Statistica Sinica **15**(2005), 1049-1068

TESTING THE ADEQUACY FOR A GENERAL LINEAR ERRORS-IN-VARIABLES MODEL

Lixing Zhu and Hengjian Cui

The University of Hong Kong and Beijing Normal University

Abstract: In testing the adequacy of a regression model, the conditional expectation of the residuals given the observed covariate is often employed to construct lack-of-fit tests. However, in the errors-in-variables model, the resiudal is biased and cannot be used directly. In this paper, by correcting for the bias, we suggest lack-of-fit tests of score type for a general linear errors-in-variables model. The polynomial model is a special case. The tests are asymptotically chi-squared under the null hypothesis. The choice of scores involved in the test statistics and the power properties are investigated. A simulation study shows that the tests perform well. Application to two data sets is also made. The approach can readily be extended to handle general parametric models.

Key words and phrases: Bias correction, errors-in-variables model, lack-of-fit test.

1. Introduction

In many applications of regression analysis, the independent variables may be observed only with measurement errors. A general linear errors-in-variables model can be written as

$$\begin{cases} Y = \alpha + h(x)^{\tau}\beta + e, \\ X = x + u, \end{cases} \tag{1.1}$$

where $E(u) = 0$, x and u are independent, $E(e|x,u) = 0$, and $E(e^2|x,u) = \sigma_e^2$. In the model, X and Y are observable, x and u are $m \times 1$ random vectors with $m \geq 1$, α and β are 1- and p-dimensional unknown parameters respectively, $h(\cdot)$ is a known vector function of p dimensions ($p \geq m$). During the last two decades, errors-in-variables models have received much attention in the literature. Readers can refer to Anderson (1984), Fuller (1987), Stefanski and Carroll (1991), Carroll, Ruppert and Stefanski (1995), and Cheng and Van Ness (1999), and the references therein for more details. Clearly, testing the adequacy of these models is important and relevant.

Consider the null hypothesis

$$H_0: \quad E[(Y - \alpha - h(x)^{\tau}\beta)|x] = 0 \;\; a.s. \quad \text{for some} \;\; \alpha \text{ and } \;\beta \tag{1.2}$$

versus the alternative, with positive probability,

$$H_1: \quad E[(Y - \alpha - h(x)^{\tau}\beta)|x] \neq 0 \quad a.s. \quad \text{for all} \quad \alpha \text{ and } \quad \beta.$$

Interestingly even when x is observable, that is, for ordinary regression models, the above testing problem received attention only in the late 1980's. There are a number of proposals in the literature, for instance Eubank and Spiegelman (1990), Hall and Hart (1990), Eubank and Hart (1993), Härdle and Mammen (1993), Müller (1993), Stute (1997), Härdle, Mammen and Müller (1998), Stute, Gonzalez Manteiga and Quindimil (1998), Stute, Thies and Zhu (1998), Stute and Zhu (2002), Zhu (2003) and Zhu and Ng (2003), among others. Hart (1997) is a good reference book in this area, especially for cases with one-dimensional covariates.

Most of the work in the errors-in-variables context is devoted to estimation rather than testing. To our knowledge, for linear errors-in-variables models with $h(x) = x$, Fuller (1987, pp.25-26) firstly recommended an informal test in terms of the residual plots; Carroll and Spiegelman (1992) considered the graphical and numerical diagnostics for nonlinearity and heteroscedasticity; Carroll, Ruppert and Stefanski (1995) had formal tests for nonlinearity in terms of the conventional way that parameters in linear models are tested to be zero or not. There are few works on lack-of-fit testing in the literature.

Due to the measurement errors, the residuals are highly correlated with the observed independent variables $X = x + u$. The conditional expectation of the residuals given the observed X is not centered: that is, even under H_0,

$$E[(Y - \alpha - h(X)^{\tau}\beta)|X] \neq 0.$$

Fuller (1987, p.23) considered a modification, but the residuals obtained by the modification are not centered either.

Zhu, Cui and Ng (2004) studied the case $h(x) = x$ and derived a necessary and sufficient condition for the linearity, with respect to X, of the above conditional expectation of the residual given X. Based on that, lack-of-fit tests can be constructed. But the normality assumption of the variables is restrictive. Cheng and Kukush (2004) and Zhu, Song and Cui (2004) independently extended Zhu, Cui and Ng's method to handle a polynomial errors-in-variables model and removed the normality restriction.

In the present article, we aim to develop lack-of-fit tests for model (1.1). The tests of score type (see, Cook and Weisberg (1982)) are defined. First, we correct for the bias of the conditional expectation given X of least squared residuals to derive centered residuals. Second, we use the modified residuals to construct score test statistics. The test statistics are asymptotically chi-squared under H_0.

In Section 3, we investigate the power of the tests and construction of further tests. Since the test statistics involve weight functions on scores, we discuss the choice of weight functions in Sections 3 and 4. A simulation study and the application to three data sets are reported in Section 4. Section 5 includes some further discussions, and proofs of our results are put in the Appendix.

2. Construction of Test and Bias Correction

2.1. The estimation of parameters

Assume that x and u have densities $f(x, \theta_1)$ and $g(u, \theta_2)$, respectively, where $f(\cdot, \theta_1)$ and $g(\cdot, \theta_2)$ are two given functions, θ_1 and θ_2 are, respectively, q_1- and q_2- dimensional unknown parameters. Write $\theta = (\theta_1, \theta_2)$ and $q = q_1 + q_2$. Then X has the density $F(\cdot, \theta) = \int f(x, \theta_1) g(\cdot - x, \theta_2) dx$. Let

$$H(X, \theta) =: E_\theta[h(x)|X] = \frac{\int h(x) f(x, \theta_1) g(X - x, \theta_2) dx}{F(X, \theta)}.$$

Hence under H_0,

$$E[(Y - \alpha - H(X, \theta)^\tau \beta)|X] = 0 \quad \text{a.s.} \tag{2.1}$$

The corrected residual $\varepsilon = (Y - \alpha - H(X, \theta)^\tau \beta)$ can be used to construct a test statistic.

Remark 2.1. In the following, we point out some H functions of useful models.

(1) *A linear model.* If $h(x) = x$, $x \sim N(0, \Sigma_x)$ and $u \sim N(0, \Sigma_u)$, then $H(X, \theta) = A(\theta) X$ with $A(\theta) = \Sigma_x (\Sigma_x + \Sigma_u)^{-1}$. This is considered in Carroll, Ruppert and Stefanski (1995). Zhu, Cui and Ng (2004) proved that this relation between h and H is a necessary and sufficient condition for normality of x and u.

(2) *A polynomial model.* If $h(x) = (x, x^2, \ldots, x^k)^\tau$, $x \sim N(0, \sigma_x^2)$ and $u \sim N(0, \sigma_u^2)$, then $H(X, \theta) = (f_1(X), \ldots, f_k(X))^\tau$, where $f_j(X) = \sum_{i=1}^j c_{ij} X^i$, c_{ij} depends on σ_x^2 and σ_u^2 only, $1 \leq j \leq k$. Refer to Cheng and Schneeweiss (1998) and Cheng and Van Ness (1999) for more details.

In the formula of $H(\cdot, \theta)$, θ_1 and θ_2 need to be estimated. When either θ_1 or θ_2 is given, the estimation is easy because we have X_i's with the distribution $F(\cdot, \theta_1, \theta_2)$ and only θ_2 or θ_1 as the unknown parameter; the estimator can be defined by, e.g., maximum likelihood. Under regularity conditions, asymptotic normality can be achieved. It is clear that if there is no constraints on θ_1 and/or θ_2, they are often inestimable because they are unidentifiable unless there are validation data $\tilde{x}_j$. See Sepanski and Lee (1995). Here we assume that θ is identifiable and that a $\sqrt{n}$ consistent estimator $\hat{\theta}$ of θ exists.

