



Simultaneous Inference for Empirical Best Predictors With a Poverty Study in Small Areas

Katarzyna Reluga, María-José Lombardía & Stefan Sperlich

To cite this article: Katarzyna Reluga, María-José Lombardía & Stefan Sperlich (2023) Simultaneous Inference for Empirical Best Predictors With a Poverty Study in Small Areas, Journal of the American Statistical Association, 118:541, 583-595, DOI: [10.1080/01621459.2021.1942014](https://doi.org/10.1080/01621459.2021.1942014)

To link to this article: <https://doi.org/10.1080/01621459.2021.1942014>



© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 09 Aug 2021.



[Submit your article to this journal](#)



Article views: 2352



[View related articles](#)





[View Crossmark data](#)



Citing articles: 4 [View citing articles](#)

Simultaneous Inference for Empirical Best Predictors With a Poverty Study in Small Areas

Katarzyna Reluga^a , María-José Lombardía^b , and Stefan Sperlich^c

^aDPMMS, University of Cambridge, Cambridge, UK; ^bCITIC, University of A Coruña, Coruña, Spain; ^cGSEM, University of Geneva, Geneva, Switzerland

ABSTRACT

Today, generalized linear mixed models (GLMM) are broadly used in many fields. However, the development of tools for performing simultaneous inference has been largely neglected in this domain. A framework for joint inference is indispensable to carry out statistically valid multiple comparisons of parameters of interest between all or several clusters. We therefore develop simultaneous confidence intervals and multiple testing procedures for empirical best predictors under GLMM. In addition, we implement our methodology to study widely employed examples of mixed models, that is, the unit-level binomial, the area-level Poisson-gamma and the area-level Poisson-lognormal mixed models. The asymptotic results are accompanied by extensive simulations. A case study on predicting poverty rates illustrates applicability and advantages of our simultaneous inference tools.

ARTICLE HISTORY

Received April 2020
Accepted June 2021

KEYWORDS

EBP; Generalized mixed models; Mixed parameters; Small area estimation; Uniform inference

1. Introduction


Generalized linear mixed models (GLMM) are suitable for modeling clustered and correlated data with categorical or count outcomes. They are ubiquitous in applied statistics, for example, in biometrics or small area estimation (SAE). In the latter, they serve to analyze surveys on a disaggregated level. Despite an increasing interest, for example, to guide resource allocation, the development of methods for simultaneous inference for predictors is missing. It is surprising as only those would make joint considerations of clusters valid. Available $(1 - \alpha)$ -confidence intervals (CI) for mixed parameters (except the credibility intervals of Ganesh 2009) are constructed such that for each study at least $\alpha 100\%$ of them do not contain the true value. Undoubtedly, practitioners do compare, but so far without valid statistical tools. We aim to close this distressing gap, not to improve any existing method.

Specifically, we introduce simultaneous confidence interval (SCI) and multiple test procedure (MTP) for the empirical best predictor (EBP) of Jiang (2003). They are based on max-type statistics combined with extreme value theory. We prove asymptotic convergence of SCI and MTP for nested (or hierarchical) GLMM within the exponential family. We study the numerical performance of our SCI and MTP for two area-level and one unit-level mixed models that are widely used, for example, for studying local poverty rates (Pratesi 2016). All introduced methods show a satisfactory performance within considered modeling frameworks. Even though our estimates under the area-level models appear to be less volatile, one can argue that EBPs are not directly comparable because different methods and model classes are used. Finally, under area-level Poisson-gamma

model, we derive a new mean squared error (MSE) estimator which is of crucial interest in SAE.

The amount of literature on estimation and testing under GLMM is considerable, see, i.a., the review of Tuerlinckx et al. (2006), the monograph of Jiang (2007), and the article of Ghosh et al. (1998) which is particularly interesting within the context of SAE. Furthermore, researchers put forward several methodologies broadly used in the analysis of count data. Molina, Saei, and Lombardía (2007) and Scealy (2010) studied the estimation of labor force status using multinomial logistic models, whereas Saei and Taylor (2012) focused on the same target parameter, and examined the performance of a bivariate random components model. Chandra, Chambers, and Salvati (2012) and Franco and Bell (2015) provided extensions for modeling proportions using logistic unit- and area-level models. Hobza and Morales (2016) implemented the EBP for unit-level, and Boubeta, Lombardía, and Morales (2016) for area-level GLMM to study poverty in small areas. Chambers, Salvati, and Tzavidis (2012) and Tzavidis et al. (2015) extended the M-quantile inference for robust estimation and prediction of count data. Yet, to the best of our knowledge, no one addresses the issue of simultaneous inference for clusters-level parameters when applying GLMM. Likewise, little research has been carried out on simultaneous inference for cluster level parameters in linear mixed effects models (LMM). Ganesh (2009) developed credibility intervals for a mixed parameter in a particular area-level model. Reluga, Lombardía, and Sperlich (2019) proposed bootstrap SCI and MTP for mixed parameter under LMM, whereas Kramlinger, Krivobokova, and Sperlich (2018) developed a framework for marginal and conditional inference with quadratic forms.

CONTACT Katarzyna Reluga  katarzyna.reluga@dpmms.cam.ac.uk  DPMMS, University of Cambridge, Wilberforce Road, Cambridge, CB3 0WB, UK

 Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

After an introduction of a model and estimators in Section 2, we propose the construction of SCI and MTP for a general EBP, followed by the theoretical justifications in Section 3. Sections 4 and 5 present simulations and a case study. Conclusions are drawn in Section 6. More details are deferred to appendix and our supplementary material (SM).

2. Best Prediction for GLMM

Let D be the number of clusters or areas with $d \in [D]$, n_d the number of sampled units in each area $j \in [n_d]$ with $n = \sum_{d=1}^D n_d$, N_d the known population sizes with $N = \sum_{d=1}^D N_d$, $[A] = \{1, \dots, A\}$. Since in our context the notion of cluster and area can be used synonymously, we proceed with the latter. Suppose that $\{v_d : d = 1, \dots, D\}$ is a set of independent and identically distributed (iid) random effects with unknown variance δ^2 , $\delta > 0$, which is often parameterized as $v_d = \delta u_d$ with $u_d \sim N(0, 1)$. The target variable y_{dj} represents the j th sample observation from the d th area. Furthermore, we consider nested data structures such that $y_{dj} \neq y_{d'j}$ for $d \neq d'$. In full generalization, we assume that random variables Y_{dj} , conditionally on a random effect u_d , are independent with a probability density function (pdf) from the exponential family $y_{dj}|u_d \sim \text{Exp.Family}(\theta)$

$$y_{dj}|u_d \sim \text{indep.} g_{Y_{dj}|u_d}(y_{dj}|u_d),$$

$$g_{dj}(y_{dj}|u_d, \theta) = \exp\{\varphi^{-1}[y_{dj}\gamma_{dj} - b(\gamma_{dj})] + c(y_{dj}, \varphi)\},$$

where $\theta = (\beta^t, \delta, \varphi)^t$ with δ the variability parameter, $\beta = (\beta_1, \dots, \beta_p)^t$ regression parameters of auxiliary variables $x_{dj} = (x_{dj1}, \dots, x_{dj p})^t$ for which typically $x_{dj1} = 1, \forall j \in [n_d] \forall d \in [D]$, and γ_{dj}, φ are canonical and scale parameters, respectively. Link function M relates $\mathbb{E}(Y_{dj}|u_d)$ to a linear mixed model such that $\gamma_{dj} = M\{\mathbb{E}(Y_{dj}|u_d)\} = x_{dj}^t \beta + \delta u_d$.

2.1. Estimation and Computation

Let $y_d = (y_{d1}, \dots, y_{dn_d})^t$ for all $d \in [D]$ be the vector of outcomes, and $y = (y_1^t, \dots, y_D^t)^t$. A conditional pdf of y and the likelihood contribution from each area d are given by

$$\mathcal{L}_d(\theta) := f_d(y_d|\theta) = \int g_d(y_d|u_d, \theta) h(u_d) du_d$$

$$= \int \prod_{j=1}^{n_d} g_{dj}(y_{dj}|u_d, \theta) h(u_d) du_d, \quad (1)$$

where θ can be derived from

$$\mathcal{L}(\theta) := \prod_{d=1}^D \mathcal{L}_d(\theta) = \prod_{d=1}^D \int \prod_{j=1}^{n_d} g_{dj}(y_{dj}|u_d, \theta) h(u_d) du_d.$$

In case of area-level models, $n_d = 1$, and Equation (1) simplifies accordingly. For a concise presentation, we assume that there is a single random effect for each area such that the integral in Equation (1) is one-dimensional. Extensions to multidimensional random effects follow immediately with some changes of notations and more complicated computation. Finding an analytical solution to Equation (1) is difficult unless the integral can be simplified. Often one evaluates the integral numerically

by Laplace approximation (LA) (De Bruijn 1981), Gaussian quadrature (GQ) (Naylor and Smith 1982) or adaptive GQ (AGQ) (Pinheiro and Bates 1995). In what follows, we proceed with AGQ as it is a higher order version of LA, that is, it gives smaller approximation errors (Bianconcini 2014). An alternative is the quasi-likelihood (Breslow and Clayton 1993) which suffers from a nondecreasing bias (Tuerlinckx et al. 2006), and the method of moments (Jiang 1998). In addition, researchers considerably advanced in developing methods to compute maximum likelihood (ML) estimators under GLMM. (Jiang 2007, sec. 4.1) proposed an expectation–maximization algorithm, whereas Lele, Nadeem, and Schmuland (2010) developed the so-called data cloning subsequently implemented by Torabi (2012).

Since we consider a prediction problem of possibly nonlinear mixed effects $\zeta_d = \zeta_d(\beta, u_d)$, we use the best predictor (BP) $\tilde{\zeta}_d$ in the sense of minimizing the area-specific MSE in Equation (3) which is actually the area-specific mean squared prediction error:

$$\tilde{\zeta}_d = \tilde{\zeta}_d(\theta) := \mathbb{E}\{\zeta_d(\beta, u_d)|y\} = \mathbb{E}\{\zeta_d(\beta, u_d)|y_d\}$$

$$= \frac{\int \zeta_d(\beta, u_d) g_d(y_d|u_d, \theta) h(u_d) du_d}{\int g_d(y_d|u_d, \theta) h(u_d) du_d}. \quad (2)$$

Simplification of Equation (2) is possible by choosing the pdf of u_d accordingly. If we replace θ by a consistent estimator, then we obtain EBP $\hat{\zeta}_d := \tilde{\zeta}_d(\hat{\theta})$. Note that in order to obtain the consistency for random effects, one needs to assume that $n_d \rightarrow \infty$ for each $\hat{\zeta}_d, d = 1, \dots, D$ (Jiang and Lahiri 2001).

