Statistics and Probability Letters 79 (2009) 765–773

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Lack-of-fit testing in errors-in-variables regression model with validation data ☆

Weixing Song

108F, Dickens Hall, Department of Statistics, Kansas State University, Manhattan, KS, 66502, United States

ARTICLE INFO

Article history:
Received 16 July 2008
Received in revised form 29 October 2008
Accepted 29 October 2008
Available online 8 November 2008

MSC:
primary 62G08
secondary 62G10

ABSTRACT

A score-type test procedure is proposed for checking the adequacy of the errors-in-variables regression model when validation data are available. Under mild conditions, the score-type test statistic is proven to be asymptotically normal. The test procedure is shown to be consistent against general fixed alternatives and it can detect local alternatives which are close to the null model at the parametric rate. Monte-Carlo simulations are conducted to evaluated the finite sample performance of the proposed test.

1. Introduction

The relationship between a random variable Y and a random vector X is often described as a regression model. The classic regression assumes that both Y and X are observable. But the real world does not often work in that way. Due to various reasons, the predictor X cannot be observed directly. Instead, a surrogate, say Z, is available. X and Z are usually assumed to have an additive relation $Z = X + U$ with U being the measurement error. There are some cases in agricultural and biological studies where the Berkson measurement error structure $X = Z + U$ is assumed. In practice, the relationship between X and Z could be more complicated. See Fuller (1987), Cheng and Van Ness (1999) for more discussion on these topics.

Any statistical inferences on the relationship between Y and X when X cannot be observed, should rely on the observations from (Y, Z). A commonly used method in the literature is the so called regression calibration technique in which the regression of Y on Z is modeled. But to proceed, we have to estimate the conditional distribution of X given Z, or the conditional expectation $E[\mu(X)|Z]$, where $\mu(x) = E(Y|X = x)$. If $Z = X + U$ is assumed, the simulation extrapolation (SIMEX) and deconvolution technique could be used to fulfill this purpose. The pros and cons of these methods can be found in Cook and Stefanski (1994), Stefanski and Cook (1995), Fan (1991), Fan and Truong (1993) and the references therein.

The technical difficulties encountered in evaluating $E[\mu(X)|Z]$ can be avoided when validation data are available. A validation data set contains observations on (X, Z) from another study, they are independent of the current data or the primary data which only contain information on (Y, Z). One economic example in Sepanski and Lee (1995) reports that earnings and work hours in the survey of the labor market are often measured with errors. A validation data set in this case can be obtained from the administrative payroll records of employees' earnings and work hours. More examples can be found in Pepe and Fleming (1991) and Pepe (1992). With help from the validation data set, the conditional distribution, hence the conditional expectation of, certain functions of X given Z could be estimated. More importantly, unlike the classical measurement error models, the predictor X and the surrogate Z do not have to be of the same dimension. This greatly expands the extent of application of the errors-in-variables model. However, the literature mainly focuses on the estimation issue. The model checking study when validation data are available seems scant in the literature. See Carroll et al. (1995), Sepanski and Carroll (1993), Wang (1999, 2003), Zhu et al. (2003), Koul and Song (forthcoming) and Song (2008). This paper will try to fill this void.

☆ The research is supported by USRG No. 2232 of Kansas State University.
E-mail address: weixing@ksu.edu.

0167-7152/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.10.032

Consider the following errors-in-variables regression model:

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

where Y is 1-dimensional random variable, X is d-dimensional random predictor which is measured with error, ε is 1-dimensional random error. Although X cannot be observed directly, we can observe the surrogate Z of p-dimension. d and p do not have to be the same. Throughout the paper, we will assume that the primary data set contains n i.i.d. copies (Y_i, Z_i), $i = 1, 2, \ldots, n$ from model (1.1), the random errors ε_i are i.i.d. with mean 0 and unknown variance σ_ε^2, the validation data set contains N i.i.d. observations on (X, Z), denoted by $(\widetilde{X}_j, \widetilde{Z}_j)$, $j = n+1, n+2, \ldots, n+N$, which are independent of the primary data set. One crucial assumption made here is that the relationship between $\widetilde{X}$ and $\widetilde{Z}$ in the validation data set is the same as the one in the primary data set. The main objective in this paper is to test the following hypothesis

$$H_0 : \mu(x) = m(x, \theta), \quad \text{for some } \theta \in \Theta, \tag{1.2}$$

where $\Theta \subset \mathbb{R}^q$, and $m(x, \theta)$ is a known function with unknown parameter θ. Some conditions on the null regression function $m(x, \theta)$ will be specified in the next section. In the sequel, if v is a vector, then v^2 will be understood as vv'.

The paper is organized as follows: assumptions on the model will be stated in Section 2; the test statistic and its asymptotic distribution under the null hypothesis will be presented in Section 3, and the consistency and the power of the test against certain fixed and a sequence of local alternatives, and the selection of weight function will be also discussed there; simulation studies will be conducted in Section 4; and all the proofs will be provided in the Appendix.

2. Estimation of θ

First we should find a way to estimate the unknown parameter θ under the null hypothesis (1.2). By conditioning on Z, (1.1) can be transformed to $Y = g(Z) + e = E(\mu(X)|Z) + e$, where $e = Y - E(\mu(X)|Z)$ is uncorrelated with Z. Let $g(z, \theta) = E(m(X, \theta)|Z = z)$. Then to test (1.2), one can test the new hypothesis

$$H_0 : g(z) = g(z, \theta), \quad \text{for all } z, \text{and for some } \theta \in \Theta \tag{2.1}$$

in the transformed model. Note that (2.1) and (1.2) are not equivalent in general, but in some cases they are. For example, if the family of densities $\{f_u(\cdot - z) : z \in \mathbb{R}\}$ is complete, then this holds, where $f_u(\cdot)$ is the density function of the measurement error U. Under (2.1), (Y_i, Z_i), $i = 1, 2, \ldots, n$ can be viewed as copies from $Y = g(Z, \theta) + e$, which is a classical nonlinear regression model. We can use the least square or quasi-likelihood procedure to estimate θ if g is known. A detailed description on least square procedure and quasi-likelihood method can be found in Jennrich (1969), Wu (1981), Carroll and Ruppert (1988), among others. However, g is unknown in our current setup, since $g(z, \theta) = E(m(X, \theta)|Z = z)$ depends on the conditional distribution of X given Z which is unknown. This difficulty can be overcome by exploiting the validation data $(\widetilde{X}_j, \widetilde{Z}_j)$, $j = n+1, n+2, \ldots, n+N$. In fact, we can estimate $g(z, \theta)$ by the kernel smoother

