Empirical best prediction under area-level Poisson mixed models

Miguel Boubeta, María José Lombardía

Departamento de Matemáticas, Universidade da Coruña.

Domingo Morales

Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández de Elche.

Abstract

The paper studies the applicability of area-level Poisson mixed models to estimate small area counting indicators. Among the available procedures for fitting generalized linear models, the method of moments (MM) and the penalized quasi-likelihood (PQL) method are employed. The empirical best predictor (EBP) of the area mean is derived using MM and compared with plug-in alternatives using MM and PQL. The plug-in estimator using PQL is computationally faster and provides competitive performance with respect to EBP that involves high complex integrals. An approximation to the mean squared error (MSE) of the EBP is given and three MSE estimators are proposed. The first two MSE estimators are plug-in estimators without and with bias correction to the second order and the third one is based on parametric bootstrap. Several simulation experiments are carried out for analyzing the behavior of the EBP and for comparing the estimators of the MSE of the EBP. A good choice in practice is the bootstrap alternative since it performs similarly to the analytical versions and is computationally faster. The developed methodology and software are applied to data from the 2008 Spanish living condition survey. The target of the application is the estimation of poverty rates at province level.

keywords: Bootstrap, Empirical best predictor, Mean squared error, Method of moments, Poisson mixed models, Poverty.

1 Introduction

One in five people is at risk of poverty or social exclusion in the European Union (EU). For reducing this amount, the EU set national targets between all its members. Most European countries use the Living Conditions Survey (LCS) to estimate poverty indicators. The Spanish Living Conditions Survey (SLCS) provides information regarding the household income received during the year prior to that of the interview. For every individual, the equivalent personal income is obtained by dividing the annual household net income by the equivalent total of household members, which is obtained as a weighted sum assigning weights 1 to the first adult, 0.5 to remaining adults and 0.3 to children under 14 years of age.

The poverty line is defined as a percentage (currently Eurostat fixed it in 60%) of the median of the equivalent personal incomes in the whole country. A person is defined as poor if his/her equivalent personal income is lower than the poverty line. Poverty rate is the proportion of people under the poverty line. This is a relative measure depending on the incomes of all the household members. Therefore, employment policies, education and welfare can have a significant impact on levels of poverty rate. Policy makers are interested in finding out which factors are more influential for poverty in order to act on them.

The SLCS planned domains are the Spanish autonomous communities. Therefore, SLCS direct estimators are not precise enough for estimating poverty rates at a lower aggregation level than autonomous communities (e.g. provinces or counties). Small area estimation (SAE) deals with this problem by introducing model-based or model assisted estimators. See the monograph Rao (2003) and the reviews of Ghosh and Rao (1994), Rao (1999), Pfeffermann (2002), Jiang and Lahiri (2006), Rao (2008) and Pfeffermann (2013) for an introduction to the SAE.

Poisson regression and binomial-logit models are generalized linear models (GLM) that are used for counts, i.e. for target variables counting some event of interest (like being under poverty line). In these models the hypothesis of linearity is relaxed in the sense that a function, called link, of the mean of the observations is linear in some set of covariates. The hypothesis of normality is also relaxed to the assumption that the distribution belongs to the exponential family.

Sometimes the GLM cannot explain the variability of the target variable through the selected auxiliary variables. It may happens that observations from different domains are independent, but observations within the same domain are dependent because they share common properties. The generalized linear mixed models (GLMM) are extensions of GLM that capture the variability between domains by introducing random effects. The random effects are usually assumed to be normally distributed.

Despite the usefulness of GLMM, inferences based on these models have some computational difficulties because the likelihood may involve high-dimensional integrals which cannot be evaluated analytically. Several methods have been proposed to overcome this problem, most of them relying on the Taylor linearization and/or on the Laplace's method for integral approximations (see the review of Jiang and Lahiri 2006). EM-type algorithms assisted by Monte Carlo methods are also applied. The penalized quasi-likelihood (PQL) algorithm (Breslow and Clayton 1993; Lin 2007; MacNab and Lin 2009) is used in combination with a Gaussian approximation of the marginal density that provides approximate maximum likelihood estimators of variance components. Unfortunately, in some cases the PQL method may lead to inconsistent and biased estimators (Jiang 1998). This paper uses the method of moments (MM) for fitting the proposed area-level Poisson mixed model, which is based on the method of simulated moments introduced by Jiang (1998). This method is computationally attractive and gives consistent estimators of model parameters.

The paper derives empirical best predictors (EBP) based on area-level Poisson mixed models for estimating count indicators. The statistical methodology is taken and adapted from Jiang and Lahiri (2001) and Jiang (2003), where EBPs of functions of fixed effects and small area specific random effects were developed in the context of logistic mixed models and GLMM respectively. In addition to the EBPs, plug-in estimators are considered and empirically studied in simulation experiments.

We consider the mean squared error (MSE) as an accuracy measure of the EBP. The estimation of the MSE is not an easy task. Prasad and Rao (1990) studied the accuracy of a second-order approximation to the MSE of empirical best linear unbiased predictor (EBLUP) for three special cases of linear mixed models: Fay-Herriot model, nested error regression model and random regression coefficient model. Jiang and Lahiri (2001) and Jiang (2003) studied the approximation of the MSE of the EBP in the context of binary and GLMM data. Their approach is based on Taylor series expansions. They further gave a second-order bias corrected estimator of the MSE. We adapt the MSE calculations given by Jiang and Lahiri (2001) and Jiang (2003) to the case of area-level Poisson mixed models. The obtained MSE approximation gives an accuracy measure for the EBP. We also give two analytical estimators of the MSE approximation, without and with bias-correction term. As the analytical estimators of MSE are computationally expensive in practice, we consider the parametric bootstrap estimator introduced by González-Manteiga et al. (2007) and González-Manteiga et al. (2008a) in the context of logistic and normal mixed models and later extended by González-Manteiga et al. (2008b) to a multivariate area-level model. We carry out a simulation experiment for empirically investigating the behavior of the MSE estimators.

For estimating small area counting indicators, area level versions of generalized linear mixed model (GLMM) with logit link function, and with combination of Penalized Quasi-Likelihood (PQL) and REML for estimation of unknown parameters have been considered by Saei and Chambers (2003), Johnson et al. (2010), López-Vizcaíno et al. (2013) and López-Vizcaíno et al. (2015). They use plug-in model predictors having analytical MSE for approximation of true MSE.

Poisson-log mixed models and binomial-logit mixed models are competitor models for count data at the area-level. For a given real data set, it is interesting to compare domain predictors (EBP or plug-in) based on these models. Note also that the Fay-Herriot model might also be a competitor. This is because of the asymptotic relationships between the Poisson, the binomial and the normal distribution. In our application to real data we are mainly interested in studying the behavior of the estimators introduced in the paper (EBP or plug-in based on the Poisson model). Nevertheless, we also include the well known EBLUP based on the Fay-Herriot model. As this is not a case-of-study paper, we do not include the other cited estimators.

The paper is organized as follows. Section 2 introduces the area-level Poisson mixed model and the employed fitting algorithm. Section 3 presents the EBP and the plug-in estimators of functions of fixed and small area specific random effects. Section 4 gives an approximation to the MSE of the EBP and three estimators. The first two MSE estimators are plug-in derivations of the MSE approximation without and with bias correction term. The third MSE estimator is based on parametric bootstrap. Section 5 presents a complete simulation study, evaluating the performance of the model-based estimators under model-based and design-based simulations. In both cases, the simulations mimic the real data study case. Section 6 applies the developed methodology to data from the SLCS2008. The target is the estimation of mean and women poverty rates at province level. Section 7 gives some conclusions. The appendix contains detailed proofs of main results.

