Simultaneous Prediction Intervals for Small Area Parameter

Katarzyna Reluga*  María José Lombardía† and Stefan Andreas Sperlich‡

Abstract

In this paper we address the construction of simultaneous prediction intervals for small area parameters in linear mixed models. Simultaneous intervals are necessary to compare areas, or to look at several areas at a time, as the presently available intervals are not statistically valid for these scenarios. We consider two frameworks to develop simultaneous intervals: the Monte Carlo approximation of the volume of a tube based intervals and bootstrap bands. Proofs of the consistency as well as the asymptotic coverage probability of the bootstrap intervals are provided. Our proposal is accompanied by simulation experiments and a data example. The simulations show which method works best under a particular scenario. We illustrate use and utility of simultaneous intervals for the analysis of small area parameters. When comparing the areas, the classical methods lead to erroneous conclusions, visible in the study of the household income distribution in Galicia in Northern Spain.

Keywords: small area estimation, uniform confidence intervals, unit level model, Fay-Herriot model, linear mixed models

*Katarzyna Reluga is a PhD candidate and a teaching assistant at the RCS, University of Geneva, Switzerland. E-mail: katarzyna.reluga@unige.ch.
†María José Lombardía is a Professor at the CITIC, University of A Coruña, Spain. E-mail maria.jose.lombardia@udc.es.
‡Stefan Andreas Sperlich is a Professor at the RCS, University of Geneva, Switzerland. E-mail: stefan.sperlich@unige.ch.

The authors gratefully acknowledge support from the MINECO grants MTM2017-82724-R and MTM2014-52876-R, the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015 and Centro Singular de Investigación de Galicia ED431G/01), all of them through the ERDF. We would like to thank the Galician Institute of Statistics (IGE) for the transfer of the data. Finally, we thank D. Flores, W. Gonzalez Manteiga, E. López-Vizcaíno, D. Morales, T. Schmid, N. Salvati and S. Ranjbar for helpful discussions.
1 Introduction

An increasing demand for reliable statistics regarding socio-demographic groups or geographical regions contributed to the development of the family of statistical methods called small area estimation (SAE) (Rao and Molina 2015). The term small area refers to any domain, such as county, school district, health service area, etc., for which direct estimates are not feasible due to poor precision. SAE is widely applied to assess, among others, the need for implementing health and educational programs, (tax) transfers, environmental planning or the allocation of subsidies in less developed regions.

The methodology of SAE borrows significantly from mixed effects modeling where the extra between area variation is captured by area-specific random effects. In linear mixed models (LMM), empirical best predictor (EBLUP) and empirical Bayes (EB) estimators are widely recognized methods to obtain small area predictions. To assess accurateness of a prediction, it is crucial to measure its variability. Traditionally, one would provide the Mean Squared Errors (MSE) which has been widely discussed (see e.g., Prasad and Rao (1990), Datta and Lahiri (2000), Das et al. (2004)). As practitioners may find prediction intervals more informative than the MSE, several authors worked on their construction, see e.g., Hall and Maiti (2006b), Chatterjee et al. (2008) or Flores Agreda (2017) for bootstrap versions, and Yoshimori and Lahiri (2014) (and references therein) for analytical approximations. In all mentioned cases, the coverage probability of $100(1 - \alpha)$ area-wise prediction intervals (APIs) refers to the mean across all areas. This implies that, by construction, about $100\alpha$ percent of the provided intervals do not contain the area parameter of interest. As a consequence, these prediction (or often called confidence) intervals are not appropriate for addressing neither a joint consideration nor a comparison of the areas. Yet such a comparison can be of great interest, for example, statistical offices providing reports to policy makers, or within public health research centers carrying out analyses on
demographic groups. The use of this practice is prevalent among practitioners, who try to find significant discrepancies between areas or make decisions about resource allocations.

In the literature of mixed modeling, several authors used the formula for the volume of a tube in order to obtain simultaneous bands in longitudinal studies (e.g., Sun et al. (1999) and Maringwa et al. (2008)). The goal of these studies was to provide bands for a fixed effect part, differing from our case, where we do so for the combination of fixed and random effects, where one must account for the variability of the latter. Krivobokova et al. (2010) extend the ideas of these authors, using a mixed model representation for penalized splines and constructing uniform bands for the regression curves for a one dimensional covariate.

Thus, the aim of this paper is to close the described gap between the needs of practitioners and what the present literature provides, constructing simultaneous prediction intervals (SPIs) for LMM. To the best of our knowledge, only Ganesh (2009) addressed this problem in SAE from a Bayesian perspective, creating credible bands, but only for area level model. We have chosen frequentist approaches and consider both unit and area level models. Starting from the idea of using the volume of a tube, similar to Krivobokova et al. (2010), we first propose a Monte Carlo (MC) method before presenting a bootstrap based approach which does not require normality of random effects and errors.

The remainder of this paper is organized as follows. In Section 2 we introduce our LMM together with the parameter used and MSE estimators. The SPI constructions, as well as the consistency of bootstrap intervals, are outlined in Section 3. A simulation study and a data example are provided in Sections 4 and 5 respectively. The conclusions can be found in Section 6. Some technical details are deferred to Appendix A and others to supplementary documents (S.D.), which also include more simulation results.
2 Small area inference

For $X$, $Z$ full column rank matrices for a fixed and a random part, consider the LMM

$$ y = X\beta + Zu + e, \quad (1) $$

with $\beta$ a vector of fixed effects, $u$ a vector of random effects, and error $e$. Assume $u$ and $e$ to be independent with $u \overset{ind}{\sim} q(0, G)$ and $e \overset{ind}{\sim} n(0, R)$. For the ease of presentation we focus on LMM with block diagonal covariance matrix (LMMb)

$$ y_d = X_d\beta + Z_d u_d + e_d, \quad d = 1, \ldots, D, \quad (2) $$

where $n_d$ is the number of units in $d^{th}$ cluster (or domain), $y_d \in \mathbb{R}^{n_d}$, $X_d \in \mathbb{R}^{n_d \times (p+1)}$ and $Z_d \in \mathbb{R}^{n_d \times q}$. Furthermore, $D$ is the total number of domains, $\beta \in \mathbb{R}^{p+1}$ is an unknown common vector of regression coefficients, $u_d \overset{ind}{\sim} q_d(0, G_d)$ and $e_d \overset{ind}{\sim} n_d(0, R_d)$. We assume that $G_d = G_d(\theta) \in \mathbb{R}^{q_d \times q_d}$ and $R_d = R_d(\theta) \in \mathbb{R}^{n_d \times n_d}$ which depend on variance parameters $\theta = (\theta_1, \ldots, \theta_h)^t$. LMM can be easily retrieved applying the notation introduced by Prasad and Rao (1990), p. 168. Under this setup we suppose that the variance-covariance matrix $V$ is nonsingular $\forall \theta$, $i = 1, \ldots, h$ and

$$ \mathbb{E}(y) = X\beta \quad \text{and} \quad \mathbb{V}(y) = R + ZGZ^t =: V(\theta) = V. \quad (3) $$

Two important examples of (2) which are extensively used in SAE are the following: the Nested Error Regression Model (NERM), see Battese et al. (1988):

$$ y_{dj} = x_{dj}^t \beta + u_d + e_{dj}, \quad d = 1, \ldots, D, \quad j = 1, \ldots, n_d, \quad (4) $$

where $y_{dj}$ is the quantity of interest for the $j^{th}$ population unit for the $d^{th}$ small area, $x_{dj} = (1, x_{d1j}, \ldots, x_{dpj})^t$, $u_d \overset{iid}{\sim} (0, \sigma_u^2)$ and $e_{dj} \overset{iid}{\sim} (0, \sigma_e^2)$ for $d = 1, 2, \ldots, D$ and $j = 1, 2, \ldots, n_d$. Here $y_d = (y_{d1}, \ldots, y_{dn_d})$, $X_d = \text{col}_{1 \leq j \leq n_d} x_{dj}^t$, $q_d = 1$, $Z_d = 1_{n_d}$ with $1_{n_d}$ which is a $n_d$ vector of ones, $e_d = (e_{d1}, \ldots, e_{dn_d})^t$, $\theta = (\sigma_e^2, \sigma_u^2)^t$, $R_d(\theta) = \sigma_e^2 I_{n_d}$ with $I_{n_d}$ which is a $n_d \times n_d$
identity matrix and $G_d(\theta) = \sigma_u^2$. Another one is the Fay-Herriot Model (FHM), see Fay and Herriot (1979):

$$y_d = x_d^T \beta + u_d + e_d, \quad d = 1, \ldots, D,$$

where $x_d = (1, x_{d1}, \ldots, x_{dp})^T$, $u_d \overset{iid}{\sim} N(0, \sigma_u^2)$ and $e_d \overset{iid}{\sim} N(0, \sigma_e^2)$ with $\sigma_e^2 (d = 1, 2, \ldots, D)$ being known. In this case, $n_d = q_d = 1$, $Z_d = 1$, $\theta = \sigma_u^2$, $R_d(\sigma_u^2) = \sigma_e^2$.