The estimation of α and β has received much attention in the literature. See Fuller (1987) for linear models; Cheng and Schneeweiss (1998) and Cheng and Van Ness (1999) for both linear and polynomial models; and Carroll and Ruppert and Stafanski (1995) for more general nonlinear models. Estimation is not the focus of this paper and we adopt the least squares estimator for ease of exposition.

Assume that $\hat{\theta}$ is a $\sqrt{n}$ consistent estimator of θ based on the sample $\{X_1, \ldots, X_n\}$. The least squares estimators of α and β are defined as follows:

$$\hat{\beta} = [S_{HH}(\hat{\theta})]^{-1} S_{HY}(\hat{\theta}), \quad \hat{\alpha} = \overline{Y} - \overline{H}(\hat{\theta})^{\tau} \hat{\beta}, \tag{2.2}$$

where

$$S_{HH}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} [H(X, \hat{\theta}) - \overline{H}(\hat{\theta})][H(X_i, \hat{\theta}) - \overline{H}(\hat{\theta})]^{\tau},$$

$$S_{HY}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} [H(X_i, \hat{\theta}) - \overline{H}(\hat{\theta})](Y_i - \overline{Y}),$$

$$\overline{H}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} H(X_i, \hat{\theta}),\ \overline{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i.$$

Under (A1) and (A2) below, we have that, if $\hat{\theta}$ is $\sqrt{n}$-consistent,

$$\begin{aligned} \hat{\beta} - \beta = & \frac{1}{n} \sum_{i=1}^{n} \Big(\mathrm{Cov}\,(H(X, \theta)) \Big)^{-1} \Big([e_i + (h(x_i) - H(X_i, \theta))^{\tau} \beta][H(X_i, \theta) - Eh(x)] \\ & - [H(X_i, \theta) - Eh(x)](\hat{\theta} - \theta)^{\tau} H'(X_i, \theta) \beta \Big) + o_p(1/\sqrt{n}) \\ = & O_p(1/\sqrt{n}). \end{aligned} \tag{2.3}$$

For a proof, see Lemma 2 in the Appendix.

2.2. The construction of tests

We have to consider methods for test construction. With ordinary regression models, there are several methodologies: the comparison between parametric and nonparametric fits to detect the alternative (see Eubank and Hart (1993) and Mammen and Härdle (1993)); the residual-marked empirical process tests (see Stute (1997) and Stute, Thies and Zhu (1998)); and score-type tests (see Cook and Weisberg (1982), Stute and Zhu (2005)). The first of these could be sensitive to the alternative, but suffers the curse of dimensionality due to the use of nonparametric smoothing; the second does not need nonparametric smoothing, but is less sensitive to the high frequency alternatives; the third is the most easily implemented and the limiting null distribution is tractable, though the choice of scores is crucial for the power performance. In this paper, we adopt score-type tests. Each residual is properly weighted by a function of the covariates. Such an approach has a long tradition in statistics. Typically, score tests are only analyzed (and optimized) after the direction from which the alternative tends to the null model has been specified. Classical examples are linear one- and two-sample rank statistics or rank correlation statistics, see Behnen and Neuhaus (1989). Also, robust tests focusing on a neighborhood of a given family of distributions are designed in this spirit. In this sense, our tests are not new. However, beyond the traditional construction, we investigate the optimal choice of scores for directional alternatives and we construct maximin tests when there are several possible departures from the null hypothesis.

To construct score-type tests we note that under H_0 of (2.1), for any weight function $w(\cdot,\theta,\beta)$,

$$E\Big([Y-\alpha-H(X,\theta)^\tau\beta]w(X,\theta,\beta)\Big)=0, \tag{2.4}$$

provided the left hand side is finite.

Suppose that $\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ is a sample of size n, where $X_i = x_i+u_i$. A test statistic is then defined as follows:

$$T_{n0}=\frac{1}{n}\sum_{j=1}^{n}[Y_j-\hat\alpha-H(X_j,\hat\theta)^\tau\hat\beta]w(X_j,\hat\theta,\hat\beta)=:\frac{1}{n}\sum_{j=1}^{n}\hat\epsilon_j w(X_j,\hat\theta,\hat\beta), \tag{2.5}$$

where $\hat\alpha$ and $\hat\beta$ are the estimators of α and β defined in (2.2), and $\hat\epsilon_j=Y_j-\hat\alpha-\hat\beta^\tau H(X_j,\hat\theta)$ are the observed residuals.

Let $\eta(X,\theta,\beta)=w(X,\theta,\beta)-E(w(X,\theta,\beta))$. Under H_0 and conditions (A1) and (A2) below, when a Taylor expansion is applied, we find

$$T_{n0}=\frac{1}{n}\sum_{j=1}^{n}\{Y_j-H(X_j,\theta)^\tau\beta\}\eta(X_j,\theta,\beta)+o_p(1/\sqrt{n}). \tag{2.6}$$

The details of the proof are in the Appendix. Note that the first term of T_{n0} is a sum of $i.i.d.$ randon variables. It is easily proved that, together with $E(e|x,u)=0$ and $Y-H(X,\theta)^\tau\beta=e+(h(x)-H(X,\theta))^\tau\beta$, the variance of this sum is

$$A^2=E\{[\sigma_e^2+\beta^\tau\,\mathrm{Cov}\,(h(x)|X)\beta]\eta^2(X,\theta,\beta)\},$$

where $\mathrm{Cov}\,(h(x)|X)$ is the conditional variance of $h(x)$ given X. Hence, applying the Central Limit Theorem and (2.6), we have

$$\sqrt{n}T_{n0}\xrightarrow{d}N(0,A^2). \tag{2.7}$$

Assumptions. We assume that the function $H(\cdot,\theta)$ and $w(\cdot)$ satisfy the following conditions.

(A1) $H'(X,\theta) =: \partial H(X,\theta)/\partial\theta$ ($q\times p$ matrix) and $w'(X,\theta,\beta) = \partial w(X,\theta,\beta)/\partial(\theta,\beta)$ are continuous with respect to θ, β and X, and

$$E[\sup_{\theta\in\mathcal{B}}(\|H(X,\theta)\| + \|H'(X,\theta)\beta\| + |w(X,\theta,\beta)| + \|w'(X,\theta,\beta)\|)^2] < +\infty,$$

where $\mathcal{B}$ is an open neighborhood of θ and $\|\cdot\|$ is the Euclidean norm.

(A2) $\mathrm{Cov}\,[H(X,\theta)]$ is a $p\times p$ positive definite matrix, $\mathrm{Cov}\,[H(X,\theta), w(X,\theta,\beta)] = 0$ and $\mathrm{Cov}\,[H'(X,\theta)\beta, w(X,\theta,\beta)] = 0$.

Remark 2.2. Condition (A1) assumes the smoothness of the functions and the finiteness of moments, there are almost necessary for the asymptotic normality of the relevant statistics. The condition $\mathrm{Cov}\,[H(X,\theta), w(X,\theta,\beta)] = 0$ of (A2) assumes the design of a weight function orthogonal to $H(\cdot,\theta)$. The rationale is as follows. Note that $\alpha + H(\cdot,\theta)^\tau\beta$ is the conditional expectation of Y given X and a linear projection of Y onto a space spanned by $H(\cdot,\theta)$. Therefore only the part of a departure from H_0 that is orthogonal to $H(\cdot,\theta)$ can be detected. The weight function $w(\cdot)$ only needs to be selected in the class of functions which are orthogonal to $H(\cdot,\theta)$. The condition that $\mathrm{Cov}\,[H'(X,\theta)\beta, w(X,\theta,\beta)] = 0$ is not necessary and can be removed at the cost of a limiting variance depends on the distribution of X.

