Statistica Sinica **13**(2003), 283-296

MODEL CHECKING OF DIMENSION-REDUCTION TYPE FOR REGRESSION

Li-Xing Zhu

The University of Hong Kong

Abstract: Residual analysis is a commonly used technique for model checking in regression. However the problem gets more complicated when the dimension of covariate is high because it is difficult to see what the residuals should be plotted against. In this paper, we propose a simple search for a good projection direction for plotting and for constructing a lack-of-fit test. We also investigate the bootstrap approximation for computing p-values.

Key words and phrases: Bootstrap approximation, dimension-reduction, empirical process, model checking, residual plot.

1. Introduction

Parametric models describe the impact of the covariate X on the response Y in a concise way. They are easy to use. But since there are usually several competing models to entertain, model checking becomes an important issue.

Suppose that $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ are *iid* observations satisfying

$$y_i = \phi(x_i) + \varepsilon_i \qquad i = 1, \ldots, n, \tag{1.1}$$

where y_i is one-dimensional, $x_i = (x_i^{(1)}, \cdots, x_i^{(d)})'$ is a d-dimensional column vector and ε_i is independent of x_i. We want to test

$$H_0 : \phi(x) = \phi_0(x, \beta) \quad \text{for some } \beta, \tag{1.2}$$

where $\phi_0(x, \beta)$ is a specified function.

For this testing problem, there are a number of non-parametric approaches available in the literature. One approach is to construct a test statistic by a suitable estimate of $\phi(\cdot) - \phi_0(\cdot, \beta)$. Local smoothing for estimating ϕ is often employed. The success of local smoothing hinges on the presence of sufficiently many data points to provide adequate local information. For one-dimensional cases, many smoothing techniques are available and obtained tests have good performance, the book by Hart (1997) gave an extensive overview and useful references. As the dimension of the covariate gets higher, however, the total number of observations needed for local smoothing escalates exponentially. Another approach is to resort to the ordinary residuals $\hat{\varepsilon}_i = y_i - \phi_0(x_i, \beta_n)$, where β_n is an estimate of β. This includes the CUSUM test (Buckley, (1991); Stute, Gonzá lez Manteiga and Presedo Quindimil (1998)) and the innovation transformation-based test (Stute, Thies and Zhu (1998)).

For a practical point of view, however, these two testing approaches suffer from the lack of flexibility in detecting subtle dependence pattern between the residuals and the covariates. As a remedy, practitioners often rely on residual plots, i.e., plots of residuals against fitted values or a selected numbers of covariates for model checking. But this poses a problem when the number of covariates is large especially when one wants to include all possible linear combinations of covariates.

In this paper, we suggest a simple approach for seeking a good projection direction for plotting and constructing a test statistic. For any fixed t, consider

$$I_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{\Sigma}^{-1/2}(x_j - \bar{x}) I(\hat{\varepsilon}_j \le t), \tag{1.3}$$

where $I(\hat{\varepsilon}_j \le t)$ is the indicator function and $\hat{\Sigma}$ is the sample covariance matrix of x_i's. For any $a \in S^d = \{a : ||a|| = 1\}$, define a test statistic T_n by

$$T_n(a) = a^\tau \Big[\frac{1}{n} \sum_{i=1}^{n} (I_n(x_i) I_n^\tau(x_i))\Big] a, \tag{1.4}$$

$$T_n := \sup_{a \in S^d} T_n(a). \tag{1.5}$$

In this paper, the estimate β_n of β is given by the least squares method, that is,

$$\beta_n = \arg\min_{\beta} \sum_{j=1}^{n} (y_j - \phi_0(x_j, \beta))^2.$$

The maximizer $\mathbf{a}$ of $T_n(a)$ over $a \in S^d$ will be used as the projection direction to plot the residuals. Note that T_n and $\mathbf{a}$ are simply the largest eigenvalue and the associated eigenvector of the matrix $[\frac{1}{n}\sum_{i=1}^{n}(I_n(x_i)I_n^\tau(x_i))]$, therefore implementation is easy.

The motivation is quite simple. If the model is correct, $e = y - \phi_0(x, \beta)$ is independent of x. Under the null hypothesis H_0, $E(\Sigma^{-1/2}(X - EX)|e) = 0$, where Σ is the covariance matrix of X. This is equivalent to $I(t) = E[\Sigma^{-1/2}(X - E(X))I(e \le t)] = 0$ for all $t \in R^1$. Consequently for any $a \in S^d$

$$T(a) := a^\tau \Big[\int (I(t))(I(t))^\tau dF_e(t)\Big] a = 0,$$

where F_e is the distribution of e. Then the test statistic $T_n = \sup_a T_n(a)$ is the empirical version of $\sup_a T(a)$. The null hypothesis H_0 is rejected for the large values of T_n.

Note that the test T_n does not involve local smoothing and the determination of the projection direction is easily done. The dimensionality problem may be largely avoided. The next section contains the limit behavior of the test statistic. For computing the p-value, the consistency of bootstrap approximations is discussed in Section 3. A simulation study on the power performance and the comparison among tests is reported in Section 4. The residual plot is also presented in this section. Section 5 contains some further remarks. Proofs of the theorems in Sections 2 and 3 are postponed to Section 6.