2 The model

This section introduces an area-level Poisson mixed model and its fitting algorithm. Let D be the number of small areas or domains, with $d = 1, \dots, D$. Let $\{v_d : d = 1, \dots, D\}$ be a set of i.i.d. $N(0, 1)$ random effects. In matrix notation, we have $\boldsymbol{v} = (v_1, \dots, v_D)' \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D)$, where $\boldsymbol{I}_D$ is the $D \times D$ unit matrix. We assume that the distribution of the target variable y_d, conditioned to the random effect v_d, is

$$y_d|_{v_d} \sim \text{Poiss}(\mu_d), \quad d = 1, \dots, D,$$

where $\mu_d > 0$. The Poisson distribution is closely related to the binomial distribution since it can be derived as a limiting case when the number of trials goes to infinity and the probability of the event of interest is sufficiently small. Therefore, we have that $\mu_d = \nu_d p_d$, where ν_d is the size variable and p_d is the binomial probability. For the natural parameter, we assume

$$\eta_d = \log \mu_d = \log \nu_d + \boldsymbol{x}_d \boldsymbol{\beta} + \phi v_d, \quad d = 1, \dots, D,$$

where $\boldsymbol{\beta} = \underset{1\leq k\leq p}{\text{col}}(\beta_k)$ is a column vector of fixed regression coefficients and $\boldsymbol{x}_d = \underset{1\leq k\leq p}{\text{col}'}(x_{dk})$ is the row vector containing the auxiliary variables. Further, we assume that the y_d's are independent conditioned to $\boldsymbol{v}$. It holds that

$$\begin{aligned} P(y_d|\boldsymbol{v}) &= P(y_d|v_d) = \frac{1}{y_d!} \exp\{-\nu_d p_d\} \nu_d^{y_d} p_d^{y_d}, \quad p_d = \exp\{\boldsymbol{x}_d \boldsymbol{\beta} + \phi v_d\}, \\ P(\boldsymbol{y}|\boldsymbol{v}) &= \prod_{d=1}^{D} P(y_d|v_d), \quad P(\boldsymbol{y}) = \int_{R^D} P(\boldsymbol{y}|\boldsymbol{v}) f_v(\boldsymbol{v})\, d\boldsymbol{v} = \int_{R^D} \psi(\boldsymbol{y}, \boldsymbol{v})\, d\boldsymbol{v}, \end{aligned}$$

where

$$\psi(\boldsymbol{y},\boldsymbol{v}) = (2\pi)^{-\frac{D}{2}} \exp\left\{\frac{-\boldsymbol{v}'\boldsymbol{v}}{2}\right\} \prod_{d=1}^{D} \frac{\exp\{-\nu_d p_d\}\nu_d^{y_d} \exp\{y_d(\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d)\}}{y_d!}$$
$$= (2\pi)^{-\frac{D}{2}} \exp\left\{\frac{-\boldsymbol{v}'\boldsymbol{v}}{2}\right\} \big(\prod_{d=1}^{D} y_d!\big)^{-1} \exp\left\{\sum_{k=1}^{p}\big(\sum_{d=1}^{D} y_d x_{dk}\big)\beta_k + \phi\sum_{d=1}^{D} y_d v_d\right\}$$
$$\cdot \exp\left\{\sum_{d=1}^{D}\big\{-\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\} + y_d \log \nu_d\big\}\right\}.$$

To fit the area-level Poisson mixed model, we derive the algorithm suggested by Jiang (1998), using the method of moments (MM). A natural set of equations for applying this method is

$$0 = f_k(\boldsymbol{\theta}) = M_k(\boldsymbol{\theta}) - \hat{M}_k = \sum_{d=1}^{D} E_\theta[y_d]x_{dk} - \sum_{d=1}^{D} y_d x_{dk}, \quad k=1,\dots,p, \quad (1)$$

$$0 = f_{p+1}(\boldsymbol{\theta}) = M_{p+1}(\boldsymbol{\theta}) - \hat{M}_{p+1} = \sum_{d=1}^{D} E_\theta[y_d^2] - \sum_{d=1}^{D} y_d^2, \quad (2)$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi)'$ is the vector of model parameters. The MM estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ is the solution of the system of nonlinear equations (1)-(2). The updating formula of the Newton-Raphson algorithm for solving this system is

$$\boldsymbol{\theta}^{(r+1)} = \boldsymbol{\theta}^{(r)} - \boldsymbol{H}^{-1}(\boldsymbol{\theta}^{(r)})\boldsymbol{f}(\boldsymbol{\theta}^{(r)}), \quad (3)$$

where $\theta_1 = \beta_1, \dots, \theta_p = \beta_p$, $\theta_{p+1} = \phi$ and

$$\boldsymbol{\theta} = \underset{1\le k\le p+1}{\text{col}}(\theta_k), \quad \boldsymbol{f}(\boldsymbol{\theta}) = \underset{1\le k\le p+1}{\text{col}}(f_k(\boldsymbol{\theta})), \quad \boldsymbol{H}(\boldsymbol{\theta}) = \left(\frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_r}\right)_{k,r=1,\dots,p+1}.$$

Appendix A.1 gives the components of vector $\boldsymbol{f}$ and matrix $\boldsymbol{H}$ appearing in (3). A good seed for the MM Newton-Raphson algorithm is $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the maximum likelihood estimator under the model without random effects. Concerning the variance parameters, we use

$$\phi^{(0)} = \left(\frac{1}{D}\sum_{d=1}^{D}(\tilde{\eta}_d - \hat{\eta}_d^{(0)})^2\right)^{1/2},$$

where $\tilde{\eta}_d = \boldsymbol{x}_d\tilde{\boldsymbol{\beta}}$, $\hat{\eta}_d^{(0)} = \log \hat{p}_d^{(0)}$ and $\hat{p}_d^{(0)} = \frac{y_d+1}{\nu_d+1}$.

The asymptotic variance of the MM estimators can be approximated by a Taylor expansion of $\boldsymbol{M}(\hat{\boldsymbol{\theta}}) = \underset{1\le k\le p+1}{\text{col}}(M_k(\hat{\boldsymbol{\theta}}))$ around $\boldsymbol{\theta}$ (Jiang 1998). This is to say,

$$\hat{\boldsymbol{M}} = \boldsymbol{M}(\hat{\boldsymbol{\theta}}) \approx \boldsymbol{M}(\boldsymbol{\theta}) + \boldsymbol{H}(\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \quad \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \approx \boldsymbol{H}^{-1}(\boldsymbol{\theta})(\hat{\boldsymbol{M}} - \boldsymbol{M}(\boldsymbol{\theta})),$$

where $\hat{\boldsymbol{M}} = \underset{1\leq k\leq p+1}{\text{col}}(\hat{M}_k)$. Under regularity conditions (Jiang 1998), it holds

$$\text{var}(\hat{\boldsymbol{\theta}}) = E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'] \approx \boldsymbol{H}^{-1}(\boldsymbol{\theta})\text{var}(\hat{\boldsymbol{M}})\boldsymbol{H}^{-1}(\boldsymbol{\theta}).$$

An estimator of $\text{var}(\hat{\boldsymbol{\theta}})$ is

$$\widehat{\text{var}}(\hat{\boldsymbol{\theta}}) = \boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})\widehat{\text{var}}(\hat{\boldsymbol{M}})\boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}}),$$

where $\widehat{\text{var}}(\hat{\boldsymbol{M}})$ is an estimator of the covariance matrix of $\hat{\boldsymbol{M}}$.

The following parametric bootstrap procedure gives estimators of $\text{var}(\hat{\boldsymbol{M}})$ and $\text{var}(\hat{\boldsymbol{\theta}})$.

1. Fit the model to the sample and calculate $\hat{\boldsymbol{\theta}}$.
2. Generate bootstrap samples $\{y_d^{*(b)} : d = 1, \dots, D\}$, $b = 1, \dots, B$, from the fitted model.
3. From bootstrap sample, calculate $\hat{\boldsymbol{M}}^{*(b)}$, $b = 1, \dots, B$, and
$$\overline{\boldsymbol{M}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{M}}^{*(b)}, \quad \widehat{\text{var}}^*(\hat{\boldsymbol{M}}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\boldsymbol{M}}^{*(b)}-\overline{\boldsymbol{M}})(\hat{\boldsymbol{M}}^{*(b)}-\overline{\boldsymbol{M}})'.$$
4. Calculate $\widehat{\text{var}}_A(\hat{\boldsymbol{\theta}}) = \boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})\widehat{\text{var}}^*(\hat{\boldsymbol{M}})\boldsymbol{H}^{-1}(\hat{\boldsymbol{\theta}})$.

We obtain an alternative estimator of $\text{var}(\hat{\boldsymbol{\theta}})$ if we replace steps 3 and 4 by

3′. Fit the model to the bootstrap samples and calculate $\hat{\boldsymbol{\theta}}^{*(b)}$, $b = 1, \dots, B$, $\overline{\boldsymbol{\theta}} = \frac{1}{B}\sum_{b=1}^{B}\hat{\boldsymbol{\theta}}^{*(b)}$.

4′. Calculate $\widehat{\text{var}}_B(\hat{\boldsymbol{\theta}}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\boldsymbol{\theta}}^{*(b)}-\overline{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}^{*(b)}-\overline{\boldsymbol{\theta}})'$.

