^X \sim U[0,2\pi]$ and $\left(\frac{\epsilon_i}{b} \bmod 2\pi\right) \xrightarrow{d} V_i^\epsilon \sim U[0,2\pi]$ as $b \to 0$ for each $i$, where $V_i^\epsilon$ 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 $\tilde{T}_n^V = \frac{1}{n}\sum_{i\neq j} h(Q_i,Q_j)$, where $Q_i = (Y_i, V_i^X, V_i^\epsilon)$ and

$$h(Q_i,Q_j) = Y_iY_j \left[\begin{array}{c} \left\{\cos(V_i^X)\cos(V_i^\epsilon) - \sin(V_i^X)\sin(V_i^\epsilon)\right\}\left\{\cos(V_j^X)\cos(V_j^\epsilon) - \sin(V_j^X)\sin(V_j^\epsilon)\right\} \\ +\left\{\sin(V_i^X)\cos(V_i^\epsilon) + \cos(V_i^X)\sin(V_i^\epsilon)\right\}\left\{\sin(V_j^X)\cos(V_j^\epsilon) + \cos(V_j^X)\sin(V_j^\epsilon)\right\} \end{array}\right].$$

Observe that $\mathrm{Cov}\left(h(Q_1,Q_2), h(Q_1,Q_3)\right) = 0$ because $E[\cos(V_i^\epsilon)] = 0$. Therefore, by applying the limit theorem for degenerate U-statistics with a fixed kernel $h$ (Serfling, 1980, Theorem 5.5.2), we obtain

$$\tilde{T}_n^V \xrightarrow{d} \sum_{k=1}^\infty \lambda_k (Z_k^2 - 1), \tag{12}$$

where $\{Z_k\}$ is an independent sequence of standard normal random variables and $\{\lambda_k\}$ are the eigenvalues of the integral operator

$$(\Lambda g)(Q_1) = \lambda g(Q_1). \tag{13}$$

where $(\Lambda g)(Q_1) = E[h(Q_1,Q_2)g(Q_2)|Q_1]$. Also, van Es and Uh (2005, Lemma 5) gives

$$\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 e^{(t/b)^2} dt \sim \frac{b}{\varphi(b)}, \tag{14}$$

where $\Gamma(\cdot)$ is the gamma function. Combining (10)-(14),

$$\varphi(b)\bar{T}_n \xrightarrow{d} \sum_{k=1}^\infty \lambda_i (Z_i^2 - 1).$$

Next, we show that $r_{1n}$ is negligible. Observe that

$$\begin{aligned}
r_{1n} &= \frac{1}{nb^3}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 \left(\int e^{\mathrm{i}sW_i/b} m^{\mathrm{ft}}\left(\frac{t-s}{b}\right)\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/b)}ds\right)\overline{\left(\int e^{\mathrm{i}sW_j/b} m^{\mathrm{ft}}\left(\frac{t-s}{b}\right)\frac{K^{\mathrm{ft}}(s)}{f_\epsilon^{\mathrm{ft}}(s/b)}ds\right)}dt \\
&= \frac{1}{nb^3}\sum_{i\neq j}\frac{1}{2\pi}\int |K^{\mathrm{ft}}(t)|^2 \left[\begin{array}{c} \left\{\int \cos\left(\frac{s_1W_i}{b}\right) m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right)\frac{K^{\mathrm{ft}}(s_1)}{f_\epsilon^{\mathrm{ft}}(s_1/b)}ds_1\right\} \\ \times\left\{\int \cos\left(\frac{s_2W_j}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right)\frac{K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(-s_2/b)}ds_2\right\} \\ +\left\{\int \sin\left(\frac{s_1W_i}{b}\right) m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right)\frac{K^{\mathrm{ft}}(s_1)}{f_\epsilon^{\mathrm{ft}}(s_1/b)}ds_1\right\} \\ \times\left\{\int \sin\left(\frac{s_2W_j}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right)\frac{K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(-s_2/b)}ds_2\right\} \end{array}\right] dt \\
&= \left(\frac{1}{2\pi}\int\int\int \frac{|K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(s_1)K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(s_1/b)f_\epsilon^{\mathrm{ft}}(-s_2/b)} m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right) ds_1 ds_2 dt\right) \\
&\quad \times \frac{1}{nb^3}\sum_{i\neq j}\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right) + \sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} + O_p(b^{2+4\beta}e^{1/b^2}),
\end{aligned}$$

