Assessing skewness, kurtosis and normality in linear mixed models

Alexandra Soberón[a], Winfried Stute[b]

[a]*Departamento de Economía, Universidad de Cantabria, Avenida Los Castros s/n, E-39005 Santander, Spain*
[b]*Mathematical Institute, University of Giessen, Arndtstr. 2, D-35392 Giessen, Germany*

Abstract

Linear mixed models provide a useful tool to fit continuous longitudinal data, with the random effects and error term commonly assumed to have normal distributions. However, this restrictive assumption can result in a lack of robustness and needs to be tested. In this paper, we propose tests for skewness, kurtosis, and normality based on generalized least squares (GLS) residuals. To do it, estimating higher order moments is necessary and an alternative estimation procedure is developed. Compared to other procedures in the literature, our approach provides a closed form expression even for the third and fourth order moments. In addition, no further distributional assumptions on either random effects or error terms are needed to show the consistency of the proposed estimators and tests statistics. Their finite-sample performance is examined in a Monte Carlo study and the methodology is used to examine changes in the life expectancy as well as maternal and infant mortality rate of a sample of OECD countries.

Keywords: Kurtosis, linear mixed model, longitudinal data, moment estimator, normality, skewness.

1. Introduction

Linear mixed models are often used to study intra-group correlation patterns. They have attracted considerable attention, e.g., in biomedical, social and economic sciences. For mathematical tractability, a common assumption in these models is that the random effects and the error terms are normally distributed. When this is the case, it is well known that maximum likelihood estimation (MLE) and restricted maximum likelihood estimation (RMLE) perform quite well; see, e.g., [2, 5, 6]. However, conclusions derived under these restrictive assumptions may not be robust to departures from Gaussianity, especially when data show multimodality and skewness; see, e.g., [16].

The impact of misspecification in the random effects distribution has been extensively investigated in the literature; see, e.g., [9, 10, 15, 16]. However, there seems to be no general consensus about its effect and the proposed alternatives are restricted to specific mixed models. Formal tests to detect mixture distributions in the random effects have also been proposed. For example, [7] provides a goodness-of-fit test for both random effects and error terms, but the statistic does not have an exact χ^2 distribution and it is a bit cumbersome to implement in practice. More recently, [11] proposed to check the normality of the random effects using gradient functions.

Therefore, it is of practical interest to develop a simple test that enables to check the normal distribution assumption of both random effects and error terms. This is not an easy task, however, because in linear mixed models the lack of Gaussianity can arise in more than one component of the regression error. The identification of which component is causing the departure from normality is then crucial. This paper is concerned with the efficient estimation and testing of linear mixed models when standard distributional assumptions such as normality or symmetry cannot be justified. Specifically, some very simple and intuitive tests to detect departures from normality in the form of skewness and/or kurtosis are proposed based on moment conditions, for which estimators of the higher order moments are necessary.

Note that despite the great importance of higher order moments for statistical inference, there are few references in the literature that define and study estimators for higher than second order moments. Two examples are [3, 8] but the methods they advocate exhibit some weaknesses: the first is not easily extended to multivariate problems, and the second is not valid if symmetry fails in practice. To overcome this situation, [14] developed an alternative method that

provides consistent estimators of higher order moments. However, their technique is somewhat problematic for the third and fourth moments. In these cases there are several choices of estimating equations that can provide consistent estimators, so finding efficient ones becomes a critical issue.

Thus, the aim of this paper is twofold: (i) to develop a new approach leading directly to efficient estimators of the higher order moments of the error term, thereby solving the efficiency issue in [14]; (ii) to propose tests for detecting departures from normality for both random effects and error terms, based on a proposal made by [4] in the context of longitudinal models. The method proposed here has the following promising features. First, it results in closed form expressions for the estimators of the higher order moments, including the third and fourth order moment. Second, using generalized least squares (GLS) residuals, no further distributional assumptions on either random effects or error terms are needed to show the consistency of these estimators. Third, the proposed tests enable one to identify departures from normality in the form of skewness and/or kurtosis of each component of the regression error, jointly or separately. Fourth, the estimators of the higher order moments allow one to study distributional properties of the estimators through their Fourier transform. To the best of our knowledge, all these results are original. For methodological reasons, we restrict our attention to the linear mixed model, but the underlying ideas are applicable to handle non-linear, semi-parametric or nonparametric models as well.

Finally, in order to illustrate the feasibility and possible gains of the proposed method, a Monte Carlo study is conducted to assess the finite-sample performance of our proposed estimators and test statistics. As a concrete example, an empirical study based on data from the Organization for Economic Co-operation and Development (OECD) is carried out to measure and assess the importance of various factors determining two commonly accepted measures of health care outcomes: life expectancy and infant mortality.

The rest of the article is organized as follows. In Section 2, we introduce the linear mixed model and describe the estimation method. In Section 3, the corresponding asymptotic properties are studied. In Section 4, we derive some tests to detect departures from normality in the form of skewness or kurtosis, and we study their asymptotic properties. Section 5 contains some simulation results and an empirical application to illustrate the usefulness of the method. Section 6 presents our main conclusions. All proofs are collected in the Appendix.

2. Statistical model and estimation procedure

Assume that data are available from a linear mixed model of the form

$$y_{ij} = \alpha + x_{ij}^{\top}\beta + b_i + \epsilon_{ij}, \tag{1}$$

where $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell_i\}$ denote the group index and the measurements within this group, respectively, ℓ_i is the sample size within group i, and n is the number of subjects. Also, y_{ij} is the response variable corresponding to the jth observation in the ith group, while x_{ij} is a $p \times 1$ vector of covariates. Further, the relation between x_{ij} and y_{ij} described in Eq. (1) contains an intercept parameter α, a $p \times 1$ fixed effect parameter vector β, and some random effects b_i, all of which are unknown. All these quantities are also perturbed by random errors ϵ_{ij}. Throughout this paper, it is assumed that ϵ_{ij} and b_i are independent and identically distributed (iid), and ϵ_{ij} is independent of all b_i and x_{ij} for all i and j. Without loss of generality, we further assume that the expectations of the random effects and errors are zero so that β is identifiable. Otherwise, unknown nonzero expectations can be incorporated in the intercept.

For each $k \in \{2, \ldots, 8\}$, let $\gamma_b^k = \mathrm{E}(b_i^k)$ and $\gamma_\epsilon^k = \mathrm{E}(\epsilon_{ij}^k)$ be the kth moments of the unobserved random effects and the idiosyncratic errors, respectively. This paper is concerned with the efficient estimation and testing of these unknown terms. Since these estimators are based on a suitable combination of the GLS residuals of (1), in this section we first obtain the estimators of the parameters of interest, β and α, and later we focus on the estimation of γ_b^k and γ_ϵ^k.

In all our developments, we are unwilling to impose any condition on the statistical relationship between b_i and the covariates of the model. Rewriting the regression model (1) in vectorial form, we have

$$y_i = \iota_{\ell_i}\alpha + x_i\beta + v_i, \quad v_i = \iota_{\ell_i}b_i + \epsilon_i, \tag{2}$$

where for any $i \in \{1, \ldots, n\}$, v_i is a composed error term, $y_i = (y_{i1}, \ldots, y_{i\ell_i})^{\top}$, $v_i = (v_{i1}, \ldots, v_{i\ell_i})^{\top}$, and $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{i\ell_i})^{\top}$ are ℓ_i-dimensional vectors for subject i, $x_i = (x_{i1}, \ldots, x_{i\ell_i})^{\top}$ is an $\ell_i \times p$ dimensional matrix, and ι_{ℓ_i} is a unitary vector of length ℓ_i.

In order to obtain consistent estimators for β, a standard solution is to sweep out b_i by the $\ell_i \times \ell_i$ idempotent transformation matrix $Q_{\ell_i} = I_{\ell_i} - \iota_{\ell_i}(\iota_{\ell_i}^\top \iota_{\ell_i})^{-1}\iota_{\ell_i}^\top$, leading, for each $i \in \{1, \ldots, n\}$, to

$$Q_{\ell_i} y_i = Q_{\ell_i} x_i \beta + Q_{\ell_i} \epsilon_i, \tag{3}$$

where I_{ℓ_i} is a $\ell_i \times \ell_i$ identity matrix.

By regressing $Q_{\ell_i} y_i$ on $Q_{\ell_i} x_i$, the ordinary least squares (OLS) estimators for β are obtained as in [14]. However, as pointed out by [13], the OLS estimators are the best linear unbiased estimators (BLUEs) under model (3), but not under the original model (2) since they ignore the correlation within units. As it is well known, for all j, the variance-covariance matrix of the composed error term v_i of Eq. (2) is of the form

$$V_{\ell_i} = \mathrm{E}(v_i v_i^\top) = \gamma_\epsilon^2 I_{\ell_i} + \gamma_b^2 \iota_{\ell_i} \iota_{\ell_i}^\top.$$

Let $P_{\ell_i} = I_{\ell_i} - Q_{\ell_i}$ be a $\ell_i \times \ell_i$ matrix and $\lambda = 1 - \{\gamma_\epsilon^2/(\gamma_\epsilon^2 + \ell_i \gamma_b^2)\}^{1/2}$. Following [12], the square root of the inverse matrix of V_{ℓ_i} is $V_{\ell_i}^{-1/2} = \gamma_\epsilon^{-1} \Omega_{\ell_i}^{-1/2}$, where $\Omega_{\ell_i} = I_{\ell_i} - \lambda P_{\ell_i}$ is a matrix of size $\ell_i \times \ell_i$.

With the aim of obtaining spherical disturbances, we premultiply (2) by $V_{\ell_i}^{-1/2}$, which leads to the expression

$$V_{\ell_i}^{-1/2} y_i = V_{\ell_i}^{-1/2} \iota_{\ell_i} \alpha + V_{\ell_i}^{-1/2} x_i \beta + V_{\ell_i}^{-1/2} v_i. \tag{4}$$

Note that the resulting estimator from (4) is infeasible since $V_{\ell_i}^{-1/2}$ depends on some unknown terms, i.e., γ_ϵ^2 and γ_b^2. Therefore, in order to get a feasible OLS estimator of β, this covariance matrix must be estimated. Based on the spectral decomposition of Ω_{ℓ_i}, we propose to replace the unknown covariance components by their consistent estimators, i.e., $\widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$, that are computed using the methods proposed in the following subsection. Then, assuming for a moment that $\widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$ are given, the generalized least squares (GLS) estimators have the form

$$\widehat{\beta} = \left(\sum_{i=1}^n \ddot{x}_i^\top \widehat{\Omega}_{\ell_i}^{-1} \ddot{x}_i \right)^{-1} \sum_{i=1}^n \ddot{x}_i^\top \widehat{\Omega}_{\ell_i}^{-1} \ddot{y}_i, \tag{5}$$

$$\widehat{\alpha} = \frac{1}{n} \sum_{i=1}^n \overline{y}_{i\bullet} - \frac{1}{n} \sum_{i=1}^n \overline{x}_{i\bullet}^\top \widehat{\beta},$$

where $\widehat{\Omega}_{\ell_i}^{-1/2} = I_{\ell_i} - \widehat{\lambda} P_{\ell_i}$, $\widehat{\lambda} = 1 - \{\widehat{\gamma}_\epsilon^2/(\widehat{\gamma}_\epsilon^2 + \ell_i \widehat{\gamma}_b^2)\}^{1/2}$, $\ddot{x}_i = x_i - \iota_{\ell_i} \overline{x}$ is an $\ell_i \times p$ matrix and $\ddot{y}_i = y_i - \iota_{\ell_i} \overline{y}$ is an ℓ_i-dimensional vector defined in terms of

$$\overline{x} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} x_{ij} \qquad \text{and} \qquad \overline{y} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} y_{ij},$$

where $N = \ell_1 + \cdots + \ell_n$ is the overall sample size,

$$\overline{x}_{i\bullet} = \frac{1}{\ell_i} \sum_{j=1}^{\ell_i} x_{ij} \qquad \text{and} \qquad \overline{y}_{i\bullet} = \frac{1}{\ell_i} \sum_{j=1}^{\ell_i} y_{ij}.$$

As stated in Section 3, under weak conditions on the group sizes and the design variables, $\widehat{\beta}$ is the BLUE estimator for β under the true model. Therefore, efficient estimators for the kth order moments of the errors can be proposed using the GLS residuals.

Since we aim at considering much more general situations than those covered by [3] and, at the same time, we want to address the efficiency issue of [14], a variant of the fundamental lemma presented by the latter is proposed. In particular, instead of looking for a suitable combination of the polynomial functions of $v_{ij} = b_i + \epsilon_{ij}$ as they propose, this alternative approach is based on the mean deviation error term of the form $\widetilde{\epsilon}_{ij} = \epsilon_{ij} - \overline{\epsilon}_{i\bullet}$, where $\overline{\epsilon}_{i\bullet} = (\epsilon_{i1} + \cdots + \epsilon_{i\ell_i})/\ell_i$.

In order to illustrate the proposed technique, we define the following nonlinear function for each $m \in \{1, \ldots, k\}$:

$$f_m^k(i) = \sum_{j=1}^{\ell_i} \widetilde{\epsilon}_{ij}^m \left(\sum_{j=1}^{\ell_i} \widetilde{\epsilon}_{ij} \right)^{k-m}.$$

The following lemma can be used to derive estimating equations for $\gamma_\epsilon^k = \mathrm{E}(\epsilon_{ij}^k)$.