3 The empirical best predictor

This section derives the best predictor (BP) and the empirical best predictor (EBP) of p_d under the area-level Poisson mixed model. The conditional distribution of $\boldsymbol{y} = (y_1, \dots, y_D)'$, given $\boldsymbol{v}$, is

$$P(\boldsymbol{y}|\boldsymbol{v}) = \prod_{d=1}^{D} P(y_d|v_d),$$

where

$$P(y_d|v_d) = \frac{\nu_d^{y_d}}{y_d!}e^{-\nu_d p_d}p_d^{y_d} = \frac{\nu_d^{y_d}}{y_d!}\exp\left\{y_d(\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d) - \nu_d\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_d\}\right\}.$$

The best predictor (BP) of p_d is the unbiased predictor minimizing the MSE. It is given by the conditional expectation $\hat{p}_d = \hat{p}_d(\boldsymbol{\theta}) = E_\theta[p_d|\boldsymbol{y}]$. In this case, we have that $E_\theta[p_d|\boldsymbol{y}] = E_\theta[p_d|y_d]$ and using Bayes's theorem, we get

$$E_\theta[p_d|y_d] = \frac{\int_R \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} P(y_d|v_d) f(v_d)\, dv_d}{\int_R P(y_d|v_d) f(v_d)\, dv_d} = \frac{N_d(y_d, \boldsymbol{\theta})}{D_d(y_d, \boldsymbol{\theta})} \triangleq \psi_d(y_d, \boldsymbol{\theta}),$$

where

$$\begin{aligned} N_d(y_d, \boldsymbol{\theta}) &= \int_R \exp\left\{(y_d+1)(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\right\}\right\} f(v_d)\, dv_d, \\ D_d(y_d, \boldsymbol{\theta}) &= \int_R \exp\left\{y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\right\} f(v_d)\, dv_d. \end{aligned}$$

The EBP of p_d is obtained by replacing the vector of unknown parameters $\boldsymbol{\theta}$ by a consistent estimator, $\hat{\boldsymbol{\theta}}$. Therefore, we can write the EBP as $\hat{p}_d = \hat{p}_d(\hat{\boldsymbol{\theta}}) = \psi_d(y_d, \hat{\boldsymbol{\theta}})$. We can approximate it by estimating the integrals with an accelerated Monte Carlo method based on the properties of the antithetic variables to reduce the variability. This algorithm can be expressed as follows.

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$ as in Section 2.

2. For $\ell = 1, \ldots, L$, generate $v_d^{(\ell)}$ i.i.d. $N(0,1)$ and calculate their antithetic variates $v_d^{(L+\ell)} = -v_d^{(\ell)}$.

3. Calculate the approximation of EBP as $\hat{p}_d(\hat{\boldsymbol{\theta}}) = \hat{N}_d/\hat{D}_d$, where the theoretical integrals are approximated by Monte Carlo, i.e.

$$\begin{aligned} \hat{N}_d &= \frac{1}{2L}\sum_{\ell=1}^{2L} \exp\left\{(y_d+1)(\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}) - \nu_d \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}\}\right\}, \\ \hat{D}_d &= \frac{1}{2L}\sum_{\ell=1}^{2L} \exp\left\{y_d(\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}) - \nu_d \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}\}\right\}. \quad (4) \end{aligned}$$

Remark 1. Since the size variable ν_d is known in practice, then the EBP of $\mu_d = \nu_d p_d$ is $\hat{\mu}_d(\hat{\boldsymbol{\theta}}) = \nu_d \hat{p}_d(\hat{\boldsymbol{\theta}})$. Further, we can consider the plug-in estimator $\hat{p}_d^P(\hat{\boldsymbol{\theta}}) = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}\hat{v}_d\}$. Section 5 studies $\hat{p}_d(\hat{\boldsymbol{\theta}})$ and $\hat{p}_d^P(\hat{\boldsymbol{\theta}})$ by calculating empirical biases and MSEs.

Remark 2. As the MM Newton-Raphson algorithm does not give a prediction of v_d, we use its EBP. The BP of v_d is

$$\hat{v}_d(\boldsymbol{\theta}) = E_\theta[v_d|y_d] = \frac{\int_R v_d P(y_d|v_d) f(v_d)\, dv_d}{\int_R P(y_d|v_d) f(v_d)\, dv_d} = \frac{N_{v,d}(y_d, \boldsymbol{\theta})}{D_d(y_d, \boldsymbol{\theta})},$$

where

$$N_{v,d}(y_d, \boldsymbol{\theta}) = \int_R v_d \exp\left\{y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\right\}\right\} f(v_d)\, dv_d.$$

The EBP of v_d is $\hat{v}_d = \hat{v}_d(\hat{\boldsymbol{\theta}})$ and it can be approximated using an accelerated Monte Carlo algorithm analogous to the previous case. The steps are:

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$.
2. For $\ell = 1, \ldots, L$, generate $v_d^{(\ell)}$ i.i.d. $N(0,1)$ and calculate their antithetic variates $v_d^{(L+\ell)} = -v_d^{(\ell)}$.
3. Calculate $\hat{v}_d(\hat{\boldsymbol{\theta}}) = \hat{N}_{v,d}/\hat{D}_d$, where $\hat{D}_d$ is defined in (4) and

$$\hat{N}_{v,d} = \frac{1}{2L}\sum_{\ell=1}^{2L} v_d^{(\ell)} \exp\left\{ y_d(\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}) - \nu_d \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}\} \right\}.$$

4 The MSE of the EBP

Appendix A.2 gives the mathematical derivations for decomposing the MSE of the EBP of $p_d = p_d(\boldsymbol{\theta}, v_d) = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}$. The MSE of the EBP is

$$MSE(\hat{p}_d) = g_d(\boldsymbol{\theta}) + \frac{1}{D}\, c_d(\boldsymbol{\theta}) + o(1/D), \tag{5}$$

where

$$c_d(\boldsymbol{\theta}) = \sum_{j=0}^{\infty} \left(\frac{\partial}{\partial \boldsymbol{\theta}}\, \psi_d(j,\boldsymbol{\theta})\right)' \boldsymbol{V}(\boldsymbol{\theta}) \left(\frac{\partial}{\partial \boldsymbol{\theta}}\, \psi_d(j,\boldsymbol{\theta})\right) p_d(j,\boldsymbol{\theta}),$$
$$\boldsymbol{V}(\boldsymbol{\theta}) = DE\left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\right].$$

A plug-in estimator of (5) is obtained replacing $\boldsymbol{\theta}$ by a consistent estimator $\hat{\boldsymbol{\theta}}$, namely

$$\widehat{MSE}^{P}(\hat{p}_d) = g_d(\hat{\boldsymbol{\theta}}) + \frac{1}{D}\, c_d(\hat{\boldsymbol{\theta}}). \tag{6}$$

By a Taylor expansion of $c_d(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ and the consistency of $\hat{\boldsymbol{\theta}}$, we have that $E[c_d(\hat{\boldsymbol{\theta}}) - c_d(\boldsymbol{\theta})] = o(1)$. However $E[g_d(\hat{\boldsymbol{\theta}}) - g_d(\boldsymbol{\theta})]$ is not of order $o(D^{-1})$. Let $\hat{\boldsymbol{\theta}}$ be a truncated MM estimator. This is to say

$$\hat{\beta}_k = \begin{cases} -L_D & \text{if } \tilde{\beta}_k < -L_D, \\ \tilde{\beta}_k & \text{if } -L_D < \tilde{\beta}_k < L_D, \\ L_D & \text{if } \tilde{\beta}_k > L_D, \end{cases} \qquad \hat{\sigma}^2 = \begin{cases} \tilde{\sigma}^2 & \text{if } \tilde{\sigma}^2 \leq L_D, \\ L_D & \text{if } \tilde{\sigma}^2 > L_D, \end{cases}$$

where $\tilde{\boldsymbol{\theta}}$ is an MM estimator. Under the assumed regularity conditions (23)-(25) of Jiang (2003), $E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = O(D^{-1})$ holds for the truncated MM estimator and $E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] = O(D^{-1})$ holds for the MM estimator. By the Taylor expansion, we have

$$g_d(\hat{\boldsymbol{\theta}}) = g_d(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\right)' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2),$$

and hence

$$E[g_d(\hat{\boldsymbol{\theta}})] = g_d(\boldsymbol{\theta}) + \frac{1}{D} b_d(\boldsymbol{\theta}) + o(D^{-1}),$$

where

$$b_d(\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\Big)' D E[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] + \frac{1}{2} E\left[D(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \Big(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\Big)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]. \quad (7)$$

Proposition 4.1 gives an approximation to the bias term b_d when $\hat{\boldsymbol{\theta}}$ is the truncated MM estimator.