2. The Limit Behavior of Test Statistic

Note that under some regularity conditions, β_n can be written as

$$\beta_n - \beta = \frac{1}{n}\sum_{j=1}^{n} L(x_j, \beta)\varepsilon_j + o_p(1/\sqrt{n}),$$

where, letting ϕ_0' be the derivative vector of ϕ_0 at β, $L(X,\beta) = (E[(\phi_0')(\phi_0')^\tau])^{-1} \times \phi_0'(X,\beta)$. Especially, when ϕ_0 is the linear function, $\beta_n = S_n^{-1}X_nY_n$ with $X_n = \{x_1 - \bar{x}, \ldots, x_n - \bar{x}\}, Y_n = \{y_1 - \bar{y}, \ldots, y_n - \bar{y}\}$ and $S_n = (X_n X_n^\tau)$.

We now state an asymptotic result for T_n. Let $V_1(X) = (E[\Sigma^{-1/2}(X - E(X))(\phi_0'(X,\beta))^\tau])L(X,\beta)$.

Theorem 2.1. *Assume that the density function f_ε of ε exists, the derivative $\phi_0'(X,\beta)$ of $\phi_0(X,\beta)$ at β is continuous and has $(2+\delta)$-th moment for some $\delta > 0$, and both the covariance matrix of $\phi_0'(X,\beta)$ and Σ are positive definite. Under H_0,*

$$\begin{aligned} I_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \Sigma^{-1/2}(x_j - E(X))\Big(I(\varepsilon_j \le t) - F_\varepsilon(t)\Big) \\ &\quad + f_\varepsilon(t)\Big(E[\Sigma^{-1/2}(X - E(X))(\phi_0'(X,\beta))^\tau]\Big)\frac{1}{\sqrt{n}}\sum_{j=1}^{n} L(x_j,\beta)\varepsilon_j + o_p(1). \end{aligned}$$

Then T_n converges in distribution to a vector Gaussian process

$$I = B - f_\varepsilon \cdot N \tag{2.1}$$

in the Skorohod space $D^d[-\infty,\infty]$, where B is a vector Gaussian process $(B_1, \ldots B_d)^\tau$ with covariance function $cov(B_i(t), B_i(s)) = F_\varepsilon(min(t,s)) - F_\varepsilon(t)F_\varepsilon(s)$, $F_\varepsilon(t)$ and $f_\varepsilon(t)$, respectively, the distribution and density functions of ε, N is a random vector with a normal distribution $N(0,\sigma^2 V)$ with $V = E(V_1V_1^\tau)$. The covariance function of each component $I^{(i)}$ of I is, for $s \le t$,

$$\begin{aligned} K^{(i)}(s,t) &= F_\varepsilon(s) - F_\varepsilon(s)F_\varepsilon(t) + f_\varepsilon(s)f_\varepsilon(t)E(V_1^{(i)})^2 \\ &\quad - f_\varepsilon(s)\int \varepsilon I(\varepsilon \le t)dF_\varepsilon E(V_1^{(i)}(\Sigma^{-1/2}(X - E(X)))^i) \\ &\quad - f_\varepsilon(t)\int \varepsilon I(\varepsilon \le s)dF_\varepsilon E(V_1^{(i)}(\Sigma^{-1/2}(X - E(X)))^i), \end{aligned} \tag{2.2}$$

where $(\Sigma^{-1/2}(X-E(X)))^i)$ *is the i-th component of* $\Sigma^{-1/2}(X-E(X))$. *The process convergence implies that* T_n *converges in distribution to* $T=\sup_a a^\tau(\int(I(t)\, I(t)^\tau)dF_\varepsilon(t))\, a$.

The distribution of T in Theorem 2.1 is intractable so re-sampling techniques will be used for determining p-values.

3. Bootstrap Approximations

Following the idea of bootstrap technique, the basic procedure for our setup is as follows: Let (x_i^*, y_i^*), $i=1,\ldots,n$ be an artificial sample to be defined later, and let β_n^* be the least squares estimator computed from this sample, a conditional counterpart of I_n given $\{(x_1,y_1),\ldots,(x_n,y_n)\}$ is defined by

$$I_n^*(t)=n^{-\frac{1}{2}}\sum_{j=1}^{n}(\hat{\Sigma}^*)^{-1/2}(x_j^*-\bar{x^*})I(\hat{\varepsilon}_j^*\le t), \tag{3.1}$$

where $\hat{\varepsilon}_j^*$'s are the residuals based on (x_i^*, y_i^*)'s, that is, $\hat{\varepsilon}_j^*=y_j^*-\phi_0(x_j^*,\beta_n^*)$ (or $\hat{\varepsilon}_j^*=y_j^*-(\beta_n^*)^\tau x_j^*$ when ϕ_0 is linear) and $(\hat{\Sigma}^*)^{-1/2}$ is the covariance matrix of x_i^*'s. The conditional counterpart of T_n is

$$T_n^*=\sup_a a^\tau\left[\int(I_n^*(t))(I_n^*(t))^\tau dF_n^*(t)\right]a, \tag{3.2}$$

where F_n^* is the empirical distribution based on ε_i^*, $i=1,\ldots,n$. For computing p-values, we generate m sets of data $\{(x_j^*,y_j^*),\ j=1,\ldots,n\}^{(i)}$, $i=1,\ldots m$, then compute m values of T_n^*. The p-value is estimated by $\hat{p}=k/m$ where k is the number of T_n^*'s larger than or equal to T_n. For the nominal level α, when $\hat{p}\le\alpha$ the null hypothesis is rejected.

