

SUPPLEMENTARY MATERIAL

Small area prediction of proportions and counts under a spatial Poisson mixed model*

July 15, 2023

A MSM equations

The MSM method as a computationally efficient alternative to procedures based on maximizing likelihood. Under regularity conditions, the MSM method gives consistent estimators of GLMM parameters. In addition, it is a computational stable and economic method that only requires solving a non linear system of $p + 2$ equations. This is why, we have applied the MSM method to estimate the parameters of Model S1, in the application to real data.

The MSM estimators of the parameters of the Poisson mixed model S1 are calculated by solving a system of $p + 2$ nonlinear equations. This section derives the MSM equations appearing in (2.4) of Section 2. First, we recall that the moment generation function of $Y \sim N(\mu, \sigma^2)$ is

$$\Psi(t; \mu, \sigma^2) = \mathbb{E} [e^{tY}] = \exp \left\{ \mu t + \frac{1}{2} \sigma^2 t^2 \right\}.$$

For ease of exposition, we write $\gamma_{d_1 d_2} = \gamma_{d_1 d_2}(\rho)$. The expectation of y_d is

$$\begin{aligned} \mathbb{E}_{\boldsymbol{\theta}}[y_d] &= \mathbb{E}_v[\mathbb{E}_{\boldsymbol{\theta}}[y_d|\mathbf{v}]] = \mathbb{E}_v[\nu_d p_d] = \int_{-\infty}^{\infty} \nu_d \exp \{ \mathbf{x}_d \boldsymbol{\beta} + \phi v_d \} f_v(v_d) dv_d \\ &= \nu_d \exp \{ \mathbf{x}_d \boldsymbol{\beta} \} \Psi(\phi; 0, \gamma_{dd}) = \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\}. \end{aligned}$$

Therefore, the first p MSM equations are

$$f_k(\boldsymbol{\theta}) = \frac{1}{D} \sum_{d=1}^D \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} x_{dk} - \frac{1}{D} \sum_{d=1}^D y_d x_{dk}, \quad k = 1, \dots, p.$$

*Supported by the Instituto Galego de Estatística, by MICINN grants PID2020-113578RB-I00 and PGC2018-096840-B-I00, by the Generalitat Valenciana grant PROMETEO/2021/063 and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C 2020/14), and by GAIN (Galician Innovation Agency) and the Regional Ministry of Economy, Employment and Industry grant COV20/00604 and Centro de Investigación del Sistema Universitario de Galicia ED431G 2019/01, all of them through the ERDF.

The derivatives of $\mathbb{E}_{\boldsymbol{\theta}}[y_d]$ are

$$\begin{aligned}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \beta_k} &= \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} x_{dk}, \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \phi} &= \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} \phi \gamma_{dd}, \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \rho} &= \frac{1}{2} \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} \phi^2 \dot{\gamma}_{dd}.\end{aligned}$$

The expectation of y_d^2 is $\mathbb{E}_{\boldsymbol{\theta}}[y_d^2] = \mathbb{E}_v[\mathbb{E}_{\boldsymbol{\theta}}[y_d^2|\mathbf{v}]]$, where

$$\mathbb{E}_{\boldsymbol{\theta}}[y_d^2|\mathbf{v}] = \text{var}_{\boldsymbol{\theta}}[y_d|\mathbf{v}] + \mathbb{E}_{\boldsymbol{\theta}}^2[y_d|\mathbf{v}] = \nu_d p_d + \nu_d^2 p_d^2.$$

Therefore

$$\mathbb{E}_{\boldsymbol{\theta}}[y_d^2] = \mathbb{E}_v[\mathbb{E}_{\boldsymbol{\theta}}[y_d^2|\mathbf{v}]] = \int_{-\infty}^{\infty} \nu_d p_d f_v(v_d) dv_d + \int_{-\infty}^{\infty} \nu_d^2 p_d^2 f_v(v_d) dv_d.$$

We have

$$\begin{aligned}\int_{-\infty}^{\infty} p_d^2 f_v(v_d) dv_d &= \int_{-\infty}^{\infty} \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi v_d \right\} f_v(v_d) dv_d \\ &= \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} \right\} \Psi(2\phi; 0, \gamma_{dd}) = \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\},\end{aligned}$$

and as a consequence

$$\mathbb{E}_{\boldsymbol{\theta}}[y_d^2] = \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} + \nu_d^2 \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\}.$$

Then, the $(p+1)$ -th MSM equation is

$$f_{p+1}(\boldsymbol{\theta}) = \frac{1}{D} \sum_{d=1}^D \left\{ \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} + \nu_d^2 \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\} \right\} - \frac{1}{D} \sum_{d=1}^D y_d^2.$$

The derivatives of $\mathbb{E}_{\boldsymbol{\theta}}[y_d^2]$ are

$$\begin{aligned}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d^2]}{\partial \beta_k} &= \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} x_{dk} + 2\nu_d^2 \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\} x_{dk}, \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d^2]}{\partial \phi} &= \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} \phi \gamma_{dd} + 4\nu_d^2 \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\} \phi \gamma_{dd}, \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d^2]}{\partial \rho} &= \frac{1}{2} \nu_d \exp \left\{ \mathbf{x}_d \boldsymbol{\beta} + \frac{1}{2} \phi^2 \gamma_{dd} \right\} \phi^2 \dot{\gamma}_{dd} + 2\nu_d^2 \exp \left\{ 2\mathbf{x}_d \boldsymbol{\beta} + 2\phi^2 \gamma_{dd} \right\} \phi^2 \dot{\gamma}_{dd}.\end{aligned}$$