Regarding the estimation of the variability of the EBP $\hat{\zeta}_d$, MSE is by far the most popular measure. Well known techniques to estimate MSE are the analytical approximation based on a Taylor expansion (Jiang 2003), and parametric bootstrap approaches (Boubeta, Lombardía, and Morales 2016; Hobza and Morales 2016). Consider the following MSE decomposition:

$$\text{MSE}(\hat{\zeta}_d) = \mathbb{E}\{[\tilde{\zeta}_d(\hat{\theta}) - \zeta_d]^2\}$$

$$= \mathbb{E}\{[\tilde{\zeta}_d(\hat{\theta}) - \tilde{\zeta}_d(\theta)]^2\} + \mathbb{E}\{[\tilde{\zeta}_d(\theta) - \zeta_d]^2\}$$

$$=: g_{2d} + g_{1d}, \quad (3)$$

which can be derived applying the law of iterated expectations (for details, see Jiang 2003, and our SM). The analytical formulas of MSE estimators are model dependent. Bootstrapping permits to obtain estimators that do not vary with the model assumed. In what follows, we denote with \mathbb{E}^* , Var^* , and MSE^* , the corresponding bootstrap operators for expected value, variance and MSE and define

$$\text{MSE}_B^*(\hat{\zeta}_d) = \mathbb{E}^*\{(\hat{\zeta}_d^* - \zeta_d^*)^2\}$$

$$\approx B_1^{-1} \sum_{b_1=1}^{B_1} (\hat{\zeta}_d^{*(b_1)} - \zeta_d^{*(b_1)})^2 =: \text{mse}_B(\hat{\zeta}_d), \quad (4)$$

which is a bootstrap equivalent of Equation (3). In their article, Hall and Maiti (2006) pointed out that (4) tends to underestimate the MSE, and propose a double-bootstrap bias-correction

$$\text{MSE}_{BC}^*(\hat{\zeta}_d) = 2\text{MSE}_B^*(\hat{\zeta}_d) - \text{MSE}_{B_2}^*(\hat{\zeta}_d)$$

$$\approx 2\text{mse}_B(\hat{\zeta}_d) - \text{mse}_{B_2}(\hat{\zeta}_d), \quad (5)$$

where $MSE_{B_2}^*(\hat{\zeta}_d)$ is the second-stage bootstrap MSE estimator, that is

$$\begin{aligned} MSE_{B_2}^*(\hat{\zeta}_d) &= \mathbb{E}^{**}\{(\hat{\zeta}_d^{**} - \zeta_d^{**})^2\} \\ &\approx B_1^{-1} B_2^{-1} \sum_{b_1=1}^{B_1} \sum_{b_2=1}^{B_2} \left(\hat{\zeta}_d^{**(b_1, b_2)} - \zeta_d^{**(b_1, b_2)}\right)^2 \\ &=: mse_{B_2}(\hat{\zeta}_d). \end{aligned}$$

The computation of $MSE_{B_2}^*(\hat{\zeta}_d)$ involves selecting B_2 bootstrap replicates from each first-stage bootstrap sample. In this article we do not aim for a precise estimation of the variability of EBP, but the construction of narrow SCI and reliable MTPs. It turns out that for doing this, the use of an estimate of g_{1d} as in Equation (3) yields better results than using an estimate of the entire MSE (see Section 4), similarly as in Chatterjee, Lahiri, and Li (2008) under LMM.

2.2. Popular Examples of GLMM and Their Properties

2.2.1. Poisson-Gamma Area-Level Model

The Poisson-gamma model is widely applied for modeling counts in the presence of overdispersion (see Cameron and Trivedi 2013, Section 4.2.2). Within the SAE context, Chen, Jiang, and Nguyen (2015) investigated the observed best prediction and bootstrap MSE estimation for small area mean counts. Among others, they also consider a Poisson-gamma specification. We propose a different model formulation, focus on the EBP of $\zeta_d := \mu_d^{PG}$ and develop a plug-in MSE estimator. Let $y_d|u_d \sim \text{Pois}(\mu_d^{PG})$, $d = 1, \dots, D$, where $\mu_d^{PG} > 0$, $n_d = 1 \forall d \in [D]$, with canonical parameter $\log \mu_d^{PG} = x_d^t \beta + u_d$, and $w_d := \exp(u_d) \sim \text{Gamma}(\delta, \delta)$ such that $\mathbb{E}(y_d|u_d) = \mu_d^{PG} = \lambda_d w_d = \exp(x_d^t \beta) w_d = \exp(x_d^t \beta + u_d)$. Since Gamma pdf is conjugate to the Poisson, their mixture yields a negative binomial $y_d \sim NB(\lambda_d, \delta^{-1})$ with likelihood

$$\begin{aligned} \mathcal{L}^{PG}(\theta) &:= f^{PG}(y|\theta) = \prod_{d=1}^D \frac{\Gamma(y_d + \delta)}{\Gamma(y_d + 1)\Gamma(\delta)} \\ &\times \left(\frac{\delta}{\delta + \lambda_d}\right)^\delta \left(\frac{\lambda_d}{\delta + \lambda_d}\right)^{y_d}, \end{aligned} \tag{6}$$

where $\mathbb{E}(y_d) = \lambda_d$ and $\text{Var}(y_d) = \lambda_d + \delta^{-1}\lambda_d^2$. The marginal mean of y_d is the same as in the Poisson case, but the random effect increases the variance. Suppose that this model holds for all areas of population \mathcal{P} of size N partitioned into subpopulations $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_D$ of sizes N_1, N_2, \dots, N_D . We can show that the BP for counts $\hat{\mu}_d^{PG}(\theta) := \mathbb{E}(\mu_d^{PG}|y_d)$ is

$$\begin{aligned} \mathbb{E}(\mu_d^{PG}|y_d) &= \frac{\int_0^\infty \lambda_d w_d g(y_d|w_d) h(w_d) dw_d}{\int_0^\infty g(y_d|w_d) h(w_d) dw_d} \\ &= \frac{A_d^{PG}(y_d, \theta)}{C_d^{PG}(y_d, \theta)} = \frac{\lambda_d(y_d + \delta)}{(\lambda_d + \delta)} \\ &=: \psi_d^{PG}(y_d, \theta). \end{aligned} \tag{7}$$

Equation (7) follows from the conjugation of the Gamma pdf to the Poisson pdf, while

$$\begin{aligned} A_d^{PG} &= \int_0^\infty \lambda_d w_d \frac{\exp(-\lambda_d w_d) \lambda_d^{y_d} w_d^{y_d} \delta^\delta w_d^{\delta-1} \exp(-w_d \delta)}{y_d! \Gamma(\delta)} dw_d \\ &= \frac{\lambda_d^{y_d+1} \delta^\delta \Gamma(y_d + 1 + \delta)}{\Gamma(\delta) y_d! (\lambda_d + \delta)^{y_d+1+\delta}}. \end{aligned}$$

The EBP $\hat{\mu}_d^{PG}$ is obtained by replacing the vector of unknown parameters θ in Equation (7) with a consistent estimator $\hat{\theta}$. Under the Poisson-gamma model $\varphi = 1$ and $\theta = (\beta, \delta)$. We derive an analytical plug-in MSE estimator to measure the variability of our EBP.

Proposition 1. Let $\text{Var}_d(\theta) = D \mathbb{E}\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^t\}$. An analytical MSE decomposition with its corresponding practical plug-in estimator are given by

$$\begin{aligned} MSE_{PG}(\hat{\mu}_d^{PG}) &= g_{PG1d} + \frac{1}{D} c_d(\theta) + o(1/D) \quad \text{and} \\ mse_{PG}(\hat{\mu}_d^{PG}) &= \hat{g}_{PG1d} + \frac{1}{D} \hat{c}_d(\hat{\theta}), \end{aligned} \tag{8}$$

$$\begin{aligned} g_{PG1d} &= \kappa_{1d}(\theta) - \kappa_{2d}(\theta), \\ \hat{g}_{PG1d} &= \kappa_{1d}(\hat{\theta}) - \hat{\kappa}_{2d}(\hat{\theta}), \quad d \in [D], \end{aligned} \tag{9}$$

$$\text{with } \kappa_{1d}(\theta) = \frac{\lambda_d^2(\delta + 1)}{\delta} \quad \text{and}$$

$$\kappa_{2d}(\theta) = \sum_{j=0}^\infty \frac{\lambda_d^2(j + \delta)^2}{(\lambda_d + \delta)^2} P(y_d = j),$$

$$\begin{aligned} \text{as well as } c_d(\theta) &= \sum_{j=0}^\infty \left\{ \frac{\partial}{\partial \theta} \psi_d^{PG}(y_d, \theta) \right\}^t \\ &\times \text{Var}_d(\theta) \left\{ \frac{\partial}{\partial \theta} \psi_d^{PG}(y_d, \theta) \right\} P(y_d = j). \end{aligned}$$

$\hat{c}_d(\theta)$ is a Monte Carlo approximation of $c_d(\theta)$, $\hat{\kappa}_{2d}$ refers to κ_{2d} with an infinite series truncated at a large term and θ replaced by $\hat{\theta}$. To estimate κ_{1d} we need only the latter.

One can estimate $\text{Var}_d(\theta)$ using any reasonable method. In Section 4 we use bootstrap estimators defined in (24). Details on the derivation of Equations (6) and (8) are deferred to our SM.

2.2.2. Poisson-Lognormal Area-Level Model

The Poisson-lognormal model has been thoroughly examined by, among others, Cameron and Trivedi (2013), Section 4.2.4, Franco and Bell (2015) and Boubeta, Lombardía, and Morales (2016). For $u_d \sim N(0, 1)$, let $y_d|u_d \sim \text{Pois}(\mu_d^{PL})$, $d = 1, \dots, D$, where $\mu_d^{PL} > 0$, $n_d = 1$ for all $d \in [D]$. In addition, $\mu_d^{PL} = v_d \rho_d$, where v_d is a known size variable and ρ_d a binomial probability. The canonical parameter is $\log \mu_d^{PL} = \log v_d + x_d^t \beta + \delta u_d$ for all $d \in [D]$. Typically $\zeta_d := \rho_d$ for which we have $\rho_d = \exp(x_d^t \beta + \delta u_d)$ with $\theta = (\beta^t, \delta)$. In this case, the likelihood is

$$\begin{aligned} \mathcal{L}^{PL}(\theta) &:= f^{PL}(y|\theta) = (2\pi)^{-D/2} \\ &\prod_{d=1}^D \int_{\mathbb{R}} \frac{\exp(-v_d \rho_d) v_d^{y_d} \exp\{y_d(x_d^t \beta + \delta u_d)\}}{y_d!} \exp\left(\frac{-u_d^2}{2}\right) du. \end{aligned}$$

Once θ is estimated, we obtain BPs $\tilde{\mu}_d^{\text{PL}}$, $\tilde{\rho}_d$, and EBPs $\hat{\mu}_d^{\text{PL}}$, $\hat{\rho}_d$ using the formulas from Boubeta, Lombardía, and Morales (2016). In Section 4.1 we estimate their MSE by bootstrap.

