Poisson mixed models:

applications to small area data

Miguel Boubeta Martínez

Tese de doutoramento UDC/2017

Poisson mixed models:

applications to small area data

Autor: Miguel Boubeta Martínez

Tese de doutoramento UDC/2017

Directores:

María José Lombardía Cortiña

Domingo Morales González

Departamento de Matemáticas

UNIVERSIDADE DA CORUÑA

UNIVERSIDADE DA CORUÑA

Los abajo firmantes hacen constar que son los directores de la Tesis Doctoral titulada "Poisson mixed models: applications to small area data", desarrollada por Miguel Boubeta Martínez en el ámbito del programa de doctorado de Estadística e Investigación Operativa ofertado por el Departamento de Matemáticas de la Universidade da Coruña, dando su consentimiento para que su autor, cuya firma también se incluye, proceda a su presentación y posterior defensa.

A Coruña a 7 de Abril de 2017.

Directores:

María José Lombardía Cortiña Domingo Morales González

Doctorando

Miguel Boubeta Martínez

A Vanessa,
mis padres y hermanos

Un viaje de mil kilómetros
comienza con un solo paso.

Lao-tsé

Agradecimientos

En primer lugar deseo expresar mi más sincero agradecimiento a los directores de esta tesis, la profesora María José Lombardía y el profesor Domingo Morales por la confianza depositada en mi y su apoyo. Sin su esfuerzo y dedicación, este trabajo no hubiese sido posible. Quiero extender el agradecimiento también al profesor Manuel Marey por su colaboración y al tribunal encargado de juzgar el Seminario de Tesis por su exhaustiva revisión.

También deseo agradecer a los compañeros del Departamento de Matemáticas y del café por los buenos momentos compartidos, especialmente a Inés, Bea y mi buen amigo Borja, compañero en este proyecto, por su constante ayuda y atención.

Y muy especialmente a Vanessa, a la que es justo hacer partícipe de esta tesis por su comprensión y su apoyo incondicional. Sus ánimos me han dado las fuerzas necesarias para alcanzar la meta. Muchas gracias también a mis padres, familia y amigos por su cariño y amistad. Sin vuestro grano de arena, este proyecto no se hubiese hecho realidad.

Quiero agradecer también al Instituto Galego de Estatística, al Instituto Nacional de Estadística, al Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente y al proyecto europeo SAMPLE por haber facilitado los datos para realizar este trabajo. Finalmente, a la Xunta de Galicia y al Ministerio de Economía y Competitividad por la financiación a través de los proyectos CN2012/130, MTM2011-22392 y MTM2014-52876-R.

Abstract

Small area estimation deals with the estimation of parameters in small subsets (small areas) of a global population. In the small areas, sample sizes are usually too small since designs are developed for the original population. Conventional modelling to high levels of disaggregation has too much error. Area-level Poisson mixed models are useful tools for estimating discrete response variables in small areas, since they can capture part of the variability not collected by the fixed effects. The basic Poisson mixed model is extended by incorporating first SAR(1) spatially correlated effects and second time effects. For the temporal extension, two models are considered depending on the assumed time correlation structure. The first model assumes that time effects are distributed independently, while the second model considers that they are distributed according to an AR(1) process. A spatio-temporal model including both spatial and time extensions is also studied. Each model is fitted by the method of moments and two predictors of functions of fixed and random effects are obtained: the empirical best predictor (EBP) and a plug-in predictor. Several simulation experiments are carried out for empirically analysing the behaviour of the estimators. As accuracy measure of the proposed EBPs, bootstrap mean squared error estimators are given. Finally, the developed methodology and software are applied in two fields of practical interest: poverty mapping and forest fires.

Resumo

A estimación en áreas pequenas ocúpase da estimación de parámetros en subconxuntos pequenos (áreas pequenas) dunha poboación global. Nas áreas pequenas, os tamaños mostrais habitualmente son demasiado pequenos, pois os deseños lévanse a cabo para a poboación orixinal. O modelado convencional a altos niveis de desagregación posúe demasiado erro. Os modelos mixtos de Poisson de área constitúen unha ferramenta útil para estimar variábeis resposta discretas en áreas pequenas, xa que poden capturar parte da variabilidade non recollida polos efectos fixos. O modelo de Poisson mixto básico exténdese incorporando primeiro efectos espaciais SAR(1) e segundo efectos temporais. Para a extensión temporal, considéranse dous modelos dependendo da estructura temporal asumida. O primeiro modelo supón que os efectos temporais distribúense de forma independente, mentras que o segundo considera que se distribúen de acordo a un proceso AR(1). Tamén se estuda un modelo espazo-temporal incluíndo ambas extensións espacial e temporal. Cada modelo axústase polo método dos momentos e obtéñense dous predictores: o predictor óptimo empírico (EBP) e un predictor plug-in. Lévanse a cabo varios experimentos de simulacións para analizar empiricamente o comportamento dos estimadores. Como medida de precisión dos EBPs propostos, dánse estimadores bootstrap do erro cadrático medio. Finalmente, a teoría e o software desenvolvidos aplícanse en dous campos de interese práctico: mapas de pobreza e incendios forestais.

Resumen

La estimación en áreas pequeñas se ocupa de la estimación de parámetros en subconjuntos pequeños (áreas pequeñas) de una población global. En las áreas pequeñas, los tamaños muestrales habitualmente son demasiado pequeños, pues los diseños se llevan a cabo para la población original. El modelado convencional a altos niveles de desagregación posee un elevado error. Los modelos mixtos de Poisson de área constituyen una herramienta útil para estimar variables respuesta discretas en áreas pequeñas, ya que pueden capturar parte de la variabilidad no recogida por los efectos fijos. El modelo de Poisson mixto básico se extiende incorporando primero efectos espaciales SAR(1) y segundo efectos temporales. Para la extensión temporal, se consideran dos modelos dependiendo de la estructura temporal asumida. El primero supone que los efectos temporales se distribuyen de forma independiente, mientras que el segundo considera que se distribuyen de acuerdo a un proceso AR(1). También se estudia un modelo espacio-temporal incluyendo ambas extensiones espacial y temporal. Cada modelo se ajusta por el método de los momentos y se obtienen dos predictores: el predictor óptimo empírico (EBP) y un predictor plug-in. Se llevan a cabo varios experimentos de simulación para analizar empíricamente el comportamiento de los estimadores. Como medida de precisión de los EBPs propuestos, se dan estimadores bootstrap del error cuadrático medio. Finalmente, la teoría y el software desarrollados se aplican en dos campos de interés práctico: mapas de pobreza e incendios forestales.

Contents

List of Tables

List of Figures

List of Abbreviations

AIC	Akaike information criterion
BLUP	Best linear unbiased predictor
BP	Best predictor
EBLUP	Empirical best linear unbiased predictor
EBP	Empirical best predictor
FH	Fay-Herriot
FS	Fisher scoring
GLM	Generalized linear models
GLMM	Generalized linear mixed models
GREG	Generalized regression
iid	Independent and identically distributed
LA	Laplace
LCS	Living condition survey
LGREG	Logistic generalized regression
LM	Linear model
LMM	Linear mixed model
ML	Maximum likelihood
MM	Method of moments
MSE	Mean squared error
MSM	Method of simulated moments
NER	Nested error regression
NR	Newton-Raphson
p.d.f.	Probability density function
PQL	Penalized quasi-likelihood
REML	Restricted maximum likelihood
SAE	Small area estimation
SLCS	Spanish living condition survey

Chapter 1

Introduction

Contents

1.1 Linear mixed models

The linear regression model (LM) can be expressed as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1.1.1}$$

where $\boldsymbol{y}$ is the response variable, $\boldsymbol{X}$ is the design matrix of known covariates, $\boldsymbol{\beta}$ is the vector of unknown regression coefficients and $\boldsymbol{\varepsilon}$ is the vector of errors. Under this model, the regression coefficients $\boldsymbol{\beta}$ are fixed. However, in practice there are many cases where the observations are correlated, and then it makes sense to assume that some of these coefficients are random. For example, in animal breading and medical experiments, data are usually collected from the same individuals over time. Then, it is necessary to take into account the existence of a correlation structure among the observations from a same individual.

Linear mixed models (LMMs) treat this type of correlation among observations. This

methodology brings statistics to the next level. That is to say, unlike the classic statistics where it is assumed that the observations belong to the same general population and are independent and identically distributed (iid), mixed model data have a more complex multilevel structure (Demidenko, 2004). LMMs may assume that observations between levels are independent, but observations within each level are dependent since they share information of the subpopulation. They deal with two sources of information: between levels and within levels. Overviews on this topic are provided by McCulloch et al. (2008); Jiang (2007) and Demidenko (2004), among others.

A contemporary application of mixed models is the analysis of longitudinal data (Diggle et al., 2002), where each time series represents a level, but they can be also used to treat data with multiple sources of variation, repeated measures, biological variety and heterogeneity, and image reconstruction problems.

Let us consider a practical example regarding economic studies (Demidenko, 2004), to show how a LMM is useful for modelling the correlation structure of the observations. It deals with the relationship between the response variable sales (y) and the auxiliary variable price (x) of several products. Figure 1.1.1 (left) presents a scatter plot of the observations $\{(x_k, y_k),\, k = 1, \ldots, n\}$, where n represents the sample size. It revels that the relation between x and y is close to the linearity and is inverse, that is to say the slope is negative. Classical linear regression models assume that the pairs of iid observations (x_k, y_k) can be modelled as

$$y_k = u + \beta x_k + \varepsilon_k, \quad k = 1, \ldots, n,$$

where $\{\varepsilon_k\}$ are iid random variables with zero mean and constant variance σ_ε^2. However, one can assume that data are clustered in factory products or clusters. In Figure 1.1.1 (right), data representing the same product are connected. This figure suggests an opposite behaviour to that observed in Figure 1.1.1 (left). That is, an increase in price leads to an increase in sales.

LMMs assume that each product d, $d = 1, \ldots, D$, has associated an unobservable specific random effect, i.e.

$$y_{dj} = u_d + \beta x_{dj} + \varepsilon_{dj}, \quad d = 1, \ldots, D,\ j = 1, \ldots, n_d, \tag{1.1.2}$$

where y_{dj} and x_{dj} denote the response and the auxiliary variable of the jth observation of the dth product, u_d is the product-specific intercept, n_d is the sample size of the dth product and ε_{dj} are the error terms. As in the previous case, we assume that they are iid

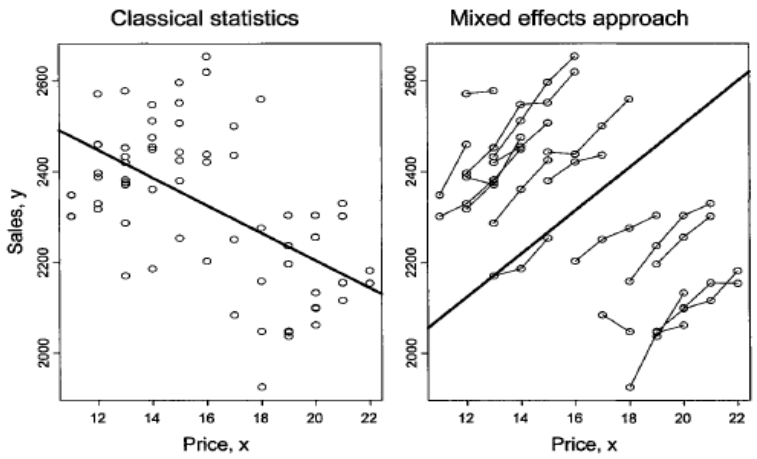

Figure 1.1.1: Relationship between the response variable sales (y) and the covariate price (x) in several products. The results under classical statistics (left) shows that the relationship between observations (iid) is negative, while under the mixed effects approach (right), where it is assumed that each product represents a cluster, the relationship is positive, i.e. an increase in price leads to an increase in sales. Source of information: Demidenko (2004).

with zero mean and variance σ_ε^2. The random effects u_d, $d = 1, \ldots, D$, are assumed to be iid with mean zero and variance σ_u^2, and independent of the error terms $\{\varepsilon_{dj}\}$. Therefore, the correlation between two observations of the same individual (product) is

$$\rho = \frac{\text{var}(u_d)}{\text{var}(u_d + \varepsilon_{dj})} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2},$$

while observations from different individuals are uncorrelated. The sample size n_d of each product d, $d = 1, \ldots, D$, fulfills $\sum_{d=1}^{D} n_d = n$.

The LMM (1.1.2) can be expressed in matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon}, \tag{1.1.3}$$

where $\boldsymbol{y}$ is the response vector of observations, $\boldsymbol{X}$ is the design matrix of covariates (known), $\boldsymbol{\beta}$ is the vector of unknown regression coefficients, $\boldsymbol{Z}$ is the incidence matrix (known), $\boldsymbol{u}$ is the vector of random effects and $\boldsymbol{\varepsilon}$ is the error vector of random perturbations. The difference between the LMM (1.1.3) and the LM (1.1.1) is $\boldsymbol{Z}\boldsymbol{u}$. The basic assumptions for the LMM (1.1.3) are that both random effects and errors have mean zero and finite variances. If $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ denote the covariances of $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ respectively, then the covariance

matrix of the response vector is

$$\boldsymbol{V} = \text{var}\left[\boldsymbol{y}\right] = \boldsymbol{Z}\boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{Z}^t + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}.$$

The covariance matrix of the random effects, $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, depends on some parameters called *variance components*.

Historically, the maximum likelihood (ML) and the restricted maximum likelihood (REML) estimators have become the most common strategies for estimating the model parameters under the LMMs.

The ML method was first used by Hartley and Rao (1967). The main reason for the delay in adapting the ML method to mixed models is because the estimation of the variance components was not easy to handle computationally in the old days (Jiang, 2007).

Under a Gaussian mixed model, the distribution of the response variable $\boldsymbol{y}$ is $N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$, and then its probability density function (p.d.f.) is

$$f(\boldsymbol{y}) = \frac{1}{(2\pi)^{n/2}\left|\boldsymbol{V}\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)\right\},$$

where n is the dimension of $\boldsymbol{y}$. The log-likelihood of $\boldsymbol{y}$ under the LMM is

$$\ell(\boldsymbol{\beta}, \boldsymbol{V}) = -\frac{1}{2}\left\{n\,\log(2\pi) + \log\left|\boldsymbol{V}\right| + \left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}\right)\right\}. \tag{1.1.4}$$

Then, the ML estimate of $(\boldsymbol{\beta}, \boldsymbol{V})$ is the one that maximizes the log-likelihood (1.1.4). Given $\boldsymbol{V}$, $\ell(\boldsymbol{\beta}, \boldsymbol{V})$ is maximized over $\boldsymbol{\beta}$ by

$$\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}. \tag{1.1.5}$$

For $\boldsymbol{y}$ having a general distribution, the estimator (1.1.5) can be shown as the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ (Rupert et al., 2008). On the other hand, if $\boldsymbol{y}$ is multivariate normal, then (1.1.5) is also the uniformly minimum variance unbiased estimator (UMVUE).

By substituting (1.1.5) in (1.1.4), we obtain the profile log-likelihood for the covariance $\boldsymbol{V}$ (see p. 101 in Rupert et al. (2008) for more details), i.e.

$$\begin{aligned}\ell_P(\boldsymbol{V}) &= -\frac{1}{2}\left\{\log|\boldsymbol{V}| + \left(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}\right)'\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}\right) + n\,\log(2\pi)\right\} \\ &= -\frac{1}{2}\left\{\log|\boldsymbol{V}| + \boldsymbol{y}'\boldsymbol{V}^{-1}\left[\boldsymbol{I}-\boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\right]\boldsymbol{y}\right\} - \frac{n}{2}\log(2\pi).\end{aligned} \tag{1.1.6}$$

The ML estimates of the variance components involved in $\boldsymbol{V}$ are obtained by maximizing the profile log-likelihood (1.1.6) over those parameters. As there is no closed expression, it is required the use of iterative methods. Among others, it is common to use Newton-Raphson (NR) or Fisher scoring (FS) algorithms.

The REML criterion (Searle et al., 1992) involves maximizing the restricted log-likelihood

$$\ell_R(\boldsymbol{V}) = \ell_P(\boldsymbol{V}) - \frac{1}{2}\log\left|\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right|.$$

The main advantage of REML criterion over the ML estimator is that the first one takes into account the degrees of freedom of $\boldsymbol{\beta}$ in the model. For small sample sizes, REML is preferable to ML, since ML is biased for the estimation of variance components, but for large sample sizes, the difference between the two approaches is negligible.

The vector of random effects $\boldsymbol{u}$ can be predicted through its best linear unbiased predictor (BLUP). The BLUP of the random effects is

$$\tilde{\boldsymbol{u}} = \boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{Z}'\boldsymbol{V}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}\right). \tag{1.1.7}$$

In practice, as the BLUE of $\boldsymbol{\beta}$ and the BLUP of $\boldsymbol{u}$ depend on the theoretical covariances $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, $\tilde{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{u}}$ are replaced by $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{u}}$, where

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}, \\ \hat{\boldsymbol{u}} &= \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\boldsymbol{Z}'\hat{\boldsymbol{V}}^{-1}\left(\boldsymbol{y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}\right).\end{aligned}$$

They are obtained replacing in (1.1.5) and (1.1.7), the theoretical covariances $\boldsymbol{V}$ and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ by their estimates, for example using ML or REML estimation procedures. We refer to the estimator $\hat{\boldsymbol{\beta}}$ and the predictor $\hat{\boldsymbol{u}}$ as empirical BLUE (EBLUE) and empirical BLUP (EBLUP) of $\boldsymbol{\beta}$ and $\boldsymbol{u}$ respectively.

1.2 Generalized linear mixed models

The generalized linear models (GLMs) extend the classical LMs (1.1.1) of normal response variables. The generalization is done in two senses:

- The distribution of the response $\boldsymbol{y}$ is allowed to be different from normal. Specifically, it is assumed that the distribution belongs to the exponential family.
- A function not necessarily linear of the mean of the response variable $\boldsymbol{y}$ is modelled linearly, i.e.

$$g\left(\mathbb{E}\left[\boldsymbol{y}\right]\right) = \boldsymbol{X}\boldsymbol{\beta},$$

where g is the link function and $\boldsymbol{X}\boldsymbol{\beta}$ is the linear predictor.

The GLMs include as particular cases, among others, the LMMs, analysis of variance, logistic regression and Poisson regression. GLMs can be derived from the exponential family distribution with probability density function

$$f_{\boldsymbol{\theta},\phi}(\boldsymbol{y}) = f(\boldsymbol{y};\boldsymbol{\theta},\phi) = \exp\left\{\frac{\boldsymbol{y}\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\phi)} - c(\boldsymbol{y},\boldsymbol{\theta})\right\},$$

where $a(.)$, $b(.)$ and $c(.)$ are specific functions of each distribution, $\phi \in \mathbb{R}$, $a(\phi) > 0$ and $\boldsymbol{\theta} \in \mathbb{R}^p$. The parameter $\boldsymbol{\theta}$ is called the linear predictor and ϕ the scale parameter. The logistic and Poisson mixed models can be obtained with specific forms of the exponential p.d.f. $f_{\boldsymbol{\theta},\phi}(\boldsymbol{y})$ and the link function g.

The generalized linear mixed model (GLMM) is an extension of the GLM that incorporates random effects. In recent years, GLMMs have achieved great popularity for modelling binary/count clustered and longitudinal data. Assuming that the response variables, y_{dj}, conditional to the random effects $\boldsymbol{u}_d$, are independent and belong to the exponential family distribution with the unit scale, where $\boldsymbol{u}_d \sim N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$, the log-likelihood under GLMM (Demidenko, 2004) takes the form

$$\begin{aligned}\ell(\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{u}}) = &-\frac{Dk}{2}\log(2\pi) - \frac{D}{2}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{u}}\right| \\ &+ \sum_{d=1}^{D}\log\int_{\mathbb{R}^k}\exp\left\{\ell_d(\boldsymbol{\beta},\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}'\boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1}\boldsymbol{u}\right\}d\boldsymbol{u},\end{aligned} \tag{1.2.1}$$

where

$$\ell_d(\boldsymbol{\beta},\boldsymbol{u}) = \sum_{j=1}^{n_d}\left[\left(\boldsymbol{\beta}'\boldsymbol{x}_{dj} + \boldsymbol{u}'\boldsymbol{z}_{dj}\right)y_{dj} - b\left(\boldsymbol{\beta}'\boldsymbol{x}_{dj} + \boldsymbol{u}'\boldsymbol{z}_{dj}\right)\right]$$

is the dth conditional log-likelihood. The ML estimator of the vector of parameters of a GLMM maximizes the log-likelihood (1.2.1). The ML estimator is consistent when the number of levels or clusters, D, goes to infinity, while the number of observations per cluster, n_d, remains uniformly bounded. An important advantage of the ML approach is that generates estimates of all model parameters (fixed effects and variance parameters) and predictions of the random effects. To maximize the log-likelihood function (1.2.1), Newton-Raphson, Fisher scoring or EM algorithms can be used. The updating equation of the first two algorithms can be expressed in the generic form

$$\hat{\boldsymbol{\beta}}_{s+1} = \hat{\boldsymbol{\beta}}_s + \lambda_s \boldsymbol{H}_s^{-1} \left(\frac{\partial \ell}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}_s} \right),$$

where $0 < \lambda_s \leq 1$ is a step length. The matrix $\boldsymbol{H}$ is the negative Hessian under NR and the expected negative Hessian under FS algorithm. It is an integral that requires numerical calculation. Integral approximation techniques as Laplace (LA), penalized quasi-likelihood (PQL), adaptive Gauss-Hermite, Monte Carlo methods and numerical integration (trapezoid, Simpson rules, etc.) can be used in combination with iterative algorithms to obtain an ML estimator (Pinheiro and Bates, 1996; Breslow and Clayton, 1993; Demidenko, 2004).

In some cases, ML-based estimators may lead to inconsistent and biased estimators. Jiang (1998) proposes the method of simulated moments (MSM) as an alternative approach to estimate the parameters in GLMMs. This method is computationally attractive and gives consistent estimators of the model parameters. This is why, Chapters 2-5 apply the MSM method to estimate the parameters of the considered Poisson mixed models.

1.3 Small area estimation

Small area estimation (SAE) is a branch of statistics involving the estimation of parameters in small subsets (called small areas or domains) of an original population. A small area usually refers to a small geographic area (such as a county, a municipality or a census division), a demographic group (e.g. a specific age × sex × race group), a demographic group within a geographic region, etc. Frequently, sample sizes within small areas are too small since sampling designs are developed for the original population. In this context, the estimators of population parameters have the desired precision at population level but not at the domain level.

The direct estimators of domain parameters use only the data information of the considered domain. Horvitz and Thompson (1952) introduced a simple design-unbiased direct estimator of a domain mean as a sum of the sample values of the target variable multiplied by the sampling weights. By using weight calibration and non response correction methods, many new weighted direct estimators can be found in the literature. The direct estimators do not use cross-sectional or temporal data. However, they are basically unbiased with respect to the sampling design distribution but with a big variance in small area estimation problems. Their estimated variances and coefficients of variation are usually greater than the ones of other more sophisticated estimators. The finding of these improved estimators is one of the main target of SAE researchers.

The first reviews on small area estimation focus on demographic methods for estimating the population in post-census periods. Purcell and Kish (1979) propose statistical methods for estimating small-domain characteristics. In the field of biomedicine, the use of maps to study disease patterns in small areas has been used for a long time (Marshall, 1991). More recently, we can find a great number of works that describe and detail exhaustively the existing theory of small area estimation. Among them, the review papers of Rao (1986, 1999, 2008), Ghosh and Rao (1994), Pfeffermann (2002), Lahiri and Meza (2002), Jiang and Lahiri (2006), Pfeffermann (2013), and the monographs of Muckhopadhyay (1998), Rao (2003) and Rao and Molina (2015).

Small area estimation techniques can be divided into three types of methodologies: design-based, model-assisted and model-based methods. The three branches introduce and study estimators that are competitors of the direct estimators. The design-based approach to small area estimation looks for indirect estimators (basic synthetic, post-stratified, sample size dependent and son on) with good properties with respect to the sampling design distribution. They employ auxiliary information from external data sources (from outside of the target domain), but they do not rely explicitly on models. For example, if the population sizes of domain crossed by sex-age groups are available from external data registers, then this information can be used for evaluating estimators that could provide better estimates than the direct ones. The design-based indirect estimators are optimized with respect to the sampling distribution. Design-based methods often use implicit models, although the bias and the variance of estimators are calculated with respect to the sampling design distribution (Lehtonen and Veijanen, 2009; Pratesi, 2016). Some indirect estimators using implicit models are discussed in Rao (2003).

In the case of having auxiliary data related to the target variable, it is possible to obtain

better accuracy for domain estimates by using explicit models, when compared to an estimation procedure not using auxiliary data. The model-assisted methodology considers the properties under the design-based distribution, but employs explicit models to motivate the choice of estimators. Important examples of model-assisted estimators are the generalized regression (GREG) estimator and the calibration estimator introduced by Deville and Sarndal (1992) or the LGREG estimator introduced by Lehtonen and Veijanen (1998). The GREG estimator uses a linear model as an assisting tool, estimates the domain mean of a continuous variable and is constructed to be design-unbiased (or approximately so) irrespectively of the fit of the model to data. The LGREG is a domain proportion estimator assisted by a logistic regression model. Different types of auxiliary data can be used in model assisted estimation. The GREG and the LGREG estimators employ auxiliary variables from survey files and their aggregated from administrative registers. They are estimators with a good balance between properties related with the design-based and the model-based distributions.

