Statistics and Probability Letters 80 (2010) 1378–1387

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Distribution-free lack-of-fit tests in balanced mixed models

Weixing Song *, Juan Du

Kansas State University, 66502 Manhattan, KS, United States

ARTICLE INFO

Article history:
Received 17 December 2009
Received in revised form 3 May 2010
Accepted 5 May 2010
Available online 21 May 2010

MSC:
primary 62G08
secondary 62G10

Keywords:
Mixed model
Lack-of-fit test
Khmaladze transformation
Brownian motion

ABSTRACT

Here we discuss the problem of fitting a parametric model to the regression function of the fixed effects in a class of balanced mixed effects models. The proposed test is based on the supremum of the Khmaladze transformation of a certain partial sum process of calibrated residuals, and the asymptotic null distribution of this transformed process turns out to be the same as that of a time transformed standard Brownian motion. Moreover, we show that this test is consistent against a large class of fixed alternatives and has non-trivial asymptotic power against a class of nonparametric local alternatives. Simulation studies are conducted to assess the finite sample performance of the proposed test.

1. Introduction

Classical statistical inferences often assume that the observations are independent, even identically distributed, but modern scientific research generates many data sets which may fail to satisfy this requirement; instead, the data present a much more complicated multilevel and hierarchical structure. For example, in clinical trial studies, a measurement may be taken from patients at different time points. Observations from different patients are independent, but the repeated measurements from the same patient are correlated. Classical inference procedures applied to these kinds of data sets are more likely to lead to misleading or invalid conclusions; see the price–sale example in Demidenko (2004), and the reading-age example in Diggle et al. (2002). Beginning with Laird and Ware (1982), mixed modeling has become more and more popular for analyzing such data sets and its theory has been well developed. However, most research is devoted to looking for efficient procedures for estimation of the regression parameters from the fixed effects, and the covariance matrix from the random effects. When doing this, the form of the regression function of the fixed effects is often predetermined. This has a potential risk that any deviation from the specified form will be likely to jeopardize the validity of estimation procedures and results. Therefore, it is very practical and also necessary to develop formal lack-of-fit test procedures for checking the adequacy of regression functions.

Graphical tools, such as the residual plots, are routinely used for model checking in various models, such as classical regression models, time series models etc.; see Brockwell and Davis (2002) and Cook and Weisberg (1994). Although these graphs are informative, drawing conclusions based merely on these is quite subjective. Pan and Lin (2005) develop some numerical methods for checking the adequacy of generalized linear mixed models. Their methods are based on the cumulative sums of residuals over covariates or predicted values of the response variable. These stochastic processes converge weakly to zero-mean Gaussian processes. Since the covariance matrix of this Gaussian process is so complicated that the resulting test statistic is not distribution-free, the application of the testing procedure is done through Monte Carlo simulation.

* Corresponding author.
E-mail address: weixing@ksu.edu (W. Song).

0167-7152/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2010.05.002

In this work, we study the following balanced mixed effects model:

$$Y_{ij} = g(X_i) + Z_{ij}'b_i + \varepsilon_{ij}, \quad i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, m,$$

where $g(\cdot)$ is the fixed effect regression function, b_i, $i = 1, 2, \ldots, n$, are i.i.d. $q \times 1$ vectors of unobservable random effects having a distribution with mean vector 0 and covariance matrix V. X_i, $i = 1, 2, \ldots, n$, are i.i.d. one-dimensional covariates. $Z_i = (Z_{i1}', Z_{i2}', \ldots, Z_{im}')'$, $i = 1, 2, \ldots, n$, are i.i.d. random vectors, Z' denotes the transpose of Z, and ε_{ij}, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$, are i.i.d. with mean 0 and variance σ^2. For the sake of identifiability, we will assume $m > 1$. Using X, Z, b and ε to denote the generic population, we shall assume that X, Z, b and ε are independent. The objective here is to test

$$H_0 : g(x) = g(x, \beta), \quad \text{for some } \beta \in \Gamma \text{ and all } x \in \mathbb{R},$$

where $g(x, \beta)$ is a parametric function with β taking values in a compact subset Γ of $\mathbb{R}^p$.

Comparing with the existing methods, the test procedure proposed in this work will be based on the Khmaladze transformation of a certain marked residual process. The transformed marked residual process converges weakly to a time transformed Brownian motion in uniform metric. Consequently, any test based on a continuous functional of this process is asymptotically distribution-free, and can be implemented at least for moderate to large samples without resorting to a resampling method.

The rest of the work is organized as follows. The marked residual process and its asymptotic null distribution are discussed in Section 2 under quite broad assumptions. Consistency and asymptotic power against $n^{-1/2}$-local nonparametric alternatives of the test are discussed in Section 3. Section 4 contains the simulation study. All the proofs will be postponed to Section 5.

Throughout the rest of the work, B denotes standard Brownian motion on $[0, 1]$, and for any random variable V, F_V, f_V denote its distribution and density functions, respectively.

2. Assumptions and main results

Define

$$S_i = \sum_{j=1}^{m}[Y_{ij} - g(X_i)] = \sum_{j=1}^{m}[Z_{ij}'b_i + \varepsilon_{ij}], \quad i = 1, 2, \ldots, n$$

and $\widetilde{Z}_i = \sum_{j=1}^{m} Z_{ij}$. Then it is easy to see that S_i are i.i.d. with mean 0 and conditional variance $E[S_i^2|X_i, Z_{ij}, j = 1, 2, \ldots, m] = m\sigma^2 + \widetilde{Z}_i'V\widetilde{Z}_i$ which depends on $\widetilde{Z}_i$ only and will be denoted as $\tau^2(\widetilde{Z}_i)$ in the sequel. By a routine argument, one can show that, as $n \to \infty$,

$$T_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{S_i}{\tau(\widetilde{Z}_i)}I[X_i \le x] \Longrightarrow B \circ (F_X(x))$$

weakly in $D(\bar{\mathbb{R}}) = D([-\infty, \infty])$, and in uniform metric. If β, V and σ^2 are all known, one can simply test the hypothesis $H_0 : g(x) = g(x, \beta)$ based on $T_n(x)$. However, these parameters are usually unknown in practice. The natural way to proceed is to replace these unknown quantities with some consistent estimators, denoted by $\hat{\beta}, \hat{V}$ and $\hat{\sigma}^2$. Define $\widetilde{Y}_i = \sum_{j=1}^{n} Y_{ij}$, $S_i(\beta) = \sum_{j=1}^{m}[Y_{ij} - g(X_i, \beta)]$; then $T_n(x)$ can be written as

$$\hat{T}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{S_i(\hat{\beta})}{\hat{\tau}(\widetilde{Z}_i)}I[X_i \le x] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widetilde{Y}_i - mg(X_i, \hat{\beta})}{\sqrt{m\hat{\sigma}^2 + \widetilde{Z}_i'\hat{V}\widetilde{Z}_i}}I[X_i \le x].$$

