



City Research Online

City, University of London Institutional Repository

Citation: González-Manteiga, W., Martínez-Miranda, M. D. & Van Keilegom, I. (2016). Goodness-of-fit test in parametric mixed effects models based on estimation of the error distribution. *Biometrika*, 103(1), pp. 133-146. doi: 10.1093/biomet/asv061

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/17075/>

Link to published version: <https://doi.org/10.1093/biomet/asv061>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Goodness-of-fit Test in Parametric Mixed Effects Models based on Estimation of the Error Distribution

BY WENCESLAO GONZÁLEZ-MANTEIGA

*Department of Statistics and Operations Research, University of Santiago de Compostela,
Faculty of Mathematics, 15782 Santiago de Compostela (Spain)*

wenceslao.gonzalez@usc.es

5

MARÍA DOLORES MARTÍNEZ-MIRANDA

*Department of Statistics and Operations Research, University of Granada
Campus Fuentenueva s/n, 18071 Granada (Spain)*

mmiranda@ugr.es

10

AND INGRID VAN KEILEGOM

*Université catholique de Louvain
Voie du Roman Pays, 20, B-1348 Louvain-la-Neuve (Belgium)*

ingrid.vankeilegom@uclouvain.be

SUMMARY

15

We address the problem of testing for a parametric function of fixed effects in mixed models. We propose a test based on the distance between two empirical error distribution functions, which are constructed from residuals calculated under the opposing hypotheses. The proposed test statistic has power against all alternatives, and its asymptotic distribution is derived. A simulation study shows that the test outperforms others in the literature. The test is applied to longitudinal data from an AIDS clinical trial and a growth study.

20

Some key words: Bootstrap; Empirical distribution; Residual; Local polynomial estimation; Mixed model.

1. INTRODUCTION

Mixed effects models assume a flexible covariance structure which allows for non-constant correlation among the observations, and have become very popular for many practical situations. A mixed effects model, or simply mixed model, contains both fixed and random effects. While the former describe the relationship between the covariates and the response for all the observations, the latter are specific to clusters or subjects within a population. This kind of model is suitable for problems related to, e.g., longitudinal data, repeated measurements, clustered data and small area estimation.

25

30

The most popular parametric mixed effects models are linear mixed models or generalized linear mixed models, which can be described as

$$g\{E(Y_{ij} | X_{ij}, b_i)\} = m(X_{ij}) + b_i^T Z_{ij}, \quad (1)$$

where, g is a known link function, Y_{ij} is the response variable, X_{ij} is a covariate vector of dimension d , Z_{ij} is a subvector of $(1, X_{ij}^T)^T$ of dimension d' , m represents the fixed effects and

b_i is a d' -dimensional vector of mean zero corresponding to the random effects. When g is the identity function and m is linear we have the linear mixed model. The nonlinear mixed model arises when the right-hand side in equation (1) is a nonlinear function of the fixed and random effects; see Pinheiro & Bates (2000).

The parametric assumption simplifies both theoretical and computational aspects, but it also provides valuable interpretations in real data applications. Therefore, it is of interest to test the adequacy of simple parametric mixed models. There are different approaches to test parametric assumptions for the function of the fixed effects, given by the function m in model (1), or for the distribution of the random effects, typically considered to be normally distributed.

Specification tests for the function m in model (1) are well-established in the literature in the absence of random effects. Different methods developed in the last twenty years (González-Manteiga & Crujeiras, 2013) can be classified into three groups: tests based on the comparison between nonparametric and parametric estimates; generalized likelihood ratio tests; and tests based on the empirical distribution of the residuals. Contributions for the case with random effects are more recent and scarce. Zhang & Lin (2003) consider a test for a semiparametric additive mixed model, where m is an additive function. In particular, when one additive component is linear and the other is nonparametric, they designed a goodness-of-fit test for polynomial regression in the nonparametric component. The authors assume clustered normal and non-normal data and base the test on nonparametric estimation by smoothing splines. Lombardía & Sperlich (2008) propose a test based on kernel smoothing to check a linear function for the nonparametric component of a generalized semiparametric additive model. See also Sperlich & Lombardía (2010), which is motivated by the small area estimation problem, and Henderson et al. (2008).

In the context of linear mixed models, or generalized linear mixed models, recent papers exploit the link between random effects and penalized regression, developing restricted likelihood ratio testing for zero variance components in linear mixed models. These methods, which are extensions of the F -test, have been applied to test whether the fixed effects are linear, quadratic, cubic, etc., in the presence of random effects, see Greven et al. (2008) and Wood (2013a,b). Huet & Kuhn (2015) suggest an omnibus test that exploits the ideas of the F -test but using a Bonferroni adjustment. Lin et al. (2002), Pan & Lin (2005) and Sánchez et al. (2009) provide omnibus tests for linear mixed models and generalized linear mixed models, based on cumulative sums of the residuals with respect to the covariates or the predicted values.

On the other hand, inference about the assumptions made for the random effects in model (1), for example the normal distribution of the b_i 's, has been considered recently in many papers. See Claeskens & Hart (2009) for an extensive review, or Meintanis & Portnoy (2011).

In this paper we propose a test based on the empirical distribution of the residuals. It extends the test of Van Keilegom et al. (2008), who consider a model without random effects representing cross-sectional independent data. This kind of method is very powerful, since it can detect alternatives at the parametric rate $n^{-1/2}$. To calibrate the distribution of the test statistics we suggest a bootstrap method suitable for the assumed mixed effects model.

2. MODEL AND ESTIMATION

In this paper we consider the semiparametric one-way model

$$Y_{ij} = m(X_{ij}) + b_i^T Z_{ij} + \epsilon_{ij} \quad (j = 1, \dots, n_i; i = 1, \dots, q), \quad (2)$$

where q is the number of levels in the model and $n = \sum_{i=1}^q n_i$ is the total number of observations. The covariate X_{ij} is a d -dimensional random vector, and Z_{ij} is a sub-vector of $(1, X_{ij}^T)^T$ of dimension d' . We assume that all X_{ij} 's are identically distributed with distribution

F_X and density f_X , and X_1, \dots, X_q are mutually independent, with $X_i = (X_{i1}, \dots, X_{in_i})^T$. We assume that the errors $\epsilon_{11}, \dots, \epsilon_{qn_q}$ are independent and identically distributed normal random variables with mean zero and variance σ^2 , and that $E(\epsilon_{ij} | X_{ij}) = 0$. The random effects b_1, \dots, b_q are independent and identically distributed d' -dimensional normal random variables with mean zero and covariance matrix V_b , which quantifies the within-subject variation. Further, assume that b_i and $X_{i'}$ are independent for $i, i' = 1, \dots, q$, so b_i is independent of X_i . Moreover, $\text{cov}(b_i, \epsilon_{i'j} | X_i, X_{i'}) = 0$ for all $i, i' = 1, \dots, q$ and $j = 1, \dots, n_{i'}$. The normal assumption made for the random effects and the errors could be relaxed (Severini & Staniswalis, 1994; Lin & Carroll, 2000). Here we make this assumption to develop simpler likelihood-based inferences for the function m , as well as for the estimation of the variances V_b and σ^2 .