Since the T_{n0} are not scale-invariant, we define standardized test statistics with the quadratic form

$$T_n^2 =: \Big(\frac{\sqrt{n}T_{n0}}{A_n}\Big)^2 = \frac{n}{A_n^2}\Big[\frac{1}{n}\sum_{j=1}^n \Big(Y_j - \hat\alpha - H(X_j,\hat\theta)^\tau\hat\beta\Big)w(X_j,\hat\theta,\hat\beta)\Big]^2, \qquad (2.8)$$

where A_n^2 is a normalizing constant and usually a consistent estimator of the asymptotic variance A^2 of $\sqrt{n}T_{n0}$, such as $A_n^2 = (1/n)\sum_{j=1}^n \eta_j^2\hat\epsilon_j^2$.

The T_n^2 are score tests that can be motivated directly as the squares of sums of weighted residuals due to the regression of $w(X_j,\hat\theta,\hat\beta)$ on the $\hat\epsilon_j$. Heuristically, the residuals $\hat\epsilon_j$ should be uncorrelated with the weights if the model is correctly fitted, T_n^2 should be large if H_0 is false.

3. Power Study and Further Construction of Tests

In this section we study the asymptotic null distribution and the power properties of T_n^2, and we construct maximin tests for cases where there are several possible alternative models. Here we give a brief description of the asymptotic behavior of the test statistics, the details of the proof can be found in the Appendix. Since A_n^2 is a consistent estimator of A^2 with $A = A(\sigma_e^2,\theta,\beta)$ and $T_n^2 = ((1/\sqrt{n})[\sum_{j=1}^n \hat\epsilon_j w(X_j,\hat\theta,\hat\beta)]/A_n)^2$, we have the following result by (2.7).

Theorem 1. *Assume* (A1) *and* (A2). *Then under* H_0, *we have*

$$T_n^2 \xrightarrow{d} \chi_1^2, \tag{3.1}$$

where χ_1^2 *is the chi-squared distribution with* 1 *degree of freedom.*

We now discuss the power properties. Consider a sequence of models

$$H_{1n}: \quad Y_{jn} = \alpha + h(x_j)^\tau \beta + C_n s(x_j) + e_j, \qquad X_j = x_j + u_j, \tag{3.2}$$

for $j = 1, \ldots, n$, with some arbitrary function $s(\cdot)$.

Theorem 2. *Assume* (A1) *and* (A2), *and* $E[s^2(x)] < +\infty$. *Then under* H_{1n}, *we have that if* $n^\gamma C_n \to a$, $0 \le \gamma < 1/2$,

$$T_n^2 \xrightarrow{d} \infty, \tag{3.3}$$

and if $n^{1/2} C_n \to 1$

$$T_n^2 \xrightarrow{d} \chi_1^2(C^2), \tag{3.4}$$

where $\chi_1^2(C^2)$ *is a non-central chi-squared random variable with* 1 *degree of freedom and non-centrality parameter* C^2, *with* $C = \mathrm{Cov}\,(s(x), w(X, \theta, \beta))/A$.

This result means that T_n^2 has asymptotic power 1 for global alternatives ($\gamma = 0$) and for local alternatives distinct from the null at rate $n^{-\gamma}$ with $0 < \gamma < 1/2$. It can also detect alternatives converging to the null at rate $n^{-1/2}$, the possibly fastest rate for lack-of-fit testing. Actually, for computation of the power at alternatives distinct from the null at the rate $n^{-1/2}$, we can determine the asymptotic p-values from the chi-squared distribution. The asymptotic power of T_n^2 is $2 - \Phi(\lambda_{\alpha/2} - C) - \Phi(\lambda_{\alpha/2} + C)$ for $\Phi(\cdot)$ being the standard normal distribution function, where $\lambda_{\alpha/2}$ is the $\alpha/2$ quantile of the standard normal distribution. This is a monotone function of $|C|$. Thus we should select the weight function $w(\cdot)$ which makes C^2 as large as possible. The lemma in the Appendix shows that the optimal choice of the weight function w is λw_0 where λ is an arbitrary positive constant and

$$w_0(X, \theta, \beta) = \{P_1(X) - B_{12}^\tau B_{22}^{-1} P_2(X)\}/B_0(X), \tag{3.5}$$

where $B_0(X) = \sigma_e^2 + \beta^\tau \,\mathrm{Cov}\,[h(x)|X]\beta$; $P_1(X), P_2(X), B_{12}$ and B_{22} can be found in Lemma 1 in the Appendix, they depend on $B_0(X)$, $B_1(X) = E(s(x)|X)$ and $B_2(X) = (H^\tau(X, \theta), \beta^\tau(H'(X, \theta))^\tau)^\tau$. Looking at B_{22} in Lemma 1, we see that it is positive semi-definite. Let B_{22}^{-1} be its Moore generalized inverse. We assume that B_{22} is non-singular. Non-singularity is satisfied when the components of

$B_2 - E(B_2)$ are almost surely linearly independent. Furthermore, for the optimal choice of weights, the non-centrality parameter becomes

$$\sup_{\substack{Ew(X,\theta)=0 \\ \mathrm{Cov}[B_2(X), w(X,\theta,\beta)]=0}} C_w = C_{w_0} = E[w_0^2(X,\theta,\beta)B_0(X)].$$

For simplicity, we can take $\lambda = 1$.

Note that $w_0(\cdot,\theta,\beta)$ involves the alternative $s(\cdot)$. For directional alternatives, that is, if $s(\cdot)$ is a known function, $w_0(\cdot,\theta,\beta)$ is estimable, hence it can be used as a weight function. Furthermore, a regression is the projection of the response Y onto the space spanned by $H(\cdot)$ for all X. Therefore, when we consider a testing problem, departures we can detect are within the space perpendicular to the projection space of Y onto X. From this observation, we can assume that departures $s(\cdot)$ are orthogonal to $H(\cdot)$. In this case, the weight function $w_0 = P_1/B_0$, where $P_1(X) = E(s(x)|X) - E(s(x)) - E[(E(s(x)|X) - E(s(x)))/B_0]/E(B_0^{-1})$, because $B_{12} = 0$. Moreover, when $B_0(X)$ is a constant function, the formula reduces further to $w_0 = (E(s(x)|X) - E(s(x)))/B_0$. From this analysis, we can see that in some cases, an optimal weight can be selected as the centered regression function $E(s(x)|X) - E(s(x))$ of $s(\cdot)$ onto the space spanned by $H(\cdot)$ for all X. However, when $s(\cdot)$ is unknown, we cannot estimate $w_0(\cdot,\theta,\beta)$ consistently. Therefore, it is not practical. We should choose $w(X,\theta,\beta)$, at the very least, to have non-zero correlation with the function $s(x)$, so that the tests have non-trivial power. Residual plots should be informative for searching a weight function, we discuss this issue in the next section.

Note that the above tests are powerful for directional alternatives. We now study an important extension of (3.2). Let $s_1,\ldots,s_d$ be any finite number of known functions, where $d \geq 1$. The s_i are possible departure functions and comprise a possible dependence of Y on X other than projection. For example, some of the s-functions may be quadratic forms, and others may look to possible interactions between coordinates of X.