$$\hat{g}(z, \theta) = \frac{N^{-1} \sum_{j=n+1}^{n+N} K_h(z - \widetilde{Z}_j) m(\widetilde{X}_j, \theta)}{N^{-1} \sum_{j=n+1}^{n+N} K_h(z - \widetilde{Z}_j)} =: \frac{Q_N(z, \theta)}{\hat{f}_Z(z)},$$

where K is a kernel function, h is the bandwidth, $K_h(\cdot) = K(\cdot/h)/h^p$. To avoid the extrapolation and the technical difficulties from edge effects, all kernel regression estimates used in the following are restricted on a compact set $\mathcal{C}$ interior to the support of Z. All sums in the following with respect to the primary data are taken only for those $Z \in \mathcal{C}$. The truncation may cause some loss in efficiency, but it can be counterbalanced by a gain in robustness and an easily described theory. See Carroll and Wand (1991), Sepanski and Carroll (1993) for a discussion on the merits of this technique.

Therefore, we can estimate θ by the modified least square procedure

$$\hat{\theta}_n = \operatorname{argmin}_{\theta \in \Theta} \sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)[Y_i - \hat{g}(Z_i, \theta)]^2, \tag{2.2}$$

where $I_{\mathcal{C}}(z)$ is the indicator function of $\mathcal{C}$. Consistency and the asymptotic normality of the above defined $\hat{\theta}_n$ has been extensively discussed by Sepanski and Lee (1995) under some general assumptions. Their results enable us to construct a lack-of-fit test procedure with much less effort.

3. Score-type test

The score-type test statistic for checking the null hypothesis (2.1) is defined by the following weighted sum of residuals,

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \hat{g}(Z_i, \hat{\theta}_n)) W(Z_i) I_{\mathcal{C}}(Z_i), \tag{3.1}$$

where W is a weight function and $\hat{\theta}_n$ is defined in (2.2). The choice of the weight function will be discussed later. Some conditions are needed to ensure a nice theory for S_n. For completeness, the assumptions used in Sepanski and Lee (1995), after some slight modifications, are reproduced here, which guarantees the asymptotic normality of $\hat{\theta}_n$. In the sequel, $\dot{g}(z,\theta)$ and $\ddot{g}(z,\theta)$ denote the first and second derivative of g with respect to θ, respectively.

Assumptions:

(C1) The set $\mathcal{C}$ of dimension p is a compact subset of the support Z or $\widetilde{Z}$ and the parameter space Θ of dimension q is compact. $\mathcal{C}$ is bounded away from the boundary of the support $\widetilde{Z}$.
(C2) The function $m(x,\theta)$ is differentiable with respect to θ. Both $\sup_{\theta\in\Theta} m(X,\theta)$ and $\sup_{\theta\in\Theta} \dot{m}(X,\theta)$ have finite second moments.
(C3) The kernel function K on $\mathbb{R}^p$ is of order k and the derivative $\partial K(v)/\partial v$ are bounded.
(C4) The density function f_Z and $g(Z,\theta)$ are in $\mathcal{D}^k$, where

$$\mathcal{D}^k = \left\{ f : \left(\frac{\partial}{\partial v_1}\right)^{i_1} \left(\frac{\partial}{\partial v_2}\right)^{i_2} \cdots \left(\frac{\partial}{\partial v_p}\right)^{i_p} f(v_1, v_2, \dots, v_p) \right.$$
$$\left. \text{are continuous and uniformly bounded for } 0 \le i_1 + i_2 + \cdots + i_p \le k \right\}.$$

(C5) $N \to \infty$, $h \to 0$, $n \to \infty$, $Nh^{2p}/\ln N \to \infty$, and $\lim_{n\to\infty} n/N = \lambda$ is finite.
(C6) Y and $g(Z,\theta)$ have finite second moments.
(C7) $g(Z,\theta_1) \ne g(Z,\theta_2)$ with positive probability on $\mathcal{C}$ for all $\theta_1 \ne \theta_2$.
(C8) $g(z,\theta)$ is third order differentiable with respect to θ at θ_0. The function $\dot{g}$ is in $\mathcal{D}^k$. Both $\sup_{\theta\in\Theta} \ddot{g}(Z,\theta)$ and $\sup_{\theta\in\mathcal{N}(\theta_0)} \partial \mathrm{vec}(\ddot{g}(Z,\theta))/\partial\theta'$ have finite second moments, where $\mathcal{N}(\theta_0)$ is a compact neighborhood of θ_0. θ_0 is the true value of θ under the null hypothesis (1.2).
(C9) $E(Y^2|Z)$ has finite $1+\delta$ moment and $g, \dot{g}, m$ and $\dot{m}$ have finite $2+\delta$ moments for some $\delta > 0$.
(C10) $\inf_{z\in\mathcal{C}} f_Z(z) > 0$.
(C11) $Nh^{2k} \to 0$ as $N \to \infty$.
(C12) $I(\theta_0)$ is nonsingular and the first moments of $\Phi(\theta_0)$ and $\Delta(\theta_0)$ exist under the null hypothesis (2.1), where

$$I(\theta) = EI_{\mathcal{C}}(Z)\dot{g}(Z,\theta)\dot{g}'(Z,\theta), \qquad \Phi(\theta,Z) = I_{\mathcal{C}}(Z)[Y - g(Z,\theta)]^2\dot{g}(Z,\theta)\dot{g}'(Z,\theta)$$
$$\Delta(\theta,X,Z) = I_{\mathcal{C}}(Z)[m(X,\theta) - g(Z,\theta)]^2\dot{g}(Z,\theta)\dot{g}'(Z,\theta).$$

The consistency of $\hat{\theta}_n$ is guaranteed by (C1)–(C7). With additional (C8)–(C12), the asymptotic normality of $\hat{\theta}_n$ can be proved. To develop an asymptotic theory for S_n, a stochastic expansion of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is needed, which is not given explicitly in Sepanski and Lee (1995), but they do point out one way to obtain such stochastic expansions. Following their suggestion, we can show that

Lemma 3.1. *Suppose Assumptions* (C1)–(C12) *hold. Then, under* H_0,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_{\mathcal{C}}(Z_i)[Y_i - g(Z_i,\theta_0)]I^{-1}(\theta_0)\dot{g}(Z_j,\theta_0)$$
$$- \frac{\sqrt{\lambda}}{\sqrt{N}} \sum_{j=n+1}^{n+N} I_{\mathcal{C}}(\widetilde{Z}_j)(m(\widetilde{X}_j,\theta_0) - g(\widetilde{Z}_j,\theta_0))I^{-1}(\theta_0)\dot{g}(\widetilde{Z}_j,\theta_0) + o_p(1),$$

where λ *is as in* (C5).

