



Minimum distance model checking in Berkson measurement error models with validation data

Pei Geng¹ · Hira L. Koul²

Received: 29 November 2017 / Accepted: 15 August 2018 / Published online: 23 August 2018
© Sociedad de Estadística e Investigación Operativa 2018

Abstract

This article studies a minimum distance regression model checking approach in the presence of Berkson measurement errors in covariates without specifying the measurement error density but when external validation data are available. The proposed tests are based on a class of minimized integrated square distances between a nonparametric estimate of the calibrated regression function and the parametric null model being fitted. The asymptotic distributions of these tests under the null hypothesis and against certain alternatives are established. Surprisingly, these asymptotic distributions are the same as in the case of known measurement error density. In comparison, the asymptotic distributions of the corresponding minimum distance estimators of the null model parameters are affected by the estimation of the calibrated regression function. A simulation study shows desirable performance of a member of the proposed class of estimators and tests.

Keywords Kernel smoothing · Integrated square distance · Two sample U statistic

Mathematics Subject Classification 62G10

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11749-018-0610-6>) contains supplementary material, which is available to authorized users.

✉ Pei Geng
pgeng@ilstu.edu
Hira L. Koul
koul@stt.msu.edu

¹ Illinois State University, Normal, IL, USA

² Michigan State University, East Lansing, MI, USA

1 Introduction

In statistical data analysis, the data are often collected subject to measurement errors. One typical way to treat the measurement error is the errors-in-variables model which assumes that the real observation Z is a surrogate of the true unobserved variable X , i.e., $Z = X + u$, where u is the measurement error. Regression models with measurement errors in covariates have received broad attention in the literature over the last century. In the last three decades, it has been the focus of numerous researchers, as is evidenced in the three monographs by Fuller (1987), Cheng and Ness (1999) and Carroll et al. (2006), and the references therein. However, as Berkson (1950) argued that in many situations it is more appropriate to treat the true unobserved variable X as the observed variable Z plus an error, i.e., $X = Z + \eta$, where now η is the measurement error. As an example, Wang (2003) mentions that in a chemical analysis, in order to study the effect of temperature to dry a sample on the resulting concentration of certain volatile matter, an oven is typically used to keep the samples at certain temperature. However, the actual temperature (X) in the oven can vary from the setup temperature (Z) due to the working mechanism of the oven. As a second example mentioned in Wang (2004), in order to study the yield of a crop in agriculture, an important covariate is the absorption amount of a fertilizer in the crop. Typically only the amount of fertilizer (Z) applied to the crop is observed. However, the actual amount of absorption (X) is not easily observed and may vary randomly around the amount Z applied to the crop. For more examples, see Carroll et al. (2006), Du et al. (2011), Schennach (2013), and the references therein.

Proceeding a bit more precisely, in the Berkson measurement error regression model of interest here, one has the triple X, Y, Z , obeying the relations

$$Y = \mu(X) + \varepsilon, \quad X = Z + \eta. \quad (1.1)$$

Here Y is a scalar response variable and ε is an error variable, independent of X , with $E\varepsilon = 0$, so that $\mu(x) = E(Y|X = x)$, $x \in \mathbb{R}^p$. The random vectors X, Z, η are p -dimensional, with X being the true unobservable covariate vector, Z representing an observation on X and η denoting the measurement error having $E\eta = 0$. We also assume that the three r.v.'s ε, Z, η are mutually independent. See Remark 2 for a further discussion on this assumption.

Let $\Theta \subset \mathbb{R}^q$ be a compact set, $\{m_\theta(x); \theta \in \Theta, x \in \mathbb{R}^p\}$ be a family of given functions and \mathcal{C} be a compact subset in \mathbb{R}^p . The problem of interest is to test

$$\begin{aligned} H_0 &: \mu(x) = m_{\theta_0}(x), \quad \text{for some } \theta_0 \in \Theta \text{ and all } x \in \mathcal{C}, \quad \text{versus} \\ H_1 &: H_0 \text{ is not true,} \end{aligned}$$

based on the primary sample $\{(Z_i, Y_i), i = 1, \dots, n\}$ and an independent validation sample $\{(\tilde{Z}_k, \tilde{X}_k), k = 1, \dots, N\}$, all satisfying (1.1). Then the empirical version of η is naturally obtained by $\tilde{\eta}_k := \tilde{X}_k - \tilde{Z}_k$, $1 \leq k \leq N$.

Koul and Song (2009) provide a class of tests for the above testing problem when f_η is known. To describe their tests, assume $E|\mu(X)| < \infty$, $E|m_\theta(X)| < \infty$, for all

$\theta \in \Theta$, and define $H(z) := E[\mu(X)|Z = z]$, $H_\theta(z) := E[m_\theta(X)|Z = z]$, $z \in \mathbb{R}^p$. Then the original model is transformed to

$$Y = H(Z) + \xi, \quad E(\xi|Z) = 0,$$

and the hypothesis H_0 implies

$$H'_0 : H(z) = H_{\theta_0}(z), \text{ for some } \theta_0 \in \Theta \text{ and all } z \in \mathcal{C}, \text{ versus } H'_1 : H'_0 \text{ is not true.}$$

Note that known f_η implies that H_θ is a known parametric function.

Next, let $w \equiv w_n = c(\log n/n)^{1/(p+4)}$, $c > 0$, and $h \equiv h_n$ be two bandwidth sequences depending on n , K be a density kernel and G be a nondecreasing right continuous real-valued function on \mathbb{R}^p and define

$$K_{hi}(z) = \frac{1}{h^p} K\left(\frac{z - Z_i}{h}\right), \quad \hat{f}_w(z) = \frac{1}{n} \sum_{i=1}^n K_{wi}(z),$$

$$M_n(\theta) = \int_{\mathcal{C}} \left[\frac{1}{n \hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z) [Y_i - H_\theta(Z_i)] \right]^2 dG(z), \quad \tilde{\theta}_n = \operatorname{argmin}_\theta M_n(\theta).$$

The class of tests, one for each K and G , proposed in Koul and Song (2009) (KS) is based on the class of minimized integrated square distances $M_n(\tilde{\theta}_n)$. They establish the consistency and asymptotic normality of suitably standardized $\tilde{\theta}_n$ and $M_n(\tilde{\theta}_n)$.

In the current paper, we extend this minimum distance (m.d.) methodology to the case when f_η is unknown, but when external validation data are available. The lack of knowledge of f_η makes H_θ to be an unknown function. The validation data $\{(\tilde{Z}_k, \tilde{X}_k), k = 1, \dots, N\}$ is used to estimate H_θ , which in turn is used to construct a class of tests analogous to M_n . More precisely, let

$$\hat{H}_\theta(z) = N^{-1} \sum_{k=1}^N m_\theta(z + \tilde{\eta}_k), \quad \tilde{\eta}_k := \tilde{Z}_k - \tilde{X}_k, \quad 1 \leq k \leq N,$$

$$\hat{M}_n(\theta) = \int_{\mathcal{C}} \left[\frac{1}{n \hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z) [Y_i - \hat{H}_\theta(Z_i)] \right]^2 dG(z), \quad \hat{\theta}_n = \operatorname{argmin}_\theta \hat{M}_n(\theta).$$

The proposed class of tests investigated in this paper is based on $\hat{M}_n(\hat{\theta}_n)$.

We establish the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$ under H_0 and that of suitably standardized $\hat{M}_n(\hat{\theta}_n)$ under H_0 and under a sequence of local alternatives. We also discuss the choice of the optimal G that maximizes the asymptotic power against a given sequence of local alternatives.

A surprising finding is that the asymptotic distributions of suitably standardized $\hat{M}_n(\hat{\theta}_n)$ under H_0 and under certain sequences of local alternatives are the same as those of the similarly standardized $M_n(\tilde{\theta}_n)$, see Theorems 2 and 5 below. In comparison, Theorem 1 below shows that the asymptotic distributions of $n^{1/2}(\hat{\theta}_n - \theta_0)$ and $n^{1/2}(\tilde{\theta}_n - \theta_0)$, under H_0 , are different, in general.

If $p = q$ and $m_\theta(x) = \theta^T x$ is linear, then $H_\theta(z) = \theta^T z$ is a known function, so there is no need to estimate it and one can use $M_n(\hat{\theta}_n)$ to fit a linear model to $\mu(x)$ regardless of the knowledge of f_η . Moreover, Proposition 1 below proves that in this case, $n^{1/2}(\hat{\theta}_n - \tilde{\theta}_n) \rightarrow_p 0$. However, as pointed out in Example 1 below, the need to estimate f_η cannot be avoided even when the regression model is a polynomial of order 2 or more, a small departure from the linearity.

The paper is organized as follows. Section 2 describes the needed assumptions for the derivation of the consistency and asymptotic normality of the proposed estimators and tests. Section 3 establishes the consistency and asymptotic normality of the m.d. estimators, while in Sect. 4 we state the main results about the m.d. tests under the null and certain fixed and local alternative hypotheses. The proofs of many results stated in Sects. 3 and 4 appear in the supplement of this paper. Section 5 reports the findings of a simulation study that assesses some finite sample properties of an estimator and a test in the proposed classes of these inference procedures.

