A comparative study on estimation methods to deal with the endogeneity in linear random-intercept models with an extension

DOI: 10.1080/00949655.2016.1196689
Handle: http://hdl.handle.net/1942/22836

A comparative study on estimation methods to deal with the endogeneity in linear random-intercept models with an extension

Reyhaneh Rikhtehgaran[a,*], Iraj Kazemi[b], Geert Verbeke[c]

[a] *Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran*
[b] *Department of Statistics, College of Science, University of Isfahan, Isfahan, Iran*
[c] *I-BioStat, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

Abstract

In this paper, we investigate estimation methods to deal with situations where random intercepts are associated to time-varying covariates in the context of linear mixed models. First, a review of previous ways to deal with this so-called endogeneity issue is present, then a new method based on shared random effects is proposed. Simulation studies and an empirical example are utilized to assess the performance of our proposed method. It is shown that our new approach is more efficient than most competitors and is robust to the misspecification of the random-effects distributions.

Keywords: Endogenous covariates, Fixed-effect approach, Longitudinal data, Mixture inference, Random-effect approach.

1. Introduction

In the analysis of longitudinal data using mixed-effects models, the main objective of inference is the estimation of longitudinal effects. These effects refer to changes over time within subjects, versus cross-sectional effects which indicate changes between subjects. The foundation of various estimation methods are usually based on making different assumptions for both cross-sectional and longitudinal effects. Using näive assumptions on the cross-sectional effects, specially on random intercepts, leads to the model misspecification and can highly influence the longitudinal inference.

The usual assumptions for the random intercepts are mainly on the distributional forms and on the independence of these effects with the covariates. Violation of the first assumption in fitting mixed models does not have large impact on the estimation of fixed effects (e.g., Neuhaus et al., 1992; Verbeke and Lesaffre, 1997) though it has sensible effects on the inferences of random effects (e.g., Verbeke and Lesaffre, 1996; McCulloch and Neuhaus, 2011). Violation of the second assumption is critical as illustrated by Palta

*Corresponding author at Department of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.
E-mail address: r_rikhtehgaran@cc.iut.ac.ir

and Yao (1991). Verbeke et al. (2001) show that the violation of this assumption can produce biases for the estimates of longitudinal effects. In the case of having some time-varying covariates, violation of this assumption is more likely to happen. This is because these covariates may be specifically stochastic and their impact time dependent. Thus, they may cause non-zero correlation with both random-effects and the error terms. In these cases, the covariates are so called endogenous in contrast with exogenous covariates that are uncorrelated with both effects and errors. According to Wooldridge (2010), there are several reasons for the generation of these correlations. Some examples include the effect of unmeasured confounding, measurement errors and the reverse causality. Details of the last two reasons are given, for example, by Diggle et al. (2002) and Wooldridge (2010).

In this paper, we focus on addressing the misspecification issue related to the independence between random intercepts and time-varying covariates that can happen because of unmeasured or omitted time-invariant covariates. The effect of omitted time-varying covariates is not our concern. To see the effect of omitted time-varying covariates one can see, for example, Palta and Yao (1991) and Wooldridge (2010).

There have been lots of efforts to handle this kind of endogeneity issue in both econometrics and biostatistics contexts for random-intercept modeling. There are two associated approaches for solving the endogeneity problem, each with advantages and shortcomings. The first one is called the fixed-effect approach and utilizes two strategies to solve the problem by treating the intercepts as fixed effects and by applying some transformations to the model for eliminating the cross-sectional effects. The second one is the random-effect approach which deals with the issue by explicitly modeling the underlying correlation.

In this paper, we combine some estimation methods suggested in the literature on modeling random intercepts with mixture concepts in fitting general linear mixed models. The motivation is originated a work done by Verbeke et al. (2001) who combine the conditional inference with mixture inference to avoid the influence of misspecifying cross-sectional effects in linear mixed models. First, a review of possible solutions for the endogeneity issue in random intercept models is provided. Then, an extension is presented to the case of general linear mixed models. Further, a new shared random-effect method will be proposed to model the correlation between random intercepts and covariates. Four simulation studies are conducted to assess the performance of the proposed method. Also, the results are confirmed by an empirical study.

The remainder of the paper is organized as follows. In Section 2, we specify the general longitudinal data model and address the endogeneity issue. Mixture inference is concisely introduced in this section. In Section 3, a brief introduction to the fixed-effect approach to deal with the endogeneity is present. Section 4 includes some possible random-effect methods proposed so far to handle the endogeneity with the introduction of a new method. Section 5 presents four simulation studies to investigate the performance of our proposed method. In Section 6, by the analysis of a real data set, we show the usefulness of the proposed method. The last section includes concluding remarks.

2. Model specification

Let y_{it} denotes the response of the t-th measurement ($t = 1, \cdots, T_i$) taken on the i-th individual ($i = 1, \cdots, n$). Consider the general linear mixed model

$$\mathbf{y}_i = \mathbf{X}_i\theta + \mathbf{Z}_i\zeta_i + \varepsilon_i, \tag{1}$$

where $\mathbf{y}_i = (y_{i1}, \cdots, y_{iT_i})'$, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $T_i \times p$ and $T_i \times q$ known design matrices which include both time-varying and time-invariant covariates, θ is a vector of regression coefficients, the ζ_i are random effects and $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{iT_i})'$ is a vector of error terms. The usual assumptions are that the parametric shape of underlying distributions for ε_i and ζ_i are known, having zero means and constant variances, and being mutually independent and uncorrelated with the covariates. We assume $\varepsilon_i \overset{\text{ind}}{\sim} N\left(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}_{T_i}\right)$, $Var\left(\zeta_i\right) = \mathbf{D}$ and ζ_i's are mutually independent and uncorrelated with the error terms. Thus, the marginal covariance matrix of $\mathbf{y}_i$ is of the form $\mathbf{V}_i = Var\left(\mathbf{y}_i\right) = \sigma_\varepsilon^2\mathbf{I}_{T_i} + \mathbf{Z}_i\mathbf{D}\mathbf{Z}_i'$. It is shown by Verbeke and Lesaffre (1997) that the assumptions related to the distribution of the random effects do not have severe effects on the estimation of fixed effects. In contrast, violation of the independence assumption between the random effects and the covariates is serious (Palta and Yao, 1991). In this paper, we assume only that the random slopes of ζ_i are independent of time-varying covariates.

While the time-invariant covariates can be endogenous as well, we focus on a more general case of endogenous time-varying covariates. Time-invariant covariates are special cases of the time-varying covariates.