The expectation of $y_{d_1} y_{d_2}$ is

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}] = \mathbb{E}_v[\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}|\mathbf{v}]] = \mathbb{E}_v[\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1}|v_{d_1}] \mathbb{E}_{\boldsymbol{\theta}}[y_{d_2}|v_{d_2}]] = \nu_{d_1} \nu_{d_2} \mathbb{E}_v[p_{d_1} p_{d_2}],$$

where

$$\begin{aligned}\mathbb{E}_v[p_{d_1} p_{d_2}] &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \exp \left\{ \mathbf{x}_{d_2} \boldsymbol{\beta} + \phi v_{d_2} \right\} f_v(v_{d_2}|v_{d_1}) dv_{d_2} \right] \exp \left\{ \mathbf{x}_{d_1} \boldsymbol{\beta} + \phi v_{d_1} \right\} f_v(v_{d_1}) dv_{d_1} \\ &= \int_{-\infty}^{\infty} \exp \left\{ \mathbf{x}_{d_2} \boldsymbol{\beta} \right\} \Psi(\phi; \mu_{d_2|d_1}, \sigma_{d_2|d_1}^2) \exp \left\{ \mathbf{x}_{d_1} \boldsymbol{\beta} + \phi v_{d_1} \right\} f_v(v_{d_1}) dv_{d_1} \\ &= \exp \left\{ (\mathbf{x}_{d_1} + \mathbf{x}_{d_2}) \boldsymbol{\beta} + \frac{1}{2} \left(\gamma_{d_2 d_2} - \frac{\gamma_{d_1 d_2}^2}{\gamma_{d_1 d_1}} \right) \phi^2 \right\} \psi \left(\left(1 + \frac{\gamma_{d_1 d_2}}{\gamma_{d_1 d_1}} \right) \phi; 0, \gamma_{d_1 d_1} \right) \\ &= \exp \left\{ (\mathbf{x}_{d_1} + \mathbf{x}_{d_2}) \boldsymbol{\beta} + \frac{1}{2} \left(\gamma_{d_2 d_2} - \frac{\gamma_{d_1 d_2}^2}{\gamma_{d_1 d_1}} \right) \phi^2 + \frac{1}{2} \left(1 + \frac{\gamma_{d_1 d_2}}{\gamma_{d_1 d_1}} \right)^2 \phi^2 \gamma_{d_1 d_1} \right\}.\end{aligned}$$

Therefore, the $(p+2)$ -th MSM equation is

$$f_{p+2}(\boldsymbol{\theta}) = \frac{1}{D(D-1)} \sum_{d_1 \neq d_2}^D \nu_{d_1} \nu_{d_2} \varphi_{d_1, d_2}(\boldsymbol{\theta}) - \frac{1}{D(D-1)} \sum_{d_1 \neq d_2}^D y_{d_1} y_{d_2},$$

where

$$\begin{aligned} \varphi_{d_1, d_2}(\boldsymbol{\theta}) &= \exp \left\{ (\mathbf{x}_{d_1} + \mathbf{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2} \left(\gamma_{d_2 d_2} - \frac{\gamma_{d_1 d_2}^2}{\gamma_{d_1 d_1}} \right) \phi^2 + \frac{1}{2} \frac{(\gamma_{d_1 d_1} + \gamma_{d_1 d_2})^2}{\gamma_{d_1 d_1}} \phi^2 \right\} \\ &= \exp \left\{ (\mathbf{x}_{d_1} + \mathbf{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2} \phi^2 (\gamma_{d_1 d_1} + \gamma_{d_2 d_2} + 2\gamma_{d_1 d_2}) \right\}. \end{aligned}$$

The derivatives of $\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}]$ are

$$\begin{aligned} \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}]}{\partial \beta_k} &= \nu_{d_1} \nu_{d_2} \varphi_{d_1, d_2}(\boldsymbol{\theta}) (x_{d_1 k} + x_{d_2 k}), \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}]}{\partial \phi} &= \nu_{d_1} \nu_{d_2} \varphi_{d_1, d_2}(\boldsymbol{\theta}) \phi (\gamma_{d_1 d_1} + \gamma_{d_2 d_2} + 2\gamma_{d_1 d_2}), \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}]}{\partial \rho} &= \frac{1}{2} \nu_{d_1} \nu_{d_2} \varphi_{d_1, d_2}(\boldsymbol{\theta}) \phi^2 (\dot{\gamma}_{d_1 d_1} + \dot{\gamma}_{d_2 d_2} + 2\dot{\gamma}_{d_1 d_2}). \end{aligned}$$

The elements of the Jacobian matrix are

$$\begin{aligned} H_{kr} &= \frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D} \sum_{d=1}^D \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \theta_r} x_{dk}, \quad k = 1, \dots, p, \quad r = 1, \dots, p+2, \\ H_{p+1r} &= \frac{\partial f_{p+1}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D} \sum_{d=1}^D \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d^2]}{\partial \theta_r}, \quad r = 1, \dots, p+2, \\ H_{p+2r} &= \frac{\partial f_{p+2}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D(D-1)} \sum_{d_1 \neq d_2}^D \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1} y_{d_2}]}{\partial \theta_r}, \quad r = 1, \dots, p+2. \end{aligned}$$

B PQL estimation

The scientific literature on GLMMs presents alternative procedures to the ML method. One of these methods consists of maximizing the joint likelihood of the fixed and random effects of the model, for each optimal selection of the variance and correlation parameters. This procedure, which we call PQL, alternately combines two optimization algorithms, the first on $\boldsymbol{\beta}$ and $\mathbf{u}$ and the second on ϕ and ρ , such that the output of one algorithm feeds the input from the other and vice versa. It is a procedure that calculates ML estimators in the case of LMMs but does not guarantee consistent estimators for GLMMs. The method has a lower computational cost than the ML-Laplace algorithm, but it presents instability problems (lack of convergence) in complex models, such as those of spatial correlation. In what follows, we describe the PQL method.

Let us first assume that the variance components ϕ and ρ are known. The PQL estimator of $\boldsymbol{\beta}$ and predictors of $\mathbf{v}$ maximizes the joint log-likelihood for $\boldsymbol{\beta}$ and $\mathbf{v}$ is

$$\begin{aligned} \ell &= \ell(\boldsymbol{\beta}, \mathbf{v}) = -\frac{D}{2} \log 2\pi - \frac{1}{2} \log |\boldsymbol{\Gamma}(\rho)|^{-1/2} - \frac{1}{2} \mathbf{v}' \boldsymbol{\Gamma}^{-1}(\rho) \mathbf{v} \\ &\quad - \sum_{d=1}^D \nu_d \exp\{\mathbf{x}_d \boldsymbol{\beta} + \phi v_d\} + \sum_{k=1}^p \left(\sum_{d=1}^D y_d x_{dk} \right) \beta_k + \phi \sum_{d=1}^D y_d v_d \end{aligned}$$