Since the observations are only dependent if they come from the same individual, we can write (2) using matrix notation. Thus we first stack the observations at the individual level, i.e., $Y_i = m(X_i) + Z_i b_i + \epsilon_i$ ($i = 1, \dots, q$), where Z_i is the $n_i \times d'$ matrix with rows $Z_{i1}^T, \dots, Z_{in_i}^T$, $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$, and $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})^T$ has diagonal covariance matrix $\sigma^2 I_{n_i}$. Also, the variance of Y_i conditionally on X_i is

$$V_i = Z_i V_b Z_i^T + \sigma^2 I_{n_i}. \quad (3)$$

The model can be compactly written for the whole set of n observations as $Y = m(X) + Zb + \epsilon$, where $Y = (Y_1^T, \dots, Y_q^T)^T$, X is the $n \times d$ matrix with rows X_i^T , and Z is the $n \times qd'$ matrix with diagonal blocks Z_i . Here, the variance of Y conditionally on X is $V = ZBZ^T + \sigma^2 I_n$, where B is the matrix with diagonal blocks V_b .

Under a parametric mixed effects model the function of the fixed effects m is commonly estimated by the global likelihood method. Under the above conditions the density of Y_i , conditionally on X_i , is normal with mean $m(X_i) = \{m(X_{i1}), \dots, m(X_{in_i})\}^T$ and covariance matrix V_i . Then the log-likelihood of Y conditional on X is

$$\ell(m, V_b, \sigma^2) = -\frac{1}{2} \sum_{i=1}^q \left[\{Y_i - m(X_i)\}^T V_i^{-1} \{Y_i - m(X_i)\} + \log |V_i| + 2n_i \log(2\pi) \right]. \quad (4)$$

Here we are interested in estimating $m(x)$, for any fixed x in \mathcal{R}^d , by a local polynomial approach. We consider an extension of the common local polynomial estimator for independent and identically distributed data, derived using local likelihood. For a given x and supposing that X_{ij} is close to x and that m is continuous at x , we have that $m(X_{ij}) \approx m(x) = \beta_{x0}$ ($j = 1, \dots, n_i$). Model (2) can then be locally approximated by the mixed model $Y_i = P_i(\beta_x, x, p) + Z_i b_i + \epsilon_i$, where $P_i(\beta_x, x, p)$ is a n_i -dimensional vector of polynomials of order p containing all products of factors of the form $X_{ij,\ell} - x_\ell$ ($\ell = 1, \dots, d; j = 1, \dots, n_i$). The vector β_x consists of all coefficients of these polynomials, and its first component is $\beta_{x0} = m(x)$. The local log-likelihood can be defined from the global log-likelihood (4), by introducing local weights for each observation. However, in the presence of within-subject correlation in the model, this should be done using blocks, i.e., considering each of the q independent components of the global log-likelihood. As Lin & Carroll (2000) pointed out, the way to introduce the kernel weights into the individual components is problem-specific, and different ways provide estimators with different theoretical and practical properties; González-Manteiga et al. (2013) give a recent discussion on this topic. In this paper we follow the approximation of Park & Wu (2006), which is simpler and has good finite-sample properties. They define the local log-likelihood by

$$\ell_{\text{loc}}(\beta_x, V_b, \sigma^2) = -\frac{1}{2} \sum_{i=1}^q \left[\{Y_i - P_i(\beta_x, x, p)\}^T W_{ih,x}^{1/2} V_i^{-1} W_{ih,x}^{1/2} \{Y_i - P_i(\beta_x, x, p)\} \right. \\ \left. + \log |V_i| + 2n_i \log(2\pi) \right], \quad (5)$$

where the matrix $W_{ih,x}$ is diagonal with elements $K_h(X_{ij} - x)$ for each independent block ($i = 1, \dots, q$). Here, for $u = (u_1, \dots, u_d) \in \mathcal{R}^d$, $K(u) = \prod_{j=1}^d k(u_j)$ is a d -dimensional product kernel, k is a univariate kernel function, $h \equiv h_n$ is a bandwidth sequence converging to zero when n tends to infinity, and $K_h(u) = \prod_{j=1}^d \{k(u_j/h)/h\}$. The estimator $\hat{m}(x)$ is then defined as the first component of the vector that maximises $\ell_{\text{loc}}(\beta_x, V_b, \sigma^2)$ over β_x .

For the special case of local linear smoothing, i.e., when $p = 1$, the estimator $\hat{m}(x)$ can be explicitly written as the first component of the 2×1 vector $(\sum_{i=1}^q X_i^T W_{ih,x}^{1/2} V_i^{-1} W_{ih,x}^{1/2} X_i)^{-1} \sum_{i=1}^q X_i^T W_{ih,x}^{1/2} V_i^{-1} W_{ih,x}^{1/2} Y_i$, with X_i being the $n_i \times 2$ matrix with rows $(1, X_{ij} - x)^T$ ($j = 1, \dots, n_i$; $i = 1, \dots, q$). The local constant case, $p = 0$, is analogous but replacing the matrix X_i by the n_i -dimensional vector of ones, 1_{n_i} . If there is no within-subject correlation, or it is ignored, the derived estimator for $m(x)$ is the local linear estimator for independent data when $p = 1$, and the Nadaraya–Watson estimator when $p = 0$.

Since the estimator of $m(x)$ depends on V_b and σ^2 , which are unknown in general, the corresponding feasible or empirical estimator at each x can be derived using a three-step procedure. The procedure accomplishes the estimation of β and the variances V_b and σ^2 , using simultaneously the local and global log-likelihoods (5) and (4), and is formulated as follows:

Step 1. For arbitrary values of σ^2 and V_b define the estimator of $m(x)$ for any x by $\hat{m}_{\sigma, V_b}(x) = \hat{\beta}_{\sigma, V_b}$, which is the first component of the maximizer of $\ell_{\text{loc}}(\beta_x, V_b, \sigma^2)$. Calculate $\hat{m}_{\sigma, V_b}(X_{ij})$ for all observed X_{ij} 's.

Step 2. Compute the estimator of (V_b, σ^2) as the maximizer $(\hat{V}_b, \hat{\sigma}^2)$ of $\ell(\hat{m}_{\sigma, V_b}, V_b, \sigma^2)$, from the global log-likelihood $\ell(m, V_b, \sigma^2)$ given in (4).