For estimation, assume that a finite population $P$ of size $N$ is partitioned into $D$ subpopulation $P_1, P_2, \ldots, P_D$ of sizes $N_1, N_2, \ldots, N_D$. Further, let $Y$ be a random value of interest and let $y_{dj}$ be a realization of $Y$ in a $j^{th}$ unit of the $d^{th}$ small area, where $j = 1, \ldots, N_d$ and $d = 1, \ldots, D$. Our target parameter is the population mean of small area $d$ which is defined as $\bar{Y}_d = \frac{1}{N_d} \sum_{j=1}^{N_d} y_{dj}$. Under LMMb, we can approximate it with

$$\mu_d = \bar{X}_d^T \beta + \bar{Z}_d^T u_d,$$

where $\bar{X}_d$ is a vector of the known population means of the covariates for the $d^{th}$ area, $\bar{Z}_d \in \mathbb{R}^q$ is composed of some constants, $d = 1, \ldots, D$. Since $\bar{X}_d$ and $\bar{Z}_d$ can be replaced by any vector, (6) is an example of a general linear combination of fixed and random effects. It can be used as a parameter of interest under NERM and FHM. Under the former model, we draw a sample of size $n_d$ from $N_d$ elements in each area and observe values $\{y_{dj}, x_{dj}\}$ for $d = 1, 2, \ldots, D$ and $j = 1, 2, \ldots, n_d$ with $n = \sum_{d=1}^{D} n_d$ the total number of units in the sample. Assume that there is no selection bias, and model (4) is valid for the population values. The assumption holds for sampling designs which do not depend on the values of $y$, but only on $x$. This includes simple random sampling. We suppose further that under NERM the sampling fraction $f_d = n_d/N_d$ is negligible. If we do not have access to the units for the whole population, but obtain $\bar{y}_d$, we can use the FHM modeling directly the area means as in (5) with $y_d \equiv \bar{y}_d$, $x_d^T \equiv \bar{X}_d^T$ and $u_d \equiv \bar{Z}_d^T u_d$ with $\bar{Z}_d = 1$. Henderson (1975) developed the best linear unbiased predictor (BLUP) of a linear combination of random effects $\mathbf{u}$ and fixed effects $\beta$ for completely known covariance matrix
V. Applying this idea we obtain the BLUP estimator of the area means \( \tilde{\mu}_d \):

\[
\tilde{\mu}_d := \tilde{\mu}_d(\theta) = \bar{X}_d^t \tilde{\beta} + \bar{Z}_d^t \tilde{u}_d,
\]

with \( \theta = (\theta_1, \ldots , \theta_h)^t \). (7)

where \( \tilde{\beta} = \tilde{\beta}(\theta) = (X'V^{-1}X)^{-1}X'V^{-1}y \), \( \tilde{u}_d = \tilde{u}_d(\theta) = G_d\bar{Z}_d^t V_d^{-1}(y_d - X_d \tilde{\beta}) \). In practice \( \theta \) is unknown, hence we use \( \hat{\theta} = \hat{\theta}(y) \) which yields the EBLUP

\[
\hat{\mu}_d := \hat{\mu}_d(\hat{\theta}) = \bar{X}_d^t \hat{\beta} + \bar{Z}_d^t \hat{u}_d,
\]

with \( \hat{\theta} = (\hat{\theta}_1, \ldots , \hat{\theta}_h)^t \). (8)

where \( \hat{\beta} = \hat{\beta}(\hat{\theta}) \), \( \hat{u} = \hat{u}(\hat{\theta}) \). Having assumed certain conditions on the distributions of random effects and errors as well as the variance components \( \theta \) (see the assumptions and regularity conditions (RC) given in Appendix A.1), Kackar and Harville (1981) proved that the two-stage procedure provides an unbiased estimator of \( \mu_d \).

Remark 1. In SAE one often assumes to have fixed design, that is complete information on \( X \) and possibly missing cases in \( y \). In practice we deal with two different frameworks, namely either one has a complete sample of units from a survey or a sample with some missing information in a response variable. In the former situation (6) is replaced with

\[
\mu^*_d = \bar{x}_d^t \beta + \bar{z}_d^t u_d
\]

with \( \bar{x}_d = (1, \bar{x}_{d1}, \ldots , \bar{x}_{dp}) \), \( \bar{x}_d = \sum_{j=1}^{n_d} x_{dj} \) and \( \bar{z} \equiv \bar{Z} \). These cases were distinguished by Lombardía and Sperlich (2012) and correspond to frameworks (a), (b) and (c) therein. In addition, if the sampling fraction is nonnegligible or there many missing responses, \( \bar{Y}_d \) is defined as \( \bar{Y}_d = f dy_d + (1 - f_d) \bar{y}_{dl} \), where \( \bar{y}_{dl} \) is the mean of the nonsampled units \( y_{dl} \), with \( l = n_d + 1, \ldots , N_d \) for \( d \)th area. Under LMMb \( y_{dl} \) is replaced with \( x_{dl}^t \beta + z_{dl}^t \hat{u}_d \) where \( x_{dl} \) are values of covariates associated with a unit \( l \). The methodology that we shall develop can be directly applied within this framework with some notational changes.

For our purpose it is important to assess the variability of the prediction and \( \text{MSE}(\hat{\mu}_d) = \mathbb{E}(\hat{\mu}_d - \mu_d) \) is the most common measurement of the uncertainty in mixed models. Here \( \mathbb{E} \)
denotes the expectation with respect to model (2). We can decompose the MSE

\[
\text{MSE}(\hat{\mu}_d) = \text{MSE}(\tilde{\mu}_d) + \mathbb{E}(\tilde{\mu}_d - \bar{\mu}_d)^2 + 2\mathbb{E}[(\tilde{\mu}_d - \mu_d)(\hat{\mu}_d - \bar{\mu}_d)].
\]  

(10)

MSE(\tilde{\mu}_d) accounts for the variability of \( \mu_d \) when the variance components \( \theta \) are known. Assuming LMMb and \( m_t^d = \bar{X}_d^t - \bar{a}_d^t X_d \) with \( \bar{a}_d^t = \bar{Z}_d^t G \bar{Z}_d V_d^{-1} \), \( \tilde{\mu}_d \) reduces to

\[
\bar{Z}_d^t (G_d - G_d \bar{Z}_d V_d^{-1} Z_d G_d) \bar{Z}_d + m_d \left( \sum_{d=1}^D X_d^t V_d^{-1} X_d \right)^{-1} m_d =: g_{1d}(\theta) + g_{2d}(\theta).
\]

Under normality the last term in (10) is zero and is therefore rarely further considered.

An accurate estimation of MSE(\( \hat{\mu}_d \)) is crucial to construct SPI. In SAE, the exact expression of MSE does not exist, because the empirical predictors are not linear statistics due to the estimation of the variance components \( \theta \). For this reason the two last terms in expression (10) are intractable and one has to approximate them. Having added some technical assumptions, one relies on the linearization and large sample techniques to approximate the unknown quantities. Kackar and Harville (1984) provided a proposal, Prasad and Rao (1990) improved on their results, studying second-order accuracy of models with block diagonal matrices (2), Datta and Lahiri (2000) derived approximations for the models with estimated variance components estimated, and Das et al. (2004) developed further expansions for general LMM. The second-order unbiased MSE(\( \hat{\mu}_d \)) estimator obtained applying the method of fitting-of-constants and REML is

\[
\text{mse}_L(\hat{\mu}_d) \approx g_{1d}(\theta) + g_{2d}(\theta) + 2g_{3d}(\theta), \quad \text{with } g_{3d}(\theta) = \text{tr} \left( (\partial a_d^t / \partial \theta) V_d (\partial a_d^t / \partial \theta)^t V_A(\theta) \right),
\]

with the asymptotic covariance matrix \( V_A(\theta) \), and where \( \mathbb{E}\left[ \text{mse}_L(\hat{\mu}_d) \right] = \text{MSE}(\hat{\mu}_d) + o(D^{-1}) \).

Prasad and Rao (1990) provided simplified expressions that account for uncertainty in NERM (4) and FHM (5). Similar analytical approximations have been obtained in the case of more general nonlinear mixed and linear multivariate models by González-Manteiga et al. (2007) and González-Manteiga et al. (2008a) respectively. Linearization based techniques
are theoretically sound, yet they are model dependent (i.e., for each class of models a new approximation is necessary). In addition, they are restricted to linear parameters and their corresponding EBLUP. Therefore, in the literature, many bootstrapping schemes have been proposed, e.g., a parametric bootstrap which is the most popular in SAE (Butar and Lahiri 2003; Hall and Maiti 2006b). On the other hand, Carpenter et al. (2003) suggested resampling with replacement from the variance-inflated errors and random effects in contrast to Hall and Maiti (2006a) who advised the use of a wild bootstrap.