The first partial derivatives of ℓ with respect to $\boldsymbol{\beta}$ and $\mathbf{v}$ are

$$\begin{aligned}\frac{\partial \ell}{\partial \beta_k} &= -\sum_{d=1}^D \nu_d \exp\{\mathbf{x}_d \boldsymbol{\beta} + \phi v_d + \phi_2 v_{2,dt}\} x_{dk} + \sum_{d=1}^D y_d x_{dk}, \quad k = 1, \dots, p, \\ \frac{\partial \ell}{\partial v_d} &= -\boldsymbol{\delta}'_d \boldsymbol{\Gamma}^{-1}(\rho) \mathbf{v} - \phi \nu_d \exp\{\mathbf{x}_d \boldsymbol{\beta} + \phi v_d\} + \phi y_d, \quad d = 1, \dots, D,\end{aligned}$$

where $\boldsymbol{\delta}_d = \text{col}_{1 \leq j \leq D}(\delta_{jd})$, $d = 1, \dots, D$. By equating to zero, we have a nonlinear system with $D + p$ equations and unknowns, i.e.

$$\text{System A: } \frac{\partial \ell}{\partial \beta_k} = 0, \quad \frac{\partial \ell}{\partial v_d} = 0, \quad k = 1, \dots, p, \quad d = 1, \dots, D.$$

We may solve System A with any available method, like Newton-Raphson or quasi-Newton Broyden. At the step i of the PQL algorithm, the solution of System A are denoted by $\boldsymbol{\beta}^{(i)}$ and $\mathbf{v}^{(i)}$. We can now update the values of ϕ and ρ by maximizing the log-likelihood of

$$\boldsymbol{\eta}_d^{(i)} = \log \nu_d + \mathbf{x}_d \boldsymbol{\beta}^{(i)} + \phi v_d^{(i)}, \quad d = 1, \dots, D, t = 1, \dots, T.$$

It holds that

$$\boldsymbol{\eta}^{(i)} = \text{col}_{1 \leq d \leq D}(\boldsymbol{\eta}_d^{(i)}) \sim N\left(\log \boldsymbol{\nu} + \mathbf{X} \boldsymbol{\beta}^{(i)}, \phi^2 \boldsymbol{\Gamma}(\delta)\right),$$

where $\boldsymbol{\nu} = \text{col}_{1 \leq d \leq D}(\nu_d)$ and $\mathbf{X} = \text{col}_{1 \leq d \leq D}(\mathbf{x}_d)$. The joint log-likelihood of $\boldsymbol{\eta}^{(i)}$ is

$$\begin{aligned}l^{(i)} &= \ell^{(i)}(\phi, \rho) = -\frac{D}{2} \log 2\pi - \frac{1}{2} \log |\phi^2 \boldsymbol{\Gamma}(\delta)| \\ &\quad - \frac{1}{2} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)})' (\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)}),\end{aligned}$$

where $\boldsymbol{\nu} = \text{col}_{1 \leq d \leq D}(\nu_d)$ and $\mathbf{X} = \text{col}_{1 \leq d \leq D}(\mathbf{x}_d)$. By treating $\boldsymbol{\eta}^{(i)}$ as known, the derivatives of $l^{(i)}$ with respect to ϕ and ρ are

$$\begin{aligned}\frac{\partial l^{(i)}}{\partial \phi^2} &= -\frac{1}{2} \text{tr}\left((\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} \boldsymbol{\Gamma}(\delta)\right) + \frac{1}{2} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)})' \\ &\quad \cdot (\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} \boldsymbol{\Gamma}(\rho) (\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)}), \\ \frac{\partial l^{(i)}}{\partial \rho} &= -\frac{1}{2} \text{tr}\left((\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} \phi^2 \dot{\boldsymbol{\Gamma}}(\delta)\right) + \frac{1}{2} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)})' \\ &\quad \cdot (\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} \phi^2 \dot{\boldsymbol{\Gamma}}(\rho) (\phi^2 \boldsymbol{\Gamma}(\delta))^{-1} (\boldsymbol{\eta}^{(i)} - \log \boldsymbol{\nu} - \mathbf{X} \boldsymbol{\beta}^{(i)}),\end{aligned}$$

where

$$\dot{\boldsymbol{\Gamma}}(\rho) = -\mathbf{C}^{-1}(\rho) \dot{\mathbf{C}}(\rho) \mathbf{C}^{-1}(\rho), \quad \mathbf{C}(\rho) = (\mathbf{I}_D - \rho \mathbf{W})' (\mathbf{I}_D - \rho \mathbf{W}), \quad \dot{\mathbf{C}}(\rho) = -2 \mathbf{W}' (\mathbf{I}_D - \rho \mathbf{W}).$$

By equating to zero the two derivatives, we have the non linear system of equations

$$\text{System B: } \frac{\partial l^{(i)}}{\partial \phi^2} = 0, \quad \frac{\partial l^{(i)}}{\partial \rho} = 0.$$

The PQL algorithm updates iteratively systems A and B until convergence.

The computational cost of the PQL method is much higher than that of the MSM method. In the MSM case, a nonlinear system of $p + 2$ equations has to be solved only once. In the PQL case, two systems of nonlinear equations have to be solved iteratively until convergence. System A has $D + p$ equations and System B has 2 equations. This is the reason why we choose the MSM method for fitting the area-level spatial Poisson mixed model S1 and not the PQL method.

C ML estimation

We recall that the marginal likelihood function of the spatial Poisson mixed model is

$$\begin{aligned}\mathbb{P}(\mathbf{y}; \boldsymbol{\theta}) &= (2\pi)^{-D/2} |\boldsymbol{\Gamma}(\rho)|^{-1/2} \left(\prod_{d=1}^D y_d! \right)^{-1} \\ &\cdot \int_{\mathbb{R}^D} \exp \left\{ \sum_{d=1}^D \left\{ -\nu_d \exp\{\mathbf{x}_d \boldsymbol{\beta} + \phi v_d\} + y_d \log \nu_d \right\} \right\} \\ &\cdot \exp \left\{ \sum_{k=1}^p \left(\sum_{d=1}^D y_d x_{dk} \right) \beta_k + \phi \sum_{d=1}^D y_d v_d - \frac{1}{2} \mathbf{v}' \boldsymbol{\Gamma}^{-1}(\rho) \mathbf{v} \right\} d\mathbf{v}.\end{aligned}$$

Given $\mathbf{y}$, the maximum likelihood (ML) estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi, \rho)$ is

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}_{\boldsymbol{\theta} \in \Theta} \mathbb{P}(\mathbf{y}; \boldsymbol{\theta}), \quad \Theta = \mathbb{R}^p \times (0, \infty) \times (-1, 1).$$

The expression of $\mathbb{P}(\mathbf{y}; \boldsymbol{\theta})$ contains integrals in $\mathbb{R}^D$. To maximize $\mathbb{P}(\mathbf{y}; \boldsymbol{\theta})$ in $\boldsymbol{\theta}$ two functions can be applied sequentially. The first function would compute the multiple integral on $\mathbf{v}$, and the second function would perform the maximization on $\boldsymbol{\theta}$. Alternatively we describe ML-Laplace approximation algorithm.