But the limiting process for $\hat{T}_n(x)$ under the null hypothesis will not be distribution-free. In fact, it depends on the CDF F_X of X, the local structure of g, given in terms of the derivative of g with respect to its parameter, and the conditional variance $\tau^2(\widetilde{Z})$. Although we can show that the limiting process will be a Gaussian process with mean 0, the covariance matrix has a complicated form, which makes the test procedure hard to implement in real applications. A similar situation occurs for the lack-of-fit test in the classical regression models or the measurement error models; see Koul and Song (2009), Stute et al. (1998) (STZ) etc. To construct a distribution-free test statistic, we introduce the so-called Khmaladze martingale transformation of $\hat{T}_n(x)$. The Khmaladze transformation was first introduced by Khmaladze (1981, 1988), then soon became a powerful tool for constructing distribution-free test statistics. Suppose a stochastic process, say $R(x)$, has the same distribution as the sum of a Brownian motion $B(x)$, and a Gaussian process $U(x)$. The Khmaladze transformation of $R(x)$ is a linear transformation L of $R(x)$ such that $L(R(x)) = L(B(x)) + L(U(x)) = B(x)$ in distribution. For more about the Khmaladze transformation, see Khmaladze (1981, 1988, 1993), Koul (2006) and Stute et al. (1998) and the references therein.

Define $\dot{g}(x, \beta) = \partial g(x, \beta)/\partial\beta$, $l(X, \widetilde{Z}) = m\tau^{-1}(\widetilde{Z})\dot{g}(X, \beta)$, $\xi = S(\beta)/\tau(\widetilde{Z}) = (\widetilde{Y} - mg(X, \beta))/\tau(\widetilde{Z})$, $a_1 = E\tau^{-1}(\widetilde{Z})$, $a_2 = E\tau^{-2}(\widetilde{Z})$, and $M_x = El(X, \widetilde{Z})l'(X, \widetilde{Z})I[X \ge x]$. By the independence of X and Z, we have $M_x = a_2m^2E\dot{g}(X, \beta)\dot{g}'(X, \beta)I[X \ge x]$. Then the Khmaladze transformation of $\hat{T}_n(x)$ has the form

$$\widehat{W}_n(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_iI(X_i \le x) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i\left[\frac{m\hat{a}_1}{n}\sum_{j=1}^{n}I(X_j \le x)\dot{g}'(X_j, \hat{\beta})\hat{M}_{X_j}^{-1}I(X_i \ge X_j)\right]\hat{l}(X_i, \widetilde{Z}_i),$$

where

$$\hat{\xi}_i = S_i(\hat{\beta})/\hat{\tau}(\widetilde{Z}_i), \qquad \hat{a}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}^{-1}(\widetilde{Z}_i), \qquad \hat{a}_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\tau}^{-2}(\widetilde{Z}_i). \tag{2.1}$$

$$\hat{l}(X_i, \widetilde{Z}_i) = \frac{m\dot{g}(X_i, \hat{\beta})}{\widehat{\tau}(\widetilde{Z}_i)}, \qquad \hat{M}_x = \frac{\hat{a}_2 m^2}{n}\sum_{i=1}^{n}\dot{g}(X_i, \hat{\beta})\dot{g}'(X_i, \hat{\beta})I[X_i \ge x].$$

The following is a list of the assumptions which we will need in the sequel to derive the asymptotic results for the process $\widehat{W}_n(x)$.

(e) $E\|Z\|^4 < \infty$.

(g1) $g(x, \beta)$ is differentiable with respect to β with $E\|\dot{g}(X, \beta_0)\|^2 < \infty$. For any $\sqrt{n}$-consistent estimator $\hat{\beta}$, $\max_{1\le i\le n} \sqrt{n}|g(X_i, \hat{\beta}) - g(X_i, \beta_0) - (\hat{\beta} - \beta_0)'\dot{g}(X_i, \beta_0)| = o_p(1)$, where β_0 is the true value of β under the null hypothesis.

(g2) For some $q \times q$ square matrix $\ddot{g}(x, \beta_0)$ and a nonnegative function $k_{\beta_0}(x)$, both measurable in the x coordinate, the following holds:

$$E\|\ddot{g}(X, \beta_0)\|^2 < \infty, \qquad E\|\ddot{g}(S, \beta_0)\|\|\dot{g}(X, \beta_0)\|^j < \infty, \qquad E\|\ddot{g}(X, \beta_0)\|^j k_{\beta_0}(X) < \infty,$$

$j = 0, 1$, and for all $\delta > 0$, there exists an $\eta > 0$ such that $\|\beta - \beta_0\| \le \eta$ implies

$$\|\dot{g}(x, \beta) - \dot{g}(x, \beta_0) - \ddot{g}(x, \beta_0)(\beta - \beta_0)\| \le \delta k_{\beta_0}(x)\|\beta - \beta_0\|, \quad \text{a.s. } F_X.$$

(m) $E\dot{g}(X, \beta_0)\dot{g}'(X, \beta_0)I(X \ge x)$ is positive definite for all $x \in \mathbb{R}$.

The following theorem gives the weak convergence result for the process $\widehat{W}_n(x)$.