From Lemma 3.1, we can show that

Theorem 3.1. *Suppose Assumptions* (C1)–(C12) *hold. Then under* H_0, $S_n \Longrightarrow N(0, \sigma_\xi^2 + \sigma_\zeta^2)$, *where*

$$\sigma_\xi^2 = EI_{\mathcal{C}}(Z)\sigma_\eta^2(Z)[W(Z) - \dot{g}'(Z,\theta_0)I^{-1}(\theta_0)EI_{\mathcal{C}}(Z)\dot{g}(Z,\theta_0)W(Z)]^2,$$
$$\sigma_\zeta^2 = \lambda EI_{\mathcal{C}}(Z)\tau^2(Z)[W(Z) - \dot{g}'(Z,\theta_0)I^{-1}(\theta_0)EI_{\mathcal{C}}(Z)\dot{g}(Z,\theta_0)W(Z)]^2.$$

The implementation of lack-of-fit test of (2.1) based on S_n needs a consistent estimator of the asymptotic variance $\sigma_\xi^2 + \sigma_\zeta^2$. Let $\rho = EI_{\mathcal{C}}(Z)\dot{g}(Z,\theta_0)W(Z)$. Simple calculation shows that

$$\sigma_\xi^2 = EI_{\mathcal{C}}(Z)(Y - g(Z,\theta_0))^2[W(Z) - \dot{g}'(Z,\theta_0)I^{-1}(\theta_0)\rho]^2, \tag{3.2}$$

$$\sigma_\zeta^2 = \lambda EI_{\mathcal{C}}(Z)(m(X,\theta_0) - g(Z,\theta_0))^2[W(Z) - \dot{g}'(Z,\theta_0)I^{-1}(\theta_0)\rho]^2. \tag{3.3}$$

A consistent estimator for $I(\theta_0)$ is given by $\widehat{I}_n(\hat\theta_n) = n^{-1}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\dot{\hat{g}}(Z_i,\hat\theta_n)\dot{\hat{g}}'(Z_i,\hat\theta_n)$, and a consistent estimator for ρ is given by $\widehat{\rho}_n = n^{-1}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\dot{\hat{g}}(Z_i,\hat\theta_n)W(Z_i)$. Therefore, a consistent estimator for σ_ξ^2 can be defined as

$$\widehat{\sigma}_\xi^2 = \frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[Y_i - \hat g(Z_i,\hat\theta_n)]^2[W(Z_i) - \dot{\hat{g}}'(Z_i,\hat\theta_n)\widehat{I}_n^{-1}(\hat\theta_n)\widehat{\rho}_n]^2. \tag{3.4}$$

Since (X, Z) can only be observed in the validation data set, so except for $I(\theta_0)$ and ρ, we will estimate σ_ζ^2 by

$$\widehat{\sigma}_\zeta^2 = \frac{\lambda}{N}\sum_{j=n+1}^{n+N} I_{\mathcal{C}}(\widetilde{Z}_j)[m(\widetilde{X}_j,\hat\theta_n) - \hat g(\widetilde{Z}_j,\hat\theta_n)]^2[W(\widetilde{Z}_j) - \dot{\hat{g}}'(\widetilde{Z}_j,\hat\theta_n)\widehat{I}_n^{-1}(\hat\theta_n)\widehat{\rho}_n]^2. \tag{3.5}$$

Therefore, a consistent estimator of the asymptotic variance can be defined as $\widehat{\sigma}_\xi^2 + \widehat{\sigma}_\zeta^2$. Under (2.1), we have

$$T_n = S_n \Big/ \sqrt{\widehat{\sigma}_\xi^2 + \widehat{\sigma}_\zeta^2} \Longrightarrow N(0, 1). \tag{3.6}$$

Thus, one can reject the null hypothesis at significance level α whenever $|T_n| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the upper $100\alpha/2$-th percentile.

In the following, we shall show that the score-type test T_n has the property of being consistent for certain alternatives. Consider the following alternative model

$$H_1 : Y = d(X_i) + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{3.7}$$

where $d(\cdot)$ does not belong to the parametric class under the null hypothesis (1.2). The calibrated regression model now takes the form of

$$Y_i = h(Z_i) + \xi_i, \quad i = 1, 2, \ldots, n \tag{3.8}$$

with $h(Z) = E[d(X)|Z]$, and $h(\cdot)$ does not belong to the parametric class under the null hypothesis (2.1). In addition to the assumptions (C1)–(C11), further assume

(C13) the second moment of $h(Z)$ is finite, or $E[h^2(Z)] < \infty$,
(C14) $EI_{\mathcal{C}}[d(X) - g(Z,\theta)]^2$ has a unique minimizer θ_a,
(C15) $I(\theta_a) = EI_{\mathcal{C}}(Z)[\dot g(Z,\theta_a)\dot g'(Z,\theta_a) - (d(X) - g(Z,\theta_a))\ddot g(Z,\theta_a)]$ is nonsingular
(C16) under alternative (3.7), the expectations $EI_{\mathcal{C}}(Z)[(Y - g(Z,\theta_a))\dot g(Z_i,\theta_a)]^2$, $EI_{\mathcal{C}}(Z)\{[h(Z) - g(Z,\theta_a)][\dot m(X,\theta_a) - \dot g(Z,\theta_a)]\}^2$, $EI_{\mathcal{C}}(Z)\{[m(X,\theta_a) - g(Z,\theta_a)]\dot g(Z,\theta_a)\}^2$ are finite.

The following theorem states that, under the alternative model (3.7) or (3.8), the asymptotic normality of the LSE $\hat\theta_n$ defined by (2.2) still holds.