Step 3. Finally, compute $\hat{m}(x) = \hat{\beta}_{\hat{\sigma}^2, \hat{V}_b}$.

From $\hat{m}(X_i) = \{\hat{m}(X_1), \dots, \hat{m}(X_q)\}^T$, the random effects b_i can be predicted using standard methods for linear mixed models by $\hat{b}_i = \hat{V}_b Z_i^T \hat{V}_i^{-1} \{Y_i - \hat{m}(X_i)\}$ ($i = 1, \dots, q$).

The above three-step procedure is suitable for model (2), where σ^2 and V_b are global parameters, and β is the only local parameter, but it can be easily adapted depending on which parameters in the model are global and which ones are local. For example, if the model is heteroscedastic, i.e., $\text{var}(\epsilon_{ij} | X_{ij}) = \sigma^2(X_{ij})$, then we could estimate $\sigma^2(\cdot)$ in the first step of the above procedure. Then we would maximize the local log-likelihood at each point x , with respect to β_x , $\sigma^2(x)$ and $\hat{V}_b(x)$. We could also consider that m belongs to a parametric class $\mathcal{M} = \{m_\theta : \theta \in \Theta\}$, in which case the estimation of m would move to the second step of the procedure. Then, we would maximize the global log-likelihood with respect to θ , σ^2 and V_b . Finally, note that our three-step procedure can be applied with any local log-likelihood and hence to any local estimator of m proposed in the literature. If the error is not supposed to be normal, steps 1 and 2 could be based on a quasi-likelihood approach and on general estimating equations, see, e.g., Liang & Zeger (1986) and Lin & Carroll (2000).

3. GOODNESS-OF-FIT TESTING

In this section we propose a test for the parametric null hypothesis about the function of the fixed effects m , formulated as follows:

$$H_0 : m \in \mathcal{M} = \{m_\theta : \theta \in \Theta\}, \quad H_1 : m \notin \mathcal{M}, \quad (6)$$

where Θ is a compact subset of \mathcal{R}^s . Let θ_0 be the true value of θ under H_0 , which is supposed to belong to the interior of Θ . Under H_1 , let us define $\bar{\theta}$ as the minimizer of $E[\{m(X) - m_\theta(X)\}^2]$ in Θ . Note that $\bar{\theta} = \theta_0$ under H_0 . The proposed test extends that of Van Keilegom et al. (2008) for cross-sectional independent data. The first step in defining it is to characterize the null hypothesis; see Theorem 2.1 in Van Keilegom et al. (2008). This result is based on the comparison between the error distributions under the null and the alternative hypotheses. When the regression model involves random effects, we can consider two approaches depending on how we define the errors: either the marginal errors, $U_{ij} = Y_{ij} - m(X_{ij})$, arising from the marginal distribution of the response Y_{ij} , or the conditional errors, $\epsilon_{ij} = Y_{ij} - m(X_{ij}) - b_i^T Z_{ij}$, arising from the whole regression structure, i.e., the distribution conditional on the random effects b_i and the covariates. Inference from conditional errors requires the estimation of both the random and the fixed effects, but the results of the tests can be affected by misspecification of the random component (Pan & Lin, 2005). The whole regression model is the target for testing procedures based on conditional residuals. However, our goal is to test the function of the fixed effects, so we suggest a test based on the marginal errors.

Consider the marginal errors $U_{ij} = Y_{ij} - m(X_{ij})$. Such errors are not independent and identically distributed, so the assumptions in Van Keilegom et al. (2008) do not apply. In order to remove the within-subject correlation, such errors have to be standardized. We consider the block transformation $V^{-1/2}U$, based on the whole vector $U = (U_1^T, \dots, U_q^T)^T$, with $U_i^T = (U_{i1}, \dots, U_{in_i})$. The elements of $V^{-1/2}U$ are then independent and identically distributed variables, so we can follow arguments like those in Van Keilegom et al. (2008) to formulate the test statistics. Let us denote the elements of the transformed vector of errors by U'_{ij} , and note that they all have the same distribution as a generic variable U' . Analogously, consider the transformed errors based on the parametric regression function under the null hypothesis, i.e., $V^{-1/2}\{Y - m_{\bar{\theta}}(X)\}$, with elements denoted by $U'_{ij,0}$, which are also independent and identically distributed variables with the same distribution as U'_0 , under H_0 . From these definitions the characterization of the null hypothesis in problem (6) follows by using arguments from Van Keilegom et al. (2008). This is formulated in the next proposition, whose proof is given in the Supplementary Material.

PROPOSITION 1. *Let m be a continuous function. The null hypothesis $H_0 : m \in \mathcal{M}$ in (6) is valid if and only if the standardized marginal errors U' and U'_0 have the same distribution.*

The next step is to estimate the distribution of the random variables U' and U'_0 , which we denote by $F_{U'}$ and $F_{U'_0}$; we assume that their corresponding densities exist and denote them by $f_{U'}$ and $f_{U'_0}$ respectively. We consider the estimators of m and the variance V resulting from the three-step estimation method presented in Section 2. Denote these estimators by \hat{m} and \hat{V} , where \hat{V} is the block diagonal matrix with blocks $\hat{V}_i = Z_i \hat{V}_b Z_i^T + \hat{\sigma}^2 I_{n_i}$. Then we can estimate the distribution of U' by the empirical distribution of the standardized nonparametric marginal residuals, $\hat{F}_{U'}(t) = n^{-1} \sum_{i=1}^q \sum_{j=1}^{n_i} I(\hat{U}'_{ij} \leq t)$. Here, the \hat{U}'_{ij} 's denote the elements of the vector $\hat{V}^{-1/2}\{Y - \hat{m}(X)\}$, and I is the indicator function. In the case of independent and identically distributed data, the asymptotic properties of the estimator have been studied by Akritas & Van

Keilegom (2001) for the special case where $d = 1$ and $p = 0$, and by Neumeyer & Van Keilegom (2010) for the general case. Also, the distribution of U'_0 is estimated by the empirical distribution of the parametric marginal residuals $\widehat{U}'_{ij,0}$, i.e., $\widehat{F}_{U'_0}(t) = n^{-1} \sum_{i=1}^q \sum_{j=1}^{n_i} I(\widehat{U}'_{ij,0} \leq t)$, where the $\widehat{U}'_{ij,0}$'s are defined as the elements of the vector $\widehat{V}^{-1/2}\{Y - \widehat{m}_0(X)\}$. The estimator \widehat{m}_0 is defined analogously to \widehat{m} , but replacing the observed responses Y_{ij} by the parametric estimator $m_{\widehat{\theta}}(X_{ij})$ of the fixed effects. Here the estimator $\widehat{\theta}$ can, e.g., be defined by maximizing the global likelihood $\ell(m_{\theta}, V_b, \sigma^2)$ with respect to θ , V_b and σ^2 , but other estimators are possible if assumptions (A7) and (A7*) below are satisfied.