2.2.3. Logit Unit-Level Model

The unit-level logit model is a popular choice for binary responses, comprehensively discussed by Hobza and Morales (2016). Under this setting, $y_{dj}|u_d \sim \text{Bin}(m_{dj}, p_{dj})$, $u_d \sim N(0, 1)$ with m_{dj} a known size parameter for a logistic regression. The natural parameter is $p_{dj}/(1-p_{dj}) = x_{dj}^t \beta + \delta u_d$, $d \in [D]$, $j \in [n_d]$ where $p_{dj} = \{\exp(x_{dj}^t \beta + \delta u_d)\} / \{1 + \exp(x_{dj}^t \beta + \delta u_d)\}$. We assume that the unit-level logit model holds for all units of population \mathcal{P} of size N , partitioned into D subpopulations \mathcal{P}_d of sizes N_d , $d \in [D]$. Let $\zeta_d := \mu_d^U = \sum_{j=1}^{n_d} p_{dj}$. As for the Poisson models, we have $\varphi = 1$ and therefore $\theta = (\beta^t, \delta)$. The likelihood is given by

$$\begin{aligned} \mathcal{L}^U(\theta) &:= f^U(y|\beta, \delta) \\ &= (2\pi)^{-D/2} \prod_{d=1}^D \int_{\mathbb{R}} \exp \left[\sum_{j=1}^{n_d} \log \binom{m_{dj}}{y_{dj}} + \sum_{j=1}^{n_d} y_{dj} (x_{dj}^t \beta + \delta u_d) \right. \\ &\quad \left. - \frac{u_d^2}{2} - \sum_{j=1}^{n_d} m_{dj} \log \{1 + \exp(x_{dj}^t \beta + \delta u_d)\} \right] du_d. \end{aligned} \quad (10)$$

We can proceed with the estimation of the BP $\tilde{p}_{dj}(\theta)$ and $\tilde{\mu}_d^U = \sum_{j=1}^{n_d} \tilde{p}_{dj}$ only if we have access to the information on each population unit. In practice, however, the auxiliary information is available only for the sample units. Then, following the suggestion of Hobza and Morales (2016), we can still estimate the population quantity of interest by using only categorical covariates. Suppose that they take a finite number of values $x_{dj} \in \{z_1, \dots, z_L\}$ for $d \in [D]$ and $j \in [n_d]$ with z_l denoting the resulting covariate class. We then define

$$\begin{aligned} \bar{\mu}_d^U &= \frac{\mu_d^U}{N_d}, \quad \mu_d^U = \sum_{j=1}^{n_d} p_{dj} = \sum_{l=1}^L N_{dl} r_{dl}, \quad \text{with} \\ r_{dl} &= \frac{\exp(z_l \beta + \delta u_d)}{1 + \exp(z_l \beta + \delta u_d)}, \end{aligned} \quad (11)$$

where $N_{dl} = \#\{l \in \mathcal{P}_d : x_{dj} = z_l\}$ is the known size of class z_l in area d . Hobza and Morales (2016) derived BP $\tilde{\mu}_d^U(\theta)$ and EBP $\hat{\mu}_d^U(\hat{\theta})$ for μ_d^U as well as for other quantities in Equation (11). Due to the computational burden of the analytical estimator, in Section 4.2 we use bootstrap for obtaining an estimate of MSE.

3. Simultaneous Intervals and Multiple Testing

To construct CI for ζ_d that account for the effect of estimates from other areas, we need to find a region $\mathcal{I}_{1-\alpha}$ such that $P(\zeta_d \in \mathcal{I}_{1-\alpha} \forall d \in [D]) = 1 - \alpha$. Define

$$S_0 = \max_{d=1, \dots, D} |S_{0d}|, \quad \text{with } S_{0d} = \frac{\hat{\zeta}_d - \zeta_d}{\hat{\sigma}(\hat{\zeta}_d)}, \quad \forall d \in [D], \quad (12)$$

$$q_{S_0}^{(1-\alpha)} = \inf\{t \in \mathbb{R} : P(S_0 \leq t) \geq 1 - \alpha\}, \quad (13)$$

with $\hat{\sigma}(\hat{\zeta}_d)$ being an estimate of the variability of EBP $\hat{\zeta}_d$. We then consider

$$\begin{aligned} \alpha &= P\left(\left|\hat{\zeta}_d - \zeta_d\right| > q_{S_0}^{(1-\alpha)} \hat{\sigma}(\hat{\zeta}_d) \text{ for some } d \in [D]\right) \\ &= P\left(\max_{d=1, \dots, D} \left| \frac{\hat{\zeta}_d - \zeta_d}{\hat{\sigma}(\hat{\zeta}_d)} \right| > q_{S_0}^{(1-\alpha)}\right). \end{aligned} \quad (14)$$

Constructing SCI boils down to the estimation of $q_{S_0}^{(1-\alpha)}$, as one can define then

$$\mathcal{I}_{1-\alpha}^S = \prod_{d=1}^D \mathcal{I}_{d,1-\alpha}^S, \quad \text{with } \mathcal{I}_{d,1-\alpha}^S = \left\{ \hat{\zeta}_d \pm q_{S_0}^{(1-\alpha)} \times \hat{\sigma}(\hat{\zeta}_d) \right\}, \quad (15)$$

where \times denotes a generalized Cartesian product. $\mathcal{I}_{1-\alpha}^S$ covers all ζ_d with probability $1 - \alpha$, that is, its joint confidence level is $1 - \alpha$. In contrast, for each $q_{S_{0d}}^{(1-\alpha)}$ defined analogously to $q_{S_0}^{(1-\alpha)}$, with S_0 replaced by $|S_{0d}|$, individual area CI (iCI) are given by

$$\mathcal{I}_{d,1-\alpha}^{\text{iCI}} = \left\{ \hat{\zeta}_d \pm q_{S_{0d}}^{(1-\alpha)} \times \hat{\sigma}(\hat{\zeta}_d) \right\} \quad \forall d \in [D]. \quad (16)$$

By construction, iCI does not contain ζ_d for at least $100\alpha\%$ of all areas.

Remark 1. $\mathcal{I}_{d,1-\alpha}^{\text{iCI}}$ is designed to cover ζ_d at an individual confidence level. Consequently, the joint coverage probability of iCIs decreases in a cumulative way for increasing D . This highlights the need to construct SCI. Nevertheless, maintaining $1 - \alpha$ simultaneous confidence level of SCI $\mathcal{I}_{1-\alpha}^S$ makes its constituents $\mathcal{I}_{d,1-\alpha}^S$ wider than corresponding iCIs $\mathcal{I}_{d,1-\alpha}^{\text{iCI}}$. This is not surprising because $\mathcal{I}_{d,1-\alpha}^{\text{iCI}}$ and $\mathcal{I}_{d,1-\alpha}^S$ were constructed to cover different sets which serve distinct inferential purposes. It is worth mentioning that the length of $\mathcal{I}_{d,1-\alpha}^S$ stabilizes as for growing D we observe two opposite trends: the increase of area parameters to cover and the decrease of MSE (see Tables 1 and 3).

The SCI defined in Equation (15) is not operational as the distribution of S_0 is unknown. The problem can be circumvented by bootstrap approximation: for $b_1 = 1, \dots, B_1$ set

$$S_B^{(b_1)} = \max_{d=1, \dots, D} |S_{Bd}^{(b_1)}|, \quad S_{Bd}^{(b_1)} = \frac{\hat{\zeta}_d^{*(b_1)} - \zeta_d^{*(b_1)}}{\hat{\sigma}^{*(b_1)}(\hat{\zeta}_d^*)}, \quad (17)$$

and approximate the critical value $q_{S_B}^{(1-\alpha)} = \inf\{t \in \mathbb{R} : P(S_B \leq t | (y, X)) \geq 1 - \alpha\}$, by a $[(1 - \alpha)B_1 + 1]$ th-order statistic of the $S_B^{(b_1)}$. Then the bootstrap equivalent of Equation (15) is

$$\begin{aligned} \mathcal{I}_{1-\alpha}^B &= \prod_{d=1}^D \mathcal{I}_{d,1-\alpha}^B, \quad \text{where} \\ \mathcal{I}_{d,1-\alpha}^B &= \left\{ \hat{\zeta}_d \pm q_{S_B}^{(1-\alpha)} \times \hat{\sigma}(\hat{\zeta}_d) \right\}. \end{aligned} \quad (18)$$

An alternative approach to Equation (12) could be to take computationally simpler nonstudentized statistics. Yet, already DiCiccio and Efron (1996) pointed out that the lack of studentization results in slower convergence rates. Since the application of nonstudentized SCI did not yield satisfactory results, we decided not to include them.

Table 1. ECP, AIW, and AIWV of SCI under area-level models.

	Poisson-gamma				Poisson-lognormal	
	B	BC	P	G	B	BC
	ECP (in %)				ECP (in %)	
$D = 26$	95.7	95.9	95.8	95.3	95.0	94.6
$D = 52$	94.7	93.7	94.9	94.6	94.9	94.8
$D = 78$	94.7	94.8	94.9	94.4	94.9	94.6
	AIW $\times 10^3$ (AIWV $\times 10^3$)				AIW $\times 10^3$ (AIWV $\times 10^3$)	
$D = 26$	24.4 (0.025)	24.5 (0.028)	24.5 (0.024)	23.9 (0.023)	18.0 (0.003)	18.0 (0.004)
$D = 52$	30.3 (0.030)	30.3 (0.034)	30.4 (0.029)	29.8 (0.028)	20.7 (0.003)	20.8 (0.004)
$D = 78$	33.4 (0.027)	33.4 (0.032)	33.5 (0.025)	33.0 (0.024)	21.9 (0.002)	22.0 (0.004)

NOTE: Nominal coverage: 95%.

Our methodology is also applicable for hypothesis testing. Consider the test problem

$$H_0 : B\zeta = \mathbf{b} \quad \text{vs.} \quad H_1 : B\zeta \neq \mathbf{b}, \quad (19)$$

where $B \in \mathbb{R}^{D' \times D}$, $D' \leq D$, $\mathbf{b} \in \mathbb{R}^{D'}$. We are interested in max-type statistic t_H such that

$$t_H = \max_{d=1, \dots, D'} |t_{Hd}|, \quad t_{Hd} = \frac{\hat{\zeta}_d^H - b_d}{\hat{\sigma}(\hat{\zeta}_d^H)},$$

$$S_{H_0} = \max_{d=1, \dots, D'} |S_{H_0d}|, \quad S_{H_0d} = \frac{\hat{\zeta}_d^H - \zeta_d^H}{\hat{\sigma}(\hat{\zeta}_d^H)}, \quad (20)$$

where $\zeta^H = (\zeta_1^H, \dots, \zeta_{D'}^H)^t = B\zeta \in \mathbb{R}^{D'}$ with $\hat{\zeta}^H$ being its estimator. One rejects H_0 at the α -level if $t_H \geq q_{H_0}^{(1-\alpha)}$ with $q_{H_0}^{(1-\alpha)} = \inf\{t \in \mathbb{R} : P(S_{H_0} \leq t) \geq 1 - \alpha\}$.

In practice, we might use such a test to examine differences between area characteristics. Similarly as for SCI, we approximate $q_{H_0}^{(1-\alpha)}$ applying bootstrap to a modified version of statistic S_B , namely

$$q_{BH_0}^{(1-\alpha)} = \inf\{t \in \mathbb{R} : P(S_{BH_0} \leq t | (y, X)) \geq 1 - \alpha\}, \quad (21)$$

where S_{BH_0} in the b_1^{th} bootstrap sample is

$$S_{BH_0}^{(b_1)} = \max_{d=1, \dots, D'} |S_{BH_0d}^{(b_1)}|, \quad S_{BH_0d}^{(b_1)} = \frac{\hat{\zeta}_d^{*H(b_1)} - \zeta_d^{*H(b_1)}}{\hat{\sigma}^*(\hat{\zeta}_d^{*H(b_1)})}, \quad (22)$$

with $\zeta^{*H(b_1)} = (\zeta_1^{*H(b_1)}, \dots, \zeta_{D'}^{*H(b_1)}) = B\zeta^{*(b_1)} \in \mathbb{R}^{D'}$ and $\hat{\zeta}^{*H(b_1)} = (\hat{\zeta}_1^{*H(b_1)}, \dots, \hat{\zeta}_{D'}^{*H(b_1)})$ its corresponding estimated version.