Lemma 1. *For any $a, b \in \mathbb{R}$, write $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. We have*

$$f_m^k(i) = \sum_{t=0}^{k} \sum_{s=(t-k+m)\vee 0}^{t \wedge m} \binom{m}{s}\binom{k-m}{t-s}\left(\sum_{j=1}^{\ell_i} \epsilon_{ij}^s\right)\left(\sum_{j=1}^{\ell_i} \epsilon_{ij}\right)^{t-s}(-\bar{\epsilon}_{i\bullet})^{k-t}\ell_i^{k-m-t+s}.$$

This lemma is slightly different from the one obtained in [14], but it still maintains its usefulness since many of the terms in the expansion of $f_m^k(i)$ will vanish when we take expectations. In addition, it is very useful from an empirical point of view. The fundamental lemma presented in [14] provides several polynomial functions of the error term, the $f_m^k(i)$'s, that have to be combined in a proper way in order to obtain efficient estimates for γ_ϵ^k. The transformation proposed here addresses this efficiency issue since it enables one to obtain unique $f_m^k(i)$'s that provide efficient estimators for each γ_ϵ^k. Also, note that this proposal can be easily extended to longitudinal, semi-parametric and nonparametric models with additive structure.

2.1. Estimation of second order moments

Focus first on the estimation of the second order moments of the random effects and error terms. From Lemma 1, we obtain the following two polynomial functions of the error

$$f_2^2(i) = \sum_{j=1}^{\ell_i} \epsilon_{ij}^2 - 2\bar{\epsilon}_{i\bullet}\sum_{j=1}^{\ell_i}\epsilon_{ij} + \ell_i\bar{\epsilon}_{i\bullet}^2, \quad f_1^2(i) = \ell_i^2\bar{\epsilon}_{i\bullet}^2 - 2\ell_i\bar{\epsilon}_{i\bullet}\sum_{j=1}^{\ell_i}\epsilon_{ij} + \left(\sum_{j=1}^{\ell_i}\epsilon_{ij}\right)^2.$$

Upon rearranging terms, we find $f_1^2(i) = 0$ and $\mathrm{E}\{f_2^2(i)\} = (\ell_i - 1)\gamma_\epsilon^2$. Averaging over all $i \in \{1, \ldots, n\}$ and replacing the unknown $\widetilde{\epsilon}_{ij}$ by the residuals $\widehat{\widetilde{\epsilon}}_{ij}$, we see that the estimator of γ_ϵ^2 is of the form

$$\widehat{\gamma}_\epsilon^2 = \frac{1}{\sum_{i=1}^n \ell_i}\sum_{i=1}^n \frac{\ell_i}{(\ell_i - 1)}\sum_{j=1}^{\ell_i}\widehat{\widetilde{\epsilon}}_{ij}^2,$$

where $\widehat{\widetilde{\epsilon}}_{ij} = (y_{ij} - \bar{y}_{i\bullet}) - (x_{ij} - \bar{x}_{i\bullet})^\top\widehat{\beta}$.

In addition, since $\mathrm{E}(v_{ij}v_{ih}) = \gamma_b^2$, when $j \neq h$, and $\mathrm{E}(v_{ij}^2) = \gamma_b^2 + \gamma_\epsilon^2$, the estimator for γ_b^2 can be written as

$$\widehat{\gamma}_b^2 = \sum_{i=1}^n \frac{2}{n\ell_i(\ell_i - 1)}\sum_{j=1}^{\ell_i - 1}\sum_{h=j+1}^{\ell_i}\widehat{v}_{ij}\widehat{v}_{ih}, \tag{6}$$

where $\widehat{v}_{ij} = y_{ij} - \widehat{\alpha} - x_{ij}^\top\widehat{\beta}$.

2.2. Estimation of third order moments

In this section we consider the estimation of higher order moments of the random error terms with minimal variance. As before, using Lemma 1 for the third order moments of the idiosyncratic error term, we obtain

$$\begin{aligned}
f_3^3(i) &= \sum_{j=1}^{\ell_i}\epsilon_{ij}^3 - 3\bar{\epsilon}_{i\bullet}\sum_{j=1}^{\ell_i}\epsilon_{ij}^2 + 3\bar{\epsilon}_{i\bullet}^2\sum_{j=1}^{\ell_i}\epsilon_{ij} - \ell_i\bar{\epsilon}_{i\bullet}^3, \\
f_2^3(i) &= \left(\sum_{j=1}^{\ell_i}\epsilon_{ij}^2\right)\left(\sum_{j=1}^{\ell_i}\epsilon_{ij}\right) - 2\bar{\epsilon}_{i\bullet}\left(\sum_{j=1}^{\ell_i}\epsilon_{ij}\right)^2 - \ell_i\bar{\epsilon}_{i\bullet}\sum_{j=1}^{\ell_i}\epsilon_{ij}^2 + 3\ell_i\bar{\epsilon}_{i\bullet}^2\sum_{j=1}^{\ell_i}\epsilon_{ij} - \ell_i^2\bar{\epsilon}_{i\bullet}^3, \\
f_1^3(i) &= \left(\sum_{j=1}^{\ell_i}\epsilon_{ij}\right)^3 - 3\ell_i\bar{\epsilon}_{i\bullet}\left(\sum_{j=1}^{\ell_i}\epsilon_{ij}\right)^2 + 3\ell_i^2\bar{\epsilon}_{i\bullet}^2\sum_{j=1}^{\ell_i}\epsilon_{ij} - \ell_i^3\bar{\epsilon}_{i\bullet}^3.
\end{aligned}$$

Rearranging terms, we find $f_2^3(i) = 0$ and $f_1^3(i) = 0$. Then, unlike what happens in [14], Lemma 1 provides a unique polynomial function of the error term that leads to an estimating equation from which the estimator for γ_ϵ^3 can be directly obtained. Taking expectations on $f_3^3(i)$, we get

$$E\{f_3^3(i)\} = \frac{(\ell_i - 1)(\ell_i - 2)}{\ell_i} \gamma_\epsilon^3$$

from which the estimator for γ_ϵ^3 has the form

$$\widehat{\gamma}_\epsilon^3 = \frac{1}{\sum_{i=1}^n \ell_i} \sum_{i=1}^n \frac{\ell_i^2}{(\ell_i - 1)(\ell_i - 2)} \sum_{j=1}^{\ell_i} \widetilde{\widehat{\epsilon}}_{ij}^3.$$

For the estimation of γ_b^3, note that $E(v_{ij} v_{ih} v_{i\ell}) = \gamma_b^3$ when $j \neq h$ and $h \neq \ell$. In this situation, the estimator for γ_b^3 can be written as

$$\widehat{\gamma}_b^3 = \sum_{i=1}^n \frac{6}{n\ell_i(\ell_i - 1)(\ell_i - 2)} \sum_{j=1}^{\ell_i-2} \sum_{h=j+1}^{\ell_i-1} \sum_{g=h+1}^{\ell_i} \widehat{v}_{ij} \widehat{v}_{ih} \widehat{v}_{ig}.$$

2.3. Estimation of fourth order moments

Finally, for the estimator of the fourth order moments of the error terms, we get from Lemma 1 that

$$f_4^4(i) \quad = \quad \sum_{j=1}^{\ell_i} \epsilon_{ij}^4 - 4\bar{\epsilon}_{i\bullet} \sum_{j=1}^{\ell_i} \epsilon_{ij}^3 + 6\bar{\epsilon}_{i\bullet}^2 \sum_{j=1}^{\ell_i} \epsilon_{ij}^2 - 4\bar{\epsilon}_{i\bullet}^3 \sum_{j=1}^{\ell_i} \epsilon_{ij} - \ell_i \bar{\epsilon}_{i\bullet}^4,$$

with $f_3^4(i) = 0$, $f_2^4(i) = 0$ and $f_1^4(i) = 0$, after rearranging terms.

Lemma 1 then yields a unique expression for the estimation of the fourth order moment of the idiosyncratic error term. Taking expectations on $f_4^4(i)$, we find

$$E\{f_4^4(i)\} = \left(\frac{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3}{\ell_i^2} \right) \gamma_\epsilon^4 + \left(\frac{6\ell_i^2 - 15\ell_i + 9}{\ell_i^2} \right) (\gamma_\epsilon^2)^2$$

so that the corresponding estimator has the form

$$\widehat{\gamma}_\epsilon^4 = \frac{1}{\sum_{i=1}^n \ell_i(\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3)} \sum_{i=1}^n \left\{ \ell_i^3 \sum_{j=1}^{\ell_i} \widetilde{\widehat{\epsilon}}_{ij}^4 - \ell_i(\ell_i - 1)(6\ell_i - 9)(\widehat{\gamma}_\epsilon^2)^2 \right\}. \tag{7}$$

Finally, for the fourth order moment estimator of the random effects, we propose

$$\widehat{\gamma}_b^4 = \sum_{i=1}^n \frac{2}{n\ell_i(\ell_i - 1)} \sum_{j=1}^{\ell_i-1} \sum_{h=j+1}^{\ell_i} \widehat{v}_{ij}^2 \widehat{v}_{ih}^2 - 2\widehat{\gamma}_b^2 \widehat{\gamma}_\epsilon^2 - (\widehat{\gamma}_\epsilon^2)^2,$$

where $\widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$ are the consistent estimators for γ_ϵ^2 and γ_b^2, respectively, obtained previously.

3. Asymptotic properties

To study the asymptotic properties of the proposed GLS estimators, we first give a set of regularity conditions.

Assumption 1. *The random variables $b_1, \ldots, b_n$ and $\epsilon_{1t}, \ldots, \epsilon_{nt}$ are iid with mean zero and variances γ_b^2 and γ_ϵ^2, respectively. Moreover, b_i and ϵ_{it} are independent for all $i \in \{1, \ldots, n\}$.*

Assumption 2. *Let $N = \ell_1 + \cdots + \ell_n$. As $n \to \infty$, $\sum_{i=1}^n (1/\ell_i)/n \to 0$ and $N/n \to \infty$.*

Assumption 3. *Let $\Sigma_n = (1/\sum_{i=1}^{n} \ell_i) \sum_{i=1}^{n} \ddot{x}_i^{\top} \Omega_{\ell_i}^{-1} \ddot{x}_i$ be a $p \times p$ matrix, $\lim_{n\to\infty} \Sigma_n = \Sigma$ for some positive definite $p \times p$ matrix Σ.*

Assumption 4. *Let $\overline{x}_{i\bullet} = (1/\ell_i) \sum_{j=1}^{\ell_i} x_{ij}$ be bounded covariates, $\max_{i,j} \|x_{ij} - \lambda \overline{x}_{i\bullet}\|/ \sqrt{N} \to 0$.*

Theorem 1. *Under Assumptions 1–4, as $n \to \infty$, $\sqrt{N} (\widehat{\beta} - \beta) \rightsquigarrow \mathcal{N}(0, \gamma_\epsilon^2 \Sigma^{-1})$ and $\sqrt{n} (\widehat{\alpha} - \alpha) \rightsquigarrow \mathcal{N}(0, \gamma_b^2)$.*

Comparing the results of this theorem with the corresponding in Theorem 2.1 in [14], it can be emphasized that a BLUE estimator for the true model can be obtained taking quasi-temporary differences instead of using a mean deviation transformation as those one in (3). More precisely, if $\widehat{\gamma}_b^2$ is relatively smaller than $\widehat{\gamma}_\epsilon^2$, it results that $\widehat{\lambda} \neq 1$ and the OLS and GLS estimators will be very different. In fact, $\widehat{\beta}_{GLS}$ will be more efficient than OLS since its variance-covariance matrix will be smaller. In the opposite case, when $\widehat{\lambda} \simeq 1$, both estimation procedures will provide similar results.

In what follows we focus on the properties of $\widehat{\gamma}_\epsilon^k$ and $\widehat{\gamma}_b^k$, for $k \in \{2, 3, 4\}$.

Theorem 2. *Under Assumptions 1–4, when γ_ϵ^4 and γ_b^4 are finite, as $n \to \infty$, $\sqrt{N} (\widehat{\gamma}_\epsilon^2 - \gamma_\epsilon^2) \rightsquigarrow \mathcal{N}(0, \mu_\epsilon^2)$ and $\sqrt{n} (\widehat{\gamma}_b^2 - \gamma_b^2) \rightsquigarrow \mathcal{N}(0, \mu_b^2)$, where $\mu_\epsilon^2 = \gamma_\epsilon^4 - (\gamma_\epsilon^2)^2$ and $\mu_b^2 = \gamma_b^4 - (\gamma_b^2)^2$.*

Theorem 2 reveals that although the estimators that we propose here for γ_ϵ^2 and γ_b^2 are a bit different from those obtained in [14], their asymptotic properties are exactly the same. This result will be corroborated in the simulations. In addition, in Appendix A it is shown that $\widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$ have iid representations, i.e.,

$$\sqrt{N} (\widehat{\gamma}_\epsilon^2 - \gamma_\epsilon^2) = \frac{1}{\sqrt{N}} \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} (\epsilon_{ij}^2 - \gamma_\epsilon^2) + o_P(1)$$

and

$$\sqrt{n} (\widehat{\gamma}_b^2 - \gamma_b^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (b_i^2 - \gamma_b^2) + o_P(1).$$

In the following theorems, the asymptotic normality of the estimators for the third and fourth order moments of the error term is established.