The model-based approach assumes that the data is generated by a true model and therefore the inferences should be based on it. The use of explicit models in SAE gives an idea of how different sources of information are combined (Fuller, 1975; Fay and Herriot, 1979; Holt et al., 1979; Datta, 2009). This approach can introduce estimators that may employ cross-sectional and temporal auxiliary information and that can take into account for temporal and spatial correlation. The estimator has optimal properties with respect to the true model distribution. An important issue of this approach is the selection and the diagnostics of the selected model. A model with a good fit to data guarantees good model-based estimators with lower mean squared errors than the design based estimators.

Based on the level of aggregation of the response variable, the small area models can be classified into two groups: (i) area level-models and (ii) unit-level models. The basic SAE unit-level model is the nested error regression (NER) model. Battese et al. (1988) applied this model to the prediction of United States county crop areas using survey and satellite data. Since then, the empirical best linear unbiased predictors (EBLUP) of domains means based on the NER model are being widely applied. Concerning poverty estimation, Molina and Rao (2010) derived empirical best predictors (EBP) of non linear parameters based on the NER model, with applications to the estimation of poverty incidences and gaps. Hobza and Morales (2016) studied EBPs of poverty incidences based on unit-level logit mixed models. Tzavidis et al. (2008) and Marchetti et al. (2012) gave M-quantile estimators for poverty mapping. The SAE literature on unit-level model-based methods covers many other estimators employing non parametric, robust or Bayesian regression procedures.

The basic area-level model is the Fay-Herriot model. Fay and Herriot (1979) used an area level linear mixed model to estimate the per-capita income in small places of U.S. Several generalization of the Fay-Herriot model have been applied to poverty estimation. For example, Esteban et al. (2012a,b), Marhuenda et al. (2013) and Morales et al. (2015) gave EBLUPs of Spanish poverty proportions based on temporal and spatio-temporal linear mixed models. Concerning GLMMs, Boubeta et al. (2016b, 2017b) and López-Vizcaíno et al. (2013, 2015) introduced EBPs of counts and proportions based on Poisson and multinomial-logit area-level mixed models with applications to Spanish data and Chandra et al. (2017) gave small area predictors of counts under a non-stationary spatial GLMM model.

The methodological developments achieved in this manuscript are obtained under the group (i) of area-level models. In practice there is a large number of applications in area-level models. One of the main reasons is due to the secrecy of confidentiality. Statistical offices usually can not display data at the individual level but they can disseminate it aggregated by regions. Area-level models are used in socioeconomic, environment, biology and health sciences, among others. In addition, they have evolved over time adapting to the needs of the data (including temporal effects or spatial correlation), giving rise to increasingly flexible models but, on the other hand, more difficult to estimate.

As we say above, the Fay-Herriot (FH) model is the basic model of the area-level approach to small area estimation. This model is defined in two stages

- Level 1 (sampling model): $y_d|\mu_d \overset{ind}{\sim} N(\mu_d, \sigma_d^2)$, $d = 1, \ldots, D$;
- Level 2 (linking model): $\mu_d \overset{ind}{\sim} N(\boldsymbol{x}_d\boldsymbol{\beta}, \sigma_u^2)$, $d = 1, \ldots, D$,

where D denotes the total number of areas or domains, d denotes a particular domain and the variances σ_d^2 are assumed to be known. Usually, in the FH model the response variable, y_d, is an estimator of the sample mean of individuals belonging to the domain d. The previous model can be expressed as an area-level linear mixed regression model, i.e.

$$y_d = \mu_d + \varepsilon_d = \boldsymbol{x}_d\boldsymbol{\beta} + u_d + \varepsilon_d,\ d = 1, \ldots, D,$$

where u_d's and ε_d's are independent with $u_d \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ and $\varepsilon_d \overset{ind}{\sim} N(0, \sigma_d^2)$, and $\boldsymbol{x}_d = (x_{d1}, \ldots, x_{dp})$ is the row vector containing the values of the aggregated auxiliary variables at domain d. The FH model can be written in the LMM matrix form

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{y} = \underset{1\leq d\leq D}{\text{col}}(y_d)$, $\boldsymbol{X} = \underset{1\leq d\leq D}{\text{col}}(\boldsymbol{x}_d)$, $\boldsymbol{Z} = \boldsymbol{I}_D$, $\boldsymbol{u} = \underset{1\leq d\leq D}{\text{col}}(u_d)$ and $\boldsymbol{\varepsilon} = \underset{1\leq d\leq D}{\text{col}}(\varepsilon_d)$. Here, $\boldsymbol{I}_a$ denotes the $a \times a$ identity matrix and $\underset{1\leq d\leq D}{\text{col}}$ denotes the column operator.

A popular estimator of μ_d, $d = 1, \ldots, D$, under the FH model is the best linear unbiased predictor. The BLUP of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_D)'$ is the one that minimizes the mean squared error (MSE) in the set of unbiased predictors and it is given by

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{X}\tilde{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{V}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}\right), \tag{1.3.1}$$

where $\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{y}$, $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \text{var}(\boldsymbol{u})$ and $\boldsymbol{V} = \text{var}(\boldsymbol{y})$. The EBLUP of $\boldsymbol{\mu}$ is obtained by substituting in (1.3.1) the unknown covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ and $\boldsymbol{V}$ by $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}$ and $\hat{\boldsymbol{V}}$ respectively, i.e.

$$\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}\hat{\boldsymbol{V}}^{-1}\left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right),$$

where $\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^t\hat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}$. Another estimator commonly used in practice is the plug-in estimator. It is given by

$$\hat{\boldsymbol{\mu}}^P = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}},$$

where $\hat{\boldsymbol{u}}$ denotes a predictor of the vector of random effects $\boldsymbol{u} = (u_1, \ldots, u_D)'$. Further, $\hat{\boldsymbol{\mu}}^P$ is the EBLUP of $\boldsymbol{\mu}$ if $\hat{\boldsymbol{u}}$ is the EBLUP of $\boldsymbol{u}$.

A simplified version of the plug-in estimators are the synthetic estimators, which use the knowledge of the population structure to improve the efficiency of estimators based only on the sample design. They assume a linear model for the data and then, the values that have not been sampled, are estimated from the implicit model using only the information of the available covariates, i.e.

$$\hat{\boldsymbol{y}}^{syn} = \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

Note that these estimators do not use the random effects, that is, they assume no variations between areas. Under implicit models, frequentist and Bayesian approaches can be used for inference (Datta and Ghosh, 1991; Ghosh et al., 1998).

The FH model can be extended to the GLMM context by assuming that the distribution of the response variable y_d belongs to the exponential family and that its expectation, transformed by the link function g, is modelled linearly, i.e.

$$g(\mathbb{E}[y_d]) = \boldsymbol{x}_d\boldsymbol{\beta} + u_d,\ d = 1, \ldots, D.$$

In particular, this manuscript deals with the calculation of the EBP of $\mathbb{E}[y_d]$ under area-level

Poisson mixed models, with the natural logarithmic link function. Therefore, the target variable counts events of interest by domains. The EBP is the natural extension of the EBLUP from LMMs to GLMMs.

1.4 Description of the databases

This section introduces the three databases that are used throughout the document. Two of them are of socio-economic nature while the other one is of environmental type. These data sets are employed to illustrate the methodology developed in this manuscript.

1.4.1 Socio-economic databases

Almost every fourth person was at risk of poverty or social exclusion in the European Union (EU) during 2014 (Eurostat, 2016). 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. This survey has a planned sample size large enough for obtaining reliable direct estimates for autonomous communities, but not for provinces or counties. Small area estimation deals with this kind of problems by introducing indirect estimators.

The poverty line is defined as a percentage (currently Eurostat fixed it to 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. 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 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.

Both employed socio-economic datasets record information from the SLCS. The first one collects the information at national level by provinces in 2008, while the second one is focused on the counties of the autonomous community of Galicia (located in the North-West of Spain) during $2010 - 2013$. The datasets at the individual level are provided by the Instituto Nacional de Estadística (INE) and the Instituto Galego de Estatística (IGE) respectively, and the area-level aggregation is of own elaboration.

Figure 1.4.1 shows the evolution of the poverty rates for men (left) and women (right) by age ranges in Spain from 2004 to 2012. Since the beginning of the crisis in 2008, the poverty proportion has increased significantly, except for people over 65, which is the age range that behaves better with respect to poverty.

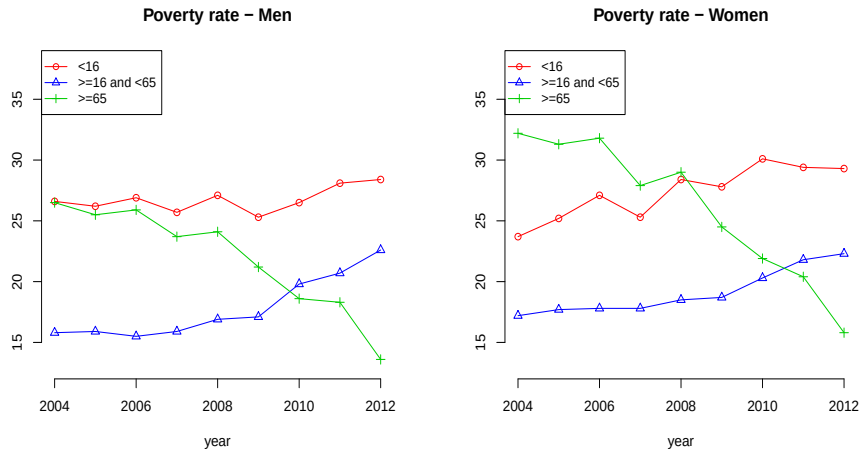

Figure 1.4.1: Poverty rate for men (left) and women (right) by age ranges in Spain from 2004 to 2012. Source of information: INE.

On the other hand, Figure 1.4.2 compares poverty rates in Galicia and Spain by sex. The results for men are represented by solid lines, while for women are in dashed lines. Until 2008, the poverty rate is approximately constant in both Galicia and Spain. In the following years, the proportion of people at risk of poverty increases significantly. Regarding the comparison with Spain, the poverty rate is close to 20% in Spain and to 15% in Galicia. In addition, the poverty rate is higher for women.

Both databases contain aggregated information regarding the household income at area level. The small areas (domains) in both databases are the Spanish provinces (50 provinces) and the Galician counties (53 counties), or the provinces and the counties crossed by sex, respectively. The first dataset also considers the autonomous cities of Ceuta and Melilla.

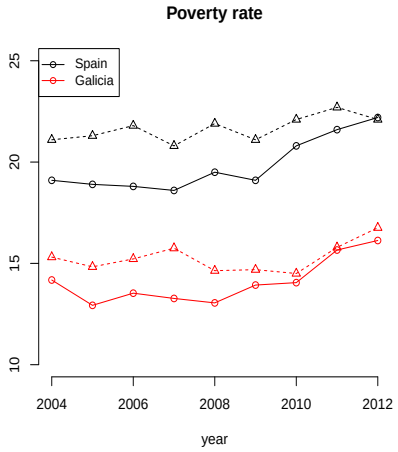

Figure 1.4.2: Poverty rate in Galicia and Spain from 2004 to 2012 (— for men and - - - for women). Sources of information: INE and IGE.

The domain sample sizes and totals are n_d and $y_d = \sum_{j \in s_d} y_{dj}$ respectively, i.e. y_d counts the number of people (by sex) under the poverty line in the domain sample s_d, $d = 1, \ldots, D$ ($D = 50$ provinces or $D = 53$ counties depending on the dataset). The available auxiliary variables in both databases are the area proportions of people (by sex) in the categories of the following classification variables.

- Age: ≤ 15 ($age0$), $16 - 24$ ($age1$), $25 - 49$ ($age2$), $50 - 64$ ($age3$) and ≥ 65 ($age4$).
- Education: less than primary ($edu0$), primary ($edu1$), secondary ($edu2$), university ($edu3$).
- Citizenship: Spanish ($cit0$), not Spanish ($cit1$).
- Labour situation: ≤ 15 ($lab0$), employed ($lab1$), unemployed ($lab2$), inactive ($lab3$).

As the proportions of people in the categories of classification variables sum up to one, we take the reference categories out of the data file of auxiliary variables. The reference categories are $age0$, $edu0$, $cit0$ 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 labour sector. As these people are in group $edu2$, we merge secondary and university education levels into a single category $edu23$. This proposal, suggested by a Spanish Office of Statistics (INE), is only carried out in the first database.

1.4.2 Forest fires database

The forest area of Galicia is $2,060,453$ hectares (ha). It represents 69% of the region, which makes it one of the Spanish communities with more woodland. Several authors have studied the characteristics of forest owners and the productive capacity of the region (Marey-Pérez et al., 2006; Marey-Pérez and Rodríguez-Vicente, 2008; Rodríguez-Vicente and Marey-Pérez, 2009a, 2010; Marey-Pérez et al., 2012). They concluded that there are many owners with small and very productive plots and a significant presence of collective forest land and no presence of public forest ownership. The first problem of the forestry sector is forest fires: there were $249,387$ wildfires registered since 1968, the year in which forest fire statistics started, until December 2012. These fires swept an area of $1,794,578$ ha, equivalent to 61% of the geographical area of the region.

The original forest fires database is provided by the *Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente* of the Spain Government (MAPAMA, 2017), and the area-level aggregation is of own elaboration. It contains a total of $85,134$ fire events registered in the database for the period $1999-2008$, regardless of their size. In addition to the spatial location and the date of occurrence of the ignition points, we use two marks: burned area and cause.

The alphanumerical information about the wildfires registered in the study area corresponded to the ignition point coordinates, which were translated to the actual land area with the aid of GIS. Subsequent data quality control confirmed the information about the attributes of the burned area (forest species composition, parish and land use) with existing data from the area in the year of fire. Therefore, Ignition Point UTM (Universal Transverse Mercator) coordinates were available for each fire and other measures of interest attached to these coordinates. These measures are related to burned material, vegetation type, fire behaviour, fire extinction, fire damage and possible fire causes.

Figure 1.4.3 plots the annual totals of forest fires occurrences in Spain, Galicia and rest of Spain during the period $1980-2008$. These data are taken from the web of *Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente* of the Spain Government (MAPAMA, 2017), and show that there have been as many fires in Galicia as in the rest of Spain during this period.

The employed database aggregate the information given in the original database by forest area and month. Forest areas constitute an administrative structure of the fire-fighting system since 1999. Currently, Galicia is divided into 63 forest areas. For each area, our data

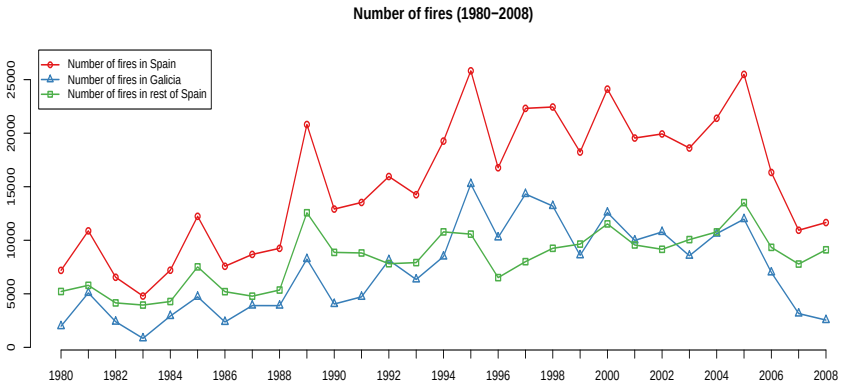

Figure 1.4.3: Fires in Spain, Galicia and rest of Spain during 1980-2008.

set contains the numerical values of the target variable *number of fires* and some auxiliary variables from 2006 to 2008. We consider two sources of auxiliary information depending on their structure. First, we consider information as population size (pop), number of cadastral parcels (par), number cadastral holders ($cadHold$), livestock units (lu), percentage of land with scrub or bush vegetation ($scrub$), percentage of wet lands (wet) and percentage of wood lands ($wood$) per forest areas. These auxiliary variables only depends on the areas in the considered period, i.e. they are constant over time. Second, we take average measurements of meteorological stations per month and forest areas. Specifically, we have accumulated rain ($acumRain$), average temperature ($averTemp$) and days without rain (dwr).

1.5 Overview of the manuscript: structure and contributions

This manuscript deals with small area estimation techniques by using area-level generalized linear mixed models. Specifically, we consider count response variables and, in consequence, our family of models is restricted to area-level Poisson mixed models. Several extensions of the basic area-level Poisson mixed model are considered incorporating spatial correlation, temporal components and spatio-temporal correlation. Different targets are addressed. The first target is developing algorithms for estimating the model parameters. We focus on the method of moments (MM), which is based on the method of simulated moments proposed in Jiang (1998). This method is a competitive alternative, since it is computationally attractive and gives consistent estimators of model parameters. The second target is obtaining the EBP of functions of fixed and random effects under the Poisson mixed models. We further compare the proposed predictor against other existing predictors in the literature such

as plug-in or direct estimators. We also provide error measures for the proposed predictors. Throughout the manuscript, two applications to real data are simultaneously carried out to illustrate the methodology. The first one has a socio-economic scope and the second one has an environmental character. Regarding the socio-economic applications, for the simplest area-level Poisson mixed model we use the Spanish socio-economic database. However, we use the Galician socio-economic database for the more complex Poisson models.

This work is organized as follows. Chapter 2 introduces the basic area-level Poisson mixed model and the employed fitting algorithms. Specifically, we consider the method of moments, Laplace and PQL. For estimating the target domain parameters, we propose the EBP and a plug-in predictor. As error measure, we introduce three MSE estimators: two analytical approximations and a bootstrap-based approach. Analytical approximations of the MSE are only proposed in this chapter due to their computational complexity. The results of this chapter are collected in the statistical journal *Test* (Boubeta et al., 2016b) and in the *Journal of Environmental Management* (Boubeta et al., 2015). The basic model is extended to SAR(1) spatial correlation in Chapter 3 and to temporal correlation in Chapter 4, giving rise to the work in *Computational Statistics and Data Analysis* (Boubeta et al., 2017b). For the temporal model, we consider independent and AR(1)-correlated time effects. Chapter 5 introduces a general model incorporating a spatio-temporal correlation structure. For each generalization of the basic area-level Poisson mixed model, we obtain the MM estimator for the vector of model parameters and the EBP and a plug-in predictor of the domain parameters of interest (counts and proportions). The behaviour of the proposed estimators and predictors are empirically investigated through simulation studies. As error measure for these EBPs, we propose algorithms based on parametric bootstrap procedures. Finally, Chapter 6 collects the main conclusions of this manuscript.

Chapter 2

The area-level Poisson mixed model

Contents

2.1 Introduction

Poisson regression and binomial-logit models are GLMs that are used for counts, i.e. for target variables counting some event of interest (like number of people under poverty line by provinces or number of fires by forest areas). In these models the assumption of linearity is relaxed in the sense that a function, called link function, of the mean of the observations is linear in some set of covariates. The normality assumption is also relaxed by assuming 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 happen that observations from different domains are

independent, but observations within the same domain are dependent because they share common properties. The GLMMs are extensions of GLMs 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 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 chapter studies an area-level Poisson mixed model. Three procedures are used for fitting the proposed GLMM: the method of moments, the PQL and ML-Laplace algorithms. Especial attention is paid to the first procedure, 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 latter two are used only in this chapter for comparative purposes. These algorithms are programmed by using functions implemented in the statistical software R. However, the application of these functions is restricted to some basic mixed models and do not cover all the fitting procedures.

Empirical best predictors, based on area-level Poisson mixed models, are derived 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 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 the 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 behaviour of the MSE estimators.

For estimating small area counting indicators, area-level versions of GLMM with logit link function, and with combination of 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 and give an analytical approximation to the 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 behaviour of the estimators introduced in this chapter (EBP or plug-in based on the Poisson model). Nevertheless, we also include the well known EBLUP based on the Fay-Herriot model.

The chapter is organized as follows. Section 2.2 introduces the area-level Poisson mixed model and the employed fitting algorithms (MM, PQL and ML-Laplace) are described in Section 2.3. Section 2.5 presents the EBP and the plug-in estimators of functions of fixed and small area specific random effects. Section 2.6 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 a parametric bootstrap. Section 2.8 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 first real data study case. Section 2.9 applies the developed methodology to data from the SLCS in 2008 and from forest fires in Galicia during 2007. Finally, Section 2.10 contains the main conclusions of

this chapter.

2.2 The model

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

$$f_v(\boldsymbol{v}) = (2\pi)^{-D/2} \exp\left\{-\frac{1}{2}\,\boldsymbol{v}'\boldsymbol{v}\right\}.$$

We assume that the distribution of the target variable y_d, conditionally on the random effect v_d, is

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

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 (known) size variable and p_d is the binomial probability. As ν_d is known, p_d is the target parameter since it univocally determines the Poisson parameter μ_d. Consequently, we focus on obtaining an estimator of p_d.

For the natural parameter, we assume

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

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

$$\mathbb{P}(y_d|\boldsymbol{v}) = \mathbb{P}(y_d|v_d) = \frac{1}{y_d!}\exp\{-\nu_d p_d\}\nu_d^{y_d} p_d^{y_d},$$

where $p_d = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}$. The probability function of the response variable $\boldsymbol{y}$ is given by

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

where

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

and $c = (2\pi)^{-\frac{D}{2}}\exp\left\{\frac{-\boldsymbol{v}'\boldsymbol{v}}{2}\right\}$.

The previous model, Model 1, can be seen as a generalization of the classical Poisson model without random effects, i.e.

$$\text{Model 0:}\quad \eta_d = \log\mu_d = \log\nu_d + \boldsymbol{x}_d\boldsymbol{\beta},\quad d = 1,\ldots,D.$$

Model 0 corresponds to a particular case of Model 1. Both match when the variance parameter ϕ in Model 1 is equal to zero.

2.3 Fitting algorithms

To fit the area-level Poisson mixed model, we consider three alternatives: the method of moments (MM), the penalized quasi-likelihood (PQL) and the ML-Laplace (LA) approximation to the likelihood.

The LA algorithm works in two steps iteratively. Step 1 approximates the model log-likelihood and Step 2 maximizes the approximated log-likelihood. The output of the algorithm gives approximate ML estimators of the model parameters and predictions (modes) of the random effects. This is an efficient algorithm when the model likelihood is a univariate integral. In that case, the Laplace approximation to the likelihood is efficient and therefore the resulting LA estimators of model parameters are close enough to the ML estimators maximizing the exact log-likelihood. This is the case of the Poisson mixed model of this chapter. However, the likelihoods of the spatial and/or temporal Poisson mixed models are multiple (high dimensional) integrals. In those cases, the Laplace integral approximation does not work well and the LA estimates of model parameters might be far from the exact ML estimates. This is the main reasons why the LA algorithm is not employed in Chapters 3-5.

The ML-PQL algorithm works in two steps iteratively. Step 1 maximizes in $\boldsymbol{\beta}$ and $\boldsymbol{v}$ the

joint likelihood of the $(\boldsymbol{y}, \boldsymbol{v})$ and Step 2 calculates the ML estimators of the variance component parameters in the LMM defined on the estimated natural parameters. If $\boldsymbol{y}$ follows a LMM then the ML-PQL algorithm gives ML estimators, but this is not true if $\boldsymbol{y}$ follows a non normal GLMM. This method avoids the need of approximating integrals and it is a good alternative to the LA algorithm when the likelihood is a multiple integral. Unfortunately, the ML-PQL estimators have unsatisfactory properties when the variance parameters are high. The corresponding estimators are inconsistent under standard asymptotic assumptions (Jiang, 1998).

The MM estimators, suggested by Jiang (1998), are obtained by solving a nonlinear system of equations. Under the Poisson mixed models considered in this manuscript, the non linear equations are explicitly calculated and there is no need of numerical approximations. Therefore, the MM algorithm for Poisson mixed models only requires running a standard Newton-Raphson algorithm. This fact makes the MM computationally attractive. Further, Section 2.4 gives regularity conditions for the consistency of MM estimator.

The MM algorithm is used in the application to poverty data of Section 2.9.1. The LA algorithm is the default option of the lme4 or nlme R packages, and it is used in the application to fire data in Section 2.9.2. Section 2.8 presents simulation experiments for investigating the behaviour of the MM, PQL and LA algorithms and for testing the programmed R codes.