To introduce bootstrap MSE estimators, consider the following analogue of LMMb (2)

\[ y_d^* = X_d^t \hat{\beta} + Z_d^t u_d^* + e_d^* \]

Algorithms to obtain vectors \( u^* \) and \( e^* \) are described in Section 3.1. Define the following bootstrap versions of \( \mu_d \) (6), \( \tilde{\mu}_d \) (7) and \( \hat{\mu}_d \) (8):

\[ \mu_d^* = X_d^t \tilde{\beta} + Z_d^t \tilde{u}_d^* \]
\[ \hat{\mu}_d^* := \hat{\mu}_d(\hat{\theta}^*) = X_d^t \tilde{\beta} + Z_d^t \tilde{u}_d^* \]
\[ \hat{\mu}_d^* := \hat{\mu}_d(\hat{\theta}^*) = X_d^t \hat{\beta} + Z_d^t \hat{u}_d^* \]

in which, with certainty, \( \theta^* := \hat{\theta} \). Then MSE*\[\hat{\mu}_d(\hat{\theta}^*)\] = \( E^* (\hat{\mu}_d - \mu_d)^2 \) is a bootstrap MSE estimator which might be approximated by averaging over the bootstrap samples

\[ \text{MSE}_{B1}^{\mu_d}(\hat{\theta}^*) \approx \text{mse}_{B1}(\hat{\mu}_d) = \frac{1}{B} \sum_{b=1}^{B} \left( \hat{\mu}_d^{*(b)} - \mu_d^{*(b)} \right)^2, \]  (12)

and \( \hat{\mu}_d^{*(b)}, \mu_d^{*(b)} \) defined in (11), calculated from the \( b^{th} \) sample. It is well known in the literature that (12) leads to \( E \{ \text{MSE}_{B1}^{\mu_d}(\hat{\theta}^*) \} = \text{MSE}(\hat{\mu}_d) + O(D^{-1}) \). To obtain a bias of order \( o(D^{-1}) \), Butar and Lahiri (2003) advocate a scheme in which, instead of approximating the whole MSE*\[\hat{\mu}_d(\hat{\theta})\], only intractable terms are estimated by bootstrap

\[ \text{MSE}_{SP}^{\mu_d}(\hat{\theta}^*) = 2 \left[ g_{1d}(\hat{\theta}) + g_{2d}(\hat{\theta}) \right] - E^* \left[ g_{1d}(\hat{\theta}^*) + g_{2d}(\hat{\theta}^*) \right] + E^*(\hat{\mu}_d - \mu_d)^2, \]  (13)

where \( g_{1d}(\hat{\theta}^*) = Z_d^t (G_d^* - G_d^* Z_d^t V_d^{-1} \hat{\mu}_d G_d^*) Z_d \) and \( g_{2d}(\hat{\theta}^*) = m_d^t \left( \sum_{d=1}^{D} X_d^t V_d^{-1} X_d \right)^{-1} m_d^* \).
with \( m^*_d = \bar{X}_d^t - a^*_d^t X_d \) and \( a^*_d = Z'_d G'_d Z'_d V^{-1} \). Since we dropped the normality assumption, we propose a second-order correct semiparametric bootstrap, i.e.,

\[
\text{MSE}_{SPA}^{*}[\hat{\mu}^*_d(\hat{\theta}^*)] = 2 \left[ g_{1d}(\hat{\theta}) + g_{2d}(\hat{\theta}) \right] - \mathbb{E}^* \left[ g_{1d}(\hat{\theta}^*) + g_{2d}(\hat{\theta}^*) \right] + \mathbb{E}^* (\hat{\mu}^*_d - \hat{\mu}_d)^2 + 2 \mathbb{E}^* [(\hat{\mu}^*_d - \mu_d)(\hat{\mu}^*_d - \hat{\mu}_d)].
\] (14)

In Section 4 we study the performance of two other MSE estimators, that is, a second-order unbiased estimator \( \text{MSE}_{BC2}^* \) proposed by Hall and Maiti (2006a) based on a double-bootstrap, and a first-order estimator \( \text{MSE}_{3T}^* \) which approximates each term of (10) using its bootstrap equivalents. Since these methods do not perform better (see simulations in Section 4), details regarding their computation are deferred to S.D.

Remark 2. In practice, similarly as for \( \text{MSE}_{B1}^*[\hat{\mu}^*_d(\hat{\theta}^*)] \), we use bootstrap approximations of (13) and (14) as well as \( \text{MSE}_{BC2}^* \) and \( \text{MSE}_{3T}^* \) that is \( \text{mse}_{SP}(\hat{\mu}^*_d) \), \( \text{mse}_{SPA}(\hat{\mu}^*_d) \), \( \text{mse}_{BC2}(\hat{\mu}^*_d) \), \( \text{mse}_{3T}(\hat{\mu}^*_d) \), respectively, which are computed means over the bootstrap samples. These approximations are consistent estimators as \( B \rightarrow \infty \).

3 SPI for small area means under two frameworks

Simultaneous confidence intervals (SCI) have been discussed extensively in nonparametric statistics, where one is interested in the estimation of a model \( y_{dj} = m(x_{dj}) + \varepsilon_{dj} \) with \( m(x_i) \in C^r([a, b]) \) a \( r \) times differentiable function. To construct SCIs, the asymptotic distribution of \( \sup_{a \leq x \leq b} |\hat{m}(x) - m(x)| \) has been tackled in the literature. Bickel and Rosenblatt (1973) consider the distribution of \( \sup_{a \leq x \leq b} |W(x)| \) where \( W(x) \) is a standard Gaussian process; Hall (1991) prove a very poor rate \(((\log n)^{-1} \text{ with } n \text{ the number of observations})\) of convergence for this. Therefore, Claeskens and Van Keilegom (2003) propose bootstrap approximation. Another method to construct SCIs is to use the formula for the volume of a tube (Sun and Loader (1994) and references therein). Last but not
least, one can obtain Bayesian simultaneous credible bands (applying Markov Chain MC), which are conceptually different from frequentest bands. Construction of Scheffe (1999) type SCIs has also been taken into consideration; these were originally developed for the models with homoscedastic independent errors. He proposed a methodology to construct simultaneous intervals for a regression space \( m(x_{dj1}, \ldots, x_{djp}), d = 1, \ldots, D, j = 1, \ldots, n_d \) for \( x_{dj} \in \mathcal{X} \equiv \mathbb{R}^p \), assuming an unconstrained domain of interest. But when dealing with a \( p \)-dimensional rectangle \( \mathcal{X} \subset \mathbb{R}^p \), any construction using a constrained domain should provide narrower bands. Hence, we do not develop this methodology any further.

For the sake of presentation we concentrate on the construction of SPI for the mean of each area (6) i.e., we consider a confidence region \( I_{1-\alpha} \) such that \( P(\hat{\mu}_d \in I_{1-\alpha} \forall d \in [D]) = 1 - \alpha, [D] = \{1, \ldots, D\} \). This is equivalent to finding a critical value \( c_{1-\alpha} \) which satisfies

\[
\alpha = P\left( \frac{|\hat{\mu}_d - \mu_d|}{\text{MSE}^{1/2}(\hat{\mu}_d)} > c_{1-\alpha} \forall d \in [D] \right) = P\left( \max_{d=1,\ldots,D} \left| \frac{\hat{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\hat{\mu}_d)} \right| > c_{1-\alpha} \right). \tag{15}
\]

We conclude from (15) that the estimation of MSE and an accurate approximation of the quantile from the distribution of

\[
S_D = \max_{d=1,\ldots,D} \left| \frac{\hat{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\hat{\mu}_d)} \right|
\tag{16}
\]

are crucial. It follows that with probability \( 1 - \alpha \) we cover all small area means with

\[
I_{1-\alpha}^S = [\hat{\mu}_d - c_{1-\alpha} \times \text{MSE}^{1/2}(\hat{\mu}_d), \hat{\mu}_d + c_{1-\alpha} \times \text{MSE}^{1/2}(\hat{\mu}_d)] \forall d \in [D], \tag{17}
\]

in which, in practice, we need to estimate \( c_{1-\alpha} \) and MSE(\( \hat{\mu} \)). The problem of simultaneous bands in nonparametric curve estimation is similar to the problem that we address.

3.1 Construction of SPI using resampling approximations

A derivation of analytical SPI based on the volume of a tube formula results in a mathematical expression that crucially hinges on unknown factors which partly are very hard to
estimate, others to be simulated. These intervals, their asymptotics including a theorem and a proof are provided in the S.D. Instead, we implemented a less sophisticated but much simpler simulation method similar to Ruppert et al. (2003) for confidence bands of nonparametric curves. It is based on assuming normality for the random predictors and therefore mainly attractive for the FHM, as under this model, normality for the random terms is usually taken for granted. More specifically, we work with

\[
\begin{bmatrix}
\hat{\beta} - \beta \\
\hat{u} - u
\end{bmatrix}
\approx \sim N \left[ 0, (C'R^{-1}C + G^+)^{-1} \right], \quad G^+ = \begin{bmatrix} 0_{(p+1) \times (p+1)} & 0_{(p+1) \times D} \\ 0_{D \times (p+1)} & G_{D \times D}^{-1} \end{bmatrix}
\]

with \( C = [X \ Z] \). We apply (18) to simulate the distribution of (16)

\[
S_D = \max_{d=1,...,D} \frac{\mid \hat{\mu}_d - \mu_d \mid}{\text{MSE}^{1/2}(\hat{\mu}_d)} \approx \max_{d=1,...,D} \frac{\mid \hat{\mu}_d - \mu_d \mid}{\text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d)} =: \hat{S}_D,
\]

where \( \hat{C}_d = (\hat{X}_d^t, \hat{Z}_d^t)^t \) and \( \text{mse}_{(\cdot)}^{(\cdot)}(\hat{\mu}_d) \) is one of the MSE estimators defined in Section 2. We draw \( B \) realizations of (18) and calculate their \( \hat{S}_D \). An estimate of the critical value \( c_{1-\alpha} \) is the \( \left( \lceil (1 - \alpha)B \rceil + 1 \right) \)th order statistic of \( \hat{S}_D \). Finally, we construct a MC SPI by

\[
I_{1-\alpha}^{MC} : \left[ \hat{\mu}_d - \hat{c}_{1-\alpha} \times \text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d), \hat{\mu}_d + \hat{c}_{1-\alpha} \times \text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d) \right] \forall d \in [D].
\]

A construction of MC SPIs resembles the derivations of the volume of a tube based SPIs in S.D., but without a correction for the variability of \( \theta \). Here an implicit correction is included due to MSE estimation.