C.1 ML-Laplace approximation algorithm

Let us start with Laplace approximation of a multiple integral of a general function $\exp(h(\mathbf{v}))$, where $h : \mathbb{R}^D \mapsto \mathbb{R}$ is a twice continuously differentiable function with a global maximum at the column vector $\mathbf{v}_0$. This is to say, let us assume that $\dot{h}(\mathbf{v}_0) = \frac{dh}{d\mathbf{v}} \Big|_{\mathbf{v}=\mathbf{v}_0} = \mathbf{0}$ and $\ddot{h}(\mathbf{v}_0) =$

$\frac{d^2 h}{d\mathbf{v}^2} \Big|_{\mathbf{v}=\mathbf{v}_0}$ is negative definite. A Taylor series expansion of $h(\mathbf{v})$ around $\mathbf{v}_0$ yields to

$$\begin{aligned}h(\mathbf{v}) &= h(\mathbf{v}_0) + \dot{h}(\mathbf{v}_0)(\mathbf{v} - \mathbf{v}_0) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_0)' \ddot{h}(\mathbf{v}_0)(\mathbf{v} - \mathbf{v}_0) + o(\|\mathbf{v} - \mathbf{v}_0\|^2) \\ &\approx h(\mathbf{v}_0) + \frac{1}{2}(\mathbf{v} - \mathbf{v}_0)' \ddot{h}(\mathbf{v}_0)(\mathbf{v} - \mathbf{v}_0).\end{aligned}$$

The Laplace approximation of the integral of the function $\exp(h(\mathbf{v}))$ is

$$\begin{aligned}\int_{\mathbb{R}^D} e^{h(\mathbf{v})} d\mathbf{v} &\approx \int_{\mathbb{R}^D} e^{h(\mathbf{v}_0)} \exp \left\{ -\frac{1}{2}(\mathbf{v} - \mathbf{v}_0)' (-\ddot{h}(\mathbf{v}_0))(\mathbf{v} - \mathbf{v}_0) \right\} d\mathbf{v} \\ &= (2\pi)^{D/2} |-\ddot{h}(\mathbf{v}_0)|^{-1/2} e^{h(\mathbf{v}_0)}.\end{aligned}$$

The marginal likelihood function of the spatial Poisson mixed model can be written in the form

$$\mathbb{P}(\mathbf{y}; \boldsymbol{\theta}) = (2\pi)^{-D/2} |\boldsymbol{\Gamma}(\rho)|^{-1/2} \left(\prod_{d=1}^D y_d! \right)^{-1} \int_{\mathbb{R}^D} \exp \{h(\mathbf{v}; \mathbf{y}, \boldsymbol{\theta})\} d\mathbf{v},$$

where

$$\begin{aligned}h(\mathbf{v}; \mathbf{y}, \boldsymbol{\theta}) &= \exp \left\{ \sum_{d=1}^D \left\{ -\nu_d \exp\{\mathbf{x}_d \boldsymbol{\beta} + \phi v_d\} + y_d \log \nu_d \right\} \right\} \\ &+ \sum_{k=1}^p \left(\sum_{d=1}^D y_d x_{dk} \right) \beta_k + \phi \sum_{d=1}^D y_d v_d - \frac{1}{2} \mathbf{v}' \boldsymbol{\Gamma}^{-1}(\rho) \mathbf{v}.\end{aligned}$$

In order to apply the Laplace approximation to the likelihood integral, we have to maximize $h(\mathbf{v}; \mathbf{y}, \boldsymbol{\theta})$ in $\mathbf{v}$, given $\mathbf{y}$ and $\boldsymbol{\theta}$. For simplicity, we write $h(\mathbf{v})$. We can carry out the maximization by applying an R function of optimization. Alternatively, we can implement a Newton-Raphson algorithm after calculating the first and second partial derivatives of h with respect to v_d , $d = 1, \dots, D$, given $\mathbf{y}$ and $\boldsymbol{\theta}$.

Let $\dot{h}$ and $\ddot{h}$ denote the $D \times 1$ vector and the $D \times D$ matrix of first and second order partial derivatives of $h(\mathbf{v})$ with respect to $\mathbf{v}$, respectively. The Newton-Raphson updating equation is

$$\mathbf{v}^{(i+1)} = \mathbf{v}^{(i)} - \ddot{h}^{-1}(\mathbf{v}^{(i)}) \dot{h}(\mathbf{v}^{(i)}). \quad (\text{C.1})$$

Let us denote by $\mathbf{v}^\circ$ the argument of maxima of the function $h(\mathbf{v})$. Thus, it holds $\dot{h}(\mathbf{v}^\circ) = \mathbf{0}$ and the matrix $\ddot{h}(\mathbf{v}^\circ)$ is negative definite. The loglikelihood of the spatial Poisson mixed model can be approximated by

$$\log P(\mathbf{y}; \boldsymbol{\theta},) \approx \frac{1}{2} \log |\boldsymbol{\Gamma}(\rho)| + h(\mathbf{v}^\circ) - \frac{1}{2} \log |-\ddot{h}(\mathbf{v}^\circ)| \triangleq g(\boldsymbol{\theta}; \mathbf{y}, \mathbf{v}^\circ).$$

The following step is to maximize $g(\boldsymbol{\theta}; \mathbf{y}, \mathbf{v}^\circ)$ in $\boldsymbol{\theta} \in \Theta$. For simplicity, we write $g(\boldsymbol{\theta})$. We can carry out the maximization by applying an R function of optimization. Alternatively, we can implement a Newton-Raphson algorithm after calculating the first and second partial derivatives of g with respect to the components of $\boldsymbol{\theta}$, given $\mathbf{y}$ and $\mathbf{v}^\circ$.