Theorem 3. *Under Assumptions 1–4, when γ_ϵ^6 and γ_b^6 are finite, as $n \to \infty$, $\sqrt{N} (\widehat{\gamma}_\epsilon^3 - \gamma_\epsilon^3) \rightsquigarrow \mathcal{N}(0, \mu_\epsilon^3)$ and $\sqrt{n} (\widehat{\gamma}_b^3 - \gamma_b^3) \rightsquigarrow \mathcal{N}(0, \mu_b^3)$, where $\mu_\epsilon^3 = \gamma_\epsilon^6 - (\gamma_\epsilon^3)^2 + 9(\gamma_\epsilon^2)^3 - 6\gamma_\epsilon^2 \gamma_\epsilon^4$ and $\mu_b^3 = \gamma_b^6 - (\gamma_b^3)^2 + 9(\gamma_b^2)^3 - 6\gamma_b^2 \gamma_b^4$.*

Theorem 4. *Under Assumptions 1–4, when γ_ϵ^8 and γ_b^8 are finite, as $n \to \infty$, $\sqrt{N} (\widehat{\gamma}_\epsilon^4 - \gamma_\epsilon^4) \rightsquigarrow \mathcal{N}(0, \mu_\epsilon^4)$ and $\sqrt{n} (\widehat{\gamma}_b^4 - \gamma_b^4) \rightsquigarrow \mathcal{N}(0, \mu_b^4)$, where $\mu_\epsilon^4 = \gamma_\epsilon^8 - (\gamma_\epsilon^4)^2 + 16(\gamma_\epsilon^3)^2 \gamma_\epsilon^2 - 8\gamma_\epsilon^3 \gamma_\epsilon^5$ and $\mu_b^4 = \gamma_b^8 - (\gamma_b^4)^2 - 8\gamma_b^3 \gamma_b^5 + 16(\gamma_b^3)^2 \gamma_b^2$.*

Based on the results of Theorems 3 and 4 we can conclude that once again the proposed estimators for $\gamma_\epsilon^3, \gamma_b^3, \gamma_\epsilon^4$, and γ_b^4 are a bit different from those obtained in [14], but their asymptotic results are similar. In addition, the proposed estimators achieve the corresponding minimum variance and they are as efficient as the moment estimators based on the true unknown terms ϵ_{ij} and b_i.

To sum up, it can be pointed out that the alternative lemma proposed in this paper effectively provides a unique estimating equation for all higher order moments of the error term even when standard distributional assumptions are violated. Thus, this method enables one to solve the efficiency issue discussed in [14]. In principle, one could go further beyond the fourth moment. For practical issues, however, when one wants to study distributional properties of our estimators through their Fourier transform, an expansion involving the first few moments is enough.

4. Test statistics

With the aim of detecting departures from normality of either random effects or error terms, jointly and separately, in this section we propose tests for skewness, kurtosis and normality based on the previous higher order estimators. First, we focus on the skewness and kurtosis tests and derive their limiting distribution under arbitrary skewness and kurtosis coefficients. Later, we extend the results to the specific normal setting. Finally, we focus on a normality test and develop its limiting properties.

4.1. Tests for skewness and kurtosis

When we are interested in testing for skewness in both random effects and error terms, the quantities of interest for each component are

$$s_b = \mathrm{E}(b^3)/\{\mathrm{E}(b^2)\}^{3/2} = \gamma_b^3/(\gamma_b^2)^{3/2}, \quad s_\epsilon = \mathrm{E}(\epsilon^3)/\{\mathrm{E}(\epsilon^2)\}^{3/2} = \gamma_\epsilon^3/(\gamma_\epsilon^2)^{3/2},$$

whereas for kurtosis the corresponding quantities are

$$\kappa_b = \mathrm{E}(b^4)/\{\mathrm{E}(b^2)\}^{4/2} = \gamma_b^4/(\gamma_b^2)^2, \quad \kappa_\epsilon = \mathrm{E}(\epsilon^4)/\{\mathrm{E}(\epsilon^2)\}^{4/2} = \gamma_\epsilon^4/(\gamma_\epsilon^2)^2.$$

Let $\widehat{\gamma}_b^k$ and $\widehat{\gamma}_\epsilon^k$ be the consistent estimators obtained previously for γ_b^k and γ_ϵ^k, respectively, when $k \in \{2, 3, 4\}$. We propose the following statistics to construct tests for skewness

$$\widehat{s}_b = \widehat{\gamma}_b^3/(\widehat{\gamma}_b^2)^{3/2}, \quad \widehat{s}_\epsilon = \widehat{\gamma}_\epsilon^3/(\widehat{\gamma}_\epsilon^2)^{3/2}, \tag{8}$$

whereas to test for kurtosis the proposed statistics are of the form

$$\widehat{\kappa}_b = \widehat{\gamma}_b^4/(\widehat{\gamma}_b^2)^2, \quad \widehat{\kappa}_\epsilon = \widehat{\gamma}_\epsilon^4/(\widehat{\gamma}_\epsilon^2)^2. \tag{9}$$

In the following, we focus on the limiting distribution of these statistics under any conjectured coefficients of skewness and kurtosis.

Theorem 5. *Under Assumptions 1–4, when γ_b^6 and γ_ϵ^6 are finite, as $n \to \infty$, $\sqrt{N}\,(\widehat{s}_\epsilon - s_\epsilon) \rightsquigarrow \mathcal{N}[0, \alpha_\epsilon^\top \Gamma_\epsilon \alpha_\epsilon/(\gamma_\epsilon^2)^3]$ and $\sqrt{n}\,(\widehat{s}_b - s_b) \rightsquigarrow \mathcal{N}[0, \alpha_b^\top \Gamma_b \alpha_b/(\gamma_b^2)^3]$, where $\alpha_\epsilon = (1, -\tfrac{3}{2}\gamma_\epsilon^2, -3\gamma_\epsilon s_\epsilon/2)^\top$ and $\alpha_b = (1, -\tfrac{3}{2}\gamma_b^2, -3\gamma_b s_b/2)^\top$ are vectors of dimension 1×3, while Γ_ϵ and Γ_b are 3×3 matrices of the form*

$$\Gamma_\epsilon = \begin{bmatrix} \gamma_\epsilon^6 - (\gamma_\epsilon^3)^2 & \gamma_\epsilon^4 & \gamma_\epsilon^5 - \gamma_\epsilon^3 \gamma_\epsilon^2 \\ \gamma_\epsilon^4 & \gamma_\epsilon^2 & \gamma_\epsilon^3 \\ (\gamma_\epsilon^5 - \gamma_\epsilon^3 \gamma_\epsilon^2) & \gamma_\epsilon^3 & \gamma_\epsilon^4 - (\gamma_\epsilon^2)^2 \end{bmatrix}, \quad \Gamma_b = \begin{bmatrix} \gamma_b^6 - (\gamma_b^3)^2 & \gamma_b^4 & \gamma_b^5 - \gamma_b^3 \gamma_b^2 \\ \gamma_b^4 & \gamma_b^2 & \gamma_b^3 \\ \gamma_b^5 - \gamma_b^3 \gamma_b^2 & \gamma_b^3 & \gamma_b^4 - (\gamma_b^2)^2 \end{bmatrix}.$$

Theorem 6. *Under Assumptions 1–4, when γ_b^8 and γ_ϵ^8 are finite, as $n \to \infty$, $\sqrt{N}\,(\widehat{\kappa}_\epsilon - \kappa_\epsilon) \rightsquigarrow \mathcal{N}[0, \beta_\epsilon^\top \Omega_\epsilon \beta_\epsilon/(\gamma_\epsilon^2)^4]$ and $\sqrt{n}\,(\widehat{\kappa}_b - \kappa_b) \rightsquigarrow \mathcal{N}[0, \beta_b^\top \Omega_b \beta_b/(\gamma_b^2)^4]$, where $\beta_\epsilon = (1, -4\gamma_\epsilon^3, -2\gamma_\epsilon^2 \kappa_\epsilon)^\top$ and $\beta_b = (1, -4\gamma_b^3, -2\gamma_b^2 \kappa_b)^\top$ are vectors of dimension 1×3, whereas Ω_b and Ω_ϵ are 3×3 dimensional matrices such as*

$$\Omega_\epsilon = \begin{bmatrix} \gamma_\epsilon^8 - (\gamma_\epsilon^4)^2 & \gamma_\epsilon^5 & \gamma_\epsilon^6 - \gamma_\epsilon^4 \gamma_\epsilon^2 \\ \gamma_\epsilon^5 & \gamma_\epsilon^2 & \gamma_\epsilon^3 \\ \gamma_\epsilon^6 - \gamma_\epsilon^4 \gamma_\epsilon^2 & \gamma_\epsilon^3 & \gamma_\epsilon^4 - (\gamma_\epsilon^2)^2 \end{bmatrix}, \quad \Omega_b = \begin{bmatrix} \gamma_b^8 - (\gamma_b^4)^2 & \gamma_b^5 & \gamma_b^6 - \gamma_b^4 \gamma_b^2 \\ \gamma_b^5 & \gamma_b^2 & \gamma_b^3 \\ \gamma_b^6 - \gamma_b^4 \gamma_b^2 & \gamma_b^3 & \gamma_b^4 - (\gamma_b^2)^2 \end{bmatrix}.$$

When the underlying distribution is symmetric, $s_b = 0$ and $s_\epsilon = 0$. Then, standardizing the above skewness test statistics we obtain the following corollary.

Corollary 1. *Under Assumptions 1–4 and assuming that γ_b^6 and γ_ϵ^6 are finite, let $sd(\widehat{s}_\epsilon) = \{\widehat{\alpha}_{\epsilon 0}^\top \widehat{\Gamma}_{\epsilon 0} \widehat{\alpha}_{\epsilon 0}/(\widehat{\gamma}_\epsilon^2)^3\}^{1/2}$ and $sd(\widehat{s}_b) = \{\widehat{\alpha}_{b0}^\top \widehat{\Gamma}_{b0} \widehat{\alpha}_{b0}/(\widehat{\gamma}_b^2)^3\}^{1/2}$, where $\widehat{\alpha}_{\epsilon 0}, \widehat{\alpha}_{b0}, \widehat{\Gamma}_{\epsilon 0}, \widehat{\Gamma}_{b0}, \widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$ are the consistent estimators for $\alpha_{\epsilon 0}, \alpha_{b0}, \Gamma_{\epsilon 0}, \Gamma_{b0}, \gamma_\epsilon^2$ and γ_b^2, respectively. Under the null hypothesis of symmetry, as $n \to \infty$, $\widehat{\pi}_{\epsilon 3} = \sqrt{N}\,\widehat{s}_\epsilon/sd(\widehat{s}_\epsilon) \rightsquigarrow \mathcal{N}(0, 1)$ and $\widehat{\pi}_{b3} = \sqrt{n}\,\widehat{s}_b/sd(\widehat{s}_b) \rightsquigarrow \mathcal{N}(0, 1)$, where $\widehat{\alpha}_{\epsilon 0}, \widehat{\alpha}_{b0}, \widehat{\Gamma}_{\epsilon 0}$ and $\widehat{\Gamma}_{b0}$ are the corresponding expressions of $\widehat{\alpha}_\epsilon, \widehat{\alpha}_b, \widehat{\Gamma}_\epsilon$ and $\widehat{\Gamma}_b$, respectively, under the null.*

The following corollary can then be exploited to test whether the distribution of the random effects and error terms is mesokurtic.

Corollary 2. *Under Assumptions 1–4 and assuming that γ_b^8 and γ_ϵ^8 are finite, let $sd(\widehat{\kappa}_\epsilon) = \{\widehat{\beta}_{\epsilon 3}^\top \widehat{\Omega}_{\epsilon 3} \widehat{\beta}_{\epsilon 3}/(\widehat{\gamma}_\epsilon^2)^4\}^{1/2}$ and $sd(\widehat{\kappa}_b) = \{\widehat{\beta}_{b3}^\top \widehat{\Omega}_{b3} \widehat{\beta}_{b3}/(\widehat{\gamma}_b^2)^4\}^{1/2}$, where $\widehat{\beta}_{\epsilon 3}, \widehat{\beta}_{b3}, \widehat{\Omega}_{\epsilon 3}, \widehat{\Omega}_{b3}, \widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$ are the consistent estimators for $\beta_{\epsilon 3}, \beta_{b3}, \Omega_{\epsilon 3}, \Omega_{b3}, \gamma_\epsilon^2$ and γ_b^2, respectively. Under the null hypothesis that $\kappa_\epsilon = 3$ and $\kappa_b = 3$, as $n \to \infty$, $\widehat{\pi}_{\epsilon 4} = \sqrt{N}\,(\widehat{\kappa}_\epsilon - 3)/sd(\widehat{\kappa}_\epsilon) \rightsquigarrow \mathcal{N}(0, 1)$ and $\widehat{\pi}_{b4} = \sqrt{n}\,(\widehat{\kappa}_b - 3)/sd(\widehat{\kappa}_b) \rightsquigarrow \mathcal{N}(0, 1)$, where $\widehat{\beta}_{\epsilon 3}, \widehat{\beta}_{b3}, \widehat{\Omega}_{\epsilon 3}$ and $\widehat{\Omega}_{b3}$ are the corresponding expressions of $\widehat{\beta}_\epsilon, \widehat{\beta}_b, \widehat{\Omega}_\epsilon$ and $\widehat{\Omega}_b$, respectively, under the null.*

Corollaries 1 and 2 imply that, under the null hypothesis, the proposed statistics for skewness and kurtosis are asymptotically standard normal. Simple Student t tests can thus be implemented by standardizing the corresponding statistics with the square root of a consistent estimator of their variances.

4.2. Test for normality

As pointed out by [1], the above skewness and kurtosis tests are useful since they can be used to test whether the data conform to any distribution of interest. In this section, we propose to extend the normality test proposed in [1] for time series data to a linear mixed model. Then, we are able to detect whether the random effects or the error terms departure from normality, jointly and separately.

Let $\widetilde{\pi}_{b3}$ and $\widetilde{\pi}_{b4}$ be the statistics for skewness and kurtosis evaluated under the null hypothesis of normality of the random effects, i.e., $s_b = 0$ and $\kappa_b = 3$, respectively. Then, the proposed statistic to construct tests for normality of the random effects is

$$\widetilde{\pi}_{b34} = \widetilde{\pi}_{b3}^2 + \widetilde{\pi}_{b4}^2.$$

Similarly, let $\widetilde{\pi}_{\epsilon 3}$ and $\widetilde{\pi}_{\epsilon 4}$ be the statistics for skewness and kurtosis evaluated under the null hypothesis of normality of the error terms, i.e., $s_\epsilon = 0$ and $\kappa_\epsilon = 3$, respectively. The proposed statistic to test for normality of the error terms is

$$\widetilde{\pi}_{\epsilon 34} = \widetilde{\pi}_{\epsilon 3}^2 + \widetilde{\pi}_{\epsilon 4}^2.$$

Theorem 7. *Under Assumptions 1–4 and assuming that γ_b^8 and γ_ϵ^8 are finite, under the null hypothesis of normality, as $n \to \infty$, $\widetilde{\pi}_{\epsilon 34} \rightsquigarrow \chi_2^2$ and $\widetilde{\pi}_{b34} \rightsquigarrow \chi_2^2$.*

Under the normality assumption, it can be proved that $\widetilde{\pi}_{\epsilon 3}$ and $\widetilde{\pi}_{\epsilon 4}$ are asymptotically independent, and similarly for $\widetilde{\pi}_{b3}$ and $\widetilde{\pi}_{b4}$. Thus, the proof of this theorem follows the same lines as the corresponding statements for Theorem 5 and it is therefore omitted.