Theorem 2.1. *Suppose* (e), (g1), (g2), (m) *hold. Then under H_0, for every $x_0 < \infty$, $\widehat{W}_n \Rightarrow B\circ F_X$, in $D([-\infty, x_0])$ and in uniform metric.*

As in STZ, it is recommended to apply the above result with x_0 equal to the 99th percentile of $\hat{F}_X$. Consequently, the test that rejects H_0 whenever

$$\sup_{x\le x_0}\left|\widehat{W}_n(x)\Big/\sqrt{\hat{F}_X(x_0)}\right| = \sup_{x\le x_0}\left|\widehat{W}_n(x)\Big/\sqrt{0.99}\right| > b_\alpha$$

will be of the asymptotic size α, where b_α is such that $P(\sup_{0\le u\le 1}|B(u)| > b_\alpha) = \alpha$.

Remark 1. The restriction of the weak convergence of $\widehat{W}_n(x)$ over $[-\infty, x_0]$ is a technical one. See more discussions on this issue in Remark 5.2. The choice of x_0 introduces some subjectivity into our test. In real applications, it is suggested that a large x_0 should be chosen to cover most of the range of X. Simulation studies show that the performance of the test is not sensitive to the choice of x_0.

Remark 2. The proposed testing procedure has some limitations with respect to the fixed design variable X. It requires that X should be the same for all of the subjects within a level or cluster. In some cases, this is not necessary. For example, if the model under consideration is $Y_{ij} = g(X_{ij}) + Z_{ij}'b_i + \varepsilon_{ij}$, and the hypothesis to test is $H_0 : g(x) = x\beta$, then our test procedure still applies on simply replacing X_i with $\widetilde{X}_i = \sum_{j=1}^{m} X_{ij}$, and the CDF of X with that of $\widetilde{X}$. Of course, that $\widetilde{X}_i$, $i = 1, 2, \ldots, n$, are i.i.d. will be assumed.

3. Consistency and local power

The ability to detect any deviation from the null hypothesis is often referred to as the consistency. More specifically, for a fixed alternative, a consistent test should have power that tends to 1 as the sample size goes to ∞. In this section, we show that our test is consistent. To this end, consider a general class of fixed alternative hypotheses: $H_a : g(x) = h(x)$, where $h(x) \notin \{g(x, \beta) : \beta \in \Gamma\}$ and $Eh^2(X) < \infty$.

In the previous section, we assume that estimators $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{V}$ are $\sqrt{n}$-consistent. Would these estimators still have a similar property under the alternative hypothesis H_a? The question is of interest in its own right. For the classical regression set-up, Jennrich (1969) and White (1981, 1982) showed that, under some mild regularity conditions, the nonlinear least squares estimator is consistent and asymptotically normal even in the presence of model misspecification. This idea can be borrowed here to show the consistency and asymptotic normality of $\hat{\beta}$, $\hat{\sigma}^2$, and $\hat{V}$ under H_a. We will not justify this rigorously here. But we should note that $\hat{\sigma}^2$ and $\hat{V}$ may not converge to the true values of σ^2 and V, say σ_0^2 and V_0. In the following, we assume $\sqrt{n}(\hat{\beta} - \beta_1) = O_p(1)$, $\sqrt{n}(\hat{\sigma}^2 - \sigma_1^2) = O_p(1)$, and $\sqrt{n}(\hat{V} - V_1) = O_p(1)$ for some β_1, σ_1^2, and V_1 under the alternative H_a.

Table 1
$\sigma^2 = \sigma_b^2 = 0.5^2$.

Model \ n	50	100	200	300	500
Model 0	0.033	0.036	0.038	0.042	0.040
Model 1	0.756	0.974	1.000	1.000	1.000
Model 2	0.724	0.967	1.000	1.000	1.000

By $\tau_1^2(\widetilde{Z}_i)$ we denote $\tau^2(\widetilde{Z}_i)$ with all the parameters replaced by β_1, σ_1^2, and V_1, a_1 and a_2 are similarly defined with τ replaced by τ_1. Define

$$\rho(x) = m^2 a_2 E\left[h(X) - g(X, \beta_1)\right]\dot{g}(X, \beta_1) I(X \geq x),$$
$$D_1(x) = m a_1 E\left[h(X) - g(X, \beta_1)\right] I(X \leq x),$$
$$D_2(x) = m a_1 E[\dot{g}'(X, \beta_1)\widetilde{M}_X^{-1}\rho(X) I(X \leq x)],$$

and $d(x_0) = \max_{x\leq x_0} |D_1(x) - D_2(x)|$. Then we have the following result:

Theorem 3.1. *Suppose all the conditions in Theorem 2.1 hold with β_0, σ_0^2, and V_0 replaced by β_1, σ_1^2, and V_1, $d(x_0) > 0$. Then for any $0 < \alpha < 1$, the test that rejects H_0 whenever $\sup_{x\leq x_0} |\widehat{W}_n(x)/\sqrt{\hat{F}_X(x_0)}| > b_\alpha$ is consistent.*

Sometimes it is desirable to investigate the performance of a test statistic at local alternatives, since the consistency tells us nothing about the power when the sample size is relatively small. Let $\delta(x)$ be a measurable function such that $E\delta^2(X) < \infty$. Consider the following sequence of local alternatives $H_{\text{Loc}} : g(x) = g(x, \beta) + \delta(x)/\sqrt{n}$. We assume that the estimators that we used in the test statistic are such that $\sqrt{n}(\hat{\beta} - \beta_0) = O_p(1)$, $\sqrt{n}(\hat{\sigma}^2 - \sigma_0^2) = O_p(1)$ and $\sqrt{n}(\hat{V} - V_0) = O_p(1)$. Define $\rho(x) = a_2 m^2 E\delta(X)\dot{g}(X, \beta_0) I(X \geq x)$, and $k(x) = m a_1 E\delta(X) I(X \leq x) - m a_1 E\dot{g}'(X, \beta_0) M_X^{-1}\rho(X) I(X \leq x)$, where a_1 and a_2 are defined like in Section 2. The power of the test against H_{Loc} can be readily obtained from the following theorem:

Theorem 3.2. *Suppose all the conditions in Theorem 2.1 hold. Then under H_{Loc}, $\widehat{W}_n(x) \Longrightarrow B \circ F_X(x) + k(x)$ weakly in $D[-\infty, x_0]$ and in uniform metric.*

Remark. It is easy to see that if the test has non-trivial asymptotic power against H_{Loc}, $\delta(x)$ has to satisfy $\sup_{x\leq x_0} |k(x)| > 0$.