Consider the sequence of models: for $1 \leq i \leq n$,

$$Y_{in} = \alpha + h(x_i)^\tau \beta + C_n \sum_{j=1}^{d} \zeta_j s_j(x_i) + e_i, \qquad X_i = x_i + u_i, \tag{3.6}$$

where $\zeta_1,\ldots,\zeta_d \in R$ are unknown parameters. The null model is $H_0 : \zeta_1 = \cdots = \zeta_d = 0$. By calibration, (3.6) can be rewritten as

$$Y_{in} = \alpha + H(X_i,\theta)^\tau \beta + C_n \sum_{j=1}^{d} \zeta_j E(s_j(x_i)|X_i) + \varepsilon_i, \qquad X_i = x_i + u_i,$$

where $E(\varepsilon|X) = 0$. Assume that the $E(s_j(x)|X)$'s are uncorrelated with $H(X_i, \theta)$ and that the conditional variance of $h(\cdot)$ given X is a constant. As described above, optimal weights can be selected as $W_j(X, \theta) = E(s_j(x)|X) - E(s(x))$, $j = 1, \ldots, d$, each corresponding to $E(s_j(x)|X)$. In the following we derive maximin tests for H_0 versus $||\zeta|| \geq c$, where $||\cdot||$ is a proper norm and $\zeta^\tau = (\zeta_1, \ldots, \zeta_d)$. The following setup provides a maximin-test. For this, similar to those of (2.6) we consider the score-statistics $\hat{T}^j_{n0}$ and A_{nj} pertaining to $W_j(X, \theta) = E(s_j(x)|X) - E(s(x))$, $j = 1, \ldots, d$. We choose the weight functions W_j because the s_j are just the departures from the hypothetical regression function. Such weight functions are natural choices and can enhance power performance, although other weight functions are also possible. Let A_j be the population version of A_{nj}. Put $\hat{T}_n = (\hat{T}^1_{n0}/A_{n1}, \ldots, \hat{T}^d_{n0}/A_{nd})^\tau$. Following the proof of Theorem 2, we can show that under (3.6), when $n^\gamma C_n \to a$ with $0 \leq \gamma < 1/2$, then $T_n \to \infty$ as $n \to \infty$; when $n^{1/2} C_n \to 1$,

$$\sqrt{n}\hat{T}_n \to \Sigma^{(1)} \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_d \end{pmatrix} + \mathcal{N}_d(\mathbf{0}, \Sigma^{(2)}). \tag{3.7}$$

Here, $\Sigma^{(k)} = (\sigma^{(k)}_{ij})_{1 \leq i,j \leq d}$, $k = 1, 2$, with $\sigma^{(1)}_{ij} = E\{[W_i(X, \theta)][W_j(X, \theta)]\}/A_j$ and $\sigma^{(2)}_{ij} = E\{(Y - \alpha - H(X, \theta)^\tau \beta)^2 [W_i(X, \theta)][W_j(X, \theta)]\}/(A_i A_j)$, and $\mathcal{N}_d$ denotes a normal distribution on R^d. Distributional characteristics of (3.6) only appear through the (estimable) covariance matrices.

We may now use existing maximin-theory to obtain optimal tests for H_0. See, e.g., Strasser (1985, Theorem 30.2). For this, define $\Sigma^{(k)}_n = (\sigma^{(k)}_{ijn})_{1 \leq i,j \leq d}$, $k = 1, 2$, through

$$\sigma^{(2)}_{ijn} = \frac{1}{n} \sum_{l=1}^n \frac{\{(Y_l - \hat{\alpha} - \hat{\beta}^\tau H(X_l, \hat{\theta}))^2 [W_i(X_l, \hat{\theta}) - (\bar{W}_i(X, \hat{\theta}))][W_j(X_l, \hat{\theta}) - (\bar{W}_j(X, \hat{\theta}))]\}}{(A_{ni} A_{nj})}.$$

Theorem 3. *For a given significance level* $0 < \alpha < 1$*, and* $\gamma = 1/2$*, the test* $t = 1_{\{\hat{T}^\tau_n (\Sigma^{(2)}_n)^{-1} \hat{T}_n \geq c_\alpha\}}$ *is a maximin* α*-test for* H_0 *versus* H_1*. Here* H_1 *is* (3.6) *with* ζ *such that* $\zeta^T \Sigma^{(1)} (\Sigma^{(2)})^{-1} \Sigma^{(1)} \zeta \geq a$ *for any user-specified value* $a > 0$*,* c_α *is the* $(1 - \alpha)$*-quantile of the chi-square random variable* χ^2_d *with* d *degrees of freedom. The asymptotic maximin power is given by* $P(\chi^2_d(a) \geq c_\alpha)$*, where* a *is the noncentrality parameter.*

Since the dimension d is arbitrary, Theorem 3 covers most examples of interest. For those who prefer omnibus tests we would consider a process-based test to handle a nonparametric class of alternatives. See Stute and Zhu (2002, 2005) for a relevant discussion where there are no measurement errors in covariates. Stute, Thies and Zhu's (1998) method may be useful when the covariate is univariate. In the multivariate covariable case the process method is still valid, but the limiting distribution is no longer distribution-free and a resampling approximation is needed, see Stute, Gonzalez Manteiga and Quindimil (1998). This problem deserves further study.

4. Simulations and Examples

4.1. Simulations

We consider the model

$$y = 1 + x + 2x^2 + cx^3 + e, \quad X = x + u, \tag{4.1}$$

where $x \sim N(0, \sigma_x^2), u \sim N(0, \sigma_u^2)$, $e \sim N(0, 0.5)$, $h(x) = (x, x^2)^\tau$, $\theta = \sigma_x^2$. That is, $c = 0$ corresponds to the null hypothesis. The function H and its derivative can be specified as

$$H(X, \theta) = E(h(x)|X) = \Big(\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}\Big)\Big[X, \frac{\sigma_x^2 X^2}{\sigma_x^2 + \sigma_u^2} + \sigma_u^2\Big]^\tau,$$
$$H'(X, \theta) = \Big(\frac{\sigma_u^2}{(\sigma_x^2 + \sigma_u^2)^2}\Big)\Big[X, \frac{2\sigma_x^2 X^2}{\sigma_x^2 + \sigma_u^2} + \sigma_u^2\Big].$$

In the simulation, we took $\sigma_x^2 = 1$, $\sigma_u^2 = 0.35$ and $n = 50$. As can be seen, the residual plots are informative in detecting alternatives. Therefore, we first tried to choose a weight function based on the plots.

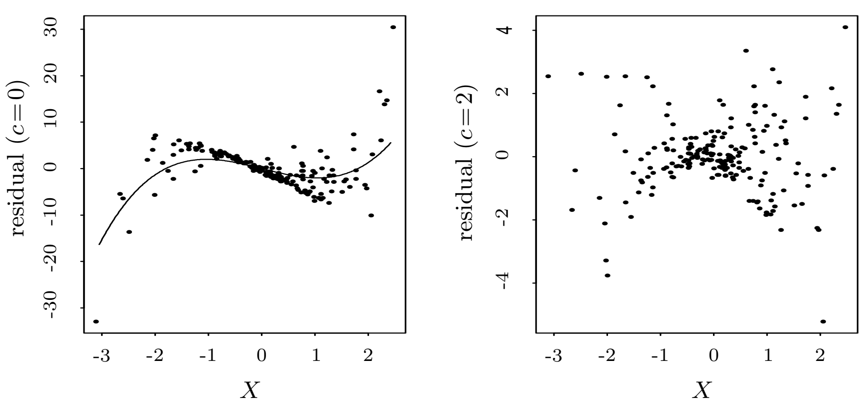


Figure 1. Simulated data with $\sigma_u = 0.35$ in model (4.1): (a). The plot of the data X_j versus the residuals with $c = 0$; (b). The plot of the data X_j versus the residuals with $c = 2$.