Looking at the results of Corollaries 1–2 and Theorem 7, it can be pointed out that the statistics proposed in this section have the following appealing features. First, the skewness test statistics are valid even if the null hypothesis is not normally distributed, albeit symmetric. Second, in [1] and [4] it is pointed that one of the main weaknesses of this type of statistics for kurtosis is that it can be influenced by skewness. However, the statistics proposed in this paper are not affected by this problem. Third, the statistics proposed for the random effects are not affected by skewness or kurtosis in the remainder terms. The same holds for the statistics proposed for the error terms. Fourth, the proposed normality tests allow to detect whether one specific term departs from normality. Fifth, all these tests can be computed directly without having to use bootstrap techniques as in [4]. The only requirement is to compute consistent estimators up the eight moment of the random effects and error terms, but they are easily obtained using Lemma 1. These consistent estimators are given in Appendix B.

5. Monte Carlo simulation and application

To study the possible gains of the estimation procedures proposed in this paper, a simulation study was carried out. In Section 5.1, we compare the performance in small samples of these estimators with respect to the ones proposed in [14]. As further evidence of the usefulness of the proposed method, Section 5.2 contains an application to a real data set.

5.1. Simulation study

Consider the linear mixed model defined, for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, \ell_i\}$, by

$$y_{ij} = \alpha + x_{ij}^\top \beta + b_i + \epsilon_{ij}, \tag{10}$$

where $\alpha = 1$ and $\beta = (1, 2)^\top$. The group values are randomly drawn from a Poisson distribution with mean 5. Also, the explanatory variables are jointly normal with zero means, unit variances and correlation 0.6.

5.1.1. Finite-sample performance of the proposed estimators. In order to estimate the previous regression model, we separately consider the following combinations of distributions:

(a) $\epsilon_{ij} \sim 0.5\,\mathcal{N}(0,1)$ and $b_i \sim 0.5\,\mathcal{N}(0,1)$;

(b) $\epsilon_{ij} \sim 0.5\,\mathcal{N}(0,1)$ and $b_i \sim 0.5\,t_8$;

(c) $\epsilon_{ij} \sim 0.5\,\mathcal{N}(0,1)$ and $b_i \sim 0.5\,\Gamma(1,1) - 0.5$;

(d) $\epsilon_{ij} \sim 0.5\,\Gamma(1,1) - 0.5$ and $b_i \sim 0.5\,\mathcal{N}(0,1)$;

(e) $\epsilon_{ij} \sim 0.5\,t_8$ and $b_i \sim 0.5\,\mathcal{N}(0,1)$.

where $\mathcal{N}(0,1)$ refers to the univariate standard normal distribution, t_8 denotes the univariate Student t distribution with 8 degrees of freedom, and $\Gamma(1,1)$ is the univariate Gamma distribution with both scale and shape parameters equal to 1, i.e., a unit exponential distribution.

The following simulation results are based on 1000 samples of data in which the number of subjects is either 50, 100 or 150. The estimated mean, standard deviation (Std) and the root mean squared error (RMSE) of the estimators of the model parameters are listed in Table 1, while the results for the higher order moments are collected in Table 2. For the sake of comparison, in Table 1 the OLS estimators are included in brackets, while in Table 2 it is the estimators of the moments of [14] that are in brackets.

As can be seen from Tables 1–2, the proposed estimators perform quite well. As the sample size n increases, both the estimated standard deviation and the RMSE are lower in the five cases considered. In addition, when we compare the behavior between the proposed estimators in this paper and those obtained in [14], one can note that our estimators perform slightly better in all cases considered. This difference is reduced as the sample size increases, as one might expect from the theoretical results.

5.1.2. Finite-sample performance of the test statistic. In order to assess the performance of the proposed tests statistics, we used the same data generating process as in (10) and explored their effectiveness under the distributional processes (a)–(e).

In Table 3 we report the proportion of rejections over 1000 Monte Carlo replications, and the results are based on either 10%, 5% or 1% nominal size. For the sake of practical implementation, the following Wald tests were used: $\widehat{\pi}_{\epsilon3}^2$ and $\widehat{\pi}_{b3}^2$ are used to test for skewness, $\widehat{\pi}_{\epsilon4}^2$ and $\widehat{\pi}_{b4}^2$ for kurtosis, and $\widetilde{\pi}_{\epsilon34}$ and $\widetilde{\pi}_{b34}$ for normality. Thus, under the corresponding null hypothesis, the skewness and kurtosis statistics have χ_1^2 asymptotic distribution, whereas $\widetilde{\pi}_{\epsilon34}$ and $\widetilde{\pi}_{b34}$ follow a χ_2^2 distribution.

Table 3, case (a), reports the results for the experiment when both random effects and error terms follow a standard normal distribution. Looking at these results, it can be pointed out that all proposed tests show a good empirical size for the different sample sizes. Specifically, tests for skewness in ϵ starts with oversized figures for $n = 50$ (0.132 when $\alpha = 0.05$) but reduces to 0.126 as the sample size increases. The test for kurtosis starts with undersized values for $n = 50$ (0.038 when $\alpha = 0.05$) but the level reaches 0.051 as n increases. Further, the tests for joint skewness and kurtosis achieve the correct empirical size as n increases in a similar way to the test of skewness. Meanwhile, for the random effects it can be seen that all tests are undersized for the smaller sample size, but all of them almost achieve the corresponding level of significance as $n \rightarrow \infty$.

In addition, as noted in [1], there are numerous distributions for which tests based on higher order moments are not valid. In order to corroborate this fact, in cases (b)–(e) the performance of the proposed tests was analyzed under different non-Gaussian distributions. In particular, in case (b) we assumed $b_i \sim 0.5\,t_8$, so the power of the kurtosis tests for this term should increase in a similar way as the sample size. In fact, looking at this result it can be seen that effectively the power of the kurtosis test in b_i increases with n, but the tests for skewness and kurtosis in ϵ_{ij} achieve the empirical size. An opposite behavior is observed in case (e), where $\epsilon_{ij} \sim 0.5\,t_8$ is assumed. These results show that kurtosis in one component does not affect tests for kurtosis in the other error component.

In case (c) we assumed $b_j \sim 0.5\,\Gamma(1,1) - 0.5$. As it is well known, this gamma function is not symmetric and presents excess of kurtosis. Then, the tests for skewness and kurtosis for b_i should have a non-trivial power. The opposite can be said for case (d), where $\epsilon_{ij} \sim 0.5\,\Gamma(1,1) - 0.5$ is assumed. As the reader can see, all these results are corroborated in Table 3.

Finally, it can be pointed out that through these Monte Carlo experiments it has been proved that the methodology proposed in this paper enables us to obtain more efficient estimators in finite samples for the higher order moments of both components of the error term which are very useful for the empirical analysis. In addition, the proposed tests enable us to assess whether the data conform to any distribution of interest. In addition, they are not affected by the presence of skewness and/or kurtosis in the other component.

Table 1: Finite-sample performance of $\widehat{\alpha}$ and $\widehat{\beta}$ in cases (a)–(e).

Case	n	Results	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\alpha}$
(a)		True	1.000	2.000	1.000
	50	Mean	1.001 (1.002)	2.000 (1.999)	0.991 (0.991)
		Std	0.043 (0.043)	0.041 (0.043)	0.074 (0.074)
		RMSE	0.044 (0.044)	0.042 (0.044)	0.075 (0.075)
	100	Mean	1.001 (1.001)	1.999 (1.999)	1.002 (1.002)
		Std	0.027 (0.028)	0.027 (0.027)	0.055 (0.055)
		RMSE	0.027 (0.028)	0.027 (0.028)	0.056 (0.056)
	150	Mean	0.999 (0.999)	1.997 (1.997)	0.999 (1.002)
		Std	0.022 (0.023)	0.022 (0.024)	0.040 (0.040)
		RMSE	0.021 (0.023)	0.021 (0.023)	0.040 (0.040)
(b)		True	1.000	2.000	1.000
	50	Mean	0.998 (0.998)	2.002 (2.002)	1.004 (1.004)
		Std	0.040 (0.041)	0.041 (0.041)	0.084 (0.084)
		RMSE	0.040 (0.041)	0.041 (0.041)	0.084 (0.085)
	100	Mean	1.002 (1.002)	2.000 (2.000)	1.001 (1.001)
		Std	0.030 (0.031)	0.031 (0.032)	0.061 (0.061)
		RMSE	0.030 (0.031)	0.031 (0.032)	0.061 (0.061)
	150	Mean	1.000 (1.000)	2.000 (1.999)	0.9998 (0.998)
		Std	0.023 (0.024)	0.024 (0.024)	0.051 (0.051)
		RMSE	0.024 (0.024)	0.024 (0.024)	0.051 (0.051)
(c)		True	1.000	2.000	1.000
	50	Mean	1.001 (1.000)	1.998 (1.998)	1.000 (1.000)
		Std	0.041 (0.042)	0.041 (0.041)	0.080 (0.080)
		RMSE	0.041 (0.042)	0.041 (0.041)	0.080 (0.080)
	100	Mean	1.000 (0.999)	2.000 (2.001)	1.001 (1.001)
		Std	0.028 (0.028)	0.029 (0.029)	0.053 (0.054)
		RMSE	0.028 (0.028)	0.029 (0.029)	0.053 (0.054)
	150	Mean	1.000 (1.000)	2.000 (2.000)	1.000 (1.000)
		Std	0.023 (0.023)	0.023 (0.024)	0.044 (0.044)
		RMSE	0.023 (0.023)	0.023 (0.024)	0.044 (0.044)
(d)		True	1.000	2.000	1.000
	50	Mean	1.001 (1.001)	2.000 (2.000)	0.999 (0.999)
		Std	0.041 (0.043)	0.040 (0.043)	0.037 (0.037)
		RMSE	0.041 (0.043)	0.040 (0.043)	0.037 (0.037)
	100	Mean	0.999 (0.999)	2.000 (2.001)	0.999 (0.999)
		Std	0.028 (0.030)	0.028 (0.030)	0.027 (0.027)
		RMSE	0.028 (0.030)	0.028 (0.030)	0.027 (0.027)
	150	Mean	0.999 (0.999)	2.000 (2.000)	1.000 (1.000)
		Std	0.022 (0.024)	0.023 (0.024)	0.022 (0.022)
		RMSE	0.022 (0.024)	0.022 (0.024)	0.022 (0.022)
(e)		True	1.000	2.000	1.000
	50	Mean	1.008 (1.008)	1.994 (1.993)	0.994 (0.994)
		Std	0.048 (0.049)	0.048 (0.050)	0.086 (0.086)
		RMSE	0.049 (0.050)	0.049 (0.051)	0.086 (0.086)
	100	Mean	0.996 (0.996)	1.999 (2.000)	1.002 (1.002)
		Std	0.035 (0.035)	0.035 (0.035)	0.058 (0.058)
		RMSE	0.035 (0.035)	0.035 (0.035)	0.058 (0.058)
	150	Mean	1.000 (1.000)	1.998 (1.998)	1.000 (1.000)
		Std	0.026 (0.028)	0.025 (0.027)	0.046 (0.046)
		RMSE	0.026 (0.028)	0.025 (0.027)	0.046 (0.046)

Table 2: Finite-sample performance of γ_ϵ^k and γ_b^k, for $k \in \{2, 3, 4\}$, in cases (a)–(e).