4. Simulation

Simulation studies are conducted in this section to evaluate the finite sample behavior of the proposed test procedure. x_0 is chosen to be the 99th percentile of the empirical distribution function $\hat{F}_X$. For the significance level $\alpha = 0.05$, the critical value b_α obtained from the distribution of $\sup_{0\leq u\leq 1} |B(u)|$ is 2.24241. See, e.g., Khmaladze and Koul (2004). In the following simulation, for various sample sizes, we repeated the above procedure 1000 times and the empirical size and power are computed by using $\#\left\{\sup_{x\leq x_0} |\widehat{W}_n(x)/\sqrt{0.99}| \geq b_\alpha\right\}/1000$. The effect of a different x_0 value on the performance of our test will be discussed at the end of this section.

Simulation 1: We generate samples from the following three models:

Model 0: $Y_i = X\beta + Zb + \varepsilon$,
Model 1: $Y_i = X\beta + Zb + 0.1X\exp(X) + \varepsilon$,
Model 2: $Y_i = X\beta + Zb + \text{sgn}(X)X^2 + \varepsilon$

where $X \sim N(0, 1)$, $Z \sim N(0, 1)$, $b \sim N(0, \sigma_b^2)$, and $\varepsilon \sim N(0, \sigma^2)$. The null model to be fitted is $g(x, \beta) = x\beta$. The true parameters are chosen to be $\beta = 1$. Data from Model 0 are used to study the empirical level, while data from Models 1, 2 are used to study the empirical power of the test. Various values of σ_b^2 and σ^2 are used in the simulation. Since the simulation results are similar for all cases, only results for $\sigma^2 = \sigma_b^2 = 0.5^2$ are reported in Table 1.

The simulation study shows that the empirical levels are all less than the nominal levels in all the chosen cases; hence the proposed test is conservative for the chosen sample sizes. The overall performance of the test against the two alternative models is satisfactory, as it has good power (almost 95%) for small sample sizes.

Simulation 2: Simulation study on a quadratic fixed effect mixed model is conducted here. The data sets are generated from the following models:

Model 0: $Y = \beta_1 X + \beta_2 X^2 + Zb + \varepsilon$,
Model 1: $Y = \beta_1 X + \beta_2 X^2 + Zb + 0.2X\exp(X) + \varepsilon$,
Model 2: $Y = \beta_1 X + \beta_2 X^2 + Zb + \text{sgn}(X)X^3 + \varepsilon$.

The null hypothesis is $H_0 : g(x, \beta) = x\beta_1 + x^2\beta_2$. Data from Model 0 are used to study the empirical level, while data from Models 1, 2 are used to study the empirical power of the test. In the simulation, $\varepsilon \sim N(0, \sigma^2)$, $X \sim N(0, 1)$, $Z \sim N(0, 1)$, and $b \sim N(0, \sigma_b^2)$. To see the effects of σ^2 and σ_b^2 on the test, we tried a variety of values. Since the simulation results are similar, we only report two cases. Table 2 gives the simulation results for $\sigma^2 = \sigma_b^2 = 0.5^2$. Table 3 is for $\sigma^2 = \sigma_b^2 = 0.8^2$.

Table 2
$\sigma^2 = \sigma_b^2 = 0.5^2$.

Model \ n	50	100	200	300	500
Model 0	0.018	0.031	0.040	0.040	0.041
Model 1	0.776	0.988	1.000	1.000	1.000
Model 2	0.761	0.987	0.999	1.000	1.000

Table 3
$\sigma^2 = \sigma_b^2 = 0.8^2$.

Model \ n	50	100	200	300	500
Model 0	0.023	0.029	0.037	0.034	0.052
Model 1	0.558	0.900	0.997	1.000	1.000
Model 2	0.621	0.890	0.998	1.000	1.000

Table 4
$\sigma_{b_1 b_2} = 0$.

Model \ n	50	100	200	300	500
Model 0	0.033	0.040	0.040	0.047	0.049
Model 1	0.478	0.815	0.991	1.000	1.000
Model 2	0.214	0.420	0.727	0.886	0.985

Table 5
$\sigma_{b_1 b_2} = 0.25$.

Model \ n	50	100	200	300	500
Model 0	0.035	0.034	0.041	0.044	0.051
Model 1	0.495	0.809	0.988	0.999	1.000
Model 2	0.240	0.422	0.736	0.903	0.989

Table 6
$\sigma_{b_1 b_2} = 0.25, x_0 = \infty$.

Model \ n	50	100	200	300	500
Model 0	0.037	0.034	0.040	0.044	0.048
Model 1	0.489	0.809	0.987	0.999	1.000
Model 2	0.238	0.428	0.685	0.898	0.995

The simulation study shows that most empirical levels tend to be less than the nominal levels in all the chosen cases, which indicates that the proposed test is conservative—in particular, when the sample size is small. But the overall performance of the test against the two alternatives is satisfactory.

Simulation 3: A simulation on a two-dimensional random effects mixed model is also conducted here. The data sets are generated from the following models:

Model 0: $Y = X\beta + Z_1 b_1 + Z_2 b_2 + \varepsilon$,
Model 1: $Y = X\beta + Z_1 b_1 + Z_2 b_2 + 0.1X \exp(X) + \varepsilon$,
Model 2: $Y = X\beta + Z_1 b_1 + Z_2 b_2 + 0.5\mathrm{sgn}(X)X^2 + \varepsilon$.