In this paper, $\mathcal{N}_q(\nu, \Sigma)$ stands for the q -variate normal distribution with mean vector ν and covariance matrix Σ , x^T denotes the transpose of an Euclidean vector x , $\|\cdot\|$ denotes the Euclidean norm, all limits are taken as $N \wedge n \rightarrow \infty$, unless mentioned otherwise, and \rightarrow_d (\rightarrow_p) denotes the convergence in distribution (probability).

2 Assumptions under H_0

In this section, we shall describe the assumptions for the asymptotic normality of suitably standardized $\hat{\theta}_n$ and $\widehat{M}_n(\hat{\theta}_n)$, under H_0 . Many of these assumptions are the same as in KS. Define, for $x, y \in \mathbb{R}^p$ and $\theta \in \Theta$,

$$\sigma_\theta(x, y) := \text{Cov}(m_\theta(x + \eta), m_\theta(y + \eta)), \quad \sigma_\theta^2(x) := \sigma_\theta(x, x) = \text{Var}(m_\theta(x + \eta)).$$

All the integrals with respect to the measure G are supposed to be over the compact set \mathcal{C} , unless specified otherwise. We are now ready to state the needed assumptions.

- (A1) $\{(Y_i, Z_i), Z_i \in \mathbb{R}^p, i = 1, \dots, n\}$ is an i.i.d. sample with regression function $H(z) = E(Y|Z = z)$ satisfying $\int H^2 dG < \infty$. The integrating measure G has continuous Lebesgue density g on \mathcal{C} . The validation data $\{(\tilde{Z}_k, \tilde{X}_k), \tilde{Z}_k \in \mathbb{R}^p, \tilde{X}_k \in \mathbb{R}^p, k = 1, \dots, N\}$ is an i.i.d. sample from Berkson measurement error model $X = Z + \eta$. Furthermore, the two samples are independent.
- (A2) $0 < \sigma_\varepsilon^2 := \text{Var}(\varepsilon) < \infty, \tau^2(z) = E[(m_{\theta_0}(X) - H_{\theta_0}(Z))^2 | Z = z]$ is continuous on \mathcal{C} .
- (A3) For some $\delta > 0, E|\varepsilon|^{2+\delta} + E|m_{\theta_0}(X) - H_{\theta_0}(Z)|^{2+\delta} < \infty$.
- (A4) $E|\varepsilon|^4 + E|m_{\theta_0}(X) - H_{\theta_0}(Z)|^4 < \infty$.
- (A5) $\int \sigma_\theta^2(z) dG(z) < \infty$, for all $\theta \in \Theta$.
- (F1) The density f_Z is uniformly continuous on \mathcal{C} and $\inf_{z \in \mathcal{C}} f(z) > 0$.
- (F2) The density f_Z is twice continuously differentiable on \mathcal{C} .
- (H1) $m_\theta(x)$ is a.e. continuous in x , for every $\theta \in \Theta$.
- (H2) The parametric function family $H_\theta(z)$ is identifiable with respect to θ , i.e., $H_{\theta_1}(z) = H_{\theta_2}(z)$ a.e. in z implies $\theta_1 = \theta_2$.

- (H3) For some positive continuous function r on \mathcal{C} , and for some $0 < \beta \leq 1$, $|H_{\theta_1}(z) - H_{\theta_2}(z)| \leq \|\theta_1 - \theta_2\|^\beta r(z)$, for all $\theta_1, \theta_2 \in \Theta$ and $z \in \mathcal{C}$.
- (H4) For each x , $m_\theta(x)$ is differentiable with respect to θ in a neighborhood of θ_0 with the derivative vector $\dot{m}_\theta(x)$ such that for any consistent estimator θ_n of θ_0 ,

$$\sup_i \frac{|\frac{1}{N} \sum_{k=1}^N [m_{\theta_n}(Z_i + \tilde{\eta}_k) - m_{\theta_0}(Z_i + \tilde{\eta}_k) - (\theta_n - \theta_0)^T \dot{m}_{\theta_0}(Z_i + \tilde{\eta}_k)]|}{\|\theta_n - \theta_0\|} = o_p(1),$$

where the supremum is taken over $1 \leq i \leq n$.

- (H5) The vector function $\dot{m}_{\theta_0}(x)$ is continuous in $x \in \mathcal{C}$ and for every $\epsilon > 0$, there are n_ϵ, N_ϵ such that for every $0 < a < \infty$, and for all $n > n_\epsilon, N > N_\epsilon$,

$$P \left(\max_{1 \leq i \leq n, 1 \leq k \leq N, (nh^p)^{1/2} \|\theta - \theta_0\| \leq a} h^{-p/2} \|\dot{m}_\theta(Z_i + \tilde{\eta}_k) - \dot{m}_{\theta_0}(Z_i + \tilde{\eta}_k)\| \geq \epsilon \right) \leq \epsilon.$$

- (H6) $\int \|\dot{H}_{\theta_0}\|^2 dG < \infty$ and $\Sigma_0 = \int \dot{H}_{\theta_0} \dot{H}_{\theta_0}^T dG$ is positive definite.

- (K) The density kernel K is positive, symmetric and square integrable on $[-1, 1]^p$.
- (W1) $nh^{2p} \rightarrow \infty$ and $N/n \rightarrow \lambda, \lambda > 0$.
- (W2) $h \sim n^{-a}$, where $0 < a < \min(1/2p, 4/(p(p + 4)))$.

We state some important facts that will be used later. From Mack and Silverman (1982), we obtain that under (F1), (K1), (W1) and (W2),

$$\begin{aligned} \sup_{z \in \mathcal{C}} |\hat{f}_h(z) - f_Z(z)| &= o_p(1), & \sup_{z \in \mathcal{C}} |\hat{f}_w(z) - f_Z(z)| &= o_p(1), \\ \sup_{z \in \mathcal{C}} \left| \frac{f_Z^2(z)}{\hat{f}_w^2(z)} - 1 \right| &= o_p(1). \end{aligned} \tag{2.1}$$

Let $d\varphi = f_Z^{-2} dG$, $d\hat{\varphi} = \hat{f}_w^{-2} dG$. We also recall the following facts from Koul and Ni (2004) (KN) and KS. For any continuous function α on \mathcal{C} , $\int |\alpha(z)| d\varphi(z) < \infty$, and by (2.1),

$$\int \alpha(z) d\hat{\varphi}(z) = \int \alpha(z) d\varphi(z) + o_p \left(\int |\alpha(z)| d\varphi(z) \right). \tag{2.2}$$

Using the equation (3.9) on page 117 of Koul and Ni (2004), for any α as above, (F1), (K1) and (W1) imply

$$\int E \left\{ \frac{1}{n} \sum_{i=1}^n K_h(z - Z_i) \alpha(Z_i) \right\}^2 d\varphi(z) = \int \alpha^2(z) dG(z) + o(1) = O(1). \tag{2.3}$$

3 Estimation of θ_0

In this section, we establish the consistency of $\hat{\theta}_n$ and asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta_0)$, under H_0 . Let

$$W_n(\theta) = \int_{\mathcal{C}} \left[\frac{1}{n \hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z) [\hat{H}_\theta(Z_i) - H_\theta(Z_i)] \right]^2 dG(z).$$

This quantity is a measure of the essential difference between $\hat{M}_n(\theta)$ and $M_n(\theta)$ as is seen in the following decomposition:

$$\begin{aligned} \hat{M}_n(\theta) &= \int \left[\frac{1}{n \hat{f}_w(z)} \sum_{i=1}^n K_{hi}(z) [Y_i - H_\theta(Z_i) + H_\theta(Z_i) - \hat{H}_\theta(Z_i)] \right]^2 dG(z) \\ &= M_n(\theta) + W_n(\theta) + 2R_n(\theta), \end{aligned}$$

where $R_n(\theta)$ is the cross product term. The following lemma about W_n is found to be useful in deriving various results in the sequel. Let

$$\gamma(\theta) = \int \int \sigma_\theta^2(x, y) dG(x) dG(y), \quad A_N(\theta) = \frac{1}{N} \int \sigma_\theta^2(z) dG(z).$$

Lemma 1 *Suppose (A1), (A2), (A5), (F1), (H1), (K), and (W1) hold. Then for every $\theta \in \Theta$ for which $\mu(x) = m_\theta(x)$, $x \in \mathcal{C}$,*

$$N(W_n(\theta) - A_N(\theta)) \rightarrow_d \mathcal{N}_1(0, 2\gamma(\theta)).$$

3.1 Consistency of $\hat{\theta}_n$

In this subsection, we shall establish the consistency of the m.d. estimators $\hat{\theta}_n$ for θ_0 . Many details below are similar to those in KN and KS. For a $v \in L_2(G)$, let

$$\rho(v, H_\theta) = \int (v - H_\theta)^2 dG, \quad T(v) = \operatorname{argmin}_\theta \rho(v, H_\theta). \quad (3.1)$$

The proof of the following lemma is similar to that of Lemma 3.2 and Corollary 3.1 of KS. Details are left out for the sake of brevity.