To construct MC SPIs, we assume normality for errors and random effects. Bootstrap can circumvent a direct application of the normal asymptotic distribution and can provide faster convergence rates. Let \( B \) now be the number of bootstrap samples \( (y^{(b)}, X, Z) \), \( b = 1, \ldots, B \) and \( c_{1-\alpha} \) the \( (1 - \alpha)^{th} \) quantile of the distribution of \( S_D \); we claim that \( c_{1-\alpha} \)

\[
\text{with } C = [X \ Z]. \quad \text{We apply } (18) \text{ to simulate the distribution of } (16)
\]

\[
S_D = \max_{d=1,...,D} \frac{\mid \hat{\mu}_d - \mu_d \mid}{\text{MSE}^{1/2}(\hat{\mu}_d)} \approx \max_{d=1,...,D} \frac{\mid \hat{\mu}_d - \mu_d \mid}{\text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d)} =: \hat{S}_D,
\]

where \( \hat{C}_d = (\hat{X}_d^t, \hat{Z}_d^t)^t \) and \( \text{mse}_{(\cdot)}^{(\cdot)}(\hat{\mu}_d) \) is one of the MSE estimators defined in Section 2. We draw \( B \) realizations of (18) and calculate their \( \hat{S}_D \). An estimate of the critical value \( c_{1-\alpha} \) is the \( \left( \lceil (1 - \alpha)B \rceil + 1 \right) \)th order statistic of \( \hat{S}_D \). Finally, we construct a MC SPI by

\[
I_{1-\alpha}^{MC} : \left[ \hat{\mu}_d - \hat{c}_{1-\alpha} \times \text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d), \hat{\mu}_d + \hat{c}_{1-\alpha} \times \text{mse}_{(\cdot)}^{1/2}(\hat{\mu}_d) \right] \forall d \in [D].
\]

A construction of MC SPIs resembles the derivations of the volume of a tube based SPIs in S.D., but without a correction for the variability of \( \theta \). Here an implicit correction is included due to MSE estimation.

To construct MC SPIs, we assume normality for errors and random effects. Bootstrap can circumvent a direct application of the normal asymptotic distribution and can provide faster convergence rates. Let \( B \) now be the number of bootstrap samples \( (y^{(b)}, X, Z) \), \( b = 1, \ldots, B \) and \( c_{1-\alpha} \) the \( (1 - \alpha)^{th} \) quantile of the distribution of \( S_D \); we claim that \( c_{1-\alpha} \)
can be estimated by $c_{1-\alpha}^*$ as the $((1-\alpha)B) + 1)^{th}$ order statistic of

$$
S_{D}^{*\{b\}} = \max_{d=1,\ldots,D} \left| \frac{\hat{\mu}_d^{*\{b\}} - \mu_d^{*\{b\}}}{\text{MSE}^{1/2}(\hat{\mu}_d^{\{b\}})} \right|^1.
$$

(20)

Then, bootstrap SPI is defined as

$$
I_{1-\alpha}^B : [\hat{\mu}_d - c_{1-\alpha}^* \times \text{MSE}^{1/2}(\hat{\mu}_d), \hat{\mu}_d + c_{1-\alpha}^* \times \text{MSE}^{1/2}(\hat{\mu}_d)] \quad \forall d \in [D].
$$

(21)

Notice that the bootstrapped samples can be used twice, to obtain a consistent estimator of MSE and to approximate the distribution of (16). We use the same $\text{MSE}^{1/2}(\hat{\mu}_d^{\{b\}})$ in (20) and (21). As pointed out in Remark 2, it is calculated using all bootstrap samples.

**Remark 3.** Choosing a suitable bootstrap in LMM is not trivial (see Flores Agreda (2017)). We tried several of his algorithms in addition to our modifications. Parametric bootstrap (PB) does not work well for NERM due to the shrinkage effect of EBLUP. Moment matching bootstrap (Hall and Maiti, 2006a) yields promising results for estimating extreme tails, but not for the (entire) p.d.f. of $S$. Regarding other schemes, some are not directly applicable (e.g., based on explicit formula of the likelihood), others (e.g., based on sampling the clusters) failed badly due to the underestimation of the variation. Note that for the FHM, the normality assumption recommends the use of PB.

We describe two bootstrap methods which yield promising results in the construction of SPIs. The first one, a random effect bootstrap (REB) is recommended for NERM. REB refers to sampling with replacement from empirical distributions of random effects and errors. $SRSWR[\delta, n]$ denotes a simple random sample with replacement of size $n$ from set $\delta$. Carpenter et al. (2003) suggest including an inflation procedure to counteract the shrinking effect which would cause undercoverage. The algorithm is as follows:

1. From the original sample, obtain consistent estimators $\hat{\beta}$ and $\hat{\theta} = (\hat{\sigma}_e^2, \hat{\sigma}_u^2)$.
2. Consider EBLUPs predictions $\hat{u}_d$ and residuals $\hat{e}_{dj} = y_{dj} - x_{dj}^T \hat{\beta} - \hat{u}_d$. To ensure that they have empirical variances $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$ respectively, scale both of them, that is

$$
\hat{u}_d^s = \hat{\sigma}_u \hat{u}_d [D^{-1} \sum_{k=1}^D \hat{u}_k^2]^{-1/2} \quad \text{and} \quad \hat{e}_{dj}^s = \hat{\sigma}_e \hat{e}_{dj} [n^{-1} \sum_{l=1}^n \sum_{k=1}^D \hat{e}_{kl}^2]^{-1/2}.
$$

12
3. Sample independently from $\hat{u}^s$ and $\hat{e}^s$, that is $u^* = SRSWR[\hat{u}^s, D]$ and $e^* = SRSWR[\hat{e}^s, n]$.

4. Create a bootstrap sample $y^* = X\hat{\beta} + u^* + e^*$.

5. Fit the nested error regression model (4) to the bootstrap sample from Step 4 and obtain bootstrap estimates $\hat{\beta}^*$, $\hat{\theta}^* = (\hat{\sigma}_e^2, \hat{\sigma}_u^2)$, $\mu^*$ and $\hat{\mu}^*$.

6. Repeat Steps 2-5 $B$ times. Calculate $S_{\mu}^{*\beta(b)}$, $b = 1, \ldots, B$ using bootstrapped MSE. Obtain $c_{1-\alpha}^*$ and $I_{1-\alpha}^B$ as defined above.

In step 2, one can additionally consider centering of the scaled residuals. In our case this does not lead to a numerical gain.

The second bootstrap procedure is PB. It is similar to REB, but based on sampling from normal distributions with suitable variance structure. PB is the most popular in SAE, especially for FHM where we use the known $\sigma_{ed}$. To implement PB, we need to slightly modify step 1 in REB and define $\hat{\theta} = \hat{\sigma}_u^2$. Additionally, we need to replace step 2 and 3 by

2. Generate $D$ independent copies of a variable $W_1 \sim N(0,1)$. Construct the vector $u^* = (u_1^*, u_2^*, \ldots, u_D^*)$ with elements $u_d^* = \hat{\sigma}_u W_1$, $d = [D]$.

3. Generate $D$ independent copies of a variable $W_2 \sim N(0,1)$. Construct the vector $e^* = (e_1^*, e_2^*, \ldots, e_D^*)$ with elements $e_d^* = \sigma_{ed} W_2$, $d = [D]$.

Extensions that include second-stage bootstrap steps to obtain bias-corrected MSE estimators are provided in the S.D. For the sake of presentation we limited ourselves to create bootstrap analogues for the units with complete observations $(y_{dj}, x_{dj})$. It is straightforward (see our S.D.) to extend this also to units with missing responses (see Remark 1).

### 3.2 Bootstrap SPI consistency

As the consistency of MC SPI follows immediately when errors and random effects are normal, and does not in any other case, we concentrate on the consistency of the bootstrap SPI. We show that the c.d.f. of $S_{D}^{*\beta\mu}$, conditional on the initial sample, converges to the same
as $S_D$. Then we can conjecture that $I^B_{-\alpha}$ has an asymptotically correct coverage probability. LMMb \cite{2} is estimated by restricted maximum likelihood (REML), but one could equally well use the method of moments (MM). \cite{Richardson and Welsh (1994)} prove that under mild conditions and without assuming normality for the random effects and errors, $\theta$ and $\beta$ obtained using REML (but using the likelihood of normal distribution) are consistent and normally distributed. Under different conditions, \cite{Jiang (1998)} expands this result to non-hierarchical LMM. Both also allow for asymmetric errors and random effects, whereas most of the SAE literature keeps the symmetry assumption to assure unbiasedness of the estimator of $\mu_d$. Otherwise we would need to correct for potential biases. While this is feasible, it would blur our contribution, making notation and procedure more complex, and deviate from what is typically assumed in SAE. Under normality \cite{González-Manteiga et al. (2008b)} derive consistency of the bootstrap for the MSE of parameter estimates obtained by Henderson’s method. Following this discussion, we use REML, assume symmetry for the NERM, but keep the normality assumptions for the FHM. Theory is first derived for NERM from which we then can conclude a corollary for the FHM.