where the first equality follows from a change of variables, the second equality follows from a direct calculation using $e^{\mathrm{i}sW_i} = \cos(sW_i) + \mathrm{i}\sin(sW_i)$, the third equality follows from Holzmann and Boysen (2006, Theorem 1) based on Assumption S (ii). By a similar argument to show (12), 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

$$\begin{aligned}
&\frac{1}{2\pi}\int\int\int \frac{|K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(s_1)K^{\mathrm{ft}}(s_2)}{f_\epsilon^{\mathrm{ft}}(s_1/b)f_\epsilon^{\mathrm{ft}}(-s_2/b)} m^{\mathrm{ft}}\left(\frac{t-s_1}{b}\right) m^{\mathrm{ft}}\left(\frac{s_2-t}{b}\right) ds_1 ds_2 dt \\
=\ & \frac{b^4 e^{1/b^2}}{2\pi}\int\int\int \left[\begin{array}{c} |K^{\mathrm{ft}}(t)|^2 K^{\mathrm{ft}}(1-b^2v_1)K^{\mathrm{ft}}(1-b^2v_2) \\ \times e^{\frac{(1-b^2v_1)^2-1}{2b^2}} e^{\frac{(1-b^2v_2)^2-1}{2b^2}} m^{\mathrm{ft}}\left(\frac{t-1+b^2v_1}{b}\right) m^{\mathrm{ft}}\left(\frac{1-b^2v_2-t}{b}\right)\end{array}\right] dv_1 dv_2 dt \\
\sim\ & \frac{A^2 b^{4+4\beta} e^{1/b^2}}{2\pi}\left(\int |K^{\mathrm{ft}}(t)|^2 m^{\mathrm{ft}}\left(\frac{t-1}{b}\right) m^{\mathrm{ft}}\left(\frac{1-t}{b}\right) dt\right)\left(\int v_1^\beta e^{-v_1} dv_1\right)\left(\int v_2^\beta e^{-v_2} dv_2\right) \\
\sim\ & \frac{A^2\Gamma(\beta+1)^2 b^{5+6\beta} e^{1/b^2}}{2\pi}\int |t|^{2\beta}|m^{\mathrm{ft}}(t)|^2 dt \\
=\ & O(b^{5+6\beta}e^{1/b^2}),
\end{aligned}$$

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 $\bar{T}_n$ defined in (9) satisfies

$$\bar{T}_n \sim \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}} Y_i Y_j\left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)+\sin\left(\frac{tW_i}{b}\right)\sin\left(\frac{tW_j}{b}\right)\right\} dt.$$

We now show that

$$\begin{aligned}
D_{\cos} \equiv\ & \frac{1}{nb}\sum_{i\neq j}\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}} Y_i Y_j\left\{\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right)\right\} dt \\
& -\left(\frac{1}{2\pi C_\epsilon^2}\int |K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0} e^{\frac{2|t|^\gamma}{\mu b^\gamma}} dt\right)\frac{1}{nb}\sum_{i\neq j} Y_i Y_j\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)\right\}
\end{aligned}$$

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

$$\left|\cos\left(\frac{tW_i}{b}\right)\cos\left(\frac{tW_j}{b}\right) - \cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)\right| \leq (1-t)\frac{(|W_i|+|W_k|)}{b}.$$

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

$$\begin{aligned}
\mathrm{Var}(D_{\cos}) &\leq O(n^{-2}b^{4\gamma_0-4})\left(\int(1-t)|K^{\mathrm{ft}}(t)|^2\,|t|^{-2\gamma_0}\,e^{\frac{2|t|^\gamma}{\mu b^\gamma}}dt\right)^2\sum_{i\neq j}E\left[|Y_i|^2|Y_j|^2(|W_i|+|W_k|)^2\right] \\
&= O\left(b^{4\gamma_0-4}\left(b^{\gamma(2+2\beta)}e^{\frac{2}{\mu b^\gamma}}\right)^2\right),
\end{aligned}$$

and we obtain $D_{\cos}=O_p\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right)$. The same argument applies to $D_{\sin}$. Note that