2.3.1 The MM algorithm

This section derives the MM algorithm for fitting Model 1. A natural set of equations for applying this method is

$$0 = f_k(\boldsymbol{\theta}) = M_k(\boldsymbol{\theta}) - \hat{M}_k = \sum_{d=1}^{D} \mathbb{E}_{\boldsymbol{\theta}}[y_d] x_{dk} - \sum_{d=1}^{D} y_d x_{dk}, \quad k = 1, \ldots, p, \tag{2.3.1}$$

$$0 = f_{p+1}(\boldsymbol{\theta}) = M_{p+1}(\boldsymbol{\theta}) - \hat{M}_{p+1} = \sum_{d=1}^{D} \mathbb{E}_{\boldsymbol{\theta}}[y_d^2] - \sum_{d=1}^{D} y_d^2, \tag{2.3.2}$$

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 (2.3.1)-(2.3.2). The updating formula of the Newton-Raphson algorithm for solving this system is

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - \boldsymbol{H}^{-1}(\boldsymbol{\theta}^{(m)}) \boldsymbol{f}(\boldsymbol{\theta}^{(m)}), \tag{2.3.3}$$

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

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

The MM Newton-Raphson algorithm is specified if we calculate the expectations appearing in $\boldsymbol{f}(\boldsymbol{\theta})$ and its partial derivatives. The expectation of y_d is

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_d] &= \mathbb{E}_{\boldsymbol{v}}\big[\mathbb{E}_{\boldsymbol{\theta}}[y_d|\boldsymbol{v}]\big] = \mathbb{E}_{\boldsymbol{v}}[\nu_d p_d] = \mathbb{E}_{\boldsymbol{v}}\left[\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\right] \\
&= \int_{-\infty}^{\infty} \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\ (2\pi)^{-1/2} \exp\big\{-\frac{1}{2}v_d^2\big\}\, dv_d \\
&= \nu_d \exp\big\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\big\} \int_{-\infty}^{\infty} (2\pi)^{-1/2} \exp\big\{-\frac{1}{2}(v_d - \phi)^2\big\}\, dv_d \\
&= \nu_d \exp\big\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\big\}.
\end{aligned}$$

Therefore, the first p MM equations are

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

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

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \beta_k} = \nu_d \exp\big\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\big\} x_{dk}, \quad \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_d]}{\partial \phi} = \nu_d \exp\big\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\big\}\phi.$$

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

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

and therefore

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

Since

$$-\frac{1}{2}(v_d - 2\phi)^2 = -\frac{1}{2}(v_d^2 - 4\phi v_d + 4\phi^2) = -\frac{1}{2}v_d^2 + 2\phi v_d - 2\phi^2,$$

we have

$$\begin{aligned}\int_{-\infty}^{\infty} p_d^2 f_{\boldsymbol{v}}(v_d)\, dv_d &= \int_{-\infty}^{\infty} \exp\left\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi v_d\right\}(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}v_d^2\right\} dv_d \\ &= \exp\left\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\right\}\int_{-\infty}^{\infty}(2\pi)^{-1/2}\exp\left\{-\frac{1}{2}(v_d - 2\phi)^2\right\} dv_d \\ &= \exp\left\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\right\}\end{aligned}$$

and consequently

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

Then, the last MM equation is

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

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

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

The elements of the Jacobian matrix in the Newton-Raphson (2.3.3) algorithm are

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

The steps of the MM algorithm are given in Algorithm 1.

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 (Model 0). 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 $\hat{p}_d^{(0)} = \frac{y_d+1}{\nu_d+1}$.

Algorithm 1 MM algorithm

1: **Input**: Set the initial values $m = 0$ and $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\beta}^{(0)}, \phi^{(0)})$
2: **Repeat**

i) Update $\boldsymbol{\theta}^{(m)}$ by using the equation

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

ii) Update the iteration index $m \leftarrow m + 1$.

3: **Until** convergence.
4: **Output**: $\boldsymbol{\theta}^{(m+1)}$.

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

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

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 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$.
2. Generate $v_d^{*(b)} \sim N(0,1)$, $d = 1, \ldots, D$. Calculate $p_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}v_d^{*(b)}\}$ and generate $y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)})$, $d = 1, \ldots, D$, $b = 1, \ldots, B$.
3. From the bootstrap resamples, calculate $\hat{\boldsymbol{M}}^{*(b)}$, $b = 1, \ldots, B$, 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 resamples and calculate $\hat{\boldsymbol{\theta}}^{*(b)}$, $b = 1, \ldots, B$, $\overline{\boldsymbol{\theta}} = \frac{1}{B}\sum_{b=1}^{B} \hat{\boldsymbol{\theta}}^{*(b)}$.

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

For the implementation in R, we use the second option, i.e. we consider Steps 3′ and 4′.

2.3.2 The PQL algorithm

The ML-PQL estimator of $\boldsymbol{\beta}$ and predictor of $\boldsymbol{v}$ (see Breslow and Clayton (1993)) maximizes the joint log-likelihood

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

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

$$U_r = \frac{\partial \ell}{\partial \beta_r} = -\sum_{d=1}^{D}\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}x_{dr} + \sum_{d=1}^{D} y_d x_{dr}, \quad r = 1, \ldots, p,$$
$$U_{p+d} = \frac{\partial \ell}{\partial v_d} = -v_d - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\phi + y_d\phi, \quad d = 1, \ldots, D.$$

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

$$H_{r_1 r_2} = \frac{\partial^2 \ell}{\partial \beta_{r_1}\partial \beta_{r_2}} = -\sum_{d=1}^{D}\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}x_{dr_1}x_{dr_2}, \quad r_1, r_2 = 1, \ldots, p,$$
$$H_{rp+d} = \frac{\partial^2 \ell}{\partial \beta_r \partial v_d} = -\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}x_{dr}\phi, \quad r = 1, \ldots, p,\ d = 1, \ldots, D,$$
$$H_{p+dp+d} = \frac{\partial^2 \ell}{\partial v_d^2} = -1 - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\phi^2, \quad d = 1, \ldots, D,$$
$$H_{p+d_1 p+d_2} = \frac{\partial^2 \ell}{\partial v_{d_1}\partial v_{d_2}} = 0, \quad d_1, d_2 = 1, \ldots, D,\ d_1 \neq d_2.$$

In matrix form, we have $\boldsymbol{\xi} = (\boldsymbol{\beta}', \boldsymbol{v}')'$, $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{\xi}) = \underset{1\le r\le p+D}{\text{col}}(U_r)$ and $\boldsymbol{H} = \boldsymbol{H}(\boldsymbol{\xi}) = (H_{rs})_{r,s=1,\ldots,p+D}$. The Newton-Raphson algorithm maximizes $\ell(\boldsymbol{\beta}, \boldsymbol{v})$, with fixed ϕ. The updating equation is

$$\boldsymbol{\xi}^{(k+1)} = \boldsymbol{\xi}^{(k)} - \boldsymbol{H}^{-1}(\boldsymbol{\xi}^{(k)})\boldsymbol{U}(\boldsymbol{\xi}^{(k)}). \tag{2.3.4}$$

At the kth iteration of the algorithm, the penalized maximum likelihood estimation of ϕ maximizes the joint likelihood of $\eta_1^{(k)}, \ldots, \eta_D^{(k)}$, where

$$\eta_d^{(k)} = \log \nu_d + \boldsymbol{x}_d \boldsymbol{\beta}^{(k)} + \phi^{(k)} v_d^{(k)}$$

and

$$\eta_d^{(k)} \sim N(\log \nu_d + \boldsymbol{x}_d \boldsymbol{\beta}^{(k)},\ \text{var} = \phi^2), \quad d = 1, \ldots, D.$$

The joint log-likelihood of $\eta_1^{(k)}, \ldots, \eta_D^{(k)}$ is

$$\ell^{(k)} = -\frac{D}{2} \log 2\pi - D \log \phi - \frac{1}{2} \frac{1}{\phi^2} \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2.$$

By taking the first derivative of $\ell^{(k)}$ with respect to ϕ and equating to zero, we get

$$0 = U^{(k)} = \frac{\partial \ell^{(k)}}{\partial \phi} = -\frac{D}{\phi} + \frac{1}{\phi^3} \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2,$$

$$\phi^2 = \frac{1}{D} \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2 = \phi^{(k)2} \frac{1}{D} \sum_{d=1}^{D} v_d^{(k)2}.$$

The ML-PQL updating equation for ϕ is

$$\phi^{(k+1)2} = \phi^{(k)2} \frac{1}{D} \sum_{d=1}^{D} v_d^{(k)2}. \tag{2.3.5}$$

The PQL algorithm calculates the predictors of $\boldsymbol{v}$ and the estimators of $\boldsymbol{\beta}$ and ϕ. The steps are:

Algorithm 2 PQL algorithm

1: Set the values $\boldsymbol{\beta}^{(0)}$, $\boldsymbol{v}^{(0)}$, $\phi^{(0)}$ and $m = 1$.
2: Run the algorithm given in equation (2.3.4). Use $\phi^{(m-1)}$ as known value and $\boldsymbol{\beta}^{(m-1)}$, $\boldsymbol{v}^{(m-1)}$ as algorithm seeds. Let $\boldsymbol{\beta}^{(m)}$ and $\boldsymbol{v}^{(m)}$ be the output of the algorithm given in equation (2.3.4).
3: Update ϕ by using the updating equation (2.3.5), i.e.

$$\phi^{(m)2} = \phi^{(m-1)2} \frac{1}{D} \sum_{d=1}^{D} v_d^{(m)2}.$$

4: Update the iteration index $m \leftarrow m + 1$.
5: Repeat the Steps 2-4 until the convergence of $\boldsymbol{\beta}^{(m)}$, $v_d^{(m)}$ and $\phi^{(m)}$.

By taking the second derivative of $\ell^{(k)}$ with respect to ϕ,

$$H^{(k)} = \frac{\partial^2 \ell^{(k)}}{\partial \phi^2} = D\phi^{-2} - 3\phi^{-4} \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2$$
$$= \phi^{-4} \Big(D\phi^2 - 3 \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2 \Big),$$

we get an alternative updating equation for the variance parameter ϕ based on the Newton-Raphson algorithm. We have that

$$\big(H^{(k)}\big)^{-1} U^{(k)} = \frac{\phi^{-3} \Big(- D\phi^2 + \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2 \Big)}{\phi^{-4} \Big(D\phi^2 - 3 \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2 \Big)}$$
$$= \phi \frac{-D\phi^2 + \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2}{D\phi^2 - 3 \sum_{d=1}^{D} (\eta_d^{(k)} - \log \nu_d - \boldsymbol{x}_d \boldsymbol{\beta}^{(k)})^2},$$

and therefore, the Newton-Raphson updating equation is

$$\phi^{(k+1)} = \phi^{(k)} - \big(H^{(k)}\big)^{-1} U^{(k)} = \phi^{(k)} - \phi^{(k)} \frac{-D\phi^{(k)2} + \sum_{d=1}^{D} \phi^{(k)2} v_d^{(k)2}}{D\phi^{(k)2} - 3 \sum_{d=1}^{D} \phi^{(k)2} v_d^{(k)2}}$$
$$= \phi^{(k)} - \phi^{(k)} \frac{-D + \sum_{d=1}^{D} v_d^{(k)2}}{D - 3 \sum_{d=1}^{D} v_d^{(k)2}} = \phi^{(k)} \Big(1 - \frac{-D + \sum_{d=1}^{D} v_d^{(k)2}}{D - 3 \sum_{d=1}^{D} v_d^{(k)2}} \Big)$$
$$= \phi^{(k)} \frac{2D - 4 \sum_{d=1}^{D} v_d^{(k)2}}{D - 3 \sum_{d=1}^{D} v_d^{(k)2}} = 2\phi^{(k)} \frac{D - 2 \sum_{d=1}^{D} v_d^{(k)2}}{D - 3 \sum_{d=1}^{D} v_d^{(k)2}}.$$

Finally, the PQL updating equation (2.3.5) can be substituted by

$$\phi^{(m)} = 2\phi^{(m-1)} \frac{D - 2 \sum_{d=1}^{D} v_d^{(m)2}}{D - 3 \sum_{d=1}^{D} v_d^{(m)2}}.$$

For the implementation of the PQL algorithm in R, we opted for the updating equation (2.3.5) since it offered a more robust behaviour.

2.3.3 The ML-Laplace algorithm

The ML-Laplace algorithm maximizes the Laplace approximation to the joint marginal log-likelihood of the target vector $\boldsymbol{y} = (y_1, \ldots, y_D)$. For this sake, let $h : \mathbb{R} \mapsto \mathbb{R}$ be a continuously twice differentiable function with a global maximum at x_0. This is to say,

let us assume that $\dot{h}(x_0) = 0$ and $\ddot{h}(x_0) < 0$, where $\dot{h}$ and $\ddot{h}$ denote the first and second derivatives of h respectively. A Taylor series expansion of $h(x)$ around x_0 yields to

$$\begin{aligned} h(x) &= h(x_0) + \dot{h}(x_0)(x - x_0) + \frac{1}{2}\ddot{h}(x_0)(x - x_0)^2 + o\big(|x - x_0|^2\big) \\ &\approx h(x_0) + \frac{1}{2}\ddot{h}(x_0)(x - x_0)^2. \end{aligned}$$

The univariate Laplace approximation is

$$\begin{aligned} \int_{-\infty}^{\infty} e^{h(x)}\,dx &\approx \int_{-\infty}^{\infty} e^{h(x_0)} \exp\Big\{ -\frac{1}{2}\big(-\ddot{h}(x_0)\big)(x - x_0)^2 \Big\}\,dx \\ &= (2\pi)^{1/2}\big(-\ddot{h}(x_0)\big)^{-1/2} e^{h(x_0)} \int_{-\infty}^{\infty} \frac{\exp\Big\{ -\frac{1}{2}\Big(\frac{x - x_0}{(-\ddot{h}(x_0))^{-1/2}}\Big)^2 \Big\}}{(2\pi)^{1/2}\big(-\ddot{h}(x_0)\big)^{-1/2}}\,dx \\ &= (2\pi)^{1/2}\big(-\ddot{h}(x_0)\big)^{-1/2} e^{h(x_0)}. \end{aligned}$$

Let us now approximate the loglikelihood of the considered Poisson mixed model. We recall that $v_1, \ldots, v_D$ are i.i.d $N(0,1)$ and that

$$y_d | v_d \underset{ind}{\sim} \text{Poiss}(\nu_d p_d), \quad p_d = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}, \quad d = 1, \ldots, D.$$

It holds that $y_1, \ldots, y_D$ are unconditionally independent with marginal probability distribution function

$$\begin{aligned} \mathbb{P}(y_d) &= \int_{-\infty}^{\infty} \mathbb{P}(y_d|v_d) f(v_d)\,dv_d = \int_{-\infty}^{\infty} \frac{1}{y_d!} \exp\{-\nu_d p_d\} \nu_d^{y_d} p_d^{y_d} (2\pi)^{-1/2} \exp\{-\frac{1}{2}v_d^2\}\,dv_d \\ &= \frac{\nu_d^{y_d}}{(2\pi)^{1/2} y_d!} \int_{-\infty}^{\infty} \exp\Big\{ -\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} + y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \frac{1}{2}v_d^2 \Big\}\,dv_d \\ &= \frac{\nu_d^{y_d}}{(2\pi)^{1/2} y_d!} \exp\{y_d \boldsymbol{x}_d\boldsymbol{\beta}\} \int_{-\infty}^{\infty} \exp\big\{h(v_d)\big\}\,dv_d, \end{aligned}$$

where

$$\begin{aligned} h(v_d) &= -\nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} + \phi y_d v_d - \frac{1}{2}v_d^2, \\ \dot{h}(v_d) &= -\nu_d \phi \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} + \phi y_d - v_d, \\ \ddot{h}(v_d) &= -(1 + \nu_d \phi^2 \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}). \end{aligned} \tag{2.3.6}$$

Let v_{0d} be the value of v_d such that $\dot{h}(v_{0d}) = 0$ and $\ddot{h}(v_{0d}) < 0$. By applying (2.3.6) in $v_d = v_{0d}$, we get

$$\mathbb{P}(y_d) \approx \frac{\nu_d^{y_d}}{y_d!}\exp\{y_d \boldsymbol{x}_d\boldsymbol{\beta}\}(1+\nu_d\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\})^{-1/2}$$
$$\cdot \exp\Big\{-\nu_d\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}+\phi y_d v_{0d}-\frac{1}{2}v_{0d}^2\}.$$

The loglikelihood can be approximated by

$$\ell = \sum_{d=1}^{D}\log\mathbb{P}(y_d) \approx \sum_{d=1}^{D}\Big\{y_d\log\nu_d - \log y_d! + y_d\boldsymbol{x}_d\boldsymbol{\beta} - \frac{1}{2}\log(1+\nu_d\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\})$$
$$-\nu_d\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}+\phi y_d v_{0d}-\frac{1}{2}v_{0d}^2\Big\} := \ell_L(\boldsymbol{\beta},\phi,v_{01},\ldots,v_{0D}) = \ell_L,$$

where $:=$ denotes equality by definition. For ease of presentation, let us define $p_{0d} = \exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\phi v_{0d}\}$ and $\xi_{0d} = 1+\nu_d\phi^2 p_{0d}$. It holds that

$$\frac{\partial p_{0d}}{\partial\beta_r} = x_{dr}p_{0d},\ \frac{\partial p_{0d}}{\partial\phi} = v_{0d}p_{0d},\quad \frac{\partial\xi_{0d}}{\partial\beta_r} = \nu_d\phi^2 x_{dr}p_{0d},\ \frac{\partial\xi_{0d}}{\partial\phi} = (2+\phi v_{0d})\nu_d\phi p_{0d}.$$

The approximated loglikelihood is

$$\ell_L = \sum_{d=1}^{D}\Big\{y_d\log\nu_d - \log y_d! + y_d\boldsymbol{x}_d\boldsymbol{\beta} - \frac{1}{2}\log\xi_{0d} - \nu_d p_{0d} + \phi y_d v_{0d} - \frac{1}{2}v_{0d}^2\Big\}.$$

The first derivatives of ℓ_L with respect to β_r and ϕ are

$$\frac{\partial\ell_L}{\partial\beta_r} = \sum_{d=1}^{D}\Big\{y_d x_{dr} - \frac{\nu_d x_{dr}}{2}\frac{\phi^2 p_{0d}}{\xi_{0d}} - \nu_d x_{dr}p_{0d}\Big\},$$
$$\frac{\partial\ell_L}{\partial\phi} = \sum_{d=1}^{D}\Big\{y_d v_{0d} - \frac{\nu_d}{2}\frac{2\phi p_{0d}+\phi^2 v_{0d}p_{0d}}{\xi_{0d}} - \nu_d v_{0d}p_{0d}\Big\}.$$

The second partial derivatives of ℓ_L are

$$\frac{\partial^2\ell_L}{\partial\beta_s\partial\beta_r} = -\sum_{d=1}^{D}\Big\{\frac{\nu_d x_{dr}\phi^2}{2}\frac{p_{0d}x_{ds}(1+\nu_d\phi^2 p_{0d}) - p_{0d}\nu_d\phi^2 x_{ds}p_{0d}}{\xi_{0d}^2} + \nu_d x_{dr}x_{ds}p_{0d}\Big\}$$
$$= -\sum_{d=1}^{D}\Big\{\frac{\nu_d x_{dr}x_{ds}\phi^2}{2}\frac{p_{0d}}{\xi_{0d}^2} + \nu_d x_{dr}x_{ds}p_{0d}\Big\} = -\sum_{d=1}^{D}\nu_d x_{dr}x_{ds}p_{0d}\Big(\frac{\phi^2}{2\xi_{0d}^2}+1\Big),$$

$$\begin{aligned}
\frac{\partial^2 \ell_L}{\partial\phi\partial\beta_r} &= -\sum_{d=1}^{D}\left\{\frac{\nu_d x_{dr}\phi p_{0d}}{2}\frac{(2+\phi v_{0d})[1+\nu_d\phi^2 p_{0d}-\nu_d\phi^2 p_{0d}]}{\xi_{0d}^2}+\nu_d x_{dr}v_{0d}p_{0d}\right\}\\
&= -\sum_{d=1}^{D}\left\{\frac{\nu_d x_{dr}\phi p_{0d}}{2}\frac{(2+\phi v_{0d})}{\xi_{0d}^2}+\nu_d x_{dr}v_{0d}p_{0d}\right\}\\
&= -\sum_{d=1}^{D}\nu_d x_{dr}v_{0d}p_{0d}\Big(\frac{\phi(2+\phi v_{0d})}{2v_{0d}\xi_{0d}^2}+1\Big),\\
\frac{\partial^2 \ell_L}{\partial\phi^2} &= -\sum_{d=1}^{D}\left\{\frac{\nu_d p_{0d}}{2}\frac{[-2+(2+\phi v_{0d})^2](1+\nu_d\phi^2 p_{0d})-(2+\phi v_{0d})^2\nu_d\phi^2 p_{0d}}{\xi_{0d}^2}+\nu_d v_{0d}^2 p_{0d}\right\}\\
&= -\sum_{d=1}^{D}\left\{\frac{\nu_d p_{0d}}{2}\frac{-2(1+\nu_d\phi^2 p_{0d})+(2+\phi v_{0d})^2(1+\nu_d\phi^2 p_{0d}-\nu_d\phi^2 p_{0d})}{\xi_{0d}^2}+\nu_d v_{0d}^2 p_{0d}\right\}\\
&= -\sum_{d=1}^{D}\left\{\frac{\nu_d p_{0d}}{2}\frac{(2+\phi v_{0d})^2-2\xi_{0d}}{\xi_{0d}^2}+\nu_d v_{0d}^2 p_{0d}\right\}\\
&= -\sum_{d=1}^{D}\nu_d v_{0d}^2 p_{0d}\Big(\frac{(2+\phi v_{0d})^2-2\xi_{0d}}{2v_{0d}^2\xi_{0d}^2}+1\Big).
\end{aligned}$$

The components of the score vector and the Hessian matrix are

$$\begin{aligned}
&U_{0r}=\frac{\partial\ell_L}{\partial\beta_r},\ U_{0p+1}=\frac{\partial\ell_L}{\partial\phi},\\
&H_{0rs}=H_{0sr}\frac{\partial^2\ell_L}{\partial\beta_s\partial\beta_r},\ H_{rp+1}=H_{p+1r}=\frac{\partial^2\ell_L}{\partial\phi\partial\beta_r},\ H_{0p+1p+1}=\frac{\partial^2\ell_L}{\partial\phi^2}.
\end{aligned}$$

$$\boldsymbol{U}_0=\boldsymbol{U}_0(\boldsymbol{\theta})=\operatorname*{col}_{1\le r\le p+1}(U_{0rs}),\quad \boldsymbol{H}_0=\boldsymbol{H}_0(\boldsymbol{\theta})=(H_{0rs})_{r,s=1,\ldots,p+1}.$$

The Newton-Raphson algorithm maximizes $\ell_L(\boldsymbol{\theta})$, with $v_d=v_{0d}$ fixed, $d=1,\ldots,D$. The updating equation is

$$\boldsymbol{\theta}^{(k+1)}=\boldsymbol{\theta}^{(k)}-\boldsymbol{H}_0^{-1}(\boldsymbol{\theta}^{(k)})\boldsymbol{U}_0(\boldsymbol{\theta}^{(k)}). \tag{2.3.7}$$

For $d=1,\ldots,D$, the Newton-Raphson algorithm also maximizes $h(v_d)=h(v_d,\boldsymbol{\theta})$, defined in (2.3.6), with $\boldsymbol{\theta}=(\boldsymbol{\beta}',\phi)=\boldsymbol{\theta}_0$ fixed. The updating equation is

$$v_d^{(k+1)}=v_d^{(k)}-\frac{\dot h(v_d^{(k)},\boldsymbol{\theta}_0)}{\ddot h(v_d^{(k)},\boldsymbol{\theta}_0)}. \tag{2.3.8}$$

The ML-Laplace algorithm combines the updating equations (2.3.7) and (2.3.8).

Algorithm 3 ML-Laplace algorithm

1: Set the initial values $k = 0$, $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(-1)} = \boldsymbol{\theta}^{(0)} + \mathbf{1}$, $v_d^{(0)} = 0$, $v_d^{(-1)} = 1$, $d = 1, \ldots, D$.

2: **Repeat**

i) Apply the iterative algorithm with updating equation (2.3.8), seeds $v_d^{(k)}$, $d = 1, \ldots, D$, convergence tolerance ε_2 and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ fixed. Output: $v_d^{(k+1)}$, $d = 1, \ldots, D$.

ii) Apply the iterative algorithm with updating equation (2.3.7), seed $\boldsymbol{\theta}^{(k)}$, convergence tolerance ε_1 and $v_{0d} = v_d^{(k+1)}$ fixed, $d = 1, \ldots, D$. Output: $\boldsymbol{\theta}^{(k+1)}$.

iii) $k \leftarrow k + 1$.

3: **Until** $\|\boldsymbol{\theta}^{(k)} - \boldsymbol{\theta}^{(k-1)}\|_2 < \varepsilon_1$, $|v_d^{(k)} - v_d^{(k-1)}| < \varepsilon_2$, $d = 1, \ldots, D$.

4: **Output**: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(k)}$, $\hat{v}_d = v_d^{(k)}$, $d = 1, \ldots, D$.

The asymptotic variance of the ML-Laplace estimators can be obtained from the diagonal of the matrix $\boldsymbol{H}_0^{-1}(\hat{\boldsymbol{\theta}})$.