Finally, we measure the distance between the empirical distributions $\widehat{F}_{U'}$ and $\widehat{F}_{U'_0}$, using Kolmogorov–Smirnov and Cramér–von Mises type statistics,

$$T_{n,\text{KS}} = n^{1/2} \sup_{-\infty < t < \infty} \left| \widehat{F}_{U'}(t) - \widehat{F}_{U'_0}(t) \right|, \quad T_{n,\text{CM}} = n \int \left\{ \widehat{F}_{U'}(t) - \widehat{F}_{U'_0}(t) \right\}^2 d\widehat{F}_{U'_0}(t).$$

To study the local power of these statistics, we consider the local alternatives

$$H_{1n} : m(\cdot) = m_{\theta_0}(\cdot) + n^{-1/2}r(\cdot)$$

for some bounded function r . These alternatives only concern the regression function and not the error distribution. The main asymptotic result of this paper provides the asymptotic distribution of the two test statistics. The result, given in Theorem 1 below, is formulated under H_{1n} , but it also covers H_0 by taking the function r equal to zero. We need the following assumptions:

- (A1) the number of blocks, q , tends to infinity, and $n_i \leq C$ ($i = 1, \dots, q$), for some $C < \infty$;
- (A2) $nh^{2p+2} \rightarrow 0$ if p is odd, $nh^{2p+4} \rightarrow 0$ if p is even, and $nh^{3d+\nu} \rightarrow \infty$ for some small $\nu > 0$;
- (A3) k is a symmetric probability density function supported on $[-1, 1]$, k is d -times continuously differentiable, and $k^{(j)}(\pm 1) = 0$ ($j = 0, \dots, d-1$);
- (A4) all partial derivatives of F_X up to order $2d+1$ exist on the interior of the compact support R_X of X , they are uniformly continuous, and $\inf_{x \in R_X} f_X(x) > 0$;
- (A5) all partial derivatives of $x \rightarrow m(x)$ up to order $p+2$ exist on the interior of R_X , and they are uniformly continuous;
- (A6) all partial derivatives of $(x, \theta) \rightarrow m_{\theta}(x)$ up to order 2 exist on the interior of $R_X \times \Theta$, and they are continuous in (x, θ) ; and
- (A7) the estimator $\widehat{\theta}$ can be written as $\widehat{\theta} - \theta_0 = n^{-1} \sum_{i=1}^q \sum_{j=1}^{n_i} \xi(X_{ij}, Y_{ij}) + n^{-1/2}\delta + o_P(n^{-1/2})$, where ξ satisfies $E\{\xi(X_{ij}, Y_{ij}) \mid X_{ij}\} = 0$ both under H_0 and H_{1n} . Moreover, the asymptotic distribution of $n^{-1/2} \sum_{i=1}^q \sum_{j=1}^{n_i} \xi(X_{ij}, Y_{ij})$ under H_0 is the same as under H_{1n} , and the constant δ depends on the direction of the alternative hypothesis determined by the function r , and equals zero under H_0 .

Assumption (A1) is common in the context of mixed effects models. In the context of longitudinal data it states that the number of individuals increases but the number of observations for each individual is bounded. Assumptions (A2)–(A5) come from Neumeyer & Van Keilegom (2010) and are required to obtain the asymptotic distribution of the process $n^{1/2}\{\widehat{F}_{U'}(\cdot) - F_{U'}(\cdot)\}$. Assumption (A6) is necessary for applying the asymptotic results in Van Keilegom et al. (2008). Finally, (A7) is needed to decompose the process $n^{1/2}\{\widehat{F}_{U'_0}(\cdot) - F_{U'}(\cdot)\}$ into a sum of independent and identically distributed terms and negligible terms, from which the weak convergence of this process will follow. See also Pan & Lin (2005), formula (2), and Theorem 3.1.2 in the 1994 Wisconsin-Madison University PhD thesis by J. C. Pinheiro, for similar decompositions, and for precise conditions under which this assumption holds true.

We are now ready to state the main result describing the limiting distribution of the test statistics $T_{n,\text{KS}}$ and $T_{n,\text{CM}}$. The proof for $p = 0$ and $p = 1$ is given in the Supplementary Material.

THEOREM 1. *Assume that conditions (A1)–(A7) are satisfied. Then, under H_{1n} ,*

$$T_{n,\text{KS}} \rightarrow \sup_{-\infty < t < \infty} |f_{U'}(t)| |W - a|, \quad T_{n,\text{CM}} \rightarrow \int f_{U'}^2(t) dF_{U'}(t) (W - a)^2$$

in distribution, where W is a zero-mean normal random variable with variance given in equation (5) in the Supplementary Material, and where the constant a depends on the direction of the alternative, defined also in the Supplementary Material, equation (4). Note that $a = 0$ under H_0 .

Since the above limiting distributions are rather complicated, we suggest using bootstrap methods to approximate the critical values. More precisely, we define a bootstrap algorithm suitable for the assumed mixed model:

1. Calculate the estimator \hat{m} of the function of the fixed effects m , the estimators \hat{V}_b and $\hat{\sigma}^2$ of the variances V_b and σ^2 , and also the estimator $m_{\hat{\theta}}$ of the parametric regression function under H_0 . These estimators are derived using the three-step method in Section 2.
2. Generate bootstrap conditional errors ϵ_{ij}^* independently from a normal distribution with mean zero and variance $\hat{\sigma}^2$, and bootstrap random effects b_i^* from a d' -dimensional normal distribution with mean zero and covariance matrix \hat{V}_b .
3. Under the null hypothesis the bootstrap responses are constructed by $Y_{ij}^* = m_{\hat{\theta}}(X_{ij}) + (b_i^*)^T Z_{ij} + \epsilon_{ij}^*$ ($j = 1, \dots, n_i$; $i = 1, \dots, q$). Then, the bootstrap sample is given by $\{(X_{ij}, Z_{ij}, Y_{ij}^*), j = 1, \dots, n_i, i = 1, \dots, q\}$.
4. Calculate the bootstrapped test statistics $T_{n,\text{KS}}^*$ and $T_{n,\text{CM}}^*$ from the bootstrap sample generated in the previous step.

Finally, the quantiles of the distribution of $T_{n,\text{KS}}^*$ and $T_{n,\text{CM}}^*$ can be approximated by repeating steps 2–4 in the bootstrap algorithm B times.

The resampling scheme could also be defined without using the normal assumption. In that case both the conditional residuals and the random effects in the second step above could be generated from the smoothed empirical distribution of the residuals (Van Keilegom et al., 2008).