From Figure 1 (a), we chose a cubic function of X as the weight function $w(X,\theta,\beta)$. This means that $w(X,\theta,\beta)$ could be taken as the third Hermite polynomial $w_1(X,\theta,\beta) = (X/\sqrt{\sigma_x^2+\sigma_u^2})^3 - 3X/\sqrt{\sigma_x^2+\sigma_u^2}$. Then $E(w_1) = E(Xw_1) = E(X^2 w_1) = 0$ and assumption (A2) is satisfied. We also use the optimal weight function $w_0(X,\theta,\beta) = \{P_1(X) - B_{12}^\tau B_{22}^{-1} P_2(X)\} B_0^{-1}(X)$. By simple calculation, we find

$$B_1(X) = E(x^3|X) = \sigma_1^6 X^3 + 3\sigma_1^2\sigma_2^2 X,\ \text{Cov}\,[h(x)|X] = \sigma_2^2 \begin{pmatrix} 1 & 2\sigma_1^2 X \\ 2\sigma_1^2 X & 2\sigma_2^2 + 4\sigma_1^4 X^2 \end{pmatrix},$$

$$B_0(X) = 0.5 + \sigma_2^2 + 8\sigma_2^4 + 8\sigma_1^2\sigma_2^2 X + 16\sigma_1^4\sigma_2^2 X^2, \quad w_0(X,\theta,\beta) = \sum_{j=0}^{3} d_j X^j / B_0(X),$$

where $\sigma_1^2 = \sigma_x^2/(\sigma_x^2+\sigma_u^2)$, $\sigma_2^2 = \sigma_x^2\sigma_u^2/(\sigma_x^2+\sigma_u^2)$,

$$\begin{aligned} d_0 &= [B_{12}^\tau B_{22}^{-1} E(B_2/B_0) - E(B_1/B_0)]/E(B_0^{-1}) - \sigma_u^2\sigma_1^2 B_{12}^\tau B_{22}^{-1} l_2 \\ d_1 &= \sigma_1^2[3\sigma_2^2 - B_{12}^\tau B_{22}^{-1} l_1], \quad d_2 = \sigma_1^4 B_{12}^\tau B_{22}^{-1} l_2, \quad d_3 = \sigma_1^6, \\ B_{22} &= E[B_2 B_2^\tau/B_0] - E[B_2/B_0]E[B_2^\tau/B_0]/E[B_0^{-1}], \\ B_{12} &= E[B_1 B_2/B_0] - E[B_1/B_0]E[B_2/B_0]/E[B_0^{-1}] \end{aligned}$$

and $l_1 = (1,0)^\tau, l_2 = (0,1)^\tau$. We replace σ_x^2 by its estimator $\hat{\sigma}_x^2 = (1/n)\sum_{i=1}^n (X_i - \overline{X})^2 - \sigma_u^2$ in practice. Looking at the form of the optimal weight function w_0, we realize that it is rather complicated, but it is essentially a ratio of a third to a second order polynomial. Therefore, we also consider a weight function w_2 which is a ratio of second and third order polynomials that satisfies conditions (A.1) and (A.2):

$$w_2(X,\theta,\beta) = \frac{\left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^3 - \frac{dX}{\sqrt{\sigma_x^2+\sigma_u^2}}}{1 + \left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^2},$$

$$d = \frac{E\Big\{\left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^4 \Big/ \Big[1 + \left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^2\Big]\Big\}}{E\Big\{\left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^2 \Big/ \Big[1 + \left(\frac{X}{\sqrt{\sigma_x^2+\sigma_u^2}}\right)^2\Big]\Big\}}.$$

The power functions against c with this choice w_0, w_1, w_2 are presented in Figure 2 (the number of replication = 500). The tests maintain the significance level well, and when the departure to the hypothetical model is not significant, that is, c is small, the tests still have an encouraging power performance. As is expected, the tests with optimal weight function fare best. But, as we know, the form of the test with optimal weight is complicated and the computational burden is heavy. In contrast, we find that the power of the test with w_2 is only slightly lower than that with w_0. The performance of the test with w_1 is also encouraging, given that it was chosen simply by checking residual plots.

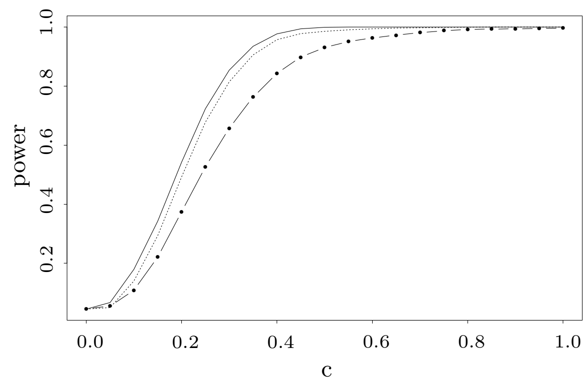


Figure 2. Simulated power functions against the values c. The solid line is with w_0, dot-dash line with w_1 and dot line with w_2.

For the sake of simplicity, using residual plots to search a proper weight can be useful. In the following examples, we adopt this approach.

4.2. Examples

Example 1. The data are the depths and locations of $n = 43$ earthquakes occurring near the Tonga trench between January 1965 and January 1966 (see Fuller (1987, pp. 214 and 289)). The variable X_1 is the perpendicular distance in hundreds of kilometers from a line that is approximately parallel to the Tonga trench. The variable X_2 is the distance in hundreds of kilometers from an arbitrary line perpendicular to the Tonga trench. The variable Y is the depth of the earthquake in hundreds of kilometers. The distribution of (X_1, X_2) is assumed to be normal. Under the plate model, the depths of the earthquakes will increase with distance from the trench and a plot of the data shows this to be the case. The location of the earthquakes is subject to error. Our intention is to check whether the plate model is a linear model or not. We assume that the variance matrix of measurement error u is known as $\Sigma_u = 0.01I_2$, which is the same as Fuller's, and $p = 2$. Figure 3 shows the plots of the response Y against $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$, the residual against $X_1/std(X_1)$, the residual against $X_2/std(X_2)$, and the weight function $w(X)$ against $X_1/std(X_1)$ respectively. Looking at Figure 3(a), a linear relationship seems visible. Thus we state as the null hypothesis that the underlying model is linear. However, the residual plot for the individual covariate x_1 shows as a quadratic. We chose the second Hermite polynomial as a weight function: $w_1(X) = ((X_1 - \mu_1)/std(X_1))^2 - 1$ with $\mu_1 = E(X_1)$. When the normality assumption holds, this weight function satisfies (A1) and (A2). According to the plots (a) and (b) in Figure 3, this is also a reasonable selection. We find $T_n^2 = 13.1059$ and a p-value of 0.00029, so the linear model is not tenable. In fact, Fuller (1987) took a working model quadratic in x_1, and linear in x_2.