In order to separate the impact of cross-sectional from the longitudinal effects, we rewrite Equation (1) as

$$\mathbf{y}_i = \mathbf{X}_i^{(1)}\theta^{(1)} + \mathbf{X}_i^{(2)}\theta^{(2)} + \mathbf{Z}_i^{(1)}\zeta_i^{(1)} + \mathbf{Z}_i^{(2)}\zeta_i^{(2)} + \varepsilon_i, \tag{2}$$

where $\mathbf{X}_i^{(1)}$ is the $T_i \times p_1$ matrix of time-invariant covariates, $\mathbf{X}_i^{(2)}$is the $T_i \times p_2$ matrix of time-varying covariates, $\zeta_i^{(1)}$ and $\zeta_i^{(2)}$ denote the random intercepts and random slopes, respectively, $\mathbf{Z}_i^{(1)} = \mathbf{1}_{T_i}$ and $\mathbf{Z}_i^{(2)}$ is equal to the $T_i \times (q-1)$ matrix of time-varying covariates corresponding to random slopes. We assume that the $\zeta_i^{(2)}$ are independent of the design matrices while the random intercepts $\zeta_i^{(1)}$ are correlated with some columns in those design matrices. In this case, we assume that there are some unmeasured or omitted time-invariant covariates, expressed by the columns of $\mathbf{W}_i$, which are correlated with the outcome and some of the covariates. Since the random intercepts $\zeta_i^{(1)}$ represent the effect of all additive unobserved subject-level covariates associated with the outcome, we assume $\zeta_i^{(1)} = \gamma'\mathbf{W}_i + \widetilde{\zeta}_i^{(1)}$, where the $\widetilde{\zeta}_i^{(1)}$ denote the rest of unobserved subject-level covariates that are uncorrelated with the design matrices. Substituting this relation in Equation (2) leads to the following model

$$\mathbf{y}_i = \mathbf{X}_i^{(1)}\theta^{(1)} + \mathbf{X}_i^{(2)}\theta^{(2)} + \gamma'\mathbf{W}_i + \mathbf{Z}_i^{(1)}\widetilde{\zeta}_i^{(1)} + \mathbf{Z}_i^{(2)}\zeta_i^{(2)} + \varepsilon_i. \tag{3}$$

For simplicity, from now on we assume a linear mixed model with a single time-varying covariate x_{it} of the form

$$y_{it} = \beta_0 + \beta_1 x_{it} + b_i x_{it} + \alpha_i + \varepsilon_{it}, i = 1, \cdots, n, t = 1, \cdots, T, \tag{4}$$

where the intercepts α_i and the covariates x_{it} are dependent and the slopes b_i are independent of x_{it}. It is assumed that there is an omitted covariate w_i which is dependent with x_{it}. Rewriting $\alpha_i = \gamma w_i + \widetilde{\alpha}_i$, where $\widetilde{\alpha}_i$ and x_{it} are independent, the model can be expressed as

$$y_{it} = \beta_0 + \beta_1 x_{it} + b_i x_{it} + \gamma w_i + \widetilde{\alpha}_i + \varepsilon_{it}. \tag{5}$$

Following Palta and Yao (1991) it can be shown for normally distributed random variables that, if the variable w_i is omitted from the model then both the mean and variance of the marginal model will change. In fact, we have

$$E\left(\mathbf{y}_i|\mathbf{x}_i\right) = \beta_0 \mathbf{1}_T + \beta_1 \mathbf{x}_i + B\overline{x}_i \mathbf{1}_T,\ i = 1, \cdots, n, \tag{6}$$

where $B{=}\psi\gamma r\sigma_w/\tau_B$, σ_w denotes the standard deviation of w_i, $\psi = T\tau_B^2/\left(T\tau_B^2 + \tau_W^2\right)$, where $\tau_B^2 = E\left(\mu_{\mathbf{x}_i} - \mu_{\mathbf{x}}\right)^2$ and $\tau_W^2 = E\left(x_{it} - \mu_{\mathbf{x}_i}\right)^2$ are, respectively, variances related to between and within variations of x_{it}, where $\mu_{\mathbf{x}_i}$ and $\mu_{\mathbf{x}}$ are subject and total means of the covariate, respectively, and $r = corr\left(w_i, \mu_{\mathbf{x}_i}\right)$. We also have

$$Var\left(\mathbf{y}_i|\mathbf{x}_i\right) = C\mathbf{J} + \mathbf{V}_i, \tag{7}$$

where $C{=}\sigma_w^2\gamma^2\left(1 - \psi r^2\right)$ and $\mathbf{J} = \mathbf{1}_T\mathbf{1}_T'$. Therefore, the omission of covariate w_i from the model affects both mean and variance of the marginal model introduced by Equations (6) and (7). Equivalently, it makes dependence of the random intercept with some covariates in Equation (4). Ignoring these facts leads to biased estimation of the coefficients of the endogenous covariates. We illustrate this fact in Section 5 using simulation studies.

It should be noted that although the endogeneity problem, due to the non-zero correlation between time-varying covariates and random-slopes, is an important case, the nature of the problem and suggested solutions are different. Most solutions in this case are based on the introduction of instrumental variables (see, e.g. ??). In this type of endogeneity, one may assume that $b_i = \alpha_0 + \alpha_1 w_i + \widetilde{b}_i$, where x_{it} is independent of $\widetilde{b}_i$ but is dependent with w_i. Indeed, the endogeneity occurs when the important interaction $w_i * x_{it}$ is omitted from the model. However, there are many situations where just an important time-invariant covariate is omitted and not its interaction term with x. Therefore, in this paper, we assume that the random slopes b_i's are independent of x.

2.1. The mixture inference

The mixture inference is commonly applied in the context of linear mixed models to obtain the estimates of longitudinal effects. In the mixture approach, a mixing distribution is utilized for the random effects α_i and b_i. The marginal density of observation $\mathbf{y}_i$ is achieved by integrating out the random effects yielding the mixture density

$$\mathbf{L}_i = \int f\left(\mathbf{x}_i, \mathbf{y}_i|\alpha_i, b_i\right) dG\left(\alpha_i, b_i\right). \tag{8}$$

Then, the likelihood function of model parameters using all observations is maximized to estimate the parameters β_0, β_1 and σ_ε^2. It is mentioned in the literature that the role of the mixing distribution G is not important for inferences about the fixed-effect parameters (Verbeke and Lesaffre, 1997) as long as the underlying model assumptions are

correctly specified. We illustrate later that the violation of these assumptions changes the effect of the mixing distribution when estimating the fixed effects. In addition, the role of the mixing distribution in inference on the random effects appears meaningful in practical applications (e.g. Austin et al., 2003; Austin et al., 2004). In fact, incorporating misspecified mixing distribution can invalidate inferences about the random-effects (Verbeke and Lesaffre, 1996).

In the next two sections, we introduce briefly two general approaches to deal with the endogeneity issue and then combine the ideas with mixture inference.

3. The fixed-effect approach

Classical techniques to handle the endogeneity issue in random intercept models treat the effects α_i's as fixed and then use a likelihood conditional on these effects, or apply certain transformations of the model to eliminate the effect of the α_i's. Then, frequently used estimation methods, such as ordinary least square, usually yield consistent estimates of the longitudinal effects under mild assumptions. One can also apply these techniques to remove the influence of random intercepts in general linear mixed models and apply the mixture approach to estimate the longitudinal effects consistently. The main advantage of this in comparison to the random-effect approach, introduced later, is to obtain estimates which are more robust with respect to misspecifications of the cross-sectional components in the model. This is important since no extra assumption is imposed on the random-intercepts distribution nor the association between random intercepts and covariates. A shortcoming is that working with likelihood conditional on α_i's or applying transformations to remove the random intercepts, eliminates the between-cluster variability which because of the endogeneity issue, contains information about the longitudinal effects. Therefore, this approach produces larger variances of estimates in comparison to those in the random-effect approach. Another shortcoming of using this approach relates to its disability to estimate cross-sectional effects. Similarly, for the time-varying covariates with having slow changes over time, the corresponding within-cluster variation becomes low and thus by applying a transformation to remove the cross-sectional effects, the between-cluster variation of these variables would also be eliminated. Therefore, the effect of variables vanishes in the model fitting process which results in imprecise estimates with large standard errors (Plumper and Troeger, 2007). Results of several published papers (e.g., Hausman-Taylor, 1981; Plumper and Troeger, 2007; Breusch et al., 2011) suggest combining the information of cross-sectional and longitudinal effects to deal with these drawbacks.