How to generate artificial data in regression settings is a crucial question. As some authors have pointed out, the classical bootstrap is sometimes not consistent in the regression setup while the wild bootstrap is applicable. The wild bootstrap was first studied by Wu (1986) in the context of variance estimation in heteroscedastic linear models, and developed by Mammen (1992). See also Härdle and Mammen (1993) and Stute, Manteiga and Quindimil (1998). It is interesting that in our case the situation is reversed, that is, the classical bootstrap is consistent while the wild bootstrap is inconsistent. We propose a variant of the wild bootstrap which is consistent. We first present the idea of the wild bootstrap so that its variant can be described.

The wild bootstrap does as follows. Define $x_i^*=x_i$ and $y_i^*=\phi_0(x_i,\beta_n)+\varepsilon_i^*$, where ε_i^* are defined as $\varepsilon_i^*=w_i^*\hat{\varepsilon}_i$ and w_i^* are *iid* artificial bounded variables with

$$E(w_i^*)=0,\quad \mathrm{Var}\,(w_i^*)=1\quad\text{and}\quad E|w^*|^3<\infty. \tag{3.3}$$

The bootstrap residuals $\hat{\varepsilon}_i^* = y_i^* - \phi_0(x_i, \beta_n^*)$ are used to construct the bootstrap process I_{n1}^* and then the test statistic T_{n1}^*, as in (3.1) and (3.2).

Option 1. This is a variant of the wild bootstrap. The new algorithm resamples the data as follows. Let

$$x_i^* = w_i^*(x_i - \bar{x}) \quad \text{and} \quad \hat{\varepsilon}_i^* = \hat{\varepsilon}_i - \Big(\frac{1}{n}\sum_{j=1}^n w_j^* L(x_j, \beta_n)\hat{\varepsilon}_j\Big) w_i^* \phi_0'(x_i, \beta_n), \tag{3.4}$$

where the weight variables w_i^* are the same as those in the wild bootstrap and $L(\cdot,\cdot)$ is defined in the estimate β_n. When the model is linear, (3.4) reduces to $x_i^* = w_j^*(x_j - \bar{x})$ and

$$\hat{\varepsilon}_i^* = \hat{\varepsilon}_i - S_n^{-1}\Big(\frac{1}{n}\sum_{j=1}^n x_j^* \hat{\varepsilon}_j\Big) x_i^* =: \hat{\varepsilon}_i - (\theta_n^*)^\tau x_i^*. \tag{3.5}$$

The bootstrap process and the resulting statistic can then be created, say I_{n2}^* and T_{n2}^*.

Option 2. Classical bootstrap. Draw the independent bootstrap data from the residuals $\hat{\varepsilon}_i$, say $e_1^*, \ldots, e_n^*$. Define $x_i^* = x_i$ and $y_i^* = \phi_0(x_j, \beta_n) + e_i^*$. The bootstrap residuals $\varepsilon_i^* = y_i^* - \phi_0(x_j, \beta_n^*)$ are used to define the bootstrap process and the test statistic as in (3.1) and (3.2), say I_{n3}^* and T_{n3}^*.

Theorem 3.1. *Under H_0 and the assumptions in Theorem* 2.1 *we have, with probability one, both I_{n2}^* and I_{n3}^* converge weakly to I^* in the Skorohod space $D^d[-\infty, \infty]$, where I^* has the distribution of I in Theorem* 2.1.

Theorem 3.2. *In addition to the assumptions in Theorem* 2.1*, assume that w^* is equally likely ± 1 and the density of ε is symmetric about the origin. Under H_0, the distribution of I_{n1}^* does not converge to that of $B - f_\varepsilon \cdot N$.*

4. Simulation Study

4.1. Power study

In order to demonstrate the performance of the proposed test procedures, small-sample simulation experiments were performed. We made a comparison among Stute, Manteiga and Quindimil's (1998) test (T_S^*), the modified wide bootstrap test (T_{n2}^*) (Option 1) and the classical bootstrap test (T_{n3}^*) (Option 2). The model was

$$y = a^\tau x + b(c^\tau x)^2 + \varepsilon, \tag{4.1}$$

where x is d-dimensional covariate, $d = 3$, 6. When $d = 3$, $a = [1, 1, 2]^\tau$ and $c = [2, 1, 1]^\tau$ and when $d = 6$, $a = [1, 2, 3, 4, 5, 6]^\tau$ and $c = [6, 5, 4, 3, 2, 1]^\tau$. Furthermore, let $b = 0.00$, 0.3, 0.7, 1.00, 1.50 and 2.00 for providing evidence on the power performance of the test under local alternatives ($b = 0.00$ corresponds to the null hypothesis H_0). Sample size is 25 or 50 and the nominal level was 0.05. In each of 1000 replicates, 1000 bootstrap samples were drawn.