Theorem 3.2. *Suppose* (C1)–(C11) *and* (C13)–(C16) *holds. Then, under the alternative hypothesis* (3.7), $\sqrt{n}(\hat\theta_n - \theta_a) \Longrightarrow N(0, I^{-1}(\theta_a)\Sigma_a I^{-1}(\theta_a))$, *where*

$$\begin{aligned}\Sigma_a &= EI_{\mathcal{C}}(Z)[(Y - g(Z,\theta_a))\dot g(Z_i,\theta_a)]^2 \\ &\quad + \lambda EI_{\mathcal{C}}(Z)\{[h(Z) - g(Z,\theta_a)][\dot m(X,\theta_a) - \dot g(Z,\theta_a)] + [m(X,\theta_a) - g(Z,\theta_a)]\dot g(Z,\theta_a)\}^2,\end{aligned}$$

and the expectation is taken under the alternative hypothesis (3.7).

Under (3.7), S_n can be written as

$$\begin{aligned}S_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i W(Z_i)I_{\mathcal{C}}(Z_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^n[\hat g(Z_i,\hat\theta_n) - g(Z_i,\theta_a)]W(Z_i)I_{\mathcal{C}}(Z_i) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n[d(X_i) - g(Z_i,\theta_a)]W(Z_i)I_{\mathcal{C}}(Z_i).\end{aligned} \tag{3.9}$$

Similar to the proof of Theorem 3.1, we can show that the first two terms on the right hand side of (3.9) are of order $O_p(1)$. If we assume

(C16) $E[d(X) - m(X,\theta)]W(Z)I_{\mathcal{C}}(Z) \neq 0$ for any $\theta \in \Theta$,

then it is easy to see the third term on the right hand side of (3.9) will go to infinity in probability. Also, applying a uniform convergence argument on $\hat\sigma_\xi^2$ and $\hat\sigma_\zeta^2$ defined in (3.4) and (3.5), one can show these two quantities also converge to σ_ξ^2 and σ_ζ^2, defined in Theorem 3.1 with θ_0 replaced by θ_a. Thus, we have

Theorem 3.3. *Suppose conditions* (C1)–(C11) *and* (C13)–(C16) *hold. Then, under the alternative hypothesis* (3.7), $|T_n| \to \infty$ *in probability*.

The power property of T_n is often analyzed for a sequence of local alternatives of the form

$$H_{an} : Y = m(X, \theta_0) + \frac{1}{\sqrt{n}}\delta(X) + \varepsilon, \tag{3.10}$$

where $\delta(x)$ is a known function which does not belong to the parametric family as $m(x, \theta_0)$ does. The following theorem states the asymptotic properties of the least square estimator $\hat{\theta}_n$ defined in (2.2) and the score-type test statistic T_n given by (3.6).

Theorem 3.4. *In addition to the assumptions* (C1)–(C12), *further assume* $E\delta^2(X) < \infty$. *Then under the local alternatives* (3.10),

(1) $\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(I^{-1}(\theta_0)EI_{\mathcal{C}}(Z)\delta(X)\dot{g}(Z, \theta_0), \Sigma_0)$, *where*

$$\Sigma_0 = EI_{\mathcal{C}}(Z)[(m(X, \theta_0) + \varepsilon - g(Z, \theta_0))\dot{g}(Z_i, \theta_0)]^2 + \lambda EI_{\mathcal{C}}(Z)\{[m(X, \theta_0) - g(Z, \theta_0)]\dot{g}(Z, \theta_0)\}^2.$$

(2) $T_n \sim N(C, 1)$, *where* $C = EI_{\mathcal{C}}(Z)\delta(X)[W(Z) - \dot{g}(Z, \theta_0)I^{-1}(\theta_0)\rho] \Big/ \sqrt{\sigma_\xi^2 + \sigma_\zeta^2}$.

One can see that, for the alternatives close to the null at $n^{-1/2}$ rate, the asymptotic power is given by $2 - \Phi(z_{\alpha/2} - C) + \Phi(z_{\alpha/2} + C)$ which is an increasing function of C^2, where Φ denotes the CDF of $N(0, 1)$. Hence, one criterion to choose the weight function W is to make C^2 larger. By Cauchy–Schwarz inequality, one has

$$C^2 = \frac{\left[EI_{\mathcal{C}}(Z)\delta(X)[W(Z) - \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho]\right]^2}{E[I_{\mathcal{C}}(Z)\sigma_\eta^2(Z) + \lambda EI_{\mathcal{C}}(Z)\tau^2(Z)][W(Z) - \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho]^2} \le E\left(\frac{I_{\mathcal{C}}(Z)d^2(X)}{\sigma_\eta^2(Z) + \lambda\tau^2(Z)}\right)$$

with equality holding if, and only if, for $Z \in \mathcal{C}$,

$$W(Z) = \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho + \delta(X)/[\sigma_\eta^2(Z) + \lambda\tau^2(Z)], \tag{3.11}$$

where $\sigma_\eta^2(Z) = E[(\varepsilon + m(X, \theta_0) - g(Z, \theta_0))^2|Z] = \sigma_\varepsilon^2 + \tau^2(Z)$. The optimal weight function given by (3.11) indeed has no explicit solution, since the ρ term also contains W. One way to select the weight function is to find a sub-optimal weight function. Replacing $W(Z)$ and $\delta(X)$ by $W^*(Z) = W(Z) - \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho$,

$$\delta^*(X) = \delta(X) - \frac{\dot{g}'(Z, \theta_0)}{\sigma^2(Z)}\left(E\left[\frac{\dot{g}(Z, \theta_0)\dot{g}'(Z, \theta_0)}{\sigma^4(Z)}\right]\right)^{-1}\left(E\left[\frac{I_{\mathcal{C}}(Z)\delta(X)\dot{g}(Z, \theta_0)}{\sigma^2(Z)}\right]\right)$$

where $\sigma^2(Z) = I_{\mathcal{C}}(Z)[\sigma_\eta^2(Z) + \lambda\tau^2(Z)]$, we have $EW^*(Z)\dot{g}(Z, \theta_0) = E\left[\delta^*(X)\dot{g}(Z, \theta_0)/\sigma^2(Z)\right] = 0$. Without loss of generality, we may assume

$$E(I_{\mathcal{C}}(Z)W(Z)\dot{g}(Z, \theta_0)) = 0, \qquad E(I_{\mathcal{C}}(Z)\delta(X)\dot{g}(Z, \theta_0)/\sigma^2(Z)) = 0.$$