Proposition 4.1. Let $\hat{\boldsymbol{\theta}}$ be the truncated MM estimator. Under regularity conditions (23)-(25) of Jiang and Lahiri (2001), it holds that $b_d(\boldsymbol{\theta}) = B_d(\boldsymbol{\theta}) + o(1)$, where

$$\begin{aligned}
B_d(\boldsymbol{\theta}) &= \frac{1}{2}\left\{E[r_{D,d}] - \Big(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\Big)' \Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1} E[\boldsymbol{q}_D]\right\}, \\
r_{D,d} &= \Delta_D' \boldsymbol{R}_d(\boldsymbol{\theta}) \Delta_D, \boldsymbol{R}_d(\boldsymbol{\theta}) = \left(\Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1}\right)' \Big(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\Big)\Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1}, \\
\boldsymbol{q}_D &= \underset{1 \le k \le p+1}{\text{col}}(q_{Dk}), \;\; \boldsymbol{M}(\boldsymbol{\theta}) = \underset{1 \le k \le p+1}{\text{col}}(M_k(\boldsymbol{\theta})), \;\; \hat{\boldsymbol{M}} = \underset{1 \le k \le p+1}{\text{col}}(\hat{M}_k), \\
q_{Dk} &= \Delta_D' \boldsymbol{Q}(\boldsymbol{\theta}) \Delta_D, \, \boldsymbol{Q}(\boldsymbol{\theta}) = \left(\Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1}\right)' \Big(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} M_k(\boldsymbol{\theta})\Big)\Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1}, \\
\Delta_D &= \sqrt{D}(\hat{\boldsymbol{M}} - \boldsymbol{M}(\boldsymbol{\theta})), \;\; \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial \boldsymbol{\theta}_{k_2}} \boldsymbol{M}_{k_1}(\boldsymbol{\theta})\Big)_{k_1, k_2 = 1, \dots, p+1}.
\end{aligned}$$

Appendix A.3 gives the proof of Proposition 4.1 and Appendix A.4 gives the computationally efficient formulas (A.3)-(A.5) for calculating the partial derivatives of $g_d(\boldsymbol{\theta})$. These formulas are taken from Lahiri et al. (2007).

The following parametric bootstrap algorithm estimates the bias correction term $B_d(\boldsymbol{\theta})$.

1. Fit the model to the sample and calculate $\hat{\boldsymbol{\theta}}$, $\boldsymbol{R}_d(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{Q}(\hat{\boldsymbol{\theta}})$.
2. Generate bootstrap samples $\{y_d^{*(b)} : d = 1, \dots, D\}$, $b = 1, \dots, B$, from the fitted model.
3. For each bootstrap sample b, calculate $\Delta_D^{*(b)} = \sqrt{D}(\hat{\boldsymbol{M}}^{*(b)} - \boldsymbol{M}(\hat{\boldsymbol{\theta}}))$, where $\hat{\boldsymbol{M}}^{*(b)} = \underset{1 \le k \le p+1}{\text{col}}(\hat{M}_k^{*(b)})$, $\hat{M}_k^{*(b)} = \sum_{d=1}^D y_d^{*(b)} x_{dk}$, $k = 1, \dots, p$, $\hat{M}_{p+1}^{*(b)} = \sum_{d=1}^D y_d^{*(b)2}$, and calculate

   $$r_{D,d}^{*(b)} = \Delta_D^{*(b)\prime} \boldsymbol{R}_d(\hat{\boldsymbol{\theta}}) \Delta_D^{*(b)}, \;\; q_{Dk}^{*(b)} = \Delta_D^{*(b)\prime} \boldsymbol{Q}(\hat{\boldsymbol{\theta}}) \Delta_D^{*(b)}, \;\; \boldsymbol{q}_D = \underset{1 \le k \le p+1}{\text{col}}(q_{Dk}^{*(b)}).$$

4. Calculate $\hat{E}_B[r_{D,d}] = \frac{1}{B} \sum_{b=1}^B r_{D,d}^{*(b)}$, $\hat{E}_B[q_D] = \frac{1}{B} \sum_{b=1}^B q_{Dk}^{*(b)}$.

5. Calculate $\hat{B}_d(\hat{\boldsymbol{\theta}}) = \frac{1}{2}\left\{\hat{E}_B[r_{D,d}] - \left(\frac{\partial}{\partial\boldsymbol{\theta}}\,\hat{g}_d(\hat{\boldsymbol{\theta}})\right)'\left(\frac{\partial}{\partial\boldsymbol{\theta}}\,\boldsymbol{M}(\hat{\boldsymbol{\theta}})\right)^{-1}\hat{E}_B[\boldsymbol{q}_D]\right\}$.

An order $o(D^{-1})$ theoretical estimator of $MSE(\hat{p}_d)$, with bias correction, is

$$\widehat{MSE}(\hat{p}_d) = \widehat{MSE}^P(\hat{p}_d) - \frac{1}{D}\,B_d(\boldsymbol{\theta}),$$

and the practical estimators, with and without bias correction, are

$$mse(\hat{p}_d) = mse^P(\hat{p}_d) - \frac{1}{D}\,\hat{B}_d(\hat{\boldsymbol{\theta}}) \quad \text{and} \quad mse^P(\hat{p}_d) = \hat{g}_d(\hat{\boldsymbol{\theta}}) + \frac{1}{D}\,\hat{c}_d(\hat{\boldsymbol{\theta}}). \qquad (8)$$

where $\hat{g}_d(\hat{\boldsymbol{\theta}})$ and $\hat{c}_d(\hat{\boldsymbol{\theta}})$ are the Monte Carlo approximations of $g_d(\hat{\boldsymbol{\theta}})$ and $c_d(\hat{\boldsymbol{\theta}})$ respectively.

The calculation of $mse(\hat{p}_d)$ is computationally expensive. An alternative MSE estimator can be introduced by applying the following parametric bootstrap approach.

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$.
2. Repeat B times $(b = 1, \dots, B)$
   (a) Generate $v_d^{*(b)} \sim N(0,1)$, $d = 1, \dots, D$. Calculate $p_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}v_d^{*(b)}\}$ and $y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)})$.
   (b) For each bootstrap sample, calculate the estimator $\hat{\boldsymbol{\theta}}^{*(b)}$ and the EBP $\hat{p}_d^{*(b)} = \hat{p}_d^*(\hat{\boldsymbol{\theta}}^{*(b)})$.
3. Calculate
$$mse^*(\hat{p}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{p}_d^{*(b)} - p_d^{*(b)}\right)^2. \qquad (9)$$

5 Simulation experiments

5.1 Model-based simulation

This subsection presents three simulation experiments based on the application to the real data from SLCS2008 (see more details in Section 6). First, we analyze the behavior of the MM and PQL fitting algorithms. Second, we compare the performances of the EBP and the plug-in estimators. Third, we empirically study the proposed MSE estimators. In the three simulation experiments, we use the same explanatory variables as those used in the case study: unemployed (*lab2*), foreign people (*cit2*), people with age in $50-64$ (*age4*) and secondary or university education completed (*edu23*) proportions.

Random effects v_d are generated from normal and Gumbel distributions with mean zero and variance one. We use the Gumbel distribution to study how the lack of normality in the random effects affects the model parameter and EBP estimates. The response variable is $y_d \sim \text{Poiss}(\nu_d p_d)$, where $p_d = \exp\{\beta_0 + lab2_d\beta_1 + cit2_d\beta_2 + age4_d\beta_3 + edu23_d\beta_4 + \phi v_d\}$, $d = 1, \ldots, D$. The model parameters, $\beta_0, \ldots, \beta_4, \phi$, and the sizes $\nu_d = n_d$ are taken from the application to real data presented in Section 6. The numbers of domains are $D = 52, 104, 150$. The x-variables are taken from provinces crossed by female if $D = 52$ and from the provinces crossed by sex if $D = 104$. In the case $D = 150$, as the data file of x-values have 104 records, we input 46 new records by doing a simple random sampling without replacement in the data file. We run the simulation experiments with $K = 1000$ Monte Carlo iterations.

For the six model parameters, $\theta \in \{\beta_0, \ldots \beta_4, \phi\}$, Table 1 presents the relative bias (RBIAS) and the relative root-MSE (RRMSE) in brackets for MM and PQL estimators, i.e.

$$RBIAS = \frac{\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta)}{|\theta|}, \;\; RRMSE = \frac{\sqrt{\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta)^2}}{|\theta|}.$$

This table suggest that RRMSE is slightly higher for Gumbel random effects. Both estimation methods (MM and PQL) behaves similarly with respect to RBIAS. PQL estimates has less variability but the MM estimator has lower bias for estimating the variance parameter. As expected, when the number of domains increases then the bias and the MSE decreases. The empirical results agree with the consistency property of the MM estimators.