Lemma 2 *Suppose (A1), (A2), (A5), (F1), (H1), (H3), (K) and (W1) hold. If $T(H)$ is unique, then $\hat{\theta}_n = T(H) + o_p(1)$. If, in addition H_0 and (H2) hold, then $\hat{\theta}_n \rightarrow_p \theta_0$.*

3.2 Asymptotic normality of $\hat{\theta}_n$

Here we present the asymptotic normality result about $\hat{\theta}_n$ under H_0 .

Theorem 1 Suppose (A1)–(A3), (A5), (F1), (F2), (H1)–(H6), (K), (W1), (W2) and H_0 hold. Then $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d \mathcal{N}_q\left(0, \Sigma_0^{-1}(\Sigma_1 + \lambda^{-1}\Sigma_2)\Sigma_0^{-1}\right)$, where Σ_0 is given in (H6) and

$$\begin{aligned} \Sigma_1 &= \int \frac{(\sigma_\varepsilon^2 + \tau^2(u))\dot{H}_{\theta_0}(u)\dot{H}_{\theta_0}^T(u)g^2(u)}{f_Z(u)}du, \\ \Sigma_2 &= \int \sigma_{\theta_0}(x, y)\dot{H}_{\theta_0}(x)\dot{H}_{\theta_0}^T(y)dG(x)dG(y). \end{aligned} \tag{3.2}$$

Note that the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is determined by Σ_1 , Σ_2 and the limit of N/n . The matrix Σ_1 represents the variation due to Berkson measurement error when f_η is known. From Koul and Song (2009), we recall that $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d \mathcal{N}_q(0, \Sigma_0^{-1}\Sigma_1\Sigma_0^{-1})$. The matrix Σ_2 represents the additional variation due to the estimation of H_θ by \hat{H}_θ when f_η is unknown. Moreover, the covariance tends to decay as N/n increases. When $N/n \rightarrow \infty$, in other words, when the validation sample size N is sufficiently large, compared to the primary sample size n , not surprisingly the above asymptotic covariance degenerates to the case of known f_η .

Remark 1 Here we shall verify that the quantities Σ_1 and Σ_2 of (3.2) are well defined under the given assumptions. By (A2) and the compactness of \mathcal{C} , $\tau^2(u)$ is bounded on \mathcal{C} . Assumption (H6) further implies that $0 < a^T \Sigma_1 a < \infty$, for all $a \in \mathbb{R}^q$.

Next, consider Σ_2 . The Cauchy–Schwarz inequality implies that $\sigma_\theta(x, y) \leq \sigma_\theta(x)\sigma_\theta(y)$ for all $x, y \in \mathbb{R}$, $\theta \in \Theta$, and that for any $a \in \mathbb{R}^q$,

$$\begin{aligned} |a^T \Sigma_2 a| &\leq \int |\sigma_{\theta_0}(x, y)| |a^T \dot{H}_{\theta_0}(x)\dot{H}_{\theta_0}^T(y)a|dG(x)dG(y) \\ &\leq \int \sigma_{\theta_0}(x)\sigma_{\theta_0}(y) \|a^T \dot{H}_{\theta_0}(x)\| \|a^T \dot{H}_{\theta_0}(y)\|dG(x)dG(y) \\ &= \left(\int \sigma_{\theta_0}(x) \|a^T \dot{H}_{\theta_0}(x)\|dG(x) \right)^2 \\ &\leq \|a\|^2 \int \sigma_{\theta_0}^2(x)dG(x) \int \|H_{\theta_0}(x)\|^2dG(x). \end{aligned}$$

Hence assumptions (A5) and (H6) ensure that the entries of Σ_2 exist and are finite. Moreover, as seen in the proof of the theorem in the supplement, Σ_2 is positive definite.

Remark 2 Here we shall discuss some of the assumptions of Sect. 2, give examples of m_θ 's that satisfy the assumptions (H3)–(H5) and provide explicit expressions of Σ_2 .

In the Berkson model (1.1), ε and η are typically assumed to be independent to ensure its identifiability, while Z and η can be correlated. One of the entities we have to estimate consistently for implementing our methodology is $H_\theta(z)$. Under the assumption of the independence of Z and η , it is relatively easy to show that the proposed estimator $\hat{H}_\theta(z)$ is consistent for $H_\theta(z)$. In the case of correlated Z and η , one could use a Nadaraya–Watson type estimator to estimate $H_\theta(z)$ instead of $\hat{H}_\theta(z)$. However we refrain from doing this in the current paper for the sake of brevity.

As for the validation sample assumed in (A1), in fact, there are two types of validation data in reality. The first type is that an external validation sample is collected after the primary study due to various factors, such as the unawareness of measurement errors in the primary study. In this case, the validation study is usually carried out on a different sample from the primary sample. Hence it is rather clear that the validation data can be treated independent of the primary data as assumed in (A1). A real data example dealing with breast cancer study can be found in Yi et al. (2015). Based on the independence, the theory developed for two sample statistics in Sepanski and Lee (1995) and Geng and Koul (2017) can be applied to derive the asymptotic results. The second type is that the validation study is carried out simultaneously with the primary study; thus, the validation sample is a subset of the primary sample. In this case, the minimum distance idea is applicable; however, the results in this paper would not hold due to the lack of independence between the two samples. Hence different assumptions and theory should be investigated for the second type of validation data. This paper focuses on the first type of validation data based on (A1).

Regarding (H2) of the identifiability of H_θ , various conditions can be imposed to ensure (H2) hold. For instance, the following two assumptions together imply (H2):

- (1) $m_{\theta_1}(x) = m_{\theta_2}(x)$, for a.e. $x \in \mathbb{R}^p$, implies $\theta_1 = \theta_2$.
- (2) The location family $\{f_\eta(\cdot - z), z \in \mathbb{R}^p\}$ is complete. More details can be found in Koul and Song (2009). Furthermore, it can be easily verified that all the examples of m_θ 's given below satisfy (1).

Example 1 (The linear and polynomial cases) Suppose $q = p$, $m_\theta(x) = \theta^T x$, $\theta, x \in \mathbb{R}^p$ and $E|X| < \infty$, where for any vector $x = (x_1, \dots, x_p)^T \in \mathbb{R}^p$, $|x| = \sum_{j=1}^p |x_j|$. Then $H_\theta(z) = \theta^T z$ is a known function. In this case, there is no need to estimate this function and one can also use $\hat{\theta}_n$ as a m.d. estimator of θ . This fact in spirit is similar to the fact that the classical least square estimators, when regressing Y on Z , continue to be unbiased and consistent in the Berkson linear model. See Remark 3 for an asymptotic equivalence between $\hat{\theta}_n$ and $\tilde{\theta}_n$.

In the polynomial regression of order p , $q = p + 1$ and $m_\theta(x) = \theta^T \ell(x)$, $x \in \mathbb{R}$, where $\theta = (\theta_1, \dots, \theta_{p+1})^T$ and $\ell(x) := (1, x, \dots, x^p)^T$ such that $E\|\ell(X)\| < \infty$. Then

$$L(z) := E(\ell(X)|Z = z) = (1, z, E(z + \eta)^2, \dots, E(z + \eta)^p)^T, \quad H_\theta(z) = \theta^T L(z).$$

This model is a simple deviation from the linear model, yet one already sees the need to estimate $H_\theta(z)$. Given the validation data, an estimate of $H_\theta(z)$ in this case is given by

$$\begin{aligned} \hat{H}_\theta(z) &= \frac{1}{N} \sum_{k=1}^N m_\theta(z + \tilde{\eta}_k) = \frac{1}{N} \sum_{k=1}^N [\theta_1 + \theta_2(z + \tilde{\eta}_k) + \theta_3(z + \tilde{\eta}_k)^2 \\ &\quad + \dots + \theta_{p+1}(z + \tilde{\eta}_k)^p]. \end{aligned}$$

Here (H3) is satisfied with $r = L$. Furthermore, $\dot{m}_\theta(x) = \ell(x)$ and $\dot{H}_\theta(z) = L(z)$ for all $\theta \in \Theta$. Therefore, similar to the linear case, (H4) and (H5) hold. Moreover,

$$\sigma_\theta(x, y) = \theta^T [E\ell(x + \eta)\ell^T(y + \eta) - L(x)L^T(y)]\theta,$$

$$\Sigma_2 = \int \theta_0^T [E\ell(x + \eta)\ell^T(y + \eta) - L(x)L^T(y)]\theta_0 L(x)L^T(y)dG(x)dG(y).$$

Example 2 (The nonlinear case) In biochemistry, one of the well-known models for enzyme kinetics relates enzyme reaction rate to the concentration of a substrate x by the formula $\alpha_0 x / (\theta + x)$, $\alpha_0 > 0, \theta > 0, x > 0$. This is the so-called Michaelis–Menten model, see Bates and Watts (1998). The ratio $\gamma_0 = \alpha_0 / \theta$ is defined as the catalytic efficiency that measures how efficiently an enzyme converts a substrate into product. When γ_0 is known, the function can be written as

$$m_\theta(x) := \frac{\gamma_0 \theta x}{\theta + x}, \quad \theta > 0, \quad x > 0. \tag{3.3}$$

We will verify that this nonlinear function satisfies (H3)–(H5).