Thus we should select the weight function W to make C^2 as large as possible, subject to the constraint $E(I_{\mathcal{C}}(Z)W(Z)\dot{g}(Z, \theta_0)) = 0$. In this case, application of the Cauchy–Schwarz inequality leads to the optimal choice of the weight function to be $W_o(Z) = c\delta(X)/\sigma^2(Z)$, which gives the maximum C^2 to be $E\left[I_{\mathcal{C}}(Z)\delta^2(X)/\sigma^2(Z)\right]$, where c is an arbitrary real constant. For simplicity, the constant c is often set to be 1. When $\delta(X)$ is known, we can just use $W_o(Z)$ with a consistent estimator of $\sigma(Z)$ as the weight function. But $\delta(X)$ is usually unknown, we cannot estimate $W_o(Z)$ consistently and, hence, it cannot be used as a weight function. In this case, we should choose $W(Z)$, at the very least, to have a non-zero correlation with $\delta(X)$ so that the test has nontrivial power. Zhu et al. (2004) suggest using a polynomial of Z as the weight function.

4. Simulation

This section presents two simulation studies to illustrate the finite sample performance of the score-type test. The nominal significance level is chosen to be 0.05 among all cases, and the empirical levels and powers are obtained by calculating the percentage of how many times that $|T_n| \ge z_{0.975}$ within 500 replications.

Simulation 1. In this simulation, the primary data (Y, Z) and the validation data are generated from: $Y = 1 + 2X + \varepsilon$, $Z = X + U$, where $\varepsilon \sim N(0, 0.5^2)$, $X \sim N(0, 1)$, $U \sim N(0, 0.5^2)$. The primary sample size n and the validation sample size N are chosen so that $N/n = 2$. In the simulation, we select $n = 100, 200, 400, 600$ and 800. Under the null hypothesis, Y and X are linear related. To investigate the power, the local alternative models are considered: $Y = 1 + 2X + cX^2 + \varepsilon$ for c between 0 and 2. The kernel function is chosen to be standard normal, the bandwidth h is set to be $N^{-1/3}$ which satisfies the conditions (C5) and (C11), and $W(Z) = Z + Z^2$ is used to be the weight function.

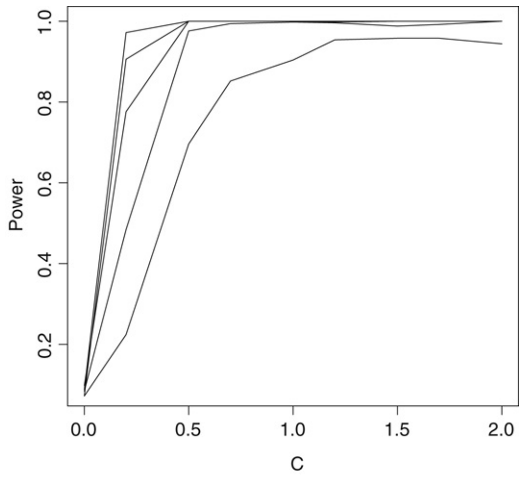

Fig. 1. Power curves in simulation 1.

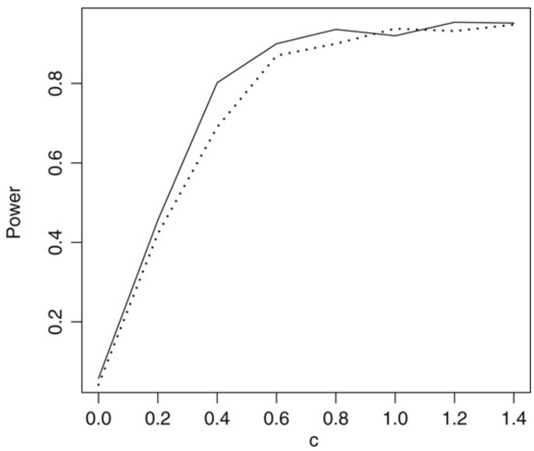

Fig. 2. Power curves in simulation 2.

Fig. 1 shows the power curves for different sample sizes when c walks from 0 to 2. The lowest one corresponds to the sample size 100, the second lowest one is for sample size 200, then 400, 600 and the highest power curve is for the sample size 800. Clearly, for all cases, the power increases when c increases. Our limited simulation studies show that the test are slightly liberal. For example, in this simulation, the empirical levels are 0.072, 0.084, 0.098, 0.072 and 0.088 for $n = 100, 200, 400, 600$ and 800. We also conduct a simulation study when X follows uniform distributions, and the same phenomenon occurs.

Simulation 2. Now we consider a nonlinear errors-in-variables regression model: $Y = \beta_1 \exp(\beta_2 X) + \varepsilon, Z = 1.25X + U$. In the simulation, the true values of the parameters are set to be $\beta_1 = 1$ and $\beta_2 = -1$, $\varepsilon \sim N(0, 0.5^2)$, $X \sim N(0, 1)$, $U \sim N(0, 0.5^2)$. n and N is chosen so that $n/N = 2$. $n = 800, 1000$ are used in the simulation. The regression function under the null hypothesis is $m(x, \theta) = \beta_1 \exp(\beta_2 x)$, where $\theta = (\beta_1, \beta_2)$. The power is investigated by considering the alternative models $Y = \beta_1 \exp(\beta_2 X) + cX^2 + \varepsilon$ with c varying between 0.2 to 1.4. The kernel function is chosen to be standard normal, the bandwidth h is chosen to be $\log(N)N^{-1/3}$ which satisfies the conditions (C5) and (C11), and again $W(Z) = Z + Z^2$ is used as the weight function.

Fig. 2 shows the power curves for different sample sizes when c walks from 0 to 1.4. The dotted line corresponds to the sample size 800, and the solid line is for sample size 1000. The power gets larger when c gets larger in both cases. The simulation study shows that the score-type test keeps the nominal significance level very well: 0.042 for $n = 800$ and 0.058 for $n = 1000$.