The null hypothesis is $H_0 : g(x, \beta) = x\beta$. Again data from Model 0 are used to study the empirical level, while data from Models 1, 2 are used to study the empirical power of the test. In the simulation, $\varepsilon \sim N(0, \sigma^2), X \sim N(0, 1), Z_1 \sim N(0, 1), Z_2 \sim N(0, 1)$, and (b_1, b_2) has a bivariate normal distribution with mean vector $(0, 0)$, variances $\mathrm{Var}(b_1) = \sigma_{b_1}^2, \mathrm{Var}(b_2) = \sigma_{b_2}^2$, and covariance $\mathrm{Cov}(b_1, b_2) = \sigma_{b_1 b_2}$. To see the effects of σ^2 and the covariance matrix V on the test, we tried a variety of values. Since the simulation results are similar, we only report two cases. Table 4 gives the simulation results for $\sigma^2 = 0.5^2, \sigma_{b_1}^2 = \sigma_{b_2}^2 = 1$, and $\sigma_{b_1 b_2} = 0$. Table 5 is for $\sigma^2 = 0.5^2, \sigma_{b_1}^2 = \sigma_{b_2}^2 = 1$, and $\sigma_{b_1 b_2} = 0.25$.

As in the previous simulation studies, the above simulation study shows that the proposed test is conservative for those sample sizes. Through simulation studies, we also find that the power decreases when the variance $\sigma^2, \sigma_{b_1}^2$, and $\sigma_{b_2}^2$ increase, but the correlation within the random effects seems to have no significant effect on the performance of the test.

To see the effects of the choice of x_0, we also conduct a simulation study in which all other set-ups are the same as in Table 5 except that x_0 is taken to be ∞. The result shows that our test is not sensitive to the choice of x_0 as far as the finite sample behavior is concerned (Table 6).

5. Proofs

To prove Theorem 2.1, we need three lemmas. Throughout this section, $u_p(1)$ stands for a sequence of stochastic processes that tends to zero, uniformly over its time domain, in probability.

Lemma 5.1. *Suppose ξ and U are random variables with $E(\xi|U) = 0$, $0 < E\xi^2 < \infty$. Let $\sigma^2(u) = E(\xi^2|U = u)$, $L(u) = E\sigma^2(U)I[U \le u]$, $u \in \bar{\mathbb{R}}$. Let (ξ_i, U_i), $1 \le i \le n$, be* i.i.d. *copies of (ξ, U). Define*

$$U_n(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i I[U_i \le u], \quad u \in \bar{R},$$

and L is assumed to be continuous. Then $U_n \Longrightarrow B \circ L$ in $D(\bar{\mathbb{R}})$ and uniform metric.

The proof of this lemma is based on Theorems 12.6, 15.5, in Billingsley (1968). The details are similar to those appearing in STZ, and therefore are omitted here.

To state the next lemma, let U be a continuous r.v. with CDF G. Let $\ell(u)$ be a vector of q functions with $E\|\ell(U)\|^2 < \infty$. Assume the matrix $C_u := E\ell(U)\ell'(U)I(U \ge u)$ is positive definite for all $u \in \mathbb{R}$. For a real valued function $\gamma \in L_2(\mathbb{R}, G)$ define the transforms

$$\mathcal{T}_\gamma(u) := \int_{y \le u} \gamma(y)\ell'(y)C_y^{-1}\mathrm{d}G(y)\ell(u), \qquad \mathcal{K}_\gamma(u) := \gamma(u) - \mathcal{T}_\gamma(u).$$

The following lemma follows from Proposition 4.1 of Khmaladze and Koul (2004) and Lemma 9.1 of Koul (2006), which in turn has its origin in Khmaladze (1988).

Lemma 5.2. *Under the above set-up,*

$$E\mathcal{K}_\gamma(U)\ell'(U) = 0, \quad \forall \gamma \in L_2(\mathbb{R}, G)$$
$$E\mathcal{K}_{\gamma_1}(U)\mathcal{K}_{\gamma_2}(U) = E\gamma_1(U)\gamma_2(U), \quad \forall \gamma_1, \gamma_2 \in L_2(\mathbb{R}, G).$$

Remark 5.1. Let ξ be a r.v. such that $E(\xi|U) = 0$, $E\xi^2 < \infty$, $\tau^2(u) := E(\xi^2|U = u) > 0$, for all u. Then the covariance of the process $W_\gamma(\xi, U) := [\xi/\tau(U)]\mathcal{K}_\gamma(U)$, as a process in $\gamma \in L_2(\mathbb{R}, G)$, is like that of $B_\gamma(G)$, where B_γ is a Brownian motion in γ. Hence, if (ξ_i, U_i), $1 \le i \le n$, are i.i.d. copies of (ξ, U), then by the classical CLT, the finite dimensional distributions of $n^{-1/2}\sum_{i=1}^n W_\gamma(\xi_i, U_i)$, as γ varies, will converge weakly to those of $B_\gamma(G)$.

To prove the main theorems, we also need the following lemma:

Lemma 5.3. *Suppose that* (e) *holds; then*

$$\max_{1\le i\le n} |\hat{\tau}^2(\widetilde{Z}_i) - \tau^2(\widetilde{Z}_i)| = o_p(1), \qquad \max_{1\le i\le n} |\hat{\tau}(\widetilde{Z}_i) - \tau(\widetilde{Z}_i)| = o_p(1),$$
$$\max_{1\le i\le n}\left|\frac{1}{\hat{\tau}(\widetilde{Z}_i)[\tau(\widetilde{Z}_i) + \hat{\tau}(\widetilde{Z}_i)]} - \frac{1}{2\tau^2(\widetilde{Z}_i)}\right| = o_p(1).$$

The proof is mainly based on the facts that $\hat{\tau}(\widetilde{Z}_i) \ge \hat{\sigma}^2$, $\tau^2(\widetilde{Z}_i) \ge \sigma^2$ and the $\sqrt{n}$-consistency of $\hat{\sigma}^2$ and $\hat{V}$. The details are omitted. Note that the above lemma also implies

$$\max_{1\le i\le n}\left|\frac{\tau(\widetilde{Z}_i)}{\hat{\tau}(\widetilde{Z}_i)} - 1\right| = o_p(1). \tag{5.1}$$