Let us define $M = p + 2$. Let $\dot{g}$ and $\ddot{g}$ denote the $M \times 1$ vector and the $M \times M$ matrix of first and second order partial derivatives of $g(\boldsymbol{\theta})$, respectively. The Newton-Raphson updating equation is

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \ddot{g}^{-1}(\boldsymbol{\theta}^{(i)}) \dot{g}(\boldsymbol{\theta}^{(i)}). \quad (\text{C.2})$$

The final ML-Laplace approximation algorithm combines the two described Newton-Raphson algorithms and can be described by the following steps:

1. Set the initial values $i = 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$, $\varepsilon_4 > 0$, $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(-1)} = \boldsymbol{\theta}^{(0)} + \mathbf{1}$, $\mathbf{v}^{(0)} = \mathbf{0}$, $\mathbf{v}^{(-1)} = \mathbf{1}$, where $\mathbf{0}$ and $\mathbf{1}$ are column vectors of zeros and ones respectively.
2. Until $\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}\|_2 < \varepsilon_1$, $\|\mathbf{v}^{(i)} - \mathbf{v}^{(i-1)}\|_2 < \varepsilon_2$, do
 - (a) Apply algorithm (C.1) with seeds $\mathbf{v}^{(i)}$, convergence tolerance ε_3 and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}$ fixed. Output: $\mathbf{v}^{(i+1)}$.
 - (b) Apply algorithm (C.2) with seed $\boldsymbol{\theta}^{(i)}$, convergence tolerance ε_4 and $\mathbf{v} = \mathbf{v}^{(i+1)}$ fixed. Output: $\boldsymbol{\theta}^{(i+1)}$.
 - (c) $i \leftarrow i + 1$.
3. Output: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(i)}$ and $\hat{\mathbf{v}} = \mathbf{v}^{(i)}$.

The ML-Laplace algorithm is apparently attractive as it provides the ML estimators of the model parameters and the modal predictions of the random effects. However, it has a very high computational cost. The algorithm sequentially runs two multivariate maximization sub-algorithms until convergence. This fact sometimes produces problems of computational stability because of overflow or non-convergence problems. On the other hand, the MSM method requires solving a nonlinear system of $p + 2$ equations. This can be done by running a Newton-Raphson algorithm only once. This is the reason why we choose the MSM method for fitting the area-level spatial Poisson mixed model S1 and not the ML-Laplace method.

D Bootstrap algorithms

Algorithm 1 gives a bootstrap test for $H_0 : \phi^2 = 0$. The steps are

1. Fit Model S1 to data and calculate $\hat{\beta}$ and $\hat{\phi}$.
2. Fit Model 0 (with $\phi = 0$) to data and calculate $\hat{\beta}^0$.
3. For $b = 1, \dots, B$, do
 - i) Generate a bootstrap resample under $H_0 : \phi^2 = 0$, i.e.

$$p_d^{*(b)} = \exp\{\mathbf{x}_d \hat{\beta}^0\}, y_d^{*(b)} \sim \text{Pois}(\nu_d p_d^{*(b)}), d = 1, \dots, D.$$

- ii) Fit Model 1 to the bootstrap data $(y_d^{*(b)}, \mathbf{x}_d)$, $d = 1, \dots, D$. Calculate $\hat{\beta}^{*(b)}$ and $\hat{\phi}^{*(b)}$.

4. Calculate the p -value: $p = \frac{\#\{\hat{\phi}^{*(b)2} > \hat{\phi}^2\}}{B}$.

Algorithm 2 gives a bootstrap test for $H_0 : \rho = 0$. The steps are

1. Fit Model S1 to data and calculate $\hat{\beta}$, $\hat{\phi}$ and $\hat{\rho}$.
2. Fit Model 1 (with $\rho = 0$) to data and calculate $\hat{\beta}^0$ and $\hat{\phi}^0$.
3. For $b = 1, \dots, B$, do

- i) Generate a bootstrap resample under $H_0 : \rho = 0$, i.e.

$$v_d^{*(b)} \sim N(0, 1), p_d^{*(b)} = \exp\{\mathbf{x}_d \hat{\beta}^0 + \hat{\phi}^0 v_d^{*(b)}\}, y_d^{*(b)} \sim \text{Pois}(\nu_d p_d^{*(b)}), d = 1, \dots, D.$$

- ii) Fit Model S1 to the bootstrap data $(y_d^{*(b)}, \mathbf{x}_d)$, $d = 1, \dots, D$. Calculate $\hat{\beta}^{*(b)}$, $\hat{\phi}^{*(b)}$ and $\hat{\rho}^{*(b)}$.

4. Calculate the p -value: $p = \frac{\#\{|\hat{\rho}^{*(b)}| > |\hat{\rho}|\}}{B}$.

E Additional simulation results

This section extends the simulation experiments in Section 4 of the main manuscript to other data generating processes. Instead of generating $u_1, \dots, u_D$ i.i.d. $N(0, 1)$, before applying the transformation $\mathbf{v} = (\mathbf{I}_D - \rho \mathbf{W})^{-1} \mathbf{u}$, with $\mathbf{u} = (u_1, \dots, u_D)'$, we generate $u_1, \dots, u_D$ i.i.d from the distributions

- $N(0, 1)$ as in Section 4 of the main manuscript.
- t -Student with degrees of freedom $df=2$ and $df=5$.
- Gumbel distribution with location parameter 0 and scale parameters $\beta = 0.8$ and $\beta = 1.2$.
- Skew normal distribution with location parameter 0, scale parameter 1 and shape parameters $\alpha = 1$ and $\alpha = 4$.

Because of the lower computational cost and the good simulations results obtained under the normality assumptions, we restrict the this study to the approximate EBP (EBP^a) and the approximate plug-in (P_{EBP}^a) predictors of p_d .

Table E.1 presents the BIAS ($\times 10^2$) of predictors of EBP^a and P_{EBP}^a based on Model S1. Random effects are generated according to a t -Student with 2 and 5 degrees of freedom (denoted by $df=2$ and $df=5$), Gumbel $G(0, \beta)$ with $\beta = 0.8$ and $\beta = 1.2$ (denoted by $\beta = 0.8$ and $\beta = 1.2$), and a skew normal $SN(0, 1, \alpha)$ with $\alpha = 1, 4$ (denoted by $\alpha = 1$ and $\alpha = 4$). In general, we observe some increase of bias when deviating from the hypothesis of normality.

Table E.1: BIAS ($\times 10^2$) of predictors of p_d (EBP^a and P_{EBP}^a) based on Model S1.