Regarding (H3), as argued in KS, it suffices to verify that (H3) holds with H_θ replaced by m_θ . Here, direct calculation shows that

$$|m_{\theta_1}(x) - m_{\theta_0}(x)| = \frac{\gamma_0 x^2 |\theta_1 - \theta_0|}{(\theta_0 + x)(\theta_1 + x)} \leq \gamma_0 |\theta_1 - \theta_0|.$$

Hence (H3) holds for the m_θ of (3.3).

Furthermore, suppose for each $x \in \mathbb{R}^p$, the $q \times q$ matrix $\ddot{m}_\theta(x) := \partial^2 m_\theta(x) / \partial \theta^2$ exists for all θ in a neighborhood U_0 of θ_0 and $\|\ddot{m}_\theta(x)\| \leq C$, for all $\theta \in U_0$ and $x \in \mathbb{R}^p$, where the constant C may depend on θ_0 . Then, by the mean value theorem, with probability 1, for all $1 \leq i \leq n, N \geq 1$,

$$\frac{|\frac{1}{N} \sum_{k=1}^N [m_\theta(Z_i + \tilde{\eta}_k) - m_{\theta_0}(Z_i + \tilde{\eta}_k) - (\theta - \theta_0)^T \dot{m}_{\theta_0}(Z_i + \tilde{\eta}_k)]|}{\|\theta - \theta_0\|} \leq C \|\theta - \theta_0\|.$$

Now apply this with $\theta = \theta_n$, where θ_n is any consistent estimator of θ_0 , to conclude that (H4) holds.

In particular, for the function m_θ of (3.3), $p = 1 = q$ and $\ddot{m}_\theta(x) = -2\gamma_0 x^2 / (\theta + x)^3$ is bounded for $\theta > 0$ and $x > 0$, so (H4) holds in this case.

As for (H5), with $\sqrt{nh^p}|\theta - \theta_0| \leq a$ and θ_1^* falling between θ and θ_0 , we have

$$\sup_{i,k,\theta} h^{-p/2} |\dot{m}_\theta(Z_i + \tilde{\eta}_k) - \dot{m}_{\theta_0}(Z_i + \tilde{\eta}_k)| = \sup_{i,k,\theta^*} h^{-p/2} |\ddot{m}_{\theta^*}(Z_i + \tilde{\eta}_k)(\theta - \theta_0)|$$

$$\leq \sup_{\theta} C' h^{-p/2} |\theta - \theta_0| = O_p(h^{-p/2} / \sqrt{nh^p}) = O_p(1 / (\sqrt{nh^p})) = o_p(1),$$

where C' is the upper bound for the second derivative $\ddot{m}_\theta(x)$. Therefore, (H5) is satisfied.

Another nonlinear example is the exponential function $m_\theta(x) = e^{\theta x}$, with $\theta, x \in \mathbb{R}$. In practice, it is reasonable to assume that both Θ and the domain of X are bounded subsets in \mathbb{R} , i.e., $|\theta| \leq C_1$ and $|x| \leq C_2$. To verify (H3), it suffices to show that this condition holds with $H_\theta(z)$ replaced by $m_\theta(x)$. With θ^* falling between θ_1 and θ_2 , we obtain

$$|m_{\theta_2}(x) - m_{\theta_1}(x)| = |\dot{m}_{\theta^*}(x)(\theta_2 - \theta_1)| \leq (|x|e^{C_1|x|})|\theta_2 - \theta_1| := r(x)|\theta_2 - \theta_1|.$$

Therefore, (H3) holds for the exponential regression function. Moreover, the second derivative $\ddot{m}_\theta(x) = x^2 e^{\theta x}$ is bounded by the constant $C_1^2 e^{C_1 C_2}$. Hence the argument similar to that for (3.3) yields that the exponential function also satisfies (H4) and (H5). The expression of Σ_2 is given in Sect. 5.1.

Remark 3 (Connection between $\hat{\theta}_n$ and $\tilde{\theta}_n$ in linear regression) Here we shall show, under some conditions, that $\hat{\theta}_n - \tilde{\theta}_n = o_p(n^{-1/2})$ in the linear regression model

$$p = q, \quad \mu(x) = m_\theta(x) = \theta^T x, \quad x \in \mathcal{C} \subset \mathbb{R}^p, \quad \text{for some } \theta_0 \in \Theta \subset \mathbb{R}^p. \quad (3.4)$$

In this case, $H_\theta(z) = \theta^T z$ and by solving the equation $\partial \widehat{M}_n(\theta) / \partial \theta = 0$, we obtain $B_n \hat{\theta}_n = A_n$, where

$$A_n = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) Y_i \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) (Z_i + \bar{\eta}) \right] d\hat{\varphi}(z),$$

$$B_n = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) (Z_i + \bar{\eta}) \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) (Z_i + \bar{\eta})^T \right] d\hat{\varphi}(z),$$

with $\bar{\eta} = N^{-1} \sum_{k=1}^N \tilde{\eta}_k$. Similarly, $\tilde{B}_n \tilde{\theta}_n = \tilde{A}_n$, where

$$\tilde{A}_n = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) Y_i \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) Z_i \right] d\hat{\varphi}(z),$$

$$\tilde{B}_n = \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) Z_i \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) Z_i^T \right] d\hat{\varphi}(z).$$

Roughly speaking, because $\bar{\eta} \rightarrow_p 0$, $A_n - \tilde{A}_n = o_p(1)$, $B_n - \tilde{B}_n = o_p(1)$ and hence $\hat{\theta}_n - \tilde{\theta}_n \rightarrow_p 0$. Furthermore, under some specific conditions, both $\hat{\theta}_n$ and $\tilde{\theta}_n$ can achieve the same asymptotic efficiency. We present two such assumptions here.

(A6) $E\eta^2 < \infty$, $\tau_1(z) := E(|\varepsilon| | Z = z)$ is a.e. (G) continuous.

(A7) $\nu_G := \int z dG(z) = 0$, $\int z z^T dG(z)$ is positive definite.

Proposition 1 Suppose (1.1) and (3.4) hold. In addition suppose (A1), (F1), (K), (W1), (A6) and (A7) hold, then $\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) \rightarrow_p 0$.

Proof For the simplicity of the exposition, we give details for the case $p = 1$ only. Then $\tilde{B}_n = \int [n^{-1} \sum_{i=1}^n K_{hi}(z)Z_i]^2 d\hat{\varphi}(z)$. By (2.1), (2.2), (2.3) and direct calculations, $\tilde{B}_n = \kappa_G + o_p(1)$, where $\kappa_G = \int z^2 dG(z)$. By (A7), $\kappa_G > 0$. Then $\tilde{\theta}_n = \tilde{B}_n^{-1} \tilde{A}_n$ is well defined for all sufficiently large n , and the consistency of $\tilde{\theta}_n$ yields that $\tilde{A}_n = O_p(1)$. We shall shortly show that

$$(a) \quad \sqrt{n}(A_n - \tilde{A}_n) = o_p(1), \quad (b) \quad \sqrt{n}(B_n - \tilde{B}_n) = o_p(n^{-1/2}). \quad (3.5)$$

Then for all sufficiently large n , $\hat{\theta}_n = B_n^{-1} A_n$ and

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) &= \frac{\sqrt{n}(A_n \tilde{B}_n - \tilde{A}_n B_n)}{B_n \tilde{B}_n} = \frac{\sqrt{n}[A_n \tilde{B}_n - \tilde{A}_n(\tilde{B}_n + o_p(n^{-1/2}))]}{\tilde{B}_n(\tilde{B}_n + o_p(n^{-1/2}))} \\ &= \frac{\sqrt{n}(A_n - \tilde{A}_n)\tilde{B}_n - o_p(\tilde{A}_n)}{\kappa_G^2 + o_p(1)} = o_p(1). \end{aligned}$$

To prove (3.5) (a), rewrite

$$\sqrt{n}(A_n - \tilde{A}_n) = \sqrt{n\bar{\eta}} \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z)Y_i \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) \right] d\hat{\varphi}(z) := \sqrt{n\bar{\eta}} \mathcal{A}_n.$$

By (A6) and the central limit theorem, $\sqrt{n\bar{\eta}} = O_p(1)$. It thus suffices to show that $\mathcal{A}_n = o_p(1)$. Let \mathcal{A}_n^* denote the \mathcal{A}_n with $\hat{\varphi}$ replaced by φ . Then (2.2), $E(|Y||Z = z) \leq |\theta_0^T z| + \tau_1(z)$, assumption (A6) and direct calculations yield that

$$|\mathcal{A}_n - \mathcal{A}_n^*| = o_p \left(\int \frac{1}{n^2} \sum_{i,j=1}^n K_{hi}(z)K_{hj}(z)|Y_i|d\varphi(z) \right) = o_p(1).$$

Next, rewrite

$$\begin{aligned} \mathcal{A}_n^* &= \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z)Y_i d\varphi(z) + \frac{1}{n^2} \sum_{i=1}^n \sum_{j \neq i=1}^n \int K_{hi}(z)K_{hj}(z)Y_i d\varphi(z) \\ &:= \mathcal{A}_{n1} + \mathcal{A}_{n2}. \end{aligned}$$

Calculation of moments shows that $E(\mathcal{A}_{n1}) = O((nh)^{-1})$, $E(\mathcal{A}_{n2}) = \theta_0 \nu_G + o(1)$, $\text{Var}(\mathcal{A}_{n1}) = O(n^{-3}h^{-2})$ and $\text{Var}(\mathcal{A}_{n2}) = O(n^{-1})$. Hence $\mathcal{A}_n^* = \theta_0 \nu_G + o_p(1)$, and (A7) implies (3.5)(a).