The consistency of this bootstrap procedure is shown in the next result and the proof is given in the Supplementary Material. For this, we need to introduce the bootstrap counterpart of (A7):

(A7*) under H_0 , the estimator $\hat{\theta}^*$ can be written as $\hat{\theta}^* - \hat{\theta} = n^{-1} \sum_{i=1}^q \sum_{j=1}^{n_i} \xi^*(X_{ij}, Y_{ij}^*) + o_{P^*}(n^{-1/2})$, in probability, where ξ^* satisfies $E\{\xi^*(X_{ij}, Y_{ij}^*) \mid X_{ij}\} = 0$ and where

$$\sup_t \left| \text{pr}^* \left\{ n^{-1/2} \sum_{i=1}^q \sum_{j=1}^{n_i} \xi^*(X_{ij}, Y_{ij}^*) \leq t \right\} - \text{pr} \left\{ n^{-1/2} \sum_{i=1}^q \sum_{j=1}^{n_i} \xi(X_{ij}, Y_{ij}) \leq t \right\} \right| \rightarrow 0,$$

in probability.

THEOREM 2. *Assume that conditions (A1)–(A7) and (A7*) are satisfied. Then, under H_0 ,*

$$\sup_s \left| \text{pr}^*(T_{n,\text{KS}}^* \leq s) - \text{pr}(T_{n,\text{KS}} \leq s) \right| \rightarrow 0, \quad \sup_s \left| \text{pr}^*(T_{n,\text{CM}}^* \leq s) - \text{pr}(T_{n,\text{CM}} \leq s) \right| \rightarrow 0,$$

in probability, where the probability pr^ is computed under the bootstrap distribution conditional on the original data (X_{ij}, Z_{ij}, Y_{ij}) ($j = 1, \dots, n_i$; $i = 1, \dots, q$).*

4. SIMULATION EXPERIMENTS

We simulate two different models: a simple model with just a random intercept,

$$Y_{ij} = m(X_{ij}) + b_i + \epsilon_{ij}, \quad (7)$$

and a model with random effects consisting of a random intercept and a random slope,

$$Y_{ij} = m(X_{ij}) + b_{i0} + b_{i1}X_{ij} + \epsilon_{ij}. \quad (8)$$

Both are particular cases of (2), with $Z_{ij} = 1$ for (7) and $Z_{ij} = (1, X_{ij})^T$ for (8). In both cases we consider X_{ij} to be scalar and generated from either a uniform distribution on $[0, 2]$ or a normal distribution with mean zero and variance 0.6. The random effects and the errors ϵ_{ij} are generated independently. The random effects b_1, \dots, b_q in (7) are generated from a normal distribution with mean zero and standard deviation σ_{b_0} . We consider the values $\sigma_{b_0} = 0.6$, case 1, and $\sigma_{b_0} = 1$, case 2. The two-dimensional random vector $(b_{i0}, b_{i1})^T$ ($i = 1, \dots, q$), in model (8), is bivariate normal with covariance matrix $V_b = \text{diag}(0.3^2, 0.3^2)$. The errors are generated from a normal distribution with mean zero and standard deviation $\sigma = 0.3$. We consider samples of sizes $n = 150, 300$ and 600 , where the number of observations per group is always $n_i = 3$, and the number of groups equals $q = 50, 100$ and 200 . We test whether the function m is linear using the test statistics $T_{n,\text{KS}}$ and $T_{n,\text{CM}}$.

The test involves a kernel estimator depending on three choices: the bandwidth parameter, h , the degree of the polynomial, p , and the kernel function k . The asymptotic analysis shows that the bandwidth h should satisfy the conditions described in assumption (A2). We have considered bandwidths of the type $h = h_0 n^{-3/10}$, which satisfy this assumption, where h_0 is a constant value chosen around the range of the covariate X . This recommendation is similar to that of Pardo-Fernández et al. (2007), who suggested the same kind of test in a different regression framework. In the Supplementary Material we analyse the sensitivity of the test to the bandwidth choice using several values for h_0 around the range of the simulated covariate values. The conclusion is that the test is quite robust to this choice so here we only report the case $h = 3n^{-3/10}$. Regarding the degree of the polynomial, it is known that the local linear estimator, $p = 1$, has better properties than the local constant estimator, $p = 0$; see for example Wand & Jones (1995). The effect of this choice is not major though; see the Supplementary Material. In this section, and in our other empirical analyses, we only consider the local linear case. Finally, it is well-known that the choice of the kernel function k does not have a major impact on the performance of the kernel estimator (Wand & Jones, 1995), and therefore on our test. In our empirical studies we consider the Epanechnikov kernel. We work with these choices of h , p and k in the rest of this section and derive the kernel estimator using the three-step estimation method presented in Section 2. We have approximated the critical values in the test using the bootstrap algorithm described in Section 3 with $B = 1000$ bootstrap samples. The bootstrapped test statistics have been calculated using the choices of h , p and k above.

We consider two other tests that can deal with the formulated problem: the omnibus test of Pan & Lin (2005), which competes with our test if the aim is to test a linear mixed model, is based on the cumulative marginal residuals and has critical values obtained using an asymptotic approximation valid for large values of q ; and the restricted likelihood ratio test of Greven et al. (2008), which is not omnibus in the sense that a single test is performed to detect deviations from a null hypothesis. One can expect that the restricted likelihood ratio test, which is an extension of the F -test, performs better than an omnibus test if the null hypothesis is linear. We have calculated this test using the function *exactRLRT* in the R-package *RLRsim* (Scheipl et al., 2008).

We compare the empirical level and power of the tests. Let us consider first model (7). To calculate the level we simulated 1000 samples under the null hypothesis, defined in this case

by $m(X) = 1 + X$. Here, the considered nominal level is $\alpha = 0.05$. The power was calculated by simulating also 1000 samples from two specific alternatives. The first consists of contaminating the null hypothesis with a sinusoidal function, by simulating $m_1(X) = 1 + (1 - a)X + a \sin(\pi X)$, with $a = 0.1$ and 0.2 . The second is harder to detect and allows us to check the power against quadratic terms by simulating $m_2(X) = 1 + (1 - a)X + aX^2$ for $a = 0.1$ and 0.2 . In both cases, the value $a = 0$ corresponds to the null hypothesis of linearity.

Table 1 shows the results obtained from each test under this scenario and considering only the normal design, which is also the most favourable design for the omnibus test of Pan & Lin (2005). We considered a randomized rule (Pearson, 1950) to determine the rejection levels for $T_{n,KS}$, which is discrete. This consists of deciding the rejection of the null hypothesis based on a random experiment when the test statistic equals the approximated critical value. In our case we define the function $\phi_\alpha(s) = \{\alpha B - \#(T_{n,KS}^* > s)\} / \#(T_{n,KS}^* = s)$, generate a uniform number $u \in (0, 1)$ and reject the null hypothesis if $u < \phi_\alpha(T_{n,KS})$, or accept it otherwise. Here the notation $\#(S)$ represents the cardinality of the set S . All the tests have similar empirical level for all sample sizes, and the average p-value of the test of Greven et al. (2008) is, in all cases, much higher than the expected value of 50%.