ρ	EBP^a							$PLUG^a$						
	N(0,1)	df=2	df=5	$\beta = 0.8$	$\beta = 1.2$	$\alpha = 1$	$\alpha = 4$	N(0,1)	df=2	df=5	$\beta = 0.8$	$\beta = 1.2$	$\alpha = 1$	$\alpha = 4$
0.1	0.189	0.277	0.197	0.329	0.430	0.288	0.281	0.177	0.272	0.190	0.321	0.405	0.266	0.267
0.3	0.205	0.301	0.224	0.340	0.542	0.302	0.362	0.194	0.317	0.230	0.321	0.528	0.291	0.347
0.5	0.222	0.336	0.263	0.424	0.704	0.431	0.543	0.213	0.336	0.255	0.405	0.677	0.411	0.528

Table E.2 presents the RMSE ($\times 10^2$) of predictors EBP^a and P_{EBP}^a based on Model S1. Random effects are generated according to the same distributions as for Table E.1. In general, we observe some increase of root mean squared errors when deviating from the hypothesis of normality.

Table E.2: RMSE ($\times 10^2$) of predictors of p_d (EBP^a and P_{EBP}^a) based on Model S1.

ρ	EBP^a							$PLUG^a$						
	N(0,1)	df=2	df=5	$\beta = 0.8$	$\beta = 1.2$	$\alpha = 1$	$\alpha = 4$	N(0,1)	df=2	df=5	$\beta = 0.8$	$\beta = 1.2$	$\alpha = 1$	$\alpha = 4$
0.1	2.170	3.846	2.632	2.526	3.659	2.096	1.818	2.159	3.831	2.609	2.498	3.617	2.061	1.784
0.3	2.236	3.862	2.722	2.668	3.923	2.179	1.958	2.226	3.846	2.711	2.630	3.866	2.138	1.907
0.5	2.380	3.896	2.864	2.885	4.364	2.466	2.239	2.354	3.885	2.843	2.853	4.313	2.393	2.161

F Additional results for the application to real data

This section presents additional results for the application to real data. Specifically, the goodness of fit of various Poisson regression models to the count data from the Spanish Survey of Living Conditions in 2013 (SLCS2013) is analyzed. We consider the following models: (1) Poisson regression model without random effects, (2) Poisson regression model with independent random effects, and (3) Poisson regression model with SAR(1) correlated random effects. In the last case, we consider three proximity matrices: (a) common borders, (b) Euclidean distance between domain centroids, and (c) k nearest neighbours.

F.1 Poisson regression model without random effects

Table F.1 presents the estimates of the regression coefficient under the area-level fixed-effect Poisson model (Model 0). Table F.2 presents a descriptive analysis of the Pearson residuals under Model 0. Figure F.1 plots the Pearson residuals for the model with only fixed effects. We observe that there too many Pearson residuals outside the interval $(-3, 3)$, which indicates that Model 0 does not have a good fit to the SLCS2013 data.

Table F.1: Coefficients estimated under Model 0.

Variable	Est.	s.e.	z -value	$P(> z)$
<i>Intercept</i>	-1.8711	0.0924	-20.254	< 0.001
<i>lab2</i>	2.9870	0.7541	3.961	< 0.001
<i>edu3</i>	-1.3815	0.3211	-4.302	< 0.001

Table F.2: Pearson residuals under Model 0.

Min	1st Qu	Median	Mean	3rd Qu	Max	Sd
-3.0582	-1.0426	0.0247	-0.0164	0.6124	4.0243	1.0135

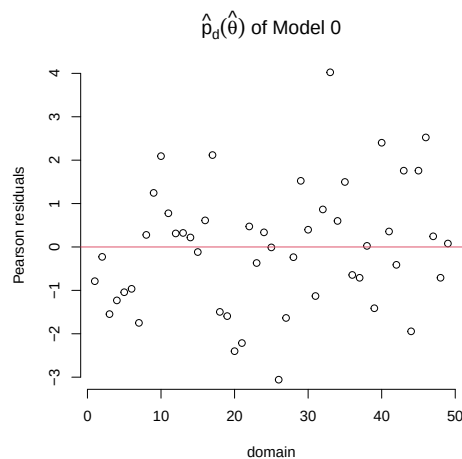


Figure F.1: Pearson residuals of Model 0.

F.2 Poisson regression model with independent random effects

Table F.3 presents the coefficient estimates under the area-level Poisson regression model with independent random effects (Model 1). The estimate of the variance parameter is $\hat{\phi} = 0.168$ and its 95% confidence interval is (0.0782, 0.2666). Table F.4 presents a descriptive analysis of the Pearson residuals under Model 1. Figure F.2 plots the Pearson residuals under Model 1. We observe that all the Pearson residuals inside the interval $(-3, 3)$, which indicates that Model 1 has a better fit to the SLCS2013 data than Model 0.

Table F.3: Coefficients estimated under Model 0.

Variable	Est.	s.e.	z-value	$P(> z)$
<i>Intercept</i>	-1.8239	0.1315	-13.869	< 0.001
<i>lab2</i>	2.6308	1.1266	2.335	0.0195
<i>edu3</i>	-1.5403	0.5610	-2.745	0.0060

Table F.4: Pearson residuals under Model 1.

Min	1st Qu	Median	Mean	3rd Qu	Max	Sd
-2.1626	-0.6826	0.0276	-0.0534	0.5013	2.3473	1.0102

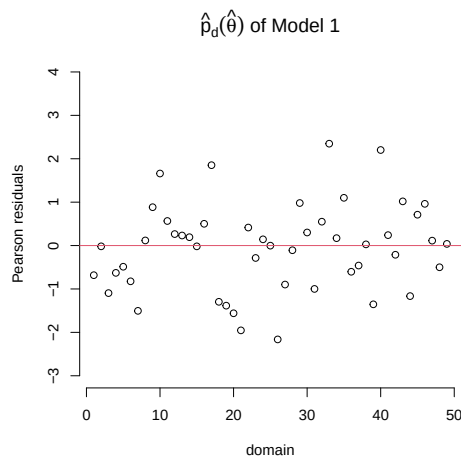


Figure F.2: Pearson residuals of Model 1.

F.3 Poisson regression models with SAR(1) correlated random effects

This section presents a comparative analysis of the behaviour of different proximity matrices $\mathbf{W}$. Table F.5 presents a descriptive analysis of the Pearson residuals under the Poisson model with random effects (Model S1) and different proximity matrices $\mathbf{W}$.