$$\begin{aligned}
\bar{T}_n &= \left(\frac{1}{b}\frac{1}{2\pi C_\epsilon^2}\int|K^{\mathrm{ft}}(t)|^2\left|\frac{t}{b}\right|^{-2\gamma_0}e^{\frac{2|t|^\gamma}{\mu b^\gamma}}dt\right)\frac{1}{n}\sum_{i\neq j}Y_iY_j\left\{\cos\left(\frac{W_i}{b}\right)\cos\left(\frac{W_j}{b}\right)+\sin\left(\frac{W_i}{b}\right)\sin\left(\frac{W_j}{b}\right)\right\} \\
&\quad +O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right) \\
&= \frac{A^2\mu^{1+2\beta}b^{\gamma-1+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\Gamma(2\beta+1)}{(2\lambda)^{1+2\beta}\pi C_\epsilon^2}\tilde{T}_n+O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right) \\
&\equiv \frac{\tilde{T}_n}{\phi(b)}+O\left(b^{2(\gamma-1)+2\gamma\beta+2\gamma_0}e^{\frac{2}{\mu b^\gamma}}\right),
\end{aligned}$$

where the second equality follows from the definition of $\tilde{T}_n$ in (11) and a modification of van Es and Uh (2005, Lemma 5). Therefore, we obtain

$$\phi(b)T_n=\tilde{T}_n+O(b^{\gamma-1}).$$

The limiting distribution of $\tilde{T}_n$ 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

$$\begin{aligned}
T_n &= \frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\xi_j(t)}dt+\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\eta_i(t)\overline{\eta_j(t)}dt \\
&\quad +\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\xi_i(t)\overline{\eta_j(t)}dt+\frac{1}{n}\sum_{i\neq j}\frac{1}{(2\pi)^d}\int\frac{|K^{\mathrm{ft}}(tb)|^2}{|f_\epsilon^{\mathrm{ft}}(t)|^2}\eta_i(t)\overline{\xi_j(t)}dt+o_p(C_{V,b}^{1/2}) \\
&\equiv \tilde{T}_n+R_{1n}^*+R_{2n}^*+R_{3n}^*+o_p(C_{V,b}^{1/2}),
\end{aligned}$$

where

$$\begin{aligned}
\eta_i(t) &= \int e^{\mathrm{i}s\cdot W_i}\{m^{\mathrm{ft}}(t-s)-m_\theta^{\mathrm{ft}}(t-s)\}\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds\,f_\epsilon^{\mathrm{ft}}(t) \\
&= c_n\int e^{\mathrm{i}s\cdot W_i}\Delta^{\mathrm{ft}}(t-s)\frac{K^{\mathrm{ft}}(sb)}{f_\epsilon^{\mathrm{ft}}(s)}ds\,f_\epsilon^{\mathrm{ft}}(t),
\end{aligned}$$

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

$$
\begin{aligned}
E[C_{V,b}^{-1/2}R_{1n}^*] &= \frac{(n-1)c_n^2}{(2\pi)^d C_{V,b}^{1/2}} \int\int\int |K^{\mathrm{ft}}(tb)|^2 K^{\mathrm{ft}}(s_1 b)K^{\mathrm{ft}}(s_2 b)\Delta^{\mathrm{ft}}(t-s_1)\Delta^{\mathrm{ft}}(s_2-t)f^{\mathrm{ft}}(s_1)f^{\mathrm{ft}}(-s_2)ds_1 ds_2 dt \\
&\equiv \Delta_n. \qquad (15)
\end{aligned}
$$

By the definition of $c_n$, $C_{V,b} = O(b^{-d(1+4\alpha)})$ (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 (4) yields

$$
\begin{aligned}
E[R_{1n}^{*2}] &= c_n^4 \int \cdots \int \frac{|K^{\mathrm{ft}}(t_1 b)|^2 |K^{\mathrm{ft}}(t_2 b)|^2 K^{\mathrm{ft}}(s_1 b)K^{\mathrm{ft}}(s_2 b)K^{\mathrm{ft}}(s_3 b)K^{\mathrm{ft}}(s_4 b)}{f_\epsilon^{\mathrm{ft}}(s_1)f_\epsilon^{\mathrm{ft}}(-s_2)f_\epsilon^{\mathrm{ft}}(s_3)f_\epsilon^{\mathrm{ft}}(-s_4)} f_W^{\mathrm{ft}}(s_1+s_3)f_W^{\mathrm{ft}}(-s_2-s_4) \\
&\quad \times \Delta^{\mathrm{ft}}(t_1-s_1)\Delta^{\mathrm{ft}}(s_2-t_1)\Delta^{\mathrm{ft}}(t_2-s_3)\Delta^{\mathrm{ft}}(s_4-t_2)ds_1\cdots ds_4 dt_1 dt_2 \\
&= O(b^{-d(1+4\alpha)}).
\end{aligned}
$$