Moreover, the estimation of the correlation between cross-sectional effects and time-varying covariates is important in some applications (Ashenfelter and Rouse, 1998). This aim cannot be achieved in using the fixed-effect approach. In the next subsections, we mention some proposed solutions to deal with the endogeneity issue in the framework of fixed-effects.

3.1. Least-squares dummy-variable method

A commonly used estimation method is based on the introduction of a dummy variable for each subject in mixed models to allow for the cross-sectional effects. In this method,

these effects are treated as model parameters. The model is written as

$$\mathbf{y} = \mathbf{X}\theta + \mathbf{A}\alpha + \varepsilon, \tag{9}$$

where $\mathbf{y}$ denotes the stacked vector including of all response vectors $\mathbf{y}_i$, $\mathbf{X}$ is the matrix of all design matrices $\mathbf{x}_i$, and ε and α are vectors of all error terms and random intercepts, respectively. The matrix $\mathbf{A}$ includes dummy variables to specify subjects and vector θ includes fixed regression parameters by noting that the intercept in Equation (9) may be removed for handling identifiability. An equivalent way to control the identifiability is to assume an intercept term in θ by imposing the restriction $\sum_{i=1}^{n} \alpha_i = 0$.

An advantage of this method is the ability of estimating both cross-sectional and longitudinal effects. A drawback comes in practice when the number of subject levels becomes large, leading to the incidental-parameter problem (Lancaster, 2000) and inconsistency of model parameters. It can easily be shown that the least-squares dummy-variable estimates of the longitudinal effects are the same as those obtained from the within-subject approach to be introduced later. Extension of this method to the general linear mixed models is straightforward by adding the term $\mathbf{Z}^{(2)}\mathbf{A}^*\mathbf{b}$ to Equation (9), where the matrix $\mathbf{A}^*$ includes dummy variables for specific slopes, $\mathbf{Z}^{(2)}$ is the matrix of all covariates and $\mathbf{b}$ is the vector of all slopes.

3.2. Time-difference method

A simple way to eliminate the cross-sectional effects in the analysis of longitudinal data is to use a transformation based on time differences. Applying this to Equation (4) leads to

$$\Delta_k\left(y_{it}\right) = \Delta_k\left(x_{it}\right)\beta_1 + \Delta_k\left(x_{it}\right)b_i + \Delta_k\left(\varepsilon_{it}\right), \tag{10}$$

where for a fixed k, the time-difference operator $\Delta_k\left(u_{it}\right) = u_{it} - u_{it-k}$. One may use any order of time-differences in specific applications. By removing the random intercepts, we can obtain consistent estimates of the longitudinal effects while these estimates are less efficient than the other methods of fixed-effect approach introduced here. This is because much information is lost by applying the operator for $T > 2$. In fact, we are only able to use information of $N(T-1)$ observations for making inference. Moreover, the differences $\Delta_k\left(\varepsilon_{it}\right)$'s are serially correlated which requires using suitable estimation methods (Cameron and Trivedi, 2005, ch. 21).

3.3. Within-subject method

Another way to eliminate the subject-specific effects is to use the within transformation method which leads to the following model

$$y_{it} - \overline{y}_i = \beta_1\left(x_{it} - \overline{x}_i\right) + b_i\left(x_{it} - \overline{x}_i\right) + \left(\varepsilon_{it} - \overline{\varepsilon}_i\right). \tag{11}$$

By removing the random intercepts and by applying the mixture approach, we achieve consistent estimates of longitudinal effects with more efficiency than with the time-difference method. The specification of this model is closed to the conditional inference method described below.

3.4. Conditional inference

A pragmatic technique for making inference on the longitudinal effects, regardless of making any assumption on the cross-sectional effects, is the application of conditional inference (Verbeke et al., 2001). In this technique, the cross-sectional effects are removed from the likelihood by conditioning on their corresponding sufficient statistics. In linear models, this can be shown to be identical to choosing a transformation of the model such that cross-sectional effects vanish from the likelihood. To do this, a full rank $T \times (T-1)$ matrix $\mathbf{A}$ with restrictions $\mathbf{A}'\mathbf{1}_T = 0$ and $\mathbf{A}'\mathbf{A} = \mathbf{I}_{T-1}$ is used. Applying this transformation to Equation (4) leads to a new linear mixed model with transformed observations $\mathbf{y}_i^* = \mathbf{A}'\mathbf{y}_i$ and $\mathbf{x}_i^* = \mathbf{A}'\mathbf{x}_i$. The new model includes the original fixed and random longitudinal effects while the cross-sectional fixed and random effects are removed from the model and the residual variance remains unchanged. Verbeke et al. (2001) show that inference does not change when using different transformations as long as they satisfy the above restrictions. Furthermore, in a Bayesian framework they illustrate that no information is lost about the longitudinal effects if nothing is known about the random intercepts. In other words, if random intercepts are assumed to be independent of other parameters and also a flat distribution is considered as prior, then working with transformed observations is sufficient for making inference about longitudinal effects. But the same as other fixed-effects methods, in the case of having endogenous covariates, this transformation removes the between-cluster variability which includes information about the longitudinal effects. Therefore, the estimates have larger variances.

4. The random-effect approach

The correlations between covariates and random effects may be expressed by incorporating these measures in the distribution of random effects (Neuhaus and McCulloch, 2006). Ignoring these correlations is equivalent to the misspecification of random-effects distributions. The likelihood of parameters for subject i is given by the mixture density

$$\mathbf{L}_i = \int f(\mathbf{x}_i, \mathbf{y}_i|\alpha_i, b_i)\, dG(\alpha_i, b_i) \tag{12}$$

$$= \int f(\mathbf{y}_i|\mathbf{x}_i, \alpha_i, b_i)\, f(\mathbf{x}_i|\alpha_i, b_i)\, dG(\alpha_i, b_i) \tag{13}$$

$$\propto \int f(\mathbf{y}_i|\mathbf{x}_i, \alpha_i, b_i)\, dG^*(\alpha_i, b_i|\mathbf{x}_i), \tag{14}$$

where $G(.)$ and $G^*(.)$ denote mixing distributions. The last proportion is made due to the deletion of parameters in distribution of $\mathbf{x}_i$ from the likelihood, since these parameters are not of direct interest. We now address several solutions derived in terms of the random-effect approach using the above likelihood specifications. Some of these methods are constructed based on Equation (13) and using $f(\mathbf{x}_i|\alpha_i, b_i)$ along with $G(\alpha_i, b_i)$ to deal with the endogeneity issue. We call these the Total-Corrected (TC) methods. Other methods are based on Equation (14) and are called Mean-Corrected (MC) methods. In these methods, the relationship between α_i and $\mathbf{x}_i$ is modeled through $E(\alpha_i|\mathbf{x}_i) = R(\mathbf{x}_i)$, with $R(.)$ a function such that it projects $\mathbf{x}_i = (x_{i1}, ..., x_{iT})'$ from a T-dimensional space into a 1-dimensional space. In other words, the relationship between α_i and $\mathbf{x}_i$ is specified

based on only the expectation of α_i through a function of all covariate observations for subject i. For these methods we may consider

$$\alpha_i = R(\mathbf{x}_i) + a_i, \tag{15}$$

where the a_i's have mean zero, constant variances and are independent of x_{it}, for all i,t. Thus, by assuming Equation (15), the joint distribution $G^*(\alpha_i, b_i|\mathbf{x}_i)$ reduces to the distribution $H(a_i, b_i)$ which does not depend on $\mathbf{x}_i$.