The power of the tests for the two alternatives are also shown in Table 1. Our test clearly outperforms that of Pan & Lin (2005). The restricted likelihood ratio test has the highest power, as expected since it incorporates model information. Taking this into account, we can conclude that our tests, in particular the Cramér–von Mises test, have good power.

Table 1. Empirical size and power (%) of tests under two types of alternatives for model (7). Under the null hypothesis, $m(X) = 1 + X$, the average p-value (%) is shown between brackets. The nominal level is 5%.

		Case 1				Case 2			
H_0	n	CM	KS	PL	RLRT	CM	KS	PL	RLRT
m	150	4.2	5.9	4.0	4.6	4.0	4.8	4.3	5
		(51)	(51)	(46)	(72)	(53)	(53)	(46)	(72)
	300	4.1	4.8	4.4	5.4	5.1	4.0	4.6	5.4
		(51)	(51)	(48)	(72)	(41)	(51)	(48)	(72)
	600	5.1	4.3	4.1	4.8	5.2	5.0	4.2	4.7
		(50)	(50)	(49)	(71)	(42)	(51)	(49)	(72)
H_1	n	CM	KS	PL	RLRT	CM	KS	PL	RLRT
m_1 $a = 0.1$	150	26.1	18.5	5.4	28.9	21.6	14.8	4.5	28.0
	300	49.1	33.3	12.3	54.2	48.8	34.1	7.0	53.9
	600	82.8	60.3	18.9	92.1	81.4	56.4	9.1	90.8
m_1 $a = 0.2$	150	75.5	54.7	16.1	90.5	71.8	49.8	6.9	89.6
	300	96.9	85.4	36.1	99.9	96.8	81.3	15.0	99.9
	600	100	97.7	68.8	100	99.6	98.0	27.2	100
m_2 $a = 0.1$	150	32.2	20.1	7.3	73.4	29.5	18.9	5.1	72.4
	300	55.0	36.5	11.2	96.2	51.6	31.3	7.4	96.2
	600	84.3	55.7	24.0	100	81.0	58.4	9.6	100
m_2 $a = 0.2$	150	81.1	59.4	19.0	99.8	75.4	54.8	8.0	99.6
	300	96.2	82.9	42.9	100	94.4	78.8	15.9	100
	600	100	97.0	82.1	100	99.5	95.6	34.0	100

CM, Cramér–von Mises; KS, Kolmogorov–Smirnov; PL, Pan & Lin (2005); RLRT, Greven et al. (2008); $m_1(X)$, sinusoidal; $m_2(X)$, quadratic.

Table 2. *Empirical size and power (%) of tests under four types of alternatives for model (8). Under the null hypothesis, $m(X) = 1 + X$, the average p-value (%) is shown between brackets. The nominal level is 5%.*

H_0	n	Uniform			Normal		
		CM	KS	RLRT	CM	KS	RLRT
m	150	4.5	4.2	5.0	4.8	4.3	6.5
		(52)	(51)	(74)	(51)	(51)	(71)
	300	4.6	5.3	5.1	5.5	5.1	4.4
		(48)	(48)	(73)	(49)	(50)	(71)
	600	5.7	5.4	5.0	3.9	3.8	5.2
		(51)	(51)	(73)	(49)	(49)	(73)
H_1	n	CM	KS	RLRT	CM	KS	RLRT
m_1 $a = 0.2$	150	17.9	14.3	35.9	24.8	21.4	58.2
	300	33.9	24.8	70.0	53.5	47.4	91.6
	600	49.9	43.3	95.4	81.9	73.2	99.6
m_2 $a = 0.2$	150	9.5	11.8	23.6	31.2	33.2	66.7
	300	16.8	19.1	47.8	65.0	66.7	93.1
	600	27.0	29.5	77.7	90.5	91.0	99.6
m_3 $a = 0.2$	150	24.1	17.5	38.2	74.2	60.3	90.5
	300	34.9	27.2	66.7	93.6	85.5	99.6
	600	49.6	42.2	93.7	99.7	98.8	100
m_4 $a = 0.2$	150	9.3	11.3	20.9	27.0	22.0	39.0
	300	12.1	13.3	42.2	43.7	37.6	68.4
	600	18.9	20.4	72.8	64.0	58.6	94.2

CM, Cramér-von Mises; KS, Kolmogorov-Smirnov; RLRT, Greven et al. (2008); $m_1(X)$, sinusoidal; $m_2(X)$, quadratic; $m_3(X)$, absolute value; $m_4(X)$, discontinuous.

To finish this section, we consider model (8). The power was calculated from four alternatives: the same $m_1(X)$ and $m_2(X)$ considered above with $a = 0.2$, and non-smooth alternatives defined by $m_3(X) = 1 - a|0.5 - X|$ and $m_4(X) = 1 - aX I(X \leq 0.5) + a I(X > 0.5)$, also with $a = 0.2$. Table 2 shows that for model (8) the size of our test is around the nominal level of 5% and the test provides reasonable power. We have not considered the test of Pan & Lin (2005) since it has been clearly beaten by our test in the simpler model (7). As expected, the test of Greven et al. (2008) exhibits the highest power but provides an average p-value, under the null hypothesis, much higher than 50%.

5. APPLICATION TO AIDS CLINICAL TRIAL

Our first application consists of CD4 counts data from an AIDS clinical trial to evaluate the efficacy of Zidovudine in treating patients with mild symptomatic HIV infection. These data have also been analysed by Lin et al. (2002) and Pan & Lin (2005). A total of 711 patients enrolled in the study, with 361 randomized to Zidovudine and 350 to placebo. Here we only consider the patients treated with Zidovudine. Experts on this type of data suggest that the CD4 counts for these patients tend to rise for the first few weeks and then decline over time (Lin et al., 2002). Hence, it seems reasonable to describe the time trend with polynomials of degree two or three. The profile plot for these patients is shown in Figure 1.