2.4 Asymptotic properties of the MM estimators

Section 2.3.1 defines the MM estimator $\boldsymbol{\theta} = (\beta_1, \ldots, \beta_p, \phi)$ as the solution to the equation

$$\boldsymbol{M}(\boldsymbol{\theta}) = \hat{\boldsymbol{M}}, \tag{2.4.1}$$

where $\boldsymbol{M}(\boldsymbol{\theta}) = (M_1(\boldsymbol{\theta}), \ldots, M_{p+1}(\boldsymbol{\theta}))$, $\hat{\boldsymbol{M}} = (\hat{M}_1, \ldots, \hat{M}_{p+1})$ and

$$\begin{aligned}
M_k(\theta) &= \sum_{d=1}^{D} \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\right\} x_{dk}, \quad k = 1, \ldots, p, \\
M_{p+1}(\theta) &= \sum_{d=1}^{D} \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\right\} + \sum_{d=1}^{D} \nu_d^2 \exp\left\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\right\}, \\
\hat{M}_k &= \sum_{d=1}^{D} y_d x_{dk}, \quad k = 1, \ldots, p, \quad \hat{M}_{p+1} = \sum_{d=1}^{D} y_d^2.
\end{aligned}$$

More generally, we may define the MM estimator $\hat{\boldsymbol{\theta}}$ as the vector $\boldsymbol{\theta}$ that minimize the Euclidean distance of the two sides of (2.4.1). This section particularizes to the area-level Poisson mixed model the asymptotic properties of the MM estimators given by Jiang (2003).

Suppose that

$$\boldsymbol{x}_d \in \mathcal{X} \subset \mathbb{R}^p, \;\; d = 1, \ldots, D, \quad \text{where } \mathcal{X} \text{ is compact.} \tag{2.4.2}$$

Furthermore, supposes that there exist $B > 0$ and $\varepsilon > 0$ such that, for large D,

$$\min \left\{ \inf_{\tilde{\boldsymbol{\theta}} \notin \Theta_B} \|\boldsymbol{M}(\tilde{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\|, \;\; \inf_{\tilde{\boldsymbol{\theta}} \in \Theta_B, \tilde{\boldsymbol{\theta}} \neq \boldsymbol{\theta}} \left\{ \frac{\|\boldsymbol{M}(\tilde{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\|}{\|\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}\|} \right\} \right\} \geq \varepsilon, \tag{2.4.3}$$

where $\boldsymbol{\theta}$ is the true parameter vector and $\Theta_B = \left\{ \tilde{\boldsymbol{\theta}} \in \mathbb{R}^p \times \mathbb{R}_+ : |\tilde{\theta}_k| \leq B, \, k = 1, \ldots, p+1 \right\}$. Under the assumptions (2.4.2)-(2.4.3), Jiang (2003) established that

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\| = O_p(D^{-1/2}). \tag{2.4.4}$$

Let $\hat{\boldsymbol{\theta}}_{d-}$ be the MM estimator based on $\boldsymbol{y}_{d-} = \left(y_{d'}\right)_{d' \neq d}$, which is the solution of the equation $\boldsymbol{M}_{d-}(\boldsymbol{\theta}) = \hat{\boldsymbol{M}}_{d-}$, where $\boldsymbol{M}_{d-}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{M}}_{d-}$ are defined similarly, i.e. with $\sum_{d'=1, d' \neq d}^{D}$ instead of $\sum_{d'=1}^{D}$. Let $\boldsymbol{M}(\Theta)$ $\left(\boldsymbol{M}_{d-}(\Theta)\right)$ be the image of the parameter space Θ under $\boldsymbol{M}(.)$ $\left(\boldsymbol{M}_{d-}(.)\right)$. For $\boldsymbol{u} \in \mathbb{R}^p \times \mathbb{R}_+$, $\mathcal{Z} \subset \mathbb{R}^p \times \mathbb{R}_+$, let us define

$$\text{dist}(\boldsymbol{u}, \mathcal{Z}) = \inf_{\boldsymbol{z} \in \mathcal{Z}} \|\boldsymbol{z} - \boldsymbol{u}\|.$$

Suppose that

$$\liminf \lambda_{min}\left(\boldsymbol{H}'(\boldsymbol{\theta})\boldsymbol{H}(\boldsymbol{\theta})\right) > 0, \tag{2.4.5}$$

where $\boldsymbol{H} = \dot{\boldsymbol{M}}$ is the matrix of first derivatives and $\lambda_{min}(.)$ is the smallest eigenvalue function. For large D, suppose that

$$\min \left\{ \text{dist}\left(\boldsymbol{M}(\boldsymbol{\theta}), \boldsymbol{M}(\Theta)^c\right), \;\; \text{dist}\left(\boldsymbol{M}_{d-}(\boldsymbol{\theta}), \boldsymbol{M}_{d-}(\Theta)^c\right) \right\} \geq \varepsilon, \quad d = 1, \ldots, D. \tag{2.4.6}$$

Under (2.4.2)-(2.4.6), Jiang (2003) proved that

$$\|\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{d-}\| = o_p(D^{-1/2}), \quad d = 1, \ldots, D. \tag{2.4.7}$$

Also, under (2.4.2)-(2.4.6), Jiang (2003) established that

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right] = O(D^{-1/2}). \tag{2.4.8}$$

2.5 The predictors

This section gives the EBP and a plug-in predictor of p_d under Model 1. The EBP is obtained from its best predictor (BP). The BP of p_d minimizes the MSE in the set of unbiased predictors. Despite its good properties, in practice this estimator is not useful since is given in terms of the theoretical parameters (unknown in practice). Therefore, it is considered its empirical version, the EBP, which is obtained from the BP replacing the unknown theoretical parameters by their estimates. Under regularity conditions, the EBPs have asymptotically the properties of the BPs. Both, the EBP and the plug-in predictor, are compared in the simulation study.

2.5.1 The empirical best predictor

This section derives the BP and the EBP of p_d under the area-level Poisson mixed model. Under Model 1, the conditional distribution of $\boldsymbol{y} = (y_1, \ldots, y_D)'$, given $\boldsymbol{v}$, is

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

where

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

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

$$\mathbb{E}_{\boldsymbol{\theta}}[p_d|y_d] = \frac{\int_{\mathbb{R}} \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\mathbb{P}(y_d|v_d) f(v_d)\, dv_d}{\int_{\mathbb{R}} \mathbb{P}(y_d|v_d) f(v_d)\, dv_d} = \frac{N_d(y_d, \boldsymbol{\theta})}{D_d(y_d, \boldsymbol{\theta})} := \psi_d(y_d, \boldsymbol{\theta}), \tag{2.5.1}$$

where

$$N_d(y_d, \boldsymbol{\theta}) = \int_{\mathbb{R}} \exp\left\{(y_d + 1)(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\right\}\right\} f(v_d)\, dv_d,$$
$$D_d(y_d, \boldsymbol{\theta}) = \int_{\mathbb{R}} \exp\left\{y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\right\} f(v_d)\, dv_d.$$

Remark 2.1. The numerator $N_d(y_d, \boldsymbol{\theta})$ can be expressed in terms of $D_d(y_d, \boldsymbol{\theta})$ as $N_d(y_d, \boldsymbol{\theta}) = D_d(y_d + 1, \boldsymbol{\theta})$.

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 approximate it by estimating the integrals with an accelerated Monte Carlo method based on the properties of the antithetic variables for reducing the variability. This algorithm is described below.

Algorithm 4 Accelerated Monte Carlo algorithm

1: Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$ as in Section 2.3.

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.

$$\hat{N}_d = \frac{1}{2L} \sum_{\ell=1}^{2L} \exp\left\{(y_d + 1)(\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}) - \nu_d \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}\}\right\},$$
$$\hat{D}_d = \frac{1}{2L} \sum_{\ell=1}^{2L} \exp\left\{y_d(\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}) - \nu_d \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{(\ell)}\}\right\}. \tag{2.5.2}$$

Since the size variable ν_d is known in practice, then the EBP of $\mu_d = \nu_d p_d$ is $\hat{\mu}_d(\hat{\boldsymbol{\theta}}) = \nu_d \hat{p}_d(\hat{\boldsymbol{\theta}})$.

2.5.2 Plug-in predictors

The plug-in predictor of p_d is obtained replacing the unknown parameters by their estimates, i.e. $\hat{p}_d^P(\hat{\boldsymbol{\theta}}) = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}\hat{v}_d\}$. As the MM Newton-Raphson algorithm does not give a prediction of v_d, we use its EBP. The BP of v_d is

$$\hat{v}_d(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_d|y_d] = \frac{\int_{\mathbb{R}} v_d \mathbb{P}(y_d|v_d) f(v_d)\, dv_d}{\int_{\mathbb{R}} \mathbb{P}(y_d|v_d) f(v_d)\, dv_d} = \frac{N_{v,d}(y_d, \boldsymbol{\theta})}{D_d(y_d, \boldsymbol{\theta})},$$

where

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

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 similar to Algorithm 4. The steps are the same, replacing Step 3 by

3. Calculate $\hat{v}_d(\hat{\boldsymbol{\theta}}) = \hat{N}_{v,d}/\hat{D}_d$, where $\hat{D}_d$ is defined in (2.5.2) and

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

Section 2.8 studies the behaviour of $\hat{p}_d(\hat{\boldsymbol{\theta}})$ and $\hat{p}_d^P(\hat{\boldsymbol{\theta}})$ by calculating empirical biases and MSEs. On the other hand, we can consider the synthetic estimator. It is similar to $\hat{p}_d^P(\hat{\boldsymbol{\theta}})$ but only uses fixed effects, i.e. $\hat{p}_d^{syn}(\hat{\boldsymbol{\theta}}) = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}}\}$. The synthetic estimator can be seen as a plug-in estimator under Model 0.

2.6 The MSE of the empirical best predictor

Theorem 2.1 gives an approximation to the MSE of the EBP $\hat{p}_d$, $d = 1, \ldots, D$.

Theorem 2.1 Assume that the condition (2.4.2) of being uniformly bounded holds for the auxiliary variables. Let $\hat{\boldsymbol{\theta}}$ be an estimator of $\boldsymbol{\theta}$ fulfilling the hypotheses (2.4.4), (2.4.7) and (2.4.8). Then the MSE of $\hat{p}_d$ can be approximated by

$$MSE(\hat{p}_d) = g_d(\boldsymbol{\theta}) + \frac{1}{D}c_d(\boldsymbol{\theta}) + o(1/D), \tag{2.6.1}$$

where

$$c_d(\boldsymbol{\theta}) = \sum_{j=0}^{\infty}\left(\frac{\partial}{\partial\boldsymbol{\theta}}\psi_d(j,\boldsymbol{\theta})\right)' \boldsymbol{V}(\boldsymbol{\theta})\left(\frac{\partial}{\partial\boldsymbol{\theta}}\psi_d(j,\boldsymbol{\theta})\right) P_d(j,\boldsymbol{\theta}), \quad \boldsymbol{V}(\boldsymbol{\theta}) = D\mathbb{E}\left[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'\right],$$

and $\psi_d(j,\boldsymbol{\theta})$ is defined from (2.5.1) by substituting y_d by j.

Proof. The MSE of the EBP of $p_d = p_d(\boldsymbol{\theta}, v_d) = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}$ can be decomposed in the following form.

$$\begin{aligned} MSE(\hat{p}_d) &= \mathbb{E}\big[(\hat{p}_d(\hat{\boldsymbol{\theta}}) - p_d(\boldsymbol{\theta}, v_d))^2\big] = \mathbb{E}\big[\big(\{\hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta})\} + \{\hat{p}_d(\boldsymbol{\theta}) - p_d(\boldsymbol{\theta}, v_d)\}\big)^2\big] \\ &= \mathbb{E}\big[\big(\hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta})\big)^2\big] + \mathbb{E}\big[\big(\hat{p}_d(\boldsymbol{\theta}) - p_d(\boldsymbol{\theta}, v_d)\big)^2\big], \end{aligned}$$

because

$$\mathbb{E}\big[\big(\hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta})\big)\big(\hat{p}_d(\boldsymbol{\theta}) - p_d(\boldsymbol{\theta}, v_d)\big)\big] = \mathbb{E}\big[\big(\hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta})\big)\mathbb{E}\big[\hat{p}_d(\boldsymbol{\theta}) - p_d(\boldsymbol{\theta}, v_d)|y_d\big]\big] = 0.$$

The second term of $MSE(\hat{p}_d)$ is the MSE of the BP, namely

$$\begin{aligned} g_d(\boldsymbol{\theta}) &= \mathbb{E}\big[\big(\hat{p}_d(\boldsymbol{\theta}) - p_d(\boldsymbol{\theta}, v_d)\big)^2\big] = \mathbb{E}\big[\hat{p}_d^2(\boldsymbol{\theta})\big] + \mathbb{E}\big[p_d^2(\boldsymbol{\theta}, v_d)\big] - 2\mathbb{E}\big[\hat{p}_d(\boldsymbol{\theta})\mathbb{E}\big[p_d(\boldsymbol{\theta}, v_d)|y_d\big]\big] \\ &= \mathbb{E}\big[p_d^2(\boldsymbol{\theta}, v_d)\big] - \mathbb{E}\big[\hat{p}_d^2(\boldsymbol{\theta})\big]. \end{aligned}$$

The first term of $g_d(\boldsymbol{\theta})$ is

$$\mathbb{E}\big[p_d^2(\boldsymbol{\theta}, v_d)\big] = \int_{\mathbb{R}} \exp\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi v_d\}\, f(v_d)\, dv_d = \exp\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\}.$$

The second term of $g_d(\boldsymbol{\theta})$ is

$$\mathbb{E}\big[\hat{p}_d^2(\boldsymbol{\theta})\big] = \mathbb{E}[\psi_d^2(y_d, \boldsymbol{\theta})] = \sum_{j=0}^{\infty} \psi_d^2(j, \boldsymbol{\theta}) P_d(j, \boldsymbol{\theta}),$$

where

$$\begin{aligned} P_d(j, \boldsymbol{\theta}) &= \mathbb{P}(y_d = j) = \int_{\mathbb{R}} \mathbb{P}(y_d = j|v_d) f(v_d)\, dv_d \\ &= \frac{\nu_d^j}{j!} \int_{\mathbb{R}} \exp\{j(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\}\, f(v_d) dv_d = \frac{\nu_d^j}{j!} D_d(j, \boldsymbol{\theta}). \end{aligned}$$

Consequently,

$$g_d(\boldsymbol{\theta}) = \exp\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\} - \sum_{j=0}^{\infty} \psi_d^2(j, \boldsymbol{\theta}) \frac{\nu_d^j}{j!} D_d(j, \boldsymbol{\theta}).$$

Concerning the first term of $MSE(\hat{p}_d)$, we expand $\hat{p}_d(\hat{\boldsymbol{\theta}})$ in Taylor series around $\boldsymbol{\theta}$ and we have

$$\begin{aligned} \hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta}) = \psi_d(y_d, \hat{\boldsymbol{\theta}}) - \psi_d(y_d, \boldsymbol{\theta}) &= \left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi_d(y_d, \boldsymbol{\theta})\right)' (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &+ \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \psi_d(y_d, \boldsymbol{\theta})\right) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2). \end{aligned}$$

As the $\boldsymbol{x}_d$'s are bounded fulfilling (2.4.2) and the regularity conditions (2.4.4) and (2.4.8) on $\hat{\boldsymbol{\theta}}$ holds, we have that

$$\mathbb{E}\left[\big(\hat{p}_d(\hat{\boldsymbol{\theta}}) - \hat{p}_d(\boldsymbol{\theta})\big)^2\right] = \frac{1}{D}\mathbb{E}\left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \psi_d(y_d, \boldsymbol{\theta})\right)' \sqrt{D}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right)^2\right] + o(1/D).$$

Now we consider $\hat{\boldsymbol{\theta}}_{d-}$, an estimator based on $\boldsymbol{y}_{d-} = (\boldsymbol{y}_{d'})_{d' \neq d}$, and write $\hat{p}_{d-} = \psi_d(y_d, \hat{\boldsymbol{\theta}}_{d-})$. Then, by the independence of y_d and $\boldsymbol{y}_{d-}$, we have

$$
\begin{aligned}
a_d(\boldsymbol{\theta}) &= \mathbb{E}\left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}}\,\psi_d(y_d,\boldsymbol{\theta})\right)'\sqrt{D}(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})\right)^2\right] \\
&= \sum_{j=0}^{\infty}\mathbb{E}\left[\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}}\,\psi_d(y_d,\boldsymbol{\theta})\right)'\sqrt{D}(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})\right)^2\middle| y_d=j\right]P_d(j,\boldsymbol{\theta}) \\
&= \sum_{j=0}^{\infty}\left(\frac{\partial}{\partial \boldsymbol{\theta}}\,\psi_d(j,\boldsymbol{\theta})\right)'\boldsymbol{V}_{d-}(\boldsymbol{\theta})\left(\frac{\partial}{\partial \boldsymbol{\theta}}\,\psi_d(j,\boldsymbol{\theta})\right)P_d(j,\boldsymbol{\theta}),
\end{aligned}
$$

where

$$
\boldsymbol{V}_{d-}(\boldsymbol{\theta}) = D\mathbb{E}\left[(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})'\middle| y_d=j\right] = D\mathbb{E}\left[(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{d-}-\boldsymbol{\theta})'\right].
$$

Therefore,

$$
MSE(\hat{p}_{d-}) = g_d(\boldsymbol{\theta}) + \frac{1}{D}\,a_d(\boldsymbol{\theta}) + o(1/D).
$$

Under the assumptions (2.4.2), (2.4.4), (2.4.8) and (2.4.7), we may replace $\hat{\boldsymbol{\theta}}_{d-}$ by $\hat{\boldsymbol{\theta}}$, an estimator of $\boldsymbol{\theta}$ based on all data, and we obtain the stated result. □

The following section derives plug-in and bias-corrected plug-in estimators of the approximation (2.6.1) for the MSE of the EBP. A bootstrap estimator is also given.

2.7 Estimation of the MSE of the empirical best predictor

Two alternatives for estimating the MSE of the EBP of p_d are presented. Section 2.7.1 provides two analytical estimators (without and with bias-correction term), assuming that the model parameters are estimated by using the MM algorithm. As the analytical estimation of the MSE is computationally expensive (specially the estimator with bias correction), Section 2.7.3 gives a bootstrap procedure.

2.7.1 Analytic estimation of the MSE

A plug-in estimator of (2.6.1) is obtained replacing $\boldsymbol{\theta}$ by a consistent estimator $\hat{\boldsymbol{\theta}}$, namely

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

By a Taylor expansion of $c_d(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ and the consistency of $\hat{\boldsymbol{\theta}}$, we have that $\mathbb{E}[c_d(\hat{\boldsymbol{\theta}}) - c_d(\boldsymbol{\theta})] = o(1)$. However $\mathbb{E}[g_d(\hat{\boldsymbol{\theta}}) - g_d(\boldsymbol{\theta})]$ is not of order $o(D^{-1})$ and therefore

$$\mathbb{E}\big[MSE(\hat{p}_d) - \big\{g_d(\hat{\boldsymbol{\theta}}) + D^{-1}c_d(\hat{\boldsymbol{\theta}})\big\}\big] \neq o(D^{-1}).$$

Let $\hat{\boldsymbol{\theta}}$ be a truncated MM estimator. This means

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

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

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

where

$$b_d(\boldsymbol{\theta}) = \Big(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\Big)' D\mathbb{E}[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] + \frac{1}{2}\mathbb{E}\left[D(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\Big(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\Big)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right]. \tag{2.7.1}$$

In this case, we have

$$\mathbb{E}\big[MSE(\hat{p}_d) - \big\{g_d(\hat{\boldsymbol{\theta}}) + D^{-1}c_d(\hat{\boldsymbol{\theta}}) - D^{-1}b_d(\boldsymbol{\theta})\big\}\big] = o(D^{-1}).$$

Proposition 2.1 gives an approximation to the bias term b_d when $\hat{\boldsymbol{\theta}}$ is the truncated MM estimator.

Proposition 2.1 Let $\hat{\boldsymbol{\theta}}$ be the truncated MM estimator. Under regularity conditions (2.4.2)-(2.4.6), it holds that

$$b_d(\boldsymbol{\theta}) = B_d(\boldsymbol{\theta}) + o(1),$$

where

$$B_d(\boldsymbol{\theta}) = \frac{1}{2}\left\{\mathbb{E}[r_{D,d}] - \Big(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\Big)' \Big(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\Big)^{-1} \mathbb{E}[\boldsymbol{q}_D]\right\},$$

$$r_{D,d} = \Delta_D' \boldsymbol{R}_d(\boldsymbol{\theta})\Delta_D, \boldsymbol{R}_d(\boldsymbol{\theta}) = \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1},$$
$$\boldsymbol{q}_D = \underset{1\le k\le p+1}{\text{col}}(q_{Dk}),\ \ \boldsymbol{M}(\boldsymbol{\theta}) = \underset{1\le k\le p+1}{\text{col}}(M_k(\boldsymbol{\theta})),\ \ \hat{\boldsymbol{M}} = \underset{1\le k\le p+1}{\text{col}}(\hat{M}_k),$$
$$q_{Dk} = \Delta_D' \boldsymbol{Q}(\boldsymbol{\theta})\Delta_D,\ \boldsymbol{Q}(\boldsymbol{\theta}) = \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} M_k(\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1},$$
$$\Delta_D = \sqrt{D}(\hat{\boldsymbol{M}} - \boldsymbol{M}(\boldsymbol{\theta})),\ \ \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta}) = \left(\frac{\partial}{\partial \boldsymbol{\theta}_{k_2}} \boldsymbol{M}_{k_1}(\boldsymbol{\theta})\right)_{k_1,k_2=1,\ldots,p+1}.$$

Proof. A first order multivariate Taylor expansion of $\boldsymbol{M}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ yields to

$$\boldsymbol{M}(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|\big).$$

Therefore

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big) + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|\big). \tag{2.7.2}$$

Let us consider a second-order Taylor expansion of the kth component of $\boldsymbol{M}(\hat{\boldsymbol{\theta}})$, denoted by $\boldsymbol{M}_k(\hat{\boldsymbol{\theta}})$, around $\boldsymbol{\theta}$, i.e.

$$\boldsymbol{M}_k(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}_k(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}_k(\boldsymbol{\theta})\right)'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} M_k(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
$$+ o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big)$$

By substituting (2.7.2) in the quadratic term, we have

$$\boldsymbol{M}_k(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}_k(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}_k(\boldsymbol{\theta})\right)'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \frac{1}{2}\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)' \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)'$$
$$\cdot \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} M_k(\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big) + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big).$$

The corresponding multivariate Taylor expansion of $\boldsymbol{M}(\hat{\boldsymbol{\theta}})$ around $\boldsymbol{\theta}$ is

$$\boldsymbol{M}(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$$
$$+ \frac{1}{2} \underset{1\le k\le p+1}{\text{col}} \left(\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)' \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} M_k(\boldsymbol{\theta})\right)\right.$$
$$\left. \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)\right) + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big).$$

The above Taylor expansion can be rewritten as

$$\boldsymbol{M}(\hat{\boldsymbol{\theta}}) = \boldsymbol{M}(\boldsymbol{\theta}) + \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + \frac{1}{2D}\boldsymbol{q}_D + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big).$$

Therefore

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \left[\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big) - \frac{1}{2D}\boldsymbol{q}_D\right] + o\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big). \tag{2.7.3}$$

By substituting (2.7.3) in the expression of $b_d(\boldsymbol{\theta})$, given in (2.7.1), we obtain

$$\begin{aligned} b_d(\boldsymbol{\theta}) = {} & \left(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\right)' D \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \left\{\mathbb{E}[\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})] - \frac{1}{2D}\mathbb{E}[\boldsymbol{q}_D]\right\} \\ & + \frac{1}{2}\mathbb{E}\left[\sqrt{D}\left[\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)' - \frac{1}{2D}\boldsymbol{q}_D'\right]\left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\right)\right. \\ & \left. \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \sqrt{D}\left[\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big) - \frac{1}{2D}\boldsymbol{q}_D\right]\right] + Do\big(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\big). \end{aligned}$$

On the one hand, we substitute $\boldsymbol{M}(\hat{\boldsymbol{\theta}})$ by $\hat{\boldsymbol{M}}$, so that $\mathbb{E}[\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})] = \mathbb{E}[\hat{\boldsymbol{M}} - \boldsymbol{M}(\boldsymbol{\theta})] = 0$ by taking expectations in the natural equations of the MM algorithm. On the other hand, all the quadratic forms in the second summand containing $\boldsymbol{q}_D$ are $o(1)$. Therefore

$$\begin{aligned} b_d(\boldsymbol{\theta}) = {} & -\frac{1}{2}\left(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\right)' \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \mathbb{E}[\boldsymbol{q}_D] \\ & + \frac{1}{2}\mathbb{E}\left[\sqrt{D}\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)' \left(\left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1}\right)' \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} g_d(\boldsymbol{\theta})\right)\right. \\ & \left. \cdot \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \sqrt{D}\big(\boldsymbol{M}(\hat{\boldsymbol{\theta}}) - \boldsymbol{M}(\boldsymbol{\theta})\big)\right] + o(1) \\ = {} & \frac{1}{2}\left\{\mathbb{E}[r_{D,d}] - \left(\frac{\partial}{\partial \boldsymbol{\theta}} g_d(\boldsymbol{\theta})\right)' \left(\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{M}(\boldsymbol{\theta})\right)^{-1} \mathbb{E}[\boldsymbol{q}_D]\right\} + o(1) = B_d(\boldsymbol{\theta}) + o(1). \end{aligned}$$

□

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 $v_d^{*(b)} \sim N(0,1)$, $d = 1, \ldots, D$. Calculate $p_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$ and generate $y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)})$, $d = 1, \ldots, D$, $b = 1, \ldots, B$.