Due to the hierarchical structure, we obtain independent sub-vectors with the data from each area such that the log-likelihood can be written as

$$
\ell(\beta, \theta) = -\frac{1}{2} \sum_{d=1}^{D} \left[ \log ||V_d|| + \frac{1}{2}(y_d - X_d\beta)'V_d^{-1}(y_d - X_d\beta) \right], \quad d = [D].
$$

For a fixed $\theta$, $\tilde{\beta}$ is defined in \cite{7} being the maximizer of a log-likelihood. If $\theta$ needs to be estimated, REML can be used. In this case the likelihood is given by

$$
\ell_R(\tilde{\beta}) = -\frac{1}{2} \log ||V|| - \frac{1}{2} \log ||X'V^{-1}X|| - \frac{1}{2} y'Py + M,
$$

with $M$ an additive constant and $P = V^{-1} - V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$. Thus, our estimating equation is

$$
\psi_i(\theta) = -\text{tr}[PZ_iZ_i'Y] + (y - X\tilde{\beta})'V^{-1}Z_iZ_i'V^{-1}(y - X\tilde{\beta}) = 0, \quad i = 1, \ldots, h, \quad (22)
$$
which corresponds to the number of variance parameters $\theta = (\theta_1, \ldots, \theta_h)$ to be estimated. Under the above setting, [Richardson and Welsh (1994)] proved the following

**Proposition 1.** Suppose that assumptions A.1-A.10 and conditions R.1-R.5 from Appendix A.1 hold. Then there exists $\hat{\theta}$ which is a solution to the estimating equations (22) such that:

(i) $||\hat{\theta} - \theta|| = O_p(n^{-1/2})$,

(ii) $||\hat{\beta} - \beta|| = O_p(n^{-1/2})$,

(iii) $n^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0_{h \times h}, B^{-1} \Sigma_\theta B)$,

(iv) $n^{-1/2}(\hat{\beta} - \beta) \xrightarrow{d} N(0_{(p+1) \times (p+1)}, \Sigma_\beta^{-1})$,

where $B$, $\Sigma_\theta$ and $\Sigma_\beta$ are block-matrices with blocks defined in Assumption A.5.

The bootstrap parameter consistency is owed to the bootstrap schemes ensuring that $e^*$ and $u^*$ are mutually independent. Furthermore

$$E^*(y^*) = X\hat{\beta} \quad \text{and} \quad \text{Var}^*(y^*) = V(\theta) = \hat{R} + \hat{GZ}\bar{Z}^\top.$$ (23)

Hence the bootstrap samples imitate the properties of the original sample yielding

**Proposition 2.** Under the assumptions of Proposition 1 it holds that

(i) $E^*(y^*) - E(y) = [o_p(1)]_{n \times 1}$,

(ii) $\text{Var}^*(y^*) - \text{Var}(y) = [o_p(1)]_{n \times n}$,

(iii) $||\hat{\theta} - \hat{\theta}|| = O_p(n^{-1/2})$,

(iv) $||\hat{\beta} - \hat{\beta}|| = O_p(n^{-1/2})$.

**Proof.** The results in (i) and (ii) follow immediately applying a definition of these moments given in (3) and (23) as well as Proposition 1. (iii) and (iv) have been proved by Carpenter et al. (2003) for REB and González-Manteiga et al. (2008b) for PB, see also S.D.

We now propose the derivation of the consistency of $I_{1-\alpha}^B$ using ideas of Hall and Pettitkel (1990) and Chatterjee et al. (2008). Define a process $\Delta_d = (\hat{\mu}_d - \mu_d)/\text{MSE}^{1/2}(\hat{\mu}_d)$ and denote its c.d.f. $F_D(v) = P(\Delta_d \leq v)$. We need to specify a critical value $c_{1-\alpha}$. Ideally, $c_{1-\alpha}$ would be determined from equation (15) which we can slightly modify using $\Delta_d$ as

$$\mathcal{P}(c_{1-\alpha}) = P(-c_{1-\alpha} \leq \Delta_d \leq c_{1-\alpha} \forall d \in [D]) = 1 - \alpha.$$ (15)

Since $\mathcal{P}(c_{1-\alpha})$ is unknown, we approximate it by bootstrap

$$\mathcal{P}^*(c_{1-\alpha}^*) = P(-c_{1-\alpha}^* \leq \Delta_d^* \leq c_{1-\alpha}^* \forall d \in [D]|W) = 1 - \alpha.$$
with $\mathcal{W} = \{(y_{dj}, x_{dj}), d \in [D], j \in [n_d]\}$, $\Delta_d^* = \frac{\hat{\mu}_d(\hat{\theta}) - \mu_d^*}{\text{MSE}^{1/2}(\hat{\mu}_d)}$ and c.d.f. $F_D(v) = P(\Delta_d^* \leq v | \mathcal{W})$. Observe that $c_{1-\alpha}$ corresponds to the high quantile of the c.d.f. of the order statistics $S_D^*$ in [20]. Hence, if we prove that $\mathcal{P}$ and $\mathcal{P}^*$ are asymptotically close up to the order $O_P(D^{-1})$, it implies the same order of accuracy for $I_{1-\alpha}^B$. Define

$$R = \{x \in \mathcal{X} : -D^{1/2}c_{1-\alpha} \leq D^{1/2}\Delta_d \leq D^{1/2}c_{1-\alpha} \forall d \in [D]\}$$

which can be represented as a finite number of unions and intersection of convex sets. This number is bounded uniformly for $D \geq 2$ and $c_{1-\alpha} > 0$. Observe that $\mathcal{P} = \int_R dF_D$. Therefore, if we can show that for all continuity points $v$ the c.d.f.’s of $\Delta_d$ and $\Delta_d^*$ converge to the same limit with a desired speed, the same speed is maintained in the convergence of $\mathcal{P}$ and $\mathcal{P}^*$. Thus, to state the final theorem we need to show (proof in Appendix A.2) that

**Lemma 1.** Under Proposition 1 and 2, it holds for all continuity points $v$ that

$$\sup_{v \in \mathbb{R}} |F_D(v) - F_D^*(v)| = O_P(D^{-1}).$$

Since $\mathcal{P}$ is defined as an integral of $dF_D$ over $R$, a direct consequence of Lemma 1 is

**Theorem 1.** Under Lemma 1, it holds for $c = c_{1-\alpha}$ with $\alpha \in (0, 1)$ that

$$\sup_{c \in \mathbb{R}} |\mathcal{P}(c) - \mathcal{P}^*(c)| = O_P(D^{-1}).$$

An important implication of this theorem is the coverage probability of $I_{1-\alpha}^B$, i.e.

**Theorem 2.** Under Theorem 1, if $c_{1-\alpha}^*$ is such that $\mathcal{P}^*(c_{1-\alpha}^*) = 1 - \alpha$, it holds that

$$P\left(\mu_d \in I_{1-\alpha}^B \forall d \in [D]\right) = 1 - \alpha + O(D^{-1}).$$

FHM is a widely used example of LMMb where explicit distributional assumptions on errors and random effects are imposed. Due to its broad applicability in SAE, we state the results for this model in a separate Corollary.
Corollary 1. Assume the FHM as defined above, $\sigma_e^2$ lies in a compact set of $(0, \infty)$ and $\hat{\sigma}_u^2$ is positive. Denote the estimate of the intraclass correlation coefficient (ICC) by $\hat{\gamma}_d = \hat{\sigma}_u^2/(\hat{\sigma}_u^2 + \sigma_e^2)$. Then for $\hat{\mu}_d^{FH} = \hat{\gamma}_d y_d + (1 - \hat{\gamma}_d)x_d^t\beta$, $\hat{\mu}_d^{FH*} = \hat{\gamma}_d^* y_d^* + (1 - \hat{\gamma}_d^*)x_d^t\beta^*$ $d \in [D]$, 

$$P \left( \hat{\mu}_d^{FH} \in I_{1-\alpha}^B \forall d \in [D] \right) = 1 - \alpha + O(D^{-1});$$

and $\epsilon_{1-\alpha}^*$ satisfies 

$$P^* \left( \hat{\mu}_d^{FH*} \in I_{1-\alpha}^B \forall d \in [D] \right) = 1 - \alpha + O_P(D^{-1}).$$

Remark 4. The asymptotics are of order $D^{-1}$ as they depend on the bias. If one is interested only in the API, and works only with variance estimates (Chatterjee et al., 2008) then it is possible to obtain faster rates. For simultaneous inference, however, it is preferable to account for the bias. Simulations show that accounting for bias is particularly important when deviating from normality.

4 Simulation study

Simulations were carried out to study the performance of SPI for various scenarios using different MSE and/or parameter estimates, and to compare it with API. The latter are calculated with the same MSE estimates as the SPI, and critical values refer to normal quantiles as it is common in the SAE literature. It might be interesting to construct API with a bootstrapped quantile and MSE, but to the best of our knowledge this has yet to be done, and is therefore beyond the scope of this paper. We present here the following scenarios (more results can be found in S.D.): MC and bootstrap SPI are based on $B = 1000$, with $C = 1$ for the double-bootstrap. In all scenarios $\forall d$ set $x_{d1} = 1$, $x_{d2} \sim U(0, 1)$ for NERM, and $\bar{x}_{d2} \sim U(0, 1)$ for FHM, with $\beta = (1, 1)$. We draw $M_s = 500$ samples for three types of sample sizes $D : n_d = \{(25:5), (50:10), (75:15)\}$. 

17
Our simulation study starts with the NERM in which we allow $u_d$ and $e_{dj}$ to deviate from normality, and even to be asymmetric. Namely, we draw them from normal, student-$t$, and chi-square distributions, always centered to zero and rescaled to variances $\sigma^2_e, \sigma^2_u$ (indicated in parenthesis in Table 1) such that the ICC equals $1/3$ or $2/3$. Results hardly differ when estimating $\theta$ using either REML or MM. Different MSE estimates can lead to different coverage probabilities, and so does the use of either MC or bootstrap SPIs. Therefore, we compare the empirical coverage probability (ECP) of SPI which is

$$ECP = \frac{1}{M_s} \sum_{k=1}^{M_s} \mathbb{1}\{\mu^{(k)}_d \in I^L_{1-\alpha} \forall d \in [D]\} \quad \text{where } L = MC \text{ and } L = B, \text{ respectively.}$$

In Table 1 ECP is compared for 95% SPI based on MC and bootstrap respectively using all mentioned MSE estimators (except $mse_L$ which never did best). Parameters $\beta$ and $\theta$ were estimated with REML; results for MM are deferred to S.D.