Now we prove (3.5)(b). Let

$$B_n := \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z)Z_i \right] \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) \right] d\hat{\varphi}(z).$$

Then, by (2.3),

$$B_n - \tilde{B}_n = 2\bar{\eta}B_n + \bar{\eta}^2 \int \left[\frac{1}{n} \sum_{i=1}^n K_{hi}(z) \right]^2 d\hat{\varphi}(z) = 2\bar{\eta}B_n + O_p(n^{-1}).$$

Argue as in the analysis of \mathcal{A}_n to obtain that $B_n = v_G + o_p(1)$. This fact and $\sqrt{n}\bar{\eta} = O_p(1)$ imply that $\sqrt{n}(B_n - \tilde{B}_n) = 2(\sqrt{n}\bar{\eta})v_G + O_p(n^{-1/2})$, which together with (A7) imply (3.5)(b). This also completes the proof of the proposition.

4 Testing

In this section, we establish the asymptotic distributions of the m.d. test statistics $\hat{M}_n(\hat{\theta}_n)$ under the null and certain fixed and local alternative hypotheses. Let

$$\begin{aligned} \xi_i &= Y_i - H_{\theta_0}(Z_i), \quad \hat{\xi}_i = Y_i - \hat{H}_{\hat{\theta}_n}(Z_i), \\ \tilde{C}_n &= \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z) \xi_i^2 d\varphi(z), \quad \tilde{\Gamma}_n = \frac{2h^p}{n^2} \sum_{i \neq j} \left(\int K_{hi}(z) K_{hj}(z) \xi_i \xi_j d\varphi(z) \right)^2, \\ \hat{C}_n &= \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z) \hat{\xi}_i^2 d\hat{\varphi}(z), \quad \hat{\Gamma}_n = \frac{2h^p}{n^2} \sum_{i \neq j} \left(\int K_{hi}(z) K_{hj}(z) \hat{\xi}_i \hat{\xi}_j d\hat{\varphi}(z) \right)^2, \\ \hat{\mathcal{T}}_n &:= nh^{p/2} \hat{\Gamma}_n^{-1/2} (\hat{M}_n(\hat{\theta}_n) - \hat{C}_n). \end{aligned}$$

Because $\xi = Y - H_{\theta_0}(Z) = \varepsilon + m_{\theta_0}(X) - H_{\theta_0}(Z)$ and because Z, η and ε are mutually independent, $E(\xi^2|Z = z) = \sigma_\varepsilon^2 + \tau^2(z)$, where τ^2 is as in (A2). Since \mathcal{C} is compact, the continuity of τ^2 implies that it is bounded on \mathcal{C} and hence $\int E(\xi^2|Z = z)dG(z) < \infty$.

4.1 Asymptotic null distribution of $\hat{\mathcal{T}}_n$

The following theorem states the asymptotic distribution of the proposed m.d. tests under the null hypothesis H_0 .

Theorem 2 *Suppose (A1), (A2), (A4), (A5), (F1)–(F2), (K), (H1)–(H6), (W1) and (W2) hold. Then, under H_0 , $\hat{\mathcal{T}}_n \rightarrow_d \mathcal{N}_1(0, 1)$.*

Consequently, the test that rejects H_0 whenever $|\hat{\mathcal{T}}_n| > z_{\alpha/2}$ is of the asymptotic size α , where z_α is the upper 100α th percentile of $\mathcal{N}_1(0, 1)$.

Theorem 2 shows that the ratio parameter N/n does not play a role in the limiting null distribution of $\hat{\mathcal{T}}_n$. This finding is also reflected in a finite sample simulation study of Sect. 5.2 below.

Define

$$\begin{aligned}
 U_{n1}(z) &= \frac{1}{n} \sum_{i=1}^n K_{hi}(z)[Y_i - H_{\theta_0}(Z_i)], \\
 U_{n2}(z) &= \frac{1}{n} \sum_{i=1}^n K_{hi}(z)[\widehat{H}_{\theta_0}(Z_i) - H_{\theta_0}(Z_i)], \\
 V_n(z, \theta) &= \frac{1}{n} \sum_{i=1}^n K_{hi}(z)[\widehat{H}_{\theta}(Z_i) - \widehat{H}_{\theta_0}(Z_i)].
 \end{aligned}
 \tag{4.1}$$

The following decomposition is important to study the asymptotic behavior of the proposed m.d. tests. Rewrite

$$\begin{aligned}
 \widehat{M}_n(\widehat{\theta}_n) &= \int \left\{ \frac{1}{n} \sum_{i=1}^n K_{hi}(z)[Y_i - H_{\theta_0}(Z_i) + H_{\theta_0}(Z_i) - \widehat{H}_{\theta_0}(Z_i) \right. \\
 &\quad \left. + \widehat{H}_{\theta_0}(Z_i) - \widehat{H}_{\widehat{\theta}_n}(Z_i)] \right\}^2 d\widehat{\varphi}(z) \\
 &= \int [U_{n1}(z) - U_{n2}(z) - V_n(z, \widehat{\theta}_n)]^2 d\widehat{\varphi}(z) \\
 &= \int [U_{n1}(z) - U_{n2}(z)]^2 d\widehat{\varphi}(z) + \int [V_n(z, \widehat{\theta}_n)]^2 d\widehat{\varphi}(z) \\
 &\quad - 2 \int [U_{n1}(z) - U_{n2}(z)]V_n(z, \widehat{\theta}_n) d\widehat{\varphi}(z) \\
 &:= J_n + \widehat{D}_n(\widehat{\theta}_n) - 2K_n(\widehat{\theta}_n), \quad \text{say.}
 \end{aligned}$$

The following three lemmas yield the conclusion of Theorem 2 in a routine fashion.

Lemma 3 *Suppose assumptions (A1), (A2), (A4), (A5), (F1)–(F2), (K), (H1)–(H6), (W1), (W2) and H_0 hold. Then*

$$nh^{p/2} \widetilde{\Gamma}_n^{-1/2} (J_n - \widetilde{C}_n) \rightarrow_d \mathcal{N}_1(0, 1).$$

Lemma 4 *Under the assumptions of Lemma 3, the following holds.*

$$\text{(a) } nh^{p/2} \widehat{D}_n(\widehat{\theta}_n) = o_p(1), \quad \text{(b) } nh^{p/2} K_n(\widehat{\theta}_n) = o_p(1).$$

Lemma 5 *Suppose assumptions (A1), (A2), (F1), (K), (H1)–(H6), (W1) with $\lambda < \infty$, (W2) and H_0 hold. Then*

$$\text{(a) } nh^{p/2} (\widehat{C}_n - \widetilde{C}_n) = o_p(1), \quad \text{(b) } \widehat{\Gamma}_n - \widetilde{\Gamma}_n = o_p(1).$$

4.2 Consistency

Next, we shall briefly discuss the consistency of these tests. Recall (3.1). Let θ_n be an consistent estimator of $T(H)$, $\xi_i = Y_i - H(Z_i)$, $\xi_{ni} = Y_i - \widehat{H}_{\theta_n}(Z_i)$, and let

$$C_n = \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z) \xi_{ni}^2 d\hat{\varphi}(z), \quad \Gamma_n = \frac{2h^p}{n^2} \sum_{i \neq j=1}^n \left(\int K_{hi}(z) K_{hj}(z) \xi_{ni} \xi_{nj} d\hat{\varphi}(z) \right)^2.$$

Let $\mathcal{T}_n := nh^{p/2} \Gamma_n^{-1/2} (\widehat{M}_n(\theta_n) - C_n)$. Then the theorem below presents the asymptotic behavior of the proposed test under certain alternative hypotheses.