3. For each bootstrap resample b, calculate $\Delta_D^{*(b)} = \sqrt{D}(\hat{\boldsymbol{M}}^{*(b)} - \boldsymbol{M}(\hat{\boldsymbol{\theta}}))$, where $\hat{\boldsymbol{M}}^{*(b)} = \underset{1\leq k\leq p+1}{\text{col}}(\hat{M}_k^{*(b)})$, $\hat{M}_k^{*(b)} = \sum_{d=1}^D y_d^{*(b)} x_{dk}$, $k = 1,\ldots,p$, $\hat{M}_{p+1}^{*(b)} = \sum_{d=1}^D y_d^{*(b)2}$, and calculate

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

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

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

Theorem 2.2 Let $\hat{\boldsymbol{\theta}}$ be the truncated MM estimator. Under regularity conditions (2.4.2)-(2.4.6), an order $o(D^{-1})$ theoretical estimator of $MSE(\hat{p}_d)$, with bias correction, is

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

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

$$mse(\hat{p}_d) = mse^P(\hat{p}_d) - \frac{1}{D}\hat{B}_d(\hat{\boldsymbol{\theta}}), \quad mse^P(\hat{p}_d) = \hat{g}_d(\hat{\boldsymbol{\theta}}) + \frac{1}{D}\hat{c}_d(\hat{\boldsymbol{\theta}}), \tag{2.7.4}$$

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.

2.7.2 Auxiliary results

Let us now calculate the partial derivatives appearing in $\hat{B}_d(\hat{\boldsymbol{\theta}})$. Concerning the derivatives of $\boldsymbol{M}(\boldsymbol{\theta}) = \underset{1\leq k\leq p+1}{\text{col}}(M_k(\boldsymbol{\theta}))$, the first order partial derivatives can be found in Section 2.3.1. The second order partial derivatives of M_k $(k = 1,\ldots,p)$, are

$$\begin{aligned}
\frac{\partial^2 M_k}{\partial\beta_s\partial\beta_r} &= \sum_{d=1}^D \nu_d x_{dk}x_{dr}x_{ds}\exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\},\\
\frac{\partial^2 M_k}{\partial\phi\partial\beta_r} &= \sum_{d=1}^D \nu_d x_{dk}x_{dr}\phi\exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\},\\
\frac{\partial^2 M_k}{\partial\phi^2} &= \sum_{d=1}^D \nu_d x_{dk}\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \frac{1}{2}\phi^2\},
\end{aligned}$$

and second order partial derivatives of M_{p+1} are

$$\frac{\partial^2 M_{p+1}}{\partial\beta_s\partial\beta_r} = \sum_{d=1}^{D}\nu_d x_{dr}x_{ds}\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\frac{1}{2}\phi^2\}+4\sum_{d=1}^{D}\nu_d^2 x_{dr}x_{ds}\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\},$$
$$\frac{\partial^2 M_{p+1}}{\partial\phi\partial\beta_r} = \sum_{d=1}^{D}\nu_d x_{dr}\phi\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\frac{1}{2}\phi^2\}+8\sum_{d=1}^{D}\nu_d^2 x_{dr}\phi\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\},$$
$$\frac{\partial^2 M_{p+1}}{\partial\phi^2} = \sum_{d=1}^{D}\nu_d\phi^2\exp\{\boldsymbol{x}_d\boldsymbol{\beta}+\frac{1}{2}\phi^2\}+16\sum_{d=1}^{D}\nu_d^2\phi^2\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\},$$

where $r,s=1,\ldots,p$.

We recall that the MSE of the BP $\hat{p}_d(\boldsymbol{\theta})$ is

$$MSE(\hat{p}_d(\boldsymbol{\theta})) = g_d(\boldsymbol{\theta}) = \exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\}-\sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})P_d(j,\boldsymbol{\theta}).$$

The first order partial derivatives of $g_d(\boldsymbol{\theta})$ are

$$\frac{\partial g_d(\boldsymbol{\theta})}{\partial\beta_r} = 2x_{dr}\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\}-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_r}P_d(j,\boldsymbol{\theta})-\sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\beta_r},$$
$$\frac{\partial g_d(\boldsymbol{\theta})}{\partial\phi} = 4\phi\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\}-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\phi}P_d(j,\boldsymbol{\theta})-\sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\phi},$$

and the second order partial derivatives of $g_d(\boldsymbol{\theta})$ are

$$\begin{aligned}\frac{\partial^2 g_d(\boldsymbol{\theta})}{\partial\beta_s\partial\beta_r} &= 4x_{dr}x_{ds}\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\}-2\sum_{j=0}^{\infty}\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_s}\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_r}P_d(j,\boldsymbol{\theta})\\ &\quad-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial^2\psi_d(j,\boldsymbol{\theta})}{\partial\beta_s\partial\beta_r}P_d(j,\boldsymbol{\theta})-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_r}\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\beta_s}\\ &\quad-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_s}\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\beta_r}-\sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})\frac{\partial^2 P_d(j,\boldsymbol{\theta})}{\partial\beta_s\partial\beta_r},\\ \frac{\partial^2 g_d(\boldsymbol{\theta})}{\partial\phi\partial\beta_r} &= 8x_{dr}\phi\exp\{2\boldsymbol{x}_d\boldsymbol{\beta}+2\phi^2\}-2\sum_{j=0}^{\infty}\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\phi}\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_r}P_d(j,\boldsymbol{\theta})\\ &\quad-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial^2\psi_d(j,\boldsymbol{\theta})}{\partial\phi\partial\beta_r}P_d(j,\boldsymbol{\theta})-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\beta_r}\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\phi}\\ &\quad-2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial\psi_d(j,\boldsymbol{\theta})}{\partial\phi}\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial\beta_r}-\sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})\frac{\partial^2 P_d(j,\boldsymbol{\theta})}{\partial\phi\partial\beta_r},\end{aligned}$$

$$\frac{\partial^2 g_d(\boldsymbol{\theta})}{\partial \phi^2} = 16\phi^2 \exp\{2\boldsymbol{x}_d\boldsymbol{\beta} + 2\phi^2\} - 2\sum_{j=0}^{\infty}\left(\frac{\partial \psi_d(j,\boldsymbol{\theta})}{\partial \phi}\right)^2 P_d(j,\boldsymbol{\theta}) - 2\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial^2 \psi_d(j,\boldsymbol{\theta})}{\partial \phi^2}P_d(j,\boldsymbol{\theta}) - 4\sum_{j=0}^{\infty}\psi_d(j,\boldsymbol{\theta})\frac{\partial \psi_d(j,\boldsymbol{\theta})}{\partial \phi}\frac{\partial P_d(j,\boldsymbol{\theta})}{\partial \phi} - \sum_{j=0}^{\infty}\psi_d^2(j,\boldsymbol{\theta})\frac{\partial^2 P_d(j,\boldsymbol{\theta})}{\partial \phi^2}.$$

The derivatives of $\hat{g}_d(\hat{\boldsymbol{\theta}})$ are obtained by substituting ψ_d, p_d, $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\theta}$ by $\hat{\psi}_d$, $\hat{p}_d$, $\hat{\boldsymbol{\beta}}$, $\hat{\phi}$ and $\hat{\boldsymbol{\theta}}$ respectively in the corresponding derivatives of $g_d(\boldsymbol{\theta})$.

Let us now calculate the unknown components appearing in the partial derivatives of $g_d(\boldsymbol{\theta})$. For ease of presentation, we use a more simple notation $N_d(j) = N_d(j,\boldsymbol{\theta})$ and $D_d(j) = D_d(j,\boldsymbol{\theta})$. The first order partial derivatives of $\psi_d(j,\boldsymbol{\theta}) = \frac{N_d(j,\boldsymbol{\theta})}{D_d(j,\boldsymbol{\theta})} = \frac{N_d(j)}{D_d(j)}$ are

$$\frac{\partial \psi_d(j,\boldsymbol{\theta})}{\partial \theta_r} = \frac{\frac{\partial N_d(j)}{\partial \theta_r}}{D_d(j)} - \frac{N_d(j)\frac{\partial D_d(j)}{\partial \theta_r}}{D_d^2(j)}, \quad r = 1,\ldots,p+1, \tag{2.7.5}$$

and the second order partial derivatives of $\psi_d(j,\boldsymbol{\theta})$ are

$$\frac{\partial \psi_d(j,\boldsymbol{\theta})}{\partial \theta_s \partial \theta_r} = \frac{\frac{\partial^2 N_d(j)}{\partial \theta_s \partial \theta_r}}{D_d(j)} - \frac{\frac{\partial N_d(j)}{\partial \theta_r}\frac{\partial D_d(j)}{\partial \theta_s} + \frac{\partial N_d(j)}{\partial \theta_s}\frac{\partial D_d(j)}{\partial \theta_r} + N_d(j)\frac{\partial^2 D_d(j)}{\partial \theta_s \partial \theta_r}}{D_d^2(j)} + \frac{2N_d(j)\frac{\partial D_d(j)}{\partial \theta_s}\frac{\partial D_d(j)}{\partial \theta_r}}{D_d^3(j)}. \tag{2.7.6}$$

Finally, the estimated bias correction term $\hat{B}_d(\hat{\boldsymbol{\theta}})$ will be determined if we calculate the partial derivatives appearing in (2.7.5) and (2.7.6). We recall that

$$D_d(j,\boldsymbol{\theta}) = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d) f(v_d)\, dv_d, \quad N_d(j,\boldsymbol{\theta}) = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j+1,v_d) f(v_d)\, dv_d = D_d(j+1,\boldsymbol{\theta}),$$

where

$$R_d(\boldsymbol{\theta},j,v_d) = \exp\left\{j(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d p_d(\boldsymbol{\theta},v_d)\right\}, \quad p_d(\boldsymbol{\theta},v_d) = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}.$$

The first order partial derivatives of $D_d(j,\boldsymbol{\theta})$ are

$$\frac{\partial D_d(j,\boldsymbol{\theta})}{\partial \beta_r} = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d)[j - \nu_d p_d(\boldsymbol{\theta},v_d)]x_{dr} f(v_d)\, dv_d, \ r = 1,\ldots,p,$$
$$\frac{\partial D_d(j,\boldsymbol{\theta})}{\partial \phi} = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d)[j - \nu_d p_d(\boldsymbol{\theta},v_d)]v_d f(v_d)\, dv_d,$$

and the second order partial derivatives of $D_d(j,\boldsymbol{\theta})$ are

$$\frac{\partial^2 D_d(j,\boldsymbol{\theta})}{\partial\beta_s\partial\beta_r} = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d)\left\{[j-\nu_d p_d(\boldsymbol{\theta},v_d)]^2 - \nu_d p_d(\boldsymbol{\theta},v_d)\right\} x_{dr}x_{ds}f(v_d)\,dv_d,$$
$$\frac{\partial^2 D_d(j,\boldsymbol{\theta})}{\partial\phi\partial\beta_r} = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d)\left\{[j-\nu_d p_d(\boldsymbol{\theta},v_d)]^2 - \nu_d p_d(\boldsymbol{\theta},v_d)\right\} x_{dr}v_d f(v_d)\,dv_d,$$
$$\frac{\partial^2 D_d(j,\boldsymbol{\theta})}{\partial\phi^2} = \int_{\mathbb{R}} R_d(\boldsymbol{\theta},j,v_d)\left\{[j-\nu_d p_d(\boldsymbol{\theta},v_d)]^2 - \nu_d p_d(\boldsymbol{\theta},v_d)\right\} v_d^2 f(v_d)\,dv_d.$$

where $r,s=1,\ldots,p$. The first and second order partial derivatives of $N_d(j,\boldsymbol{\theta})$ are obtained by changing j by $(j+1)$ in the corresponding derivatives of $D_d(j,\boldsymbol{\theta})$. It is immediate to see that the partial derivatives of $N_d(j,\boldsymbol{\theta})$ and $D_d(j,\boldsymbol{\theta})$ can be approximated by an accelerated Monte Carlo algorithm analogous to that seen in Section 2.5.

As the partial derivatives of $g_d(\boldsymbol{\theta})$ are computationally demanding, we propose alternative efficient formulas (2.7.7) – (2.7.9) taken from Lahiri et al. (2007). For that, let $h:\mathbb{R}^n\mapsto\mathbb{R}$ be a twice continuously differentiable real-valued function. Let us define the column vectors $\boldsymbol{\theta}=\underset{1\le r\le n}{\text{col}}\,(\theta_r)\in\mathbb{R}^n$, $\boldsymbol{e}_r=(0,\ldots,0,1^{(r},0,\ldots,0)'=\underset{1\le i\le n}{\text{col}}\,(\delta_{ir})$, $\boldsymbol{e}_{rs}=\boldsymbol{e}_r+\boldsymbol{e}_s$, where $\delta_{ij}=0$ if $i\neq j$ and $\delta_{ij}=1$ if $i=j$. For $\varepsilon>0$, a first order Taylor expansion of $h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)$ and $h(\boldsymbol{\theta}-\varepsilon\boldsymbol{e}_r)$ around θ yields to

$$h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)=h(\boldsymbol{\theta})+\frac{\partial h(\boldsymbol{\theta})}{\partial\theta_r}\,\varepsilon+o(\varepsilon),\quad h(\boldsymbol{\theta}-\varepsilon\boldsymbol{e}_r)=h(\boldsymbol{\theta})-\frac{\partial h(\boldsymbol{\theta})}{\partial\theta_r}\,\varepsilon+o(\varepsilon).$$

By subtraction, we get

$$h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)-h(\boldsymbol{\theta}-\varepsilon\boldsymbol{e}_r)=2\varepsilon\frac{\partial h(\boldsymbol{\theta})}{\partial\theta_r}+o(\varepsilon).$$

Therefore

$$\frac{\partial h(\boldsymbol{\theta})}{\partial\theta_r}=\frac{1}{2\varepsilon}\big\{h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)-h(\boldsymbol{\theta}-\varepsilon\boldsymbol{e}_r)\big\}+o(\varepsilon). \tag{2.7.7}$$

For $\varepsilon>0$, a second order Taylor expansion of $h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)$ around θ yields to

$$h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)=h(\boldsymbol{\theta})+\frac{\partial h(\boldsymbol{\theta})}{\partial\theta_r}\,\varepsilon+\frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_r^2}\,\varepsilon^2+o(\varepsilon^2).$$

By applying (2.7.7), we get

$$h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)=h(\boldsymbol{\theta})+\frac{1}{2}\big\{h(\boldsymbol{\theta}+\varepsilon\boldsymbol{e}_r)-h(\boldsymbol{\theta}-\varepsilon\boldsymbol{e}_r)\big\}+\frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial\theta_r^2}\,\varepsilon^2+o(\varepsilon^2).$$

Therefore

$$\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r^2} = \frac{1}{\varepsilon^2}\big\{h(\boldsymbol{\theta}+\varepsilon \boldsymbol{e}_r) + h(\boldsymbol{\theta}-\varepsilon \boldsymbol{e}_r) - 2h(\boldsymbol{\theta})\big\} + o(\varepsilon^2). \tag{2.7.8}$$

For $\varepsilon > 0$, a second order Taylor expansion of $h(\boldsymbol{\theta}+\varepsilon \boldsymbol{e}_{rs})$ and $h(\boldsymbol{\theta}-\varepsilon \boldsymbol{e}_{rs})$ around θ yields to

$$h(\boldsymbol{\theta}+\varepsilon \boldsymbol{e}_{rs}) = h(\boldsymbol{\theta}) + \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_r}\varepsilon + \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_s}\varepsilon + \frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r^2}\varepsilon^2 + \frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_s^2}\varepsilon^2 + \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_r}\varepsilon^2 + o(\varepsilon^2),$$
$$h(\boldsymbol{\theta}-\varepsilon \boldsymbol{e}_{rs}) = h(\boldsymbol{\theta}) - \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_r}\varepsilon - \frac{\partial h(\boldsymbol{\theta})}{\partial \theta_s}\varepsilon + \frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r^2}\varepsilon^2 + \frac{1}{2}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_s^2}\varepsilon^2 + \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_r}\varepsilon^2 + o(\varepsilon^2).$$

By summation, we get

$$h(\boldsymbol{\theta}+\varepsilon \boldsymbol{e}_{rs}) + h(\boldsymbol{\theta}-\varepsilon \boldsymbol{e}_{rs}) = 2h(\boldsymbol{\theta}) + \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r^2}\varepsilon^2 + \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_s^2}\varepsilon^2 + 2\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_r}\varepsilon^2 + o(\varepsilon^2).$$

For $r \neq s$, we obtain

$$\begin{aligned}\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r \partial \theta_r} = \frac{1}{2\varepsilon^2}\Big[&\big\{h(\boldsymbol{\theta}+\varepsilon \boldsymbol{e}_{rs}) + h(\boldsymbol{\theta}-\varepsilon \boldsymbol{e}_{rs}) - 2h(\boldsymbol{\theta})\big\} \\ &-\varepsilon^2\left\{\frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_r^2} + \frac{\partial^2 h(\boldsymbol{\theta})}{\partial \theta_s^2}\right\}\Big] + o(\varepsilon^2).\end{aligned} \tag{2.7.9}$$

2.7.3 Bootstrap estimation of the MSE

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 $(b = 1, \ldots, B)$

   i) Generate $v_d^{*(b)} \sim N(0,1)$, $d = 1, \ldots, D$. Calculate $p_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$ and $y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)})$.

   ii) For each bootstrap resample, calculate the estimator $\hat{\boldsymbol{\theta}}^{*(b)}$ and the EBP $\hat{p}_d^{*(b)} = \hat{p}_d^*(\hat{\boldsymbol{\theta}}^{*(b)})$.

3. Calculate

$$mse^*(\hat{p}_d) = \frac{1}{B}\sum_{b=1}^{B}\big(\hat{p}_d^{*(b)} - p_d^{*(b)}\big)^2. \tag{2.7.10}$$

2.8 Simulation experiments

Several simulation experiments are carried out for analysing the fitting algorithms, the EBP of p_d and the MSE estimators. Two approaches are considered: simulations based on the model distribution and simulations based on a sampling design distribution.

2.8.1 Model-based simulation

This subsection presents three simulation experiments related to the application to the real data from SLCS in 2008 (see more details in Section 2.9.1). First, we analyse the behaviour of the MM, PQL and LA fitting algorithms. The simulation is also employed for checking that the developed R codes work efficiently. The function glmer of the R package lme4 gives the ML-Laplace approximation algorithm. 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 (*cit1*), people with age in $50-64$ (*age3*) 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 $p_d = \exp\{\beta_0 + lab2_d\beta_1 + cit1_d\beta_2 + age3_d\beta_3 + edu23_d\beta_4 + \phi v_d\}$, $d = 1, \ldots, D$. 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 2.9.1. 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 2.8.1 (for normal random effects) and Table 2.8.2 (for Gumbel random effects) present the bias and the root-MSE (RMSE) in brackets for MM, PQL and LA estimators, i.e.

$$BIAS = \frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta), \;\; RMSE = \left(\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta)^2\right)^{1/2}.$$

These tables suggest that RMSE is slightly higher for Gumbel random effects. In general, BIAS is lower for the MM and LA estimators. On the other hand, the RMSEs for PQL and LA estimators are similar, but they are higher for MM estimators. As expected, when the number of domains increases then the bias and the RMSE decreases. The empirical results agree with the consistency property of the MM estimators. Tables 2.8.1 and 2.8.2 suggest that the variance is the component that contributes most to the MSE since bias is much smaller than the RMSE.

Table 2.8.1: BIAS and RMSE (in brackets) for MM, PQL and LA estimators taking normal random effects.

D	$\hat{\theta}$	MM	PQL	LA
52	$\hat{\beta}_0$	0.0284 (0.6204)	0.0318 (0.5742)	0.0113 (0.5736)
	$\hat{\beta}_1$	0.0001 (2.4308)	0.0009 (2.2285)	0.0610 (2.2211)
	$\hat{\beta}_2$	-0.0212 (0.7144)	-0.0265 (0.6687)	-0.0296 (0.6681)
	$\hat{\beta}_3$	-0.0817 (3.3998)	-0.0247 (3.0507)	0.0047 (3.0431)
	$\hat{\beta}_4$	-0.0227 (0.8213)	-0.0316 (0.7638)	-0.0233 (0.7659)
	$\hat{\phi}$	-0.0362 (0.1175)	-0.0817 (0.0872)	-0.0167 (0.0328)
104	$\hat{\beta}_0$	-0.0033 (0.4252)	0.0159 (0.3877)	0.0018 (0.3861)
	$\hat{\beta}_1$	-0.0534 (1.5785)	-0.1036 (1.4599)	-0.0552 (1.4522)
	$\hat{\beta}_2$	0.0184 (0.4723)	0.0189 (0.4281)	0.0160 (0.4237)
	$\hat{\beta}_3$	0.0771 (2.1812)	0.0634 (1.9584)	0.0590 (1.9479)
	$\hat{\beta}_4$	-0.0124 (0.4708)	-0.0225 (0.4331)	-0.0157 (0.4311)
	$\hat{\phi}$	-0.0313 (0.1006)	-0.0787 (0.0820)	-0.0091 (0.0230)
150	$\hat{\beta}_0$	0.0094 (0.3542)	0.0235 (0.3280)	0.0078 (0.3275)
	$\hat{\beta}_1$	-0.0242 (1.2636)	-0.0717 (1.1803)	-0.0232 (1.1773)
	$\hat{\beta}_2$	-0.0036 (0.3979)	-0.0087 (0.3643)	-0.0079 (0.3635)
	$\hat{\beta}_3$	-0.0163 (1.8200)	-0.0203 (1.6452)	-0.0201 (1.6445)
	$\hat{\beta}_4$	-0.0083 (0.4018)	-0.0115 (0.3676)	-0.0043 (0.3678)
	$\hat{\phi}$	-0.0257 (0.0897)	-0.0747 (0.0770)	-0.0057 (0.0177)

The second simulation studies the behaviour 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), as well as the EBP, 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 Step 2 of the EBP algorithm given in Section 2.5.

Table 2.8.2: BIAS and RMSE (in brackets) for MM, PQL and LA estimators taking Gumbel random effects.

D	$\hat{\theta}$	MM	PQL	LA
52	$\hat{\beta}_0$	-0.0468 (0.6884)	-0.0129 (0.5809)	-0.0293 (0.5741)
	$\hat{\beta}_1$	0.1136 (2.6604)	0.0385 (2.2745)	0.0888 (2.2505)
	$\hat{\beta}_2$	0.0048 (0.7285)	0.0078 (0.6600)	0.0052 (0.6583)
	$\hat{\beta}_3$	0.1046 (3.7879)	0.0701 (3.1537)	0.0749 (3.1160)
	$\hat{\beta}_4$	0.0376 (0.8753)	0.0109 (0.7621)	0.0200 (0.7566)
	$\hat{\phi}$	-0.0255 (0.1288)	-0.0725 (0.0829)	-0.0090 (0.0371)
104	$\hat{\beta}_0$	-0.0036 (0.4601)	0.0207 (0.4109)	0.0053 (0.4077)
	$\hat{\beta}_1$	0.0003 (1.6610)	-0.0805 (1.4962)	-0.0278 (1.4791)
	$\hat{\beta}_2$	-0.0076 (0.4916)	-0.0092 (0.4298)	-0.0147 (0.4260)
	$\hat{\beta}_3$	-0.0189 (2.3000)	-0.0429 (2.0067)	-0.0461 (1.9935)
	$\hat{\beta}_4$	0.0091 (0.5306)	-0.0007 (0.4703)	0.0076 (0.4660)
	$\hat{\phi}$	-0.0246 (0.1141)	-0.0719 (0.0777)	-0.0031 (0.0262)
150	$\hat{\beta}_0$	0.0029 (0.3726)	0.0185 (0.3263)	0.0014 (0.3240)
	$\hat{\beta}_1$	-0.0354 (1.3296)	-0.0743 (1.1802)	-0.0167 (1.1677)
	$\hat{\beta}_2$	-0.0045 (0.4351)	-0.0045 (0.3821)	-0.0038 (0.3778)
	$\hat{\beta}_3$	-0.0168 (1.9787)	-0.0023 (1.7116)	-0.0037 (1.7009)
	$\hat{\beta}_4$	-0.0002 (0.4470)	-0.0104 (0.3963)	-0.0017 (0.3943)
	$\hat{\phi}$	-0.0143 (0.1008)	-0.0667 (0.0711)	0.0013 (0.0214)

Table 2.8.3 for normal and Table 2.8.4 for Gumbel random effects compare these estimators through the bias, B_d, and the root mean squared error RE_d (in brackets), i.e.

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

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 RMSEs (in brackets), i.e.

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

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 RE_d's of the EBP are close to PLUG2. If we move from normal to Gumbel distribution we get a moderate increase of RMSE for the three consider estimators of p_d. Again B_d is much smaller than RE_d, so the variance is, by far, the most important part of the MSE.

Table 2.8.3: B_d and RE_d in brackets (both $\times 10^3$) for the estimators of p_d using normal random effects.