As is shown by Palta and Yao (1991) and illustrated by Equations (6) and (7), methods not taking into account the covariance structure to deal with the endogeneity problem, are less efficient than others. This means that the fixed-effect and MC approaches are less efficient than proper TC methods.

It is noted that the validity of inferences in a random-effect approach depends on the correct specification of the correlation between α_i and x_{it}'s through their joint distribution. Therefore, a misspecified relation which leads to a misspecified distribution, can cause invalid results. This fact is the main drawback of this approach. In contrast, this approach enables inferences on both longitudinal and cross-sectional effects.

4.1. Chamberlain approach

Chamberlain (1982, 1984) introduced an MC method by assuming a function $R(\mathbf{x}_i) = \sum_{j=1}^{T} \delta_j x_{ij}$ in Equation (15). This yields

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + \sum\nolimits_{j=1}^{T} \delta_j x_{ij} + a_i + b_i x_{it} + \varepsilon_{it} && (16) \\ &= \beta_0 + \sum\nolimits_{j=1}^{T} \pi_{tj} x_{ij} + a_i + b_i x_{it} + \varepsilon_{it}, && (17) \end{aligned}$$

where the coefficients π_{tj} are elements of the matrix $\Pi = \beta_1 \mathbf{I}_T + \delta_j \mathbf{1}_T \mathbf{1}_T'$. We can then use the above equation together with the mixing distribution $H(a_i, b_i)$, instead of $G^*(\alpha_i, b_i|\mathbf{x}_i)$, which does not depend on $\mathbf{x}_i$. A major advantage of this method is that it allows a different relation between different subject levels and time-varying covariates. A serious problem however occurs in some specific applications where the number of parameters becomes large as long as the number of longitudinal observations increases. Special methods based on minimum-distance methodology (Malinvaud, 1970) have been developed to overcome the estimation problems due to the restriction $\Pi = \beta_1 \mathbf{I}_T + \delta_j \mathbf{1}_T \mathbf{1}_T'$ in the context of econometrics (e.g., Hsiao, 2003).

4.2. Between- and within-cluster covariate model

In using the MC methods, Mundlak (1978) assumed $R(\mathbf{x}_i) = \lambda \overline{x}_i$ which leads to the random intercept model

$$y_{it} = \beta_0 + \beta x_{it} + \lambda \overline{x}_i + a_i + \varepsilon_{it}, \tag{18}$$

where the a_i and the covariates are assumed to be independent. This method can be seen as a special case of the previous approach when $\delta_j = 1/T$ for all j. Neuhaus and

Kalbfleisch (1998) proposed the following decomposition to avoid inconsistency due to model misspecification

$$y_{it} = \beta_0 + \beta_W \left(x_{it} - \overline{x}_i\right) + \beta_B \overline{x}_i + a_i + \varepsilon_{it}, \tag{19}$$

where β_W and β_B measure the effects of within- and between-cluster covariates on the response expectations. They show that incorrectly assuming these effects to be equal is an endogenous issue. Then, they rewrite Equation (19) as

$$y_{it} = \beta_0 + \beta_W x_{it} + \left(\beta_B - \beta_W\right) \overline{x}_i + a_i + \varepsilon_{it}, \tag{20}$$

where the term $R\left(\mathbf{x}_i\right) = \left(\beta_B - \beta_W\right) \overline{x}_i$ is involved to handle the problem of omitted covariates. This method can be applied to the general linear mixed model as

$$y_{it} = \beta_0 + x_{it}\beta_W + \left(\beta_B - \beta_W\right) \overline{x}_i + a_i + b_i x_{it} + \varepsilon_{it}. \tag{21}$$

Similar to the other random-effects methods, a mild shortcoming of this is that it requires full specification of the mixing distribution $H\left(a_i, b_i\right)$. Like the other MC methods, the main disadvantage relates to the assumed expression showing the relation between α_i and the time-varying covariate x_{it}, which is formed only based on the expectation of α_i through the mean of the covariate over time. Therefore, the estimates are not as precise as proper TC methods. Another important point is that the within-cluster covariate effect, β_W, is estimated. Therefore, in situations where the endogenous time-varying covariate x_{it} shows little variability over time, we may simply lose the efficiency by working with within-cluster covariate instead of the original covariate.

4.3. A shared random-effect model

Neuhaus and McCulloch (2006) use a TC method by assuming the time-varying covariates being related to the random intercept α_i through a function h. This relation is expressed as $x_{it} = h\left(\alpha_i, b_i\right) + \eta_{it}$, where η_{it} is assumed independent of both α_i and b_i. They show that Equation (4) can be rewritten as

$$y_{it} = \beta_0 + \beta_1 \left(x_{it} - \overline{x}_i\right) + b_i \left(x_{it} - \overline{x}_i\right) + d_i + \varepsilon_{it}, \tag{22}$$

where $d_i = \alpha_i + \left(\beta_1 + b_i\right) \overline{x}_i = \alpha_i + \left(\beta_1 + b_i\right) \left(h\left(\alpha_i, b_i\right) + \overline{\eta}_i\right)$ is now uncorrelated with $x_{it} - \overline{x}_i$. This fact is achieved because α_i, b_i and $\overline{\eta}_i$ are uncorrelated with $\eta_{it} - \overline{\eta}_i = x_{it} - \overline{x}_i$. Using this approach, however, causes some difficulties when working with $f\left(\mathbf{x}_i | \alpha_i, b_i\right)$, since it depends on the function h. Although they proposed some choices for h and discussed their effects on the distribution of d_i, finding the true h remains still a statistical issue. The same as the between- and within-cluster covariate method, the within-cluster covariate effect is estimated with this method. Therefore, the results would be imprecise for the covariate with the slow changes over time.

4.4. An improved shared random-effect model

We now propose a shared random-effect model by specifying two groups of correlated latent subject effects. It means that if ζ_i represents the effect of all unobserved subject-level covariates associated with the outcome and ζ_i^* represents the effect of those unobserved subject-level covariates associated with the time-varying covariates, then some of

these unobserved covariates have effects on both response and covariates. Therefore, we can consider the following model

$$\begin{aligned} \mathbf{y}_i &= \mathbf{X}_i\theta + \mathbf{Z}_i\zeta_i + \varepsilon_i, & (23)\\ \widetilde{\mathbf{X}}_i &= \lambda + \zeta_i^* + \eta_i, & (24) \end{aligned}$$

where the $\widetilde{\mathbf{X}}_i$ include some columns of the design matrices that are endogenous. It is assumed that ζ_i^* and η_i are independent with zero means, $Cov\left(\zeta_i, \zeta_i^*\right) = \mathbf{\Lambda}$ and $Var\left(\eta_i\right) = \mathbf{\Sigma}$ is a diagonal matrix. Then, for subject i, the likelihood is specified as

$$\begin{aligned} \mathbf{L}_i &= \int f\left(\mathbf{x}_i, \mathbf{y}_i | \zeta_i, \zeta_i^*\right) dG\left(\zeta_i, \zeta_i^*\right) & (25)\\ &= \int f\left(\mathbf{y}_i | \mathbf{x}_i, \zeta_i\right) f\left(\mathbf{x}_i | \zeta_i^*\right) dG\left(\zeta_i, \zeta_i^*\right). & (26) \end{aligned}$$