Remark. One question often raised in nonparametric smoothing is that of bandwidth selection. Conditions (C5) and (C11) provide us with some guidelines to choose the bandwidth, but this only guarantees the validity of the theoretical results when the sample size is large enough. It will be very interesting to find some automatic methods for choosing the bandwidths that assure good small-sample performance of the tests. Another question raised in all kernel related research, including the score-type test procedure developed in our current paper, is the so called "curse of dimensionality", which simply states the performance of the kernel regression degrades when the predictor lies in a higher dimensional space. This is reconfirmed by some simulation studies we conducted when X has two or more dimensions.

Appendix

Proof of Lemma 3.1. The Taylor expansion and the consistency of $\hat{\theta}_n$ imply

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = I_n^{-1}(\bar{\theta}_n)A_n(\theta_0), \tag{A.1}$$

where $A_n(\theta_0) = n^{-1/2}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[Y_i - \hat{g}(X_i,\theta_0)]\dot{\hat{g}}(Z_i,\theta_0)$,

$$I_n(\bar{\theta}_n) = \frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\left[\dot{\hat{g}}(Z_i,\bar{\theta}_n)\dot{\hat{g}}'(Z_i,\bar{\theta}_n) - (Y_i - \hat{g}(X_i,\bar{\theta}))\ddot{\hat{g}}(Z_i,\bar{\theta}_n)\right],$$

and $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0. By Lemma 2.2 and Theorem 2.1 in Sepanski and Lee (1995), $I_n(\bar{\theta}_n) \to I(\theta_0)$ in probability. Although they showed the asymptotic normality of $\hat{\theta}_n$, we need an stochastic expansion of $A_n(\theta_0)$ here. Similar to Sepanski and Lee (1995)'s arguments, $A_n(\theta_0)$ can be written as the sum $V_{1n} + V_{2n} - V_{3n} - V_{4n} + o_p(1)$, where

$$V_{1n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[Y_i - g(Z_i,\theta_0)]\dot{g}(Z_j,\theta_0);$$

$$V_{2n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{I_{\mathcal{C}}(Z_i)[Y_i - g(Z_i,\theta_0)]}{f_Z(Z_i)}[\dot{Q}_N(Z_i,\theta_0) - f_Z(Z_i)\dot{g}(Z_i,\theta_0)];$$

$$V_{3n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{I_{\mathcal{C}}(Z_i)[Y_i - g(Z_i,\theta_0)]\dot{g}(Z_i,\theta_0)}{f_Z(Z_i)}[\hat{f}_Z(Z_i) - f_Z(Z_i)];$$

$$V_{4n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{I_{\mathcal{C}}(Z_i)\dot{g}(Z_i,\theta_0)}{f_Z(Z_i)}[Q_N(Z_i,\theta_0) - \hat{f}_Z(Z_i)g(Z_i,\theta_0)].$$

By calculating the covariances, one can show that V_{2n} and V_{3n} are of order $o_p(1)$.

Now let's consider V_{4n}. Let

$$\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h) = \frac{I_{\mathcal{C}}(Z_i)\dot{g}(Z_i,\theta_0)}{h^d f_Z(Z_i)} K\left(\frac{Z_i - \widetilde{Z}_j}{h}\right)[m(\widetilde{X}_j,\theta_0) - g(Z_i,\theta_0)].$$

Then V_{4n} can be written as a two sample U-statistic

$$V_{4n} = \frac{1}{N\sqrt{n}}\sum_{i=1}^n\sum_{j=n+1}^{n+N}\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h).$$

The mean μ_n of $\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h)$ is

$$\mu_n = E\left[I_{\mathcal{C}}(Z_i)\dot{g}(Z_i,\theta_0)f_Z^{-1}(Z_i)E([Q_N(Z_i,\theta_0) - g(Z_i,\theta_0)\hat{f}_Z(Z_i)]|Z_i)\right].$$

One can show that $|\mu_n| = O(h^k)$. By assumption (C12), $\sqrt{N}\mu_n = O(\sqrt{N}h^k)$ which is of order $o(1)$. To use Lemma B.1 in Sepanski and Lee (1995), we have to explore the conditional expectation of ψ_n given the primary and the validation data. In fact, we can show $E[\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h)|X_i, Z_i] \to 0$, and

$$E[\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h)|\widetilde{X}_j, \widetilde{Z}_j] \to I_{\mathcal{C}}(\widetilde{Z}_j)(m(\widetilde{X}_j,\theta_0) - g(\widetilde{Z}_j,\theta_0))\dot{g}(Z_j,\theta_0).$$

Therefore, we have $V_{4n} = \lambda^{1/2}N^{-1/2}\sum_{j=n+1}^{n+N} I_{\mathcal{C}}(\widetilde{Z}_j)(m(\widetilde{X}_j,\theta_0) - g(\widetilde{Z}_j,\theta_0))\dot{g}(\widetilde{Z}_j,\theta_0) + o_p(1)$, which, together with V_{1n}, implies the lemma. □

Proof of Theorem 3.1. Subtracting and adding $g(Z_i,\theta_0)$ from Y_i in (3.1), and rearranging the terms, S_n can be rewritten as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - g(Z_i,\theta_0))W(Z_i)I_{\mathcal{C}}(Z_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^n (\hat{g}(Z_i,\hat{\theta}_n) - g(Z_i,\theta_0))W(Z_i)I_{\mathcal{C}}(Z_i). \tag{A.2}$$

The second term of (A.2) can be written as the sum of

$$S_{n1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\hat{g}(Z_i, \hat{\theta}_n) - \hat{g}(Z_i, \theta_0)]W(Z_i)I_{\mathcal{C}}(Z_i), \tag{A.3}$$

$$S_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\hat{g}(Z_i, \theta_0) - g(Z_i, \theta_0)]W(Z_i)I_{\mathcal{C}}(Z_i). \tag{A.4}$$