The second simulation studies the behavior of the EBP and two plug-in estimators of p_d: the first one (PLUG1) uses PQL (see Saei and Chambers (2003) for more details) while the second one (PLUG2) uses the method of moments as fitting algorithm. For approximating the EBP of p_d, we generate $L = 2500$ independent random variables with $N(0, 1)$ distribution and we apply the step 2 of the EBP algorithm given in Section 3. Table 2 for normal and Table 3 for Gumbel random effects compare these estimators through the bias and the MSE (in brackets), i.e.

$$B_d = \frac{1}{K}\sum_{k=1}^{K}(\hat{p}_d^{(k)} - p_d^{(k)}), \;\; E_d = \frac{1}{K}\sum_{k=1}^{K}(\hat{p}_d^{(k)} - p_d^{(k)})^2, \quad d = 1, \ldots, D.$$

In both cases, results are presented for the quintiles of the set $\{1, \ldots, D\}$, where the domains are sorted by sample sizes. The last row of each subtable, $D = 52, 104, 150$, contains the average absolute biases and the average MSEs (in brackets), i.e.

$$B = \frac{1}{D}\sum_{d=1}^{D}|B_d|, \;\; E = \frac{1}{D}\sum_{d=1}^{D}E_d.$$

These tables suggest that plug-in estimator PLUG1 has the best performance in the simulation experiment and that PLUG2 and EBP behave similarly. We also observe that E_d's of the EBP are close to PLUG2. If we move from normal to

Table 1: RBIAS and RRMSE (in brackets) for MM and PQL estimators, taking normal (N) and gumbel (G) random effects.

		N		G	
D	$\hat{\theta}$	MM	PQL	MM	PQL
52	$\hat{\beta}_0$	0.0181 (0.3960)	0.0203 (0.3665)	-0.0298 (0.4393)	-0.0083 (0.3707)
	$\hat{\beta}_1$	0.0001 (0.3527)	0.0001 (0.3233)	0.0165 (0.3860)	0.0056 (0.3300)
	$\hat{\beta}_2$	-0.0071 (0.2394)	-0.0089 (0.2241)	0.0016 (0.2441)	0.0026 (0.2212)
	$\hat{\beta}_3$	-0.0108 (0.4518)	-0.0033 (0.4054)	0.0139 (0.5033)	0.0093 (0.4190)
	$\hat{\beta}_4$	-0.0063 (0.2281)	-0.0088 (0.2122)	0.0105 (0.2432)	0.0030 (0.2117)
	$\hat{\phi}$	-0.1982 (0.6425)	-0.4464 (0.4764)	-0.1395 (0.7039)	-0.3963 (0.4530)
104	$\hat{\beta}_0$	-0.0021 (0.2714)	0.0101 (0.2474)	-0.0023 (0.2936)	0.0132 (0.2623)
	$\hat{\beta}_1$	-0.0077 (0.2290)	-0.0150 (0.2118)	0.0001 (0.2410)	-0.0117 (0.2171)
	$\hat{\beta}_2$	0.0062 (0.1582)	0.0063 (0.1434)	-0.0026 (0.1647)	-0.0031 (0.1440)
	$\hat{\beta}_3$	0.0102 (0.2898)	0.0084 (0.2602)	-0.0025 (0.3056)	-0.0057 (0.2666)
	$\hat{\beta}_4$	-0.0034 (0.1308)	-0.0063 (0.1203)	0.0025 (0.1474)	-0.0002 (0.1306)
	$\hat{\phi}$	-0.1708 (0.5500)	-0.4300 (0.4479)	-0.1347 (0.6240)	-0.3931 (0.4244)
150	$\hat{\beta}_0$	0.0060 (0.2261)	0.0150 (0.2094)	0.0019 (0.2378)	0.0118 (0.2083)
	$\hat{\beta}_1$	-0.0035 (0.1833)	-0.0104 (0.1712)	-0.0051 (0.1929)	-0.0108 (0.1712)
	$\hat{\beta}_2$	-0.0012 (0.1333)	-0.0029 (0.1221)	-0.0015 (0.1458)	-0.0015 (0.1280)
	$\hat{\beta}_3$	-0.0022 (0.2418)	-0.0027 (0.2186)	-0.0022 (0.2629)	-0.0003 (0.2274)
	$\hat{\beta}_4$	-0.0023 (0.1116)	-0.0032 (0.1021)	-0.0001 (0.1242)	-0.0029 (0.1101)
	$\hat{\phi}$	-0.1404 (0.4901)	-0.4084 (0.4207)	-0.0779 (0.5507)	-0.3643 (0.3883)

Table 2: B_d and E_d (in brackets) for the estimators of p_d using normal random effects.

D	d	p_d	PLUG1	PLUG2	EBP
52	12	0.1358	-0.0007 (0.0005)	-0.0021 (0.0006)	-0.0013 (0.0006)
	22	0.2199	-0.0008 (0.0010)	-0.0049 (0.0013)	-0.0040 (0.0013)
	32	0.1473	-0.0001 (0.0004)	-0.0008 (0.0005)	-0.0003 (0.0005)
	42	0.1390	-0.0012 (0.0003)	-0.0019 (0.0004)	-0.0015 (0.0004)
$B\,(E)$			0.0010 (0.0011)	0.0030 (0.0014)	0.0023 (0.0014)
104	22	0.2043	-0.0005 (0.0011)	-0.0022 (0.0011)	-0.0010 (0.0011)
	43	0.2902	-0.0026 (0.0013)	-0.0041 (0.0016)	-0.0030 (0.0016)
	63	0.3341	-0.0012 (0.0012)	-0.0003 (0.0017)	0.0006 (0.0017)
	84	0.1346	-0.0002 (0.0003)	-0.0010 (0.0003)	-0.0006 (0.0003)
$B\,(E)$			0.0008 (0.0009)	0.0021 (0.0010)	0.0014 (0.0010)
150	31	0.2980	-0.0003 (0.0018)	-0.0007 (0.0021)	0.0007 (0.0021)
	61	0.2883	-0.0010 (0.0013)	-0.0012 (0.0016)	-0.0001 (0.0016)
	91	0.1183	-0.0014 (0.0003)	-0.0023 (0.0003)	-0.0017 (0.0003)
	121	0.1364	-0.0011 (0.0002)	-0.0022 (0.0003)	-0.0018 (0.0003)
$B\,(E)$			0.0009 (0.0009)	0.0021 (0.0010)	0.0014 (0.0010)

Gumbel distribution we get a moderate increase of MSE for the three consider estimators of p_d.

The third simulation investigates the behavior of the MSE estimators of the EBP. This simulation requires, as input, very accurate empirical approximations of the variance-covariance matrix of the MM estimator $\hat{\boldsymbol{\theta}}$ and of the true MSE, E_d, of $\hat{p}_d$. We do these calculations in advance by running a Monte Carlo experiment with 10^4 iterations.

Three estimators of the MSE are compared. They are the two plug-in estimators given in (8), mse^P and mse without and with bias correction respectively, and the parametric bootstrap estimator, mse^*, introduced in (9). The calculation of mse^P and mse is computationally intensive and requires Monte Carlo approximations. We generate $L = 2500$ independent random variables with distribution $N(0,1)$ for approximating $\hat{g}_d(\hat{\boldsymbol{\theta}})$ and $\hat{c}_d(\hat{\boldsymbol{\theta}})$. Furthermore, we approximate the infinite sums appearing in the definitions of these two terms by the corresponding finite sums with the first 300 summands. In this way, we guarantee an approximation of the infinite sum with an error lower than the precision of the computer.

Figure 1 plots the MSE estimators for each domain $d = 1, \ldots, D$ and for $D = 52$ (left), $D = 104$ (right) and $D = 150$ (bottom). They are sorted by sample size. The results for small values of d are quite similar. However, the bootstrap estimator shows a more stable behavior when d increases. We note that the estimator with bias correction, mse, is a good alternative despite not

Table 3: B_d and E_d (in brackets) for the estimators of p_d using Gumbel random effects.

D	d	p_d	PLUG1	PLUG2	EBP
52	12	0.1370	-0.0010 (0.0006)	-0.0024 (0.0007)	-0.0016 (0.0007)
	22	0.2217	-0.0031 (0.0011)	-0.0062 (0.0016)	-0.0054 (0.0016)
	32	0.1488	-0.0009 (0.0004)	-0.0018 (0.0006)	-0.0013 (0.0006)
	42	0.1382	0.0006 (0.0003)	-0.0009 (0.0005)	-0.0005 (0.0005)
$B\,(E)$			0.0009 (0.0011)	0.0029 (0.0016)	0.0022 (0.0016)
104	22	0.2046	-0.0003 (0.0010)	-0.0028 (0.0012)	-0.0016 (0.0012)
	43	0.2912	-0.0022 (0.0016)	-0.0044 (0.0022)	-0.0032 (0.0022)
	63	0.3331	-0.0014 (0.0013)	-0.0017 (0.0018)	-0.0009 (0.0018)
	84	0.1358	-0.0003 (0.0003)	-0.0013 (0.0005)	-0.0008 (0.0005)
$B\,(E)$			0.0009 (0.0010)	0.0027 (0.0013)	0.0018 (0.0012)
150	31	0.2990	-0.0007 (0.0020)	-0.0014 (0.0025)	0.0001 (0.0025)
	61	0.2888	-0.0021 (0.0014)	-0.0029 (0.0017)	-0.0017 (0.0017)
	91	0.1172	-0.0009 (0.0003)	-0.0019 (0.0003)	-0.0013 (0.0003)
	121	0.1347	-0.0001 (0.0003)	-0.0014 (0.0004)	-0.0010 (0.0004)
$B\,(E)$			0.0010 (0.0009)	0.0028 (0.0012)	0.0019 (0.0012)

being able to capture the bias of the plug-in estimator in the last domains. For the bootstrap approach, we consider $B = 500$ resamples.