D	d	p_d	PLUG1	PLUG2	EBP
52	12	0.1358	-0.6625 (22.9513)	-2.0825 (23.9080)	-1.2805 (23.8648)
	22	0.2199	-0.7860 (31.7712)	-4.8846 (35.9817)	-4.0407 (35.9576)
	32	0.1473	-0.0736 (20.2905)	-0.8316 (22.3833)	-0.2971 (22.3528)
	42	0.1390	-1.1504 (17.0833)	-1.8658 (20.5645)	-1.5098 (20.5429)
$B\,(RE)$			0.9537 (30.2549)	2.8992 (35.0194)	2.3260 (35.0673)
104	22	0.2043	-0.5402 (32.7631)	-2.2054 (33.5076)	-1.0222 (33.4617)
	43	0.2902	-2.5915 (36.4277)	-4.1313 (40.2427)	-3.0267 (40.1638)
	63	0.3341	-1.2321 (34.6773)	-0.2562 (40.9947)	0.5668 (40.9594)
	84	0.1346	-0.2059 (16.1346)	-1.0437 (17.4878)	-0.6061 (17.4812)
$B\,(RE)$			0.8679 (27.6212)	2.2133 (30.5543)	1.4937 (30.6393)
150	31	0.2980	-0.2759 (42.7971)	-0.6607 (46.3047)	0.7313 (46.3008)
	61	0.2883	-0.9757 (36.6614)	-1.1722 (40.0554)	-0.0712 (40.0093)
	91	0.1183	-1.3843 (16.0990)	-2.2849 (17.2439)	-1.7292 (17.2047)
	121	0.1364	-1.1466 (15.5221)	-2.2030 (17.7481)	-1.7813 (17.7208)
$B\,(RE)$			1.1842 (27.6926)	2.3748 (30.5481)	1.7497 (30.6419)

The third simulation investigates the behaviour 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

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

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

Table 2.8.4: B_d and RE_d in brackets (both $\times 10^3$) for the estimators of p_d using Gumbel random effects.

D	d	p_d	PLUG1	PLUG2	EBP
52	12	0.1370	-1.0361 (24.6185)	-2.3661 (27.0481)	-1.5873 (27.0729)
	22	0.2217	-3.1392 (33.2708)	-6.2361 (40.4207)	-5.4267 (40.3495)
	32	0.1488	-0.8530 (21.0478)	-1.8381 (24.9230)	-1.2984 (24.8738)
	42	0.1382	0.6079 (16.2585)	-0.8915 (21.6641)	-0.5114 (21.6722)
$B\,(RE)$			1.1290 (31.2405)	3.0348 (37.6639)	2.4642 (37.7298)
104	22	0.2046	-0.2642 (32.0103)	-2.8320 (34.2867)	-1.5970 (34.2506)
	43	0.2912	-2.2316 (40.3293)	-4.3607 (47.3573)	-3.2026 (47.3121)
	63	0.3331	-1.3900 (35.4902)	-1.6936 (42.9036)	-0.8724 (42.8879)
	84	0.1358	-0.3168 (16.7957)	-1.3039 (21.3683)	-0.8217 (21.3281)
$B\,(RE)$			0.9791 (28.9698)	2.4634 (28.9698)	1.7554 (33.4948)
150	31	0.2990	-0.6858 (44.9199)	-1.3983 (49.8233)	0.1416 (49.9213)
	61	0.2888	-2.0584 (36.8686)	-2.8549 (41.5932)	-1.6780 (41.5041)
	91	0.1172	-0.9483 (16.9548)	-1.8923 (18.4209)	-1.2805 (18.3762)
	121	0.1347	-0.1022 (16.6420)	-1.4287 (20.5855)	-0.9755 (20.5620)
$B\,(RE)$			1.0275 (28.6226)	2.4008 (32.3556)	1.6388 (32.4816)

(2.7.4), mse^P and mse without and with bias correction respectively, and the parametric bootstrap estimator, mse^*, introduced in (2.7.10). 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 2.8.1 plots the logarithm of the MSE estimators for each domain $d = 1, \ldots, D$ and for $D = 52$ (left), $D = 104$ (center) and $D = 150$ (right). They are sorted by sample size. The logarithm scale is used to improve the visualization of the estimators. The results for small values of d are quite similar. However, the bootstrap estimator shows a more stable behaviour when d increases. We note that the estimator with bias correction, mse, is a good alternative despite not being able to capture the bias of the plug-in estimator in the last domains. For the bootstrap approach, we consider $B = 500$ resamples.

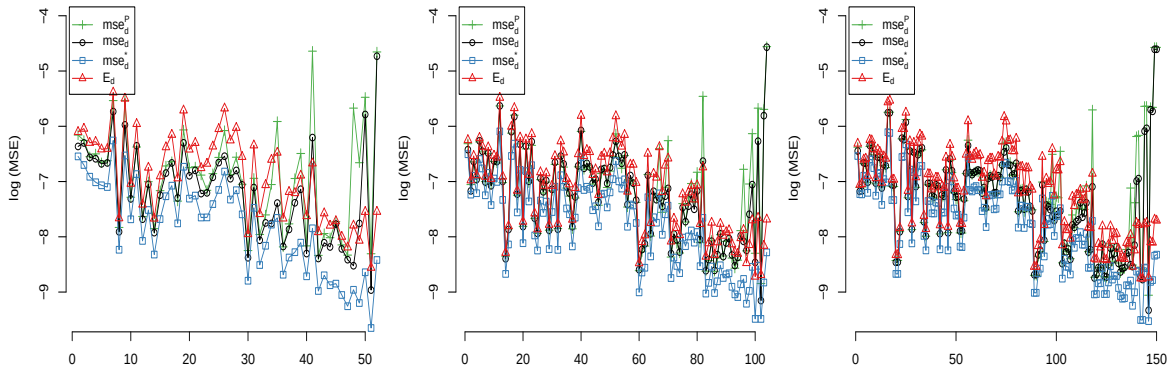

Figure 2.8.1: MSE estimators, in logarithmic scale, for $D = 52$ (left), $D = 104$ (center) and $D = 150$ (right).

Figure 2.8.2 shows 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 contain very few outliers.

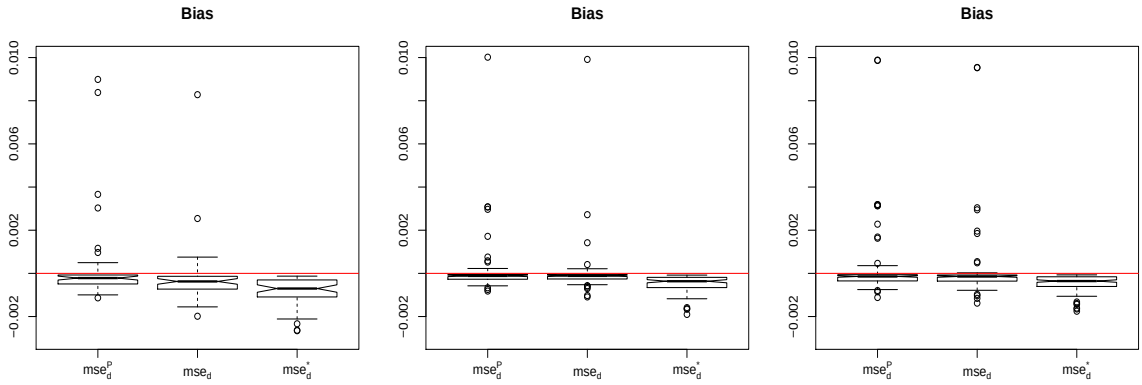

Figure 2.8.2: Bias of MSE estimators for $D = 52$ (left), $D = 104$ (center) and $D = 150$ (right).

Table 2.8.5 presents the bias and the root mean squared error (both $\times 10^3$) 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 root mean squared error. The

bootstrap MSE estimator has a similar root mean squared error to the analytic ones, it is computationally faster and it is easy to implement. However, it has a higher bias. The obtained results suggest that bias is the most important part of the MSE in the bootstrap approach, specially for small values of D.

Table 2.8.5: Bias and root mean squared error in brackets (both $\times 10^3$) of the MSE estimators.

D	d	E_d	mse^P	mse	mse^*
52	12	0.0006	-0.1052 (0.3291)	-0.1117 (0.4516)	-0.2904 (0.4056)
	22	0.0012	-0.1297 (0.3018)	-0.3782 (0.7953)	-0.6890 (0.7947)
	32	0.0005	-0.1559 (0.2238)	-0.1624 (0.3591)	-0.3059 (0.3459)
	42	0.0004	-0.1130 (0.1820)	-0.0991 (0.3100)	-0.2399 (0.2589)
104	22	0.0010	-0.1641 (0.4482)	-0.1325 (0.5495)	-0.3971 (0.5582)
	43	0.0016	-0.4271 (0.6183)	-0.3831 (0.7954)	-0.7590 (0.8740)
	63	0.0015	-0.4892 (0.6793)	-0.4398 (0.9407)	-0.8273 (0.8946)
	84	0.0003	-0.0857 (0.1269)	-0.0716 (0.1699)	-0.1533 (0.1707)
150	31	0.0019	-0.4192 (0.8080)	-0.3879 (0.9702)	-0.7737 (1.0145)
	61	0.0014	-0.3835 (0.6396)	-0.3201 (0.7839)	-0.6646 (0.8082)
	91	0.0003	-0.0527 (0.1280)	-0.0455 (0.1540)	-0.1109 (0.1541)
	121	0.0003	-0.1029 (0.1458)	-0.0934 (0.1756)	-0.1505 (0.1719)

The estimation of the MSE of small area predictors under different models is a relevant issue. In particular, the development of methods for obtaining bias-corrected MSE estimators is an open topic. In the SAE literature, the first bootstrap procedure for correcting the bias of MSE estimators is due to Hall and Maiti (2006). These authors proposed a parametric double-bootstrap procedure based on the conventional additively and multiplicatively bias-corrected estimators, such that the first step bootstrap estimator has a bias of order $O(1/D)$ and the double-bootstrap estimator has a bias of order $o(1/D)$ under some regularity conditions. Despite having bias of lower order, these corrections may leave appreciable bias for small samples. Pfeffermann and Correa (2012) developed a general method of bias correction, which models the MSEs of the target small area estimators as a function of the corresponding bootstrap MSE estimators, the small area estimators and the bootstrap estimators of the parameters of the model fitted to the sample data. The method is flexible and can be applied to complicated problems without new theoretical derivations, but the application under our particular Poisson mixed model involves an extensive simulation

study which is not the subject of the work at hand. Other procedures in the SAE literature are the jackknife method (Jiang et al., 2002) or the resampling methods for estimating the MSE of M-quantile estimators (Chambers et al., 2011), among others.

2.8.2 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_{dj}. The artificial population is built by repeating $\lfloor 10^{-3} w_{dj} \rfloor$ times each sampling unit j of domain d.

We implement a simplified version of the SLCS sampling design in 2008. 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

$$\hat{p}_d^{dir} = \frac{1}{\hat{N}_d} \sum_{j \in s_d} w_{dj}\, y_{dj},\ \widehat{\mathrm{var}}_\pi(\hat{p}_d^{dir}) = \frac{1}{\hat{N}_d^2} \sum_{j \in s_d} w_{dj}(w_{dj} - 1) \left(y_{dj} - \hat{p}_d^{dir} \right)^2, \tag{2.8.1}$$

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 and Rao (1990).

Table 2.8.6 gives the results of the bias B_d and the root mean squared error RE_d (in brackets) for 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 of the real population, where $D = 104$. As expected, the direct estimator has lower bias but its RMSE is higher than the model-based-estimators.

The FH has lower bias and greater RMSE than the PLUG1 predictor in most cases. The three Poisson mixed model predictors (EBP, PLUG1 and PLUG2) have a similar behaviour,

in both bias and root mean squared error. If we compare these results with those obtained in Table 2.8.3 under model-based simulation, they increase slightly. This fact is somehow expected but gives more realistic information about the behaviour of the considered predictors in practice.

Table 2.8.6: B_d and RE_d in brackets (both $\times 10^2$) for the estimators of p_d.

D	d	p_d	Dir	FH	PLUG1	PLUG2	EBP
104	22	0.3632	-0.3635	-5.8868	-6.2216	-8.1702	-8.1666
			(9.9666)	(7.3783)	(6.6154)	(8.4582)	(8.4548)
	43	0.1657	-0.1274	-0.5098	1.9462	0.6762	0.6786
			(6.3085)	(4.0864)	(2.3150)	(1.3859)	(1.3862)
	63	0.1010	-0.1625	0.0465	2.3047	1.8443	1.8461
			(3.8086)	(3.0336)	(2.4213)	(1.9763)	(1.9778)
	84	0.2182	0.0718	1.0257	1.8058	-0.0772	-0.0742
			(4.2012)	(3.3773)	(2.3373)	(1.4629)	(1.4637)
B			0.1877	2.6396	4.9789	5.0036	5.0036
(RE)			(6.9890)	(5.3960)	(5.2454)	(5.2875)	(5.2874)

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 (2.3.1)-(2.3.2) by Newton-Raphson and *evd* package to generate random effects according to a Gumbel distribution. Regarding the computational burden, the MM estimator is faster than PQL since its computing times were 0.02, 0.03 and 0.04 seconds for $D = 52, 104$ and 150 respectively while for the PQL were 0.02, 0.04 and 0.08. On the other hand, the PQL algorithm already provides the prediction of the random effects $\hat{v}_d$ $(d = 1, \ldots, D)$. Thus, the plug-in estimator of p_d obtained using PQL (PLUG1) is less demanding computationally. Its runtimes for $D = 52, 104, 150$ were 0.07, 0.10 and 0.17 seconds respectively. As the p_d estimators based on the MM fitting algorithm (PLUG2 and EBP) require the use of EBPs, their runtimes are higher (0.15, 0.20 and 0.36 seconds for PLUG2 and 0.15, 0.22 and 0.41 for the EBP). Finally, regarding the MSE estimators, the calculation of mse^P and mse is computationally intensive. Specially the mse estimator where the bias correction term, calculated by using a bootstrap approach, increases considerably the runtimes (from 151.14 seconds to 2231.80 for $D = 52$ and $B = 500$ resamples). In addition, the bootstrap estimator mse^* is easy to implement, it has a similar behaviour to the previous ones and requires a much lower runtime (15.84 seconds for D=52 and $B = 500$ resamples).

2.9 Applications to real data

2.9.1 Poverty 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 Spain by provinces during 2008. The data is taken from the SLCS in 2008. As domains, we consider provinces crossed by sex. In Spain there are 50 provinces. In addition, we consider as provinces the autonomous cities of Ceuta and Melilla. Then, the number of domains is $D = 52 \times 2 = 104$. The SLCS planned domains are the 17 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 deals with this problem by introducing model-based or model assisted estimators. The response variable y_d counts the number of people under the poverty line in the domain d. We assume that y_d can be described by an area-level Poisson mixed model and some explanatory variables (see Chapter 1 for a detailed explanation of the considered auxiliary information).

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 labour 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 the Spanish Office of Statistics.

An area-level Poisson mixed model (Model 1) is fitted to the 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 2.9.1 presents the estimates of the regression parameters and their standard errors, z-values and p-values.

The signs of the regression parameters in Table 2.9.1 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 $cit1$ 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

Table 2.9.1: MM estimates of regression parameters under Model 1.

Variable	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	1.5669	0.5030	3.7653	< 0.001
*lab*2	6.8923	1.8939	2.9949	0.0027
*cit*1	-2.9844	0.5860	-4.9693	< 0.001
*age*3	-7.5259	2.6311	-3.8857	< 0.001
*edu*23	-3.5998	0.5913	-5.3807	< 0.001

et al. (2012b) found the same result when fitting Fay-Herriot temporal models to data from the SLCS of 2006.

Each domain (province-sex) d, $d = 1, \ldots, 104$, has a random intercept with distribution $N(0, \phi^2)$. The estimate of ϕ is $\hat{\phi} = 0.183$ and its 95% percentile bootstrap confidence interval is $(0.144, 0.271)$. To test the null hypothesis $H_0 : \phi^2 = 0$, we use a bootstrap procedure. The steps are:

Algorithm 5 A bootstrap test for $H_0 : \phi^2 = 0$

1: Fit the Model 1 (see Section 2.2) to data and calculate $\hat{\boldsymbol{\beta}}$ and $\hat{\phi}$.

2: Fit the Model 0 (see Section 2.2) to data and calculate $\hat{\boldsymbol{\beta}}^0$.

3: For $b = 1, \ldots, B$, do

i) Generate a bootstrap resample under $H_0 : \phi^2 = 0$, i.e.

$$p_d^{*(b)} = \exp\{\boldsymbol{x}_d \hat{\boldsymbol{\beta}}^0\}, \; y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)}), \; d = 1, \ldots, D.$$

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

4: Calculate the p-value

$$p = \frac{\#\left\{\hat{\phi}^{*(b)2} > \hat{\phi}^2\right\}}{B}.$$

The obtained bootstrap p-value is 0.007. Then, we conclude that the variance parameter is significantly different from 0, taking $\alpha = 0.05$

For the sake of comparison, we also fit a Poisson regression model with the same auxiliary variables of Table 2.9.1 but without any random effect (Model 0). Figure 2.9.1 plots the Pearson residuals of the Poisson regression models without (left) and with (right) domain random effects. In both cases the behaviour is symmetrical around 0 and a clear improve-

ment is observed when we use the more complex model including random effects, as they capture the variability between domains.

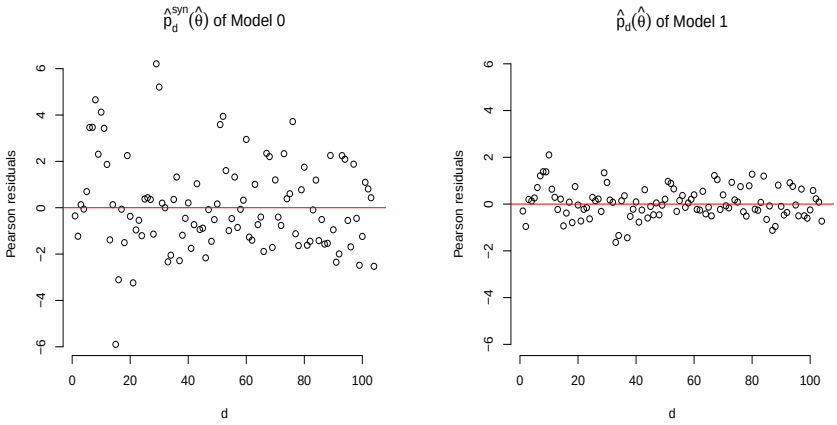

Figure 2.9.1: Pearson residuals for the fixed effects model (left) and mixed effects model (right).

The objective of this analysis 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 a Fay-Herriot model (Fay and Herriot, 1979) fitted by the REML method to the set of auxiliary variables described in Table 2.9.1. The MSE of the EBP is estimated by parametric bootstrap and the MSE of the EBLUP by the g_1-g_3 formula (see eq(4) in Datta and Lahiri (2000)). Finally, direct estimators of p_d and of its design-based variance are calculated following (2.8.1), 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 2.9.2 (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 2.9.2 (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 2.9.2 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.

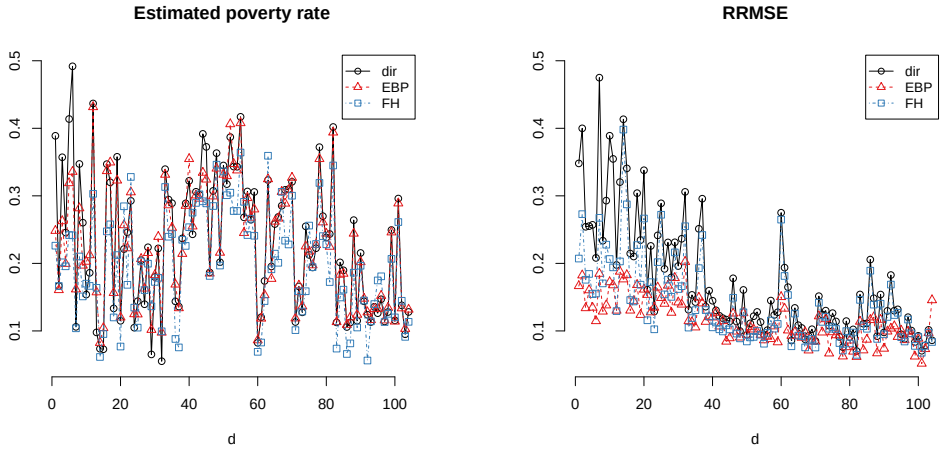

Figure 2.9.2: Direct, EBP and EBLUP estimates of p_d (left) and relative root-MSE (right) for the three estimators.

Table 2.9.2 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 (note that direct estimator is unbiased). 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, all estimates of p_d and their corresponding errors show a similar behaviour. The displayed results are in accordance with those shown in Figure 2.9.2.

Table 2.9.2: 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 2.9.3 (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 2.9.3 (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 behaviour of the direct estimators, but they could be in troubles for estimating the lowest or the highest poverty rates.

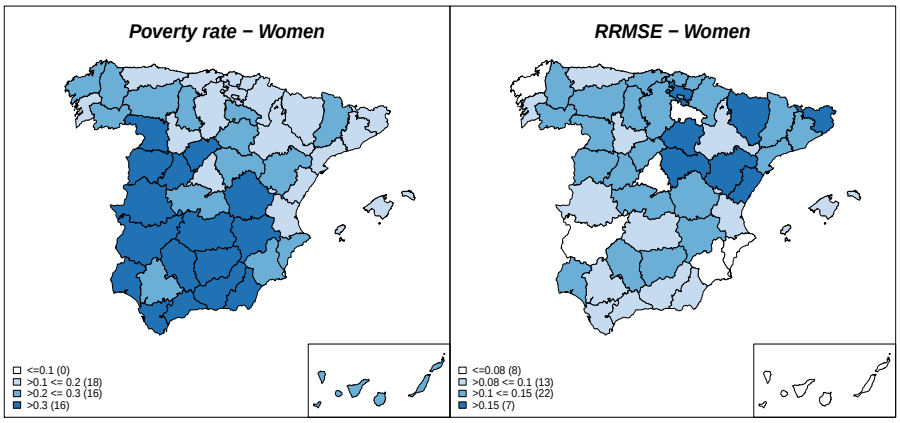

Figure 2.9.3: Poverty rate EBPs for women (left) and RRMSE (right) in 2008.

The Moran's test was applied to the residuals of the EBP of p_d to study a possible spatial autocorrelation. We use *moran.test* function in *spdep* 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 spatial autocorrelation is tested. The obtained p-value for women is 0.061. Taking as significance level $\alpha = 0.05$, the null hypothesis of no spatial autocorrelation is not rejected.

2.9.2 Forest fires data

The region of Galicia is in the northwest of Spain. It has a population of $2,795,422$ (5.9% of the Spanish population) and a surface area of $29,574$ km^2 (5.8% of Spain). Its forest area is $2,060,453$ ha. The ratio of the forest surface is notably higher than the national average (69%). Regarding the property regime of the land, most Galician forest is private (97.2%), and this percentage is much higher than the national average at 67.7% (Rodríguez-Vicente and Marey-Pérez, 2009b). Following the study of Marey-Pérez and Gómez-Vázquez (2010),

private forest ownership is subdivided into either particular ownership or communal ownership in collective woodlands (Montes Vecinales en Mano Común, MVMC), an ownership typology almost exclusive to Galicia.

Wildfires in Galicia are a recurrent problem and show increasing levels of severity. There were 249,387 wildfires in Galicia since 1968, the year in which forest fire statistics started, until December 2012 (MAPAMA, 2017). These fires burned an area of 1,794,578 ha, equivalent to 61% of the geographical area of the region (Boubeta et al., 2016a). Wildfires mainly affect rural municipalities in the south of the region with low population densities and regressive demographic dynamics due to low birth rates and an aged population (Balsa-Barreiro and Hermosilla, 2013; Fuentes-Santos et al., 2013). These municipalities have also been unaffected by recent foreign immigration patterns, which, combined with last century's rural flight has led to strong declines in population. Additionally, economic structures are based on primary sectors (González et al., 2007). We use the described methodology for modelling the number of fires by forest areas ($D = 63$) in Galicia during the summer 2007. That summer may be taken as representing the wildfire problems of recent years. The objective is to estimate the number of fires by forest areas using the plug-in estimator and to provide their bootstrap MSEs. This tool allows to construct fire risk maps with their error measures. We also extend the statistical methodology to the prediction of fire counts and the estimation of the corresponding uncertainties under new related scenarios.

Figure 2.9.4 shows the number of fires by forest areas of Galicia during the summer of 2007. Galicia is divided into 63 forest areas, which are the basic territorial structure of the fight against wildfires. At the same time, these areas are grouped into 19 districts. In that summer there were a high amount of fires, 15 areas had more than 19 fires, 15 areas had between 14 and 19 fires, 16 areas had between 8 and 13 fires and 17 areas had less than 7 fires. Most fires were concentrated in the coastal and south-east regions of Galicia.

We select the model covariates by taking into account an exploratory analysis, the Akaike information criterion (AIC) under Model 0, correlation studies and some expert judgments (see Chapter 1 for further details about the considered auxiliary information). Table 2.9.3 shows the parameter estimates of the selected best Poisson mixed model, in terms of simplicity and performance. They are estimated by using the ML-Laplace algorithm included in the lme4 package of R. All regression parameters, except *woods*, have positive slopes and therefore an increase in these covariates causes an increase in the number of fires. Thus, the number of fires is greater in those forest areas with higher values of the variables *dwr, pop, par, wet* and *lu*.