Case	n	Results	$\widehat{\gamma}_\epsilon^2$	$\widehat{\gamma}_\epsilon^3$	$\widehat{\gamma}_\epsilon^4$	$\widehat{\gamma}_b^2$	$\widehat{\gamma}_b^3$	$\widehat{\gamma}_b^4$
(a)		True	0.250	0.000	0.187	0.250	0.000	0.187
	50	Mean	0.248 (0.248)	0.000 (0.000)	0.179 (0.179)	0.239 (0.239)	0.000 (0.000)	0.186 (0.186)
		Std	0.022 (0.022)	0.026 (0.026)	0.038 (0.040)	0.064 (0.064)	0.034 (0.034)	0.131 (0.136)
		RMSE	0.011 (0.011)	0.028 (0.028)	0.027 (0.030)	0.042 (0.042)	0.069 (0.071)	0.079 (0.084)
	100	Mean	0.250 (0.250)	0.001 (0.001)	0.186 (0.188)	0.244 (0.244)	0.004 (0.004)	0.176 (0.176)
		Std	0.017 (0.017)	0.018 (0.018)	0.035 (0.037)	0.036 (0.036)	0.021 (0.021)	0.071 (0.075)
		RMSE	0.006 (0.006)	0.020 (0.020)	0.019 (0.020)	0.025 (0.025)	0.048 (0.048)	0.046 (0.047)
	150	Mean	0.250 (0.250)	0.001 (0.001)	0.189 (0.189)	0.245 (0.245)	0.000 (0.000)	0.181 (0.182)
		Std	0.013 (0.013)	0.014 (0.014)	0.027 (0.026)	0.031 (0.031)	0.030 (0.033)	0.056 (0.056)
		RMSE	0.005 (0.005)	0.015 (0.015)	0.016 (0.016)	0.021 (0.021)	0.039 (0.039)	0.045 (0.045)
(b)		True	0.250	0.000	0.187	0.333	0.000	0.500
	50	Mean	0.248 (0.248)	0.001 (0.001)	0.184 (0.185)	0.327 (0.327)	0.000 (0.000)	0.477 (0.477)
		Std	0.023 (0.023)	0.025 (0.025)	0.045 (0.047)	0.097 (0.097)	0.044 (0.051)	0.598 (0.594)
		RMSE	0.040 (0.041)	0.041 (0.041)	0.084 (0.085)	0.048 (0.048)	0.124 (0.112)	0.209 (0.210)
	100	Mean	0.248 (0.248)	0.000 (0.000)	0.185 (0.185)	0.327 (0.327)	0.003 (0.006)	0.463 (0.464)
		Std	0.017 (0.017)	0.021 (0.021)	0.035 (0.036)	0.066 (0.066)	0.031 (0.035)	0.335 (0.335)
		RMSE	0.030 (0.031)	0.031 (0.032)	0.061 (0.061)	0.034 (0.034)	0.119 (0.106)	0.177 (0.177)
	150	Mean	0.249 (0.249)	0.000 (0.000)	0.185 (0.186)	0.327 (0.327)	0.000 (0.000)	0.468 (0.469)
		Std	0.013 (0.013)	0.014 (0.014)	0.026 (0.027)	0.053 (0.053)	0.027 (0.028)	0.358 (0.360)
		RMSE	0.024 (0.024)	0.024 (0.024)	0.051 (0.051)	0.028 (0.028)	0.103 (0.104)	0.136 (0.137)
(c)		True	0.250	0.000	0.187	0.250	0.250	0.562
	50	Mean	0.247 (0.247)	0.000 (0.000)	0.182 (0.182)	0.240 (0.240)	0.226 (0.227)	0.489 (0.489)
		Std	0.023 (0.023)	0.024 (0.024)	0.044 (0.046)	0.102 (0.102)	0.234 (0.234)	0.800 (0.800)
		RMSE	0.010 (0.010)	0.028 (0.028)	0.027 (0.030)	0.049 (0.049)	0.128 (0.129)	0.368 (0.366)
	100	Mean	0.249 (0.249)	0.000 (0.000)	0.187 (0.187)	0.249 (0.249)	0.246 (0.248)	0.560 (0.560)
		Std	0.016 (0.016)	0.019 (0.019)	0.032 (0.034)	0.072 (0.072)	0.180 (0.180)	0.691 (0.690)
		RMSE	0.007 (0.007)	0.021 (0.022)	0.019 (0.021)	0.034 (0.034)	0.093 (0.093)	0.330 (0.328)
	150	Mean	0.250 (0.250)	0.000 (0.000)	0.187 (0.187)	0.250 (0.250)	0.242 (0.244)	0.534 (0.533)
		Std	0.013 (0.013)	0.014 (0.014)	0.024 (0.026)	0.057 (0.057)	0.133 (0.134)	0.463 (0.464)
		RMSE	0.006 (0.006)	0.016 (0.016)	0.016 (0.016)	0.028 (0.028)	0.074 (0.076)	0.231 (0.234)
(d)		True	0.250	0.250	0.562	0.250	0.000	0.187
	50	Mean	0.252 (0.252)	0.246 (0.246)	0.542 (0.545)	0.237 (0.237)	0.000 (0.000)	0.181 (0.1739)
		Std	0.042 (0.042)	0.116 (0.116)	0.375 (0.383)	0.059 (0.059)	0.028 (0.029)	0.122 (0.104)
		RMSE	0.012 (0.012)	0.045 (0.045)	0.091 (0.111)	0.039 (0.039)	0.067 (0.074)	0.104 (0.082)
	100	Mean	0.250 (0.250)	0.249 (0.249)	0.545 (0.545)	0.245 (0.245)	0.000 (0.000)	0.176 (0.182)
		Std	0.031 (0.031)	0.078 (0.078)	0.259 (0.262)	0.046 (0.046)	0.022 (0.022)	0.092 (0.086)
		RMSE	0.007 (0.007)	0.024 (0.024)	0.052 (0.062)	0.029 (0.029)	0.048 (0.049)	0.072 (0.072)
	150	Mean	0.250 (0.250)	0.249 (0.249)	0.548 (0.548)	0.217 (0.217)	0.000 (0.000)	0.151 (0.157)
		Std	0.022 (0.022)	0.060 (0.060)	0.245 (0.245)	0.037 (0.037)	0.019 (0.019)	0.078 (0.079)
		RMSE	0.006 (0.006)	0.029 (0.029)	0.069 (0.079)	0.021 (0.021)	0.038 (0.040)	0.052 (0.045)
(e)		True	0.333	0.000	0.500	0.250	0.000	0.187
	50	Mean	0.330 (0.330)	−0.010 (−0.010)	0.445 (0.448)	0.241 (0.241)	0.000 (0.001)	0.176 (0.176)
		Std	0.044 (0.044)	0.055 (0.055)	0.171 (0.178)	0.060 (0.060)	0.018 (0.061)	0.104 (0.103)
		RMSE	0.014 (0.014)	0.054 (0.054)	0.064 (0.073)	0.035 (0.035)	0.069 (0.072)	0.078 (0.071)
	100	Mean	0.326 (0.326)	0.001 (0.001)	0.466 (0.469)	0.242 (0.242)	0.000 (0.000)	0.186 (0.184)
		Std	0.025 (0.025)	0.053 (0.053)	0.158 (0.168)	0.044 (0.044)	0.013 (0.043)	0.082 (0.079)
		RMSE	0.009 (0.009)	0.036 (0.036)	0.058 (0.063)	0.029 (0.029)	0.049 (0.054)	0.061 (0.060)
	150	Mean	0.329 (0.329)	0.004 (0.004)	0.487 (0.492)	0.246 (0.246)	0.000 (0.000)	0.182 (0.189)
		Std	0.025 (0.025)	0.046 (0.046)	0.164 (0.164)	0.037 (0.037)	0.012 (0.039)	0.072 (0.071)
		RMSE	0.007 (0.007)	0.028 (0.028)	0.057 (0.054)	0.026 (0.026)	0.041 (0.044)	0.054 (0.052)

Table 3: Finite-sample performance of skewness, kurtosis and normality tests in cases (a)–(e).

Case	n	Results	Error Term Significance Level			Random Effects Significance Level		
			0.10	0.05	0.01	0.10	0.05	0.01
(a)	50	Skewness	0.221	0.132	0.046	0.035	0.020	0.017
		Kurtosis	0.073	0.038	0.013	0.049	0.025	0.003
		Normality	0.147	0.078	0.005	0.024	0.013	0.002
	100	Skewness	0.225	0.148	0.036	0.043	0.028	0.011
		Kurtosis	0.095	0.041	0.007	0.082	0.049	0.014
		Normality	0.160	0.087	0.006	0.045	0.023	0.002
	150	Skewness	0.201	0.126	0.029	0.086	0.041	0.006
		Kurtosis	0.102	0.051	0.009	0.116	0.051	0.012
		Normality	0.133	0.067	0.010	0.083	0.041	0.002
(b)	50	Skewness	0.271	0.183	0.054	0.201	0.153	0.086
		Kurtosis	0.106	0.062	0.024	0.059	0.025	0.010
		Normality	0.192	0.093	0.020	0.097	0.073	0.026
	100	Skewness	0.200	0.136	0.022	0.182	0.125	0.080
		Kurtosis	0.071	0.031	0.015	0.189	0.107	0.040
		Normality	0.224	0.086	0.015	0.264	0.154	0.015
	150	Skewness	0.140	0.100	0.020	0.157	0.112	0.005
		Kurtosis	0.076	0.076	0.026	0.242	0.175	0.109
		Normality	0.155	0.099	0.013	0.281	0.180	0.032
(c)	50	Skewness	0.255	0.154	0.074	0.362	0.241	0.081
		Kurtosis	0.090	0.033	0.016	0.127	0.059	0.029
		Normality	0.168	0.071	0.014	0.197	0.134	0.051
	100	Skewness	0.177	0.198	0.062	0.548	0.391	0.161
		Kurtosis	0.137	0.057	0.010	0.307	0.180	0.066
		Normality	0.192	0.085	0.012	0.343	0.278	0.016
	150	Skewness	0.151	0.133	0.062	0.616	0.464	0.188
		Kurtosis	0.102	0.045	0.015	0.441	0.290	0.072
		Normality	0.123	0.079	0.017	0.384	0.335	0.062
(d)	50	Skewness	0.525	0.394	0.224	0.032	0.021	0.011
		Kurtosis	0.670	0.493	0.186	0.037	0.029	0.009
		Normality	0.268	0.211	0.044	0.027	0.014	0.001
	100	Skewness	0.587	0.278	0.112	0.038	0.021	0.011
		Kurtosis	0.657	0.517	0.240	0.077	0.050	0.016
		Normality	0.142	0.095	0.022	0.043	0.028	0.006
	150	Skewness	0.741	0.403	0.052	0.086	0.044	0.008
		Kurtosis	0.682	0.555	0.335	0.081	0.060	0.022
		Normality	0.243	0.099	0.007	0.054	0.035	0.002
(e)	50	Skewness	0.196	0.099	0.024	0.034	0.024	0.006
		Kurtosis	0.386	0.252	0.074	0.069	0.045	0.019
		Normality	0.302	0.163	0.020	0.037	0.023	0.006
	100	Skewness	0.161	0.086	0.022	0.042	0.025	0.010
		Kurtosis	0.557	0.411	0.145	0.089	0.057	0.020
		Normality	0.414	0.256	0.025	0.054	0.034	0.005
	150	Skewness	0.146	0.075	0.017	0.087	0.042	0.007
		Kurtosis	0.676	0.526	0.226	0.102	0.065	0.019
		Normality	0.485	0.329	0.043	0.051	0.030	0.002

5.2. Application to a real data set

In order to illustrate our proposed method, in this section we perform an empirical analysis of the health care delivery system of the OECD countries. Furthermore, we compare our GLS results with the standard OLS estimators.

These data are from the OECD Health Data 2015 and the Education Statistics of The World Bank. With the aim of having enough sample variability for all covariates, we consider 19 OCDE countries, i.e., all OECD nations except Belgium, Canada, Chile, France, Germany, Greece, Ireland, Italy, Korea, Norway, Portugal, Sweden, Switzerland, the Netherlands, and Turkey. After removing the missing observations we obtain a dataset for which we do not have the same number of time observations per country. The total sample size of this study is 225 observations. The variables included in this study are defined in Table 4.

Table 4: Variables description.

Variable	Description	Source
LIFE	Life expectancy at birth, for the total population	OECD Health Data 2015
MORT	Maternal and infant mortality rate, per 1000 live births	OECD Health Data 2015
TECH	Number of medical technology*, per million population	OECD Health Data 2015
PHYS	Density of practicing physicians, per 1000 population	OECD Health Data 2015
PATIENT	In-patient care average length of stay in hospital	OECD Health Data 2015
INMUNI	Percentage of children immunized against 2000 measles	OECD Health Data 2015
HEXP	Total expenditure on health per capital, (US$ PPP)	OECD Health Data 2015
ALCO	Alcohol consumption litres per capita (age 15+)	OECD Health Data 2015
EDU	School life expectancy years, primary and secondary	The World Bank

*Medical technology: Computed tomography scanners, magnetic resonance imaging, positron emission tomography scanners, gamma cameras, digital subtraction angiography, mammograph, radiation therapy equipment, and lithrotriptors

Before specifying the structural model of health outcome, we wish to stress issues related to the characteristics of the health care delivery system of the OECD countries. In the last decade, expenditures on health and health care have increased considerably mainly due to medical advancement, population aging, and rising public demand. Specifically, as described in the OECD database, health expenditure as a percentage of gross domestic product (GDP) was around 7.23% in 2000, and this figure has increased in recent years reaching 8.99% in 2013. Nevertheless, in recent years there has been a slight reduction in this growth mainly due to the global economic crisis. In this situation, it is of great interest to be able to identify the relative importance of various factors in key aspects of health outcomes such as life expectancy or maternal and infant mortality in order to develop efficient public policies. Nevertheless, these estimators can lead to unreliable inference in those cases in which life expectancy or maternal and infant mortality follow a skewed distribution. Therefore, testing for skewness or normality is an interesting issue.

Specifically, in this application we propose to assess the impact of economic, institutional, and social factors in the health outcome, approximated through the life expectancy and the infant and maternity mortality rate. With this aim, we consider two different regression models.

In the first regression model, the dependent variable is life expectancy (LIFE), which represents the typical number of years that a person at that age can be anticipated to live, presuming that age-specific mortality levels remain constant. In the second regression model, the dependent variable is the maternal and infant mortality rate (MORT) that encompasses the number of deaths of children under one year of age that occurred in a given year.

Assuming that LIFE and MORT are functions of the countries' medical investment, the quality of the health care system, and the educational level of the population, the regression models to estimate are the following:

$$\text{LIFE}_{ij} = \alpha + \beta_1 \text{TECH}_{ij} + \beta_2 \text{PHYS}_{ij} + \beta_3 \text{PATIENT}_{ij} + \beta_4 \text{INMUNI}_{ij}$$
$$+ \beta_5 H\text{EXP}_{ij} + \beta_6 \text{ALCO}_{ij} + \beta_7 \text{EDU}_{ij} + b_i + \epsilon_{ij},$$

$$\text{MORT}_{ij} = \alpha + \beta_1 \text{TECH}_{ij} + \beta_2 \text{PHYS}_{ij} + \beta_3 \text{PATIENT}_{ij} + \beta_4 \text{INMUNI}_{ij}$$
$$+ \beta_5 \text{HEXP}_{ij} + \beta_6 \text{ALCO}_{ij} + \beta_7 \text{EDU}_{ij} + b_i + \epsilon_{ij},$$

13

where $i \in \{1, \dots, 19\}$ denotes the individuals (i.e., countries) and $j \in \{1, \dots, \ell_i\}$ is the number of observations per individual.