Figure 2 prints the boxplots of the biases B_d, $d = 1, \ldots, D$, of the three MSE estimators for $D = 52$ (left), $D = 104$ (center) and $D = 150$ (right). The MSE estimators are the two plug-in estimators mse_d and mse_d^P (with and without bias correction, respectively) and the parametric bootstrap estimator mse_d^*. We observe that all MSE estimators under-estimate the true MSE, specially the bootstrap estimator. On the other hand, bootstrap estimates are more stable because they do not contain many outliers.

Table 4 presents the bias and mean squared error ($\times 10^5$) of the three considered estimators of MSE for quintiles of $\{1, \ldots, D\}$. Analytic estimators (without and with bias correction term) perform well in both bias and mean squared error. The bootstrap MSE estimator has a similar mean squared error to the analytic ones, it has a higher bias, it is computationally faster and it is easy to implement.

5.2 Design-based simulation

Model-based simulations depend on the model used for data generation. However, in practice, we do not know what is the model that generates the population. The target of this simulation experiment is to analyse if the proposed estimator under a Poisson mixed model performs well even if the population under study is not Poisson distributed. For this sake, we generate a population based on the real data by using the sampling weights w_j. The artificial

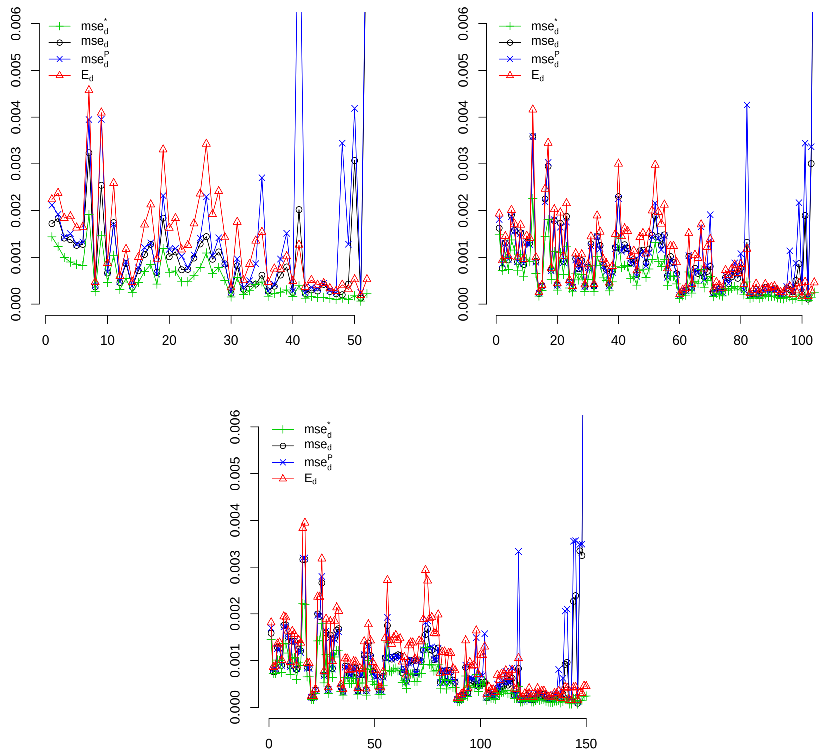

Figure 1: MSE estimators for $D = 52$ (left), $D = 104$ (right) and $D = 150$ (bottom).

population is built by repeating $\lfloor 10^{-3} w_j \rfloor$ times each sampling unit j.

We implement a simplified version of the SLCS2008 sampling design. Within each autonomous community, the units are selected with a simple random sampling design. As sample size for each autonomous community, we take $n_c = \lfloor N_c 10^{-1} \rfloor + 1$, where N_c denotes the population size of each autonomous community. For each drawn sample k ($k = 1, \ldots, K = 1000$), we evaluate the direct estimator (Dir), the EBLUP based on the Fay-Herriot model (FH), the two considered plug-in estimators (PLUG1 and PLUG2) and the EBP. For the direct estimators of p_d and its design-based variance we take

$$\hat{p}_d^{dir} = \frac{1}{\hat{N}_d} \sum_{j \in s_d} w_{dj}\, y_{dj},\ \widehat{\text{var}}_\pi(\hat{p}_d^{dir}) = \frac{1}{\hat{N}_d^2} \sum_{j \in s_d} w_{dj}(w_{dj} - 1) \left(y_{dj} - \hat{p}_d^{dir} \right)^2, \quad (10)$$

where $w_{dj} = N_c/n_c$ and $\hat{N}_d = \sum_{j \in s_d} w_{dj} = n_d \frac{N_c}{n_c}$. The variance estimator is taken from Särndal et al. (1992), pp. 43, 185 and 391, with the simplifications $w_{dj} = 1/\pi_{dj}$, $\pi_{dj,dj} = \pi_{dj}$ and $\pi_{di,dj} = \pi_{di}\pi_{dj}$, $i \neq j$ in the second order inclusion probabilities. The EBLUP of p_d is taken from Fay and Herriot (1979) or Prasad

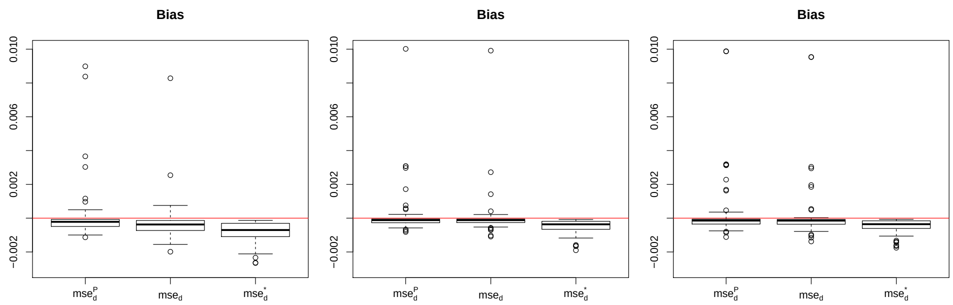

Figure 2: Bias of MSE estimators for $D = 52$ (left), $D = 104$ (center) and $D = 150$ (right).

Table 4: Bias and mean squared error in brackets ($\times 10^5$) of the MSE estimators.

D	d	E_d	mse^P	mse	mse^*
52	12	0.0006	-10.5209 (0.0108)	-11.1735 (0.0204)	-29.0402 (0.0164)
	22	0.0012	-12.9678 (0.0091)	-37.8213 (0.0633)	-68.9004 (0.0631)
	32	0.0005	-15.5940 (0.0050)	-16.2415 (0.0129)	-30.5855 (0.0120)
	42	0.0004	-11.2966 (0.0033)	-9.9097 (0.0096)	-23.9860 (0.0067)
104	22	0.0010	-16.4116 (0.0201)	-13.2513 (0.0302)	-39.7102 (0.0312)
	43	0.0016	-42.7149 (0.0382)	-38.3071 (0.0633)	-75.9042 (0.0764)
	63	0.0015	-48.9221 (0.0461)	-43.9790 (0.0885)	-82.7329 (0.0800)
	84	0.0003	-8.5716 (0.0016)	-7.1582 (0.0029)	-15.3339 (0.0029)
150	31	0.0019	-41.9215 (0.0653)	-38.7860 (0.0941)	-77.3717 (0.1029)
	61	0.0014	-38.3529 (0.0409)	-32.0055 (0.0615)	-66.4599 (0.0653)
	91	0.0003	-5.2724 (0.0016)	-4.5544 (0.0024)	-11.0868 (0.0024)
	121	0.0003	-10.2886 (0.0021)	-9.3420 (0.0031)	-15.0528 (0.0030)

and Rao (1990).

Table 5 gives the results of the bias B_d and the MSE E_d (in brackets) for the the direct estimator (Dir), the EBLUP based on the Fay-Herriot model (FH), the two considered plug-in estimators (PLUG1 and PLUG2) and the EBP. The results are presented for the quintiles to real population, where $D = 104$. As expected, the direct estimator has lower bias but its MSE is higher than the model-based-estimators.

The FH has lower bias and greater MSE than the PLUG1 predictor in most cases. The three Poisson mixed model predictors (EBP, PLG1 and PLUG2) have a similar behavior, in both bias and mean squared error. If we compare these results with those obtained in Table 2 under model-based simulation, they increase slightly. This fact is somehow expected but gives more realistic information about the behavior of the considered predictors in practice.