We provide the consistency of our bootstrap-based CI and tests, as well as asymptotic convergence and coverage probability. Proofs are deferred to Appendix A.2 and A.3. Suppose that $\hat{\theta}$ is consistent such that $\|\hat{\theta} - \theta\| = O_p(n^{-c})$, $c > 0$. Since for the GLMM with clustered random effects the log-likelihood can be expressed as the sum of independent random components, the consistency of $\hat{\theta}$ estimated by ML follows assuming a classical theory. The consistency under a general GLMM had been an open problem for many years until it was solved by Jiang (2013). Bianconcini (2014) and Huber, Ronchetti, and Victoria-Feser (2004) investigated the consistency of $\hat{\theta}$ once we compute it using AGQ and LA respectively. For our purpose, we need to prove the bootstrap consistency

Proposition 2. Under Assumptions 1–5 from Appendix A.1 it holds that

$$\mathbb{E}^*(y_{dj}^*) - \mathbb{E}(y_{dj}) = o_p^*(1),$$

$$\text{Var}^*(y_d^*) - \text{Var}(y_d) = [o_p^*(1)]_{n_d \times n_d},$$

$$\|\hat{\theta}^* - \hat{\theta}\| = O_{p^*}(n^{-c}).$$

Given Proposition 2, we can derive the consistency of $\mathcal{I}_{1-\alpha}^B$ based on results from extreme value theory and asymptotic expansions of the standardized statistics using ideas from Chatterjee, Lahiri, and Li (2008). Let us assume $\hat{\sigma}(\hat{\zeta}_d) = \sqrt{\hat{g}_{1d}(\hat{\zeta}_d)}$, though similar results are immediate for $\hat{\sigma}(\hat{\zeta}_d) = \sqrt{\text{mse}_{(\cdot)}(\hat{\mu}_d^{(\cdot)})}$ where (\cdot) stands for different types of estimators. We use $q := q_{S_0}^{(1-\alpha)}$ where unambiguous, and denote the cumulative distribution function (cdf) of S_{0d} and S_{Bd} by $G_d(w) = P(S_{0d} \leq w)$ and $G_{Bd}(w) = P(S_{Bd} \leq w)$. In Appendix A.3 we provide asymptotic expansions for both. Define $(S_{0(d+1)}, \dots, S_{0(2D)}) = (-S_{01}, \dots, -S_{0D})$, and observe that $\max_{d=1, \dots, D} |S_{0d}| = \max_{d=1, \dots, 2D} (S_{01}, \dots, S_{0D}, -S_{01}, \dots, -S_{0D})$. From Equation (14), we have

$$\begin{aligned} \mathcal{T}_D(q) &= P(S_0 \leq q) \\ &= P(S_{01} \leq q, \dots, S_{0D} \leq q, -S_{01} \leq q, \dots, -S_{0D} \leq q) \\ &= \prod_{d=1}^{2D} G_d(q). \end{aligned} \quad (23)$$

As $D \rightarrow \infty$, unless standardized, the distribution in Equation (23) converges to 0 or 1. In Appendix A.3, we show that $G_d(w)$ is asymptotically normal, such that $P(S_0 \leq q) \approx \Phi^{2D}(q)$. Since the cdf of the maxima of the standard normal random variables is in the domain of attraction of the Gumbel law, it follows that $\lim_{D \rightarrow \infty} \Phi^{2D}(q/b_D + b_D) = \exp(\exp(-q)) = \mathcal{T}_0(q)$, for all $q \in \mathbb{R}$ where b_D is a sequence of constants (see Leadbetter, Lindgren, and Rootzén 2012, theor.1.5.3). Unfortunately, this approximation has a poor convergence rate, but bootstrap is again a remedy. Notice that a similar representation holds for S_B , substituting P with P^* and replacing the true parameters by their estimates. Application of Poyla’s theorem that combines the convergence in distribution with a convergence in sup norm results in our next proposition.

Proposition 3. Define $\mathcal{T}_D^*(w) = P^*(S_B \leq q)$ which is a bootstrap analogue of $\mathcal{T}_D(w)$ in Equation (23). Under Assumptions

1–5 from [Appendix A.1](#) it holds that

$$\sup_{w \in \mathbb{R}} |\mathcal{T}_D(w) - \mathcal{T}_D^*(w)| = o_P(1).$$

Corollary 1. [Proposition 3](#) implies that under the same assumptions,

$$P(\zeta_d \in \mathcal{I}_{1-\alpha}^B \forall d \in [D]) \rightarrow 1 - \alpha.$$

Since we use almost identical max-type statistics in [Equations \(12\) and \(20\)](#), the construction of MTP follows almost immediately from the correspondence between tests and CI. In fact, the acceptance region of our test is $\mathcal{I}_{1-\alpha}^{H_0} = \times_{d=1}^D \mathcal{I}_{d,1-\alpha}^{H_0}$, where

$$\mathcal{I}_{d,1-\alpha}^{H_0} = \left\{ \hat{\zeta}_d - q_{S_0}^{(1-\alpha)} \times \hat{\sigma}(\hat{\zeta}_d) \leq b_d \leq \hat{\zeta}_d + q_{S_0}^{(1-\alpha)} \times \hat{\sigma}(\hat{\zeta}_d) \right\},$$

that is, we reject H_0 if $b \notin \mathcal{I}_{1-\alpha}^{H_0}$. We can write $P(h_d \in \mathcal{I}_{1-\alpha}^{H_0} \forall d \in [D]) = 1 - \alpha$. Since this probability statement is true for any $h_d = \zeta_d$, we obtain the CI defined in [Equation \(15\)](#) by inverting the test.

Corollary 2. Let H_0 be the null hypothesis defined in [Equation \(19\)](#) and $\alpha \in (0, 1)$. Under [Proposition 3](#), we have

$$P(t_H > q_{BH_0}) \leq \alpha + o(1).$$

Remark 2. Our single-step testing procedure in [Equation \(19\)](#) with a bootstrap critical value in [Equation \(21\)](#) controls weakly for the family-wise error rate (FWER), and might be limited in detecting false null hypotheses once we deal with a large D' . Yet, we can readily extend our test to a bootstrap-based step-down procedure of [Romano, Shaikh, and Wolf \(2008\)](#) which controls the false discovery rate with a better power to detect false H_{0d} than FWER.

4. Empirical Reliability Study

We performed intensive simulation studies to assess the reliability of our methods. SCI and MTP for EBP were constructed with different estimators of variability under the models presented in [Sections 2.2.1–2.2.3](#). First, we examined the relative bias and relative root-MSE of fixed effects $\hat{\beta}$ and variability parameter $\hat{\delta}$. Then, the performance of EBP was evaluated comparing bias, average absolute bias and MSE for $D = 26, 52$, and 78 . Since they did not show any atypical pattern, the results under Poisson area-level models and logistic unit-level model were deferred to the SM. Regarding SCIs, we calculated empirical coverage probability (ECP), average interval width (AIW), and the AIW variation (AIWV):

$$\text{ECP} = \frac{1}{K} \sum_{k=1}^K \mathbb{1}\{\zeta_d^{(k)} \in \mathcal{I}_{1-\alpha}^S \forall d \in [D]\},$$

$$\text{AIW} = \frac{1}{DK} \sum_{d=1}^D \sum_{k=1}^K \omega_d^{(k)}, \quad \omega_d^{(k)} = 2q_{(\cdot)}^{(1-\alpha)(k)} \hat{\sigma}(\hat{\zeta}_d^{(k)}),$$

$$\text{AIWV} = \frac{1}{D(K-1)} \sum_{d=1}^D \sum_{k=1}^K \left(\omega_d^{(k)} - \bar{\omega}_d \right)^2,$$

$$\bar{\omega}_d = \frac{1}{K} \sum_{k=1}^K \omega_d^{(k)}, \quad d = 1 \dots, D.$$

For each simulation run k we record the widths of the SCI and check whether they cover all EBP. ECP is then computed by averaging over K simulation runs and is aimed to be close to $1 - \alpha$. AIW is obtained by averaging over the simulation runs and areas. Narrower intervals are preferable if its ECP is close to the nominal level. These are standard measures to assess the quality of interval estimators ([Chatterjee, Lahiri, and Li 2008; Ganesh 2009](#)). Lower AIWV values indicate that the length is stable and does not depend on the simulation run.

4.1. Finite Sample Performance Under Area-Level Models

Under the Poisson-gamma model we set $y_d \sim \text{Pois}(\mu_d^{\text{PL}})$, $\mu_d^{\text{PL}} = \lambda_d w_d$. Covariates, parameters and sample sizes are taken from our case study in [Section 5](#), that is, we set $\theta = (\beta^t, \delta) = (10.038, 7.747, -3.136, 11.317, -2.466, 2.480)^t$, and $D = \{26, 52, 78\}$, $n_d = 1, \forall d \in [D]$, $n = D$. For $D = 52$ we take covariates from the original sample, for $D = 26$ we randomly select the areas using simple random sampling without replacement, and for $D = 78$, we take the original sample plus 26 randomly selected areas, that is, these areas enter at most twice. Parameter of interest is the area proportion of individuals below the poverty line, $\bar{\mu}_d^{\text{PL}} = \mu_d^{\text{PL}}/N_d$. The EBP for μ_d^{PL} is given in [Equation \(7\)](#). Since N_d is usually unknown, in practice it is replaced by its estimate, see [Equation \(25\)](#) in [Section 5](#). We apply double bootstrap with $B_1 = 1000$ first-stage and $B_2 = 1$ second-stage bootstrap replicates (the choice of the latter is motivated by [Erculescu and Fuller 2014](#)). We generate $K = 1000$ samples with the same areas and fixed covariates, but randomly drawn w_d and y_d . SCIs and iCIs are constructed as follows:

1. Fit the model to the data and obtain estimates $\hat{\theta} = (\hat{\beta}, \hat{\delta})$.
2. For $b_1 = 1, \dots, B_1$ bootstrap samples, generate $w_d^{*(b_1)} \sim \text{Gamma}(\hat{\delta}, \hat{\delta})$ iid and set

$$\mu_d^{\text{PG}*(b_1)} = \hat{\lambda}_d w_d^{*(b_1)} \quad \text{and} \quad y_d^{*(b_1)} \sim \text{Poisson}(\mu_d^{\text{PG}*(b_1)}).$$
3. For each bootstrap sample calculate $\hat{\theta}^{*(b_1)}, \hat{\mu}_d^{\text{PG}*(b_1)}(\hat{\theta}^{*(b_1)})$, $AD_{\text{PG},d}^{(b_1)} = \left| \hat{\mu}_d^{\text{PG}*(b_1)} - \mu_d^{\text{PG}*(b_1)} \right|$.
 - (a) For $b_2 = 1, \dots, B_2$ generate samples $w_d^{**(b_1, b_2)} \sim \text{Gamma}(\hat{\delta}^{*(b_1)}, \hat{\delta}^{*(b_1)})$ iid and

$$\mu_d^{\text{PG}**(b_1, b_2)} = \hat{\lambda}_d^{*(b_1)} w_d^{**(b_1, b_2)},$$

$$y_d^{**(b_1, b_2)} \sim \text{Poisson}(\mu_d^{\text{PG}**(b_1, b_2)}).$$
 - (b) For each bootstrap sample calculate $\hat{\theta}^{**(b_1, b_2)}$ and $\hat{\mu}_d^{\text{PG}**(b_1, b_2)}(\hat{\theta}^{**(b_1, b_2)})$.
 - (c) Set $\text{mse}_d^{(b_1)} = \frac{1}{B_2} \sum_{b_2=1}^{B_2} \left(\hat{\mu}_d^{\text{PG}**(b_1, b_2)} - \mu_d^{\text{PG}**(b_1, b_2)} \right)^2$.
4. Calculate bootstrap estimates $\hat{g}_{\text{PG}1d}(\hat{\theta}^{*(b_1)})$ as in [Equation \(9\)](#) as well as

$$\text{mse}_B(\hat{\mu}_d^{\text{PG}}) = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \left(\hat{\mu}_d^{\text{PG}*(b_1)} - \mu_d^{\text{PG}*(b_1)} \right)^2,$$

$$\text{mse}_{\text{BC}}(\hat{\mu}_d^{\text{PG}}) = 2\text{mse}_B(\hat{\mu}_d^{\text{PG}}) - \frac{1}{B_1} \sum_{b_1=1}^{B_1} \text{mse}_d^{(b_1)}.$$