Theorem 3 *Suppose (A1), (A2), (A4), (A5), (F1), (F2), (H3), (K), (W1) and (W2) hold and the alternative hypothesis $H_1 : \mu(x) = m(x)$, $x \in \mathcal{C}$ satisfies that $\inf_{\theta} \rho(H, H_{\theta}) > 0$ and $T(H)$ is unique. Then $|\mathcal{T}_n| \rightarrow_p \infty$ for any consistent estimator θ_n of $T(H)$.*

By Lemma 2, $\hat{\theta}_n$ is consistent for $T(H)$; therefore, the above theorem implies that $|\widehat{\mathcal{T}}_n| \rightarrow \infty$ in probability under the same regularity conditions, and the test based on $\widehat{\mathcal{T}}_n$ is consistent against the alternative m for which $\inf_{\theta} \rho(H, H_{\theta}) > 0$. The proof of Theorem 3 is similar to that of Theorem 5.1 in KS with slight modifications. The techniques used for analyzing $W_n(\theta)$ in Lemma 1 and $\widehat{D}_n(\theta)$ in the proof of Theorem 1 are enough to produce the conclusions. Details are skipped for the sake of brevity.

4.3 Power at local alternatives

We further investigate the asymptotic power of the proposed test against certain local alternatives. Let a be a known real-valued function with continuous derivative. Define $A(z) = E(a(X)|Z = z)$ and $A_2(z) = E([a(X)]^2|Z = z)$, $z \in \mathcal{C}$. Furthermore, suppose both A and A_2 are continuous on \mathcal{C} and

$$\int H_{\theta} A dG = 0, \quad \forall \theta \in \Theta. \tag{4.2}$$

We consider a sequence of local alternatives

$$\mathcal{H}_{1,n} : \mu(x) = m_{\theta_0}(x) + b_n a(x), \quad b_n = 1/\sqrt{nh^{p/2}}. \tag{4.3}$$

The asymptotic distribution of $\hat{\theta}_n$ under $\mathcal{H}_{1,n}$ is given in the following theorem.

Theorem 4 *Assume (A1)–(A3), (A5), (F1), (F2), (H1)–(H6), (K), (W1) and (W2) hold. Then under (4.2) and (4.3), $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightarrow_d \mathcal{N}_q(0, \Sigma_0^{-1}(\Sigma_1 + \lambda^{-1} \Sigma_2) \Sigma_0^{-1})$, where Σ_0 is given in (H6), Σ_1 and Σ_2 are defined in (3.2).*

The asymptotic distribution of the proposed m.d. tests against the local alternatives $\mathcal{H}_{1,n}$ in (4.3) is presented in the following theorem. Define

$$K_2(v) = \int K(v + u)K(u)du, \quad \Gamma = 2 \int K_2^2(v)dv \int [\sigma_{\varepsilon}^2 + \tau^2(z)]^2 g(z) d\varphi(z).$$

Theorem 5 Suppose (A1)–(A3), (A5), (F1), (F2), (H1)–(H6), (K), (W1) and (W2) hold. Then under (4.2) and (4.3), $\widehat{T}_n \rightarrow_d \mathcal{N}_1(\Gamma^{-1/2} \int A^2 dG, 1)$.

Similar to Theorem 2, under the chosen local alternative sequences, the limit λ of sample ratio N/n does not play a critical role in the asymptotic property of the m.d. test.

Remark 4 (Optimal G) Let $\mathcal{K}(g) := \Gamma^{-1/2} \int A^2 dG$. From the above theorem, the asymptotic power of the level α test based on \widehat{T}_n is

$$1 - \Phi(z_{\alpha/2} - \mathcal{K}(g)) + \Phi(-z_{\alpha/2} - \mathcal{K}(g)).$$

This power is an increasing function of $\mathcal{K}(g)$. Thus it will be maximized by that g which maximizes $\mathcal{K}(g)$. Let $c = 2 \int K_Z^2(v)dv$ and $\kappa(z) := \sigma_\epsilon^2 + \tau^2(z)$. Note that $\kappa(z) \geq \sigma_\epsilon^2 > 0$. Under (F1), $f_Z(z) > 0$ for all $z \in \mathcal{C}$. Then by the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} \mathcal{K}(g) &= \frac{\int A^2(z)g(z)dz}{\sqrt{c \int \kappa^2(z)g^2(z)f_Z^{-2}(z)dz}} = c^{-1/2} \frac{\int A^2(z)f_Z(z)\kappa^{-1}(z)\kappa(z)g(z)f_Z^{-1}(z)dz}{\sqrt{\int \kappa^2(z)g^2(z)f_Z^{-2}(z)dz}} \\ &\leq c^{-1/2} \left(\int \frac{A^4(z)f_Z^2(z)}{\kappa^2(z)} dz \right)^{1/2}, \end{aligned}$$

with equality holding if, and only if, $g(z) \propto A^2(z)f_Z^2(z)/\kappa^2(z)$, for all $z \in \mathcal{C}$. Since $\mathcal{K}(g)$ is scale invariant, i.e., $\mathcal{K}(bg) = \mathcal{K}(g)$, for all $b > 0$, we can take the optimal $g(z) = A^2(z)f_Z^2(z)/\kappa^2(z)$.

5 Simulation study

In this section, we present the results of a Monte Carlo study of the proposed estimation and testing procedures for $p = 1, 2$. For $p = 1$, a nonlinear regression model is considered. For $p = 2$, a linear regression is assumed. Three different values of the ratio $N/n = (4, 1, 1/4)$ are selected to assess its effect on the performance of these inference procedures. Throughout the simulation, $K(u) = 0.75(1 - u^2)I_{(|u| \leq 1)}$ for $p = 1$ and $K(u) = 0.75^2(1 - u_1^2)(1 - u_2^2)I_{(|u_1| \leq 1, |u_2| \leq 1)}$ for $p = 2$. The set \mathcal{C} and the integrating measure G are specified later. All of the results obtained are based on 1000 replications.

We need to determine the two bandwidths for the implementation of the above inference procedures. As mentioned in Sect. 1, one bandwidth used for estimating f_Z is $w = c(\log n/n)^{1/(p+4)}$, $c > 0$. We propose to obtain c by minimizing, w.r.t. c , the unbiased cross-validation criterion $UCV(w)$ of Härdle et al. (1990), where

$$UCV(w) = \frac{(R(K))^p}{nw^p} + \frac{1}{n(n-1)w^p} \sum_{i \neq j=1}^n (K * K - K) \left(\frac{Z_i - Z_j}{w} \right),$$

with $R(K) = \int K^2(x)dx$ and $K * K(x) = \int K(y)K(x - y)dy$. We applied a grid search to choose the optimal coefficient c starting from 0.1 with step 0.02, i.e.,

$$c_n^* := \operatorname{argmin}_{0.1 \leq c \leq 10} UCV \left(c(\log n/n)^{1/(p+4)} \right), \quad w_{opt} = c_n^*(\log n/n)^{1/(p+4)}.$$

In order for the bandwidth h to satisfy (W2), we used $h = dn^{-1/\Delta}$, $\Delta = \max\{2p, p(p + 4)/4\} + 1$. We further adopted the leave-one-out cross-validation approach for the nonparametric regression function estimator $\hat{\mu}$ based on $\{(Y_i, Z_i), 1 \leq i \leq n\}$ to select the optimal coefficient d , i.e.,

$$d_n^* = \operatorname{argmin}_{0.1 \leq d \leq 10} \sum_{i=1}^n \{Y_i - \hat{\mu}_{-i}(Z_i)\}^2, \quad \hat{\mu}_{-i}(z) = \frac{\sum_{j=1, j \neq i}^n K_{hj}(z)Y_j}{\sum_{j=1, j \neq i}^n K_{hj}(z)},$$

$$h_{opt} = d_n^*n^{-1/\Delta}.$$

In order to interpret the performance of the proposed estimator $\hat{\theta}_n$, we also present the performance of the KS estimator $\tilde{\theta}_n$. Recall that in KS, f_η is assumed to be known.

Both the absolute bias and square root of mean square error (RMSE) of the two estimators are reported. In both linear and nonlinear cases, the absolute bias and RMSE decrease as the sample sizes increase. In the linear case, as shown in Example 1, there is no need to estimate the regression function and the asymptotic variance of $\hat{\theta}_n$ is the same as that of $\tilde{\theta}_n$. This is also reflected in this finite sample study as in the case of $p = 2$ the RMSEs of the components of $\hat{\theta}_n$ and $\tilde{\theta}_n$ in Table 2 are very similar to each other for all the chosen values of N/n . In the nonlinear case, Table 1 shows that the obtained RMSE of $\hat{\theta}_n$ is larger than $\tilde{\theta}_n$ and it decreases as N/n increases from 1/4 to 4.