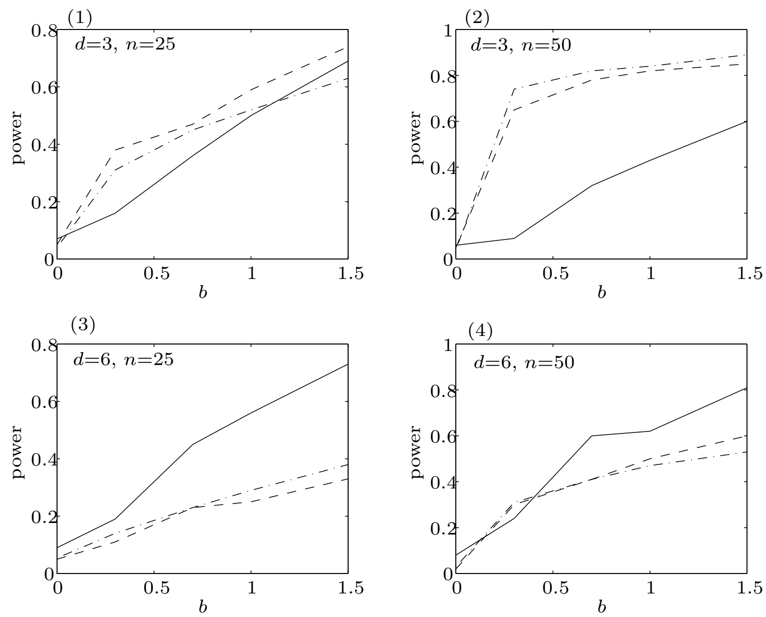

Figure 1. Plots (1) and (2) are with a 3-dimensional covariate and plots (3) and (4) are for 6-dimensional case. The solid, dot dashed and dashed lines are, respectively, for the power of T_S^*, T_{n2}^* and T_{n3}^*.

Figure 1 presents the power of the tests. First, looking at Figures 1(1) and 1(2), we find that with increased sample size, T_{n2}^* and T_{n3}^* improve their performance more quickly than does T_S^*. Figures 1(3) and 1(4) give the same indication. Second, in the 6-dimensional cases, T_S^* has much higher power but cannot maintain the size of the test (the sizes are 0.09 for $d = 6$, $n = 25$, and 0.083 for $d = 6$, $n = 50$). Both T_{n2}^* and T_{n3}^* are a bit conservative. Third, T_{n2}^* and T_{n3}^* have are comparable performance. One might recommend T_{n2}^* on grounds of computational efficiency.

4.2. Residual plots

In addition to the formal test, we also consider the plots of $\hat{\varepsilon}_i$ against the projected covariate $\alpha^\tau x_i$ along the direction α selected by (1.5). We use model (4.1) with $b = 0$ and $b = 1$ to generate $n = 50$ data points. Here $b = 0$ and $b = 1$ correspond, respectively, to linear and nonlinear models. Figure 2 presents the plots of the residuals versus the projected covariates for linear and nonlinear models when the models are fitted linearly. Plots (1) and (3) show that there is no clear relationship while plots (2) and (4) are more suggestive.

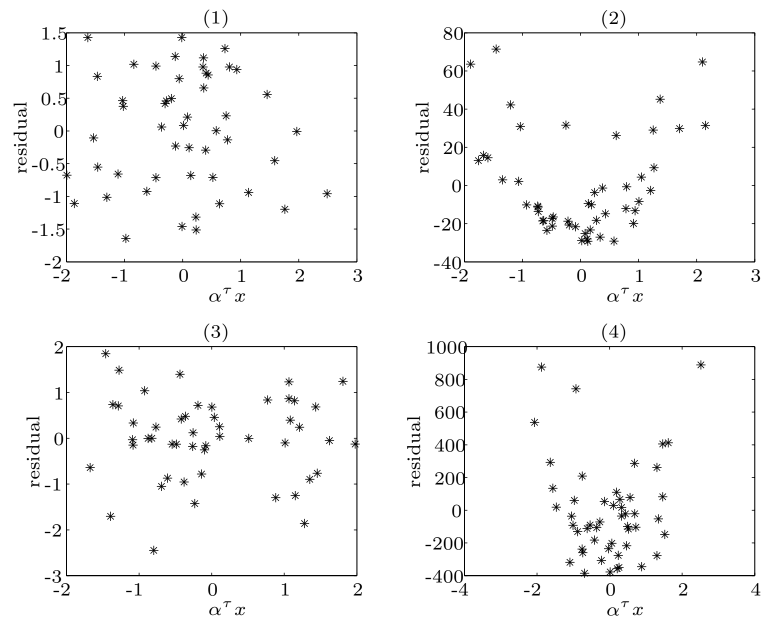

Figure 2. The residual plots against $\alpha^\tau x$ where α is determined by (1.5). Plots (1) and (2) are for the 3-dimensional cases with $b = 0$ and $b = 1$ and plots (3) and (4) are for 6-dimensional cases with $b = 0$ and $b = 1$.

4.3. A real example

We consider the 1984 Olympic records data on various track events as reported by Dawkins (1989). Principal component analysis has been applied to study the athletic excellence of a given nation and the relative strength of the nation at the various running distances. For 55 countries, winning times for men's running events at 100, 200, 400, 800, 1,500, 5,000 and 10,000 meters and the Marathon distance are reported in Dawkins. It is of interest to study whether a nation whose performance is better in running long distances may also have greater strength at short running distances. It may be more reasonable to use speed rather than the winning time for the study, see Naik and Khattree (1996). Let these speeds be $x_1, \ldots, x_8$. We regard 100, 200 and 400 meters as short running distances, 1,500 meters and longer as long running distances. A linear model was fitted by considering the speed of the 100 meters running event (x_1) as the response and the speed of the 1,500, 5000 and 10,000 meters and Marathon running events ($x_5, \ldots, x_8$) as covariates. The p-values of T_S^*, T_{n2}^* and T_{n3}^* are 0.02, 0.08 and 0.01. We may have to reject the null hypothesis that a linear relationship exists between the speed of the 100 meters running event and the long distance running events. Looking at Figure 3, which presents the plots of the residuals versus $\alpha^\tau x$, we find that Figure 3(1) shows some relationship. But after removing the Cook Islands, no clear indication of relationship is visible. Using T_S^*, T_{n2}^* and T_{n3}^* again for the data without the Cook Islands, the p-values are 0.57, 0.64 and 0.06, respectively, so the linear model may be tenable.