Therefore, $\mathrm{Var}(C_{V,b}^{-1/2}R_{1n}^*) \to 0$ and we obtain $C_{V,b}^{-1/2}R_{1n}^* \xrightarrow{p} \lim_{n\to\infty}\Delta_n$. Finally, using similar arguments combined with $E[\xi_i(t)] = 0$, we can show that $C_{V,b}^{-1/2}R_{2n}^* \xrightarrow{p} 0$ and $C_{V,b}^{-1/2}R_{3n}^* \xrightarrow{p} 0$. 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

$$
\begin{aligned}
E[\varphi(b)R_{1n}^*] &= \varphi(b)(n-1)c_n^2 \int\int\int |K^{\mathrm{ft}}(tb)|^2 K^{\mathrm{ft}}(s_1 b)K^{\mathrm{ft}}(s_2 b) \\
&\quad \times \Delta^{\mathrm{ft}}(t-s_1)\Delta^{\mathrm{ft}}(s_2-t)f^{\mathrm{ft}}(s_1)f^{\mathrm{ft}}(-s_2)ds_1 ds_2 dt \\
&\equiv \Upsilon_n, \qquad (16)
\end{aligned}
$$

and the limit of $\Upsilon_n$ exists from the definition of $c_n$. Also, by similar treatment to $r_{1n}$ in the proof of Theorem 1 (ii), we can show $\mathrm{Var}(\varphi(b)R_{1n}^*) \to 0$ and thus $\varphi(b)R_{1n}^* \to \lim_{n\to\infty}\Upsilon_n$. Using similar arguments combined with $E[\xi_{1i}(t)] = 0$, 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 $c_n = b^{(\lambda-1)/2+\lambda\beta+\lambda_0}e^{\frac{1}{\mu b^\lambda}}$.

## References


[1] Azzalini, A., Bowman, A. W. and W. Härdle (1989) On the use of nonparametric regression for model checking, *Biometrika*, 76, 1-11.

[2] Bickel, P. J. and A. J. C. Ritov (1987) Efficient estimation in the errors-in-variables model, *Annals of Statistics*, 15, 513-540.

[3] Butucea, C. (2007) Goodness-of-fit testing an quadratic functional estimation from indirect observations, *Annals of Statistics*, 35, 1907-1930.

[4] Butucea, C. and M.-L. Taupin (2008) New M-estimators in semi-parametric regression with errors in variables, *Annales de l'Institut Henri Poincaré-Probabilités et Statistiques*, 44, 393-421.

[5] Carroll, R. J., Spiegelman, C. H., Lan, K. K. G., Bailey, K. T. and R. D. Abbott (1984) On errors-in-variables for binary regression models, *Biometrika*, 71, 19-25.

[6] Cheng, C.-L. and A. G. Kukush (2004) A goodness-of-fit test for a polynomial errors-in-variables model, *Ukrainian Mathematical Journal,* 56.4 641-661.

[7] Cheng, C.-L. and H. Schneeweiss (1998) Polynomial regression with errors in the variables, *Journal of the Royal Statistical Society: Series B,* 60, 189–199.

[8] Delaigle, A., Hall, P. and A. Meister (2008) On deconvolution with repeated measurements, *Annals of Statistics*, 36, 665-685.

[9] Delgado, M. A., Dominguez, M. A. and P. Lavergne (2006) Consistent tests of conditional moment restrictions, *Annales D'Economie et de Statistique*, 81, 33-67.

[10] Eubank, R. L. and C. H. Spiegelman (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques, *Journal of the American Statistical Association*, 85, 387-392.

[11] Fan, J. (1992) Deconvolution with Supersmooth Distributions, *Canadian Journal of Statistics*, 20, 155-169.

[12] Fan, J. and L. S. Huang (2001) Goodness-of-fit tests for parametric regression models, *Journal of the American Statistical Association*, 96, 640-652.