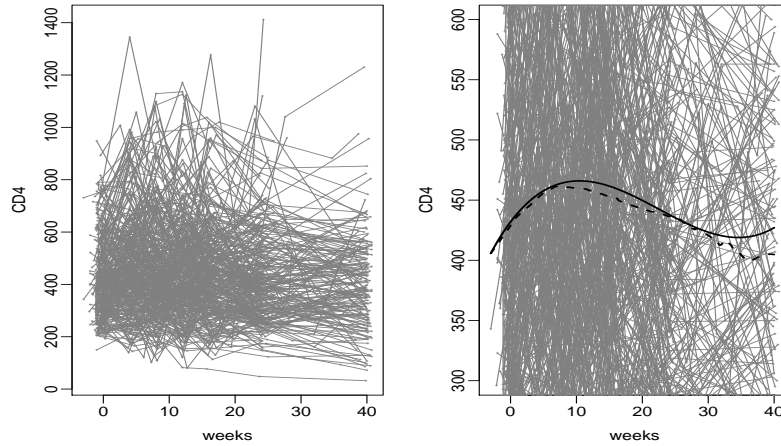


Fig. 1. CD4 count data. Observed individual profiles (gray lines) for patients treated with Zidovudine. The estimated function of the fixed effects using the local linear kernel estimator is shown by a dashed curve using a bandwidth of 8 weeks, and a cubic parametric estimator is plotted by a solid curve.

From these plots it is difficult to extract any useful information, because the individual CD4 cell counts are quite noisy. However, the nonparametric estimator proposed in Section 2 is able to capture the underlying structure in the data with $p = 1$. This local linear kernel estimator is shown by the black solid curve in Figure 1. We have calculated this estimator by assuming model (2) with the response, Y_{ij} , being the CD4 cell counts, and with the covariate, X_{ij} , being the time in weeks. We chose a bandwidth of 8 weeks by eye. This choice considers the variability within the data, and it is sufficient to provide a first visual impression about the underlying function. However, González-Manteiga et al. (2013) describe an automatic data-driven bandwidth selector for this type of kernel estimator.

To choose the covariance structure in model (2) we considered two candidates: a simple model with only a random intercept, that is, with $Z_{ij} = 1$; and a more complex model with both random intercept and slope, $Z_{ij} = (1, X_{ij})^T$. The second was used to calculate the local linear estimator plotted in Figure 1. These models can be written in the forms (7) and (8), respectively. To decide which model is more appropriate, we calculated the AIC considering a quadratic polynomial for the function of the fixed effects. The AIC values are 27344 and 27359, suggesting that the model with just a random intercept describes the random variations better. Therefore in the following we work under model (7).

The local linear estimator in Figure 1 shows that the underlying time trend in the CD4 cell counts could be modelled by a quadratic or a cubic polynomial. However the impression from this graph depends on the degree of smoothness considered in the kernel estimator. We therefore consider the tests proposed in this paper to decide between these two parametric models. First we consider a quadratic polynomial as the null hypothesis. The resulting p-values are 1.2% using $T_{n,KS}$, and 0.1% using $T_{n,CM}$. These statistics were calculated using the local linear estimator and bandwidth parameter $h = h_0 n^{-3/10}$, with h_0 equal to the range of the covariate. With the same type of kernel estimator and bandwidth choice, we now consider a cubic polynomial as the null hypothesis. In this case, $T_{n,KS}$ and $T_{n,CM}$ provide p-values equal to 16.9% and 14.1%,

respectively. The tests confirm that the time trend in the CD4 cell counts, for the patients treated with Zidovudine, can be described by a cubic polynomial.

6. TESTING NONLINEAR FIXED EFFECTS ASSUMING A MORE GENERAL MODEL

We now consider a model motivated by a data application to growth studies with longitudinal data. Our objective is to provide a suitable extension of the methods proposed in Sections 2 and 3. The data analysis itself and a brief simulation experiment show the practicability of the tests and their good performance in these settings. Further research is still necessary to derive the theoretical properties of the test under the new model.

Our motivating data are the orange tree dataset described by Draper & Smith (1998). The data arise from an experiment in which trunk circumference in millimeter was measured for $q = 5$ orange trees on $n_i = 7$ different occasions, over roughly a 4-year period of growth defined at $(x_{i1}, \dots, x_{in_i}) = (118, 484, 664, 1004, 1231, 1372, 1582)$ days for each tree. The interest in growth studies lies, among others, in characterizing the average growth pattern in the population. Thus testing whether a parametric function is appropriate for a particular growth study is of interest. Figure 2 shows the profile plot. Previous studies suggest that the marginal mean can be described by a logistic model

$$E(Y | X) = \beta_1 [1 + \exp \{-(X - \beta_2)/\beta_3\}]^{-1}, \quad (9)$$

which represents many common growth patterns (Draper & Smith, 1998). Neither the test of Pan & Lin (2005) nor that of Greven et al. (2008) can be used to check the suitability of this model, but our test can be easily extended to it.

The methods in Sections 2 and 3 were described under the semiparametric model (2), which assumes that the function of the random effects is linear, inducing the marginal covariance structure given in (3). In order to apply the bootstrap method proposed in Section 3, the structure of the random effects needs to be specified. Serroyen et al. (2009) consider the mean structure model (9) for this dataset and suggest different models to describe the covariance structure. Among several candidates, the following nonlinear mixed effects model, also suggested by Pinheiro & Bates (2000), provides a suitable representation of the underlying structure:

$$Y_{ij} = (\beta_1 + b_i) [1 + \exp \{-(X_{ij} - \beta_2)/\beta_3\}]^{-1} + \epsilon_{ij}. \quad (10)$$

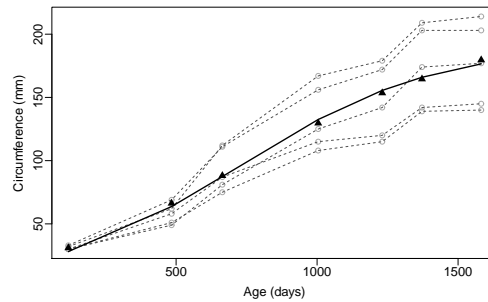


Fig. 2. Orange tree dataset. The dashed curves show the observed individual profiles. The parametric logistic estimator (solid curve) and the local linear estimator (triangle-points) of the marginal mean are also shown.

Here, b_i are independent and identically distributed normal variables with mean zero and variance τ^2 , and ϵ_{ij} are independent and identically distributed normal variables with mean zero and variance σ^2 ($j = 1, \dots, n_i; i = 1, \dots, q$). The marginal mean from this model is indeed the logistic model (9), and it induces the covariance structure

$$V_i = \text{var}(Y_i | X_i) = \tau^2 s_i s_i^T + \sigma^2 I_{n_i}, \quad (11)$$

with $s_i = ([1 + \exp\{-(X_{i1} - \beta_2)/\beta_3\}]^{-1}, \dots, [1 + \exp\{-(X_{in_i} - \beta_2)/\beta_3\}]^{-1})^T$.