Tables 5 and 6 collect the regression results of life expectancy and maternal and infant mortality, respectively, whereas Table 7 reports the skewness, kurtosis, and normality tests for both regression models.

Analyzing the results in Tables 5 and 6, it can be pointed out that the signs of the parameter estimates of all the variables in the life expectancy and mortality regression are as expected. From the life expectancy point of view, more medical resources (PHYS), a higher level of immunization (INMUNI), a higher health expenditure (HEXP), a higher literacy rate (EDU), and a better access to medical technology (TECH) have a positive impact. Contrary, high in-patient utilization (PATIENT) and high alcohol consumption (ALCO) reduce life expectancy. Also, higher in-patient utilization (PATIENT) and a greater number of children immunized (INMUNI) are related with an increment in the mortality rate, whereas the rest of the explanatory variables of this regression reduce the mortality rate.

In addition, in the first regression all the explanatory variables considered, except ALCO, are statistically significant at the 10% level or better using the statistical test proposed previously. For the second regression, all the explanatory variables, except PATIENT and INMUNI, are statistically significant even at the 1% level. Then, higher supply of physicians and the inversion in medical training and education should be the primary focus of the OECD national health care policy. Further, several differences can be noticed when we compare the results of the OLS and GLS estimators. Specifically, the OLS estimators show more significant coefficients, while the GLS method proposed in this paper is more conservative.

In addition, from Table 7 it is found that the idiosyncratic error term for the life expectancy regression model is symmetric but with large kurtosis, and the joint test for skewness and normality is rejected at the 5% significance level. Meanwhile, for the infant and maternity mortality rate, one finds that ϵ_{ij} is symmetric with kurtosis close to the normal value of 3 and the joint test is rejected at the 5% significance level. As to the country level component, in both regression models it is shown to be symmetric with a kurtosis coefficient close to the normal value. Then, based on these results we can conclude that alternative estimation techniques for generalized linear mixed models should be used in order to provide more reliable results. This issue is left for future work.

6. Conclusion

Linear mixed models provide a useful tool to fit continuous longitudinal data, and the random effects and error terms are commonly assumed to have a normal distribution. However, these restrictive assumptions may result in a lack of robustness against departures from normality and invalid statistical inferences, especially when data exhibit multimodality and skewness. Therefore, it is of practical interest to test for normality. In this paper, we propose a very simple and intuitive test for skewness, kurtosis, and normality based on GLS residuals. This requires the estimation of higher order moments, and an alternative estimation procedure is developed to this end. Compared to other procedures in the literature, our method is very appealing since it results in closed form expressions for the estimators even for the

Table 5: Estimated results for life expectancy

	OLS		GLS	
	$\widehat{\beta}_{OLS}$	$\widehat{se}(\widehat{\beta}_{OLS})$	$\widehat{\beta}_{GLS}$	$\widehat{se}(\widehat{\beta}_{GLS})$
INTERCEPT	5.679***	0.041	5.924***	0.042
PHYS	0.764**	0.351	0.677*	0.354
PATIENT	−0.149**	0.069	−0.175**	0.069
INMUNI	0.051**	0.021	0.037*	0.021
HEXP	0.001***	0.000	0.001***	0.000
ALCO	−0.040	0.059	−0.049	0.059
EDU	0.292**	0.124	0.212*	0.125
TECH	0.0083***	0.002	0.008***	0.002

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Table 6: Estimated results for maternal and infant mortality

	OLS		GLS	
	$\widehat{\beta}_{OLS}$	$\widehat{se}(\widehat{\beta}_{OLS})$	$\widehat{\beta}_{GLS}$	$\widehat{se}(\widehat{\beta}_{GLS})$
INTERCEPT	0.917***	0.055	1.697***	0.054
PHYS	−1.433***	0.461	−1.709***	0.462
PATIENT	0.128***	0.088	0.052	0.091
INMUNI	0.044*	0.026	0.001	0.028
HEXP	−0.000	0.000	−0.000	0.000
ALCO	−0.179**	0.078	−0.211***	0.078
EDU	−0.286*	0.149	−0.542***	0.164
TECH	−0.007***	0.002	−0.006**	0.003

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level.

Table 7: Estimated results for maternal and infant mortality

	$\widehat{\pi}^2_{\epsilon_3}$	$\widehat{\pi}^2_{b_3}$	$\widehat{\pi}^2_{\epsilon_4}$	$\widehat{\pi}^2_{b_4}$	$\widetilde{\pi}_{\epsilon_{34}}$	$\widetilde{\pi}_{b_{34}}$
LIFE	1.806	0.055	9.033***	0.794	9.194**	0.774
MORT	1.774	0.061	1.381	0.001	8.577**	0.001

* Significant at the 10% level. ** Significant at the 5% level. *** Significant at the 1% level. *Note:* Using the results of Corollaries 1–2 and Theorem 4, the proposed test statistics follow a chi-square distribution even when standard distributional assumptions for the disturbances do not hold.

third and fourth order moments. In addition, no further distributional assumptions on either random effects or error terms are needed to show the consistency of the proposed estimators and tests statistics.

Finally, to illustrate the value and practicality of this new methodology, the finite-sample behavior of the proposed estimators and tests statistics has been analyzed through several simulation experiments. A study of the health care delivery system of the OECD countries was also carried out. As a result it was shown that health care expenditures, higher supply of physicians, and the inversion in medical training and education should be priority aspects of the OECD national health care policies because of their positive impact in both lengthening life expectancy and reducing infant mortality. In addition, the test for skewness of the error terms suggests that alternative estimation techniques would provide more reliable results.

Appendix A

Proof of Theorem 1. In order to prove the claims in Theorem 1, we follow the standard proofs scheme as in [14]. Fix $a \in \mathbb{R}^p$. Under the conditions of Theorem 1 we have to show $\sqrt{N}(\widetilde{\beta} - \beta) \rightsquigarrow \mathcal{N}(0, a^\top \Sigma_2^{-1} a)$. Using the results of Theorem 2, it can be proved that $\widehat{\Omega}_{\ell_i}$ is a consistent estimator for Ω_{ℓ_i}. Also, it is easy to show that the jth element of $\ddot{y}_i$ is $y_{ij} - \lambda \overline{y}_{i\bullet}$ and similarly for $\ddot{x}_i$. Then, the estimator (5) can be rewritten as

$$\widehat{\beta} - \beta = \left\{ \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} (x_{ij} - \lambda \overline{x}_{i\bullet})(x_{ij} - \lambda \overline{x}_{i\bullet})^\top \right\}^{-1} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} (x_{ij} - \lambda \overline{x}_{i\bullet})(y_{ij} - \lambda \overline{y}_{i\bullet}) + o_P(1).$$

According to this expression, $\widehat{\beta} - \beta$ is a sum of zero-mean independent random vectors. Then, to obtain the convergence in distribution of this estimator it is necessary to check the variance term and Lindeberg's condition.

Focus first on the variance of $\sqrt{N} a^\top (\widehat{\beta} - \beta)$. Under Assumption 3, we get

$$\frac{\gamma_\epsilon^2}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} \{a^\top \Sigma_n^{-1}(x_{ij} - \lambda \overline{x}_{i\bullet})\}^2 = \gamma_\epsilon^2 a^\top \Sigma_{2n}^{-1} a \to \gamma_\epsilon^2 a^\top \Sigma_2^{-1} a.$$

Furthermore, to verify Lindeberg's condition, we fix $\delta > 0$. Then, the Lindeberg function equals

$$L_n(\delta) = \frac{1}{N} \sum_{i=1}^n \{a^\top \Sigma_n^{-1}(x_{ij} - \lambda \overline{x}_{i\bullet})\}^2 \int_{\{|a^\top \Sigma_n^{-1}(x_{ij} - \lambda \overline{x}_{i\bullet})||\epsilon_{ij} - \lambda \overline{\epsilon}_{i\bullet}| \geq \delta \sqrt{N}\}} (\epsilon_{ij} - \lambda \overline{\epsilon}_{i\bullet})^2 dP.$$

Then, under Assumption 3 and by the Cauchy–Schwarz inequality, we can conclude that

$$L_n(\delta) \leq a^\top \Sigma_n^{-1} a \int_{\{|a^\top \Sigma_n^{-1}||\epsilon_{ij} - \lambda \overline{\epsilon}_{i\bullet}| \geq \delta \widetilde{C}_n^{-1}\}} (\epsilon_{ij} - \lambda \overline{\epsilon}_{i\bullet})^2 dP \to 0,$$

where $\widetilde{C}_n = (1/\sqrt{N}) \max_{ij} \|x_{ij} - \lambda \overline{x}_{i\bullet}\|$. Remember that under Assumption 4, $\widetilde{C}_n \to 0$ as $n \to \infty$ so Lindeberg's condition is proved.

Focus now on the behavior of $\widehat{\alpha}$. We can write

$$\widehat{\alpha} - \alpha = \frac{1}{n} \sum_{i=1}^n b_i + \overline{\epsilon} - \frac{1}{n} \sum_{i=1}^n \overline{x}'_{i\bullet}(\widehat{\beta} - \beta).$$

Under the assumptions of the theorem, it is straightforward to show that

$$\sqrt{n}(\widehat{\alpha} - \alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^n b_i + o_P(1) \rightsquigarrow \mathcal{N}(0, \gamma_b^2).$$

Thus the proof of Theorem 1 is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 2. We first study the asymptotic properties of $\widehat{\gamma}_\epsilon^2$ and later we focus on $\widehat{\gamma}_b^2$. Remember that $\widehat{\widetilde{\epsilon}}_{ij} = (y_{ij} - \overline{y}_{i\bullet}) - (x_{ij} - \overline{x}_{i\bullet})^\top \widehat{\beta}$, so it follows from (6) that

$$\widehat{\gamma}_\epsilon^2 = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^{\ell_i} \epsilon_{ij}^2 - \frac{1}{N} \sum_{i=1}^n \frac{1}{(\ell_i - 1)} \sum_{j=1}^{\ell_i} \sum_{h \neq j}^{\ell_i} \epsilon_{ij} \epsilon_{ih} + \frac{1}{N} \sum_{i=1}^n \frac{\ell_i}{(\ell_i - 1)} \sum_{j=1}^{\ell_i} \{(x_{ij} - \overline{x}_{i\bullet})^\top (\widehat{\beta} - \beta)\}^2$$

$$- \frac{2}{N} \sum_{i=1}^n \frac{\ell_i}{(\ell_i - 1)} \sum_{j=1}^{\ell_i} \widetilde{\epsilon}_{ij}(x_{ij} - \overline{x}_{i\bullet})^\top (\widehat{\beta} - \beta) = \mathbb{I}_\epsilon - \mathbb{III}_\epsilon + \mathbb{IIII}_\epsilon - \mathbb{IV}_\epsilon, \quad \text{say.} \quad \text{(A1)}$$

As it is going to be shown below, only the first one of these four terms matters; the others are negligible. Analyzing each of these terms separately and proceeding as in [14], under Assumption 4 and using the results of Theorem 1, it is straightforward to obtain that $\mathbb{III}_\epsilon/\sqrt{N}$, $\mathbb{IIII}_\epsilon/\sqrt{N}$, and $\mathbb{IV}_\epsilon/\sqrt{N}$ converges in probability to 0.

Using these results in (A1) we can write

$$\sqrt{N}(\widehat{\gamma}_\epsilon^2 - \gamma_\epsilon^2) = \frac{1}{\sqrt{N}} \sum_{i=1}^n \sum_{j=1}^{\ell_i} (\epsilon_{ij}^2 - \gamma_\epsilon^2) + o_P(1),$$

and by the Central Limit Theorem it can be proved that, as $n \to \infty$, $\sqrt{N}(\widehat{\gamma}_\epsilon^2 - \gamma_\epsilon^2) \rightsquigarrow \mathcal{N}(0, \mu_\epsilon^2)$, where $\mu_\epsilon^2 = \gamma_\epsilon^4 - (\gamma_\epsilon^2)^2$. Thus, the first part of Theorem 1 is proved.