[13] Fan, Y. and Q. Li (2000) Consistent model specification tests: Kernel-based versus Bierens' ICM tests, *Econometric Theory*, 16, 1016-1041.

[14] Fan, J. and Y. K. Truong (1993) Nonparametric regression with errors in variables, *Annals of Statistics*, 4, 1900-1925.

[15] Fuller, W. A. (1987) *Measurement Error Models*, Wiley.

[16] Gleser, L. J. (1981) Estimation in a multivariate "errors in variables" regression model: large sample results, *Annals of Statistics*, 9, 24-44.

[17] Hall, P. (1984) Central limit theorem for integrated square error of multivariate nonparametric density estimators, *Journal of Multivariate Analysis*, 14, 1-16.

[18] Hall, P. and Y. Ma (2007) Testing the suitability of polynomial models in error-in-variables problems, *Annals of Statistics*, 35, 2620–2638.

[19] Härdle, W. and E. Mammen (1993) Comparing nonparametric versus parametric regression fits, *Annals of Statistics*, 21, 1926-1947.

[20] Hausman, J., Newey, W. and J. Powell (1995) Nonlinear Errors in Variables. Estimation of Some Engel Curves, *Journal of Econometrics*, 65, 205–233.

[21] Holzmann, H., Bissantz, N. and A. Munk (2007) Density testing in a contaminated sample, *Journal of multivariate analysis*, 98, 57-75.

[22] Holzmann, H. and L. Boysen (2006) Integrated square error asymptotics for supersmooth deconvolution, *Scandinavian Journal of Statistics*, 33, 849-860.

[23] Horowitz, J. L. and V. G. Spokoiny (2001) An adaptive, rate optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69, 599-631.

[24] Koul, H. L. and W.-X. Song (2009) Minimum distance regression model checking with Berkson measurement errors, *Annals of Statistics*, 37, 132-156.

[25] Koul, H. L. and W.-X. Song (2010) Model checking in partial linear regression models with Berkson measurement errors, *Statistica Sinica*, 20, 1551-1579.

[26] Leser, C. (1963) Forms of Engel Functions, *Econometrica*, 31, 694–703.

[27] Li, T. and Q. Vuong (1998) Nonparametric estimation of the measurement error model using multiple indicators, *Journal of Multivariate Analysis*, 65, 139-165.

[28] Ma, Y., Hart, J. D., Janicki, R. and R. J. Carroll (2011) Local and omnibus goodness-of-fit tests in classical measurement error models, *Journal of the Royal Statistical Society: Series B,* 73, 81–98

[29] Meister, A. (2009) *Deconvolution Problems in Nonparametric Statistics*, Springer.

[30] Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence, *Annals of Statistics*, 1-14.

[31] Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, Wiley.

[32] Schennach, S. (2004) Estimation of nonlinear models with measurement error, *Econometrica* 72, 33–75.

[33] Song, W.-X. (2008) Model checking in errors-in-variables regression, *Journal of Multivariate Analysis*, 99, 2406-2443.

[34] Song, W.-X. (2009) Lack-of-fit testing in errors-in-variables regression model with validation data, *Statistics & Probability Letters*, 79, 765-773.

[35] Taupin, M.-L. (2001) Semi-parametric estimation in the nonlinear structural errors-in-variables model, *Annals of Statistics*, 29, 66-93.

[36] van der Vaart, A. W. (1988) Estimating a real parameter in a class of semiparametric models, *Annals of Statistics*, 16, 1450-1474.

[37] van Es, A. J. and H.-W. Uh (2005) Asymptotic normality for kernel type deconvolution estimators, *Scandinavian Journal of Statistics*, 32, 467-483.

[38] Xu, W.-L. and L.-X. Zhu (2015) Nonparametric check for partial linear errors-in-covariables models with validation data, *Annals of Institute of Statistical Mathematics*, 67, 793-815.

[39] Zhu, L.-X. and H.-J. Cui (2005) Testing the adequacy for a general linear errors-in-variables model, *Statistica Sinica*, 15, 1049-1068.

[40] Zhu, L.-X. Song, W.-X. and H.-J. Cui (2003) Testing lack-of-fit for a polynomial errors-in-variables model, *Acta Mathematicae Applicatae Sinic*a, 19, 353-362.

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