Table 5: B_d and E_d (in brackets) for the estimators of p_d.

D	d	p_d	Dir	FH	PLUG1	PLUG2	EBP
104	22	0.3632	-0.0036	-0.0589	-0.0622	-0.0817	-0.0817
			(.0099)	(.0054)	(.0044)	(.0072)	(.0071)
	43	0.1657	-0.0013	-0.0051	0.0195	0.0068	0.0068
			(.0040)	(.0017)	(.0005)	(.0002)	(.0002)
	63	0.1010	-0.0016	0.0005	0.0230	0.0184	0.0185
			(.0015)	(.0009)	(.0006)	(.0004)	(.0004)
	84	0.2182	0.0007	0.0103	0.0181	-0.0008	-0.0007
			(.0018)	(.0011)	(.0005)	(.0002)	(.0002)
B			0.0019	0.0264	0.0498	0.0500	0.0500
(E)			(.0066)	(.0048)	(.0048)	(.0051)	(.0051)

All these simulation experiments have been carried out using the statistical software R 3.1.1. We use *nleqslv* package to solve the system of nonlinear equations (1)-(2) by Newton-Raphson and *evd* package to generate random effects according to a Gumbel distribution.

6 Application to real data

Policy makers are interested in finding out which factors are more influential for poverty in order to act on them and achieve a decrease of their consequences, especially in poor regions where a greater commitment to the competent authorities is necessary.

This section estimates the poverty rate, p_d, in 2008 by domains (provinces crossed by sex) using the data from the SLCS. The number of domains is $D = 104$. At the unit level, the target variable is dichotomic and takes the values $y_{dj} = 1$ if individual j of domain d is under the poverty line and $y_{dj} = 0$ otherwise. The domains are the provinces crossed by sex. The domain sample sizes and totals are n_d and $y_d = \sum_{j\in s_d} y_{dj}$ respectively. As y_d counts the number of people under the poverty line in the domain sample s_d, we assume that y_d can be described by an area-level Poisson mixed model with size parameters n_d, $d = 1,\dots,D$, and some explanatory variables. The available auxiliary variables are the domain proportions of people in the categories of the following classification variables.

- Age: ≤ 15 ($age1$), $16-24$ ($age2$), $25-49$ ($age3$), $50-64$ ($age4$) and ≥ 65 ($age5$).
- Education: less than primary ($edu0$), primary ($edu1$), secondary ($edu2$), university ($edu3$).
- Citizenship: Spanish ($cit1$), not Spanish ($cit2$).

- Labor situation: ≤ 15 ($lab0$), employed ($lab1$), unemployed ($lab2$), inactive ($lab3$).

As the proportions of people in the categories of a classification variables sum up to one, we take the reference categories out of the data file of auxiliary variables. The reference categories are $age1$, $edu0$, $cit1$ and $lab0$. Regarding the level of education, we note that people that have passed the national programme of professional training courses typically have good job opportunities at the industry and services labor sector. As these people are in group $edu2$, we merge secondary and university education levels into a single category $edu23$. This proposal was suggested by a Spanish Office of Statistics.

An area-level Poisson mixed model is fitted to data. The MM Newton-Raphson algorithm is employed for estimating the model parameters and their asymptotic variances. A subset of significant auxiliary variables is selected, i.e. with p-value lower than 0.05. Table 6 presents the estimates of the regression parameters and their standard errors, z-values and p-values. Each domain (province-sex) d, $d = 1, \ldots, 104$, has a random intercept with distribution $N(0, \phi^2)$. The estimate of ϕ is $\hat{\phi} = 0.183$.

Table 6: MM estimates of regression parameters.

Coefficient	Estimate	Std. Error	z-value	P(> |z|)
$Intercept$	1.5669	0.5030	3.7653	0.0002
$lab2$	6.8923	1.8939	2.9949	0.0027
$cit2$	-2.9844	0.5860	-4.9693	0.0000
$age4$	-7.5259	2.6311	-3.8857	0.0001
$edu23$	-3.5998	0.5913	-5.3807	0.0000

The signs of the regression parameters in Table 6 show that unemployment ($lab2$) contributes to increase the poverty since its sign is positive, while the remaining covariates are protective in the sense that an increase in them causes a reduction in the number of people below poverty line, assuming that the other auxiliary variables are fixed. The sign of $cit2$ appears because the foreign people tend to establish in provinces with higher economical activity, given that they can find better living conditions and job opportunities. Esteban et al. (2012) found the same result when fitting Fay-Herriot temporal models to data from the Spanish Living Conditions survey of 2006.

For the sake of comparison, we also fit a Poisson regression model with the same auxiliary variables of Table 6 but without any random effect. Figure 3 plots the Pearson residuals of the Poisson regression models without (left) and with (right) domain random effects. In both cases the behavior is symmetrical around 0 and a clear improvement is observed when we use the more complex model including random effects, as they capture the variability between domains.

The objective of this work is to study the EBP. We are also interested in comparing the EBP of p_d with the direct estimator and the EBLUP based on

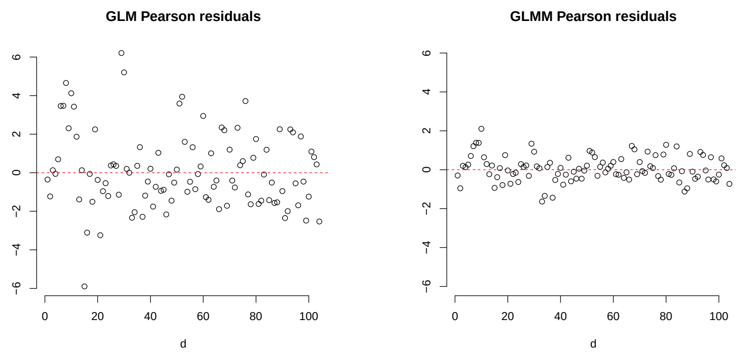

Figure 3: Pearson residuals for the fixed effects model (left) and mixed effects model (right).

a Fay-Herriot model (Fay and Herriot 1979) fitted by the REML method to the set of auxiliary variables described in Table 6. The MSE of the EBP is estimated by parametric bootstrap and the MSE of the EBLUP by the g_1-g_3 formula given by Datta and Lahiri (2000).

Direct estimators of p_d and of its design-based variance are calculated following (10), where $\hat{N}_d = \sum_{j \in s_d} w_{dj}$ and the w_{dj}'s are the official calibrated SLCS sampling weights which take into account for non response.

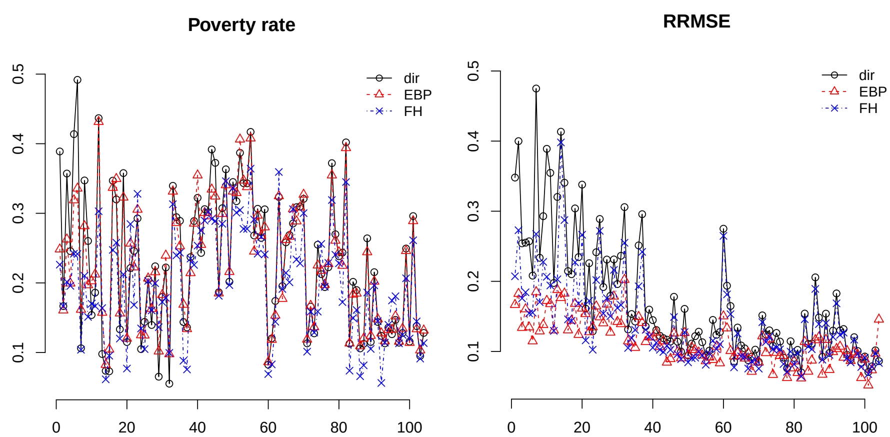

Figure 4: Direct and EBP estimates of p_d (left) and relative root-MSE (right) for both estimators.

Figure 4 (left) plots the EBP, direct and EBLUP estimates of p_d, $d = 1, \ldots, D$. We note that all estimates follow the same patterns. Figure 4 (right) plots the relative squared-root MSE (RRMSE) estimates of the EBPs (EBP) and of the EBLUPs (FH). It also plots the relative squared-root design-based variance (RRvar) estimates of the direct estimators (dir). The domains are sorted by sample size. Figure 4 shows that the RRMSEs of the EBPs are in most domains smaller than the RRvars of the direct estimators and than the RRMSEs of the EBLUPs. The performances of the RRMSE of the EBLUP and of the RRvar of the direct estimator are similar. We observe a greater accuracy when the sample size increases. We are cautious in claiming that the EBP has better performance than the Fay-Herriot EBLUP as the estimated MSEs are derived under the assumption that the model is correct and they are not comparable. Nevertheless, we conclude that the Poisson mixed-model EBP is a good alternative for estimating p_d.