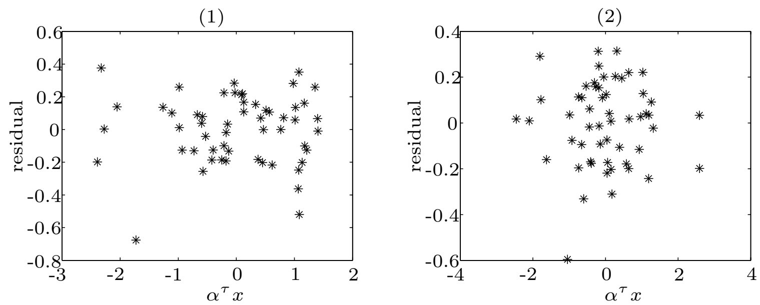

Figure 3. (1) and (2) are the residual plots against $\alpha^\tau x$ where α is determined by (1.5). Plots are for the countries with and without Cook Islands, respectively.

5. Concluding Remarks

In the present paper, we recommend a dimension-reduction approach to model checking for regression models. The formal test and the residual plots can be constructed in terms of the projected covariate. The implementation is easy. A negative aspect of our approach is that the test may rely too highly on the assumption of independence between the covariate X and the error ε. This suggests that our test may be difficult to apply when only $E(\varepsilon|x) = 0$ is assumed. This deserves further study.

6. Proofs

To simplify the proofs, we assume with no loss of generality that X is scalar and $\phi_0(X, \beta)$ is a linear function, $\beta^\tau X$, since asymptotically $\phi_0(X, \beta_n) - \phi_0(X, \beta) = (\beta_n - \beta)^\tau \phi_0'(X, \beta)$ and $\beta_n - \beta$ has an asymptotically linear representation like that in the linear model case. Hence the proof for the general ϕ_0 is almost the same as that for ϕ_0 linear. We also assume that Σ is the identity and $\hat{\Sigma}$ is as well. This replacement does not affect the asymptotic results.

We first present a lemma. Its proof serves as a guide to the proofs of the theorems to follow.

Lemma 6.1. *Assume that for any sequence $\theta_n = O(n^{-c})$, $c > 1/4$, the density* *function f_ε is bounded and $E||x||^{2+\delta} < \infty$ for some $\delta > 0$. Then*

$$\begin{aligned}\sup_{\theta_n,t} R_n(\theta_n,t) &= \sup_{\theta_n,t}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}(x_j - Ex)\{I(\varepsilon_j - \theta_n^\tau(x_j - Ex) \le t) - I(\varepsilon_j \le t) \\ &\quad -F_\varepsilon(t + \theta_n^\tau(x_j - Ex)) + F_\varepsilon(t)\}\Big| \\ &=: \sup_{\theta_n,t}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{n} g_n(x_j,\varepsilon_j,\theta_n,t)\Big| \longrightarrow 0, a.s. \qquad (6.1)\end{aligned}$$

as $n \to \infty$.

Proof. For any $\eta > 0$, application of Pollard's Symmetrization Inequality (Pollard (1984, p.14)) yields, for large n,

$$P\Big\{\sup_{\theta_n,t}|R_n(\theta_n,t)| \ge \eta\Big\} \le 4P\Big\{\sup_{\theta_n,t}|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sigma_j g(x_j,\varepsilon_j,\theta_n,t)| \ge \frac{\eta}{4}\Big\}, \qquad (6.2)$$

provided that for each t and θ_n, $P\{|R_n(\theta_n,t)| \ge \frac{\eta}{2}\} \le \frac{1}{2}$. By the Chebychev Inequality and the conditions imposed, the LHS of (6.2) is less than or equal to $4\theta_n \operatorname{Cov}(x)/\eta$. Hence, (6.2) holds for all large n. To further bound the RHS of (6.2) we recall that the class of all functions $g(\cdot,\cdot,\theta_n,t)$ discriminates finitely many points at a polynomial rate, see Gaenssler (1983). An application of the Hoeffding Inequality (e.g., see Pollard (1984, p.16)) yields for some $w > 0$,

$$\begin{aligned}&P\Big\{\sup_{\theta_n,t}|\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\sigma_j g(x_j,\varepsilon_j,\theta_n,t)| \ge \frac{\eta}{4}|x_1\cdots x_n,\varepsilon_1\cdots\varepsilon_n\Big\} \\ &\le \Big(cn^w \sup_{\theta_n,t}\exp\Big[-\frac{\eta^2}{32\sup_t \frac{1}{n}\sum_{j=1}^{n} g^2(x_j,\varepsilon_j,\theta_n,t)}\Big]\Big) \wedge 1. \qquad (6.3)\end{aligned}$$

To bound the denominator in the power, similar to Lemma II. 33 in Pollard (1984, p.31), we see that for any $c_1 > 0$ there exists a $c_2 > 0$ with

$$\sup_{\theta_n,t}\frac{1}{n}\sum_{j=1}^{n}|I(\varepsilon_j - \theta_n^\tau(x_j - Ex) \le t) - I(\varepsilon_j \le t)|^{c_1} = o_p(n^{-c_2}). \qquad (6.4)$$

By the Hölder Inequality, the sample mean of $g^2(\cdot,\cdot,\theta_n,t)$ is less than or equal to a power of $n^{-1}\sum_{j=1}^{n}(x_j - Ex)^{2+\delta}$ times n^{-c_2} for some $c_2 > 0$. This shows that the RHS of (6.3) goes to zero. Integrate out to get the result. Lemma 6.1 is proved.