<table>
<thead>
<tr>
<th>$e_{dj}$</th>
<th>$u_d$</th>
<th>$mse^*_{B1}$</th>
<th>$mse^*_{BC2}$</th>
<th>$mse^*_{ST}$</th>
<th>$mse^*_{SP}$</th>
<th>$mse^*_{SPA}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(1)$</td>
<td>$N(0.5)$</td>
<td>MC 92.2 94.2 94.4 91.6 94.6 94.6 91.6 94.4 94.4 92.4 94.2 94.6 92.4 94.2 94.6</td>
<td>B 92.2 94.0 94.2 92.0 93.8 94.4 92.4 94.4 92.4 92.6 94.6 94.6 92.8 94.4 94.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N(0.5)$</td>
<td>$\chi^2_5(1)$</td>
<td>MC 93.6 94.8 95.6 93.6 94.2 95.6 93.8 95.0 95.6 93.0 94.6 95.8 92.8 94.6 95.8</td>
<td>B 93.6 94.4 95.6 93.6 94.0 95.2 93.6 94.4 95.6 93.0 94.0 95.8 93.0 94.0 95.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2_5(1)$</td>
<td>$N(0.5)$</td>
<td>MC 90.4 88.2 90.6 90.8 88.6 90.2 90.4 88.4 90.6 89.8 88.6 90.2 89.8 89.0 90.4</td>
<td>B 92.2 91.2 93.2 91.8 91.2 93.2 92.4 91.2 93.4 91.4 93.4 92.6 91.6 93.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_6(0.5)$</td>
<td>$N(1)$</td>
<td>MC 89.4 91.0 92.2 89.8 90.8 91.8 89.4 91.0 92.4 89.6 90.0 92.0 89.4 90.0 92.0</td>
<td>B 92.6 93.2 95.4 92.2 92.8 95.4 92.8 93.2 95.6 93.6 95.8 93.6 93.6 95.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$t_6(1)$</td>
<td>$t_6(0.5)$</td>
<td>MC 87.8 90.6 91.6 88.8 91.0 91.8 88.2 90.6 91.6 87.8 90.0 91.2 87.4 90.2 90.8</td>
<td>B 90.8 92.6 94.2 91.0 92.8 93.6 91.2 92.6 94.0 91.2 93.4 95.0 91.2 93.4 95.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: $ECP$ (in %) for MC and Bootstrap SPI using different MSE estimators under NERM with sample sizes $D : n_d$ as indicated. Nominal coverage probability is 95%.

We conclude that, even for growing sample size MC based SPI performs well if errors and random effects are normally distributed, and sample size is large. This does not
change for different MSE estimators; the results are actually quite similar. Nonetheless, the performance of bootstrap based SPI is more satisfactory. While errors and random effects are symmetrically distributed, they attain the nominal level for moderate sample sizes. They suffer from a slight undercoverage in the case of skewed errors, which is probably due to the bias in the two stage estimator $\hat{\mu}_d$. In contrast, an asymmetric distribution of $u_d$ hardly affects ECP. Simulations for 90% an 99% SPI show the same features.

We also studied the length increase when switching from API to SPI in order to verify their usefulness. In practice, intervals which are too wide are not desirable, and the additional length may not be justified if API and SPI provide essentially the same information. Figure 1 displays 95% bootstrap SPI (gray) with API (black) for small area means with $mse_{\xi PA}$. White dots represent the true means $\mu_d$. Out of the 50 $\mu_d$, 3 are far outside of their API and another 7 are lying on the boundaries such that just a slight underestimation of the MSE provokes that for one given data set 20% of all API no longer contain the true parameter. The same pattern appears under the scenario with normally distributed random effects and errors (see simulation results in S.D.). Recall that this does not happen by chance, but by construction: for $100(1 - \alpha)$% API about $100\alpha$% (often many more) of the area parameter are not in their intervals. In contrast, our SPI contain all of the true area means. Moreover, the SPI do not seem to be excessively wide but just as wide as necessary: one area mean is right at the upper boundary, and another one at the lower boundary of its SPI.

Additional tables and figures can be found in our S.D., but in brief we can summarize the main findings as follows: one cannot claim that a certain MSE estimator for constructing SPI is more efficient than another, in the sense that one obtains the same ECP with a narrower SPI. The increase of length from API to SPI is the smallest when the $e_{dj}$ are normal. Moreover, under normality the length varies much less over different samples. For $(D : n_d) = (25 : 5)$ the average ratio of length(SPI)/length(API) is between 1.51 to 1.64,
Figure 1: API and bootstrap SPI for small area means estimated using REML, $mse_{SPA}^*$, $e_{dj} \sim t(0.5)$, $u_d \sim N(1)$, $D = 50$.

for (75 : 15) between 1.71 to 1.83. Surprisingly, when we fix the distributions of $u_d$ and $e_{dj}$ up to their variances and increase the ICC from 0.1 to 0.8, these values hardly change.

Under FHM, we apply a similar setting as under NERM; namely the sample size $D$, and the number of bootstrap and MC samples remain unchanged. The random effects and error terms are centered and normally distributed with unknown variance $\sigma_u^2$ and known heteroscedastic $\sigma_{e_d}^2$. Following the simulation study of Datta et al. (2005), each fifth part of the total number of areas is assigned to a different value for $\sigma_{e_d}^2$: in Scenario 1: 0.7, 0.6, 0.5, 0.4, 0.3; in Scenario 2: 2.0, 0.6, 0.5, 0.4, 0.2; and in Scenario 3: 4.0, 0.6, 0.5, 0.4, 0.1. Variance $\sigma_u^2$ is estimated using REML, Henderson’s method 3 (Prasad and Rao, 1990), and the method introduced by Fay and Herriot (1979). The results for using the latter are not shown as they typically lie between those for the former, but closer to that under REML which seems to yield the most promising outcomes. Results for $mse_L$ and $mse_{SPA}^*$ are not included because the former never perform best and the latter are equivalent to those with $mse_{SP}$ under normality. In Table 2 we can see that for small samples, the MC SPIs provide more accurate results, where bootstrap SPIs suffer from a slight overcoverage. In
contrast, the latter is more reliable for larger samples. Generally, there is once again no clear winner, all our proposed methods seem to also work well for the FHM under known heteroscedasticity and normality.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>mse_{B1}</th>
<th>mse_{BC2}</th>
<th>mse_{3T}</th>
<th>mse_{SP}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95.4</td>
<td>92.6</td>
<td>94.0</td>
<td>95.0</td>
</tr>
<tr>
<td>1</td>
<td>92.6</td>
<td>94.0</td>
<td>95.0</td>
<td>92.6</td>
</tr>
<tr>
<td>1</td>
<td>94.0</td>
<td>95.0</td>
<td>92.6</td>
<td>94.2</td>
</tr>
<tr>
<td>1</td>
<td>92.6</td>
<td>94.0</td>
<td>95.0</td>
<td>92.6</td>
</tr>
<tr>
<td>1</td>
<td>95.0</td>
<td>92.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>92.4</td>
<td>94.0</td>
<td>95.2</td>
<td>92.8</td>
</tr>
<tr>
<td>1</td>
<td>92.8</td>
<td>94.6</td>
<td>95.0</td>
<td>92.2</td>
</tr>
<tr>
<td>1</td>
<td>94.6</td>
<td>95.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>95.0</td>
<td>92.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: ECP (in %) for MC and bootstrap SPI using different MSE estimators under FHM for $D = 25, 25, 75$ as indicated. Nominal coverage probability is 95%.

5 Application to income data

We apply our method to construct SPI for average household income in 52 counties of Galicia, Northern Spain. It is of great interest for the Statistical Office and politicians alike to gather information about the average income of individuals and households. In particular, it is important to extend statistical analysis to the county level of so-called comarcas to be able to adjust regional policies and implement new programs. In this study we focus on the construction of an interval estimate on the household level. We make use of the general part of the Structural Survey for Homes in Galicia in 2015 with 9203
households in total, but in some areas \( n_d < 20 \). The survey contains information about the total income as well as different characteristics on individual and household level. The dependent variable refers to the total, yearly income which consists of paid work, own professional activity and miscellaneous benefits (like employment). To fit the model we used 18 covariates – 11 of them contain the information on the household level (composition, type, financial difficulties, the size of municipality where the household is located) and 7 on the individual level of the main sustainer (nationality, age, education, gender). A detailed definition of variables, descriptive statistics, estimated REML coefficients, standard deviations, p-values, and the discussion of the model selection process is deferred to S.D.