Taylor expansion gives $\hat{g}(Z_i, \hat{\theta}_n) - \hat{g}(Z_i, \theta_0) = (\widehat{\theta}_n - \theta_0)'\dot{\hat{g}}(Z_i, \tilde{\theta}_n)$, where $\tilde{\theta}_n$ lies between $\widehat{\theta}_n$ and θ_0. Therefore, the asymptotic normality of $\hat{\theta}_n$ implies

$$S_{n1} = (\widehat{\theta}_n - \theta_0)'\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\dot{\hat{g}}(Z_i, \widetilde{\theta}_n)W(Z_i)I_{\mathcal{C}}(Z_i) = \sqrt{n}(\widehat{\theta}_n - \theta_0)'\rho + o_p(1), \tag{A.5}$$

where $\rho = E\dot{g}(Z, \theta_0)W(Z)I_{\mathcal{C}}(Z)$. By the definition of $\hat{g}(z, \theta_0)$ and the uniform convergence of $\hat{f}_Z(z)$ to $f_Z(z)$ over the compact set $\mathcal{C}$, S_{n2} can be written as

$$S_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{I_{\mathcal{C}}(Z_i)}{f_Z(Z_i)}[Q_N(Z_i, \theta_0) - \hat{f}_Z(Z_i)g(Z_i, \theta_0)]W(Z_i) + o_p(1).$$

Similar to the discussion on V_{4n}, one can get

$$S_{n2} = \frac{\sqrt{\lambda}}{\sqrt{N}}\sum_{j=n+1}^{n+N}I_{\mathcal{C}}(\widetilde{Z}_j)W(\widetilde{Z}_j)[m(\widetilde{X}_j, \theta_0) - g(\widetilde{Z}_j, \theta_0)] + o_p(1). \tag{A.6}$$

Finally, combining (A.1), (A.2), (A.5) and (A.6), we can show that $S_n = n^{-1/2}\sum_{i=1}^{n}\xi_i + N^{-1/2}\sum_{j=n+1}^{n+N}\zeta_j + o_p(1)$ with $\xi_i = I_{\mathcal{C}}(Z_i)(Y_i - g(Z_i, \theta_0))[W(Z_i) - \dot{g}'(Z_i, \theta_0)I^{-1}(\theta_0)\rho]$, and $\zeta_j = \sqrt{\lambda}I_{\mathcal{C}}(\widetilde{Z}_j)[m(\widetilde{X}_j, \theta_0) - g_{\theta_0}(\widetilde{Z}_j)][W(\widetilde{Z}_j) - \dot{g}(\widetilde{Z}_j, \theta_0)I^{-1}(\theta_0)\rho]$. Then a double array central limit theorem implies the result. □

Proof of Theorem 3.2. The consistency of $\hat{\theta}_n$ to θ_a under the alternative (3.7) can be proven in the similar way as in the null case. A Taylor expansion on the normal equation with respect to θ_a yields

$$\sqrt{n}(\hat{\theta}_n - \theta_a) = I_n^{-1}(\bar{\theta}_n)A_n(\theta_a), \tag{A.7}$$

where $I_n(\theta)$ and $A_n(\theta)$ are the same as in the proof of Lemma 3.1 except for the replacement of θ_0 by θ_a, with $\bar{\theta}_n$ now sits between $\hat{\theta}_n$ and θ_a. Using the same arguments as in proving Lemma 2.2 and Theorem 2.1 in Sepanski and Lee (1995), we can show that $I_{1n}(\bar{\theta}_n) \to I_1(\theta_a)$ in probability. As for the term $A_{1n}(\theta_a)$, we have a similar expansion as the one in the proof of Lemma 3.1 with θ_0 replaced by θ_a in those V_n-terms. But under the alternative model (3.7), V_{2n} and V_{3n} are no longer negligible. Indeed, $V_{2n} - V_{3n}$ can be written as

$$V_{2n} - V_{3n} = \frac{1}{N\sqrt{n}}\sum_{i=1}^{n}\sum_{j=n+1}^{n+N}\psi_n(Y_i, Z_i; \widetilde{X}_j, \widetilde{Z}_j, h)$$

where

$$\psi(Z_i, Y_i, \widetilde{X}_j, \widetilde{Z}_j, h) = \frac{I_{\mathcal{C}}(Z_i)(Y_i - g(Z_i, \theta_a))}{h^q f_Z(Z_i)}K\left(\frac{Z_i - \widetilde{Z}_j}{h}\right)[\dot{m}(\widetilde{X}_j, \theta_a) - \dot{g}(Z_i, \theta_a)].$$

The same argument as in the proof of Lemma 3.1 leads to

$$V_{2n} - V_{3n} = \frac{\sqrt{\lambda}}{\sqrt{N}}\sum_{j=n+1}^{n+N}I_{\mathcal{C}}(\widetilde{Z}_j)(h(\widetilde{Z}_j, \theta_a) - g(\widetilde{Z}_j, \theta_a))(\dot{m}(\widetilde{X}_j, \theta_a) - \dot{g}(\widetilde{Z}_j, \theta_a)) + o_p(1),$$

which, together with the V_{1n} and the asymptotic expansion of V_{4n}, implies that $A_n(\theta_a)$ is stochastically equal to

$$A_n(\theta_a) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i - \frac{\sqrt{\lambda}}{\sqrt{N}}\sum_{j=n+1}^{n+N}\zeta_j + o_p(1), \tag{A.8}$$

where $\xi_i = I_{\mathcal{C}}(Z_i)[Y_i - g(Z_i, \theta_a)]\dot{g}(Z_i, \theta_a)$,

$$\zeta_j = I_{\mathcal{C}}(\widetilde{Z}_j)\{[h(\widetilde{Z}_j) - g(\widetilde{Z}_j, \theta_a)][\dot{m}(\widetilde{X}_j, \theta_a) - \dot{g}(\widetilde{Z}_j, \theta_a)] + [m(\widetilde{X}_j, \theta_a) - g(\widetilde{Z}_j, \theta_a)]\dot{g}(\widetilde{Z}_j, \theta_a)\}.$$

Notice that $E\xi_i = 0, E\zeta_j = 0$ for all $i = 1, 2, \ldots, n, j = n+1, n+2, \ldots, n+N$, so (C4), (C6), (C9) and (C13) implies that $\{\xi_i\}_{i=1}^n$ and $\{\zeta_j\}_{j=n+1}^{n+N}$ are two sequences of i.i.d. random vectors with finite second moments. Then, an application of a double array central limit theorem on (A.8), together with the result on $I_n(\theta_a)$, implies the theorem. □

Proof of Theorem 3.4. The consistency of $\hat{\theta}_n$ is implied by the uniform convergence of $n^{-1}\sum_{i=1}^n [Y_i - \hat{g}(Z_i, \theta)]^2 I_{\mathcal{C}}(Z_i)$ to $E\varepsilon^2 I_{\mathcal{C}}(Z) + E[m(X, \theta_0) - g(Z_i, \theta)]^2 I_{\mathcal{C}}(Z)$ with respect to $\theta \in \Theta$ and the unique minimizer of $E[m(X, \theta_0) - g(Z_i, \theta)]^2 I_{\mathcal{C}}(Z)$ is θ_0.