5. Calculate statistic $S_{PG,B}$ with the critical value $q_{PG,S_B}^{(1-\alpha)}$ obtained from the bootstrap sample $S_{PG,B} = (S_{PG,B}^{(1)} \dots S_{PG,B}^{(B_1)})^t$, where

$$S_{PG,B}^{(b_1)} = \max_{d=1,\dots,D} AD_{PG,d}^{*(b_1)} / \hat{\sigma}^{*(b_1)} \left(\hat{\mu}_d^{PG*(b_1)} \right) \quad \text{and}$$

$$q_{PG,S_B}^{(1-\alpha)} = Q_{1-\alpha}(S_{PG,B})$$

as well as a variance estimate for $\hat{\theta}$:

$$\widehat{\text{var}}(\hat{\theta}) = \frac{1}{B_1} \sum_{b_1=1}^{B_1} (\hat{\theta}^{*(b_1)} - \bar{\theta})(\hat{\theta}^{*(b_1)} - \bar{\theta})^t \quad \text{with}$$

$$\bar{\theta} = \frac{1}{B_1} \sum_{b_1=1}^{B_1} \hat{\theta}^{*(b_1)}. \tag{24}$$

We compare the performance of SCI and MTP for different variability estimates $\hat{\sigma}(\hat{\mu}_d^{PG})$ and their bootstrap equivalents $\hat{\sigma}^*(\hat{\mu}_d^{PG*})$, namely for $\hat{\sigma}(\hat{\mu}_d^{PG}) = \sqrt{\hat{g}_{PG1d}}$ and $\hat{\sigma}(\hat{\mu}_d^{PG}) = \sqrt{\text{mse}_{(\cdot)}(\hat{\mu}_d^{PG})}$. Here, $\text{mse}_{(\cdot)}$ refers to either the plug-in mse_P , the mse_B or the mse_{BC} , defined in Equations (8), (4), and (5). Steps 3(a)–(c) of the algorithm refer to the second-stage bootstrap which is only necessary to obtain bias-corrected mse_{BC} . Under the Poisson-gamma model, we are interested in the estimation of poverty rates. We thus consider $\hat{\mu}_d^{PG} = \hat{\mu}_d^{PG}/N_d$, $\hat{g}_{PG1d} = \hat{g}_{PG1d}/N_d^2$ and $\text{mse}_{(\cdot)}(\hat{\mu}_d^{PG}) = \text{mse}_{(\cdot)}(\hat{\mu}_d^{PG})/N_d^2$.

For the Poisson-lognormal model with $y_d \sim \text{Poisson}(\mu_d^{PL})$, $\mu_d^{PL} = \nu_d \rho_d$, the parameter of interest is ρ_d with $N_d = \nu_d$ estimated by EBP derived by Boubeta, Lombardía, and Morales (2016). We take the fixed parameters from Section 5, that is, $(\beta^t, \delta) = (-2.264, 3.480, -0.870, 4.842, 0.125, 0.322)^t$. Covariates, sample sizes, number of simulation runs and bootstrap replicates are the same as in case of the Poisson-gamma model. The variability of $\hat{\rho}_d$ was estimated using bootstrap MSEs, that is mse_B and mse_{BC} . To obtain estimates of SCI and iCI one can use almost the same algorithm as above by changing the way we generate $y_d^{*(b_1)}$.

Table 1 summarizes the performance of 95% SCI for $\hat{\mu}_d^{PG}$ constructed with mse_B (B), mse_{BC} (BC), plug-in mse_P (P) and \hat{g}_{PG1d} (G). For $\hat{\rho}_d$, they were constructed using mse_B (B) and mse_{BC} (BC). All methods perform very well regarding the coverage ECP, even for $D = 26$. In contrast, SCIs constructed using a Bonferroni procedure yield unacceptably low ECP. For instance, for $D = 52$ and mse_B it equals 78% for the Poisson-gamma, and 88% for the Poisson-lognormal model. Therefore, we do not further report them.

Figure 1 presents 95% SCI and iCI estimates for a randomly selected simulation under the Poisson-gamma model. The plot is divided into five panels according to the number of units \hat{N}_d^{dir} defined in Equation (25) with the first presenting the results for the areas with the fewest observations. The black and red dots represent the true proportions known in a simulation. The color red indicates true parameters not covered by their iCIs. In Figure 1, that holds for four of the true values ($\approx 7.7\%$). This illustrates well the difference between individual and simultaneous inference as well as a particular relevance of the latter. We obtain similar figures for other simulations (see our SM).

Finally, we studied the performance of our test (19) under Poisson-gamma and Poisson-lognormal models. Results for the latter are in our SM as they reveal the same features. Consider $H_0 : \bar{\mu}^{PG} = \mathbf{b}$ vs. $H_1 : \bar{\mu}^{PG} = \mathbf{b} + \mathbf{1}_D \Delta$, where $\mathbf{b} := \bar{\mu}$ for the same data-generating processes as before. Critical values are obtained from the bootstrap analogues of S_{H_0} calculated similarly as in Step 5 of the algorithm above. Figure 2 shows the power functions of our test based on different variability estimates. They are visibly indistinguishable, which is not surprising given the similar ECPs and AIWs in Table 1. For $D = 52$, that is, the sample size of the real data, the nominal level of 5% is attained almost exactly under H_0 .

4.2. Finite Sample Performance Under the Unit-Level Model

Under the unit-level model we assume $y_{dj} \sim \text{Bin}(m_{dj}, p_{dj})$ with $p_{dj} = \{\exp(x_{dj}^t \beta + \delta u_d)\} / \{1 + \exp(x_{dj}^t \beta + \delta u_d)\}$, $m_{dj} = 1$, $u_d \sim N(0, 1)$. In our context, y_{dj} is binary and value 1 indicates an individual below the poverty threshold defined in Section 5. The regression parameters are taken from our case study: $\theta = (\beta^t, \delta) = (-2.048, 0.989, 0.172, 0.760, 0.100, 0.348)^t$. Four categorical covariates result in 16 covariate classes $x_{dj} \in \{z_1, \dots, z_{16}\}$ for which we need to estimate N_{dl} using Equation (25), $l = 1, \dots, 16$. We considered $D = \{26, 52, 78\}$ containing unit-level information with $n = \{11423, 23628, 35818\}$, respectively. Summary statistics for all samples are presented in Table 2. Furthermore, for $D = 52$, n_d, N_d, x_{dj}, z_l are the same as in our case study. The areas for $D = \{26, 78\}$ were selected in the same way as in Section 4.1. In addition, for $D = 78$, within each of the additional area we sampled with replacement n_d units (i.e., 26 newly sampled areas contained different units in comparison to the original sample). The parameter of interest is the area poverty proportion $\bar{\mu}_d^U$ defined in Equation (11). Given that the original sample size was $n = 23,628$, under the unit-level model we restrict our simulations to $K = 200$, $B_1 = 500$, and $B_2 = 1$. As far as the algorithm for constructing SCI and iCI is concerned, it follows almost the same steps as in Section 4.1. The exact algorithm can be found in the SM.

Table 3 presents the performance of SCI constructed using mse_B (B) and mse_{BC} (BC). The coverage probability is somewhat lower than the nominal level. In addition, it slightly decreases with increasing D , whereas the AIW increase stabilizes as expected (see Remark 1). The undercoverage might be related to the simulation design. Even though the latter is popular in SAE, it is suboptimal for random effects from the asymptotic point of view ($n_d \rightarrow \infty, d \in [D]$, recall Section 2.1). The results in Table 3 do not demonstrate any inconsistencies with respect to the theoretical developments, nor they exhibit unexpected findings. Due to their limited impact, the equivalents of Figures 1 and 2 for this simulation are deferred to our SM. In comparison to the area-level models, the coverage probability is worse and the average width of SCIs is much larger (it is also the case for the iCI, see our SM). Moreover, fitting unit-level models is computationally more expensive. In our case the estimation of MSE and construction of intervals took about 900–1000 times longer. Since the data-generation processes are different, the numerical results in our simulations

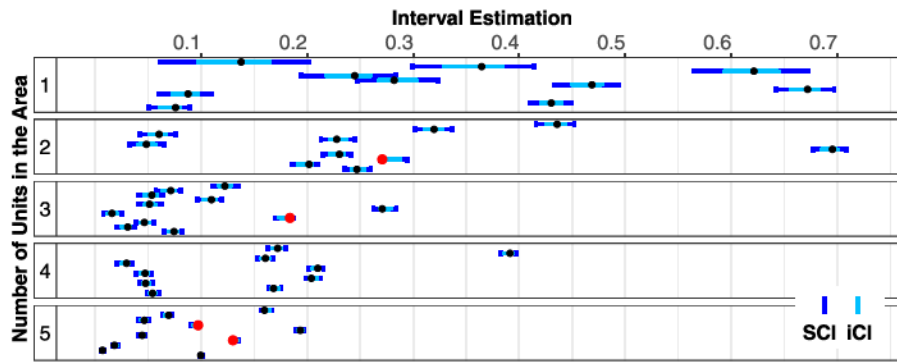


Figure 1. 95% iCI and SCI for proportions with $D = 52$. Red dots indicate true parameters outside iCI, whereas black dots indicate true parameters inside their iCI.

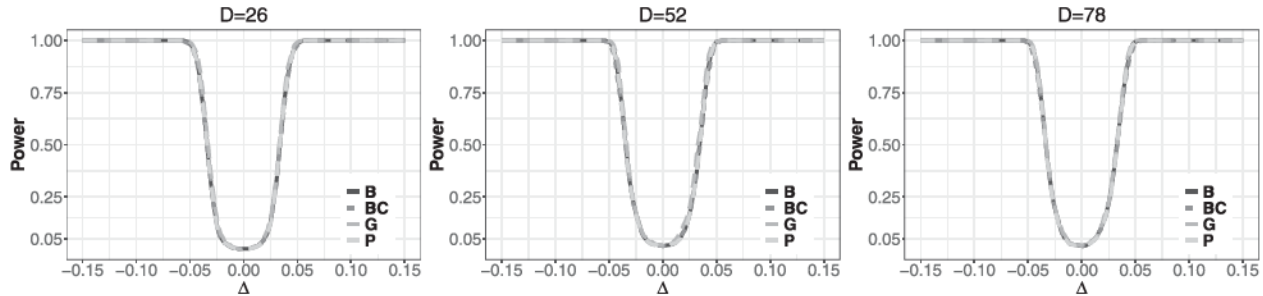


Figure 2. Simulated powers for multiple test $H_0 : \mu^{PG} = b$ versus $H_1 : \mu^{PG} = b + 1_D \Delta$ under the area-level Poisson-gamma model; (left) $D = 26$, (middle) $D = 52$, (right) $D = 78$.