Focus now on the asymptotic properties of $\widehat{\gamma}_b^2$. Let us denote $\overline{b} = \sum_{i=1}^n b_i/n$, $\overline{\epsilon} = \sum_{i=1}^n \sum_{j=1}^{\ell_i} \epsilon_{ij}/N$ and $\overline{x} = \sum_{i=1}^n \sum_{j=1}^{\ell_i} x_{ij}/N$. Using these definitions, we have

$$\widehat{\alpha} - \alpha = \overline{b} + \overline{\epsilon} - \overline{x}^\top (\widehat{\beta} - \beta).$$

16

Hence, the residual term $\widehat{v}_{ij}$ is

$$\widehat{v}_{ij} = (b_i + \epsilon_{ij}) - (\overline{b} + \overline{\epsilon}) - (x_{ij} - \overline{x})^\top (\widehat{\beta} - \beta). \tag{A2}$$

Using the results of Theorem 1, the higher order terms of $(x_{ij} - \overline{x}_{i\bullet})^\top (\widehat{\beta} - \beta)$ can be ignored. Then, replacing (A2) in (6) the estimator for γ_b^2 has the form

$$\widehat{\gamma}_b^2 = \frac{1}{n} \sum_{i=1}^n b_i^2 + \frac{2}{n} \sum_{i=1}^n \frac{1}{\ell_i} \sum_{j=1}^{\ell_i} b_i \epsilon_{ij} + \frac{1}{n} \sum_{i=1}^n \frac{2}{\ell_i(\ell_i - 1)} \sum_{j=1}^{\ell_i - 1} \sum_{h=j+1}^{\ell_i} \epsilon_{ij} \epsilon_{ih} - \overline{\epsilon}^2 + o_P(n^{-1/2})$$

$$= \mathbb{I}_b + \mathbb{II}_b + \mathbb{III}_b - \mathbb{IV}_b + o_P(n^{-1/2}), \text{ say.} \tag{A3}$$

By arguments similar to those used previously, it is straightforward to show that $\mathbb{II}_b$, $\mathbb{III}_b$, and $\mathbb{IV}_b$ are independent variables with zero mean and finite variance. Then, we can conclude that $n^{1/2}\mathbb{II}_b$, $n^{1/2}\mathbb{III}_b$, and $n^{1/2}\mathbb{IV}_b$ converge in probability to zero. Using these results in (A3) and under Assumption 4 we obtain

$$\sqrt{n}\,(\widehat{\gamma}_b^2 - \gamma_b^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (b_i^2 - \gamma_b^2) + o_P(1).$$

By the Central Limit Theorem, as $n \to \infty$, $\sqrt{n}\,(\widehat{\gamma}_b^2 - \gamma_b^2) \rightsquigarrow \mathcal{N}(0, \mu_b^2)$, where $\mu_b^2 = \gamma_b^4 - (\gamma_b^2)^2$. $\qquad\square$

Proof of Theorem 3. As in the previous proof, using the results of Theorem 1 we can show the higher order terms of $(x_{ij} - \overline{x}_{i\bullet})^\top (\widehat{\beta} - \beta)$ to be negligible. Therefore, replacing $\widehat{v}_{ij}$ with (A2) in (7) we obtain

$$n\widehat{\gamma}_b^3 = \sum_i \{b_i^3 - 3b_i^2(\overline{b} + \overline{\epsilon})\} + \sum_i \frac{6}{\ell_i(\ell_i - 1)(\ell_i - 2)} \sum_{j \neq h \neq g} \epsilon_{ij} \epsilon_{ih} \epsilon_{ig}$$

$$+ \sum_i \frac{2}{\ell_i(\ell_i - 1)} \sum_{j \neq h} \{4b_i \epsilon_{ij} \epsilon_{ih} - 3\epsilon_{ij} \epsilon_{ih}(\overline{b} + \overline{\epsilon})\} + \sum_i \frac{1}{\ell_i} \sum_j \{3b_i^2 \epsilon_{ij} - 6b_i \epsilon_{ij}(\overline{b} + \overline{\epsilon}) + 3\epsilon_{ij}(\overline{b} + \overline{\epsilon})^2\}$$

$$+ 3n\overline{b}(\overline{b} + \overline{\epsilon})^2 - n(\overline{b} + \overline{\epsilon})^3 + o_P(n^{1/2}).$$

Under Assumption 4, the first expression is the leading term since the rest of the elements are negligible and the resulting equation to study is

$$\widehat{\gamma}_b^3 - \gamma_b^3 = \frac{1}{n} \sum_i \{b_i^3 - \gamma_b^3 - 3b_i^2(\overline{b} + \overline{\epsilon})\} + o_P(n^{-1/2}) = \frac{1}{n} \sum_i \{b_i^3 - \gamma_b^3 - 3\gamma_b^2 b_i\} + o_P(n^{-1/2}).$$

Then, by the Central Limit Theorem, the first part of Theorem 3 is proved.

On the other hand, for $\widehat{\gamma}_\epsilon^3$, we can use the previous arguments. Then, ignoring once again the higher order terms of $(x_{ij} - \overline{x}_{i\bullet})^\top (\widehat{\beta} - \beta)$, we can write

$$N\widehat{\gamma}_\epsilon^3 = \sum_i \frac{\ell_i^2}{(\ell_i - 1)(\ell_i - 2)} \sum_j \widetilde{\epsilon}_{ij}^3 + o_P(N^{1/2}).$$

Then, by the expansion of $\widetilde{\epsilon}_{ij}^3$ the expression to study is

$$N\widehat{\gamma}_\epsilon^3 = \sum_i \sum_j \epsilon_{ij}^3 - \sum_i \frac{3}{(\ell_i - 1)} \sum_{j \neq h} \epsilon_{ij}^2 \epsilon_{ih} + \sum_i \frac{2}{(\ell_i - 1)(\ell_i - 2)} \sum_{j \neq h \neq g} \epsilon_{ij} \epsilon_{ih} \epsilon_{ig} + o_P(N^{1/2})$$

$$= \sum_i \sum_j (\epsilon_{ij}^3 - \gamma_\epsilon^2 \epsilon_{ij}) + o_P(N^{1/2}).$$

So it is proved that after centering ϵ_{ij}^3,

$$N(\widehat{\gamma}_\epsilon^3 - \gamma_\epsilon^3) = \sum_i \left(\sum_j \epsilon_{ij}^3 - \ell_i \gamma_\epsilon^3 - 3 \sum_j \gamma_\epsilon^2 \epsilon_{ij} \right) + o_P(N^{1/2}). \tag{A4}$$

17

Finally, the conclusion for $\widehat{\gamma}_\epsilon^3$ now readily follows from the Central Limit Theorem and Theorem 3 is proved. $\qquad\square$

Proof of Theorem 4. For $\widehat{\gamma}_b^4$ we need arguments similar to before. Using the results of Theorem 1 we neglect again all terms that contain $(x_{ij} - \overline{x}_{i\bullet})^\top(\widehat{\beta} - \beta)$. Then, replacing $\widehat{v}_{ij}$ with (A2), we get that the expression to study, after rearranging terms, is

$$
n\widehat{\gamma}_b^4 = \sum_i \{b_i - (\overline{b} + \overline{\epsilon})\}^4 + \sum_i \frac{4}{\ell_i} \sum_j \{b_i - (\overline{b} + \overline{\epsilon})\}^3 \epsilon_{ij} + \sum_i \frac{2}{\ell_i} \sum_j \{b_i - (\overline{b} + \overline{\epsilon})\}^2 \epsilon_{ij}^2
$$
$$
+ \sum_i \frac{8}{\ell_i(\ell_i - 1)} \sum_{j \neq h} \{b_i - (\overline{b} + \overline{\epsilon})\}^2 \epsilon_{ij}\epsilon_{ih} + \sum_i \frac{8}{\ell_i(\ell_i - 1)} \sum_{j \neq h} \{b_i - (\overline{b} + \overline{\epsilon})\} \epsilon_{ij}^2 \epsilon_{ih}
$$
$$
- \sum_i \frac{2}{\ell_i(\ell_i - 1)} \sum_{j \neq h} \epsilon_{ij}^2 \epsilon_{ih}^2 - 2\widehat{\gamma}_b^2\widehat{\gamma}_\epsilon^2 - (\widehat{\gamma}_\epsilon^2)^2 + o_P(n^{1/2}).
$$

Based on the previous results for $\widehat{\gamma}_\epsilon^2$ and $\widehat{\gamma}_b^2$, it is easy to show $\widehat{\gamma}_\epsilon^2 \overset{p}{\to} \gamma_\epsilon^2$ and $\widehat{\gamma}_b^2 \overset{p}{\to} \gamma_b^2$. Using these results and the assumptions of the theorem, after rearranging terms, we find that the equation to analyze is

$$
n(\widehat{\gamma}_b^4 - \gamma_b^4) = \sum_i [\{b_i - (\overline{b} + \overline{\epsilon})\}^4 - \gamma_b^4] + o_P(n^{1/2}) = \sum_i (b_i^4 - \gamma_b^4 - 4\gamma_b^3 b_i) + o_P(n^{1/2}) \tag{A5}
$$

and by the Central Limit Theorem the first part of this theorem is proved.

Finally, for $\widehat{\gamma}_\epsilon^4$, it can be shown that

$$
N\widehat{\gamma}_\epsilon^4 = \sum_i \sum_j \epsilon_{ij}^4 - \sum_{i=1}^n \left(\frac{4\ell_i^2 - 12\ell_i + 9}{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3} \right) \sum_{j \neq h} \epsilon_{ij}^3 \epsilon_{ih} + \sum_i \left(\frac{6\ell_i - 9}{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3} \right) \sum_{j \neq h} \epsilon_{ij}^2 \epsilon_{ih}^2
$$
$$
+ \sum_i \left(\frac{6\ell_i - 9}{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3} \right) \sum_{j \neq h \neq g} \epsilon_{ij}^2 \epsilon_{ih} \epsilon_{ig} - \sum_i \left(\frac{3}{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3} \right) \sum_{j \neq h \neq g \neq r} \epsilon_{ij} \epsilon_{ih} \epsilon_{ig} \epsilon_{ir}
$$
$$
- \sum_i \left\{ \frac{\ell_i(\ell_i - 1)(6\ell_i - 9)}{\ell_i^3 - 4\ell_i^2 + 6\ell_i - 3} \right\} (\widehat{\gamma}_\epsilon^2)^2 + o_P(N^{1/2})
$$
$$
= \sum_i \sum_j (\epsilon_{ij}^4 - 4\gamma_\epsilon^3 \epsilon_{ij}) + o_P(N^{1/2}).
$$

Then, after centering we get

$$
N(\widehat{\gamma}_\epsilon^4 - \gamma_\epsilon^4) = \sum_i \sum_j (\epsilon_{ij}^4 - \gamma_\epsilon^4 - 4\gamma_\epsilon^3 \epsilon_{ij}) + o_P(N^{1/2})
$$

and the proof of Theorem 4 follows from the Central Limit Theorem. $\qquad\square$

Proof of Theorem 5. Following the standard proof scheme as in [1], we first focus on the behavior of $\widehat{s}_\epsilon$ and later on the properties of $\widehat{s}_b$. From (8), we obtain

$$
\widehat{s}_\epsilon - s_\epsilon = \frac{\widehat{\gamma}_\epsilon^3 - \gamma_\epsilon^3}{(\widehat{\gamma}_\epsilon^2)^{3/2}} - s_\epsilon \left\{ \frac{(\widehat{\gamma}_\epsilon^2)^{3/2} - (\gamma_\epsilon^2)^{3/2}}{(\widehat{\gamma}_\epsilon^2)^{3/2}} \right\}. \tag{A6}
$$

Also, by the delta method we know that

$$
\sqrt{N} \{(\widehat{\gamma}_\epsilon^2)^{k/2} - (\gamma_\epsilon^2)^{k/2}\} = \frac{k}{2} (\gamma_\epsilon^2)^{k/2 - 1} \sqrt{N} (\widehat{\gamma}_\epsilon^2 - \gamma_\epsilon^2) + o_P(1). \tag{A7}
$$

By inserting (A4) and (A7), for $k = 3$, into (A6) we get

$$
\widehat{s}_\epsilon - s_\epsilon = \frac{1}{(\widehat{\gamma}_\epsilon^2)^{3/2}} \left\{ \frac{1}{N} \sum_{ij} (\epsilon_{ij}^3 - \gamma_\epsilon^3) - \frac{3\gamma_\epsilon^2}{N} \sum_{ij} \epsilon_{ij} \right\} - \frac{3 s_\epsilon \gamma_\epsilon}{2(\widehat{\gamma}_\epsilon^2)^{3/2}} \left\{ \frac{1}{N} \sum_{ij} (\epsilon_{ij}^2 - \gamma_\epsilon^2) \right\} + o_P\left(\frac{1}{\sqrt{N}} \right)
$$

and rearranging the expression to analyze leads to

$$\sqrt{N}\,(\widehat{s}_\epsilon - s_\epsilon) = \frac{\alpha_\epsilon^\top}{(\widehat{\gamma_\epsilon^2})^{3/2}}\,\frac{1}{\sqrt{N}}\sum_i\sum_j s_{ij}^\epsilon + o_P(1),$$

where $\alpha_\epsilon = (1, -3\gamma_\epsilon^2, -3s_\epsilon\gamma_\epsilon/2)^\top$ and $s_{ij}^\epsilon = (\epsilon_{ij}^3 - \gamma_\epsilon^3, \epsilon_{ij}, \epsilon_{ij}^2 - \gamma_\epsilon^2)^\top$ are 3×1 vectors.

Under the assumptions of Theorem 5, it is easy to show that $\mathrm{E}(\sum_{ij} s_{ij}^\epsilon/\sqrt{N}) = 0$ and its variance-covariance matrix is of the form

$$\Gamma_\epsilon = \left[\begin{array}{ccc} \gamma_\epsilon^6 - (\gamma_\epsilon^2)^3 & \gamma_\epsilon^4 & \gamma_\epsilon^5 - \gamma_\epsilon^3\gamma_\epsilon^2 \\ \gamma_\epsilon^4 & \gamma_\epsilon^2 & \gamma_\epsilon^3 \\ \gamma_\epsilon^5 - \gamma_\epsilon^3\gamma_\epsilon^2 & \gamma_\epsilon^3 & \gamma_\epsilon^4 - (\gamma_\epsilon^2)^2 \end{array}\right]. \tag{A8}$$

Using the result of Theorem 2, $\widehat{\gamma_\epsilon^2} \overset{p}{\longrightarrow} \gamma_\epsilon^2$, and by the Central Limit Theorem the first part of this theorem is proved.