We compared the proposed test \hat{T}_n with the test \tilde{T}_n of KS, where

$$C_n = \frac{1}{n^2} \sum_{i=1}^n \int K_{hi}^2(z) \tilde{\xi}_i^2 d\hat{\varphi}(z), \quad \tilde{\xi}_i = Y_i - H_{\tilde{\theta}_n}(Z_i),$$

$$\tilde{T}_n = \frac{2h^p}{n^2} \sum_{i \neq j} \left(\int K_{hi}(z)K_{hj}(z) \tilde{\xi}_i \tilde{\xi}_j d\hat{\varphi}(z) \right)^2, \quad \tilde{T}_n := nh^{p/2} \tilde{T}_n^{-1/2} (M_n(\tilde{\theta}_n) - C_n).$$

The \tilde{T}_n test rejects H_0 at the significance level α whenever $|\tilde{T}_n| \geq z_{\alpha/2}$. With the nominal level 0.05, the empirical levels and powers of these two tests are obtained by computing $\#\{|\hat{T}_n| \geq 1.96\}/1000$ and $\#\{|\tilde{T}_n| \geq 1.96\}/1000$.

5.1 Finite sample performance of $\hat{\theta}_n$

In this subsection, we report the findings of a finite sample performance of the estimator $\hat{\theta}_n$ in nonlinear and linear cases.

The nonlinear case with $q = 1 = p$. Here, data are generated from the model (1.1) with

Table 1 Performance of $\hat{\theta}_n$ and $\tilde{\theta}_n$ in the nonlinear case (5.1) with $p = 1$

$N/n = 4$					
(n, N)	(60, 240)	(100, 400)	(200, 800)	(300, 1200)	(400, 1600)
$ \text{BIAS}(\hat{\theta}_n) $	0.0010	0.0030	0.0008	0.0017	0.0007
$\text{RMSE}(\hat{\theta}_n)$	0.0716	0.0552	0.0393	0.0311	0.0274
$N/n = 1$					
(n, N)	(60, 60)	(100, 100)	(200, 200)	(300, 300)	(400, 400)
$ \text{BIAS}(\hat{\theta}_n) $	0.0012	0.0036	0.0021	0.0015	0.0009
$\text{RMSE}(\hat{\theta}_n)$	0.0768	0.0591	0.0424	0.0338	0.0293
$N/n = 1/4$					
(n, N)	(60, 15)	(100, 25)	(200, 50)	(300, 75)	(400, 100)
$ \text{BIAS}(\hat{\theta}_n) $	0.0063	0.0048	0.0027	0.00014	0.0008
$\text{RMSE}(\hat{\theta}_n)$	0.0954	0.0730	0.0503	0.0417	0.0355
$\tilde{\theta}_n$					
n	60	100	200	300	400
$ \text{BIAS}(\tilde{\theta}_n) $	0.0029	0.0044	0.0012	0.0009	0.0005
$\text{RMSE}(\tilde{\theta}_n)$	0.0686	0.0552	0.0392	0.0325	0.0264

$$\mu(x) = m_{\theta_0}(x) = e^{\theta_0 x}, \quad \theta_0 = -1, \tag{5.1}$$

where $\varepsilon \sim \mathcal{N}_1(0, 0.2^2)$, $\eta \sim \mathcal{N}_1(0, 0.2^2)$, $Z \sim U[-1, 1]$. Then

$$H_{\theta_0}(z) = e^{\theta_0^2 \sigma_\eta^2 / 2} e^{\theta_0 z}, \quad \widehat{H}_{\theta_0}(z) = \frac{1}{N} \sum_{k=1}^N e^{\theta_0(z + \tilde{\eta}_k)} = e^{\theta_0 z} \frac{1}{N} \sum_{k=1}^N e^{\theta_0 \tilde{\eta}_k},$$

and the second term Σ_2 in the asymptotic variance is calculated as

$$\sigma_{\theta_0}(x, y) = e^{\sigma_\eta^2} (e^{\sigma_\eta^2} - 1) e^{\theta_0(x+y)}, \quad \Sigma_2 = e^{2\sigma_\eta^2} (e^{\sigma_\eta^2} - 1) \left[\int (x + \sigma_\eta^2 \theta_0) e^{\theta_0 x} dG(x) \right]^2.$$

We used $\mathcal{C} = [-1, 1]$ and G equal to the uniform measure on $[-1, 1]$.

Table 1 shows little empirical bias in $\hat{\theta}_n$ and its RMSE decreases as the sample sizes and N/n increase. For $N/n = 4, 1$, the RMSE of $\hat{\theta}_n$ is similar to that of $\tilde{\theta}_n$, while for $N/n = 1/4$, RMSE of $\hat{\theta}_n$ is much larger than that of $\tilde{\theta}_n$, because of the smaller size of the validation data set.

The linear case with $q = 2 = p$. Here we consider the model $m_\theta(x) = \theta_1 x_1 + \theta_2 x_2$, $\theta = (\theta_1, \theta_2)^T \in \mathbb{R}^2$, $x = (x_1, x_2)^T \in \mathbb{R}^2$. The true parameter $\theta_0 = (0.5, 1)$ is used to generate the data. Denote $Z_i = (Z_{i1}, Z_{i2})^T$ and $\eta_i = (\eta_{i1}, \eta_{i2})^T$ for $1 \leq i \leq n$. Both Z_{i1} and Z_{i2} are generated independently from $U[-1, 1]$, while η_i are generated from a bivariate normal distribution $\mathcal{N}_2(0, \Sigma_\eta)$ with $\Sigma_\eta = (\sigma_{ij})_{i,j=1,2}$, $\sigma_{11} = 0.1^2$, $\sigma_{22} = 0.2^2$, $\sigma_{12} = \sigma_{21} = 0.01$. Then $X_i = (X_{i1}, X_{i2})^T = Z_i + \eta_i$. The primary data $\{(Y_i, Z_i), 1 \leq i \leq n\}$ are generated from the above regression function with the error

Table 2 Performance of $\hat{\theta}_n$ and $\tilde{\theta}_n$ in the linear case with $p = 2 = q$

$N/n = 4$					
(n, N)	(60, 240)	(100, 400)	(200, 800)	(300, 1200)	(400, 1600)
$ \text{BIAS}(\hat{\theta}_{n,1}) $	0.0007	0.0031	0.0007	0.0004	0.0009
$\text{RMSE}(\hat{\theta}_{n,1})$	0.1069	0.0911	0.0515	0.0434	0.0345
$ \text{BIAS}(\hat{\theta}_{n,2}) $	0.0020	0.0003	0.0034	0.0020	0.0004
$\text{RMSE}(\hat{\theta}_{n,2})$	0.1048	0.0863	0.0511	0.0428	0.0356
$N/n = 1$					
(n, N)	(60, 60)	(100, 100)	(200, 200)	(300, 300)	(400, 400)
$ \text{BIAS}(\hat{\theta}_{n,1}) $	0.0012	0.0032	0.0009	0.0003	0.0009
$\text{RMSE}(\hat{\theta}_{n,1})$	0.1064	0.0895	0.0516	0.0434	0.0345
$ \text{BIAS}(\hat{\theta}_{n,2}) $	0.0004	0.0014	0.0032	0.0016	0.0001
$\text{RMSE}(\hat{\theta}_{n,2})$	0.1049	0.0844	0.0516	0.0427	0.0355
$N/n = 1/4$					
(n, N)	(60, 15)	(100, 25)	(200, 50)	(300, 75)	(400, 100)
$ \text{BIAS}(\hat{\theta}_{n,1}) $	0.0042	0.0041	0.0015	0.0002	0.0005
$\text{RMSE}(\hat{\theta}_{n,1})$	0.1073	0.0916	0.0516	0.0435	0.0344
$ \text{BIAS}(\hat{\theta}_{n,2}) $	0.0040	0.0040	0.0012	0.0002	0.0009
$\text{RMSE}(\hat{\theta}_{n,2})$	0.1079	0.0882	0.0518	0.0429	0.0357
$\tilde{\theta}_n$					
n	60	100	200	300	400
$ \text{BIAS}(\tilde{\theta}_1) $	0.0070	0.0005	0.0028	0.0021	0.0011
$\text{RMSE}(\tilde{\theta}_1)$	0.1162	0.0952	0.0560	0.0497	0.0339
$ \text{BIAS}(\tilde{\theta}_2) $	0.0023	0.0006	0.0012	0.0022	0.0002
$\text{RMSE}(\tilde{\theta}_2)$	0.1086	0.0877	0.0513	0.0438	0.0357

ε following $\mathcal{N}_1(0, 0.3^2)$. The validation data $\{\tilde{\eta}_k, 1 \leq k \leq N\}$ are independently simulated from $\mathcal{N}_2(0, \Sigma_\eta)$. The bandwidths h and w are obtained based on the criteria mentioned above. In this case, $\mathcal{C} = [-1, 1]^2$ and G is the uniform measure on $[-1, 1]^2$. The choices of N/n are the same as in the previous case.