The asymptotic normality of $\hat{\theta}_n$ is proved by considering the similar Taylor expansion (A.1) as in the null case. The only difference from the null case is that $A_n(\theta_0)$ has an extra term under the local alternatives. In fact

$$A_n(\theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[\varepsilon_i + m(X_i, \theta_0) - \hat{g}(Z_i, \theta_0)]\dot{\hat{g}}(Z_i, \theta_0) + \frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\delta(X_i)\dot{\hat{g}}(Z_i, \theta_0). \tag{A.9}$$

The first term can be proceed exactly as in the null case by treating $\varepsilon_i + m(X_i, \theta_0)$ as the response. It then has the asymptotic normal distribution with mean 0 and covariance matrix Σ_0. The second term in (A.9), by law of large numbers, converges to $EI_{\mathcal{C}}(Z)\delta(X)\dot{g}(Z, \theta_0)$ in probability. The first result in Theorem 3.4 follows. Under (3.10),

$$\begin{aligned} S_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[\varepsilon_i + m(X_i, \theta_0) - g(Z_i, \theta_0)]W(Z_i) + \frac{1}{n}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)\delta(X_i)W(Z_i) \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[\hat{g}(Z_i, \hat{\theta}_n) - g(Z_i, \theta_0)]W(Z_i) = S_{nA} + S_{nB} + S_{nC}. \end{aligned}$$

The law of large numbers implies that $S_{nB} \to EI_{\mathcal{C}}(Z)\delta(X)W(Z)$. For S_{nC}, we have $S_{nC} = \sqrt{n}(\hat{\theta}_n - \theta_0)\rho + S_{n2}$, where S_{n2} is given by (A.4). Therefore

$$\begin{aligned} S_n &= \frac{1}{\sqrt{n}}\sum_{i=1}^n I_{\mathcal{C}}(Z_i)[\varepsilon_i + m(X_i, \theta_0) - g(Z_i, \theta_0)][W(Z_i) - \dot{g}'(Z_j, \theta_0)I^{-1}(\theta_0)\rho] \\ &\quad - \frac{\sqrt{\lambda}}{\sqrt{N}}\sum_{j=n+1}^{n+N} I_{\mathcal{C}}(\widetilde{Z}_j)(m(\widetilde{X}_j, \theta_0) - g(\widetilde{Z}_j, \theta_0))[W(\widetilde{Z}_j) - \dot{g}'(\widetilde{Z}_j, \theta_0)I^{-1}(\theta_0)\rho] \\ &\quad + EI_{\mathcal{C}}(Z)\delta(X)[W(Z) - \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho] + o_p(1). \end{aligned}$$

Applying a double array central limit theorem on above S_n shows that S_n is asymptotically normal with mean $EI_{\mathcal{C}}(Z)\delta(X)[W(Z) - \dot{g}'(Z, \theta_0)I^{-1}(\theta_0)\rho]$ and variance $\sigma_\xi^2 + \sigma_z^2$, where σ_ξ^2 and σ_z^2 given by (3.2) and (3.3), with Y in σ_ξ^2 replaced by $m(X, \theta_0) + \varepsilon$. Finally, one can show that $\hat{\sigma}_\xi^2$ and $\hat{\sigma}_z^2$ defined in (3.4) and (3.5) converge to σ_ξ^2 and σ_z^2 respectively under the local alternatives. The second result of theorem follows. □

References

Carroll, R.J., Ruppert, D., 1988. Transformation and Weighting in Regression. Chapman & Hall, New York.
Carroll, R.J., Ruppert, D., Stefanski, L.A., 1995. Measurement Error in Nonlinear Models. Chapman & Hall/CRC, Boca Raton.
Carroll, R.J., Wand, M.P., 1991. Semiparametric estimation in logistic measure error models. J. Roy. Statist. Soc. Ser.B 53, 652–663.
Cheng, C., Van Ness, J.W., 1999. Statistical Regression with Measurement Error. Arnold, London.
Cook, J., Stefanski, L.A., 1994. A simulation extrapolation method for parameter measurement error models. JASA 89, 1314–1328.
Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19, 1257–1272.
Fan, J., Truong, K.T., 1993. Nonparametric regression with errors in variables. Ann. Statist. 21, 1900–1925.
Fuller, W.A., 1987. Measurement Error Models. Wiley, New York.
Jennrich, R.I., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Koul, H.L., Song, W., 2008. Minimum distance regression model checking with Berkson error. Ann. Statist. (Forthcoming).
Pepe, M.S., 1992. Inference using surrogate outcome data and a validation sample. Biometrika 79, 355–365.
Pepe, M.S., Fleming, T.R., 1991. A general nonparametric method for dealing with errors in missing or surrogate covariate data. JASA 86, 108–113.
Sepanski, J.H., Lee, L.F., 1995. Semiparametric estimation of nonlinear errors-in-variables models with validation study. J. Nonparametr. Stat. 4, 365–394.
Sepanski, J.H., Carroll, R.J., 1993. Semiparametric quasilikelihood and variance function estimation in measurement error models. J. Econometircs 58, 223–256.
Song, W.X., 2008. Model checking in errors-in-variables regression. JMVA 99, 2406–2443.
Stefanski, L.A., Cook, J., 1995. Simulation extrapolation: The measurement error jackknife. JASA 90, 1247–1255.
Wang, Q.H., 1999. Estimation of partial linear error-in-variables model with validation data. J. Multivariate Anal. 69, 30–64.
Wang, Q.H., 2003. Dimension reduction in partly linear error-in-response models with validation data. J. Multivariate Anal. 85, 234–252.
Wu, Chien-Fu, 1981. Asymptotic theory of nonlinear least squares estimation. Ann. Statist. 9, 501–513.
Zhu, L.X., Cui, H.J., Ng, K.W., 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., Song, W.X., Cui, H.J., 2003. Testing lack-of-fit for a polynomial errors-in-variables model. Acta Math. Appl. Sin. Engl. Ser. 19, 353–362.