Table 2. Summary statistics of n_d under different scenarios in the simulation study under the unit-level model.

$D = 26$						$D = 52$						$D = 78$					
Min.	Q_1	Q_2	Mean	Q_3	Max.	Min.	Q_1	Q_2	Mean	Q_3	Max.	Min.	Q_1	Q_2	Mean	Q_3	Max.
47	139	260	439	454	1714	34	108	260	454	479	2631	34	114	289	459	482	2631

Table 3. ECP, AIW and AIWV of 95% SCI under the unit-level model.

D	ECP (in %)		AIW $\times 10^3$ (AIWV $\times 10^3$)	
	B	BC	B	BC
26	94.0	93.0	134.0 (0.193)	135.0 (0.268)
52	93.5	93.0	150.3 (0.139)	150.9 (0.189)
78	92.0	91.5	153.9 (0.125)	155.1 (0.234)

NOTE: Nominal coverage: 95%.

are not directly comparable. However, our empirical studies suggest to give some preference to the area-level modeling in the considered GLMM settings.

Our simulations lead us to following conclusions. First, for a given sample size and data, our SCI attains the nominal coverage probability, almost independently from the choice of the estimator of variability. In particular, the area-level models yield very accurate results even for small samples. Second, the distinction between SCI and iCI is crucial, and the latter should not be employed in comparative studies. Third, the numerical performance of our test for comparative studies is satisfying. Given the simplicity of SCI and tests based on $\sqrt{\hat{g}_{1d}}$, we restrict further presentations to them.

Remark 3. In our simulation study, we do not analyze the performance of direct estimators for proportions, because our goal is to study the numerical performance of our MTP and SCIs, and to compare them to existing iCIs. Since MTP and

SCI are the first tools for simultaneous inference with GLMM-based mixed parameter, we concentrate on their implementation and application to the well-known model-based estimators. These have been thoroughly examined in comparative analyses which included direct estimators (see for instance Boubeta, Lombardía, and Morales 2016; Hobza and Morales 2016). In our case study in Section 5, we include direct estimators in order to have an almost model-free benchmark.

5. Predicting Poverty Rates in Galicia

Poverty prediction is of great interest for statistical offices. It provides a basis on which local or central authorities can decide about resource allocation and related polices. The interest is not in individual, randomly chosen small areas but in the total picture. Resource distribution requires comparative statistics, and one would thus provide SCI instead of iCI. We illustrate our methodology calculating point estimates, iCIs and SCIs for the poverty rates in each county of Galicia, that is, the proportions of inhabitants who live under a poverty line. We make use of a general part of the Structural Survey for Homes (SSH) in Galicia in 2015 with 23,628 individuals within 9203 households located in 52 counties (small areas). The survey does not produce official estimates at the area level, but we managed to recover the direct estimates of the totals of people below the poverty line (Y_d), as well as the number of inhabitants (N_d) for each county. For the area-level models, we need to calculate the number of

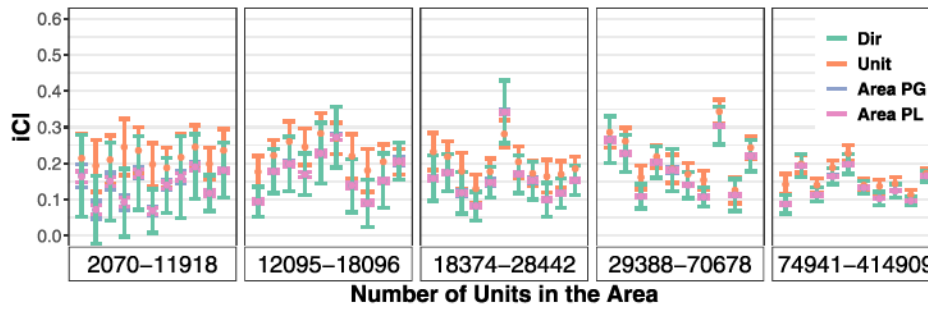


Figure 3. Design and model-based 95% iCIs.

units which fall into a particular category (X_{di}), for example, number of employees or of graduates in each county of Galicia, $i = 1, \dots, p$. The latter are used to obtain the proportions of individuals in each category $\bar{X}_{di} = X_{di}/N_d$. For the unit-level model, we need to obtain the number of units N_{dl} falling into artificially created categories z_{dl} , $d = 1, \dots, D$, $l = 1, \dots, L$, see Section 2.2.3. The explicit formulas are

$$\begin{aligned} \hat{Y}_d^{\text{dir}} &= \sum_{j \in \mathcal{R}_d} w_j y_j, & \hat{N}_d^{\text{dir}} &= \sum_{j \in \mathcal{R}_d} w_j, \\ \hat{N}_{dl}^{\text{dir}} &= \sum_{j \in \mathcal{R}_d} w_j x_j \mathbb{1}_{\{x_j = z_l\}}, & & \\ \hat{X}_{di}^{\text{dir}} &= \sum_{j \in \mathcal{R}_d} w_j x_{ji}, & \hat{\hat{X}}_{di}^{\text{dir}} &= \hat{X}_{di}^{\text{dir}} / \hat{N}_d^{\text{dir}}, \end{aligned} \quad (25)$$

where $\mathcal{R}_d \in \mathcal{P}_d$ are the sample elements belonging to area d , $d \in [D]$, w_{dj} sampling weights, and y_{dj} binary variables with 1 indicating that an individual is below the poverty line. The poverty threshold is calculated from the survey. It is set to 0.6 of the median household income per capita in Galicia, that is, we do not use county specific poverty lines. This income is calculated in each household according to scale developed by the Organisation for Economic Co-operation and Development (the same technique is used by Eurostat). The model-based approach of this paper assumes that the estimates in Equation (25) are considered to be known, nonrandom quantities, following López-Vizcaíno, Lombardía, and Morales (2015). SSH provides many categorical, auxiliary variables. Under the unit-level model these are binary variables with 1 indicating that a person belongs to a particular category, whereas under area-level models we use the county proportions. We considered four variables for labor status: children (ls0), employed (ls1), unemployed (ls2), inactive (ls3), and four covariates for education: less than primary (ed0), primary (ed1), first- and second-level secondary (ed2), higher education (ed3). Furthermore, we analyzed three variables for the size of municipality: less than 10,000 (sm1), 10,000–50,000 (sm2), more than 50,000 (sm3). We have also investigated the effect of two variables indicating the nationality, that is, Spanish (n1), not Spanish (n2). Finally, we examined five age variables: < 15 (age1), 15 – 24 (age2), 25 – 49 (age3), 50 – 64 (age4), >= 65 (age5). We are interested in $\bar{\mu}_d^{(\cdot)} := \mu_d^{(\cdot)} / N_d$ with (\cdot) standing for PG or U in case of Poisson-gamma and binomial model, respectively, and in ρ_d in case of the Poisson-lognormal model. We first compute estimates of proportions and their variances using the same formulas as Boubeta, Lombardía, and Morales (2016)

$$\hat{p}_d^{\text{dir}} = \hat{Y}_d^{\text{dir}} = \frac{\bar{Y}_d^{\text{dir}}}{\hat{N}_d^{\text{dir}}},$$

$$\widehat{\text{var}}(\hat{p}_d^{\text{dir}}) = \frac{1}{(\hat{N}_d^{\text{dir}})^2} \sum_{j \in \mathcal{R}_d} w_j (1 - w_j) (y_j - \hat{p}_d^{\text{dir}})^2. \quad (26)$$

We used estimates in Equation (26) to construct design-based iCI intervals (Dir) displayed in Figure 3. Following López-Vizcaíno, Lombardía, and Morales (2015), we then proceed with a variable selection inspired by the simulation results. More specifically, under the Poisson-gamma model we check if any of the levels of categorical variables for labor status, education and age are significant at the $\alpha = 0.05$ level. We examined these covariates in the first place, because they turned out to be important in earlier studies on poverty rates (see, for instance, Boubeta, Lombardía, and Morales 2016). In this way, we selected ls2, ed2, and age2. Afterwards, we tested the levels of variables nationality and the size of the municipality and we additionally retained sm1 which was significant after the selection of ls2, ed2, and age2. The same categories were then used to other models, see Table 4. As we do not carry out a causality analysis, we refrain ourselves from a discussion of the magnitude or signs of estimates. We only notice that under the Poisson-gamma model, the signs are consistent with our expectations; unemployment and young age are associated with higher poverty rates, whereas higher level of studies or living in a small municipality is associated with lower poverty rates.

Figure 3 shows point and iCI estimates of proportions under Poisson-gamma (PG), Poisson-lognormal (PL), and binomial (Unit) models together with direct estimates (Dir). In this plot, we compare point estimates within four modeling frameworks; we do not compare them across different areas within the same model. First, the variability reflected by the width of iCIs decreases with the number of units in each area \hat{N}_d^{dir} defined in Equation (25). Second, even though the area sample sizes n_d , $d \in [D]$ are not that small, the iCI of direct estimates are wider than model-based estimates, which is in accordance with the literature. The width difference is especially pronounced when comparing area-level-based with design-based direct estimates—the latter entirely cover the former. Unit-level model-based point and interval estimates are different with much wider iCIs than under area-level models, but still overlapping with the direct estimates. Only in one case (sixth area in the third panel), the iCIs under area-level models do not overlap with the iCI under the unit-level model which indicates a possible bias in one of the approaches. In contrast, both area-level models produce almost identical estimates.

Figure 4 presents bootstrap iCI and SCI for $\hat{\mu}_d^{\text{PG}}$, $d = 1, \dots, D$ constructed with $\hat{\sigma}(\hat{\zeta}_d) = \sqrt{\hat{\sigma}_{\text{PG}1d}}$ as defined in Equa-

Table 4. Estimates of regression parameters under the area- and the unit-level models with $\delta^{PG} = 2.48$, $\delta^{PL} = 0.32$ and $\delta^U = 0.35$, respectively.

	Poisson-gamma				Poisson-lognormal				Unit-level model			
	$\hat{\beta}$	SE	z-value	$P(> z)$	$\hat{\beta}$	SE	z value	$Pr(> z)$	$\hat{\beta}$	SE	z-value	$P(> z)$
Int	10.038	0.669	15.005	0.000	-2.264	0.341	-6.633	0.000	-2.048	0.067	-30.415	0.000
ls2	7.747	3.091	2.506	0.012	3.480	1.577	2.207	0.027	0.989	0.052	19.160	0.000
ed2	-3.136	1.201	-2.611	0.009	-0.870	0.612	-1.422	0.155	0.172	0.039	4.442	0.000
age2	11.317	4.023	2.813	0.005	4.842	2.057	2.354	0.019	0.760	0.058	13.033	0.000
sm1	-2.466	0.267	-9.224	0.000	0.125	0.136	0.918	0.358	0.100	0.050	1.993	0.046

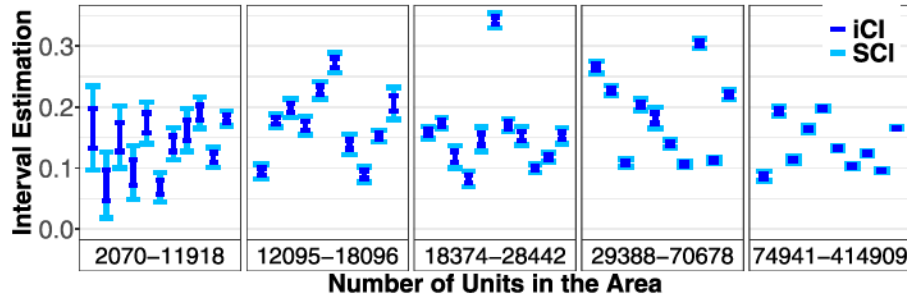


Figure 4. 95% bootstrap iCI and SCI estimates for poverty rates in counties of Galicia.