For the simple linear mixed model (4) we can write

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + b_i x_{it} + \alpha_i + \varepsilon_{it}, & (27)\\ x_{it} &= \lambda_0 + c_i + \eta_{it}, & (28) \end{aligned}$$

where $var\left(\eta_{it}\right) = \sigma_\eta^2$ and $\sigma_{\alpha c} = cov\left(\alpha_i, c_i\right)$ is nonzero among covariance components of matrix Λ. An important feature of this model is that the variability of x_{it} is taken into account by introducing the error term η_{it}. Therefore, the estimates are expected to be more precise than those obtained from the other mentioned methods, since no information related to the longitudinal effects are removed from the model. Moreover, with this method, the total effect of the covariate, not only the within-cluster covariate effect, is estimable. A drawback of this random-effect model is that the distribution of the random effects needs to be specified. But the interesting point is that results are somehow robust to the misspecification of this distribution. This is shown with two simulation studies in Section 5. An unappealing feature is that we deal with the estimation of parameters in the endogenous covariates model which is not of direct interest.

The proposed method extends the Palta and Yao (1991) approach to the general linear mixed model. They express the endogeneity problem in the structure of compound-symmetry models with misspecified mean and variance structures due to omitted covariates. They derive a formula for the optimal compound-symmetry structure by minimizing the mean squared error (MSE) of a generalized estimating equations type estimator for the coefficient of the endogenous covariate. We here propose using latent variables and applying the mixture approach to solve the endogeneity problem. This method is investigated and compared with some of the discussed methods by means of four simulation studies in Section 5.

4.5. Instrumental-variable based methods

The instrumental variable (IV) based estimators or more generally generalized method of moments estimators, introduced by Hansen (1982), are commonly used in the econometric contexts. These methods take into account the endogeneity by introducing IVs into the estimation process. These IVs must be highly correlated with the endogenous covariates and not be correlated with error terms, subject-level effects and other covariates. Selection of proper IVs is a crucial issue in these methods. Estimation results can

be unstable in practice and the validity of the corresponding inference highly depends on the proper choice of IVs. There is a wide literature on using these methods in various models. A comprehensive literature review is given by Verbeek (2004).

5. Evaluating the proposed method

As was previously mentioned the omitted variables that are dependent on covariates affect both the mean and the variance of $\mathbf{y}_i$ given $\mathbf{x}_i$. Therefore, TC methods which take into consideration these impacts can be more efficient than the MC and the fixed-effect methods. In the following, we conduct four simulation studies to investigate the properties of the proposed method in comparison to some other solutions for the endogeneity problem. We use the conditional inference because it covers most of the fixed-effect methods. We also consider the between- and within-cluster covariate method as an MC method, because of it can be implemented easily. Furthermore, the shared-random effect is considered as a TC method. Since the performance of IV-based methods depends on the proper selection of IVs and these IVs can be different in each application, we do not include these methods in our simulations. We compare all mentioned methods with the method proposed in Section 4.4. The performance of the various methods is measured by means of biases, standard errors, and MSE values.

5.1. Simulation studies

In order to investigate the performance of the described methods, four simulation studies are conducted. The first three simulations are derived for the random-intercept model and in the fourth simulation, a general model with random slope is considered.

The random-intercept model

For each simulation, a number of 1000 data sets were simulated from the model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 w_i + a_i + \varepsilon_{it}, \tag{29}$$

for $i = 1, ..., 100$ and $t = 1, ..., 6$, where $\varepsilon_{it} \overset{\text{iid}}{\sim} N\left(0, \sigma_\varepsilon^2\right)$ and $a_i \overset{\text{iid}}{\sim} N\left(0, \sigma_a^2\right)$. To impose a correlation between x_{it} and w_i, we assume $x_{it} = \kappa_0 + \kappa_1 w_i + \eta_{it}$ where $\eta_{it} \overset{\text{iid}}{\sim} N\left(0, \sigma_\eta^2\right)$ and $w_i \overset{\text{iid}}{\sim} N\left(0, \sigma_w^2\right)$ for the first simulation. We set $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = 10$, $\kappa_0 = 0$ and $\kappa_1 = 5$. The variance components will be set later. Let Equation (29) be the true model and the variable w_i being omitted when fitting this model. In fact we fit the working model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + \varepsilon_{it}, \tag{30}$$

where the random intercept α_i, which includes the variable w_i, is assumed not to be correlated with x_{it}. Therefore, the effect of the endogenous covariate x_{it} is confounded and the estimate of β_1 expected to be biased.

According to our simulation strategy, the correlation between the time-varying covariate x_{it} and the random intercept α_i in the fitted model, depends on the ratios $\sigma_\eta^2/\kappa_1^2\sigma_w^2$ and $\sigma_a^2/\beta_2^2\sigma_w^2$, which decrease as the correlation increases. In the first simulation, we set $\sigma_\varepsilon^2 = 1$, $\sigma_a^2 = 1$ and $\sigma_w^2 = 2$ for various σ_η^2. The aim is to investigate the ability of the methods to deal with the endogeneity issue in the presence of variation among observations of the endogenous time-varying covariate. To do this, we use different values for

Table 1: Estimation results of the longitudinal effect β_1 in the random-intercept model for various σ_η^2.

Method	σ_η^2	0.01	0.05	0.1	0.5	1	1.5	3	10	30
Model (30) with	Bias	199.697	198.863	198.022	193.743	181.945	114.042	4.8891	0.9432	0.3012
$Corr(\alpha_i, x_{it}) = 0$	StD	1.550	1.568	1.593	1.822	2.200	2.786	2.5613	1.4113	0.8159
	MSE	398.837	395.514	392.177	375.449	343.954	208.178	3.2785	0.0481	0.0139
Conditional	Bias	-1.5534	-0.6947	-0.4912	-0.2197	-0.1553	-0.1268	-0.0897	-0.0491	-0.0284
	StD	44.7239	20.0011	14.1429	6.3249	4.4724	3.6517	2.5821	1.4143	0.8165
	MSE	38.9269	7.7854	3.8927	0.7785	0.3893	0.2595	0.1298	0.0389	0.0130
Proposed	Bias	-1.4240	-0.6885	-0.4858	-0.2164	-0.1534	-0.1253	-0.0889	-0.0490	-0.0283
	StD	1.5963	1.5632	1.5795	1.6854	1.7687	1.8107	1.7934	1.3252	0.8087
	MSE	16.8185	3.8018	1.9138	0.4064	0.2204	0.1589	0.0952	0.0365	0.0129
Between&	Bias	-1.5522	-0.6939	-0.4907	-0.2207	-0.1559	-0.1270	-0.0881	-0.0455	-0.0258
Within cluster	StD	44.7255	20.0044	14.1476	6.3368	4.4912	3.6766	2.6189	1.4523	0.8408
	MSE	38.9290	7.7873	3.8946	0.7805	0.3914	0.2617	0.1319	0.0401	0.0134
Shared random-	Bias	-1.5534	-0.6947	-0.4912	-0.2197	-0.1553	-0.1268	-0.0897	-0.0491	-0.0284
effects	StD	44.7239	20.0011	14.1429	6.3249	4.4724	3.6517	2.5821	1.4143	0.8165
	MSE	38.9269	7.7854	3.8927	0.7785	0.3893	0.2595	0.1298	0.0389	0.0130

Results are reported in percentages.