Proof of Theorem 2.1. Slightly modifying the argument of Koul's (1992) Theorem 2.3.3, or by applying Lemma 6.1, we can prove the theorem. Details are omitted.

Proof of Theorem 3.1. We deal with I^*_{n2} first. Recall I^*_{n2} has the form, together with (3.5),

$$I^*_{n2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x^*_j \{I(\hat{\varepsilon}_j - (\theta^*_n)^\tau x^*_j \le t) - F^*_n(t)\},$$

$$F^*_n(t) = \frac{1}{n} \sum_{j=1}^{n} I(\hat{\varepsilon}_j - (\theta^*_n)^\tau x^*_j \le t).$$

First of all, we can obtain that for any θ_n satisfying $||\theta_n|| \le c \log n/n^{\frac{1}{2}}$ and for almost all sequences $\{(x_1, y_1), \ldots, (x_n, y_n), \ldots\}$,

$$\begin{aligned} R^*_n(t) \;=\; & \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \{[x^*_j (I(\hat{\varepsilon}_j - \theta^\tau_n x^*_j \le t) - I(\hat{\varepsilon}_j \le t)) \\ & \qquad - E_w [x^*_j [I(\hat{\varepsilon}_j - \theta^\tau_n x^*_j \le t) - I(\hat{\varepsilon}_j \le t))]\} \\ \longrightarrow 0 \quad & \text{a.s.} \end{aligned} \tag{6.5}$$

uniformly on t, where E_w stands for the integration over the variable w^*. The argument is similar to that used to prove Lemma 6.1, noticing that $\theta^*_n = O_p(1/\sqrt{n})$ and letting $\theta_n = \theta^*_n$. Decompose I^*_{n2} as $I^*_{n2}(t) = R^*_n(t) + R^*_{n1}(t) + R^*_{n2}(t) - R^*_{n3}(t)$, where

$$\begin{aligned} R^*_{n1}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} x^*_j \{I(\hat{\varepsilon}_j \le t) - F_n(t)\}, \\ R^*_{n2}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E_w x^*_j [I(\hat{\varepsilon}_j - (\theta^*_n)^\tau x^*_j \le t) - I(\hat{\varepsilon}_j \le t)]\}, \\ R^*_{n3}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} E_w \{x^*_j [F^*_n(t) - F_n(t)]\}. \end{aligned} \tag{6.6}$$

Then we need to show that, combining with (6.5), R^*_{n1} converges in distribution to the Gaussian process B, R^*_{n2} converges in distribution to $f_\varepsilon \cdot N$ and R^*_{n3} tends to zero in probability. Invoking Theorem VII 21 of Pollard (1984, p.157), the convergence of R^*_{n1} can be derived for almost all sequences $\{(x_1, y_1), \ldots, (x_n, y_n), \ldots\}$. The basic steps are as follows. First, we show that the covariance function of R^*_{n1} converges almost surely to that of B. This is easy to do via elementary calculation. Second, we check that the conditions in Pollard's Theorem VII. 21 are satisfied, mainly condition (22) on page 157 (Pollard (1984)). Similar to Lemma 6.1, the class (depending on n) of all functions $x^*(I(\hat{\varepsilon} \le \cdot) - F_n(\cdot))$ discriminates finitely many point at a polynomial rate, see Gaenssler (1983). The condition (22) is satisfied by applying Lemma VII 15 of Pollard (1984, p.150). We omit the full details of the proof. The convergence of R^*_{n3} is much easier to obtain as long as we notice that $\sqrt{n}(F^*_n - F_n)$ and $1/\sqrt{n}\sum_{j=1}^n x^*_j$ have finite limits. The remaining work is to deal with R^*_{n2}. Let

$$\begin{aligned}
&R^*_{n21}(t)\\
&= E_w\Big\{\frac{1}{\sqrt{n}}\sum_{j=1}^n x^*_j[(I(\hat\varepsilon_j - (\theta^*_n)^\tau x^*_j \le t) - F_\varepsilon(t + (\beta_n - \beta)^\tau(x_j - \bar x) + (\theta^*_n)^\tau x^*_j))\\
&\qquad -(I(\hat\varepsilon_j \le t) - F_\varepsilon(t + (\beta_n - \beta)^\tau(x_j - \bar x)))]\Big\}\\
&=: E_w\Big\{\frac{1}{\sqrt{n}}\sum_{j=1}^n x^*_j[\ldots]\Big\}.
\end{aligned}$$