We propose the following extension of our model:

$$Y_{ij} = m(X_{ij}) + v_i^\xi(Z_{ij}) + \epsilon_{ij} \quad (j = 1, \dots, n_i; i = 1, \dots, q), \quad (12)$$

where Y_{ij} , X_{ij} , Z_{ij} , m and ϵ_{ij} are specified exactly as before, but now the random effects function $v_i^\xi(Z_{ij})$ can be considered as a realization of a zero-mean nonlinear process depending on a parameter vector ξ , with covariance function $\gamma_\xi(Z_{ij_1}, Z_{ij_2}) = E\{v_i^\xi(Z_{ij_1})v_i^\xi(Z_{ij_2})\}$. Assume also that $v_i^\xi(Z_{ij})$ is independent of ϵ_{ij} , conditionally on X_{ij} . The nonlinear mixed model defined in (10) is a particular case of the semiparametric model (12), where m is the logistic function defined in (9), $v_i^\xi(Z_{ij}) = b_i [1 + \exp\{-(X_{ij} - \beta_2)/\beta_3\}]^{-1}$, and the b_i are independent and identically distributed normal variables with mean zero and variance τ^2 .

Under model (12) our tests can be calculated in a straightforward way. The parametric and nonparametric estimators of m can be derived using the three-step method of Section 2. Here it is necessary to specify the marginal covariance V_i involved in the global and local log-likelihoods, (4) and (5), respectively. For the orange trees dataset we consider the structure V_i defined in (11). Figure 2 shows two estimators of m : the parametric logistic estimator and the local linear estimator with bandwidth $h = 500$ days. This figure suggests the adequacy of the logistic model. To confirm this impression we perform the tests proposed in Section 3, and find that the p-values for $T_{n,KS}$ and $T_{n,CM}$ are 58.7% and 55.7%, respectively. To derive these p-values we have considered a modification of the bootstrap algorithm, given in the Supplementary Material. The finite sample performance of the tests described above is investigated in the Supplementary Material under a scenario which represents the performed data analysis.

ACKNOWLEDGEMENT

The authors thank Dr. Sánchez-Sellero for helpful discussions about computational issues, Dr. Pan for supplying the AIDS dataset, and the Universidad de Granada for providing the computing time. This research was supported by the European Community's Seventh Framework Programme, by the Spanish Ministry of Economy and Competitiveness and by the Belgian government.

SUPPLEMENTARY MATERIAL

Supplementary Material available at *Biometrika* online includes the proofs of Proposition 1 and Theorems 1 and 2 in Section 3, and additional simulations and details for Sections 4 and 6.

REFERENCES

- AKRITAS, M. G. & VAN KEILEGOM, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Stat.* **28**, 549–567.
- CLAESKENS, G. & HART, J. D. (2009). Goodness-of-fit tests in mixed models. *Test* **18**(2), 213–239.
- DRAPER, N. R. & SMITH, H. (1998). *Applied Regression Analysis* 3rd ed. New York: Wiley.

GONZÁLEZ-MANTEIGA, W. & CRUJEIRAS, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test* **22**, 361–411.

GONZÁLEZ-MANTEIGA, W., LOMBARDÍA, M. J., MARTÍNEZ-MIRANDA, M. D. & SPERLICH, S. (2013). Kernel smoothers and bootstrapping for semiparametric mixed effects models. *J. Multivar. Anal.* **114**, 288–302.

GREVEN, S., CRAINICEANU, C., KUECHENHOFF, H. & PETERS, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *J. Comp. Graph. Stat.* **17**, 870–891.

HENDERSON, D. J., CARROLL, R. J. & LI, Q. (2008). Nonparametric estimation and testing of fixed effects panel data models. *J. Econometrics* **144**, 257–275.

HUET, S. & KUHN, E. (2015). Goodness-of-fit test for Gaussian regression with block correlated errors. *Statistics* **49**, 239–266.

LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

LIN, X. & CARROLL, R. J. (2000). Nonparametric function estimation for clustered data when predictor is measured without/with error. *J. Am. Statist. Assoc.* **95**, 520–534.

LIN, D. Y., WEI, L. J. & YING, Z. (2002). Model-checking techniques based on cumulative residuals. *Biometrics* **58**, 1–12.

LOMBARDÍA, M. J. & SPERLICH, S. (2008). Semiparametric inference in generalized mixed effects models. *J. R. Statist. Soc. B* **70**, 913–930.

MEINTANIS, S. G. & PORTNOY, S. (2011). Specification tests in mixed effects models. *J. Stat. Plan. Inference* **141**, 2545–2555.

NEUMEYER, N. & VAN KEILEGOM, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multiv. Anal.* **101**, 1067–1078.

PAN, Z. & LIN, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics* **61**, 1000–1009.

PARDO-FERNÁNDEZ, J. C., VAN KEILEGOM, I. & GONZÁLEZ-MANTEIGA, W. (2007). Testing for the equality of k regression curves. *Stat. Sin.* **17**, 1115–1137.

PARK, J. G. & WU, H. (2006). Backfitting and local likelihood methods for nonparametric mixed-effects models with longitudinal data. *J. Stat. Plan. Inference* **136**, 3760–3782.

PEARSON, E. S. (1950). On questions raised by the combination of tests based on discontinuous distributions. *Biometrika* **37**, 383–398.

PINHEIRO, J. C. & BATES, D. M. (2000). *Mixed-Effects Models in S and S-Plus*, Springer, New York.

SÁNCHEZ, B. N., HOUSEMAN, E. A. & RYAN, L. M. (2009). Residual-based diagnostic for structural equation models. *Biometrics* **65**, 104–115.

SCHEIPL, F., GREVEN, S. & KUECHENHOFF, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comp. Stat. Dat. Anal.* **52**, 3283–3299.

SERROYEN, J., MOLENBERGHS, G., VERBEKE, G. & DAVIDIAN, M. (2009). Non-linear models for longitudinal data. *Am. Stat.* **63**, 378–388.

SEVERINI, T. A. & STANISWALIS, J. G. (1994). Quasilikelihood estimation in semiparametric models. *J. Am. Statist. Assoc.* **89**, 501–511.

SPERLICH, S. & LOMBARDÍA, M. J. (2010). Local polynomial inference for small area statistics: estimation, validation and prediction. *J. Nonparametr. Stat.* **22**, 633–648.

VAN KEILEGOM, I., GONZÁLEZ-MANTEIGA, W. & SÁNCHEZ-SELLERO, C. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17**, 401–415.

WAND, M. P. & JONES, M. C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

WOOD, S. N. (2013a). A simple test for random effects in regression models. *Biometrika* **100**, 1005–1010.

WOOD, S. N. (2013b). On p-values for smooth components of an extended generalized additive model. *Biometrika* **100**, 221–228.

ZHANG, D. & LIN, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4**, 57–74.

[Received . Revised]