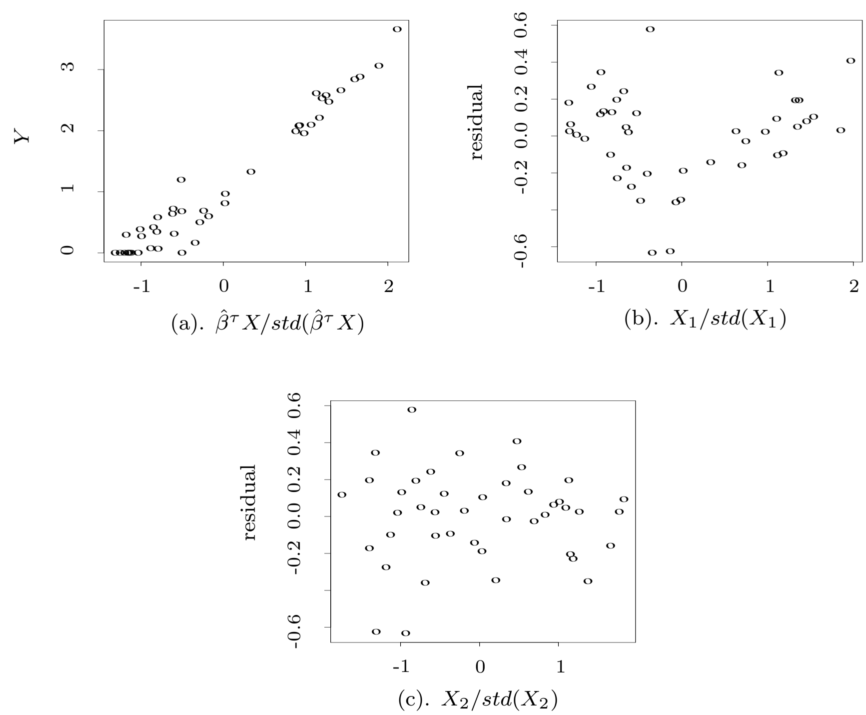
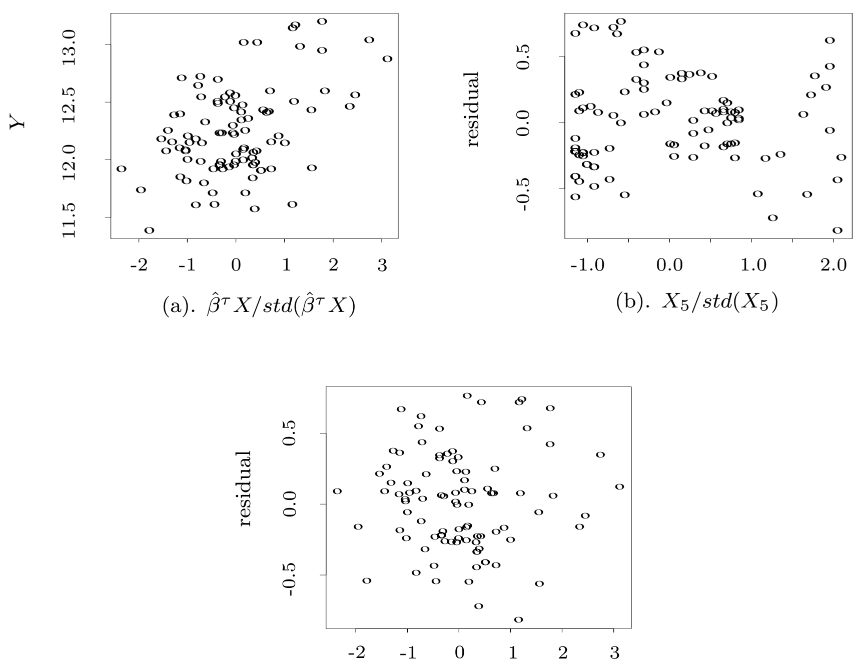


Figure 3. (a). The plot of the data $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$ versus the response Y; (b). The plot of the data $X_1/std(X_1)$ versus the residuals; (c). The plot of the data $X_2/std(X_2)$ versus the residuals.

Example 2. The data are houses liquidated by Saving and Loans in the California area, as functions of the number (x_1) of rooms, the number (x_2) of bedrooms, the number (x_3) of bathrooms, total living area x_4 in hundreds of square feet, and age (x_5) of the house. There are 99 observations collected by the Resolution Trust Corporation (R.T.C.) during the period of 1990-1992. The response variable Y is the logarithm of the appraised value in thousands of dollars (see He and Ng (1999)). Here the observed age ($X_5 = x_5 + u_5$) of house is subject to measurement error because of survey reasons. Write the observed vector $X = x + u$, where x and u are multinormal and independent with $p = 5$, the mean and var-covariance matrix of measurement error u are assumed to be known as 0 and

$\Sigma_u = diag(0,0,0,0,25)$, respectively. We still use the notation X to denote the centralization of X (i.e., $X - mean(X)$) in the following.

Figure 4 shows the plots of the response Y against $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$, the residual against $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$, the residual against $X_5/std(X_5)$, and $w_1(X)$ against $X_5/std(X_5)$, respectively. When we look at Figure 4 (a), a linear model seems tenable. Thus, the null model is also assumed to be linear. However, Figure 4 (c) provides evidence of nonlinearity. We tried a second order Hermite polynomial first: $w_1(X) = (X_5/std(X_5))^2 - 1$. As before, when normality is true, this weight function satisfies (A1) and (A2). We find $T_n^2 = 5.852$ and a p-value of 0.015. Linearity was rejected for other choices as well. We do not report these results here.



(a). $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$

(b). $X_5/std(X_5)$

(c). $\hat{\beta}_n^\tau X/std(\hat{\beta}_n^\tau X)$

Figure 4. (a). The plot of the data $\hat{\beta}^\tau X/std(\hat{\beta}^\tau X)$ versus the response Y; (b). The plot of the data $X_5/std(X_5)$ versus the residuals; (c). The plot of the data $\hat{\beta}_n^\tau X/std(\hat{\beta}_n^\tau X)$ versus the residuals.

From the above study, quadratic departure of x_5 from a linear model may be considered, although it is not obvious. Therefore, we test a more general linear model, namely

$$E(Y|x) = \alpha + h(x)^\tau \beta, \qquad X = x + u, \tag{4.1}$$

where $h(x) = (x^\tau, x_5^2)^\tau$. That is, x_5 is an order-two polynomial. We chose a cubic Hermite polynomial as the weight function $w_2(X) = (X_5/std(X_5))^3 - 3(X_5/std(X_5))$. This weight function satisfies (A1) and (A2) under normality of X. Here $T_n^2 = 1.84$ and the p-value is equal to 0.175, insufficient to reject (4.1).

5. Conclusion

In this paper, we study lack-of-fit tests of score type for a general linear errors-in-variables model. Due to measurement errors, we have to consider bias correction of the residuals when constructing the tests. As is well known, score tests benefit from appropriate weights for power performance. We discuss in detail how to define an optimal weight for directional alternatives, and suggest empirical ways such as the use of residual plots to select weights. For general alternatives, maximin-tests are defined to handle the cases where we have several possible candidates of alternatives. Our approach can readily be extended to handle model checking for general parametric models as long as the parameters can be well estimated with asymptotic normality. We will study this in the further research.

Appendix

Lemma 1. *Assume that $B_0(X) > 0$ and $E[B_0(X) + B_1^2(X) + \|B_2(X)\|^2] < +\infty$. Then*

$$\max_{\substack{Ew(X,\theta,\beta)=0 \\ \mathrm{Cov}[B_2(X),w(X,\theta,\beta)]=0}} \frac{\{E[B_1(X)w(X,\theta,\beta)]\}^2}{E[B_0(X)w^2(X,\theta,\beta)]} = E[w_0^2(X,\theta,\beta)B_0(X)],$$

where $w_0(X,\theta,\beta) = \{P_1(X) - B_{12}^\tau B_{22}^{-1} P_2(X)\}/B_0(X)$ with $P_l =: B_l - EB_l - E[(B_l - E(B_l))/B_0]/E[B_0^{-1}]$, $l = 1, 2$, $B_{22} = E\{P_2(B_2 - EB_2)^\tau/B_0\}$, $B_{12} = E\{P_1(B_2 - EB_2)/B_0\}$, where B_0, B_1 are scalars and B_2 is a column vector, provided B_{22} is invertible.