![Figure 2: Cholesky REML residuals: (left) kernel density estimation and (middle) QQ plot; REML empirical Bayes estimates of random effects: (right) QQ plot.](image)

Before addressing the construction of SPI, we focus on the normality assumptions of errors and random effects. It is well known that income data are right skewed and, unsurprisingly, our dependent variable exhibits this feature. We tried different transformations to income (natural logarithm, square root and fourth root) to obtain an approximately symmetric density function. We found that the square root led to the best model adequacy which can be seen in the plot of the kernel density function of Cholesky residuals (Jacqmin-Gadda et al., 2007) in the left panel of Figure 2. The uncorrelated Cholesky residuals are
constructed by multiplying \( y - X\hat{\beta} \) by the Cholesky square root of the variance matrix. The departure from normality is visualized using a QQ plot displayed in the middle panel of Figure 2. The pattern of heavy-tails is obvious which implies that the residuals deviate from normality. This is also detected by the Shapiro-Francia test and a high value of the Royston’s \( V' \) statistics ([Royston 1993] with a p-value \( p < 0.0001 \) and \( V' = 39.117 \). As deviations from normality of errors has a strong impact on the coverage probability of MC SPI, these results suggest to focus on bootstrap SPI only. As long as the normality of random effects is considered, the right panel of Figure 2 displays a diagnostic plot of [Lange and Ryan 1989] using standardized empirical Bayes estimates of random effects in a weighted normal QQ plot; it supports the adequacy of the normality assumption. [Ritz 2004] constructed a test based on the weighted empirical distribution function of the standardized random effects and their expectations. We applied three goodness-of-fit statistics based on this empirical process; that is Kolmogorov-Smirnov, Cramer von Mises and Anderson-Darling. Their p-values were 0.278, 0.087 and 0.156 and are therefore far from significant.

Figure 3: 95% API and bootstrap SPI for mean income of comarcas in Galicia.

Figure 3 presents API and bootstrap SPI with \( mse_{SPA}^* \) (our choice from the simulation results - see Table 1) of the square root of means of income in the counties of Galicia. Tables with the average widths of the SPIs and APIs can be found in S.D. We divided the
plot in Figure 3 into 5 panels based on the number of units in each area. We can see a lot of variability over the estimates in the areas. Evaluating the results of the API (dark grey) versus the SPI (light grey), it is apparent that the prediction intervals are not adequate to address neither a joint consideration nor a comparison of the areas. If we consider, for example, the areas of A Mariña Oriental and Chantada (5th and 6th regions of the second panel), the APIs indicate significantly different incomes, whereas the SPIs do not confirm this. In Figure 3 it can be clearly seen that there are many other cases in which the API would insinuate significant differences in mean income between comarcas whereas the SPI indicates that this difference is indeed insignificant. This does not mean that SPIs are generally too wide for a practical use; we detect significant differences between several comarcas – but now such comparison is valid, whereas it is not when considering APIs (recall our discussions in the introduction and simulation study). In Figure 4 we use SPIs to construct maps of comarcas which present the lower and upper limits of the bootstrap SPIs. One detects a substantial variation of the household income over the counties. The comarcas of A Coruña and Santiago de Compostela are the richest (with a large number of units) and they are indicated in the last panel of Figure 3. A Paradanta and O Ribeiro, being located in the inner zone where inhabitants live off the agriculture, are the poorest regions; we marked them in the first and the second panel of Figure 3. Furthermore, we can see that in the south-eastern region there is a group of relatively poor comarcas, with the exception of Ourense (a large area). Similar conclusions can be drawn from Lombardía et al. (2018) and from the publications of the Galician Institute of Statistics. Observe that APIs should not be used to make such maps as this would suggest that we were allowed to compare them. Moreover, APIs would not contain at least 3 of true area parameters.
6 Conclusions

We develop two frameworks to construct SPIs for small area means under LMMb. We derive theoretical formulas based on the volume of a tube and use MC approximations to make them operational. Furthermore, we construct SPIs based on bootstrap and prove its consistency. We study the sensitivity of SPI to different MSE estimators and deviations from normality. The results confirm that MC SPI does not perform well unless errors and random effects are normally distributed. On the contrary, this problem is alleviated by bootstrap SPI with an approximately correct ECP for any kind of distribution of random effects and errors. However, for the FHM the MC SPI is still attractive as it can outperform bootstrap SPI if the normality assumptions hold. Therefore a practitioner should always verify, at least using graphical tools such as QQ plots, whether there is a suspicion of the violation of this assumption.

It is clear that accounting for the variability of all areas makes SPI larger than API,
yet only with SPIs comparisons between areas are statistically valid. Moreover, we can be
assured that if we conduct several surveys, SPIs would contain all true area parameters in
$100(1 - \alpha)\%$ of studies whereas by construction about $D\alpha$ of true area parameters are not
inside their APIs for each survey. Finally, we illustrate the use of this methodology along
a small study on average household income in the comarcas of Galicia.

The proposed techniques can be extended to any hierarchical model with several nested
random effects for each subdomain as well as to the longitudinal models with temporal cor-
relation (e.g. to study the average levels of lymphocytes of patients or average precipitation
at meteorological stations). Moreover, our methodology can be developed to be applied
for modeling more complex data structures using generalized linear, semiparametric mixed
models or models which account for spatial correlation or heteroscedastic errors.

A Appendix

A.1 Assumptions and regularity conditions

A.1 LMMb is nested, that is, each row of $n \times D$ matrix $Z$ contains exactly one 1 and
zeros and no column is composed only of zeros.
A.2 LMMb has a hierarchical structure, therefore if we suppose that $z(k)$ denotes the $k^{th}$
column of the matrix $Z$; then, there exists a constant $Q$ such that $z(k)^{t}z(k) \leq Q \forall n$.

We need some conditions which guarantee continuity and control the asymptotic behavior
of the score equations (22) and their derivatives. Since for $\|A\|, |Ax| \leq \|A\||x|$, it follows
that $|x^{t}Ax| \leq \|A\||x|^{2}$ and $\|AB\| \leq \|A\||B|$.
A.3 $\exists \delta > 0$ and a constant $M$ s.th. $E|y_{d} - X_{d}\beta|^{4+\delta} \leq M \forall d$, $n^{-1}\sum_{d=1}^{D}||X||^{2+\delta} = O(1)$.
A.4 All elements of $\theta$ are positive and finite. Furthermore, there exists a neighbourhood
\(N_{\theta,\varepsilon} = \{\theta_{x} \mid \|\theta - \theta_{x}\| \leq \varepsilon\}\) such that $\forall d$, $V_{d}$ is nonsingular $\forall \theta \in N_{\theta,\varepsilon}$ and $||V_{d}^{-1}||$ is
uniformly bounded in $\theta \in N_{\theta,\varepsilon}$.
A.5 For a nonsingular \((p + 1) \times (p + 1)\) matrix \(\Sigma_\beta\) and for \(i, k = 1, \ldots, h\), we have

\[
\begin{align*}
(a) & \quad \frac{1}{n} \sum_{d=1}^{D} X_d^t V_d^{-1} X_d \to \Sigma_\beta \\
(b) & \quad \frac{1}{n} \sum_{d=1}^{D} \text{tr}\{V_d^{-1}[Z_d Z_d^t]_d V_d^{-1}[Z_k Z_k^t]_d\} \to B_{i,k} \text{ is } h \times h \text{ positive definite matrix } \\
(c) & \quad \frac{1}{n} \sum_{d=1}^{D} \left\{ \mathbb{E}(y_d - X_d \beta)^t V_d^{-1}[Z_d Z_d^t]_d V_d^{-1}(y_d - X_d \beta)(y_d - X_d \beta)^t \right. \\
& \quad \quad \left. \times V_d^{-1}[Z_k Z_k^t]_d V_d^{-1}(y_d - X_d \beta) \right\} - \text{tr}\{V_d^{-1}[Z_d Z_d^t]_d \text{tr}\{V_d^{-1}[Z_k Z_k^t]_d\}\} \to \Sigma_{\theta_i,k}
\end{align*}
\]

A.6 For each fixed \(y\), a score equation (22) is continuously differentiable and \(\mathbb{E}[\psi_i(\theta)] = 0\) if \(\theta\) is a true parameter value.

A.7 \(\lim \inf_{n} \lambda[n^{-1} \text{Var}(s_n(\theta))] > 0\) and \(\lim \inf_{n} \lambda[-n^{-1} \mathbb{E}(\nabla s_n(\theta))] > 0\) where \(s_n(\theta) = \sum_{i}^{h} \psi_i\), \(\nabla s_n(\theta) = \frac{\partial s_n(\theta)}{\partial \theta}\) and \(\lambda[A]\) indicates the smallest eigenvalue of a matrix \(A\).

Moreover, to assure unbiasedness of the two-stage estimator \(\hat{\mu}_d := t(\hat{\theta}, y)\), i.e. \(\mathbb{E}[t(\hat{\theta}, y) - \mu] = 0\) and the validity of the asymptotic expansion of the c.d.f. of \(\hat{\mu}_d(\hat{\theta})\) (cfr. Hall (1992)), we assume as follows:

A.8 \(\mathbb{E}(t(\hat{\theta}))\) is finite.

A.9 The elements of \(\hat{\theta}\) are even translation-invariant functions of \(y\), that is \(\hat{\theta}(\hat{\theta}) = \hat{\theta}(y)\), and \(\hat{\theta}(y - X a) = \hat{\theta}(y)\), for any \(a \in \mathbb{R}^{p+1}\) and for all \(y\).

A.10 The distributions of random effect \(u\) and errors \(e\) are both symmetric with mean 0.

R.1 \(X_d\) and \(Z_d\) are uniformly bounded s.t. \(X_d^t V_d^{-1} X_d = [O(D)]_{(p+1) \times (p+1)}\) ∀\(d\).