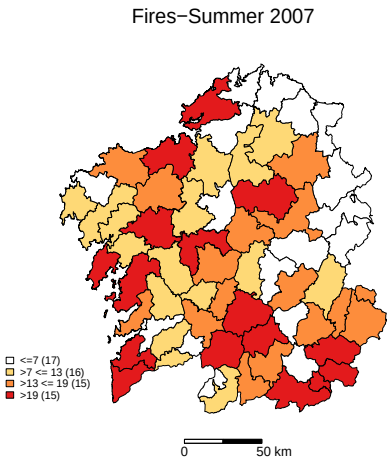

Figure 2.9.4: Fires in Galicia during summer 2007.

The coefficient of *woods* has a negative sign and hence this covariate protects against the increase of the response variable. Most forested areas experience fewer fires due to the greater economic interest in forestry. It is in scrublands, according to Bajocco and Ricotta (2008); Díaz-Delgado et al. (2004); González et al. (2006); González and Pukkala (2007); Koutsias et al. (2009); Montané et al. (2009); Moreira et al. (2009); Mouillot et al. (2005); Nunes et al. (2005) and Sebastian-Lopez et al. (2008), where greater intensity and number of fires is. The positive coefficient of *wet* is related to the location and the surface of reservoirs, built in the mid to late 20th century to produce electricity. They are mainly found in the headwaters in the mountains of the interior of the region, where an extensive livestock causes a significant number of fires for obtaining pasture, both in early spring and late summer.

Each forest area d ($d = 1, \ldots, 63$) has a random intercept with estimated normal distribution $N(0, \hat{\phi}^2)$, where $\hat{\phi} = 0.329$. The 95% percentile bootstrap confidence interval is $(0.162, 0.383)$. We approximate the sampling distribution of $\hat{\phi}$ by the resampling distribution of $\hat{\phi}^*$ using the $B = 1000$ bootstrap resamples obtained by applying the algorithm described in Section 2.7.3. This is, $\mathbb{P}(\hat{\phi} \leq a) \approx \mathbb{P}^*(\hat{\phi}^* \leq a)$. Therefore, if $t^*_{(\alpha)}$ is the α-quantile of $\hat{\phi}^*$, then $CI^* = (t^*_{(\alpha/2)}, t^*_{(1-\alpha/2)})$. Shao and Tu (1995) give mathematical details about the construction of the employed bootstrap confidence interval. We use the Algorithm 5 to test the null hypothesis $H_0 : \phi^2 = 0$. The obtained p-value is 0. Then, we conclude that ϕ is different from 0 for any level of significance α and, in consequence, we propose a Poisson GLMM with the random effects v_d's instead of the corresponding Poisson GLM without the random effects.

Table 2.9.3: ML-Laplace estimates of regression parameters under Model 1.

Variable	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	2.510	0.057	43.941	< 0.001
dwr	0.322	0.071	4.503	< 0.001
pop	0.449	0.064	7.004	< 0.001
par	0.250	0.059	4.247	< 0.001
scrub	0.274	0.078	3.521	< 0.001
wet	0.244	0.057	4.305	< 0.001
woods	-0.148	0.063	-2.363	0.018
lu	0.324	0.066	4.896	< 0.001

For the set of auxiliary variables appearing in Table 2.9.3, Figure 2.9.5 plots the Pearson residuals for the synthetic $\hat{\mu}_d^{syn}(\hat{\boldsymbol{\theta}})$ and plug-in $\hat{\mu}_d^{P}(\hat{\boldsymbol{\theta}})$ estimators under Model 0 (left) and Model 1 (right). We observe that the residuals of the Poisson mixed model (right) are closer to zero than the ones of the Poisson model with only fixed effects (left), so that we again prefer the model with random effects. The obtained p-values for Shapiro-Wilk test are 0.054 (GLM) and 0.860 (GLMM), and thus the normality assumption is not rejected in both cases.

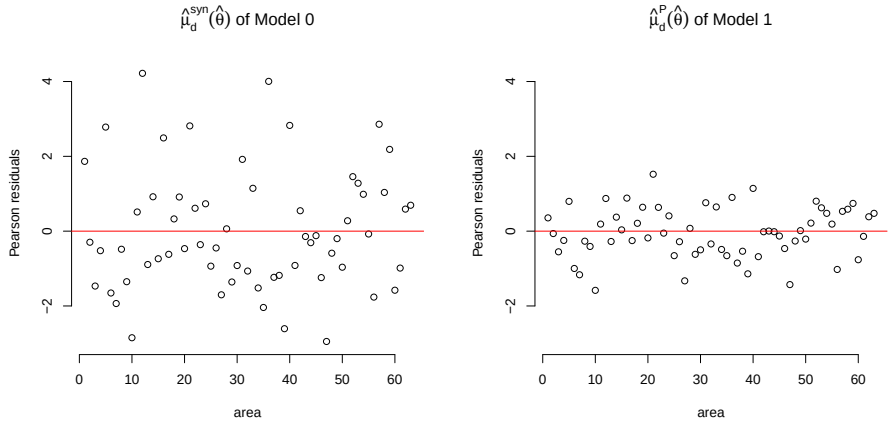

Figure 2.9.5: Pearson residuals under Model 0 (left) and Model 1 (right).

Figure 2.9.6 plots the observed number of fires versus the predicted number of fires under the models without (left) and with (right) random effects. We observe that Poisson mixed model has a greater prediction strength. This is why we confirm our selection of the introduced Poisson GLMM.

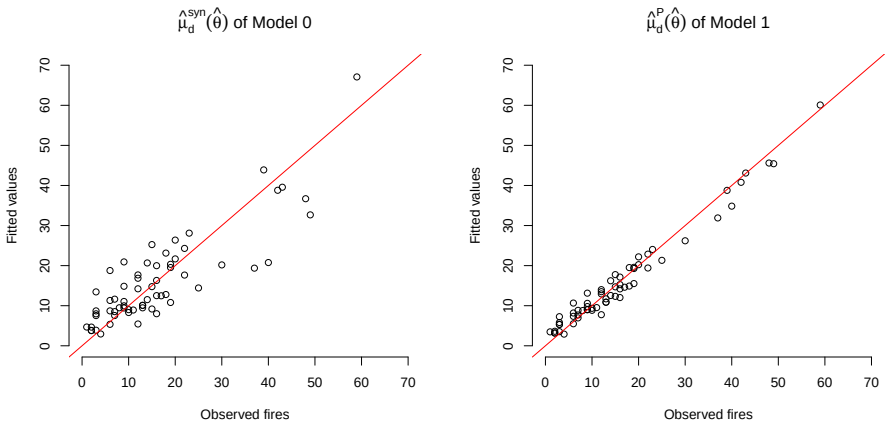

Figure 2.9.6: Observed versus predicted number of fires under Model 0 and Model 1.

Figure 2.9.7 plots the histogram (left) and the normal qqplot (right) of the predicted random effects, $\hat{v}_d$, $d = 1, \ldots, D$. Based on the two plots and on the p-value of the Shapiro-Wilk test (0.266), we can assume that the normality hypothesis on the model random effects is not severely violated.

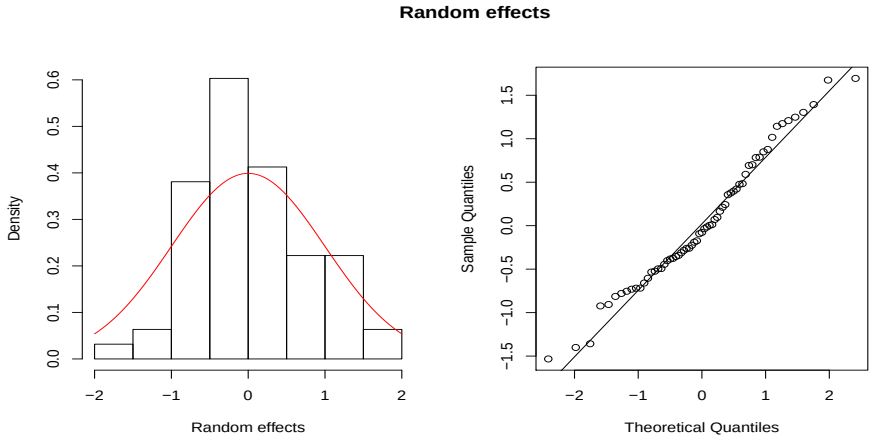

Figure 2.9.7: Histogram (left) and normal qqplot (right) of predicted random effects.

It is interesting to note that the auxiliary variable *par* has a regression parameter with positive sign. This fact is related to the small size of the rural parcels in Galicia. Formerly these parcels were cultivated and were the livelihood of families living from agriculture and livestock. Currently many of these parcels are no longer worked, as young people preferred to migrate to cities in search of better opportunities. Many parcels are then abandoned and overgrown allow fast expansion of summer fires. Therefore, the authorities would be interested in promoting the creation of cooperatives or land consolidation to get more out

of the land and to reduce the number of fires. Below we illustrate the reduction that would occur in the number of fires if we reduce by 5% the number of parcels and we let the remaining variables as they were (scenario $0.95 \times par$).

Figure 2.9.8 (left) shows the predicted number of fires per forest areas of Galicia under the scenario $0.95 \times par$. Reducing by 5% the number of cadastral parcels, the model predicts 15 areas with $y_d > 19$, 14 areas with $13 < y_d \leq 19$, 23 areas with $7 < y_d \leq 13$ and 11 areas with $y_d \leq 7$, where y_d is the number of fires in the forest area d. Therefore, if the number of cadastral parcels were reduced by 5% then an important reduction of the number of fires might occur. The Poisson mixed model predicts a reduction from 1001 to 970 fires in summers with similar environmental conditions than the one of 2007. This is because today one of the main problems of the Galician forest sector is the large number of plots (5 plots by ha) (Marey-Pérez et al., 2006; Rodríguez-Vicente and Marey-Pérez, 2009b) and the large number of forest owners (average of 4 ha for owner) (Marey-Pérez and Rodríguez-Vicente, 2009; Rodríguez-Vicente and Marey-Pérez, 2009a), that is at the origin of a percentage of fires by unprofitability (Rodríguez-Vicente and Marey-Pérez, 2010; Barreal et al., 2011), conflict of ownership and parcel boundaries (Gómez-Vázquez et al., 2009; Bruña García and Marey-Pérez, 2014). Reducing the number of plots contributes to increasing the profitability (Rodríguez et al., 2013) and decreasing much of those fires specially in the most conflictive areas.

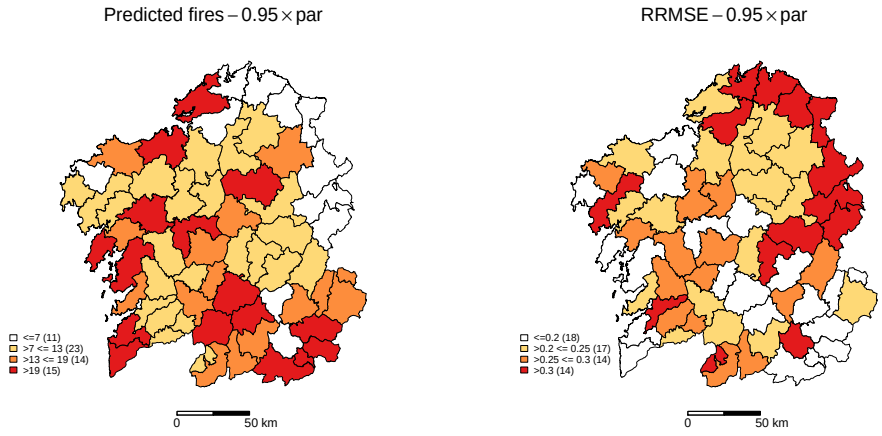

Figure 2.9.8: Predicted number of fires and relative root-MSEs under the scenario $0.95 \times par$.

Figure 2.9.8 (right) plots the estimated relative root-MSEs (RRMSE), which are obtained as the ratio of the square root of the MSE estimators $mse^*(\hat{\mu}_d^P)$ and the model-based predictors $\hat{\mu}_d^P$. In this framework, we take the MSE estimator based on a parametric bootstrap (2.7.10)

with $B = 1000$. The average RRMSE across all the forest areas is 24.66% and its behaviour is more satisfactory in the western region.

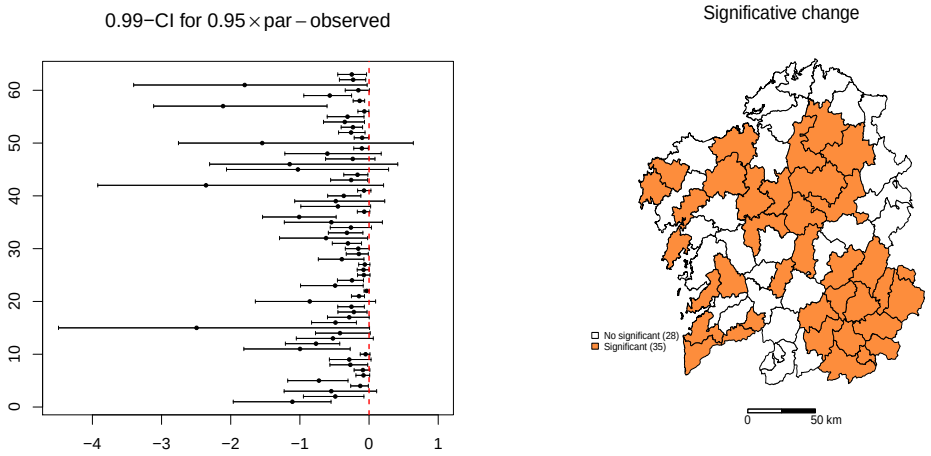

Figure 2.9.9: Confidence intervals for the variation of the number of fires if we reduce by 5% the number of parcels (left) and its significance (right), with 99% confidence.

Figure 2.9.9 presents the basic bootstrap confidence intervals for the estimates of the difference between the number of fires in the scenarios $0.95 \times par$ ($\tilde{\mu}_d$) and observed (μ_d, summer 2007), with a 99% confidence. We denote by $\tilde{\mu}_d$ the plug-in estimator of μ_d, $\hat{\mu}_d^P$, under the new scenario to simplify the notation. Now the construction of the bootstrap confidence intervals is a little different. It follows the classic idea of Davison and Hinkley (2007) but we consider the random effect in the bootstrap world to estimate μ_d. So, the $(1-\alpha)\%$ confidence limits are $\hat{L} = (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t_{(1-\alpha/2)}$ and $\hat{U} = (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t_{(\alpha/2)}$, being $\widehat{\tilde{\mu}}_d$ the predictor in the new scenario. We approximate the $t_{(\alpha)}$ quantiles by bootstrap following the steps of the previous algorithm in Section 2.7.3:

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi})$.
2. Repeat B times ($b = 1, \ldots, B$)
   i) Do $v_d^{*(b)} \sim N(0,1)$, $\mu_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$, $y_d^{*(b)} \sim \text{Poiss}(\mu_d^{*(b)})$, $d = 1, \ldots, D$.
   ii) From $\{\boldsymbol{x}_d, y_d^{*(b)}\}$, calculate $\hat{\boldsymbol{\theta}}^{*(b)}$, $\hat{v}_d^{*(b)}$, $\hat{\mu}_d^{*(b)} = \hat{\mu}_d^*(\boldsymbol{x}_d, \hat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_d^{*(b)})$, $d = 1, \ldots, D$.
   iii) For the new scenario $\{\tilde{\boldsymbol{x}}_d, y_d^{*(b)}\}$, calculate $\tilde{\mu}_d^{*(b)} = \exp\{\tilde{\boldsymbol{x}}_d\hat{\boldsymbol{\beta}} + \hat{\phi} v_d^{*(b)}\}$ and $\widehat{\tilde{\mu}}_d^{*(b)} = \hat{\mu}_d^*(\tilde{\boldsymbol{x}}_d, \hat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_d^{*(b)})$, $d = 1, \ldots, D$. Note that, the model is fitted from the bootstrap initial sample and then we study the effect of changing the values of the auxiliary variable $\boldsymbol{x}_d$ for $\tilde{\boldsymbol{x}}_d$, as in the real world.

3. Calculate the quantiles $t^*_{(\alpha/2)}$ and $t^*_{(1-\alpha/2)}$ as the $(\alpha/2)$-quantile and $(1-\alpha/2)$-quantile of $\left\{(\widehat{\tilde{\mu}}_d^{*(b)} - \hat{\mu}_d^{*(b)}) - (\tilde{\mu}_d^{*(b)} - \mu_d^{*(b)})\right\}_{b=1}^{B}$, respectively.

4. Finally, the bootstrap confidence interval is

$$\left((\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t^*_{(1-\alpha/2)},\ (\widehat{\tilde{\mu}}_d - \hat{\mu}_d) - t^*_{(\alpha/2)}\right).$$

We have a significant reduction in the number of fires in the forest regions where the confidence interval does not cut the dashed line at the origin. For example, in area 61 (located on the south-west coast), a significant decrease in the number of fires is obtained. In this case we can achieve a reduction of up to 3 fires with a confidence level of 99%. It also corresponds to one of the biggest cities in the community, so a reduction in the number of fires there would be positive for its socio-economic impact and risk of casualties. Figure 2.9.9 (right) shows that the above difference is not significant in 28 forest areas and significant in 35. The greatest changes are observed in the interior and southeast zones of Galicia.

Another important auxiliary variable for the occurrence of forest fires is *scrub area*. If we reduce this covariate in the same percentage as in the previous case (5%), the Poisson mixed model predict a reduction of 17 fires. Although this alternative involves a smaller reduction in the number of fires than in the previous case, its economic impact is also lower.

Remark 2.2. This application gives rise to a work published in Journal of Environmental Management (Boubeta et al., 2015). In this case, we infer about the individual hypotheses by using individual confidence intervals. Alternatively, we could construct confidence regions or use multiple testing procedures to investigate the differences between the two scenarios in all the areas simultaneously. Following this idea, we could use Bonferroni procedures where the level of significance is divided by the number of performed tests. However, there are more flexible and powerful methods. Benjamini and Hochberg (1995) and Benjamini and Yekutieli (2001) give different approaches to how the errors in multiple testing should be treated; for example, by controlling the expected proportion of erroneous rejections among all rejections.

However, what would be the most interesting procedure in this particular case? We could consider two approaches: a global focus with interest in the general population, which leads to the prevention of fires in the Galicia Community. In this case we have to study the effect of fire jointly in all forest areas. We can use the function *p.adjust* from the R package stats, and adapt the bootstrap algorithm to the test of difference of means. Or a second

focus, in which each local government (municipality, town hall, ...) makes its decisions independently, according to their needs and the characteristics of the area. Following the idea of SAE, the interest is each particular area, therefore we study the mean difference in each area independently according to as shown in this application.

2.10 Concluding remarks

Poisson regression models are quite simple but flexible enough for modelling count variables. This chapter analyses the number of people under the poverty line in Spanish provinces and the number of forest fires by areas using an area-level Poisson mixed model. In this framework, it has carried out a comparative study between the MM, the LA and the PQL fitting algorithms. PQL and LA perform better for the fixed effect coefficients but MM and LA capture the variance component more precisely.

The chapter considers that the EBP (using MM) is a good alternative for describing the target variable due to the good performance shown in the design-based simulation experiment, where a comparison against two plug-in estimators (using MM and PQL) is given. 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 0.07 seconds in our computer while for the plug-in using MM and EBP (taking $L = 2500$) was 0.15.

For the EBP, the chapter calculates the MSE and introduces 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. It analyses the behaviour 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, it suggests the bootstrap procedure, easy to implement and with similar results.

Two applications to real data are presented for applying the developed methodology. The first one deals with a socioeconomic topic (poverty in Spain by provinces and sex) and the second one treats with an environmental problem (number of forest fires in Galicia by areas). The application to poverty data from the SLCS in 2008, published in the journal *Test* (Boubeta et al., 2016b), proposes the EBPs for estimating poverty rates since their results are more satisfactory than the ones obtained by the direct estimators. It concludes that the south and center-west provinces of Spain have highest levels of poverty. As performance

measure it takes the RRMSE estimated by parametric bootstrap. The RRMSE estimates are lower than 20.25% in all provinces.

Finally, Poisson mixed models are also employed in Boubeta et al. (2015) to predict forest fires in Galicia by areas. Two contributions are presented. The first one is the methodology for predicting fire counts and for estimating the corresponding efficiency measures approximated by bootstrap. The second one is the construction of bootstrap confidence intervals for the variation of number of fires between observed data and data coming from new scenarios. This tool allows us to study whether the small changes in some auxiliary variables may produce significant reduction of fires in some domains of interest. The introduced statistical methodology gives a useful decision-making tool for policy makers. It can be also extended to target variables with distributions from the exponential family and, therefore, it can be used under Binomial and Poisson distributions, among others. The use of these Poisson mixed models is the first stage towards the integration of a higher number of variables and data about a longer wildfire period. These improvements can increase the predictive capacity which explains the presence of arson wildfires in a conflictive area, answering thus the demands of policy makers and technicians.

Chapter 3

The area-level Poisson mixed model with SAR(1) domain effects

Contents

3.1 Introduction

When auxiliary variables related to the target count variable are available at the area level, the area-level Poisson Model 1 links all the domains to enhance the estimation at a particular area, that is, it borrows strength from other areas. Model 1 has random effects taking into account the between-domain variability that is not explained by the auxiliary variables. This model assumes that the domain random effects are independent. However, in socioeconomic, environmental and epidemiological applications, estimates for areas that are spatially close may be more alike than estimates for areas that are further apart. In fact,

Cressie (1993) shows that not employing spatial models may lead to inefficient inferences when the auxiliary variables does not explain the spatial correlation of the study variable.

In small area estimation, modelling the spatial correlation among data from different areas allows to borrow even more strength from the areas. This recommendation was applied to the basic Fay-Herriot model by Singh et al. (2005). Later, several authors have proposed new spatial area-level linear models. Petrucci and Salvati (2006); Pratesi and Salvati (2008); Molina et al. (2009); Marhuenda et al. (2013) and Chandra et al. (2015) consider extensions of the Fay-Herriot model by assuming that area effects follow a simultaneously autoregressive process of order 1 or SAR(1).

In the Bayesian framework, Moura and Migon (2002) and You and Zhou (2011) consider spatial stationary mixed models and Sugasawa et al. (2015) study an empirical Bayesian estimation method with spatially non-stationary hyperparameters for area-level discrete and continuous data having a natural exponential family distribution.

Concerning nonparametric and robust methods, Opsomer et al. (2008) give a small area estimation procedure using penalized spline regression with applications to spatially correlated data. Ugarte et al. (2006) and Ugarte et al. (2010) study the geographical distribution of mortality risk, that is an important area of research in disease mapping, using small areas techniques and penalized splines. Chandra et al. (2012) introduce a geographical weighted empirical best linear unbiased predictor for a small area average and give an estimator of its conditional mean squared error. Baldermann et al. (2016) describe robust small area estimation methods under spatial non-stationarity linear mixed models. Chandra et al. (2017) develop a geographically weighted regression extension of the logistic-normal and the Poisson-normal generalized linear mixed models allowing for spatial nonstationarity.

The above cited papers introduce small area estimation procedures that borrows strength from spatial correlation. However, none of them deals with empirical best predictors under spatial GLMMs. This is why Chapter 3 studies an area-level Poisson mixed model with SAR(1) correlated random effects. A fitting algorithm based on the method of moments is proposed in Section 3.3. Empirical best predictors of domain proportions and counts are given in Section 3.4 and a parametric bootstrap method for estimating its mean squared error is introduced in Section 3.5. The introduced methodology is empirically investigated in Section 3.6 by means of simulation experiments. Section 3.7 gives two relevant applications to real data. The first one has socio-economic interest and the second one is in the field of environmental sciences. Finally, Section 3.8 collects the main conclusions of this chapter.