Proof. Let $B^*(X) = P_1(X) - B_{12}^\tau B_{22}^{-1} P_2(X)$. It is easy to see that $E[B_1(X)w(X,\theta,\beta)] = E[B^*(X)w(X,\theta,\beta)]$, because $Ew(X,\theta,\beta) = 0$ and $\mathrm{Cov}\,[B_2(X), w(X,\theta,\beta)] = 0$. By the Cauchy-Schwarz inequality, we have

$$\frac{\{E[B_1(X)w(X,\theta,\beta)]\}^2}{E[B_0(X)w^2(X,\theta,\beta)]} = \frac{\{E[B^*(X)w(X,\theta,\beta)]\}^2}{E[B_0(X)w^2(X,\theta,\beta)]} \le E\Big(\frac{B^{*2}(X)}{B_0(X)}\Big).$$

Note that $E(P_l/B_0) = 0$, $l = 1, 2$, and $\mathrm{Cov}\,(B_0^{-1}P_1, B_2) = B_{12}^\tau B_{22}^{-1}\,\mathrm{Cov}\,(B_0^{-1}P_2, B_2)$. Hence $E[B^*/B_0] = 0$ and $\mathrm{Cov}\,[B^*/B_0, B_2] = 0$. Then when we choose $w_0 = B^*/B_0$, equality holds in the above inequality. In other words, W_0 is a maximizer. The proof is finished.

Lemma 2. *Under* (A1), (A2) *and* $\hat\theta-\theta=O_p(1/\sqrt{n})$, $\hat\beta-\beta=O_p(1/\sqrt{n})$.

Proof. Invoking $E[\sup_{\theta\in\mathcal{B}}\|H(X,\theta)\|^2]<+\infty$, the continuity of $H(X,\theta)$ at θ, $\|\hat\theta-\theta\|=O_p(1/\sqrt{n})$, and the Weak Law of Large Numbers, we can show that $S_{HH}(\hat\theta)-\text{Cov}\,[H(X,\theta)]=o_p(1)$, and

$$S_{HY}(\hat\theta)-S_{HH}(\hat\theta)\beta=\frac{1}{n}\sum_{i=1}^n[H(X_i,\hat\theta)-\overline{H}(\hat\theta)][Y_i-H(X_i,\hat\theta)^\tau\beta]$$
$$=\frac{1}{n}\sum_{i=1}^n[H(X_i,\hat\theta)-\overline{H}(\hat\theta)]\{[h(x_i)-H(X_i,\hat\theta)]^\tau\beta+e_i\}.$$

Using a Taylor expansion for $H(X_i,\cdot)$ around θ, we get

$$S_{HY}(\hat\theta)-S_{HH}(\hat\theta)\beta=\frac{1}{n}\sum_{i=1}^n[H(X_i,\theta)-\overline{H}(\theta)[e_i+(h(x_i)-H(X_i,\theta))^\tau\beta]]$$
$$-\frac{1}{n}\sum_{i=1}^n[H(X_i,\theta)-\overline{H}(\theta)](\hat\theta-\theta)^\tau(H'(X_i,\theta))\beta$$
$$+\frac{1}{n}\sum_{i=1}^n[H'(X_i,\theta)-\overline{H}'(\theta)]^\tau(\hat\theta-\theta)\Big(e_i+[h(x_i)-H(X_i,\theta)]^\tau\beta\Big)+o_p(\frac{1}{\sqrt{n}}).$$

Furthermore, noting that $E\Big((h(x)-H(X,\theta))|X\Big)=0$, we have

$$E\{[H'(X_i,\theta)-\overline{H}'(\theta)]\Big(e+[h(x)-H(X,\theta)]^\tau\beta]\Big)=0.$$

The last sum is $o_p(1/\sqrt{n})$ and, replacing $\overline{H}(\theta)$ by $Eh(x)$, we have

$$\begin{aligned}\hat\beta-\beta&=S_{HH}^{-1}(\hat\theta)[S_{HY}(\hat\theta)-S_{HH}(\hat\theta)\beta]\\&=\frac{1}{n}\sum_{i=1}^n\text{Cov}\,[H(X,\theta)]^{-1}\Big\{[H(X_i,\theta)-Eh(x)][e_i+(h(x_i)-H(X_i,\theta))^\tau\beta]\\&\qquad-[H(X_i,\theta)-Eh(x)](\hat\theta-\theta)^\tau H'(X_i,\theta)\beta\Big\}+o_p(\frac{1}{\sqrt{n}})\\&=O_p(\frac{1}{\sqrt{n}}).\end{aligned}\tag{A.1}$$

The proof is finished.

Proof of Theorem 1. Based on the description in Section 2 and (2.3), we only need to prove (2.6). The conclusion of Theorem 1 is then implied by the Central Limit Theorem.

Deal with (2.6). Note that

$$\begin{aligned}\hat{\epsilon}_j w(X_j,\hat{\theta},\hat{\beta}) &= \Big(Y_j-\hat{\alpha}-H(X_j,\hat{\theta})^\tau\hat{\beta}\Big)w(X_j,\hat{\theta},\hat{\beta})\\ &= \Big(Y_j-\overline{Y}-[H(X_j,\hat{\theta})-\overline{H}(\hat{\theta})]^\tau\hat{\beta}\Big)w(X_j,\hat{\theta},\hat{\beta})\\ &= \Big(e_j-\overline{e}+[h(x_j)-\overline{h}-(H(X_j,\theta)-\overline{H}(\theta))]^\tau\beta\Big)w(X_j,\hat{\theta},\hat{\beta})\\ &\quad -[H(X_j,\hat{\theta})-\overline{H}(\hat{\theta})]^\tau(\hat{\beta}-\beta)w(X_j,\hat{\theta},\hat{\beta})\\ &\quad -[H(X_j,\hat{\theta})-\overline{H}(\hat{\theta})-(H(X_j,\theta)-\overline{H}(\theta))]^\tau\beta w(X_j,\hat{\theta},\hat{\beta})\\ &=: I_{jn}^{(1)}-I_{jn}^{(2)}-I_{jn}^{(3)}. \end{aligned}\tag{A.2}$$

By $E[\sup_{\theta\in\mathcal{B}}\|w(X,\theta,\beta)\|^2]<+\infty$, $E[(h(x_j)-H(X_j,\theta))|X_j]=0$ and the continuity of $w'(X,\theta)$ at θ, invoking a Taylor expansion for both H and w at θ and β, we have

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n I_{jn}^{(1)} &= \frac{1}{n}\sum_{j=1}^n\Big(e_j-\overline{e}+[h(x_j)-\overline{h}-(H(X_j,\theta)-\overline{H}(\theta)]^\tau\beta\Big)w(X_j,\hat{\theta},\hat{\beta})\\ &= \frac{1}{n}\sum_{j=1}^n[e_j+(h(x_j)-H(X_j,\theta))^\tau\beta][w(X_j,\theta,\beta)-Ew(X,\theta,\beta)]\\ &\quad +o_p(\frac{1}{\sqrt{n}}).\end{aligned}\tag{A.3}$$

For $I_{jn}^{(2)}$, by using the root-n consistency of $\hat{\beta}-\beta$ in Lemma 2 and (A.2), it follows that

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^n I_{jn}^{(2)} &= \frac{1}{n}\sum_{j=1}^n[H(X_j,\hat{\theta})-\overline{H}(\hat{\theta})]^\tau(\hat{\beta}-\beta)w(X_j,\hat{\theta},\hat{\beta})\\ &= \Big(\operatorname{Cov}[H(X,\theta),w(X,\theta,\beta)]+o_p(1)\Big)^\tau O_p(\frac{1}{\sqrt{n}})=o_p(\frac{1}{\sqrt{n}}).\end{aligned}\tag{A.4}$$

Applying the Mean Value Theorem, $E[\sup_{\theta\in\mathcal{B}}\|\beta^\tau H'(X,\theta)w(X,\theta,\beta)\|]<+\infty$ and $\mathrm{ECov}[H'(X,\theta)\beta,w(X,\theta,\beta)]=0$, it follows from $\|\hat{\theta}-\theta\|=O_p(1/\sqrt{n})$ that

$$\frac{1}{n}\sum_{j=1}^n I_{jn}^{(3)}=o_p(\frac{1}{\sqrt{n}}).\tag{A.5}$$

From (6.2)–(6.5), we obtain (2.6). Since A_n^2 is a consistent estimator of A^2, T_n^2 is asymptotically distributed as χ_1^2, which completes the proof.