σ_η^2 ranging from 0.01 to 30, resulting in variations changing from 99.7% to 78.9% for the correlations between α_i and x_{it}.

Table 1 shows absolute values of biases, standard deviations and MSE's for the estimate β_1, reported in percentages. The first method considers the misspecified model (30) by assuming zero correlation between the random intercept and covariate. Other methods include conditional inference, the proposed method in Section 4.4, the between- and within-cluster covariate method and the shared random-effect method, respectively. It should be mentioned that normal distributions are assumed for both random-effects and error terms in fitting all models. Clearly, the maximum likelihood estimate of β_1 in Equation (30) is biased. These biases increase as σ_η^2 decreases which highlights the seriousness of the endogeneity problem. The same happens for the MSE values. The considerable point is that the fitting result of the conditional inference, the between- and within-cluster covariate method and the shared random-effect method are approximately similar. The reason is that as these three methods use centered covariates, $x_{it} - \overline{x}_i$, for solving the endogeneity problem and as the model being balanced, the estimates and their standard deviations are exactly the same. This fact is illustrated also in Verbeke and Fieuws (2007). Therefore, the between- and within-cluster covariate method and the shared random-effect method, by ignoring the cross-sectional part of the model, perform the same as conditional inference. Certainly the results of these three methods differ in the case of unbalanced models.

It is also seen that competitive methods used in the simulation study, by ignoring the between-cluster variability, lead to larger variances for the estimate of β_1 while the proposed method by taking this variability into consideration, results to more precise estimates. In general, the proposed random-effect method seems to perform better than other discussed methods.

In the second and the third simulation studies, we assess the performance of the proposed method in the case when the mixing distribution is misspecified. More specifically, we simulate data sets when the true distribution of w_i, in the second simulation, is

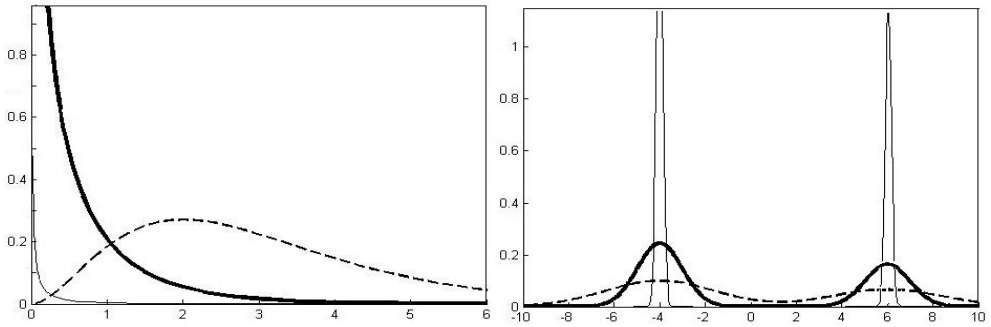

Figure 1: Curves of the Gamma distribution and the mixture of normal distributions with the specifications used for the second and the third simulation studies are depicted, respectively, in the left and the right panels. Values of 0.01, 0.5 and 3 are set for σ_w^2 and the related graphs are respectively shown by solid, bold-solid and dashed lines.

Gamma with the probability density $w_i^{\sigma_w^2-1}e^{-w_i}/\Gamma\left(\sigma_w^2\right)$, where σ_w^2 equals to both mean and variance and corresponds also to the skewness measure. In the third simulation, the true distribution of w_i is assumed to be mixture of two normal distributions, i.e. $0.4N\left(6, \frac{\sigma_w^2}{0.6^2+0.4^2}\right) + 0.6N\left(-4, \frac{\sigma_w^2}{0.6^2+0.4^2}\right)$. Other stochastic components are assumed normal in both simulations. We allow parameter σ_w^2 to vary between 0.01 to 3, implying the correlations between α_i and x_{it} will range between 31.6% and 99.2%. Other variance-component parameters are set to 1, as before. Figure 1, depicts curves related to Gamma and the mixture distributions for three values of σ_w^2=0.01, 0.5 and 3. Results including absolute values of biases, standard deviations and MSEs for the estimate of β_1 are shown, in percentages, in Tables 2 and 3, respectively, for the second and the third simulation studies. Normal distributions are assumed for both random effects and error terms in fitting all models.

Table 2: Estimation results of the longitudinal effect β_1 in the random-intercept model for various σ_w^2, where the omitted variable w_i is generated by the Gamma distribution.

Method	σ_w^2	0.01	0.05	0.1	0.5	1	1.5	3
Model(30) with	Bias	2.4808	7.1176	8.6381	11.4373	34.4149	113.3130	193.3380
$Corr(\alpha_i, x_{it}) = 0$	StD	4.4071	4.3913	4.3879	4.3707	4.1200	3.1820	1.6780
	MSE	0.5181	0.9425	1.1719	3.8484	48.3303	204.2370	374.2200
Conditional	Bias	-0.1049	-0.1049	-0.1049	-0.1049	-0.1049	-0.1049	-0.1049
	StD	4.4813	4.4813	4.4813	4.4813	4.4813	4.4813	4.4813
	MSE	0.4010	0.4010	0.4010	0.4010	0.4010	0.4010	0.4010
Proposed	Bias	-0.1044	-0.1046	-0.1043	-0.1037	-0.1033	-0.1030	-0.1027
	StD	4.3685	4.1645	3.9643	2.9343	2.3410	2.0088	1.4917
	MSE	0.3913	0.3740	0.3579	0.2870	0.2552	0.2406	0.2223
Between&	Bias	-0.0934	-0.0934	-0.1127	-0.1055	-0.0997	-0.1030	-0.1015
Within cluster	StD	4.7485	4.7701	4.6838	4.5178	4.4992	4.4910	4.4857
	MSE	0.4303	0.4308	0.4229	0.4043	0.4026	0.4019	0.4016
Shared random-	Bias	-0.1049	-0.10493	-0.1049	-0.1049	-0.1049	-0.1049	-0.1049
effects	StD	4.4813	4.4813	4.4813	4.4813	4.4813	4.4813	4.4813
	MSE	0.4010	0.4010	0.4010	0.4010	0.4010	0.4010	0.4010

Results are reported in percentages.

Table 3: Estimation results of the longitudinal effect β_1 in the random-intercept model for various σ_w^2, where the omitted variable w_i is generated by mixture of normal distributions.