Table F.5: Pearson residuals under Model S1.

Proximity matrix	Min	1st Qu	Median	Mean	3rd Qu	Max	Sd
common borders	-2.1792	-0.6790	0.0263	-0.0552	0.5088	2.3485	1.0057
distances	-2.9243	-1.1440	-0.0076	0.1332	0.9512	4.1968	1.6116
knn	-2.9243	-1.1440	-0.0076	0.1332	0.9512	4.1969	1.6117

F.3.1 Common-borders proximity matrix

Table F.6 presents the coefficient estimates under the area-level Poisson mixed model with SAR(1) domain effects and a common-borders proximity matrix $\mathbf{W}$. The estimates of the variance and correlation parameters are $\hat{\phi} = 0.130$ and $\hat{\rho} = 0.324$, respectively. The corresponding 95% confidence interval are $(0.001, 0.331)$ and $(0.194, 0.407)$. Figure F.3 plots the Pearson residuals for the common-borders proximity matrix. We note that the confidence interval for ρ does not contain the zero. Further, we observe that all the Pearson residuals are inside the interval $(-3, 3)$. These two facts indicates that common-borders proximity matrix gives a good representation of the correlation structure of the data. This is why, this approach is selected for the application to real data.

Table F.6: Coefficients estimated under Model S1.

Variable	Est.	s.e.	z-value	$P(> z)$
<i>Intercept</i>	-1.8803	0.1515	-12.4086	< 0.001
<i>lab2</i>	2.9848	1.2097	2.4689	0.0136
<i>edu3</i>	-1.3809	0.5033	-2.7445	0.0061

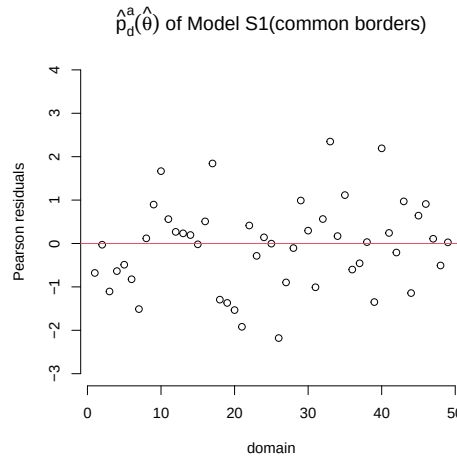


Figure F.3: Pearson residuals of Model S1 with common-borders proximity matrix.

F.3.2 Distance-based proximity matrix

Table F.7 presents the coefficient estimates under the area-level Poisson mixed model with SAR(1) domain effects and a distance-based proximity matrix $\mathbf{W}$. The estimates of the variance

and correlation parameters are $\hat{\phi} = 0.0008$ and $\hat{\rho} = 0.1174$, respectively. The corresponding 95% confidence interval are $(0.0001, 0.2663)$ and $(-0.0490, 0.1960)$. Figure F.4 plots the Pearson residuals for the distance-based proximity matrix. We note that the confidence interval for ρ contains the zero. Further, we observe that there are Pearson residuals outside the interval $(-3, 3)$. These two facts indicates that distance-based proximity matrix does not give a good representation of the correlation structure of the data.

Table F.7: Estimates of model coefficients.

Variable	Est.	s.e.	z-value	$P(> z)$
<i>Intercept</i>	-1.8090	0.0225	-80.3928	< 0.001
<i>lab2</i>	0.0004	0.0012	0.3114	0.7555
<i>edu3</i>	-0.0018	0.0012	-1.4423	0.1492

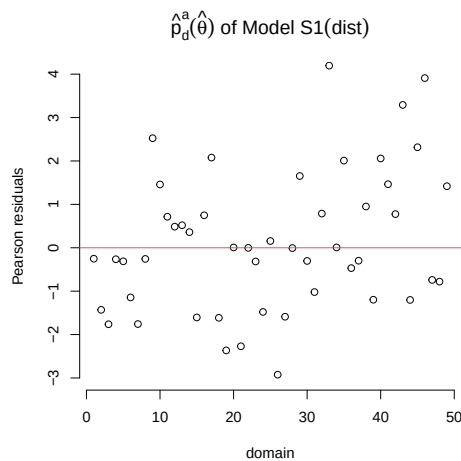


Figure F.4: Pearson residuals of Model S1 with distance-based proximity matrix.

F.3.3 k nearest neighbour proximity matrix

Table F.9 presents the coefficient estimates under the area-level Poisson mixed model with SAR(1) domain effects and a k nearest neighbour (knn) proximity matrix $\mathbf{W}$ with $(k = 3)$. The estimates of the variance and correlation parameters are $\hat{\phi} = 0.0007$ and $\hat{\rho} = 0.4552$, respectively. The corresponding 95% confidence interval are $(0.0001, 0.2657)$ and $(-0.2160, 0.5850)$. Figure F.6 plots the Pearson residuals for the knn proximity matrix. We note that the confidence interval for ρ contains the zero. Further, we observe that there Pearson residuals outside the interval $(-3, 3)$. These two facts indicates that knn proximity matrix does not give a good representation of the correlation structure of the data.

Table F.8: Coefficients estimated under Model S1.

Variable	Est.	s.e.	z-value	$P(> z)$
<i>Intercept</i>	-1.8090	0.0228	-79.2666	< 0.001
<i>lab2</i>	0.0004	0.0012	0.3104	0.7562
<i>edu3</i>	-0.0018	0.0013	-1.4151	0.1570

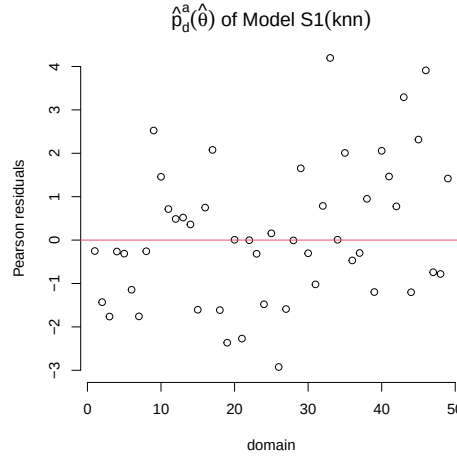


Figure F.5: Pearson residuals of Model S1 with knn proximity matrix.