Both bias and RMSE of the estimators $\hat{\theta}_n = (\hat{\theta}_{n,1}, \hat{\theta}_{n,2})^T$ and $\tilde{\theta}_n = (\tilde{\theta}_1, \tilde{\theta}_2)^T$ are presented in Table 2. It shows small estimation bias and reduced RMSE for increased sample sizes. As noted in Proposition 1, in this linear setup, the asymptotic variances of $\hat{\theta}_n$ and $\tilde{\theta}_n$ are equivalent to each other. This fact is also reflected in this finite sample study by observing that the RMSEs of the components of $\hat{\theta}_n$ are very similar to those of the components of $\tilde{\theta}_n$ for the different chosen values of N/n .

5.2 Test performance

Here we present the performance of a member of the proposed class of m.d. tests based on \hat{T}_n and a member of the KS tests based on \tilde{T}_n .

Table 3 Empirical levels and powers of $\widehat{\mathcal{T}}_n$ and $\widetilde{\mathcal{T}}_n$ tests for the nonlinear null model (5.1) with $p = 1 = q$

<i>N/n = 4</i>					
(<i>n, N</i>)	(60, 240)	(100, 400)	(200, 800)	(300, 1200)	(400, 1600)
Model 0	0.043	0.042	0.048	0.045	0.047
Model 1	0.153	0.183	0.462	0.724	0.878
Model 2	0.113	0.196	0.438	0.680	0.866
Model 3	0.163	0.288	0.689	0.936	0.990
<i>N/n = 1</i>					
(<i>n, N</i>)	(60, 60)	(100, 100)	(200, 200)	(300, 300)	(400, 400)
Model 0	0.043	0.045	0.052	0.044	0.048
Model 1	0.170	0.199	0.481	0.722	0.861
Model 2	0.130	0.201	0.437	0.680	0.870
Model 3	0.187	0.325	0.668	0.922	0.990
<i>N/n = 1/4</i>					
(<i>n, N</i>)	(60, 15)	(100, 25)	(200, 50)	(300, 75)	(400, 100)
Model 0	0.062	0.054	0.059	0.055	0.053
Model 1	0.185	0.227	0.464	0.724	0.851
Model 2	0.146	0.217	0.464	0.672	0.856
Model 3	0.228	0.339	0.690	0.914	0.985
$\widetilde{\mathcal{T}}_n$					
<i>n</i>	60	100	200	300	400
Model 0	0.074	0.060	0.044	0.043	0.055
Model 1	0.145	0.219	0.469	0.680	0.849
Model 2	0.144	0.230	0.474	0.705	0.902
Model 3	0.180	0.291	0.646	0.880	0.986

The case $q = 1 = p$. The finite sample performance of these tests is assessed for the nonlinear model (5.1) as the null. Three different alternatives defined below are chosen to obtain the empirical power of a member of the class of the proposed tests.

$$\text{Model 0: } Y = e^{-X} + \varepsilon.$$

$$\text{Model 1: } Y = e^{-X} - 0.2X^2 + \varepsilon.$$

$$\text{Model 2: } Y = e^{-X} + 0.2 \sin(2X) + \varepsilon.$$

$$\text{Model 3: } Y = e^{-X} I_{(X \leq 0.4)} + e^{-0.4} I_{(X > 0.4)} + \varepsilon.$$

The entities G, K, f_Z, η and ε are as in the case of $q = 1 = p$ in Sect. 5.1.

The empirical levels under model 0 and the empirical powers under models 1, 2, 3 for both the proposed and the KS tests are shown in Table 3 with increasing sample sizes. With the nominal level 0.05, the empirical level of the $\widehat{\mathcal{T}}_n$ test is well controlled for the larger sample sizes when $N/n = 1, 4$. For $N/n = 1/4$, the empirical level is slightly inflated for small and moderate sample sizes due to the limited validation data, and it decreases toward 0.05 when the sample size increases. The empirical levels of

Table 4 Empirical levels and powers of \widehat{T}_n and \widetilde{T}_n tests under the linear null model with $p = 2 = q$

$N/n = 4$					
(n, N)	(60, 240)	(100, 400)	(200, 800)	(300, 1200)	(400, 1600)
Model \emptyset	0.045	0.038	0.042	0.050	0.048
Model I	0.205	0.470	0.865	0.968	0.996
Model II	0.066	0.129	0.303	0.519	0.686
Model III	0.222	0.488	0.901	0.984	0.997
$N/n = 1$					
(n, N)	(60, 60)	(100, 100)	(200, 200)	(300, 300)	(400, 400)
Model \emptyset	0.048	0.035	0.043	0.053	0.049
Model I	0.218	0.468	0.859	0.970	0.996
Model II	0.073	0.128	0.313	0.521	0.688
Model III	0.223	0.476	0.884	0.979	0.998
$N/n = 1/4$					
(n, N)	(60, 15)	(100, 25)	(200, 50)	(300, 75)	(400, 100)
Model \emptyset	0.060	0.047	0.044	0.056	0.045
Model I	0.234	0.497	0.883	0.975	0.996
Model II	0.086	0.159	0.347	0.558	0.716
Model III	0.242	0.522	0.867	0.971	0.995
\widetilde{T}_n					
n	60	100	200	300	400
Model \emptyset	0.042	0.036	0.042	0.056	0.047
Model I	0.199	0.464	0.869	0.975	0.997
Model II	0.058	0.124	0.302	0.516	0.690
Model III	0.212	0.477	0.902	0.984	0.997

the test \widehat{T}_n are very similar to those of the \widetilde{T}_n test for the chosen larger sample sizes, $n \geq 200$. This finding is consistent with the theoretical result that the asymptotic null distribution of \widehat{T}_n does not depend on the ratio N/n and is the same as that of \widetilde{T}_n .

We also find that the empirical powers of the \widehat{T}_n test for all chosen alternatives are very similar to the empirical powers of the \widetilde{T}_n test for all the three choices of N/n and for moderate-to-large sample sizes. This finding is in some sense consistent with the results in Sect. 4 that the asymptotic distribution of \widehat{T}_n does not depend on the limit of N/n for certain fixed and local alternatives.

The linear case $q = 2 = p$. In this case, the setup is the same as Sect. 5.1 for $p = 2$. We investigate the empirical level of the proposed test under models defined below. With $\theta_0 = (0.5, 1)^T$ and $X = (X_1, X_2)^T$,

$$\text{Model } \emptyset : Y = \theta_0^T X + \varepsilon,$$

$$\text{Model I} : Y = \theta_0^T X + 0.2X_1X_2 + \varepsilon,$$

$$\text{Model II} : Y = \theta_0^T X + 0.5 \sin(2X_1X_2) + \varepsilon,$$

$$\text{Model III} : Y = \theta_0^T X I_{(\theta_0^T X \leq 0.5)} + 0.5 I_{(\theta_0^T X > 0.5)} + \varepsilon.$$

The numerical findings are summarized in Table 4. In this linear case, it is observed that the empirical levels of the proposed test are close to the nominal level 0.05 for the larger chosen sample sizes and for all three choices of N/n . The empirical power performance pattern is similar to that in the nonlinear case. Again, not surprisingly, the empirical powers of the \widehat{T}_n test are similar to those of the \widetilde{T}_n test for the larger chosen sample sizes and all the three choices of N/n .

Acknowledgements The authors thank the two referees for their constructive comments that helped improve the presentation.

References

- Bates DM, Watts DG (1998) Nonlinear regression analysis and its applications. Wiley, New York
- Berkson J (1950) Are there two regressions? *J Am Stat Assoc* 45:164–180
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. Chapman and Hall/CRC, Boca Raton
- Cheng CL, Van Ness JW (1999) Statistical regression with measurement error. Wiley, New York
- Du L, Zou C, Wang Z (2011) Nonparametric regression function estimation for errors-in-variables models with validation data. *Stat Sin* 21:1093–1113
- Fuller WA (1987) Measurement error models. Wiley, New York
- Geng P, Koul HL (2017) Model checking in tobit regression with measurement errors using validation data. *J Stat Plan Inference* 190:15–31
- Härdle W, Marron JS, Wand MP (1990) Bandwidth choice for density derivatives. *J R Stat Soc B* 52(1):223–232
- Koul HL, Ni P (2004) Minimum distance regression model checking. *J Stat Plan Inference* 119:109–141
- Koul HL, Song W (2009) Minimum distance regression model checking with Berkson measurement errors. *Ann Stat* 37(1):132–156
- Mack YP, Silverman BW (1982) Weak and strong uniform consistency of kernel regression estimates. *Z Wahrsch Gebiete* 61:405–415
- Schennach SM (2013) Regressions with berkson errors in covariates a nonparametric approach. *Ann Stat* 41(3):1642–1668
- Sepanski JH, Lee L (1995) Semiparametric estimation of nonlinear errors-in-variables models with validation study. *J Nonparametr Stat* 4(4):365–394
- Wang L (2003) Estimation of nonlinear Berkson-type measurement error models. *Stat Sin* 13:1201–1210
- Wang L (2004) Estimation of nonlinear models with Berkson measurement errors. *Ann Stat* 32(6):2559–2579
- Yi GY, Ma Y, Spiegelman D, Carroll RJ (2015) Functional and structural methods with mixed measurement error and misclassification in covariates. *J Am Stat Assoc* 110(510):681–696