Noticing $\hat\varepsilon = \varepsilon - (\beta_n - \beta)^\tau(x_j - \bar x)$, and following Lemma II. 33 of Pollard (1984, p.31), we have that $\sup_t \frac{1}{n}\sum_{j=1}^n (x^*_j)^2[\ldots]^2 = o(n^{-c_2})$ a.s., for almost all sequences $\{(x_1, y_1), \ldots, (x_n, y_n), \ldots\}$. We then easily derive that, similar to Lemma 6.1, R^*_{n21} converges in probability to zero uniformly on t. Note that $E_w[x^*_j(I(\hat\varepsilon_j \le t) - F_\varepsilon(t + (\beta_n - \beta)(x_j - \bar x)))] = 0$. Hence

$$\begin{aligned}
&R^*_{n2}(t) - R^*_{n21}(t)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n E_w[x^*_j\{F_\varepsilon(t+(\beta_n-\beta)^\tau(x_j-\bar x)+(\theta^*_n)^\tau x^*_j) - F_\varepsilon(t+(\beta_n-\beta)^\tau(x_j-\bar x))\}]\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^n E_w[x^*_j(\theta^*_n)^\tau x^*_j)]f_\varepsilon(t) + o_p(1)\\
&= E_w\Big[\frac{1}{\sqrt{n}}\sum_{j=1}^n (w^*)^2(x_j - \bar x)(x_j - \bar x)^\tau(\theta^*_n)\Big]f_\varepsilon(t) + o_p(1)\\
&=: E_w[\ldots]f_\varepsilon(t) + o_p(1)
\end{aligned}$$

converges in distribution to $f_\varepsilon \cdot N$ as long as we note the fact that the sum $[\ldots]$ is asymptotically equal to $\frac{1}{\sqrt{n}}\sum_{j=1}^n x^*_j\hat\varepsilon_j$, and then is asymptotically normal by the CLT for almost all sequences $\{(x_1, y_1), \ldots, (x_n, y_n), \ldots\}$. The convergence of I^*_{n2} is proved.

We now turn to the proof of the convergence of I^*_{n3}. Note that

$$I^*_{n3}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^n (x_j - Ex)\{I(\hat\varepsilon^*_j - (\beta^*_n - \beta_n)^\tau(x_j - \bar x) \le t) - F^*_n(t)\},$$

$$F^*_n(t) = \frac{1}{n}\sum_{j=1}^n I(\hat\varepsilon^*_j - (\beta^*_n - \beta_n)(x_j - \bar x) \le t).$$

Similar to (6.6), decompose I^*_{n3} as $I^*_{n2}(t) = J^*_n(t) + J^*_{n1}(t) + J^*_{n2}(t) - J^*_{n3}(t)$ where

$$\begin{aligned}
J^*_n(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \{(x_j - Ex)(I(\hat{\varepsilon}^*_j - \theta^\tau_n(x_j - \bar{x}) \le t) - I(\hat{\varepsilon}^*_j \le t) \\
&\qquad - E^*[I(\hat{\varepsilon}^*_j - \theta^\tau_n(x_j - \bar{x}) \le t) - I(\hat{\varepsilon}^*_j \le t)])\}, \\
J^*_{n1}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - Ex)\{I(\hat{\varepsilon}^*_j \le t) - F_n(t)\}, \\
J^*_{n2}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - Ex)E^*[I(\hat{\varepsilon}^*_j - \theta^\tau_n(x_j - \bar{x}) \le t) - I(\hat{\varepsilon}^*_j \le t)]\}, \qquad (6.7) \\
J^*_{n3}(t) &= \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - Ex)E^*[F^*_n(t) - F_n(t)]\}.
\end{aligned}$$

Similar to (6.5) we can derive that for any θ_n satisfying $||\theta_n|| \le c \log\, n/n^{\frac{1}{2}}$ and for almost all sequences $\{(x_1, y_1), \ldots, (x_n, y_n), \ldots\}$, $J^*_n(t) \longrightarrow 0$ a.s. uniformly on t, where E^* stands for the integration on the bootstrap variable $\hat{\varepsilon}^*_i$. Along with the arguments used for proving I^*_{n2}, we can verify that J^*_{n1} converges in distribution to B, J^*_{n2} converges in distribution to $f_\varepsilon \cdot N$ and J^*_{n3} tends to zero in probability. We omit the details of the proof.

Proof of Theorem 3.2. The argument is similar to that for Theorem 3.1, hence we only present an outline. Let

$$R^*_{n4} = \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - \bar{x})\{I(\varepsilon^*_j - (\beta^*_n - \beta_n)^\tau (x_j - \bar{x}) \le t) - I(\varepsilon^*_j \le t)\}.$$

Consider $R^*_{n4} - E_{w^*} R^*_{n4}$ first, where

$$E_{w^*} R^*_{n4} = \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - \bar{x}) E_{w^*_j}\{I(w^*_j \hat{\varepsilon}_j - (\beta^*_n - \beta_n)^\tau (x_j - \bar{x}) \le t) - I((w^*_j \hat{\varepsilon}_j \le t)\},$$

and E_{w^*} is the expectation over w^*. Since $\beta^*_n - \beta_n = O(\log n/\sqrt{n})$ a.s., similar to Lemma 6.1, we can verify that, for almost all sequences $\{(x_1, y_1) \ldots, (x_n, y_n), \ldots\}$, $R^*_{n4}(t) - E_{w^*} R^*_{n4}(t) \longrightarrow 0$ a.s. uniformly on $t \in R^1$. Decompose $I^*_{n1}(t)$ as $I^*_{n1}(t) = R^*_{n4}(t) - E_{w^*} R^*_{n4}(t) + R^*_{n5}(t) + E_{w^*} R^*_{n4}(t)$ where $R^*_{n5}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - \bar{x}) I(\varepsilon^*_j \le t)$. We now show that $E_{w^*} R^*_{n4}$ converges to $-f_\varepsilon \cdot N$. Let E_{ε, w^*} denote the expectation on ε and w^*. Define

$$\begin{aligned}
&E_{\varepsilon, w^*} R^*_{n41}(t) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^n (x_j - \bar{x})[E_{\varepsilon_j, w^*_j} I(w^*_j \hat{\varepsilon}_j - (\beta^*_n - \beta_n)(x_j - \bar{x} \le t) - E_{\varepsilon_j w^*_j} I(w^*_j \hat{\varepsilon}_j \le t)].
\end{aligned}$$