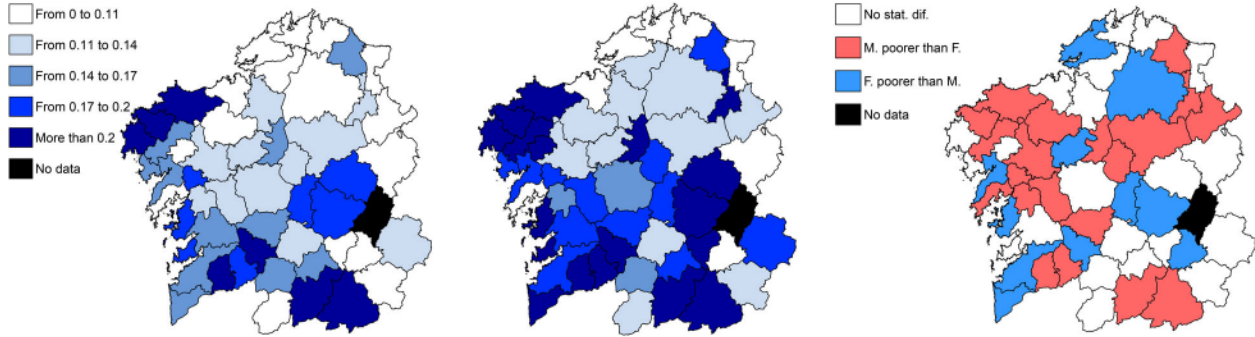


Figure 5. SCI of EBP poverty proportions: (left) lower boundary, (middle) upper boundary; (right) significant differences in poverty rates between women (F) and men (M).

tion (9). The plot is divided into five panels according to the numbers of units in each area obtained by direct estimates of county inhabitants N_d^{dir} in (25). Figure 4 serves as an illustration of the differences between individual and simultaneous inference. When comparing iCI and SCI, in many cases (e.g., first and second county of the first panel in Figure 4) iCI would insinuate statistically different poverty rates, whereas SCI does not confirm this claim. Such multiple comparisons are valid only if we use SCI. In addition, at least 5% of the true poverty rates are not contained in their iCIs. Analogous figures under the Poisson-lognormal and binomial models lead to the same conclusions. They are thus deferred to the SM. Further model selection and specification testing might be interesting but they are beyond the scope of this article.

Since we do not know which model is closer to the real data-generating process, we proceed with the Poisson-gamma area-level model, as it is reliable and the least computer intensive. Left and middle panel of Figure 5 depict the resulting maps of the counties with the corresponding lower and upper bounds of our SCI. We observe a higher rate of poverty in the interior and a south-western part of the region whereas a lower level is typical for the northern part. These conclusions are similar to those drawn by Boubeta, Lombardía, and Morales (2017).

Finally, we investigate whether men and women are equally affected by poverty. We wish to test for equality on the county level across Galicia. Testing for each county individually at $\alpha = 5\%$ error level results in rejection of at least 5% of the hypotheses of no significant difference. We thus use our MTP and consider clusters created from the cross section of sex and county such that $\zeta \in \mathbb{R}^{104}$. We test $H_0 : B\zeta = \mathbf{0}_{52}$ vs. $H_1 : B\zeta \neq \mathbf{0}_{52}$ where $B \in \mathbb{R}^{52 \times 104}$ with rows being vectors with 1 on the $2d - 1$ place, -1 on $2d$ place, and 0 elsewhere. The max-type test statistic yields $t_H = \max_{d=1, \dots, 104} |B\hat{\zeta}| / \hat{\sigma}(\hat{\zeta}) \approx 20.489$ while the bootstrap critical value under H_0 is $q_{BH_0}^{(1-\alpha)} \approx 2.999$. Thus, we strongly reject H_0 . However, our test does not support the hypothesis that women are more affected than men, or vice versa, see the right panel of Figure 5. Additional results are deferred to our SM.

Remark 4. Imagine that Galician counties were considered as a part of a macro region, for example, Spain with D_S counties, and consider two inferential problems: (a) the calculation of SCI for the poverty rates in all D_S Spanish counties, (b) the calculation of SCI only for D Galician counties, but using all data. Following Remark 1, we expect that the widths of our SCI in Figure 4 would increase in case (a) to maintain the joint coverage probability of

95% for all $D_S > D$ counties. In contrast, they would most likely slightly decrease in case (b). In fact, the simultaneous coverage probability of 95% would be requested for the set of D counties, but SCI would be constructed using a more precise estimate of MSE computed using a larger dataset with D_S counties.

6. Conclusions

We developed a methodology that allows for statistically valid simultaneous inference for EBPs under GLMM. We constructed SCI and MTP applying a combination of max-type statistics and consistent bootstrap estimation of its distribution. These tools enable practitioners to make comparisons between areas. In contrast, the iCIs are not suitable for such comparative analyses because they are constructed at individual confidence level and disregard an additional variation which arises in joint studies. We do not claim that SCI and MTP are better than iCIs or t -tests. The former simply complete the toolbox for statistical inference for mixed parameter ζ_d . Similarly, the simultaneous inference completes the individual inference for fixed parameters. We introduced various versions of statistics to construct SCI and MTP. Within our framework, all of them exhibited similar performances without indicating a clear winner.

Our methodology can be extended to more complicated data structures such as GLMM with spatial or temporal correlation (see, e.g., Hobza, Morales, and Santamaría 2018; Chandra, Chambers, and Salvati 2019). One could also consider spatio-temporal or nonparametric models to build SCI by adjusting the statistic S_0 and choosing a bootstrap procedure accordingly. Apart from a mathematical challenge to develop a valid asymptotic theory, these extensions would require a construction of an appropriate bootstrap scheme and its computationally efficient implementation.

Supplementary Materials

The supplementary materials consist of: (a) a document with further developments, in particular additional MSE decomposition, a proof of Proposition 1, the derivations of estimators under area-level Poisson and unit-level binomial models, additional numerical results and a data analysis which completes the case study in Section 5, (b) codes for replicating the results in the main document, and (c) a document which contains additional information on the data set and the description of the codes.

Appendix A: Technical details

A.1. Regularity conditions

In this section, we state the regularity conditions used in our derivations.

- $\hat{u}_d = \arg \max_{u_d \in \mathbb{R}} \{\log g_d(y_d|u_d, \theta) + \log h(u_d)\}$.
- $l(\theta)$ exists and is well-defined if: (a) $l(\theta)$ is continuous, uniquely maximized and $\theta_0 \in \Theta$, where θ_0 is a true parameter value; (b) $l(\theta)$ and $\hat{l}(\theta)$ are concave; (c) θ_0 is an interior point of the parameter space and the estimator $\hat{\theta}$ is an interior point of the neighborhood of θ_0 ; (d) $\hat{l}(\theta)$ converges uniformly in probability to $l(\theta)$.
- x_{dj} are bounded and $\mathbb{E}(y_{dj}^m) < \infty$ for all $d \in [D], j \in [n_d]$, where m is suitable large.
- For each fixed y , a score equation is continuously differentiable and $\mathbb{E}\{R(\theta_0)\} = 0$.

- $\liminf_n \lambda[n^{-1}\text{Var}\{R(\theta)\}] > 0$ and $\liminf_n \lambda[-n^{-1}\mathbb{E}\{\nabla R(\theta)\}] > 0$ where $\nabla R(\theta) = \frac{\partial R(\theta)}{\partial \theta}$ and $\lambda[A]$ indicates the smallest eigenvalue of matrix A .

The first two conditions refer to the log-likelihood function (see, e.g., Bianconcini 2014), whereas conditions 3–5 are needed for the derivation of the MSE estimators.

A.2. Proof of Proposition 2

Let $y_{dj}^* \sim \text{Exp.Family}(\theta)$. If u_d^* is sampled from a suitable distribution, then we have $\gamma_{dj}^* = M\{\mathbb{E}(y_{dj}^*|u_d)\} = x_{dj}^t \hat{\beta} + u_d^*$. Furthermore, $\text{Var}^*(y_d^*) = \text{Var}^*(\mathbb{E}^*(y_d^*|u_d^*)) + \mathbb{E}^*(\text{Var}^*(y_d^*|u_d^*))$. The first part of the Proposition follows from the way we generate the random effects as well as the results on the consistency of $\hat{\theta}$. To show the second part we consider a general score equation. Replace y by y^* and set $\theta = \hat{\theta}$, that is, $R^*(\theta) = \frac{\partial l^*(\hat{\theta})}{\partial \theta} = \sum_{d=1}^D \frac{\partial \log f_d(y_d^*|\hat{\theta})}{\partial \theta} = 0$. Then $\mathbb{E}^*\{R^*(\theta)\} = 0$ at $\theta = \hat{\theta}$ which yields consistency of $\hat{\theta}^*$. \square

A.3. Proof of Proposition 3

Let ζ_d be a general EBP, $g_d := g_d(\theta)$ and $\hat{g}_d := \hat{g}_d(\hat{\theta})$. Assume that $\|\hat{g}_d - g_d\| = O_p(n^{-c}), c > 0$. The proof uses ideas of Chatterjee, Lahiri, and Li (2008). We investigate the properties of $G_d(a)$.

$$\begin{aligned} G_d(a) &= P\left(\frac{\hat{\zeta}_d - \zeta_d}{\sqrt{\hat{g}_d}} \leq a\right) \\ &= \mathbb{E}\left(P\left[\frac{\hat{\zeta}_d - \zeta_d}{\sqrt{\hat{g}_d}} \leq a + \left\{\frac{a(\sqrt{\hat{g}_d} - \sqrt{g_d}) + \hat{\zeta}_d - \zeta_d}{\sqrt{\hat{g}_d}}\right\}\right] \middle| y_d\right) \\ &= \mathbb{E}[\Phi(a + Q(a, y_d))] \\ &= \Phi(a) + \phi(a)\mathbb{E}\{Q(a, y_d)\} - 2^{-1}a\phi(a)\mathbb{E}\{Q^2(a, y_d)\} \\ &\quad + 2^{-1}\mathbb{E}\left[\int_a^{a+Q(a, y_d)} \{a + Q(a, y_d) - x\}^2 (x^2 - 1)\phi(x)dx\right]. \end{aligned}$$

Applying some classical results and a triangle inequality, it follows that the last term is bounded by $\mathbb{E}|Q|^3$, and is of smaller order than the first three terms. Therefore, the first step toward the consistency of SCI is to quantify the asymptotic expansions of $\mathbb{E}\{Q(a, y_d)\}$ and $\mathbb{E}\{Q^2(a, y_d)\}$. We decompose $Q(a, y_d)$ into

$$Q(a, y_d) = g_d^{-1/2}(\hat{\zeta}_d - \zeta_d) + a g_d^{-1/2}(\hat{g}_d^{1/2} - g_d^{1/2}) = Q_1 + Q_2.$$