Proof of Theorem 2. By arguments like those used to prove Lemma 2, we find that, under H_{1n},

$$\begin{aligned}
&\hat{\beta}-\beta = S_{HH}^{-1}(\hat{\theta})[S_{HY}(\hat{\theta}) - S_{HH}(\hat{\theta})\beta] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Cov}\,[H(X,\theta)]^{-1}\Big([H(X_i,\theta) - Eh(x)][e_i + (h(x_i) - H(X_i,\theta))^{\tau}\beta] \\
&\qquad -[H(X_i,\theta) - Eh(x)](\hat{\theta}-\theta)^{\tau}H'(X_i,\theta)\beta\Big) \\
&\quad + \frac{C_n}{n}\sum_{i=1}^{n} \mathrm{Cov}\,[H(X,\theta)]^{-1}\Big([s(x_i) - \overline{s}][H(X_i,\theta) - Eh(x)]\Big) + o_p(\frac{1}{\sqrt{n}}). \qquad (A.6)
\end{aligned}$$

Now

$$\begin{aligned}
\hat{\epsilon}_j w(X_j,\hat{\theta},\hat{\beta}) &= (Y_j - \hat{\alpha} - H(X_j,\hat{\theta})^{\tau}\hat{\beta})w(X_j,\hat{\theta},\hat{\beta}) \\
&=: I_{jn}^{(1)} - I_{jn}^{(2)} - I_{jn}^{(3)} + I_{jn}^{(4)} + o_p(\frac{1}{\sqrt{n}}), \qquad (A.7)
\end{aligned}$$

where $I_{jn}^{(4)} = -C_n(H(X_j,\hat{\theta}) - \overline{H}(X_j,\hat{\theta}))^{\tau}w(X_j,\hat{\theta},\hat{\beta})((1/n)\sum_{i=1}^{n}\mathrm{Cov}\,[H(X,\theta)]^{-1}([s(x_i) - \overline{s}][H(X_i,\theta) - Eh(x)])) + C_n(s(x_j) - \overline{s})w(X_j,\hat{\theta},\hat{\beta})$. By the conditions of Theorem 2, we obtain that

$$\Big(\frac{1}{C_n\sqrt{n}}\Big)\frac{1}{\sqrt{n}}\sum_{j=1}^{n} I_{jn}^{(4)} = \mathrm{Cov}\,[s(x), w(X,\theta)] + o_p(1). \qquad (A.8)$$

In other words, $(1/\sqrt{n})\sum_{j=1}^{n} I_{jn}^{(4)}$ converges in probability to infinity at rate $n^{1/2-\gamma}$ for $0 \le \gamma < 1/2$, and to a constant when $\gamma = 1/2$. The proof is concluded from (6.7), (6.8), and the proof of Theorem 1.

Acknowledgements

Zhu's research was supported by a CRGC grant from the University of Hong Kong, and a grant from Research Grant council of Hong Kong (#HKU7060/04P). Cui's research was partially supported by the EYTP, RFDP (No: 20020027010) and NNSF (No: 10071009) of China. The authors thank the referee for constructive suggestions and generous help in editing.

References

Anderson, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12**, 1-45.

Behnen, K. and Neuhaus, G. (1989). *Rank Tests with Estimated Scores and Their Application.* B. G. Neubner Stuttgart, Germany.

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error on Nonlinear Models.* Monographs on Statistics and Applied Probability. Chapman and Hall, Now York.

Carroll, R. J. and Spiegelman, C. H. (1992). Diagnostics for nonlinearity and heteroscedasticity in errors-in-variables regression. *Technometrics* **34**, 186-196.

Cheng, C. L. and Schneeweiss H. (1998). Polynomial regression with errors in the variables. *J. R. Statist. Soc. B* **60**, 189-199.

Cheng, C. L. and Van Ness, J. W.(1999). *Statistical Regression with Measurement Error.* Arnold, London.

Cheng, C. L. and Kukush, A. (2004). A goodness-of-fit test for a polynomial errors-in-variables model. *Ukrainian Math. J.* **56**, 527-543.

Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression.* Chapman and Hall, New York

Eubank, R. L. and Hart, J. D. (1993). Commonalty of cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* **80**, 89-98.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85**, 387-392.

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley, New York.

Hall, P. and Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.* **85**, 1039-1049.

Härdle, W. and Mammen, E. (1993). Comparing non-parametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.

Härdle, W., Mammen, E. and Müller, M. (1998). Testing parametric versus semiparametric modeling in generalized linear models. *J. Amer. Statist. Assoc.* **93**, 1461-1474.

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer Series in Statistics, New York.

He, X. and Ng, P. (1999). Quantile splines with several covariates. *J. Statist. Plann. Inference* **75**, 343-352.

Stefanski, L. A. and Carroll, R. J. (1991). Deconvolution-based score tests in measurement error models. *Ann. Statist.* **19**, 249-259.

Sepanski, J. H. and Lee, L. F. (1995). Semiparametric estimation of nonlinear error-in-variables models with validation study. *J. Nonparametr. Statist.* **4**, 365–394.

Strasser, H. (1985). *Mathematical Theory of Statistics.* De Gruyter, Berlin.

Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.* **25**, 613-641.

Stute, W., Gonzalez Manteiga, G. and Quindimil, M. P. (1998). Bootstrap approximations in model checks for regression. *J. Amer. Statist. Asso.* **93**, 141-149.

Stute, W., Thies, S. and Zhu, L. X. (1998). Model checks for regression: An innovation process approach. *Ann. Statist.* **26**, 1916-1934.

Stute, W. and Zhu, L. X.(2002). Model checks for generalized linear models. *Scan. J. Statist.* **29**, 535-546.

Stute, W. and Zhu, L. X.(2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048-1085.

Zhu, L. X. (2003). Model checking of dimension-reduction type for regression. *Statist. Sinica* **13**, 283-296.

Zhu, L. X., Cui, H. J. and Ng, K. W. (2004). Testing lack-of-test for linear errors-in-variables model. *Acta Math. Appl. Sinica* **20**, 533-540.

Zhu, L. X. and Ng, K. W. (2003). Checking the adequacy of a partial linear model. *Statist. Sinica* **13**, 763-781.

Zhu, L. X., Song, W. X. and Cui, H. J. (2004). Testing lack-of-fit for a polynomial errors-in-variables model, *Acta Math. Appl. Sinica* **19**, 353-362.

Department of Statistics and Actuarial Science, The University of Hong Kong, P.R. China.

E-mail: lzhu@hkbu.edu.hk

Department of Statistics and Financial Mathematics, School of Mathematical Science, Beijing Normal University, P.R. China.

E-mail: hjcui@bnu.edu.cn

(Received May 2003; accepted December 2004)