Proof of Theorem 2.1. We will use β_0, V_0, σ_0^2 to denote the true values of the corresponding parameters under H_0; $\tau_0(\widetilde{Z}_i)$ is the same as $\tau(\widetilde{Z}_i)$ with all parameters replaced by the null values. Let $\hat{U}_n(x) = n^{-1/2}\sum_{i=1}^n \hat{\xi}_i\hat{l}(X_i, \widetilde{Z}_i)I(X_i \ge x)$, where $\hat{\xi}_i$ and $\hat{l}(X_i, \widetilde{Z}_i)$ are defined in (2.1). Then $\widehat{W}_n(x)$ can be written as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \hat{\xi}_i I(X_i \le x) - m\hat{a}_1\int_{s\le x} \dot{g}'(s, \hat{\beta})\hat{M}_s^{-1}\hat{U}_n(s)\mathrm{d}\hat{F}_X(s) =: W_{n1}(x) - W_{n2}(x).$$

Let $\Delta_n(\widetilde{Z}_i) = \tau_0(\widetilde{Z}_i)/\widehat{\tau}(\widetilde{Z}_i) - 1$. By adding and subtracting $mg(X_i, \beta_0)$ from $mg(X_i, \hat{\beta})$ in the numerator of $\hat{\xi}_i$, $W_{n1}(x)$ can be written as the sum of four terms:

$$W_{n11}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_i I(X_i \le x),$$

$$W_{n12}(x) = -\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\hat{\beta})-g(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}I(X_i\le x),$$

$$W_{n13}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\Delta_n(\widetilde{Z}_i)I(X_i\le x),$$

$$W_{n14}(x) = -\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\hat{\beta})-g(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}\Delta_n(\widetilde{Z}_i)I(X_i\le x).$$

We already know that $W_{n11}(x)\Longrightarrow B\circ F_X(x)$ weakly in $D[-\infty,\infty]$ and in uniform metric. Note that

$$W_{n12}(x) = -\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\hat{\beta})-g(X_i,\beta_0)-\dot{g}'(X_i,\beta_0)(\hat{\beta}-\beta_0)}{\tau_0(\widetilde{Z}_i)}I(X_i\le x)+\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{\dot{g}'(X_i,\beta_0)}{\tau(\widetilde{Z}_i)}I(X_i\le x)(\hat{\beta}-\beta_0);$$

then from (g1) and the fact that $\tau_0(\widetilde{Z}_i)>m\sigma_0^2>0$, we know the first term on the right is $u_p(1)$. By a Glivenko–Cantelli type argument, we also have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\dot{g}(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}I(X_i\le x) = E\frac{\dot{g}(X,\beta_0)}{\tau(\widetilde{Z})}I(X\le x)+u_p(1).$$

Hence

$$W_{n12}(x) = -ma_1E[\dot{g}'(X,\beta_0)I(X\le x)]\sqrt{n}(\hat{\beta}-\beta_0)+u_p(1).$$

By definition of $\hat{\tau}^2(\widetilde{Z}_i)$ and $\tau^2(\widetilde{Z}_i)$,

$$\begin{aligned}W_{n13}(x) &= (\sigma_0^2-\hat{\sigma}^2)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_i}{2\tau^2(\widetilde{Z}_i)}I(X_i\le x)+(\sigma_0^2-\hat{\sigma}^2)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_iI(X_i\le x)\left[\frac{1}{\hat{\tau}(\widetilde{Z}_i)[\tau_0(\widetilde{Z}_i)+\hat{\tau}(\widetilde{Z}_i)]}-\frac{1}{2\tau^2(\widetilde{Z}_i)}\right]\\&\quad+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\xi_i\widetilde{Z}_i'(V-\hat{V})\widetilde{Z}_i}{2\tau^2(\widetilde{Z}_i)}I(X_i\le x)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\widetilde{Z}_i'(V-\hat{V})\widetilde{Z}_iI(X_i\le x)\left[\frac{1}{\hat{\tau}(\widetilde{Z}_i)[\tau_0(\widetilde{Z}_i)+\hat{\tau}(\widetilde{Z}_i)]}-\frac{1}{2\tau^2(\widetilde{Z}_i)}\right].\end{aligned}$$

The first and third terms are $u_p(1)$ which can be proved by a Glivenko–Cantelli argument and the $\sqrt{n}$-consistency of $\hat{\sigma}^2$ and $\hat{V}$. By (5.1), one can further show that the second and fourth terms are $u_p(1)$.

For $W_{n14}(x)$, we have

$$\begin{aligned}W_{n14}(x) &= -\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\hat{\beta})-g(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}\left[\frac{\tau_0(\widetilde{Z}_i)}{\hat{\tau}(\widetilde{Z}_i)}-1\right]I(X_i\le x)\\&= -\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\hat{\beta})-g(X_i,\beta_0)-(\hat{\beta}-\beta_0)'\dot{g}(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}\left[\frac{\tau_0(\widetilde{Z}_i)}{\hat{\tau}(\widetilde{Z}_i)}-1\right]I(X_i\le x)\\&\quad-\frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{(\hat{\beta}-\beta_0)'\dot{g}(X_i,\beta_0)}{\tau_0(\widetilde{Z}_i)}\left[\frac{\tau_0(\widetilde{Z}_i)}{\hat{\tau}(\widetilde{Z}_i)}-1\right]I(X_i\le x).\end{aligned}$$

From (g1) and (5.1), it follows that $W_{n14}(x)=u_p(1)$. Therefore, we show that

$$W_{n1}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_iI(X_i\le x)-ma_1E\dot{g}'(X,\theta_0)I(X\le x)\cdot\sqrt{n}(\hat{\beta}-\beta_0)+u_p(1). \tag{5.2}$$