Similarly, $\widehat{s}_b$ can be written as

$$\sqrt{n}(\widehat{s}_b - s_b) = \frac{\alpha_b^\top}{(\widehat{\gamma_b^2})^{3/2}}\,\frac{1}{\sqrt{n}}\sum_i s_i^b + o_P(1),$$

where $\alpha_b = (1, -3\gamma_b^2, -3s_b\gamma_b/2)^\top$ and $s_i^b = (b_i^3 - \gamma_b^3, b_i, b_i^2 - \gamma_b^2)^\top$ are 3×1 vectors. Again, under the assumptions of Theorem 5 $\mathrm{E}(\sum_{i=1}^n s_i^\epsilon/\sqrt{n}) = 0$ whereas its variance-covariance matrix is

$$\Gamma_b = \left[\begin{array}{ccc} \gamma_b^6 - (\gamma_b^2)^3 & \gamma_b^4 & \gamma_b^5 - \gamma_b^3\gamma_b^2 \\ \gamma_b^4 & \gamma_b^2 & \gamma_b^3 \\ \gamma_b^5 - \gamma_b^3\gamma_b^2 & \gamma_b^3 & \gamma_b^4 - (\gamma_b^2)^2 \end{array}\right].$$

This concludes the proof of Theorem 5. $\qquad\square$

Proof of Theorem 6. Focus on the asymptotic properties of the kurtosis statistics and insert (A6) and (A8), for $k = 4$, into (9). Then, $\widehat{\kappa}_\epsilon$ can be written as

$$\widehat{\kappa}_\epsilon - \kappa_\epsilon = \frac{1}{(\widehat{\gamma_\epsilon^2})^2}\left\{\frac{1}{N}\sum_{ij}(\epsilon_{ij}^4 - \gamma_\epsilon^4) - \frac{4\gamma_\epsilon^3}{N}\sum_{ij}\epsilon_{ij}\right\} - \frac{2\kappa_\epsilon\gamma_\epsilon^2}{(\widehat{\gamma_\epsilon^2})^2}\left\{\frac{1}{n}\sum_{ij}(\epsilon_{ij}^2 - \gamma_\epsilon^2)\right\} + o_P\left(\frac{1}{\sqrt{N}}\right)$$

and rearranging terms leads to

$$\sqrt{N}\,(\widehat{\kappa}_\epsilon - \kappa_\epsilon) = \frac{\beta_\epsilon^\top}{(\widehat{\gamma_\epsilon^2})^2}\,\frac{1}{\sqrt{N}}\sum_i\sum_j \kappa_{ij}^\epsilon + o_P(1),$$

where $\beta_\epsilon = (1, -4\gamma_\epsilon^3, -2\kappa_\epsilon\gamma_\epsilon^2)^\top$ and $\kappa_{ij}^\epsilon = (\epsilon_{ij}^4 - \gamma_\epsilon^4, \epsilon_{ij}, \epsilon_{ij}^2 - \gamma_\epsilon^2)^\top$ are 3×1 vectors.

Again, under the assumptions of Theorem 6 it is straightforward to show that $\mathrm{E}(\sum_{ij}\kappa_{ij}^\epsilon/\sqrt{N}) = 0$ and its variance-covariance matrix is

$$\Omega_\epsilon = \left[\begin{array}{ccc} \gamma_\epsilon^8 - (\gamma_\epsilon^4)^2 & \gamma_\epsilon^5 & \gamma_\epsilon^6 - \gamma_\epsilon^4\gamma_\epsilon^2 \\ \gamma_\epsilon^5 & \gamma_\epsilon^2 & \gamma_\epsilon^3 \\ \gamma_\epsilon^6 - \gamma_\epsilon^4\gamma_\epsilon^2 & \gamma_\epsilon^3 & \gamma_\epsilon - (\gamma_\epsilon^2)^2 \end{array}\right].$$

Then, using the results of Theorem 4 and the Central Limit Theorem, we can conclude the proof of the first part of the theorem.

Finally, focusing on the behavior of $\widehat{\kappa}_b$ and inserting (A5) and (A8), for $k = 4$, into (9), we find that the expression to analyze in closed form is

$$\sqrt{n}\,(\widehat{\kappa}_b - \kappa_b) = \frac{\beta_b^\top}{(\widehat{\gamma_b^2})^2}\,\frac{1}{\sqrt{n}}\sum_i \kappa_i^b + o_P(1),$$

where $\beta_b = (1, -4\gamma_b^3, -2\kappa_b\gamma_b^2)^\top$ and $\kappa_i^b = (b_i^4 - \gamma_b^4, b_i, b_i^2 - \gamma_b^2)^\top$ are 3×1 vectors.

As before, under the assumptions of this theorem, we have $E(\sum_i \kappa_i^b / \sqrt{n}) = 0$ whereas its variance matrix is of the form

$$\Omega_b = \begin{bmatrix} \gamma_b^8 - (\gamma_b^4)^2 & \gamma_b^5 & \gamma_\epsilon^6 - \gamma_\epsilon^4\gamma_\epsilon^2 \\ \gamma_b^5 & \gamma_b^2 & \gamma_b^3 \\ \gamma_b^6 - \gamma_b^4\gamma_b^2 & \gamma_b^3 & \gamma_b - (\gamma_b^2)^2 \end{bmatrix}.$$

The conclusion for $\widehat{\kappa}_b$ follows easily from the Central Limit Theorem and Theorem 4. This completes the proof of Theorem 6. $\qquad\square$

Appendix B

In this Appendix we collect the proposed estimators for the higher order moments up to the eight moment of both the random effects and the error terms. As before, using Lemma 1 and rearranging terms the proposed estimators for the error term are of the form

$$\widehat{\gamma}_\epsilon^5 = \frac{1}{\sum_{i=1}^n \ell_i(\ell_i^4 - 5\ell_i^3 + 10\ell_i^2 - 10\ell_i + 4)} \sum_{i=1}^n \left\{ \ell_i^4 \sum_{j=1}^{\ell_i} \widetilde{\widehat{\epsilon}}_{ij}^5 - 10\ell_i(\ell_i - 1)(\ell_i - 2)^2\widehat{\gamma}_\epsilon^3\widehat{\gamma}_\epsilon^2 \right\},$$

$$\widehat{\gamma}_\epsilon^6 = \frac{1}{\sum_{i=1}^n \ell_i(\ell_i^5 - 6\ell_i^4 + 15\ell_i^3 - 20\ell_i^2 + 15\ell_i - 5)} \sum_{i=1}^n \left\{ \ell_i^5 \sum_{j=1}^{\ell_i} \widetilde{\widehat{\epsilon}}_{ij}^6 - 15\ell_i(\ell_i - 1)(\ell_i^3 - 4\ell_i^2 + 7\ell_i - 5)\widehat{\gamma}_\epsilon^4\widehat{\gamma}_\epsilon^2 \right.$$
$$\left. + 10\ell_i(\ell_i - 1)(2\ell_i^2 - 6\ell_i + 5)(\widehat{\gamma}_\epsilon^3)^2 - 15\ell_i(\ell_i - 1)(\ell_i - 2)(3\ell_i - 10)(\widehat{\gamma}_\epsilon^2)^3 \right\},$$

$$\widehat{\gamma}_\epsilon^8 = \frac{1}{\sum_{i=1}^n \ell_i(\ell_i^7 - 8\ell_i^6 + 28\ell_i^5 - 56\ell_i^4 + 70\ell_i^3 - 56\ell_i^2 + 28\ell_i - 7)} \sum_{i=1}^n \left\{ \ell_i^7 \sum_{j=1}^{\ell_i} \widetilde{\widehat{\epsilon}}_{ij}^8 \right.$$
$$- 28\ell_i(\ell_i - 1)(\ell_i^5 - 6\ell_i^4 + 15\ell_i^3 - 20\ell_i^2 + 16\ell_i - 7)\widehat{\gamma}_\epsilon^6\widehat{\gamma}_\epsilon^2 + 56\ell_i(\ell_i - 1)(\ell_i^4 - 5\ell_i^3 + 11\ell_i^2 - 13\ell_i + 7)\widehat{\gamma}_\epsilon^5\widehat{\gamma}_\epsilon^3$$
$$35\ell_i(\ell_i - 1)(2\ell_i^3 - 8\ell_i^2 + 12\ell_i - 7)(\widehat{\gamma}_\epsilon^4)^2 - 70\ell_i(\ell_i - 1)(\ell_i - 2)(3\ell_i^3 - 24\ell_i^2 + 42\ell_i - 42)\widehat{\gamma}_\epsilon^4(\widehat{\gamma}_\epsilon^2)^2$$
$$\left. + 280\ell_i(\ell_i - 1)(\ell_i - 2)(2\ell_i^2 - 7\ell_i + 14)\widehat{\gamma}_\epsilon^2(\widehat{\gamma}_\epsilon^3)^2 - 210\ell_i(\ell_i - 1)(\ell_i - 2)(\ell_i - 3)(4\ell_i - 21)(\widehat{\gamma}_\epsilon^2)^4 \right\}$$

Furthermore, the proposed estimators for the random effects are of the form

$$\widehat{\gamma}_b^5 = \frac{1}{n}\sum_{i=1}^n \frac{48}{\ell_i(\ell_i - 1)(\ell_i - 2)(\ell_i - 3)} \sum_{j=1}^{\ell_i-3}\sum_{h=j+1}^{\ell_i-2}\sum_{g=h+1}^{\ell_i-1}\sum_{r=g+1}^{\ell_i} \widehat{v}_{ij}^2\widehat{v}_{ih}\widehat{v}_{ig}\widehat{v}_{ir} - \frac{1}{n}\sum_{i=1}^n \frac{6}{\ell_i(\ell_i - 1)(\ell_i - 2)} \sum_{j=1}^{\ell_i-2}\sum_{h=j+1}^{\ell_i-1}\sum_{g=h+1}^{\ell_i} \widehat{v}_{ij}^2\widehat{v}_{ih}^2\widehat{v}_{ig},$$

$$\widehat{\gamma}_b^6 = \frac{1}{n}\sum_{i=1}^n \frac{72}{\ell_i(\ell_i - 1)(\ell_i - 2)(\ell_i - 3)} \sum_{j=1}^{\ell_i-3}\sum_{h=j+1}^{\ell_i-2}\sum_{g=h+1}^{\ell_i-1}\sum_{r=g+1}^{\ell_i} \widehat{v}_{ij}^2\widehat{v}_{ih}^2\widehat{v}_{ig}\widehat{v}_{ir}$$
$$- \frac{1}{n}\sum_{i=1}^n \frac{12}{\ell_i(\ell_i - 1)(\ell_i - 2)} \sum_{j=1}^{\ell_i-2}\sum_{h=j+1}^{\ell_i-1}\sum_{g=h+1}^{\ell_i} \widehat{v}_{ij}^2\widehat{v}_{ih}^2\widehat{v}_{ig}^2 + 3\widehat{\gamma}_b^2(\widehat{\gamma}_\epsilon^2)^2 + 2(\widehat{\gamma}_\epsilon^2)^3,$$

$$\widehat{\gamma}_b^8 = \frac{1}{n} \sum_{i=1}^{n} \frac{240}{\ell_i(\ell_i-1)(\ell_i-2)(\ell_i-3)(\ell_i-4)} \sum_{j=1}^{\ell_i-4} \sum_{h=j+1}^{\ell_i-3} \sum_{g=h+1}^{\ell_i-4} \sum_{r=g+1}^{\ell_i-1} \sum_{l=r+1}^{\ell_i} \widehat{v}_{ij}^2 \widehat{v}_{ih}^2 \widehat{v}_{ig}^2 \widehat{v}_{ir} \widehat{v}_{il}$$

$$- \frac{1}{n} \sum_{i=1}^{n} \frac{24}{\ell_i(\ell_i-1)(\ell_i-2)(\ell_i-3)} \sum_{j=1}^{\ell_i-3} \sum_{h=j+1}^{\ell_i-2} \sum_{g=h+1}^{\ell_i-1} \sum_{r=g+1}^{\ell_i} \widehat{v}_{ij}^2 \widehat{v}_{ih}^2 \widehat{v}_{ig}^2 \widehat{v}_{ir}^2 - 2\widehat{\gamma}_b^6 \widehat{\gamma}_\epsilon^2 + 2\widehat{\gamma}_b^2 (\widehat{\gamma}_\epsilon^2)^3 + (\widehat{\gamma}_\epsilon^2)^4.$$

Under the assumptions of Theorems 5 and 6, it is straightforward to show that $\widehat{\gamma}_\epsilon^k \xrightarrow{p} \gamma_\epsilon^k$ and $\widehat{\gamma}_b^k \xrightarrow{p} \gamma_b^k$ for $k \in \{2, 6\}$.

References

[1] J. Bai, S. Ng, Tests for skewness, kurtosis, and normality for time series data, J. Bus. Econom. Statist. 23 (2005) 49–60.

[2] O. Barndorff-Nielsen, On a formula for the distribution of the maximum likelihood estimator, Biometrika 70 (1983) 343–365.

[3] D.R. Cox, P. Hall, Estimation in a simple random effects model with nonnormal distributions, Biometrika 89 (2002) 831–840.

[4] A. Galvao, G. Montes-Rojas, W. Sosa-Escudero, L. Wang, Tests for skewness and kurtosis in the one-way error component model, J. Multivariate Anal. 122 (2013) 35–52.

[5] D.A. Harville, Bayesian inference for variance components using only error contrasts, Biometrika 61 (1974) 383–385.

[6] J. Jiang, REML estimation: Asymptotic behavior and related topics, Ann. Statist. 25 (1996) 255–286.

[7] J. Jiang, Goodness-of-fit tests for mixed model diagnostics, Ann. Statist. 29 (2001) 1137–1164.

[8] J. Jiang, Empirical method of moments and its applications, J. Statist. Plann. Inf. 115 (2003) 69–84.

[9] G. Verbeke, E. Lesaffre, A linear mixed-effects model with heterogeneity in the random effects population, J. Amer. Statist. Assoc. 91 (1996) 217–221.

[10] G. Verbeke, E. Lesaffre, The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data, Comput. Statist. Data Anal. 23 (1997) 541–556.

[11] G. Verbeke, G. Molenberghs, The gradient function as an explanatory goodness-of-fit assessment of the random-effects distribution in mixed models, Biostatistics 14 (2013) 477–790.

[12] T.J. Wansbeek, A. Kapteyn, A simple way to obtain the spectral decomposition of variance components models for balanced data, J. Econom. 41 (1989) 341–361.

[13] P. Wu, L. Zhu, An orthogonality-based estimation of moments for linear mixed models, Scand. J. Statist. 37 (2010) 253–263.

[14] P. Wu, W. Stute, L. Zhu, Efficient estimation of moments in linear mixed models, Bernoulli 18 (2012) 206–228.

[15] R.D. Ye, T. Wang, A.K. Gupta, Distribution of matrix quadratic forms under skew-normal settings, J. Multivariate Anal. 131 (2014) 229–239.

[16] D. Zhang, M. Davidian, Linear mixed models with flexible distributions of random effects for longitudinal data, Biometrics 57 (2001) 795–802.