Let ψ be a twice differentiable function with respect to $\theta, y_d = y_{d1} + \dots + y_{dn_d}, \forall d \in [D]$. Observe that $y_d = y_d$ under an area-level model. The specific form of ψ depends on the choice of the GLMM (for instance, under the Poisson-gamma model we spelled it out in Equation (7)). Function ψ satisfies the decomposition

$$\begin{aligned} \hat{\zeta}_d - \zeta_d &= \psi_d(y_d, \hat{\theta}) - \psi_d(y_d, \theta) = \left\{\frac{\partial}{\partial \theta} \psi_d(y_d, \theta)\right\}^t (\hat{\theta} - \theta) \\ &\quad + \frac{1}{2} (\hat{\theta} - \theta)^t \left\{\frac{\partial^2}{\partial \theta^2} \psi_d(y_d, \theta)\right\} (\hat{\theta} - \theta) + o_p(\|\hat{\theta} - \theta\|^2). \end{aligned} \tag{A.1}$$

Let $C = 2c$ where $c > 0$. Since we assume $\|\hat{\theta} - \theta\| = O_p(n^{-c})$, we have

$$\mathbb{E}[(\hat{\zeta}_d(\hat{\theta}) - \zeta_d(\theta))^2] = \frac{1}{n^C} \mathbb{E}\left[\left[\left\{\frac{\partial}{\partial \theta} \psi_d(y_d, \theta)\right\}^t n^c (\hat{\theta} - \theta)\right]^2\right] + o(n^{-C}). \tag{A.2}$$

As for Q_1 , it has been found in Equation (A.1) that

$$\mathbb{E}(\hat{\zeta}_d - \zeta_d) = \frac{1}{n^C} \mathbb{E}\left[\left\{\frac{\partial}{\partial \theta} \psi_d(y_d, \theta)\right\}^t n^c (\hat{\theta} - \theta)\right] + o(n^{-C}),$$

and $\mathbb{E}\left\{\left(\hat{\xi}_d - \tilde{\xi}_d\right)^2\right\} = O(n^{-C})$, thanks to the result in Equation (A.2). Furthermore, observe that g_d is of order $O(1)$ which leads to $\mathbb{E}(Q_1) = O(D^{-c})$ as well as $\mathbb{E}(Q_2^2) = O(D^{-C})$. When we turn to Q_2 , we have an immediate simplification $Q_2 = ag_d^{-1/2}(\hat{g}_d^{1/2} - g_d^{1/2}) = a\left\{\left(\hat{g}_d/g_d\right)^{1/2} - 1\right\}$. Let g_d be twice differentiable with respect to θ . Similarly to the computations above, we have the expansion

$$\begin{aligned} \hat{g}_d(\hat{\theta}) &= g_d(\theta) + \left(\frac{\partial}{\partial\theta}g_d(\theta)\right)^t (\hat{\theta} - \theta) + \frac{1}{2} (\hat{\theta} - \theta)^t \\ &\quad \times \left(\frac{\partial^2}{\partial^2\theta}g_d(\theta)\right) (\hat{\theta} - \theta) + o_p(\|\hat{\theta} - \theta\|^2). \end{aligned}$$

Therefore, we obtain


$$\mathbb{E}\{\hat{g}_d(\hat{\theta})\} = g_d(\theta) + \frac{1}{n^C} \mathbb{E}\left[\left\{\frac{\partial}{\partial\theta}g_d(\theta)\right\}^t n^C (\hat{\theta} - \theta)\right] + O(n^{-C}).$$

It follows that $\mathbb{E}\left\{\left(\hat{g}/g\right)^{1/2}\right\} = O(n^{-c})$, $\mathbb{E}(Q_2) = O(n^{-C})$ and $\mathbb{E}(Q_2^2) = O(n^{-C})$. We can deduce that $G_d(a)$ attains the asymptotic expansion $G_d(a) = \Phi(a) + n^{-c}\gamma(a, \theta) + O(n^{-C})$. A similar expansion can be established for $G_d^*(a)$ if we replace θ with $\hat{\theta}$ and P with P^* .

Funding

The authors gratefully acknowledge the support from the Swiss National Science Foundation for the project 200021-192345. In addition, they acknowledge the 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. The computations were performed at the University of Geneva on the Baobab cluster.

ORCID

Katarzyna Reluga  <http://orcid.org/0000-0001-9651-5030>

María-José Lombardía Cortiña  <http://orcid.org/0000-0001-9452-9818>

References

- Bianconcini, S. (2014), "Asymptotic Properties of Adaptive Maximum Likelihood Estimators in Latent Variable Models," *Bernoulli*, 20, 1507–1531. [584,587,593]
- Boubeta, M., Lombardía, M. J., and Morales, D. (2016), "Empirical Best Prediction Under Area-Level Poisson Mixed Models," *Test*, 25, 548–569. [583,584,585,586,589,590,591]
- (2017), "Poisson Mixed Models for Studying the Poverty in Small Areas," *Computational Statistics & Data Analysis*, 107, 32–47. [592]
- Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25. [584]
- Cameron, A. C., and Trivedi, P. K. (2013), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press. [585]
- Chambers, R., Salvati, N., and Tzavidis, N. (2012), "M-Quantile Regression for Binary Data With Application to Small Area Estimation," Centre for Statistical and Survey Methodology, University of Wollongong, Working Paper, 1–24. [583]
- Chandra, H., Chambers, R., and Salvati, N. (2012), "Small Area Estimation of Proportions in Business Surveys," *Journal of Statistical Computation and Simulation*, 82, 783–795. [583]
- (2019), "Small Area Estimation of Survey Weighted Counts Under Aggregated Level Spatial Model," *Survey Methodology*, 45, 31–59. [593]
- Chatterjee, S., Lahiri, P., and Li, H. (2008), "Parametric Bootstrap Approximation to the Distribution of EBLUP and Related Prediction Intervals in Linear Mixed Models," *Annals of Statistics*, 36, 1221–1245. [585,587,588,593]
- Chen, S., Jiang, J., and Nguyen, T. (2015), "Observed Best Prediction for Small Area Counts," *Journal of Survey Statistics and Methodology*, 3, 136–161. [585]
- De Bruijn, N. G. (1981), *Asymptotic Methods in Analysis*, New York: Dover Publications, Inc. [584]
- DiCiccio, T. J., and Efron, B. (1996), "Bootstrap Confidence Intervals," *Statistical Science*, 11, 189–228. [586]
- Erciulescu, A. L., and Fuller, W. A. (2014), "Parametric Bootstrap Procedures for Small Area Prediction Variance," in *Proceedings of the Joint Statistical Meeting-Survey Research Methods Section*, Boston, pp. 3307–3318. [588]
- Franco, C., and Bell, W. R. (2015), "Borrowing Information Over Time in Binomial/Logit Normal Models for Small Area Estimation," *Statistics in Transition New Series*, 16, 563–584. [583,585]
- Ganesh, N. (2009), "Simultaneous Credible Intervals for Small Area Estimation Problems," *Journal of Multivariate Analysis*, 100, 1610–1621. [583,588]
- Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. P. (1998), "Generalized Linear Models for Small-Area Estimation," *Journal of the American Statistical Association*, 93, 273–282. [583]
- Hall, P., and Maiti, T. (2006), "On Parametric Bootstrap Methods for Small Area Prediction," *Journal of the Royal Statistical Society, Series B*, 68, 221–238. [584]
- Hobza, T., and Morales, D. (2016), "Empirical Best Prediction Under Unit-Level Logit Mixed Models," *Journal of Official Statistics*, 32, 661–692. [583,584,586,590]
- Hobza, T., Morales, D., and Santamaría, L. (2018), "Small Area Estimation of Poverty Proportions Under Unit-Level Temporal Binomial-Logit Mixed Models," *Test*, 27, 270–294. [593]
- Huber, P., Ronchetti, E., and Victoria-Feser, M.-P. (2004), "Estimation of Generalized Linear Latent Variable Models," *Journal of the Royal Statistical Society, Series B*, 66, 893–908. [587]
- Jiang, J. (1998), "Consistent Estimators in Generalized Linear Mixed Models," *Journal of American Statistical Association*, 93, 720–729. [584]
- (2003), "Empirical Best Prediction for Small-Area Inference Based on Generalized Linear Mixed Models," *Journal of Statistical Planning Inference*, 111, 117–127. [583,584]
- (2007), *Linear and Generalized Linear Mixed Models and Their Applications*, Springer Science & Business Media, New York: Springer. [583,584]
- (2013), "The Subset Argument and Consistency of MLE in GLMM: Answer to an Open Problem and Beyond," *Annals of Statistics*, 41, 177–195. [587]
- Jiang, J., and Lahiri, P. (2001), "Empirical Best Prediction for Small Area Inference With Binary Data," *Annals of the Institute of Statistical Mathematics*, 53, 217–243. [584]
- Kramlinger, P., Krivobokova, T., and Sperlich, S. (2018), "Marginal and Conditional Multiple Inference in Linear Mixed Models," arXiv:1812.09250. [583]
- Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012), *Extremes and Related Properties of Random Sequences and Processes*, Springer Science & Business Media, New York: Springer. [587]
- Lele, S. R., Nadeem, K., and Schmuland, B. (2010), "Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning," *Journal of the American Statistical Association*, 105, 1617–1625. [584]
- López-Vizcaino, E., Lombardía, M. J., and Morales, D. (2015), "Small Area Estimation of Labour Force Indicators Under a Multinomial Model With Correlated Time and Area Effects," *Journal of Royal of Statistical Society, Series A*, 178, 535–565. [591]
- Molina, I., Saei, A., and Lombardía, M. J. (2007), "Small Area Estimates of Labour Force Participation Under a Multinomial Logit Mixed Model," *Journal of the Royal Statistical Society, Series A*, 170, 975–1000. [583]
- Naylor, J. C., and Smith, A. F. (1982), "Applications of a Method for the Efficient Computation of Posterior Distributions," *Journal of Royal Statistical Society, Series C*, 31, 214–225. [584]

- Pinheiro, J. C., and Bates, D. M. (1995), "Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model," *Journal of Computational and Graphical Statistics*, 4, 12–35. [584]
- Pratesi, M. (2016), *Analysis of Poverty Data by Small Area Estimation*, New Jersey: Wiley. [583]
- Reluga, K., Lombardía, M. J., and Sperlich, S. A. (2019), "Simultaneous Inference for Mixed and Small Area Parameters," arXiv:1903.02774. [583]
- Romano, J. P., Shaikh, A. M., and Wolf, M. (2008), "Control of the False Discovery Rate Under Dependence Using the Bootstrap and Subsampling," *Test*, 17, 417–442. [588]
- Saei, A., and Taylor, A. (2012), "Labour Force Status Estimates Under a Bivariate Random Components Model," *Journal of the Indian Society of Agricultural Statistics*, 66, 187–201. [583]
- Scealy, J. (2010), "Small Area Estimation Using a Multinomial Logit Mixed Model With Category Specific Random Effects," Research Paper, Australian Bureau of Statistics. [583]
- Torabi, M. (2012), "Likelihood Inference in Generalized Linear Mixed Models With Two Components of Dispersion Using Data Cloning," *Computational Statistics & Data Analysis*, 56, 4259–4265. [584]
- Tuerlinckx, F., Rijmen, F., Verbeke, G., and De Boeck, P. (2006), "Statistical Inference in Generalized Linear Mixed Models: A Review," *British Journal of Mathematical and Statistical Psychology*, 59, 225–255. [583,584]
- Tzavidis, N., Ranalli, M. G., Salvati, N., Dreassi, E., and Chambers, R. (2015), "Robust Small Area Prediction for Counts," *Statistical Methods in Medical Research*, 24, 373–395. [583]