To study the asymptotic property of the process $W_{n2}(x)$, let's look carefully into the process $\hat{U}_n(x)$. By routine adding and subtracting arguments, we can write $\hat{U}_n(x)$ as the sum of $U_n(x)$ and the other seven terms:

$$B_{n1}(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_il(X_i,\widetilde{Z}_i)\Delta_n(\widetilde{Z}_i)I(X_i\ge x),$$

$$B_{n2}(x) = \frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\beta_0)-g(X_i,\hat{\beta})}{\tau_0(\widetilde{Z}_i)}l(X_i,\widetilde{Z}_i)I(X_i\ge x),$$

$$B_{n3}(x) = \frac{m}{\sqrt{n}}\sum_{i=1}^{n}\frac{g(X_i,\beta_0)-g(X_i,\hat{\beta})}{\tau_0(\widetilde{Z}_i)}l(X_i,\widetilde{Z}_i)\Delta_n(\widetilde{Z}_i)I(X_i\ge x),$$

$$B_{n4}(x) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi_i[\dot{g}(X_i, \hat{\beta}) - \dot{g}(X_i, \beta_0)] I(X_i \geq x)}{\tau_0(\widetilde{Z}_i)},$$

$$B_{n5}(x) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \frac{\xi_i[\dot{g}(X_i, \hat{\beta}) - \dot{g}(X_i, \beta_0)] I(X_i \geq x)}{\tau_0(\widetilde{Z}_i)} \Delta_n(\widetilde{Z}_i),$$

$$B_{n6}(x) = \frac{m^2}{\sqrt{n}} \sum_{i=1}^{n} \frac{[g(X_i, \beta_0) - g(X_i, \hat{\beta})][\dot{g}(X_i, \hat{\beta}) - \dot{g}(X_i, \beta_0)]}{\tau_0^2(\widetilde{Z}_i)} I(X_i \geq x),$$

$$B_{n7}(x) = \frac{m^2}{\sqrt{n}} \sum_{i=1}^{n} \frac{[g(X_i, \beta_0) - g(X_i, \hat{\beta})][\dot{g}(X_i, \hat{\beta}) - \dot{g}(X_i, \beta_0)]}{\tau_0^2(\widetilde{Z}_i)} I(X_i \geq x) \Delta_n(\widetilde{Z}_i).$$

Since

$$B_{n2}(x) = \frac{m}{\sqrt{n}} \sum_{i=1}^{n} \frac{g(X_i, \beta_0) - g(X_i, \hat{\beta}) + (\hat{\beta} - \beta_0)' \dot{g}(X_i, \beta_0)}{\tau_0(\widetilde{Z}_i)} l(X_i, \widetilde{Z}_i) I(X_i \leq x)$$
$$- \frac{m}{n} \sum_{i=1}^{n} \frac{\dot{g}'(X_i, \beta_0)}{\tau_0(\widetilde{Z}_i)} l(X_i, \widetilde{Z}_i) I(X_i \leq x) \sqrt{n}(\hat{\beta} - \beta_0),$$

it follows from (g1) that $B_{n2}(x) = -M_x \sqrt{n}(\hat{\beta} - \beta_0) + u_p(1)$. All other terms are $u_p(1)$ on the basis of (g2) and the routine arguments used to deal with $W_{n1}(x)$ earlier. Combining these results gives

$$\sup_{x \in \bar{\mathbb{R}}} \left| \hat{U}_n(x) - U_n(x) + M_x \cdot \sqrt{n}(\hat{\beta} - \beta_0) \right| = o_p(1). \tag{5.3}$$

Similarly, one can show that $\sup_{x \in \bar{\mathbb{R}}} \|\hat{M}_x - M_x\| = o_p(1)$. Note that M_x is positive definite by condition (m); hence for any $x_0 < \infty$,

$$\sup_{x \leq x_0} \|\hat{M}_x^{-1} - M_x^{-1}\| = o_p(1). \tag{5.4}$$

Finally, by the consistency of $\hat{a}_1$, $W_{n2}(x)$ can be written as the sum of three terms:

$$C_{n1}(x) = m a_1 \int_{s \leq x} \dot{g}'(s, \beta_0) M_s^{-1} U_n(s) \mathrm{d}F(s),$$

$$C_{n2}(x) = m a_1 \int_{s \leq x} \dot{g}'(s, \beta_0) M_s^{-1} [\hat{U}_n(s) - U_n(s)] \mathrm{d}F(s),$$

and a remainder term which can be shown to be $u_p(1)$. From (5.3), it follows that

$$C_{n2}(x) = -m a_1 E g(X, \theta_0) I(X \leq x) \cdot \sqrt{n}(\hat{\beta} - \beta_0).$$

(5.2), together with the above result, yields

$$\widehat{W}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i I(X_i \leq x) - m a_1 \int_{s \leq x} \dot{g}'(s, \beta_0) M_s^{-1} U_n(s) \mathrm{d}F(s) + u_p(1) = W_n(x) + u_p(1).$$

Using Lemma 5.2, we can obtain the desired result of Theorem 2.1. □

Remark 5.2. Note that in the proof above, the only place where we need to restrict the supremum to $[-\infty, x_0]$ is for establishing (5.4). If $\ell(x, z)$, and hence M_x, are free from the null parameters and depend on the other unknown entities in a simpler fashion, then one does not need this result and a suitable analog of $\widehat{W}_n$ will converge weakly to $B \circ F_X$ in $D([-\infty, \infty])$. Consider, for example, the case where $l(x, \tilde{z}) = c \cdot h(x, \tilde{z})$ is known completely except for the constant c which can be consistently estimated, and the CDFs of $F_X, F_{\widetilde{Z}}$ are known.