Table 7 presents the estimates of p_d using the direct, EBLUP and EBP estimators, and their corresponding errors: the MSE of the EBLUP and EBP (E_d^{eblup} and E_d^{ebp}) and the design-based variance (E_d^{dir}) of the direct estimator. Due to limited space, we only show the results for women. Further, we order the results by sample size and we show the results for the minimum, maximum and sixtiles of ν_d. For small sample sizes the EBP estimates have a minor error and when they increase, both estimates of p_d and their corresponding errors show a similar behaviour. The displayed results are in accordance with those shown in Figure 4.

Table 7: Direct (p_d^{dir}), Fay-Herriot EBLUP (p_d^{eblup}) and EBP (p_d^{ebp}) estimates of p_d for women and MSE estimates.

Sex	ν_d	p_d^{dir}	p_d^{eblup}	p_d^{ebp}	E_d^{dir}	E_d^{eblup}	E_d^{ebp}
Women	18	0.5303	0.2262	0.2483	0.0341	0.0021	0.0017
	124	0.1355	0.1345	0.1249	0.0011	0.0007	0.0004
	162	0.3484	0.3131	0.3314	0.0021	0.0010	0.0014
	247	0.3976	0.3641	0.4078	0.0014	0.0009	0.0012
	424	0.2996	0.3002	0.3269	0.0007	0.0005	0.0008
	501	0.1759	0.1724	0.2248	0.0003	0.0003	0.0002
	1491	0.1122	0.1135	0.1317	0.0001	0.0001	0.0004

Figure 5 (left) maps the EBP estimates of p_d for women. We observe that highest levels of poverty are found in the south and center-west of the country. On the other hand, the northeastern provinces offer better living conditions. Figure 5 (right) maps the bootstrap relative root-MSE estimates of the EBP of p_d for women with $B = 1000$ resamples. In general, the estimation error is low. The number of provinces where the estimated RRMSE is greater than 15% is seven. The maximum value of the estimated RRMSE is 20.24%, which is achieved in a province with very low level of poverty. In general, the model-based estimators smooth the behavior of the direct estimators, but they could be in troubles for estimating the lowest or the highest poverty rates.

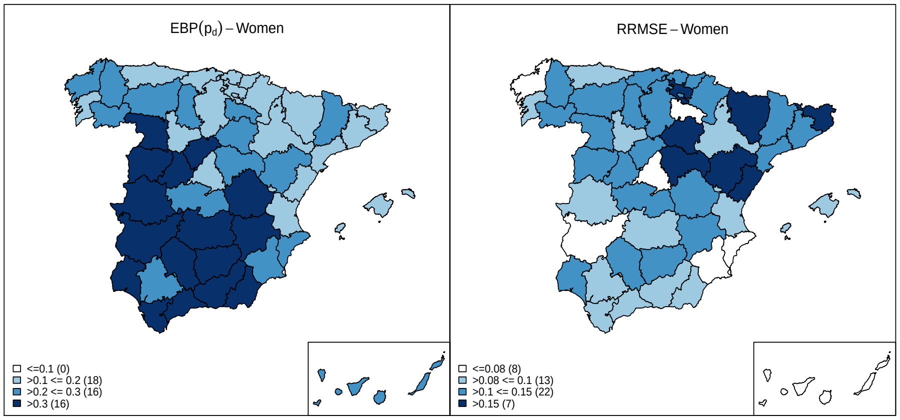

Figure 5: Poverty rate EBPs for women (left) and RRMSE (right) in 2008.

The Moran test was applied to the residuals of the EBP of p_d to study a possible spatial correlation. We use *Moran.I* function in *ape* package of R. The matrix of weights was calculated by using the Euclidean distance between the centroid of the provinces. The null hypothesis of no phylogenetic correlation is tested assuming normality. The obtained p-value for women is 0.098. Taking as significance level $\alpha = 0.05$, the null hypothesis of no correlation is not rejected.

7 Conclusions

Poisson regression models are quite simple but flexible enough for modelling count variables. This work analyzes the number of people under the poverty line in Spanish provinces by using an area-level Poisson mixed model. In this framework, we have carried out a comparative study between the MM and PQL fitting algorithms. PQL performs better for the fixed effect coefficients but MM captures the variance component more precisely.

We consider that the EBP is a good alternative for describing the target variable due to the good performance shown in the design-based simulation experiment, where we have compared it against two plug-in estimators (using MM and PQL). Despite the inconsistency of PQL, the plug-in estimator of p_d using this fitting algorithm is very attractive, specially when the variance parameter is small. Further, it has a lower runtime. For example, taking $D = 52$ its runtime was 0.02 seconds in our computer while for the plug-in using MM and EBP (taking $L = 2500$) was 0.07.

For the EBP, we calculate the MSE and we introduce three estimators. The first two ones are plug-in estimators without and with bias correction of the second order. The third estimator is based on a parametric bootstrap. We analyze the behavior of the proposed estimators in a simulation study. The bias correction term is computationally intensive and the results of the plug-in estimators without and with bias correction are quite similar. As a good alternative, we suggest the bootstrap procedure, easy to implement and with similar results.

In the application to poverty data from the SLCS2008, we use the EBPs for estimating poverty rates since their results are more satisfactory than the ones obtained by the direct estimators. We conclude that the south and center-west provinces of Spain have highest levels of poverty. As performance measure we take the RRMSE estimated by parametric bootstrap. The RRMSE estimates are lower than 20.25% in all provinces.

acknowledgements

The work was supported by the grants MTM2013-41383-P, MTM2014-52876-R, MTM2011-22392, MTM2008-03010 and MTM2012-37077-C02-01 of the Spanish "Ministerio de Ciencia e Innovación" and by "Xunta de Galicia" CN2012/130.

References

Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Am Statist Assoc 88:9-25

Datta GS, Lahiri P (2000) A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. Stat Sinica 10:613-627

Esteban MD, Morales D, Pérez A, Santamaría L (2012) Small area estimation of poverty proportions under area-level time models. Comput Stat Data Anal 56:2840-2855

Fay RE, Herriot RA (1979) Estimation of income from small places: An application of james-stein procedures to census data. J Am Statist Assoc 74:269-277

Ghosh M, Rao JNK (1994) Small area estimation: An appraisal (with discussion). Statist Science 9:55-93

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2007) Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Comput Stat Data Anal 51:2720-2733

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008b) Bootstrap mean squared error of small-area EBLUP. J Stat Comput Simul 78:443-462

González-Manteiga W, Lombardía MJ, Molina I, Morales D, Santamaría L (2008a) Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. Comput Stat Data Anal 52:5242-5252

Jiang J (1998) Consistent estimators in generalized linear models. J Am Statist Assoc 93:720-729

Jiang J (2003) Empirical best prediction for small-area inference based on generalized linear mixed models. J Stat Plan Inference 111:117-127

Jiang J, Lahiri P (2001) Empirical best prediction for small area inference with binary data. A Inst Stat Math 53:217-243

Jiang J, Lahiri P (2006) Mixed model prediction and small area estimation. Test 15:1-96

Johnson FA, Chandra H, Brown J, Padmadas S (2010) Estimating district-level births attended by skilled attendants in Ghana using demographic health survey and census data: An application of small area estimation technique. J Official Stat 26 (2):341-359

Lahiri SN, Maiti T, Katzoff M, Parsons V (2007) Resampling-based empirical prediction: an application to small area estimation. Biometrika 94:469-485

Lin X (2007) Estimation using penalized quasilikelihood and quasi-pseudolikelihood in poisson mixed models. Lifetime Data Anal 13:533-544

López-Vizcaíno E, Lombardía MJ and Morales D (2013) Multinomial-based small area estimation of labour force indicators. Stat Modelling 13: 153-178

López-Vizcaíno E, Lombardía MJ and Morales D (2015) Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. J R Stat Soc Ser A 178:535-565.

MacNab YC, Lin X (2009) On empirical bayes penalized quasi-likelihood inference in GLMMs and in Bayesian disease mapping and ecological modelling. Comput Stat Data Anal 53:2950-2967

Pfeffermann D (2002) Small area estimation - new developments and directions. International Stat Review 70:125-143

Pfeffermann D (2013) New important developments in small area estimation. Stat Science 28:40-68

Prasad NGN, Rao JNK (1990) The estimation of the mean squared error of small-area estimators. J Am Statist Assoc 85:163-171

Rao JNK (1999) Some recent advances in model-based small area estimation. Survey Methodology 25:175-186

Rao JNK (2003) Small area estimation. John Wiley & Sons, New York

Rao JNK (2008) Some methods for small area estimation. Riv Intern Sc Sociali 4:387-406

Saei A, Chambers R (2003) Small area estimation under linear and generalized linear mixed models with time and area effects. Methodology Working Paper No M03/15 Southampton Statistical Sciences Research Institute, University of Southampton, UK

Särndal C, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer