Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 138 (2008) 1615–1628
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Regression model checking with Berkson measurement errors

Hira L. Koul*[,1], Weixing Song

Department of Statistics and Probability, Michigan State University, A435 Wells Hall, East Lansing, MI 48824-1027, USA

Received 9 March 2006; received in revised form 8 May 2007; accepted 8 May 2007
Available online 15 August 2007

Abstract

This paper discusses asymptotically distribution free tests for the lack-of-fit of a parametric regression model in the Berkson measurement error model. These tests are based on a martingale transform of a certain marked empirical process of calibrated residuals. A simulation study is included to assess the effect of measurement error on the proposed test. It is observed that empirical level is more stable across the chosen measurement error variances when fitting a linear model compared to when fitting a nonlinear model, while, in both cases, the empirical power decreases as this error variance increases, against all chosen alternatives.

MSC: Primary 62G08; secondary 62G10

Keywords: Marked empirical process; Asymptotically distribution free

1. Introduction

This paper is concerned with developing lack-of-fit tests for fitting a parametric model to the regression function in the Berkson measurement error regression model. In this model one regresses the response variable Y on a non-observable predictor variable X, but where a surrogate Z on X is observable. More precisely, here the variables X, Y and Z satisfy

$$Y = \mu(X) + \varepsilon, \quad X = Z + \eta, \tag{1.1}$$

where the random variable Z and the random errors η and ε are assumed to be mutually independent, with $E\varepsilon = 0 = E\eta$, $0 < \sigma_\varepsilon^2 := E(\varepsilon^2) < \infty$ and $\mu(X) = E(Y|X)$.

This model has been found to be useful in agricultural and medical studies. As an example, consider the herbicide study of Rudemo et al. (1989) in which a nominal measured amount Z of herbicide was applied to a plant but the actual amount absorbed by the plant X is unobservable. As another example, from Wang (2004), an epidemiologist studies the severity of a lung disease, Y, among the residents in a city in relation to the amount of certain air pollutants, X. The amounts of the air pollutants Z can be measured at certain observation stations in the city, but the actual exposure of the residents to the pollutants, X, is unobservable and may vary randomly from the Z values. In both cases, X can

* Corresponding author. Tel.: +1 517 353 7170; fax: +1 517 432 1405.
E-mail addresses: koul@stt.msu.edu (H.L. Koul), songweix@msu.edu (W. Song).
[1] Research of this author partly supported by the NSF DMS Grant 0701430.

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2007.05.048

be expressed as Z plus a random error. For some other examples see, e.g., Fuller (1987), Carroll et al. (1995), among others.

Observe that under the model assumptions, $\nu(z) := E(Y|Z=z) = E(\mu(X)|Z=z)$. Thus, one may think of the new regression model $Y = \nu(Z) + \zeta$, where $E(\zeta|Z) = 0$, a.s., so that the error ζ is uncorrelated with Z. This is the so-called calibrated regression model widely used in the measurement errors model literature, see, e.g., Carroll et al. (1995).

Berkson (1950) pointed out that in linear regression models the least squares estimators based on (Z_i, Y_i), $1 \leqslant i \leqslant n$, are unbiased for the regression parameters, without assuming the knowledge of the densities of ε, Z or η. But if regression is nonlinear in X or if there are other parameters in (1.1) that need to be estimated, then extra information about these densities should be supplied to ensure identifiability of the underlying model. A standard assumption in the literature is to assume that density f_η of η is known or unknown only up to an Euclidean parameter vector, cf. Carroll et al. (1995), Huwang and Huang (2000) and Wang (2003, 2004), among others. In this paper, we shall assume that f_η is known.

Now, let $\mathscr{M} := \{m_\theta(x) : x \in \mathbb{R}, \theta \in \Theta \subset \mathbb{R}^q\}$, $q \geqslant 1$, be a given class of real valued parametric functions, where Θ is an open subset of $\mathbb{R}^q$. The problem of interest is to test the hypothesis that $\mu \in \mathscr{M}$, i.e., to test

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

based on n i.i.d. observations (Z_i, Y_i), $1 \leqslant i \leqslant n$, from model (1.1).

In the case of no measurement error in X, statistics literature is replete with tests of H_0. See Hart (1997) and references therein and Koul and Ni (2004), among others, for tests based on nonparametric regression function estimates. An and Cheng (1991), Stute (1997), Stute, Thies, and Zhu (1998) (STZ) and Khmaladze and Koul (2004) base their tests on a certain marked empirical process of residuals. An advantage of these tests over those based on nonparametric regression function estimates is that they have nontrivial asymptotic power against $n^{-1/2}$-nonparametric alternatives. The latter two papers provide a martingale-type transformation of the underlying marked empirical process whose asymptotic null distribution is free from the underlying model, error and the design d.f.'s. The focus of this paper is to develop similar tests for the above testing problem for the measurement errors model (1.1).

Let $s_n(x)$ be a positive consistent estimator of $[E\{(Y-\mu(X))^2|X=x\}]^{1/2}$. The marked empirical process used by STZ to construct tests of H_0 in the classical heteroscedastic regression model when X_i are observable is

$$n^{-1/2} \sum_{i=1}^{n} \frac{(Y_i - \mu(X_i))}{s_n(X_i)} I(X_i \leqslant x), \quad x \in \mathbb{R}.$$

Using regression calibration we modify this process to where X_i is replaced by Z_i and $\mu(X_i)$ is replaced by $\nu(Z_i)$. One may argue that after the regression calibration, the testing problem becomes a classical one. Although the lack-of-fit testing problem in classical regression setup is well studied and these methods, without any doubt, can be adapted to test for lack-of-fit in the Berkson model after regression calibration, no theoretical results of these procedures under (1.1) are available in the literature currently. This paper is an attempt towards filling this void.

In order to implement the above-mentioned modification of the above process, we need a consistent estimator of the conditional variance $\sigma_\nu^2(z) := E\{(Y-\nu(Z))^2|Z=z\}$ when testing for a simple hypothesis and a composite hypothesis. Because of the independence of Z and η, under (1.1), $\nu(z) = \int \mu(x) f_\eta(x-z)\,dx$. Let

$$\tau_\nu^2(z) := E[(\mu(X) - \nu(Z))^2|Z=z], \quad z \in \mathbb{R}.$$

Because Z is uncorrelated with $Y - \nu(Z)$ and independent of ε, we obtain, with $\sigma_\varepsilon^2 := \mathrm{Var}(\varepsilon)$,

$$\sigma_\nu^2(z) = \sigma_\varepsilon^2 + \tau_\nu^2(z), \quad z \in \mathbb{R}. \tag{1.2}$$

Extend the definition of ν, ν_θ and τ_ν to $\bar{\mathbb{R}} := [-\infty, \infty]$ by assigning the value zero to these functions at $\pm\infty$. This convention will apply to analogs of these functions in the sequel. Note that then $\sigma_\nu^2(z) \geqslant \sigma_\varepsilon^2 > 0$, for all $z \in \bar{\mathbb{R}}$.

Consider the problem of testing the simple hypothesis H: $\mu = \mu_0$, where μ_0 is a known regression function. Set $\nu_0(z) := E(\mu_0(X)|Z = z)$ and write $\sigma_0(z)$, $\tau_0(z)$ for $\sigma_{\nu_0}(z)$, $\tau_{\nu_0}(z)$, respectively. Then,

$$\tau_0^2(z) = \int [\mu_0(y) - \nu_0(z)]^2 f_\eta(y - z)\,\mathrm{d}y.$$

Observe that under H, $\tau_0^2(z)$ is known for all z, and that $E(Y - \nu_0(Z))^2 = E\tau_0^2(Z) + \sigma_\varepsilon^2$. Hence, a consistent estimator of σ_ε^2, under H, is given by

$$s_{n0}^2 := \left| n^{-1} \sum_{i=1}^n (Y_i - \nu_0(Z_i))^2 - n^{-1} \sum_{i=1}^n \tau_0^2(Z_i) \right|.$$

This in turn gives a consistent estimator of $\sigma_0^2(z)$ to be

$$\sigma_{n0}^2(z) := \tau_0^2(z) + s_{n0}^2, \quad z \in \mathbb{R}.$$

The analog of the above-marked empirical process suitable here for testing H is

$$V_n^0(z) := n^{-1/2} \sum_{i=1}^n \frac{Y_i - \nu_0(Z_i)}{\sigma_{n0}(Z_i)} I(Z_i \leqslant z), \quad z \in \bar{\mathbb{R}}.$$

Let G denote the d.f. of Z assumed to be continuous. In view of Theorem 2.1 below, under H, V_n^0 converges weakly to $B \circ G$ in $D(\bar{\mathbb{R}})$ and uniform metric. This result, for example, implies that the test that rejects H whenever

$$\sup_{-\infty \leqslant z \leqslant \infty} |V_n^0(z)| > b_\alpha \tag{1.3}$$

is of the asymptotic size α, where b_α is such that $P(\sup_{0 \leqslant t \leqslant 1} |B(t)| > b_\alpha) = \alpha$. Consistency of this test and its asymptotic power against $n^{-1/2}$-nonparametric alternatives are discussed in Section 2.1 below.

Now consider the more interesting problem of testing H_0. Let P_θ and E_θ denote the probability measure and expectation, respectively, under model (1.1) when $\mu = m_\theta$, $\theta \in \Theta$. Let, for $\theta \in \Theta$, $z \in \mathbb{R}$,

$$\nu_\theta(z) := E(m_\theta(X)|Z = z) = \int m_\theta(x) f_\eta(x - z)\,\mathrm{d}x,$$
$$\sigma_\theta^2(z) := E_\theta[(Y - \nu_\theta(Z))^2|Z = z],$$
$$\tau_\theta^2(z) := E[(m_\theta(X) - \nu_\theta(Z))^2|Z = z] = \int [m_\theta(y) - \nu_\theta(z)]^2 f_\eta(y - z)\,\mathrm{d}y.$$

Arguing as for (1.2), we obtain that under (1.1) and P_θ,

$$\sigma_\theta^2(z) = \sigma_\varepsilon^2 + \tau_\theta^2(z) \geqslant \sigma_\varepsilon^2 > 0, \quad z \in \bar{\mathbb{R}}, \quad \theta \in \Theta. \tag{1.4}$$

Since f_η is known, under H_0, $\tau_{\theta_0}^2(z)$ is known except for θ_0. Let θ_n be an $n^{1/2}$-consistent estimator of θ_0 under H_0 and define

$$s_n^2 := \left| n^{-1} \sum_{i=1}^n (Y_i - \nu_{\theta_n}(Z_i))^2 - n^{-1} \sum_{i=1}^n \tau_{\theta_n}^2(Z_i) \right|.$$

This in turn suggests an estimator of $\sigma_{\theta_0}^2(z)$ to be

$$\sigma_{\theta_n}^2(z) := s_n^2 + \tau_{\theta_n}^2(z), \quad z \in \bar{\mathbb{R}}. \tag{1.5}$$

The analog of the above process suitable for testing H_0 here is $\hat{V}_n(z) := V_n(z, \theta_n)$, where

$$V_n(z, \theta) := n^{-1/2} \sum_{i=1}^n \frac{Y_i - \nu_\theta(Z_i)}{\sigma_\theta(Z_i)} I(Z_i \leqslant z), \quad \theta \in \Theta, \quad z \in \bar{\mathbb{R}}.$$

However, tests based on this process are not generally asymptotically distribution free (ADF).

But, under some additional assumptions on the null model, tests based on certain martingale transforms of the process $\hat{V}_n$ are shown to be ADF. This transformation is analogous to the one given in STZ and is described in Section 2.2 below after describing a generic transformation and needed additional assumptions. Section 2.1 discusses a test of the simple hypothesis H, its consistency against a fixed alternative and the asymptotic power against $n^{-1/2}$-nonparametric alternatives.

Section 3 contains a simulation study that assesses the effect of measurement error variance σ_η^2 on the finite sample level and power of the proposed test when fitting a linear model and a nonlinear model. In both cases $q = 2$. From this study one sees that for the selected linear model and for each chosen sample size, the empirical level is relatively stable for all the selected values of σ_η. But when fitting the selected nonlinear model, it is stable for sample sizes 300 and larger. In both cases, the empirical power decreases as σ_η increases. For more details see Section 3 below. Section 4 contains some proofs.

An alternative errors-in-variable model is the one where $Z = X + u$, with the error u being independent of X. Like in Berkson model, the lack-of-fit testing problem in this model is also not fully investigated in the literature. Zhu et al. (2003) and Zhu and Cui (2004, 2005) construct score-type tests for fitting certain parametric regression functions in these models when densities of X and u are of known parametric forms. In the case of unknown density of X, there is no ready to use lack-of-fit test in the literature for this errors-in-variable model. Part of the difficulty is the irregular behavior of the deconvoluted density estimators of the density of X, cf. Fan (1991a,b) and Fan and Truong (1993). Analyzing the asymptotic behavior of the lack-of-fit test based on an analog of $\hat{V}_n$ where one would have to use a deconvoluted density estimator to estimate density of X in these errors-in-variables models is beyond the scope of this paper.

2. Main results

The first subsection discusses the tests of H while those for H_0 are discussed in Section 2.2.

2.1. Tests of a simple hypothesis

To prove the claimed weak convergence of V_n^0 to a time transformed Brownian motion under H and to discuss the consistency and asymptotic power of test (1.3) we first state a weak convergence result.

Lemma 2.1. *Suppose ξ and Z are random variables with $E(\xi|Z) = 0$, $0 < E(\xi^2) < \infty$. Let $\sigma^2(z) := E(\xi^2|Z = z)$, $L(z) := E\sigma^2(Z)I(Z \leqslant z)$, $z \in \bar{\mathbb{R}}$. Let (ξ_i, Z_i), $1 \leqslant i \leqslant n$, be i.i.d. copies of (ξ, Z), and define*

$$U_n(z) := n^{-1/2} \sum_{i=1}^{n} \xi_i I(Z_i \leqslant z), \quad z \in \bar{\mathbb{R}}.$$

Assume L to be continuous. Then,

$$U_n \Longrightarrow B \circ L \quad \text{in } D(\bar{\mathbb{R}}) \text{ and uniform metric.} \tag{2.1}$$

The proof of this lemma uses Theorem 12.6 in Billingsley (1968). Details are similar to those appearing in STZ. Next, we focus on the V_n^0 process. Recall $\sigma_0^2(z) = E\{(Y - v_0(Z))^2|Z = z\}$ and let

$$S_n(z) := n^{-1/2} \sum_{i=1}^{n} \frac{Y_i - v_0(Z_i)}{\sigma_0(Z_i)} I(Z_i \leqslant z), \quad z \in \bar{\mathbb{R}}.$$

The following theorem gives its weak convergence result.

Theorem 2.1. *Suppose model* (1.1) *and* H *hold. In addition, assume that G is continuous and the following condition holds*:

$$E\varepsilon^4 + E(\mu_0(X) - v_0(Z))^4 < \infty. \tag{2.2}$$

Then,

$$\sup_{z\in\bar{\mathbb{R}}} |V_n^0(z) - S_n(z)| = o_p(1), \tag{2.3}$$

$$V_n^0 \Longrightarrow B\circ G \quad \textit{in } D(\bar{\mathbb{R}}) \textit{ and uniform metric}. \tag{2.4}$$

Proof. Clearly, S_n is like a U_n process, with $\xi := (Y - v_0(Z))/\sigma_0(Z)$, $\sigma^2(z) \equiv 1$ and $L = G$. In view of (1.2), $\sigma_0^2(z) \geqslant \sigma_\varepsilon^2$, for all $z \in \bar{\mathbb{R}}$, so that ξ is a well-defined r.v. Under (1.1), H and the continuity of G, assumptions for (2.1) are satisfied and we readily obtain that $S_n \Longrightarrow B\circ G$, in $D(\bar{\mathbb{R}})$ and uniform metric. This fact and (2.3) readily imply (2.4).

To prove (2.3), let

$$\xi_i := \frac{Y_i - v_0(Z_i)}{\sigma_0(Z_i)}, \quad d_{ni} := \sigma_{n0}(Z_i)[\sigma_{n0}(Z_i) + \sigma_0(Z_i)], \quad 1\leqslant i\leqslant n, \tag{2.5}$$

$$W_n(z) := n^{-1/2}\sum_{i=1}^n \xi_i\left[\frac{\sigma_0(Z_i)}{\sigma_{n0}(Z_i)} - 1\right] I(Z_i\leqslant z), \quad z\in\bar{\mathbb{R}}, \quad \Delta_n := \sigma_\varepsilon^2 - s_{n0}^2.$$

We have the decomposition $V_n^0 = S_n + W_n$. Because $\sigma_0^2(z) - \sigma_{n0}^2(z) \equiv \Delta_n$, we obtain

$$\begin{aligned} W_n(z) &= \frac{n^{1/2}\Delta_n}{2} n^{-1}\sum_{i=1}^n \frac{\xi_i}{\sigma_0^2(Z_i)} I(Z_i\leqslant z) + n^{1/2}\Delta_n n^{-1}\sum_{i=1}^n \xi_i\left[\frac{1}{d_{ni}} - \frac{1}{2\sigma_0^2(Z_i)}\right] I(Z_i\leqslant z) \\ &= \frac{n^{1/2}\Delta_n}{2}\, T_{n1}(z) + n^{1/2}\Delta_n T_{n2}(z) \quad \text{say.} \end{aligned}$$

Under (1.1), (2.2) and H, $n^{1/2}|\Delta_n| = O_p(1)$. Moreover,

$$\sup_z |T_{n2}(z)| \leqslant n^{-1}\sum_{i=1}^n |\xi_i|\left|\frac{1}{d_{ni}} - \frac{1}{2\sigma_0^2(Z_i)}\right| \leqslant \max_{1\leqslant i\leqslant n} \frac{|2\sigma_0^2(Z_i) - d_{ni}|}{d_{ni}\sigma_0^2(Z_i)} n^{-1}\sum_{i=1}^n |\xi_i|.$$

Now, note that from the definitions,

$$\begin{aligned} &d_{ni}\sigma_0^2(Z_i) = \sigma_{n0}\{(s_n^2 + \tau_0^2(Z_i))^{1/2} + (\sigma_\varepsilon^2 + \tau_0^2(Z_i))^{1/2}\}(\sigma_\varepsilon^2 + \tau_0^2(Z_i)) \geqslant \sigma_\varepsilon^4 > 0, \\ &|2\sigma_0^2(Z_i) - d_{ni}| = |\sigma_0^2(Z_i) - \sigma_{n0}^2(Z_i) + \sigma_0(Z_i)(\sigma_0(Z_i) - \sigma_{n0}(Z_i))| \leqslant 2|\Delta_n|, \quad \forall 1\leqslant i\leqslant n. \end{aligned}$$

Hence, $\max_{1\leqslant i\leqslant n}(|2\sigma_0^2(Z_i) - d_{ni}|/d_{ni}\sigma_0^2(Z_i)) \leqslant 2\sigma_\varepsilon^{-4}|\Delta_n| = o_p(1)$. These facts together with the fact that $n^{-1}\sum_{i=1}^n |\xi_i| = O_p(1)$ yield

$$\sup_z |T_{n2}(z)| = o_p(1). \tag{2.6}$$

Next consider $T_{n1}(z)$. Let $L_0(z) := E\sigma_0^{-2}(Z)I(Z\leqslant z)$. Because $\sigma_0^2(z) \geqslant \sigma_\varepsilon^2 > 0$, for all $z\in\bar{\mathbb{R}}$, the measure induced by L_0 is a finite measure on $\bar{\mathbb{R}}$. Also, $ET_{n1}(z) \equiv 0$, $\mathrm{Var}(T_{n1}(z)) = n^{-1}L_0(z) \leqslant n^{-1}\sigma_\varepsilon^{-2} \to 0$, for all $z\in\bar{\mathbb{R}}$. This fact and a Glivenko–Cantelli type argument, where the interval is partitioned according to the measure induced by $L_0(z)$, shows that $\sup_{z\in\bar{\mathbb{R}}}|T_{n1}(z)| = o_p(1)$. This completes the proof of (2.3), and hence of the theorem. □

To establish the *consistency* of the test (1.3) against a fixed alternative H_1: $\mu = \mu_1$, $\mu_1 \neq \mu_0$, proceed as follows. Let E_1 denote the expectation under H_1. Assume μ_1 satisfies $E\mu_1^2(X) < \infty$. Let $v_1(z) := E(\mu_1(X)|Z = z)$ and suppose, additionally, that

$$0 < \sigma_1^2 := E_1(Y - v_1(Z))^2 < \infty, \quad d := \sup_{z\in\bar{\mathbb{R}}}\left|E\left\{\frac{v_1(Z) - v_0(Z)}{\sigma_0(Z)} I(Z\leqslant z)\right\}\right| \neq 0. \tag{2.7}$$

Write $V_n^0 = V_n^1 + n^{1/2}\tilde{D}_n$, where

$$V_n^1(z) := n^{-1/2}\sum_{i=1}^n \frac{Y_i - v_1(Z_i)}{\sigma_{n0}(Z_i)} I(Z_i \leqslant z), \quad \tilde{D}_n(z) := n^{-1}\sum_{i=1}^n \frac{v_1(Z_i) - v_0(Z_i)}{\sigma_{n0}(Z_i)} I(Z_i \leqslant z).$$

Let D_n denote the average like $\tilde{D}_n$ but with $\sigma_{n0}(Z_i)$'s replaced by $\sigma_0(Z_i)$'s. An argument similar to the one that yielded (2.6) shows that $\sup_z |\tilde{D}_n(z) - D_n(z)| = o_p(1)$. By a Glivenko–Cantelli type argument, we also have $\sup_{z\in\bar{\mathbb{R}}} |D_n(z)| \longrightarrow d$ a.s. Hence, $\sup_{z\in\bar{\mathbb{R}}} |\tilde{D}_n(z)| \longrightarrow d$ in probability. Next, let

$$S_n^1(z) := n^{-1/2}\sum_{i=1}^n \frac{Y_i - v_1(Z_i)}{\sigma_0(Z_i)} I(Z_i \leqslant z), \quad \psi(z) := E_1 \frac{(Y - v_1(Z))^2}{\sigma_0^2(Z)} I(Z \leqslant z), \quad z \in \bar{\mathbb{R}}.$$

Note that $\psi(z) \leqslant \sigma_\varepsilon^{-2}\sigma_1^2 < \infty$, for all $z \in \bar{\mathbb{R}}$. With (2.1) applied to $\xi_i = (Y_i - v_1(Z_i))/\sigma_0(Z_i)$, we have $L = \psi$, $S_n^1 = U_n$, and, hence, $S_n^1 \Longrightarrow B \circ \psi$ under H_1. Moreover, an argument like the one used in concluding (2.3) shows that $\sup_{z\in\bar{\mathbb{R}}} |V_n^1(z) - S_n^1(z)| = o_p(1)$, under H_1, so that we also have $V_n^1 \Longrightarrow B \circ \psi$, under H_1. These facts and a routine argument show that test (1.3) is consistent against H_1, under (2.2) and (2.7).

Moreover, the asymptotic power of this test against the local nonparametric alternatives $\mu(x) = \mu_0(x) + n^{-1/2}\delta(x)$, where $\delta(x) \neq 0$ and $E\delta^2(X) < \infty$, can be shown to be $P(\sup_{0\leqslant t\leqslant 1} |B(t) + \int_0^t [\delta(u)/\sigma_0(u)]G^{-1}(u)\,du| > b_\alpha)$. Similar facts can be established for any test based on a continuous function of V_n^0.

2.2. *Tests of* H_0

For the sake of completeness we shall first describe a generic transformation of a vector of functions that preserves Brownian motion and that is orthogonal to certain kinds of drifts in some Gaussian processes. This transformation has its roots in Khmaladze (1981, 1993), STZ and Khmaladze and Koul (2004). Then we state needed assumptions on $\mathscr{M}$ and other underlying entities under which the application of this generic transformation to the process $\hat{V}_n$ and its weak convergence to a time transformed Brownian motion is justified. These results in turn are then used to construct ADF tests for H_0 based on the process $\hat{V}_n$.

To describe the generic transformation, let U be a continuous r.v. with d.f. K on $\mathbb{R}$ and ℓ be a vector of q functions such that $E\|\ell(U)\|^2 < \infty$. Let

$$C_u := E\ell(U)\ell(U)' I(U \geqslant u), \quad u \in \mathbb{R}.$$

Suppose C_u is positive definite for all $u \in \mathbb{R}$. Let C_u^{-1} denote its inverse matrix. For a function γ from $\mathbb{R}$ to $\mathbb{R}$, define the operators

$$\mathscr{T}\gamma(u) := \int_{y\leqslant u} \gamma(y)\ell(y)' C_y^{-1}\,dK(y)\ell(u), \quad \mathscr{K}\gamma(u) := \gamma(u) - \mathscr{T}\gamma(u), \quad u \in \mathbb{R},$$

Let $\mathscr{L} := \{\gamma \in L_2(\mathbb{R}, K) : \int \gamma\ell\,dK = 0\}$. Then the transformation $\mathscr{K}$ is norm preserving from $L_2(\mathbb{R}, K)$ to $\mathscr{L}$: For any $\gamma \in L_2(\mathbb{R}, K)$, $\int \mathscr{K}\gamma\ell'\,dK = 0$, $\int (\mathscr{K}\gamma)^2\,dK = \int \gamma^2\,dK$.

Moreover,

$$E\{\mathscr{K}\gamma_1(U)\mathscr{K}\gamma_2(U)\} = E\{\gamma_1(U)\gamma_2(U)\}, \quad \forall \gamma_1, \gamma_2 \in L_2(\mathbb{R}, K). \tag{2.8}$$

The proof of these facts is similar to that of Proposition 6.1 in Khmaladze and Koul (2004). See also Koul (2006).

Next, let ζ be a r.v. such that $E(\zeta|U) = 0$, $\tau^2(u) := E(\zeta^2|U = u) > 0$, for all u. Then from (2.8) we obtain that the process $W_\gamma(\zeta, U) := [\zeta/\tau(U)]\mathscr{K}\gamma(U)$, as a process in $\gamma \in L_2(\mathbb{R}, K)$, is like $B_\gamma(K)$, where B_γ is a Brownian motion in γ. Consequently, if (ζ_i, U_i), $1 \leqslant i \leqslant n$, are i.i.d. copies of (ζ, U), then by the classical CLT, the finite dimensional distributions of $n^{-1/2}\sum_{i=1}^n W_\gamma(\zeta_i, U_i)$, as γ varies, will converge weakly to those of $B_\gamma(K)$. These are some of the basic observations that are useful in constructing ADF tests for H_0 based on $\hat{V}_n$.

We now turn to the problem of testing for H_0. Throughout, the true parameter θ_0 under H_0 is assumed to be in the interior of Θ. Consider the following assumptions:

(e) $E\varepsilon^4 + E(\mu_{\theta_0}(X) - v_{\theta_0}(Z))^4 < \infty$.

(m) For some positive continuous function $r(x)$ with $Er^4(X) < \infty$,

$$|m_{\theta_1}(x) - m_{\theta_2}(x)| \leqslant \|\theta_1 - \theta_2\| r(x), \quad \forall \theta_1, \theta_2 \in \Theta, \quad x \in \mathbb{R}.$$

(v1) For every z, $v_\theta(z)$ is differentiable in θ in a neighborhood of θ_0 with the vector of derivatives $\dot{v}_\theta(z)$, such that for every $0 < k < \infty$,

$$\sup_{1 \leqslant i \leqslant n,\, n^{1/2}\|\theta - \theta_0\| \leqslant k} n^{1/2} |v_\theta(Z_i) - v_{\theta_0}(Z_i) - (\theta - \theta_0)' \dot{v}_{\theta_0}(Z_i)| = o_p(1) \quad (P_{\theta_0}).$$

(v2) For some $q \times q$ square matrix $\ddot{v}_{\theta_0}(z)$ and a nonnegative function $g_{\theta_0}(z)$, both measurable in the z-coordinate, the following holds:

$$E\|\ddot{v}_{\theta_0}(z)\|^j g_{\theta_0}(z) < \infty, \quad E\|\ddot{v}_{\theta_0}(z)\| \|\dot{v}(z, \theta_0)\|^j < \infty, \quad j = 0, 1,$$

and, for all $\delta > 0$, there exists an $\eta > 0$ such that $\|\theta - \theta_0\| < \eta$ implies

$$\|\dot{v}_\theta(z) - \dot{v}_{\theta_0}(z) - \ddot{v}_{\theta_0}(z)(\theta - \theta_0)\| \leqslant \delta g_{\theta_0}(z) \|\theta - \theta_0\|, \quad \text{for } G - \text{almost all } z.$$

(v3) $E\|\dot{v}_{\theta_0}(Z)\|^2 < \infty$, and $L_{\theta_0}(z) := \int_{y \geqslant z} \sigma_{\theta_0}^{-2}(y) \dot{v}_{\theta_0}(y) \dot{v}_{\theta_0}(y)' \, dG(y)$ is positive definite for all $z \in \mathbb{R}$.

We note here that (v3) and (1.4) imply that $E\|\dot{v}_{\theta_0}(Z)/\sigma_{\theta_0}(Z)\|^2 \leqslant \sigma_\varepsilon^{-2} E\|\dot{v}_{\theta_0}(Z)\|^2 < \infty$ and the matrix $\Gamma_{\theta_0} := \int \sigma_{\theta_0}^{-2}(y) \dot{v}_{\theta_0}(y) \dot{v}_{\theta_0}(y)' \, dG(y)$ is positive definite.

The conditions (v1)–(v3) are trivially satisfied by the model $m_\theta(x) = \theta' h(x)$, where $h = (h_1, \ldots, h_q)'$ is vector of real q functions such that $E\|h(X)\|^2 < \infty$, and the matrix $\int_{y \geqslant z} E(h(X)|Z = y) E(h(X)|Z = y)' \, dG(y)$ is positive definite for all $z \in \mathbb{R}$.

A set of somewhat stronger conditions in terms of the given model $\mathscr{M}$ that imply (v1)–(v3) are as follows:

(m1) For all x, $m_\theta(x)$ is differentiable in θ in a neighborhood of θ_0 with the vector of differential $\dot{m}_{\theta_0}$ such that $E\|\dot{m}_{\theta_0}(X)\|^2 < \infty$ and for every $k < \infty$,

$$\sup_{x \in \mathbb{R}, n^{1/2}\|\theta - \theta_0\| \leqslant k} n^{1/2} |m_\theta(x) - m_{\theta_0}(x) - (\theta - \theta_0)' \dot{m}_{\theta_0}(x)| = o_p(1) \quad (P_{\theta_0}).$$

(m2) For some $q \times q$ square matrix $\ddot{m}_{\theta_0}(x)$ and a nonnegative function $g_1(x, \theta_0)$, both measurable in the x-coordinate, the following holds: For every $\delta > 0$, there exists an $\eta > 0$ such that $\|\theta - \theta_0\| < \eta$ implies

$$\|\dot{m}_\theta(x) - \dot{m}_{\theta_0}(x) - \ddot{m}_{\theta_0}(x)(\theta - \theta_0)\| \leqslant \delta g_1(x, \theta_0) \|\theta - \theta_0\|, \quad \forall x,$$

$$E\{\|\ddot{m}_{\theta_0}(X)\|^j E(g_1(X, \theta_0)|Z)\} < \infty, \quad E\{\|\ddot{m}_{\theta_0}(X)\| \|E(\dot{m}_{\theta_0}(X)|Z)|^j\} < \infty, \quad j = 0, 1.$$

Because $v_\theta(z) = \int m_\theta(y) f_\eta(y - z) \, dy$, and $\int f_\eta(y - z) \, dy \equiv 1$, (m1) readily implies ($v$1) with $\dot{v}_{\theta_0}(z) \equiv E(\dot{m}_{\theta_0}(X)| Z = z)$. Also, by the Cauchy–Schwarz inequality applied to the conditional expectation, given Z, $E\|E(\dot{m}_{\theta_0}(X)|Z)\|^2 \leqslant E\|\dot{m}_{\theta_0}(X)\|^2 < \infty$, by (m1), so that the first condition of (v3) holds. Similarly, (m2) readily implies (v2) with $\ddot{v}_{\theta_0}(z) \equiv E(\ddot{m}_{\theta_0}(X)|Z = z)$ and $g_{\theta_0}(z) = E(g_1(X, \theta_0)|Z = z)$.

We are now ready to describe an ADF test of H_0. Apply the transformation $\mathscr{K}$ to $\ell = \dot{v}_{\theta_0}/\sigma_{\theta_0}$, $U = Z$, $\xi = (Y - v_{\theta_0}(Z))/\sigma_{\theta_0}(Z)$, $K = G$, and with $\gamma(u) := I(u \leqslant z)$. Note that now $C_y \equiv L_{\theta_0}(y)$ of (v3). Denote the corresponding $\mathscr{K}\gamma$ by $\mathscr{K}_{\theta_0}(z)(\xi, Z, G)$. In view of the above assumptions, this is well defined and

$$\mathscr{K}_{\theta_0}(z)(\xi, Z, G) := \xi \left[I(Z \leqslant z) - \int_{y \leqslant z} \frac{\dot{v}'_{\theta_0}}{\sigma_{\theta_0}}(y) L_{\theta_0}^{-1}(y) \frac{\dot{v}_{\theta_0}}{\sigma_{\theta_0}}(Z) I(Z \geqslant y) \, dG(y) \right].$$

Let $\xi_i := (Y_i - v_{\theta_0}(Z_i))/\sigma_{\theta_0}(Z_i)$ and define

$$\mathscr{W}_{\theta_0,G}(z) := n^{-1/2}\sum_{i=1}^{n}\mathscr{K}_{\theta_0}(z)(\xi_i, Z_i, G).$$

From the above discussion and the classical CLT we readily obtain that all finite dimensional distributions of $\mathscr{W}_{\theta_0,G}$ converge weakly to those of $B \circ G$. However, because this process depends on the parameters θ_0 and G, it is not useful for inference. The process that is useful for testing for H_0 is $\mathscr{W}_n := \mathscr{W}_{\theta_n,G_n}$, where G_n is the empirical of Z_i, $1 \leqslant i \leqslant n$. Let $\xi_{ni} := (Y_i - v_{\theta_n}(Z_i))/\sigma_{\theta_n}(Z_i)$, $1 \leqslant i \leqslant n$. Recalling definition (1.5) and of $\hat{V}_n$, we obtain

$$\begin{aligned}\mathscr{W}_n(z) &= n^{-1/2}\sum_{i=1}^{n}\mathscr{K}_{\theta_n}(z)(\xi_{ni}, Z_i, G_n)\\ &= \hat{V}_n(z) - \int_{x\leqslant z}\frac{\dot{v}'_{\theta_n}}{\sigma_{\theta_n}}(x)L^{-1}_{\theta_n}(x)\int_{y\geqslant x}\frac{\dot{v}_{\theta_n}}{\sigma_{\theta_n}}(y)\hat{V}_n(\mathrm{d}y)\,\mathrm{d}G_n(x).\end{aligned}$$

Similar to the classical regression case discussed in STZ and Koul (2002), the matrices $L^{-1}_{\theta_n}(x)$ are unstable for x in the right tail closer to infinity and often not uniformly continuous in the underlying parameter θ; so; for a given sample size, $\mathscr{W}_n(z)$ may become very unstable for an arbitrarily large z. Hence, we need to assume that

$$\text{For some } z_0 < \infty\ L_{\theta_0}(z_0) \text{ is non-singular.} \tag{2.9}$$

Consequently, we have to restrict $\mathscr{W}_n(z)$ to the intervals $[-\infty, z_0]$. A practical choice of z_0 depends on the data. The following theorem gives the needed weak convergence result.

Theorem 2.2. *Suppose the model* (1.1) *and* H_0 *hold. In addition, suppose the conditions* (m) *and* (v1)–(v3), *and* (2.9) *hold, and the estimator* θ_n *satisfies*

$$\|n^{1/2}(\theta_n - \theta_0)\| = O_P(1) \quad (H_0). \tag{2.10}$$

Then, $\mathscr{W}_n \Longrightarrow B \circ G$, *in* $D([-\infty, z_0])$ *and uniform metric.*

Consequently, the test that rejects H_0 whenever $\sup_{z\leqslant z_0}|\mathscr{W}_n(z)/0.995| > b_\alpha$ will be of the asymptotic size α. As in STZ, it is recommended to take z_0 to be the 99% quantile of the Z-data in applications.

The main idea of the proof of the above theorem is similar to that appearing in STZ, but some details, which are necessarily different, are given in the last section.

3. Simulations

This section contains a simulation study of the proposed lack-of-fit test for the two cases: *Case* 1: $q = 2$ and m_θ is linear; *Case* 2: $q = 2$ and m_θ is nonlinear. In all cases, $\{Z_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ are generated at random from the uniform distribution on $[-1, 1]$ and $\mathscr{N}(0, (0.1)^2)$ distribution, respectively. To assess the effect of measurement error on the finite sample level and power, $\{\eta_i\}_{i=1}^n$ are generated as a random sample from $\mathscr{N}(0, \sigma_\eta^2)$, for various values of σ_η. Then (X_i, Y_i) are generated from the model $Y_i = m_\theta(X_i) + \varepsilon_i$, $X_i = Z_i + \eta_i$, $i = 1, 2, \ldots, n$.

The sample sizes chosen are $n = 50, 100, 200, 300, 500$ in Case 1, while in Case 2 we also chose $n = 800$. Each simulation is repeated 1000 times. In each case θ_n is taken to be the least squares estimator and $\alpha = 0.05$. The test statistic is $\widehat{\mathscr{D}}_n = \sup_{z\leqslant z_0}|\mathscr{W}_n(z)/0.995|$ with z_0 being the 99th quantile of the Z-data. Under H_0, $\sup_{z\leqslant z_0}|\mathscr{W}_n(z)/0.995| \Longrightarrow \sup_{0\leqslant t\leqslant 1}|B(t)|$. The critical value $b_{0.05} = 2.24241$ is obtained from the fact that

$$P\left(\sup_{0\leqslant t\leqslant 1}|B(t)| < b\right) = P(|B(1)| < b) + 2\sum_{i=1}^{\infty}(-1)^i P((2i-1)b < B(1) < (2i+1)b).$$

The empirical size and power are computed by using $\#\{\widehat{D}_n > 2.24241\}/1000$.

Table 1
Levels and powers of the test in Case 1

		σ_η						
		0.03	0.05	0.08	0.1	0.2	0.4	0.8
$n = 50$	Model 0	0.018	0.017	0.015	0.017	0.019	0.021	0.021
	Model 1	0.865	0.711	0.400	0.258	0.056	0.019	0.012
	Model 2	0.939	0.843	0.591	0.446	0.154	0.080	0.049
	Model 3	1.000	1.000	1.000	0.999	0.976	0.499	0.142
$n = 100$	Model 0	0.029	0.027	0.035	0.031	0.031	0.030	0.029
	Model 1	1.000	0.982	0.860	0.704	0.187	0.045	0.019
	Model 2	1.000	0.997	0.936	0.827	0.375	0.140	0.081
	Model 3	1.000	1.000	1.000	1.000	1.000	0.830	0.243
$n = 200$	Model 0	0.033	0.028	0.029	0.031	0.023	0.027	0.028
	Model 1	1.000	1.000	0.999	0.975	0.466	0.091	0.027
	Model 2	1.000	1.000	1.000	0.996	0.673	0.260	0.119
	Model 3	1.000	1.000	1.000	1.000	1.000	0.995	0.388
$n = 300$	Model 0	0.036	0.036	0.042	0.043	0.035	0.037	0.036
	Model 1	1.000	1.000	1.000	0.998	0.703	0.154	0.032
	Model 2	1.000	1.000	1.000	1.000	0.842	0.346	0.139
	Model 3	1.000	1.000	1.000	1.000	1.000	1.000	0.520
$n = 500$	Model 0	0.048	0.043	0.043	0.049	0.047	0.047	0.049
	Model 1	1.000	1.000	1.000	1.000	0.920	0.227	0.060
	Model 2	1.000	1.000	1.000	1.000	0.968	0.537	0.204
	Model 3	1.000	1.000	1.000	1.000	1.000	1.000	0.702

Case 1: The data were simulated from the following four models:

$$\text{Model 0:} \quad Y_i = 1 + 2X_i + \varepsilon_i, \quad \text{Model 1:} \quad Y_i = 1 + 2X_i + 1.4\exp(-0.2X_i^2) + \varepsilon_i,$$
$$\text{Model 2:} \quad Y_i = 1 + 2X_i + 0.3X_i^2 + \varepsilon_i, \quad \text{Model 3:} \quad Y_i = 1 + 2X_i I(X_i \geqslant 0.2) + \varepsilon_i.$$

Data from Model 0 are used to study the empirical level, while from Models 1–3 are used to study the empirical power of the test.

From Table 1we see that for a fixed sample size, the empirical level is relatively stable for all the selected values of σ_η, while the empirical power decreases when σ_η increases for all alternative models. Secondly, for every fixed value of σ_η, empirical level approaches to the nominal level 0.05 when n gets larger, and empirical power increases when n increases for all selected alternatives.

Case 2. In this case, the data were simulated from the following four models:

$$\text{Model 0:} \quad Y_i = 2X_i + \exp(X_i) + \varepsilon_i,$$
$$\text{Model 1:} \quad Y_i = 2X_i + \exp(X_i) + 0.5X_i^2[\exp(-0.2X_i) + \exp(1.2X_i^2)] + \varepsilon_i,$$
$$\text{Model 2:} \quad Y_i = 2X_i + \exp(X_i) + 0.5 + \varepsilon_i,$$
$$\text{Model 3:} \quad Y_i = 2X_i + \exp(X_i) + 0.5\exp(-0.2X_i) + \varepsilon_i.$$

Data from Model 0 are used to study the empirical level, and from Models 1–3 are used to study the empirical powers of the test.

From Table 2,we see that empirical level for fitting the selected nonlinear model is not as stable as in Case 1 for the selected σ_η values when sample size is small, but it becomes more stable and approaches to the nominal level 0.05 when sample size gets larger. Secondly, for each selected σ_η, the empirical power increases when n increases for all alternative models, while, for each fixed sample size, it decreases when σ_η increases for all selected alternatives.

Table 2
Levels and powers of the test in Case 2

		σ_η						
		0.03	0.05	0.08	0.1	0.2	0.4	0.8
$n=50$	Model 0	0.022	0.021	0.021	0.019	0.038	0.128	0.216
	Model 1	0.725	0.715	0.660	0.655	0.500	0.334	0.134
	Model 2	1.000	1.000	0.988	0.946	0.688	0.561	0.452
	Model 3	1.000	1.000	0.993	0.954	0.728	0.588	0.468
$n=100$	Model 0	0.031	0.026	0.021	0.022	0.029	0.074	0.185
	Model 1	0.991	0.964	0.891	0.844	0.590	0.345	0.144
	Model 2	1.000	1.000	1.000	1.000	0.935	0.685	0.558
	Model 3	1.000	1.000	1.000	1.000	0.948	0.702	0.590
$n=200$	Model 0	0.030	0.037	0.037	0.035	0.034	0.039	0.123
	Model 1	1.000	1.000	0.995	0.980	0.707	0.253	0.108
	Model 2	1.000	1.000	1.000	1.000	0.997	0.853	0.712
	Model 3	1.000	1.000	1.000	1.000	0.998	0.889	0.743
$n=300$	Model 0	0.038	0.045	0.039	0.038	0.035	0.041	0.091
	Model 1	1.000	1.000	1.000	0.998	0.806	0.021	0.094
	Model 2	1.000	1.000	1.000	1.000	1.000	0.908	0.770
	Model 3	1.000	1.000	1.000	1.000	1.000	0.938	0.802
$n=500$	Model 0	0.044	0.045	0.050	0.051	0.053	0.055	0.066
	Model 1	1.000	1.000	1.000	0.999	0.905	0.164	0.107
	Model 2	1.000	1.000	1.000	1.000	1.000	0.978	0.826
	Model 3	1.000	1.000	1.000	1.000	1.000	0.983	0.874
$n=800$	Model 0	0.046	0.045	0.046	0.044	0.049	0.048	0.044
	Model 1	1.000	1.000	1.000	1.000	0.966	0.146	0.114
	Model 2	1.000	1.000	1.000	1.000	1.000	0.996	0.884
	Model 3	1.000	1.000	1.000	1.000	1.000	0.997	0.916

4. Some proofs

Here we sketch the proof of Theorem 2.2. Transform $\mathscr{W}_n$ is an analog of the one given in STZ with a major difference. In the no measurement error setup, STZ uses a part of the sample to estimate the conditional variance and the other part is used to implement their test. But here because of (1.4) we estimate this variance based on the entire sample. This then means that some of the arguments of the proofs though similar are necessarily different.

The proof consists of the following two steps:

(a) $\sup_{z\leqslant z_0} |\mathscr{W}_n(z) - \mathscr{W}_{\theta_0,G}(z)| = o_p(1) \quad (P_{\theta_0})$.

(b) $\mathscr{W}_{\theta_0,G} \Longrightarrow B\circ G$ in $D([-\infty, z_0])$ and in the uniform metric.

Step (b) has been proved in Section 2, so it remains to prove step (a) only.

To begin with rewrite $V_n(z)$ for $V_n(z,\theta_0)$, and $v_n, \dot v_n, \ddot v_n, \tau_n, \sigma_n, L_n$ for $v_{\theta_n}, \dot v_{\theta_n}, \ddot v_{\theta_n}, \tau_{\theta_n}, \sigma_{\theta_n}, L_{\theta_n}$, respectively. Also, let $\Delta_n := \theta_n - \theta_0$. Then, by direct algebra, we obtain

$$\mathscr{W}_{\theta_0,G}(z) = V_n(z) - \int_{x\leqslant z} \frac{\dot v'_{\theta_0}}{\sigma_{\theta_0}}(x) L^{-1}_{\theta_0}(x) \int_{y\geqslant x} \frac{\dot v_{\theta_0}}{\sigma_{\theta_0}}(y) V_n(\mathrm{d}y)\,\mathrm{d}G(x),$$

$$\mathscr{W}_n(z) = \hat V_n(z) - \int_{x\leqslant z} \frac{\dot v'_n}{\sigma_n}(x) L^{-1}_n(x) \int_{y\geqslant x} \frac{\dot v_n}{\sigma_n}(y) \hat V_n(\mathrm{d}y)\,\mathrm{d}G_n(x).$$

Rewrite

$$\hat{V}_n(z) = V_n(z) - \Delta_n' n^{-1/2} \sum_{i=1}^n \frac{\dot{v}_{\theta_0}(Z_i)}{\sigma_{\theta_0}(Z_i)} I(Z_i \leqslant z) + n^{-1/2} \sum_{i=1}^n \zeta_i \left[\frac{\sigma_{\theta_0}(Z_i)}{\sigma_n(Z_i)} - 1 \right] I(Z_i \leqslant z)$$
$$- n^{-1/2} \sum_{i=1}^n \sigma_n^{-1}(Z_i)[v_n(Z_i) - v_{\theta_0}(Z_i) - \Delta_n' \dot{v}_{\theta_0}(Z_i)] I(Z_i \leqslant z)$$
$$- \Delta_n' n^{-1/2} \sum_{i=1}^n \frac{\dot{v}_{\theta_0}(Z_i)}{\sigma_{\theta_0}(Z_i)} \left[\frac{\sigma_{\theta_0}(Z_i)}{\sigma_n(Z_i)} - 1 \right] I(Z_i \leqslant z). \tag{4.1}$$

We first prove that

$$\sup_{1 \leqslant i \leqslant n} |\sigma_n^2(Z_i) - \sigma_{\theta_0}^2(Z_i)| = o_p(1). \tag{4.2}$$

To see this, note that $\sigma_n^2(Z_i) - \sigma_{\theta_0}^2(Z_i) = \tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i) + s_n^2 - \sigma_\varepsilon^2$, where s_n^2 is defined in (1.4). Since $s_n^2 - \sigma_\varepsilon^2 = o_p(1)$, so to prove (4.2) it suffices to show that

$$\sup_{1 \leqslant i \leqslant n} |\tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i)| = o_p(1). \tag{4.3}$$

Using (m), one verifies that for some constant c,

$$|\tau_n^2(Z_i) - \tau_{\theta_0}^2(Z_i)| \leqslant c \|\theta_n - \theta_0\|^2 \int r^2(y) f_\eta(y - Z_i)\, dy. \tag{4.4}$$

Also from (m), we know that $Er^4(X) < \infty$, therefore, $\max_{1 \leqslant i \leqslant n} |\int r^2(y) f_\eta(y - Z_i)\, dy| = o_p(n^{1/2})$. This fact, together with (2.10), implies (4.3), and hence (4.2).

Now, we shall show that the third term in the r.h.s. of (4.1) is of the order $u_p(1)$. Using (1.5), this term can be written as the sum

$$n^{-1/2} \sum_{i=1}^n \frac{(\sigma_\varepsilon^2 - s_n^2)\zeta_i I(Z_i \leqslant z)}{(\sigma_{\theta_0}(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)} + n^{-1/2} \sum_{i=1}^n \frac{(\tau_{\theta_0}^2(Z_i) - \tau_n^2(Z_i))\zeta_i I(Z_i \leqslant z)}{(\sigma_{\theta_0}(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)}. \tag{4.5}$$

But the first term in (4.5) can be rewritten as $\sqrt{n}(\sigma_\varepsilon^2 - s_n^2)$ multiplied by

$$\frac{1}{n} \sum_{i=1}^n \zeta_i I(Z_i \leqslant z) \left(\frac{1}{(\sigma_{\theta_0}(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)} - \frac{1}{2\sigma_{\theta_0}^2(Z_i)} \right) + \frac{1}{n} \sum_{i=1}^n \frac{\zeta_i I(Z_i \leqslant z)}{2\sigma_{\theta_0}^2(Z_i)}.$$

By (4.2), $\max_{1 \leqslant i \leqslant n} |[(\sigma_{\theta_0}(Z_i) + \sigma_n(Z_i))\sigma_n(Z_i)]^{-1} - (2\sigma_{\theta_0}^2(Z_i))^{-1}| = o_p(1)$. By a Glivenko–Cantelli-type argument, $\sup_{z \in \bar{\mathbb{R}}} |n^{-1} \sum_{i=1}^n \zeta_i I(Z_i \leqslant z)/2\sigma_{\theta_0}^2(Z_i)| = o_p(1)$. Also, the condition (e) ensures that $\sqrt{n}(\sigma_\varepsilon^2 - s_n^2) = O_p(1)$. These facts together then imply that the first term in (4.5) is of the order $u_p(1)$.

From (4.4), the second term in (4.5) is bounded above by

$$c\sqrt{n} \|\Delta_n\|^2 \frac{1}{2n\sigma_\varepsilon^2} \sum_{i=1}^n |\zeta_i| \int r^2(y) f_\eta(y - Z_i)\, dy = o_p(1).$$

Thus, we have proved that the third term in the r.h.s. of (4.1) is of the order $u_p(1)$. The fourth term in the r.h.s. of (4.1) is bounded above by the l.h.s. of the assumption $(v1)$ multiplied by $\sigma_\varepsilon^{-1}\sqrt{n}\Delta_n$, and hence $u_p(1)$. Similarly, $(v3)$, (2.10) and (4.2) imply that the fifth term is $u_p(1)$. These facts and a Glivenko–Cantelli-type argument shows that

$$\sup_{z \leqslant z_0} \left| \hat{V}_n(z) - V_n(z) - \sqrt{n}\Delta_n' E \frac{\dot{v}_{\theta_0}(Z)}{\sigma_{\theta_0}(Z)} I(Z \leqslant z) \right| = o_p(1). \tag{4.6}$$

For convenience, let $A_{n1}(z)$ and $A_{n2}(z)$ denote the second terms in the r.h.s of $\mathscr{W}_{\theta_0,G}(z)$ and $\mathscr{W}_n(z)$, respectively. In order to analyze the difference $A_{n2}(z)-A_{n1}(z)$, let

$$U_n(x)=\int_{y\geqslant x}\frac{\dot{v}_{\theta_0}(y)}{\sigma_{\theta_0}(y)}V_n(\mathrm{d}y),\quad \hat{U}_n(x)=\int_{y\geqslant x}\frac{\dot{v}_n(y)}{\sigma_n(y)}\hat{V}_n(\mathrm{d}y).$$

Then,

$$A_{n1}(z)=\int_{x\leqslant z}\frac{\dot{v}'_{\theta_0}(x)}{\sigma_{\theta_0}(x)}L^{-1}_{\theta_0}(x)U_n(x)G(\mathrm{d}x),\quad A_{n2}(z)=\int_{x\leqslant z}\frac{\dot{v}'_n(x)}{\sigma_n(x)}L^{-1}_n(x)\hat{U}_n(x)G_n(\mathrm{d}x). \tag{4.7}$$

Rewrite the difference $\hat{U}_n(x)-U_n(x)$ as

$$n^{-1/2}\sum_{i=1}^n I(Z_i\geqslant x)\left(\frac{\dot{v}_n(Z_i)}{\sigma^2_n(Z_i)}\cdot(Y_i-v_n(Z_i))-\frac{\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\cdot(Y_i-v_{\theta_0}(Z_i))\right).$$

This can be further decomposed into the sum of the following seven terms:

$$B_{n1}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_n(Z_i)-\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\left(\frac{\sigma^2_{\theta_0}(Z_i)}{\sigma^2_n(Z_i)}-1\right)(Y_i-v_{\theta_0}(Z_i))\right],$$

$$B_{n2}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_n(Z_i)-\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}(Y_i-v_{\theta_0}(Z_i))\right],$$

$$B_{n3}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_n(Z_i)-\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\left(\frac{\sigma^2_{\theta_0}(Z_i)}{\sigma^2_n(Z_i)}-1\right)(v_{\theta_0}(Z_i)-v_n(Z_i))\right],$$

$$B_{n4}(x)=-n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_n(Z_i)-\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}(v_n(Z_i)-v_{\theta_0}(Z_i))\right],$$

$$B_{n5}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\left(\frac{\sigma^2_{\theta_0}(Z_i)}{\sigma^2_n(Z_i)}-1\right)(Y_i-v_{\theta_0}(Z_i))\right],$$

$$B_{n6}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\left(\frac{\sigma^2_{\theta_0}(Z_i)}{\sigma^2_n(Z_i)}-1\right)(v_{\theta_0}(Z_i)-v_n(Z_i))\right],$$

$$B_{n7}(x)=n^{-1/2}\sum_{i=1}^n I[Z_i\geqslant x]\left[\frac{\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}(v_{\theta_0}(Z_i)-v_n(Z_i))\right].$$

Using standard but somewhat lengthy arguments one verifies that under assumptions (m), ($v1$), ($v2$) and (2.10), $\sup_{x\in\bar{\mathbb{R}}}|B_{nj}(x)|=o_{\mathrm{p}}(1)$, for $j=1,\ldots,6$. Fact (1.4) is often used in these calculations.

Now consider $B_{n7}(x)$. Using conditions ($v1$) and ($v3$), we obtain

$$\begin{aligned}B_{n7}(x)&=\frac{1}{\sqrt{n}}\sum_{i=1}^n I[Z_i\geqslant x]\frac{\dot{v}_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}(v_{\theta_0}(Z_i)-v_n(Z_i)+\Delta'_n\dot{v}_{\theta_0}(Z_i))\\&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^n I[Z_i\geqslant x]\frac{\dot{v}_{\theta_0}(Z_i)\dot{v}'_{\theta_0}(Z_i)}{\sigma^2_{\theta_0}(Z_i)}\Delta_n=-L_{\theta_0}(x)\sqrt{n}\Delta_n+u_{\mathrm{p}}(1).\end{aligned}$$

Hence,

$$\sup_{x\leqslant z_0}\|\hat{U}_n(x)-U_n(x)+L_{\theta_0}(x)\sqrt{n}\Delta_n\|=u_{\mathrm{p}}(1). \tag{4.8}$$

Next, we consider the asymptotic behavior of the matrix $L_n(x)$. Note that

$$\|L_n(x)-L_{\theta_0}(x)\| = \left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\dot v_n(Z_i)\dot v_n'(Z_i)}{\sigma_n^2(Z_i)}I[Z_i\geqslant x]-L_{\theta_0}(x)\right\|$$

$$\leqslant \frac{1}{n}\sum_{i=1}^{n}\frac{\|\dot v_n(Z_i)-\dot v_{\theta_0}(Z_i)\|^2}{\sigma_n^2(Z_i)}+\frac{2}{n}\sum_{i=1}^{n}\frac{\|\dot v_{\theta_0}(Z_i)\|\|\dot v_n(Z_i)-\dot v_{\theta_0}(Z_i)\|}{\sigma_n^2(Z_i)}$$

$$+\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\dot v_{\theta_0}(Z_i)\dot v_{\theta_0}'(Z_i)}{\sigma_{\theta_0}^2(Z_i)}\left(\frac{\sigma_{\theta_0}^2(Z_i)}{\sigma_n^2(Z_i)}-1\right)\right\|$$

$$+\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{\dot v_{\theta_0}(Z_i)\dot v_{\theta_0}'(Z_i)}{\sigma_{\theta_0}^2(Z_i)}I[Z_i\geqslant x]-L_{\theta_0}(x)\right\|.$$

Using conditions ($v2$) and (4.2), we can show that the first three terms are $u_{\mathrm{p}}(1)$. The last term is also $u_{\mathrm{p}}(1)$ by showing the tightness of $n^{-1}\sum_{i=1}^{n}\dot v_{\theta_0}(Z_i)\dot v_{\theta_0}'(Z_i)I[Z_i\leqslant x]/\sigma_{\theta_0}^2(Z_i)$. Consequently,

$$\sup_{x\leqslant z_0}\|L_n^{-1}(x)-L_{\theta_0}^{-1}(x)\| = o_{\mathrm{p}}(1). \tag{4.9}$$

We also need to show that

$$\sup_{z\leqslant z_0}\left|A_{n2}(z)-A_{n1}(z)+\sqrt{n}\Delta_n'\int_{x\leqslant z}\frac{\dot v_{\theta_0}(x)}{\sigma_{\theta_0}(x)}G(\mathrm{d}x)\right| = o_{\mathrm{p}}(1), \tag{4.10}$$

where $A_{n1}(z)$ and $A_{n2}(z)$ are defined in (4.7). This is achieved by rewriting

$$A_{n2}(z)=\int_{x\leqslant z}\frac{\dot v_{\theta_0}'(x)}{\sigma_{\theta_0}(x)}L_{\theta_0}^{-1}(x)(U_n(x)-L_{\theta_0}(x)\sqrt{n}\Delta_n)G_n(\mathrm{d}x)+\mathscr{R}_n(z).$$

Using conditions ($v2$), (4.2), (4.9) and (4.8), one shows that $\sup_{z\leqslant z_0}|\mathscr{R}_n(z)| = o_{\mathrm{p}}(1)$, while the first term equals $A_{n1}(z)-\sqrt{n}\Delta_n'\int_{x\leqslant z}(\dot v_{\theta_0}(x)/\sigma_{\theta_0}(x))G_n(\mathrm{d}x)+u_{\mathrm{p}}(1)$. Using a Glivenko–Cantelli-type argument, one further concludes that

$$\int_{x\leqslant z}\frac{\dot v_{\theta_0}(x)}{\sigma_{\theta_0}(x)}G_n(\mathrm{d}x)=\int_{x\leqslant z}\frac{\dot v_{\theta_0}(x)}{\sigma_{\theta_0}(x)}G(\mathrm{d}x)+u_{\mathrm{p}}(1).$$

These facts then imply (4.10).

Finally, claim (a) follows from $\mathscr{W}_n(z)-\mathscr{W}_{\theta_0,G}(z)=(\hat V_n(z)-V_n(z))-(A_{n2}(z)-A_{n1}(z))$, (4.6) and (4.10). For any details of the above proof contact the authors of this paper.

Acknowledgment

The authors would like to thank the Associate Editor for critical comments that helped to improve the presentation.

References

An, H.Z., Cheng, B., 1991. A Kolmogorov–Smirnov type statistic with application to test for nonlinearity in time series. Internat. Statist. Rev. 59, 287–307.

Berkson, J., 1950. Are these two regressions? J. Amer. Statist. Assoc. 45, 164–180.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York, London, Sydney.

Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement Error in Nonlinear Models. Chapman & Hall/CRC Press, Boca Raton, FL.

Fan, J., 1991a. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19, 1257–1272.

Fan, J., 1991b. Asymptotic normality for deconvolution kernel density estimators. Sankhyā Ser. A. 53, 97–110.

Fan, J., Truong, K.T., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21, 1900–1925.

Fuller, W.A., 1987. Measurement Error Models. J Wiley, New York.
Hart, J.D., 1997. Nonparametric Smoothing and Lack-of-Fit Tests. Springer, New York.
Huwang, L., Huang, Y.H.S., 2000. On errors-in-variables in polynomial regression–Berkson case. Statist. Sinica 10, 923–936.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240–257.
Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovation martingales. Ann. Statist. 21, 798–829.
Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32 (3), 995–1034.
Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models, second ed. Lecture Notes Series in Statistics, vol. 166. Springer, New York.
Koul, H.L., 2006. Model diagnostics via martingale transforms: a brief review. In: Fan, J., Koul, H.L. (Eds.), Frontiers in Statistics. Imperial College Press, London, UK, pp. 183–206.
Koul, H.L., Ni, P., 2004. Minimum distance regression model checking. J. Statist. Plann. Inference 119, 109–141.
Rudemo, M., Ruppert, D., Streibig, J., 1989. Random-effect models in nonlinear regression with applications to bioassay. Biometrics 45, 349–362.
Stute, W., 1997. Nonparametric model checks for regression. Ann. Statist. 25, 613–641.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: an innovation process approach. Ann. Statist. 26, 1916–1934.
Wang, L., 2003. Estimation of nonlinear Berkson-type measurement errors models. Statist. Sinica 13, 1201–1210.
Wang, L., 2004. Estimation of nonlinear models with Berkson measurement errors. Ann. Statist. 32, 2559–2579.
Zhu, L.X., Cui, H., 2004. Some properties of a lack-of-fit for a linear errors in variables model. Acta Math. Appl. Sinica 20, 533–540.
Zhu, L.X., Cui, H., 2005. Testing lack-of-fit for general linear errors in variables models. Statist. Sinica 15, 1049–1068.
Zhu, L.X., Song, W.X., Cui, H.J., 2003. Testing lack-of-fit for a polynomial errors-in-variables model. Acta Math. Appl. Sinica (English Ser.) 19, 353–362.