Method	σ_w^2	0.01	0.05	0.1	0.5	1	1.5	3
Model(30) with	Bias	199.339	199.318	199.309	199.332	199.375	199.389	199.463
$Corr(\alpha_i, x_{it}) = 0$	StD	0.557	0.558	0.556	0.545	0.536	0.527	0.501
	MSE	397.367	397.285	397.249	397.337	397.509	397.568	397.860
Conditional	Bias	-0.1366	0.0604	0.2266	-0.0520	-0.1191	0.0661	0.0481
	StD	4.4826	4.4864	4.4761	4.4846	4.4732	4.4654	4.4691
	MSE	0.4067	0.4008	0.4163	0.3930	0.4023	0.4021	0.3918
Proposed	Bias	1.5437	1.9057	2.2339	1.2696	0.7870	2.8325	2.6102
	StD	0.6133	0.6133	0.6121	0.6115	0.6124	0.6114	0.6013
	MSE	0.1814	0.2126	0.2740	0.2142	0.2069	0.3148	0.2199
Between&	Bias	-0.1359	0.0574	0.2260	-0.0514	-0.1198	0.0698	0.0493
Within cluster	StD	4.4872	4.4910	4.4804	4.4883	4.4765	4.4682	4.4714
	MSE	0.4073	0.4011	0.4169	0.3933	0.4026	0.4025	0.3919
Shared random-	Bias	-0.1366	0.0604	0.2266	-0.0520	-0.1191	0.0661	0.0481
effects	StD	4.4826	4.4864	4.4761	4.4846	4.4732	4.4654	4.4691
	MSE	0.4067	0.4008	0.4163	0.3930	0.4023	0.4021	0.3918

Results are reported in percentages.

It is seen that, in the second simulation study which random intercepts are generated from a unimodal skewed distribution, the proposed method performs nicely even if the misspecified mixing distribution and estimates are approximately as accurate as other competitors. More precise estimates are available for the proposed method rather than the other discussed methods. The good performance of the proposed method is highlighted when σ_w^2 tends to large values which means that the endogeneity problem is more serious.

However, in the third simulation study when random intercepts are generated by a mixture distribution with two distinct modes, the proposed method produces larger biases but still smaller standard errors, and in general smaller MSE's in comparison to other discussed methods.

The random-slope model

In this simulation, we apply our proposed method in the more general case of random-slope model. We follow the same scenario as the first simulation but we now assume Equations (29) and (30) include the term $b_i x_{it}$. We also assume $b_i \overset{\text{iid}}{\sim} N\left(0, \sigma_b^2\right)$ which are independent with the random intercepts, α_i's. We set the true parameters as follows: $\beta_0 = -1$, $\beta_1 = 1$, $\beta_2 = 5$, $\kappa_0 = 0$ and $\kappa_1 = 2$, and for the variance components we set $\sigma_\varepsilon^2 = 1$, $\sigma_a^2 = 1$, $\sigma_b^2 = 1$ and $\sigma_w^2 = 5$ for σ_η^2 ranging from 0.01 to 30 which causes the correlation between α_i and x_{it} to vary from 99.6% to 63.0%. Results, not reported here, are in the same direction as in the random-intercept model. By decreasing σ_η^2 which yields to small within-cluster variations, it is seen that the precision of other methods decreases in comparison with the proposed method.

6. An illustrative example

We reanalyze the Georgia birth weight data, used by Vittinghoff et al. (2012), including 200 women, each of whom had 5 children. Data are collected from a study on the

birth weight by Centers for Disease Control in Georgia. This data set, in a larger scale, was previously studied to assess the edogeneity problem by Neuhaus and Kalbfleisch (1998) and Neuhaus and McCulloch (2006). The response of interest is birth weight in 10 kilogram and the time-varying and time-invariant covariates are the mother's age at each birth and at the first birth, respectively. We fit the mixed-effects linear model

$$y_{it} = \beta_0 + \beta_1 x_{it} + \beta_2 x_{i0} + \xi_i + \varepsilon_{it},\ i = 1, \cdots, 200, \tag{31}$$

where ε_{it}'s and ξ_i's are independent and normally distributed with means zero and variances σ^2_ε and σ^2_ξ, respectively. Results of parameter estimates are reported at the top part of Tables 4 and 5. It is obvious that mother's age at each birth depends on her age at the first birth (the sample covariance between x and x_0 is 0.58). We show that when x_0 is omitted from the model then the proposed method, introduced in Section 4.4, can more accurately retrieve the estimate of the coefficient of the time-varying covariate x in Model (31) rather than other discussed methods. It is seen from the first top part of Table 5 that all methods similarly estimate the effect of time-varying covariate x while in the proposed method the estimate of standard error is smaller.

As already mentioned in Section 4.3, the between- and within-cluster covariate method and the shared random-effects method estimate the within-cluster covariate effect. Consequently, the results would be imprecise when the time-varying covariate has slow changes over time. To show this, for the i-th individual, $i = 1, \cdots, n$, we have omitted observations with covariate values outside the interval $(\bar{x}_i - R_{x_i}, \bar{x}_i + R_{x_i})$, where $\bar{x}_i$ and R_{x_i} are respectively the sample mean and the range of covariate observations for the i-th individual. Therefore, the covariate x would have slower changes over time. By doing this omission, 17.9% of observations are discarded in the reduced data set. Figure 2 shows box-plots of covariate x for 10 randomly selected individuals, before and after the mentioned omission, respectively in the left and the right panels.

Results of parameters' estimates for the reduced data set are reported in the bottom parts of Tables 4 and 5. It is seen that the between- and within-cluster covariate method and the shared random-effects method produce large biases for the estimation of β_1. The conditional and the proposed methods produce smaller biases for the estimation of β_1 and the proposed method is more accurate.

We conclude that the proposed method can estimate the total effect of the time-varying covariate x and the estimate is more accurate than the comparable methods mentioned in this paper. However, this statement is more clear by the simulation studies than the empirical study, since as is clear from the reported p-values of parameter β_2 in the original models, the omitted time-invariant covariate x_2 is not significant in the presence of x.

7. Concluding remarks

A review of suggested methods for solving the endogeneity problem in the random-intercept models based on two strategies, the fixed-effect and the random-effect is presented and extended to the case of the general linear mixed model. For the random-effect approach, the methods are categorized as MC and TC methods in which the former expresses the relation between random intercepts and covariates through $E(\alpha_i|\mathbf{x}_i)$ and the

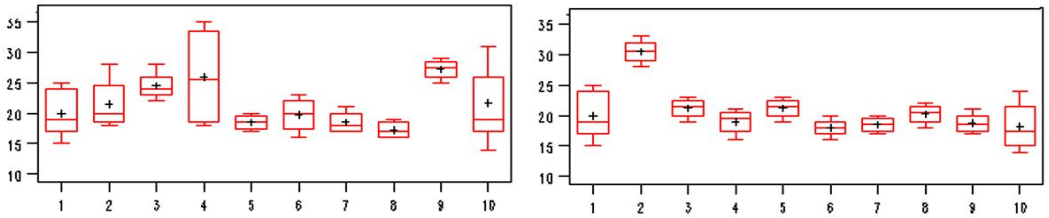

Figure 2: Box-plots of covariate x for 10 randomly selected individuals, in the original and the reduced data sets are shown in the left and the right panels, respectively.

latter assumes a more general structure by working with $f(\mathbf{x}_i|\alpha_i, b_i)$ which also takes into consideration the covariance structure of the model.

In the context of TC methods, a new method is proposed. With simulation studies we have shown that the proposed method leads to more precise estimates by taking the variability of x_{it} into consideration while the method is still good enough in terms of MSE values. This method can be applied even in situations where there are covariates with slow changes over time, a case where those methods that are based on centered covariates to solve the endogeneity problem are not capable in producing precise estimates.