R.2 Covariance matrices \(G_d\) and \(R_d\) have a linear structure, that is \(G_d = \sum_{j=0}^{h} \theta_i J_{dj} J_{dj}^t\) and \(R_d = \sum_{j=0}^{h} \theta_i T_{dj} T_{dj}^t\), where \(d = 1, \ldots, D, j = 0, \ldots, s \theta_0 = 1, J_{dj}\) and \(T_{dj}\) are known of order \(n_d \times q_d\) and \(q_d \times q_d\) respectively; in addition, the elements of \(J_{dj}\) and \(T_{dj}\) are uniformly bounded known constants such that \(G_d\) and \(R_d\) are positive definite matrices. In certain cases, \(J_{dj}\) and \(T_{dj}\) can be null matrices.
R.3 Rate of convergence: $D \rightarrow \infty$ such that $D = o(n)$ and $n_d \rightarrow \infty$ such that $n_d = o(D/n)$.

R.4 To ensure the nonsingularity of $\Sigma_{\theta}$, $0 < \inf_{d \leq 1} \sigma^2_{e_d} \leq \sup_{d \leq 1} \sigma^2_{e_d} < \infty$ and $\sigma^2_u \in (0, \infty)$.

R.5 $E[e^8_{d_j}] < \infty$ and $E[u^8_{d_j}] < \infty$.

The conditions regarding continuity and score equations are as in Richardson and Welsh (1994) and Jiang (1998), whereas the conditions concerning $\theta$ are justified by Kackar and Harville (1981) and Datta and Lahiri (2000). Kackar and Harville (1981) proved that many of the standard methods of estimation (like REML) yield even translation invariant estimators of $\theta$. In the following we focus on $\theta$ obtained using REML, but similar results (i.a. regarding bootstrap consistency) can be found for other estimators (González-Manteiga et al., 2008b).  

A.2 Proof of Lemma 1

We develop an asymptotic expansion of the c.d.f. $F_D$. Given an initial sample and assuming the consistency of the bootstrap parameters, the same expansion holds for $F_D^*$ if we replace $\beta$ and $\theta$ with $\hat{\beta}$ and $\hat{\theta}$ respectively. For now, we drop a subscript $1 - \alpha$ and we denote $c_{1-\alpha} = c$. Using calculations similar to those in Chatterjee et al. (2008) and Yoshimori and Lahiri (2014), we have:

$$F_D(c) = P\left(\frac{\hat{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\hat{\mu}_d)} \leq c\right) = P\left(\frac{\hat{\mu}_d - \hat{\mu}_d + \bar{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\hat{\mu}_d)} \leq c\right)$$

$$= P\left(\frac{\bar{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\bar{\mu}_d)} \leq c + \frac{c \left(\text{MSE}^{1/2}(\hat{\mu}_d) - \text{MSE}^{1/2}(\bar{\mu}_d)\right) - (\hat{\mu}_d - \bar{\mu}_d)}{\text{MSE}^{1/2}(\bar{\mu}_d)}\right)$$

$$= \mathbb{E}\left[P\left(\frac{\bar{\mu}_d - \mu_d}{\text{MSE}^{1/2}(\bar{\mu}_d)} \leq c + \frac{c \left(\text{MSE}^{1/2}(\hat{\mu}_d) - \text{MSE}^{1/2}(\bar{\mu}_d)\right) - (\hat{\mu}_d - \bar{\mu}_d)}{\text{MSE}^{1/2}(\bar{\mu}_d)}\right) \mid \mathbf{y}_d\right]\right]$$

$$=: \mathbb{E}[F(c + T_d(c, \mathbf{y}))]$$
\[
F(c) + E[T_d(c, y)] = F'(c) + 2^{-1} E \left[ \int_c^{c+T_d(c, y)} (c + T_d(c, y) - x) F''(x) dx \right] =: F(c) + F'(c)V_1(c) + V_2(c).
\]

Note that under normality assumption for errors and random effects, \(F(c)\) is a c.d.f. of the normal distribution \(\Phi(c)\). To this end, using \(x \in (c, c + T_d(c, y))\), we have \(0 \leq |c + T_d(c, y) - x| \leq |T_d(c, y)|\). Therefore, it follows that

\[
2^{-1} E \left[ \int_c^{c+T_d(c, y)} (c + T_d(c, y) - x) F''(x) dx \right] \leq 2^{-1} E \left[ \int_c^{c+T_d(c, y)} |(c + T_d(c, y) - x)||F''(x)| dx \right] \leq 2^{-1} E \left[ \int_c^{c+T_d(c, y)} |T_d(c, y)||F''(x)| dx \right] \leq 2^{-1} E \left[ |T_d(c, y)|^2 \sup_{x \in (c, c + T_d(c, y))} |F''(x)| \right] \leq E \left[ |T_d(c, y)|^2 \right] M.
\]

where \(M\) is some constant. In the following calculations, we will simplify the expression for \(T_d(c, y)\) to show that \(\sup_c V_2(c) = O(D^{-1})\). First of all, notice that

\[
\hat{\mu}_d - \bar{\mu}_d = \bar{X}_d^t(\hat{\beta} - \bar{\beta}) + \bar{Z}_d^t(\hat{u} - \bar{u}) = \bar{X}_d^t \left[ (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \bar{X}^t \bar{V}^{-1} \bar{y} - (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \bar{X}^t \bar{V}^{-1} \bar{y} \right] + \bar{Z}_d^t \left[ \bar{G}_d \bar{Z}_d^t \hat{\beta} - \bar{G}_d \bar{Z}_d^t \bar{V}^{-1} (y_d - \bar{X}_d \bar{\beta}) \right].
\]

In a view of the above, let us write \(T_d(c, y) = T_{1d}(y) + T_{2d}(y) + T_{3d}(c, y)\) where

\[
T_{1d}(y) = \bar{X}_d^t \left[ (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \bar{X}^t \bar{V}^{-1} \bar{y} - (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \bar{X}^t \bar{V}^{-1} \bar{y} \right] / \text{MSE}^{1/2}(\hat{\mu}_d),
\]

\[
T_{2d}(y) = \bar{Z}_d^t \left[ \bar{G}_d \bar{Z}_d^t \hat{\beta} - \bar{G}_d \bar{Z}_d^t \bar{V}^{-1} (y_d - \bar{X}_d \hat{\beta}) - \bar{G}_d \bar{Z}_d^t \bar{V}^{-1} (y_d - \bar{X}_d \hat{\beta}) \right] / \text{MSE}^{1/2}(\hat{\mu}_d),
\]

\[
T_{3d}(c, y) = c \left[ \text{MSE}^{1/2}(\hat{\mu}_d) - \text{MSE}^{1/2}(\bar{\mu}_d) \right] / \text{MSE}^{1/2}(\hat{\mu}_d).
\]

Following [Richardson and Welsh (1994)] we decompose \(T_{1d}(y)\)

\[
T_{1d}(y) = \bar{X}_d^t \left\{ \left[ (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} - (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \right] \bar{X}^t \bar{V}^{-1} (\bar{y} - \bar{X} \hat{\beta}) + (\bar{X}^t \bar{V}^{-1} \bar{X})^{-1} \bar{X}^t \left[ \bar{V}^{-1} - \bar{V}^{-1} \right] (\bar{y} - \bar{X} \hat{\beta}) \right\} / \text{MSE}^{1/2}(\hat{\mu}_d).
\]
Observe that \( T_{1d}(y) = o_P(D^{-1/2}) \) and \( \mathbb{E}[y] = X\beta \) which leads to \( \mathbb{E}[y - X\beta] = 0 \). Hence, we can immediately conclude that \( \mathbb{E}[T_{1d}(y)] = 0 \) and \( \mathbb{E}[T_{1d}^2(y)] = O(D^{-1}) \). When it comes to \( T_{2d}(y) \), we have a following expansion

\[
T_{2d}(y) = Z_d^t \left[ (\hat{G}_d Z_d^t \hat{V}_d^{-1} - G_d Z_d^t V_d^{-1})(y_d - X_d \hat{\beta}) + \hat{G}_d Z_d^t \hat{V}_d^{-1} X_d (\beta - \hat{\beta}) \right] / \text{MSE}^{1/2}(\hat{\mu}_d).
\]

Similarly to the above arguments, for the first part of the expression in a bracket we have \( \mathbb{E}[y] = X\beta \) and in the second part we can use the same derivation as for \( T_{1d}(y) \). Therefore, using the above arguments we have \( \mathbb{E}[T_{2d}(y)] = 0 \) and \( \mathbb{E}[T_{2d}^2(y)] = O(D^{-1}) \). When it comes to \( T_{3d}(c, y) \), we pointed out that \( \mathbb{E}[\hat{\mu}_d - \mu_d] \) can be approximated up to the order \( O(D^{-1/2}) \) or \( o(D^{-1/2}) \) and \( \mathbb{E}[\hat{\mu}_d - \hat{\mu}_d] \) up to \( o(D^{-1/2}) \). Thus, \( \mathbb{E}[T_{3d}^2(c, y)] = O(D^{-1}) \). The above derivations for \( T_{1d}, T_{2d} \) and \( T_{3d} \) leads to the following statement

\[
F_d(c) = F(c) + D^{-1} g(c, \beta, \theta) + O(D^{-1})
\]

where \( g \) is a smooth function of \( O(1) \). A similar representation can be derived for \( F_d^*(c) \) replacing \( \beta \) and \( \theta \) with \( \hat{\beta} \) and \( \hat{\theta} \). The result holds due to the consistency of \( \hat{\beta}^* \) and \( \hat{\theta}^* \).

References


simultaneous confidence bands in a cardiovascular safety experiment with longitudinal


likelihood (REML) estimates for hierarchical mixed linear models. *Australian & New


Royston, P. (1993). A pocket-calculator algorithm for the Shapiro-Francia test for non-

University Press.


Sun, J. and Loader, C. R. (1994). Simultaneous confidence bands for linear regression and