F.4 Spatial Fay-Herriot model

This section presents a summary of the application of the statistical methodology of Pratesi and Salvati (2008) to the estimation of poverty proportions. We fit a mixed linear area model with SAR(1) spatially correlated random effects (spatial Fay-Herriot model) to the aggregated data by using the function `eblupSFH` of R package `sae`. We carry out some residual analysis for model diagnosis. Based on the fitted model, we calculate the Spatial Empirical Best Linear Unbiased Predictor of Pratesi and Salvati (2008) (SEBLUP) of the poverty proportions. The functions `eblupSFH` and `mseSFH` of R package `sae` delivers the SEBLUPs and the corresponding root-MSEs, respectively. We finally compare the results given by the SEBLUPs based on the spatial Fay-Herriot model with those obtained by the approximate EBPs based on the spatial Poisson mixed model (Model S1).

Table F.9 presents the coefficient estimates under the spatial Fay-Herriot model. The estimates of the variance and correlation parameters are $\hat{\phi} = 0.0013$ and $\hat{\rho} = 0.7369$, respectively.

Table F.9: Coefficients estimated under Model S1.

Variable	Est.	s.e.	z-value	$P(> z)$
<i>Intercept</i>	0.1622	0.0306	5.3047	< 0.001
<i>lab2</i>	0.1115	0.2227	0.5004	0.6168
<i>edu3</i>	-0.1285	0.0967	-1.3294	0.1837

Figure F.6 plots the Pearson residuals of the EBP approximation under Model S1, i.e.

$$r_d^P = \frac{y_d - \nu_d \hat{p}_d^{seblup}}{rmse(\hat{p}_d^{seblup})}, \quad d = 1, \dots, D.$$

The distribution of the Pearson residuals is symmetrical around 0 and takes values in the interval $(-2, 2)$.

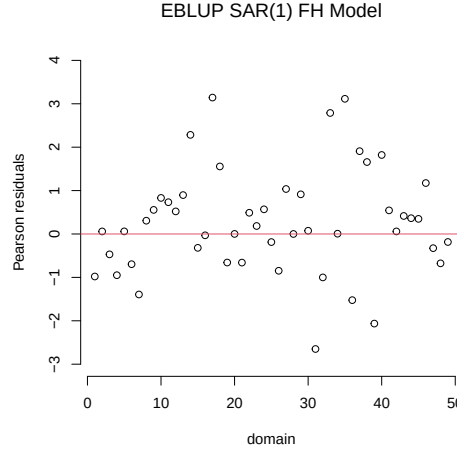


Figure F.6: Pearson residuals of Spatial Fay-Herriot model.

Figure F.7 plots the direct estimates (dir), the approximate EBPs (EBP^a) and the SEBLUPs of poverty proportions p_d (left) and relative root-MSEs (right) for women in 2013, showing SEBLUP as the best alternative.

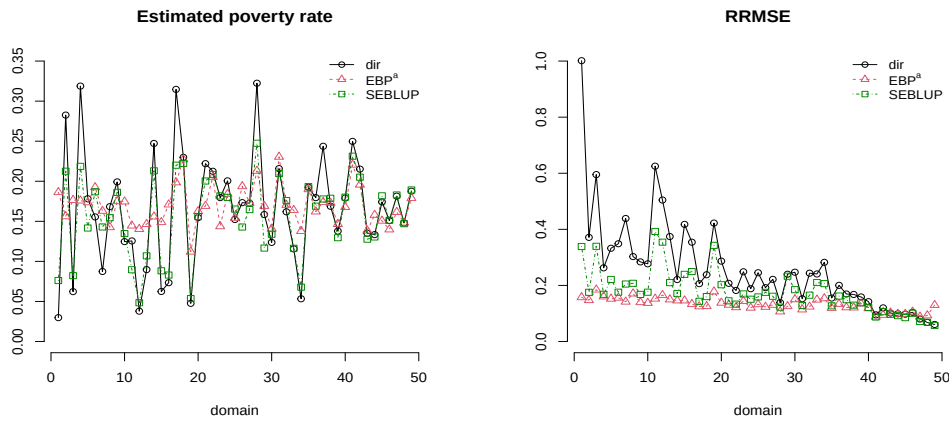


Figure F.7: Direct estimates, approximate EBPs and SEBLUPs of poverty proportions p_d (left) and relative root-MSEs (right) for women in 2013.

Figure F.8 maps the SEBLUPs of poverty proportions (left) and their relative root-MSEs (right) for women in Galicia during 2013.

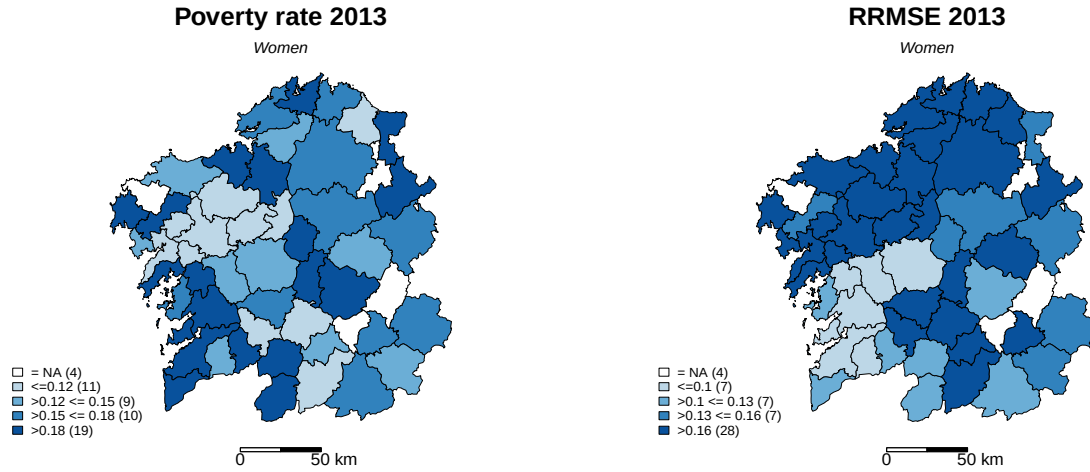


Figure F.8: SEBLUPs of poverty proportions (left) and their RRMSEs (right) for women in Galicia during 2013

References

- Pratesi, M. and Salvati, N. (2008). Small area estimation: the EBLUP estimator based on spatially correlated random area effects. *Statistical Methods and Applications*, 17:113-171.