It is also shown by two simulation studies that the proposed method is roughly robust against the misspecification of mixing distribution. Indeed, we show that when the random-intercepts follow a skew unimodal distribution, the assumption of mixing normality does not have much impact on the performance of our proposed method. While when random-intercepts are generated from a mixture of two normal distributions with two distinct modes, the proposed method has larger biases but still smaller standard deviations than the MC method. The proposed method can also be extended to the application of other models but performance needs to be investigated further.

Table 4: Results of parameter estimates for the birth weight data set.

		Model(31)	Conditional	B&W cluster-Covariate	Shared random-effects	Proposed
For the original data set						
β_0	Estimate	26.1977	31.8692	26.9480	31.3546	28.1908
	StD	(1.6061)	(2.7516)	(1.5124)	(0.2954)	(0.7302)
	p-value	<.0001	<.0001	<.0001	<.0001	<.0001
β_1	Estimate	0.1434	0.1463	0.1463	0.1463	0.1463
	StD	(0.0332)	(0.0348)	(0.0348)	(0.0348)	(0.0310)
	p-value	<.0001	<.0001	<.0001	<.0001	<.0001
β_2	Estimate	0.1171	-	-	-	-
	StD	(0.0964)				
	p-value	0.2246				
λ^*	Estimate	-	-	0.05745	-	-
	StD			(0.0769)		
	p-value			0.4553		
σ^2_ε	Estimate	19.9051	19.9070	19.9070	19.9070	19.9019
	StD	(0.9958)	(0.9960)	(0.9960)	(0.9960)	(0.9955)
	p-value	<.0001	<.0001	<.0001	<.0001	<.0001
$\sigma^2_{\xi_1}$	Estimate	12.7383	-	12.8089	13.4681	12.8018
	StD	(1.6921)		(1.6992)	(1.7606)	(1.6958)
	p-value	<.0001		<.0001	<.0001	<.0001
For the reduced data set						
β_0	Estimate	23.5341	29.7730	21.9909	31.4204	26.6281
	StD	(6.0979)	(2.8915)	(1.5592)	(0.3098)	(1.1373)
	p-value	<.0001	<.0001	<.0001	<.0001	<.0001
β_1	Estimate	0.2353	0.2236	0.1276	0.0343	0.2230
	StD	(0.0610)	(0.0679)	(0.0595)	(0.0589)	(0.0511)
	p-value	<.0001	0.0011	0.0323	0.5600	<.0001
β_2	Estimate	0.0402	-	-	-	-
	StD	(0.1116)				
	p-value	0.7187				
λ^*	Estimate	-	-	0.3099	-	-
	StD			(0.0801)		
	p-value			<.0001		
σ^2_ε	Estimate	18.4513	18.4579	17.8849	18.7687	18.4431
	StD	(1.0476)	(1.0483)	(1.0222)	(1.0655)	(1.0468)
	p-value	<.0001	<.0001	<.0001	<.0001	<.0001
$\sigma^2_{\xi_1}$	Estimate	13.6243	-	14.2163	14.5969	13.5928
	StD	(1.8402)		(1.9256)	(1.9381)	(1.8359)
	p-value	<.0001		<.0001	<.0001	<.0001

∗Coefficient of $\bar{x}_i$.

8. References

Ashenfelter, O., Rouse, C. 1998. Income, schooling, and ability: evidence from a new sample of identical twins. The Quarterly Journal of Economics **113** (1), 253–284.

Austin, P.C., Alter, D.A., Anderson, G.M., Tu, J.V. 2004. Impact of the choice of benchmark on the conclusions of hospital report cards. American Heart Journal **148**, 1041–1046.

Austin, P. C., Alter, D. A., Tu, J.V. 2003. The use of fixed and random-effects models for

Table 5: Results of parameter estimates of the endogenous covariate model.

	β_{00}	σ^2_η	$\sigma^2_{\xi_2}$	$\sigma_{\xi_1\xi_2}$
Estimate	21.6330	20.6070	13.7988	1.0295
StD	(0.2993)	(1.0303)	(1.8083)	(1.3799)
p-value	<.0001	<.0001	<.0001	0.4556
Estimate	21.4791	6.4442	13.5369	0.8497
StD	(0.2749)	(0.3657)	(1.5180)	(1.5592)
p-value	<.0001	<.0001	<.0001	0.5858

classifying hospitals as mortality outliers: A Monte Carlo assessment. Medical Decision Making **23**, 526–539.

Breusch, T., Ward, M.B., Nguyen, H.T.M., Kompas, T. 2011. On the fixed-effects vector decomposition. Political Analysis **19** (2), 123–134.

Cameron, A.C., Trivedi, P.K. 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Chamberlain, G. 1982. Multivariate regression models for panel data. Journal of econometrics **18**, 5-46.

Chamberlain, G. 1984. Panel data. In: Z. Griliches and M.D. Intriligator (Eds.) Handbook of Econometrics, Amsterdam: North-Holand, Volume 2, chapter 22.

Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S. 2002. Analysis of Longitudinal Data. 2nd edition. Oxford Statistical Science Series 25.

Hansen, L. 1982. Large sample properties of generalized method of moments estimators. Econometrica **50** (3), 1029–1054.

Hausman, J., Taylor, W. 1981. Panel data and unobservable individual effects. Econometrica **49** (6), 1377–1398.

Hsiao, C. 2003. Analysis of Panel Data. 2nd edition. Cambridge: Cambridge University Press.

Lancaster, T. (2000). The incidental parameter problem since 1948. Journal of Econometrics **95**, 391–413.

Malinvaud, E. 1970. Statistical Methods of Econometrics. 2nd edition. Amsterdam: North-Holland.

Mundlak, Y. 1978. On the pooling of time series and cross section data. Econometrica, **46**, 69–85.

McCulloch, C.E., Neuhaus, J.M. 2011. Prediction of random effects in linear and generalized linear models under model misspecification. Biometrics **67** (1), 270–279.

Neuhaus, J. M., Hauck, W. W., Kalbfleisch, J. D. 1992. The effects of mixture distribution misspecification when fitting mixed-effects logistic models. Biometrika **7**(9), 755–762.

Neuhaus, J. M., Kalbfleisch, J. D. 1998. Between- and within-cluster covariate effects in the analysis of clustered data. Biometrics **54**, 638–645.

Neuhaus, J. M., McCulloch, C. E. 2006. Separating between- and within-cluster covariate effects by using conditional and partitioning methods. Journal of the Royal Statistical Society, B **68** (5), 859–872.

Palta, M., Yao, T.-J. 1991. Analysis of longitudinal data with unmeasured confounders. Biometrics **47**, 1355–1369.

Plumper, T., Troeger, V. 2007. Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. Political Analysis **15** (2), 124–39.

Verbeek, M. 2004. A Guide to Modern Econometrics. 2nd edition. John Wiley & Sons.

Verbeke, G., Fieuws, S. 2007. The effect of miss-specified baseline characteristics on inference for longitudinal trends in linear mixed models. Biostatistics **8**(4), 772–783.

Verbeke, G., Lesaffre, E. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association **91,** 217–221.

Verbeke, G., Lesaffre, E. 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics and Data Analysis **23**, 541–556.

Verbeke, G., Spiessens, B., Lesaffre, E. 2001. Conditional linear mixed models. The American Statistician **55**, 25-34.

Vittinghoff, E., Glidden, D.V., Shiboski, S.C., McCulloch, C.E. 2012. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. 2nd edition, Springer-Verlag, Inc.

Wooldridge, J.M. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd edition, MIT Press.