Then $E_{w^*}R^*_{n4}(t) = \{E_{w^*}R^*_{n4}(t) - E_{\varepsilon,w^*}R^*_{n41}(t)\} + E_{\varepsilon,w^*}R^*_{n41}(t)$. With the argument of Lemma 6.1 again, $\{E_{w^*}R^*_{n4}(t) - E_{\varepsilon,w^*}R^*_{n41}(t)\} \longrightarrow 0$ a.s. uniformly on $t \in R^1$. Now consider $E_{\varepsilon,w^*}R^*_{n41}(t)$. Note that for each j, $E_{w^*_j}I(w^*_j\hat{\varepsilon}_j \le t) = 1/2I(\hat{\varepsilon}_j \le t) + 1/2I(-\hat{\varepsilon}_j \le t)$ and then

$$\begin{aligned}&E_{\varepsilon_j,w^*_j}I(w^*_j\hat{\varepsilon}_j \le t)\\ &= \frac{1}{2}F_\varepsilon\Big(\frac{t-(\beta_n-\beta)^\tau_{(j)}(x_j-\bar{x})}{1-(x_j-\bar{x})^\tau S_n^{-1}(x_j-\bar{x})}\Big) + \frac{1}{2}\Big(1-F_\varepsilon\Big(\frac{-t+(\beta_n-\beta)^\tau_{(j)}(x_j-\bar{x})}{1-(x_j-\bar{x})^\tau S_n^{-1}(x_j-\bar{x})}\Big)\Big),\end{aligned}$$

where $(\beta_n-\beta)_{(j)} = S_n^{-1}\sum_{i\ne j}(x_j-\bar{x})\varepsilon_i$. Taylor expansion yields that

$$\begin{aligned}E_{\varepsilon w^*}(R^*_{n41}(t)) &= -\frac{1}{2\sqrt{n}}\sum_{j=1}^n(x_j-\bar{x})(x_j-\bar{x})^\tau(\beta_n-\beta)(f_\varepsilon(t)+(f_\varepsilon(-t))+o_p(1) \quad \text{a.s.}\\ &= -f_\varepsilon(t)\cdot N + o_p(1) \quad \text{a.s.}\end{aligned}$$

The last equation is due to the symmetry of f_ε. Now we are in the position to show that R^*_{n5} does not converge in distribution to the Gaussian process B so that the conclusion of the theorem is reached. We can see this immediately by calculating the variance of R^*_{n5} at each t. Actually, $\lim_{n\to\infty}\mathrm{Var}\,(R^*_{n5}(t)) = 1/4E(I(\varepsilon \le t)-I(-\varepsilon \le t))^2$ which is not equal to $\mathrm{Var}\,(B(t)) = (F_\varepsilon(t)(1-(F_\varepsilon(t))$. The proof is completed.

Acknowledgements

Research support from the NSF of China, a grant from the University of Hong Kong, and a grant from the Research Grant Council of Hong Kong is gratefully acknowledged. The author thanks the Editor and an associate editor for their suggestions which greatly improved the presentation of the article as well as, a referee for providing a key reference for the proof of Theorem 2.1.

References

Buckley, M. J. (1991). Detecting a smooth signal: optimality of cusum based on procedures. *Biometrika* **78**, 253-262.

Dawkins, B. (1989). Multivariate analysis of national track records. *Amer. Statist.* **43**, 110-112.

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

Gaenssler, P. (1983). *Empirical Processes.* Lecture Notes-Monograph Series 3: Institute of Mathematical Statistics, Hayward, California.

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

Hart, J. D. (1997). *Nonparametric Smoothing and Lack-of-fit Tests.* Springer, New York.

Koul, H. L. (1992). *Weighted Empiricals and Linear Models.* Lecture Notes—Monograph Series, 21. Institute of Mathematical Statistics, Hayward, California.

Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16**, 1696-1708.

Müller, H. G. (1992). Goodness-of-fit diagnostics for regression models. *Scand. J. Statist.* **19**, 157-172.

Naik, D. N. and Khattree, R. (1996). Revisiting Olympic track records: Some practical considerations in the principal component analysis. *Amer. Statist.* **50**, 140-144.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag, New York.

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

Stute, W. and Manteiga, G. W. (1995). NN goodness-of-fit tests for linear models. *J. Statist. Plann. Inference* **53**, 75-92.

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

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

Su, J. Q. and Wei, L. J.(1991). A lack-of-fit test for the mean function in a generalized linear model. *J. Amer. Statist. Assoc.* **86**, 420-426.

Wu, C. F. J. (1986). Jackknife, bootstrap and other re-sampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1295.

Institute of Applied Mathematics, Chinese Academy of Sciences, Beijing, China and Department of Statistics, The University of Hong Kong, Pokfulam Road, Hong Kong.

E-mail: lzhu@hkucc.hku.hk

(Received December 2000; accepted October 2002)