Proof of Theorem 3.1. Let b_i^a, $i = 1, 2, \ldots, n$, be random vectors, i.i.d. with mean 0 and covariance matrix V_1, and be independent of other random entities. Define $Y_{ij}^a = g(X_i, \beta_1) + Z_{ij}' b_i^a + \varepsilon_{ij}$ and $\widetilde{Y}_i^a = \sum_{j=1}^{m} Y_{ij}^a$; then

$$\hat{\xi}_i = \hat{\xi}_i^a + \frac{\widetilde{Y}_i - \widetilde{Y}_i^a}{\hat{\tau}(\widetilde{Z}_i)}, \quad \hat{\xi}_i^a = \frac{\widetilde{Y}_i^a - m g(X_i, \hat{\beta})}{\hat{\tau}(\widetilde{Z}_i)}.$$

Hence, the test statistic $\widehat{W}_n(x)$ can be decomposed into two terms:

$$\hat{W}_n^a(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\xi}_i^a\left[I(X_i \le x) - \frac{1}{n}\sum_{j=1}^{n}\hat{l}'(X_j,\widetilde{Z}_j)\widehat{M}_{X_j}^{-1}I(X_j \le x \wedge X_i)\hat{l}(X_i,\widetilde{Z}_i)\right],$$

$$\hat{R}_n^a(x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widetilde{Y}_i - \widetilde{Y}_i^a}{\hat{\tau}(\widetilde{Z}_i)}\left[I(X_i \le x) - \frac{1}{n}\sum_{j=1}^{n}\hat{l}'(X_j,\widetilde{Z}_j)\widehat{M}_{X_j}^{-1}I(X_j \le x \wedge X_i)\hat{l}(X_i,\widetilde{Z}_i)\right].$$

Using an argument similar to that for the null case, we can show that $W_n^a(x) \Longrightarrow B\circ\psi(x)$ in $D[-\infty, x_0]$ for any $x_0 < \infty$ and in uniform metric with $\psi(x) = P(X \le x)E\tau_0^2(\widetilde{Z})/\tau_1^2(\widetilde{Z})$.

Write $\widehat{R}_n^a(x)$ as $\widehat{R}_{n1}^a(x) - \widehat{R}_{n2}^a(x)$, where

$$\begin{aligned}\widehat{R}_{n1}^a(x) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widetilde{Y}_i - \widetilde{Y}_i^a}{\hat{\tau}(\widetilde{Z}_i)}I(X_i \le x)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{mg_1(X_i) + \widetilde{Z}_i'b_i - mg(X_i,\beta_1) - \widetilde{Z}_i'b_i^a}{\sqrt{m\hat{\sigma}^2 + \widetilde{Z}_i'\hat{V}\widetilde{Z}_i}}I(X_i \le x).\end{aligned}$$

Therefore, the law of large numbers and a Glivenko–Cantelli type argument imply

$$n^{-1/2}\widehat{R}_{n1}^a(x) \Longrightarrow D_1(x) + u_p(1). \tag{5.5}$$

Let

$$\widehat{V}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{(\widetilde{Y}_i - \widetilde{Y}_i^a)\hat{l}(X_i,\widetilde{Z}_i)}{\hat{\tau}(\widetilde{Z}_i)}I(X_i \ge x).$$

Then

$$n^{-1/2}\widehat{R}_{n2}^a(x) = m\hat{a}_1\int_{s\le x}\dot{g}'(s,\hat{\beta})\widehat{M}_s^{-1}\widehat{V}_n(s)\mathrm{d}\hat{F}_X(s).$$

Again, the law of large numbers and a Glivenko–Cantelli type argument imply $\hat{V}_n(x) = \rho(x) + u_p(1)$. Using the same argument as in the null case, one can verify that, under H_a, for any fixed $x_0 < \infty$, $\sup_{x\le x_0}\|\widehat{M}_x^{-1} - A_x^{-1}\| = o_p(1)$, where A_x is the same as M_x with β_0, σ_0^2, V_0 replaced by β_1, σ_1^2, and V_1, respectively. Therefore, we obtain $n^{-1/2}\widehat{R}_{n2}^a(x) = D_2(x) + u_p(1)$. This, together with (5.5), yields

$$\sup_{x\le x_0}|n^{-1/2}R_n^a(x) - [D_1(x) - D_2(x)]| = o_p(1). \tag{5.6}$$

Finally, the consistency of our test is derived by combining (5.6), the inequality

$$\sup_{x\le x_0}|\widehat{W}_n(x)| \ge \sqrt{n}\sup_{x\le x_0}|n^{-1/2}\widehat{R}_n^a(x)| - \sup_{x\le x_0}|\widehat{W}_n^a(x)|,$$

and the condition $d(x_0) = \sup_{x\le x_0}|D_1(x) - D_2(x)| > 0$. □

Proof of Theorem 3.2. The proof of this theorem is similar to that of Theorem 2.1 with obvious modifications. □

Acknowledgements

The authors are grateful to Dr. Hira L. Koul and the two anonymous referees for providing a constructive criticism that helped to improve the presentation of the work.

References

Billingsley, P., 1968. Convergence of Probability Measure. Wiley.
Brockwell, P.J., Davis, R.A., 2002. Introduction to Time Series and Forecasting. Springer.
Cook, D.R., Weisberg, S., 1994. An Introduction to Regression Graphics. Wiley.
Demidenko, E., 2004. Mixed Models: Theory and Applications. Wiley.
Diggle, P.J., Heagerty, P., Liang, K.Y., Zeger, S.L., 2002. Analysis of Longitudinal Data, 2nd ed. Oxford.
Jennrich, R.J., 1969. Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, 633–643.
Khmaladze, E.V., 1981. Martingale approach in the theory of goodness-of-fit tests. Theory Probab. Appl. 26, 240–257.
Khmaladze, E.V., 1988. An innovative approach to goodness-of-fit tests in R^m. Ann. Statist. 16, 1503–1516.
Khmaladze, E.V., 1993. Goodness of fit problem and scanning innovation martingales. Ann. Statist. 21, 798–829.
Khmaladze, E.V., Koul, H.L., 2004. Martingale transforms goodness-of-fit tests in regression models. Ann. Statist. 32, 995–1034.

Koul, H.L., 2006. Model diagnostics via martingale transforms: a brief review. In: Frontiers in Statistics. Imp. Coll. Press, London, pp. 183–206.
Koul, H.L., Song, W., 2009. Minimum distance regression model checking with Berkson error. Ann. Statist. 37, 132–156.
Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.
Pan, Z.Y., Lin, D.Y., 2005. Goodness-of-fit methods for generalized linear mixed models. Biometrics 61, 1000–1009.
Stute, W., Thies, S., Zhu, L.X., 1998. Model checks for regression: an innovation process approach. Ann. Statist. 26, 1916–1934.
White, H., 1981. Consequences and detection of misspecified nonlinear regression models. JASA 76, 419–433.
White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrics 50, 1–25.