3.2 The model

This section extends the area-level Poisson mixed model (Model 1), proposed in Chapter 2 (Boubeta et al., 2016b), to the context of spatial correlation. Specifically, we assume a SAR(1) process. Let us consider a population partitioned into D domains and let us denote each particular domain by d, $d = 1, \ldots, D$. Let $\boldsymbol{v} = (v_1, \ldots, v_D)'$ be a vector of spatially correlated random effects following a SAR(1) process with unknown autoregression parameter ρ and known proximity matrix $\boldsymbol{W}$. This means that the vector of random effects $\boldsymbol{v}$ fulfills the linear combination

$$\boldsymbol{v} = \rho \boldsymbol{W} \boldsymbol{v} + \boldsymbol{u}, \tag{3.2.1}$$

where $\boldsymbol{u} \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D)$, $\boldsymbol{0}$ is the $D \times 1$ zero vector and $\boldsymbol{I}_D$ denotes the $D \times D$ identity matrix. Assuming that $(\boldsymbol{I}_D - \rho \boldsymbol{W})$ is non-singular, the equation (3.2.1) can be expressed as

$$\boldsymbol{v} = (\boldsymbol{I}_D - \rho \boldsymbol{W})^{-1} \boldsymbol{u}. \tag{3.2.2}$$

For the proximity matrix $\boldsymbol{W}$, we assume that it is row stochastic. Then, the autoregression parameter ρ is a correlation and is called spatial autocorrelation parameter. Some of the most used proximity matrices are based on: (i) common borders, (ii) distances and (iii) k-nearest neighbours. In all cases, the proximity matrix $\boldsymbol{W}$ is obtained from an original proximity matrix $\boldsymbol{W}^0$ with diagonal elements equal to zero and remaining entries depending on the employed option. In option (i), the non diagonal elements of $\boldsymbol{W}^0$ are equal to 1 when the two domains corresponding to the row and the column indices are regarded as neighbours and zero otherwise. In Option (ii), the nondiagonal elements of the proximity matrix $\boldsymbol{W}^0$ are defined by applying a monotonously decreasing function to the domain distances; for example, by using the inverse function. Finally, the non diagonal elements of $\boldsymbol{W}^0$ in option (iii) are 1 if they correspond to the k-nearest neighbours of a given domain and zero otherwise. For each option, the row standardization is carried out by dividing each entry of $\boldsymbol{W}^0$ by the sum of the elements in its row. Consequently, $\boldsymbol{W}$ is row stochastic. Equation (3.2.2) implies that $\boldsymbol{v} \sim N_D(\boldsymbol{0}, \boldsymbol{\Gamma}(\rho))$, where

$$\boldsymbol{\Gamma}(\rho) = \left(\gamma_{d_1 d_2}(\rho)\right)_{d_1, d_2 = 1, \ldots, D} = \boldsymbol{C}^{-1}(\rho) \tag{3.2.3}$$

and $\boldsymbol{C}(\rho) = (\boldsymbol{I}_D - \rho \boldsymbol{W})'(\boldsymbol{I}_D - \rho \boldsymbol{W})$. Therefore, the density function of the random effects is

$$f_v(\boldsymbol{v}) = (2\pi)^{-D/2} |\boldsymbol{\Gamma}(\rho)|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{v}' \boldsymbol{\Gamma}^{-1}(\rho) \boldsymbol{v} \right\}.$$

Further, we have $v_d \sim N\big(0, \gamma_{dd}(\rho)\big)$ and $v_{d_2}|v_{d_1} \sim N\big(\mu_{d_2|d_1}, \sigma^2_{d_2|d_1}\big)$, where

$$\mu_{d_2|d_1} = \frac{\gamma_{d_1 d_2}(\rho)}{\gamma_{d_1 d_1}(\rho)}\, v_{d_1},$$
$$\sigma^2_{d_2|d_1} = \gamma_{d_2 d_2}(\rho) - \frac{\gamma^2_{d_1 d_2}(\rho)}{\gamma_{d_1 d_1}(\rho)}.$$

The first partial derivatives of $\boldsymbol{C}$ and $\boldsymbol{\Gamma}$ with respect to ρ are

$$\dot{\boldsymbol{C}}(\rho) = \frac{\partial \boldsymbol{C}}{\partial \rho} = -\boldsymbol{W} - \boldsymbol{W}' + 2\rho \boldsymbol{W}'\boldsymbol{W},$$
$$\dot{\boldsymbol{\Gamma}}(\rho) = \frac{\partial \boldsymbol{\Gamma}}{\partial \rho} = -\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \rho}\boldsymbol{C}^{-1} = \big(\dot{\gamma}_{d_1 d_2}(\rho)\big)_{d_1, d_2 = 1, \ldots, D}.$$

The vector of response variables $\boldsymbol{y} = (y_1, \ldots, y_D)'$ follows an area-level Poisson mixed model with a SAR(1) vector of domain random effects $\boldsymbol{v}$ if the conditioned distribution of y_d, given v_d, is

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

where μ_d denotes the mean of the Poisson distribution. As in the previous chapter, we consider that the Poisson parameter μ_d can be expressed as $\nu_d p_d$, where ν_d and p_d are size and probability parameters respectively. We assume that the natural parameter, $\log \mu_d$, can be expressed in terms of a set of p covariates $\boldsymbol{x}_d = (x_{d1}, \ldots, x_{dp})$ by a regression model, i.e.

$$\text{Model S1:} \quad \log \mu_d = \log \nu_d + \log p_d = \log \nu_d + \boldsymbol{x}_d \boldsymbol{\beta} + \phi v_d, \quad d = 1, \ldots, D,$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ and ϕ are the regression and variance parameters. We denote the vector of all model parameters by $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \rho)$.

Further, we assume that the y_d's are independent conditionally on $\boldsymbol{v}$. It holds that

$$\mathbb{P}(y_d|\boldsymbol{v}) = \mathbb{P}(y_d|v_d) = \frac{1}{y_d!}\exp\{-\nu_d p_d\}\nu_d^{y_d} p_d^{y_d},$$

where $p_d = \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}$. The probability function of the response variable $\boldsymbol{y}$ is

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

where

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

3.3 The MM algorithm

The method of moments is derived, under Model S1, to estimate the vector of parameters $\boldsymbol{\theta}$. A natural set of equations for applying the MM algorithm is

$$\begin{aligned}0 = f_k(\boldsymbol{\theta}) &= M_k(\boldsymbol{\theta})-\hat{M}_k = \frac{1}{D}\sum_{d=1}^{D}\mathbb{E}_{\boldsymbol{\theta}}[y_d]x_{dk}-\frac{1}{D}\sum_{d=1}^{D}y_d x_{dk},\quad k=1,\ldots,p,\\ 0 = f_{p+1}(\boldsymbol{\theta}) &= M_{p+1}(\boldsymbol{\theta})-\hat{M}_{p+1} = \frac{1}{D}\sum_{d=1}^{D}\mathbb{E}_{\boldsymbol{\theta}}[y_d^2]-\frac{1}{D}\sum_{d=1}^{D}y_d^2,\\ 0 = f_{p+2}(\boldsymbol{\theta}) &= M_{p+2}(\boldsymbol{\theta})-\hat{M}_{p+2} = \frac{1}{D(D-1)}\sum_{d_1\neq d_2}\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1}y_{d_2}]-\frac{1}{D(D-1)}\sum_{d_1\neq d_2}y_{d_1}y_{d_2},\end{aligned}\tag{3.3.1}$$

where $d_1, d_2 = 1,\ldots,D$. The MM estimator of $\boldsymbol{\theta}$ is the solution of the system of nonlinear equations (3.3.1). The updating formula of the Newton-Raphson algorithm follows the equation (2.3.3), where now $\theta_1=\beta_1,\ldots,\theta_p=\beta_p$, $\theta_{p+1}=\phi$, $\theta_{p+2}=\rho$ and

$$\boldsymbol{\theta} = \underset{1\le k\le p+2}{\text{col}}(\theta_k),\quad \boldsymbol{f}(\boldsymbol{\theta}) = \underset{1\le k\le p+2}{\text{col}}(f_k(\boldsymbol{\theta})),\quad \boldsymbol{H}(\boldsymbol{\theta}) = \left(\frac{\partial f_k(\boldsymbol{\theta})}{\partial\theta_r}\right)_{k,r=1,\ldots,p+2}.\tag{3.3.2}$$

Let us now calculate the expectations appearing in $\boldsymbol{f}(\boldsymbol{\theta})$ and $\boldsymbol{H}(\boldsymbol{\theta})$. First, we recall that the moment generation function of $Y\sim N(\mu,\sigma^2)$ is

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

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

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

Therefore, the first k MM equations are

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

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

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

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

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

Therefore

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

We have

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

and as a consequence

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

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

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

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

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

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

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

where

$$\begin{aligned}
\mathbb{E}_v\big[p_{d_1}p_{d_2}\big] &= \int_{-\infty}^{\infty}\Big[\int_{-\infty}^{\infty}\exp\big\{\boldsymbol{x}_{d_2}\boldsymbol{\beta} + \phi v_{d_2}\big\}f_v(v_{d_2}|v_{d_1})\,dv_{d_2}\Big]\exp\big\{\boldsymbol{x}_{d_1}\boldsymbol{\beta} + \phi v_{d_1}\big\}f_v(v_{d_1})dv_{d_1}\\
&= \int_{-\infty}^{\infty}\exp\{\boldsymbol{x}_{d_2}\boldsymbol{\beta}\}\Psi(\phi;\mu_{d_2|d_1},\sigma^2_{d_2|d_1})\exp\big\{\boldsymbol{x}_{d_1}\boldsymbol{\beta} + \phi v_{d_1}\big\}\,f_v(v_{d_1})dv_{d_1}\\
&= \exp\Big\{(\boldsymbol{x}_{d_1}+\boldsymbol{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2}\Big(\gamma_{d_2d_2} - \frac{\gamma^2_{d_1d_2}}{\gamma_{d_1d_1}}\Big)\phi^2\Big\}\psi\Big(\big(1+\frac{\gamma_{d_1d_2}}{\gamma_{d_1d_1}}\big)\phi;0,\gamma_{d_1d_1}\Big)\\
&= \exp\Big\{(\boldsymbol{x}_{d_1}+\boldsymbol{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2}\Big(\gamma_{d_2d_2} - \frac{\gamma^2_{d_1d_2}}{\gamma_{d_1d_1}}\Big)\phi^2 + \frac{1}{2}\Big(1+\frac{\gamma_{d_1d_2}}{\gamma_{d_1d_1}}\Big)^2\phi^2\gamma_{d_1d_1}\Big\}.
\end{aligned}$$

Therefore, the $(p+2)$-th MM 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

$$\begin{aligned}
\varphi_{d_1,d_2}(\boldsymbol{\theta}) &= \exp\Big\{(\boldsymbol{x}_{d_1}+\boldsymbol{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2}\Big(\gamma_{d_2d_2} - \frac{\gamma^2_{d_1d_2}}{\gamma_{d_1d_1}}\Big)\phi^2 + \frac{1}{2}\frac{(\gamma_{d_1d_1}+\gamma_{d_1d_2})^2}{\gamma_{d_1d_1}}\phi^2\Big\}\\
&= \exp\Big\{(\boldsymbol{x}_{d_1}+\boldsymbol{x}_{d_2})\boldsymbol{\beta} + \frac{1}{2}\,\phi^2(\gamma_{d_1d_1}+\gamma_{d_2d_2}+2\gamma_{d_1d_2})\Big\}.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1}y_{d_2}]}{\partial \beta_k} &= \nu_{d_1}\nu_{d_2}\varphi_{d_1,d_2}(\boldsymbol{\theta})(x_{d_1k}+x_{d_2k}),\\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1}y_{d_2}]}{\partial \phi} &= \nu_{d_1}\nu_{d_2}\varphi_{d_1,d_2}(\boldsymbol{\theta})\phi(\gamma_{d_1d_1}+\gamma_{d_2d_2}+2\gamma_{d_1d_2}),\\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d_1}y_{d_2}]}{\partial \rho} &= \frac{1}{2}\nu_{d_1}\nu_{d_2}\varphi_{d_1,d_2}(\boldsymbol{\theta})\phi^2(\dot{\gamma}_{d_1d_1}+\dot{\gamma}_{d_2d_2}+2\dot{\gamma}_{d_1d_2}).
\end{aligned}$$

The elements of the Jacobian matrix are

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

Under Model S1, the MM algorithm keeps the steps of Algorithm 1 (see Section 2.3), replacing $\boldsymbol{\theta}$, $\boldsymbol{H}$ and $\boldsymbol{f}$ for those given in (3.3.2).

As algorithm seeds for $\boldsymbol{\beta}$ and ϕ, we take the MM estimator under the model with no spatial correlation ($\rho = 0$). Concerning the parameter ρ, we propose to use the Moran's I measure of spatial autocorrelation, i.e.

$$I = \frac{D}{\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1d_2}}\,\frac{\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1d_2}(\tilde{v}_{d_1}-\tilde{v})(\tilde{v}_{d_2}-\tilde{v})}{\sum_{d=1}^{D}(\tilde{v}_d-\bar{v})^2}, \tag{3.3.3}$$

where $\tilde{v}_d$, $d=1,\dots,D$, are the predicted random effects under Model 1 ($\rho=0$), $\tilde{v} = \frac{1}{D}\sum_{d=1}^{D}\tilde{v}_d$ and $w_{d_1d_2}$, $d_1,d_2=1,\dots,D$, are the elements of the proximity matrix $\boldsymbol{W}$.

The asymptotic variance of the MM estimator under the area-level Poisson mixed model with SAR(1) domain effects can be approximated by a similar bootstrap algorithm to that described in Section 2.3.1.

3.4 The predictors

This section gives the EBP and a plug-in predictor of p_d under the area-level Poisson mixed model with SAR(1) spatially correlated random effects. As the EBP involves high-

dimensional integrals, we propose a computationally less demanding approximation.

3.4.1 The empirical best predictor

In this section we obtain EBPs for the area-level Poisson mixed model with SAR(1) domain effects. Specifically, we focus on the calculation of the EBP of p_d, given its relationship with μ_d and assuming that the size parameter ν_d is known. Therefore, for the rest, we will refer to p_d as target parameter. Let δ_{ij} be the Kronecker's delta; i.e. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise.

The EBP of p_d is obtained from the BP by replacing the vector of all model parameters $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$. The BP of p_d is the unbiased estimator that minimizes the mean squared error and is given by

$$\hat{p}_d(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[p_d|\boldsymbol{y}] = \frac{\int_{\mathbb{R}^D} \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} \prod_{i=1}^D \mathbb{P}(y_i|v_i) f_v(\boldsymbol{v})\, d\boldsymbol{v}}{\int_{\mathbb{R}^D} \prod_{i=1}^D \mathbb{P}(y_i|v_i) f_v(\boldsymbol{v})\, d\boldsymbol{v}} = \frac{N_d(\boldsymbol{y}, \boldsymbol{\theta})}{D_d(\boldsymbol{y}, \boldsymbol{\theta})}, \tag{3.4.1}$$

where

$$N_d(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}^D} \exp\left\{\sum_{i=1}^D \left[(y_i + \delta_{id})(\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i) - \nu_i \exp\left\{\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i\right\}\right]\right\} f_v(\boldsymbol{v})\, d\boldsymbol{v},$$

$$D_d(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}^D} \exp\left\{\sum_{i=1}^D \left[y_i(\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i) - \nu_i \exp\left\{\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i\right\}\right]\right\} f_v(\boldsymbol{v})\, d\boldsymbol{v}.$$

Remark 3.1. The component $N_d(\boldsymbol{y}, \boldsymbol{\theta})$ can be expressed in terms of $D_d(\boldsymbol{y}, \boldsymbol{\theta})$ as $N_d(\boldsymbol{y}, \boldsymbol{\theta}) = D_d(\boldsymbol{y} + \boldsymbol{e}_d, \boldsymbol{\theta})$, where $\boldsymbol{e}_d = (\delta_{1d}, \ldots, \delta_{Dd})'$.

The EBP of p_d is $\hat{p}_d(\hat{\boldsymbol{\theta}})$ and it can be approximated by using an antithetic Monte Carlo algorithm. The steps are:

1. Generate $\boldsymbol{v}^{(\ell)}$ i.i.d. $N_D(\boldsymbol{0}, \boldsymbol{\Gamma}(\hat{\rho}))$ and calculate their antithetics $\boldsymbol{v}^{(L+\ell)} = -\boldsymbol{v}^{(\ell)}$, $\ell = 1, \ldots, L$.

2. Calculate $\hat{p}_d(\hat{\boldsymbol{\theta}}) = \hat{N}_d / \hat{D}_d$, where

$$\hat{N}_d = \frac{1}{2L} \sum_{\ell=1}^{2L} \exp\left\{\sum_{i=1}^D \left[(y_i + \delta_{id})(\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \hat{\phi} v_i^{(\ell)}) - \nu_i \exp\left\{\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \hat{\phi} v_i^{(\ell)}\right\}\right]\right\},$$

$$\hat{D}_d = \frac{1}{2L} \sum_{\ell=1}^{2L} \exp\left\{\sum_{i=1}^D \left[y_i(\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \hat{\phi} v_i^{(\ell)}) - \nu_i \exp\left\{\boldsymbol{x}_i\hat{\boldsymbol{\beta}} + \hat{\phi} v_i^{(\ell)}\right\}\right]\right\}.$$

As the above BP involves high-dimensional integrals, we propose a less computationally demanding approach. For that, let us divide the response variable $\boldsymbol{y}$ and the vector of random effects $\boldsymbol{v}$ into two parts $(y_d, \boldsymbol{y}_{d-})$ and $(v_d, \boldsymbol{v}_{d-})$, where $\boldsymbol{y}_{d-} = \underset{1\leq i\leq D,\, i\neq d}{\text{col}}(y_i)$ and $\boldsymbol{v}_{d-} = \underset{1\leq i\leq D,\, i\neq d}{\text{col}}(v_i)$. The conditional distribution of $\boldsymbol{y}$, given $\boldsymbol{v}$, is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) = \prod_{i=1}^{D} \mathbb{P}(y_i|v_i) = \mathbb{P}(y_d|v_d) \prod_{i=1, i\neq d}^{D} \mathbb{P}(y_i|v_i) = \mathbb{P}(y_d|v_d)\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}). \tag{3.4.2}$$

Using (3.4.2), the component $D_d(\boldsymbol{y}, \boldsymbol{\theta})$ of (3.4.1) can be rewritten as

$$D_d(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} \Big[\int_{\mathbb{R}^{D-1}} \mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}) f(\boldsymbol{v}_{d-}|v_d)\, d\boldsymbol{v}_{d-} \Big] \mathbb{P}(y_d|v_d) f(v_d)\, dv_d,$$

and since $\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}) f(\boldsymbol{v}_{d-}|v_d) = \mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}, v_d) f(\boldsymbol{v}_{d-}|v_d)$, the inner integral is

$$\int_{\mathbb{R}^{D-1}} \mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}, v_d) f(\boldsymbol{v}_{d-}|v_d)\, d\boldsymbol{v}_{d-} = \mathbb{P}(\boldsymbol{y}_{d-}|v_d).$$

Therefore

$$D_d(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} \mathbb{P}(\boldsymbol{y}_{d-}|v_d)\mathbb{P}(y_d|v_d) f(v_d)\, dv_d.$$

Taking into account the Remark 3.1 and reasoning analogously with the component $N_d(\boldsymbol{y}, \boldsymbol{\theta})$ of (3.4.1), we have that

$$N_d(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\} \mathbb{P}(\boldsymbol{y}_{d-}|v_d)\mathbb{P}(y_d|v_d) f(v_d)\, dv_d.$$

Under the assumption that $\mathbb{P}(\boldsymbol{y}_{d-}|v_d) \approx \mathbb{P}(\boldsymbol{y}_{d-})$, $d = 1, \ldots, D$, the BP of p_d, $\hat{p}_d(\boldsymbol{\theta})$, can be approximated by

$$\hat{p}_d^a(\boldsymbol{\theta}) = N_d^a(\boldsymbol{y}, \boldsymbol{\theta}) / D_d^a(\boldsymbol{y}, \boldsymbol{\theta}), \tag{3.4.3}$$

where

$$N_d^a(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} \exp\left\{(y_d + 1)(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\right\}\right\} f(v_d)\, dv_d,$$
$$D_d^a(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} \exp\left\{y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\left\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\right\}\right\} f(v_d)\, dv_d.$$

The approximated BP, $\hat{p}_d^a(\boldsymbol{\theta})$, involves integrals with a complex analytical solution. We propose to approximate them by using an antithetic Monte Carlo algorithm. In practice, as one does not know the true vector of model parameters $\boldsymbol{\theta}$, the corresponding EBP, $\hat{p}_d^a(\hat{\boldsymbol{\theta}})$, is required. It can be approximated as follows.

1. Generate $\boldsymbol{v}^{(\ell)}$ i.i.d. $N_D(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho}))$ and calculate $\boldsymbol{v}^{(L+\ell)} = -\boldsymbol{v}^{(\ell)}$, $\ell = 1, \ldots, L$.

2. Approximate the EBP of p_d as $\hat{p}_d^a(\hat{\boldsymbol{\theta}}) = \hat{N}_d^a / \hat{D}_d^a$ $(d = 1, \ldots, D)$, where

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

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

This approximation is similar to the Algorithm 4. The difference lies in the process for generating the random effects. In Algorithm 4, random effects are $N(0, 1)$, while in the above approximation they are SAR(1)-correlated. As an immediate consequence, by applying equation (3.4.1) we have that the EBP of the Poisson parameter $\mu_d = \nu_d p_d$ is $\hat{\mu}_d(\hat{\boldsymbol{\theta}}) = \nu_d \hat{p}_d(\hat{\boldsymbol{\theta}})$. From the equation (3.4.3), its approximated version is $\hat{\mu}_d^a(\hat{\boldsymbol{\theta}}) = \nu_d \hat{p}_d^a(\hat{\boldsymbol{\theta}})$.

3.4.2 A plug-in predictor

Another estimator of p_d, commonly used in this context, is the plug-in estimator. It is obtained by replacing, in the theoretical expression of p_d, the unknown parameters by their estimates, i.e.

$$\hat{p}_d^P = \exp\left\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}\hat{v}_d\right\}.$$

It is important to note that the MM algorithm only provides estimates for the fixed effects $\boldsymbol{\beta}$, the variance ϕ and the autocorrelation parameter ρ and that for obtaining $\hat{p}_d^P$ it is necessary to predict the vector of random effects $\boldsymbol{v} = (v_1, \ldots, v_D)$. Therefore, we propose to use its EBP. As above, this predictor is obtained from the corresponding BP. The BP of v_d is

$$\hat{v}_d(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_d|\boldsymbol{y}] = \frac{\int_{\mathbb{R}^D} v_d \prod_{i=1}^D \mathbb{P}(y_i|v_i) f_v(\boldsymbol{v})\, d\boldsymbol{v}}{\int_{\mathbb{R}^D} \prod_{i=1}^D \mathbb{P}(y_i|v_i) f_v(\boldsymbol{v})\, d\boldsymbol{v}} = \frac{N_{v,d}(\boldsymbol{y}, \boldsymbol{\theta})}{D_d(\boldsymbol{y}, \boldsymbol{\theta})},$$

where

$$N_{v,d}(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}^D} v_d \exp\left\{\sum_{i=1}^D y_i(\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i) - \nu_i \exp\{\boldsymbol{x}_i\boldsymbol{\beta} + \phi v_i\}\right\} f_v(\boldsymbol{v})\, d\boldsymbol{v}.$$

If the assumption $\mathbb{P}(\boldsymbol{y}_{d-}|v_d) \approx \mathbb{P}(\boldsymbol{y}_{d-})$, holds for $d = 1, \ldots, D$, similar mathematical developments as those presented in Section 3.4.1 yield to an approximation to $\hat{v}_d(\boldsymbol{\theta})$ equivalent

to that obtained for the target parameter p_d, i.e. $\hat{v}_d^a(\boldsymbol{\theta}) = N_{v,d}^a(\boldsymbol{\theta})/D_d^a(\boldsymbol{\theta})$, where

$$N_{v,d}^a(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}} v_d \exp\{y_d(\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d) - \nu_d \exp\{\boldsymbol{x}_d\boldsymbol{\beta} + \phi v_d\}\} f(v_d)\, dv_d.$$

The EBP of v_d is $\hat{v}_d = \hat{v}_d(\hat{\boldsymbol{\theta}})$ and can be approximated by $\hat{v}_d^a = \hat{v}_d^a(\hat{\boldsymbol{\theta}})$. As above, we propose approximating the analytical integrals by using an antithetic Monte Carlo algorithm. The steps for $\hat{v}_d^a$ are:

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi}, \hat{\rho})$.
2. For $\ell = 1, \ldots, L$, generate $\boldsymbol{v}^{(\ell)}$ i.i.d. $N_D(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho}))$ and calculate their antithetics $\boldsymbol{v}^{(L+\ell)} = -\boldsymbol{v}^{(\ell)}$.
3. Calculate $\hat{v}_d^a(\hat{\boldsymbol{\theta}}) = \hat{N}_{v,d}^a/\hat{D}_d^a$ $(d = 1, \ldots, D)$, where

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

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

The difference between this approximation and the one described in Section 2.5.2 (for approximating the EBP of v_d), comes from the process of generating the random effects. Before, the random effects followed a $N(0, 1)$ distribution while now they are SAR(1)-correlated.

3.5 MSE estimation

The MSE is a measure of the accuracy of the proposed EBP of p_d under Model S1. As the analytical approach is computationally demanding, this section introduces an estimation procedure of the MSE of $\hat{p}_d$ by using a parametric bootstrap procedure based on the one given in González-Manteiga et al. (2007). The steps are:

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi}, \hat{\rho})$.
2. For each domain d, $d = 1, \ldots, D$, repeat B times $(b = 1, \ldots, B)$:
   i) Generate the bootstrap random effects $\boldsymbol{v}^{*(b)} = (v_1^{*(b)}, \ldots, v_D^{*(b)})' \sim N_D(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho}))$, where $\boldsymbol{\Gamma}(\hat{\rho})$ is the plug-in version of the covariance matrix (3.2.3).

ii) Calculate the theoretical bootstrap parameter $p_d^{*(b)} = \exp\{\boldsymbol{x}_d\hat{\boldsymbol{\beta}} + \hat{\phi}v_d^{*(b)}\}$.

iii) Generate the response variables $y_d^{*(b)} \sim \text{Poiss}(\nu_d p_d^{*(b)})$.

iv) For each bootstrap resample, calculate the estimator $\hat{\boldsymbol{\theta}}^{*(b)}$ and the EBP $\hat{p}_d^{*(b)} = \hat{p}_d^{*(b)}(\hat{\boldsymbol{\theta}}^{*(b)})$.

3. Output:

$$mse^*(\hat{p}_d) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{p}_d^{*(b)} - p_d^{*(b)}\right)^2.$$

3.6 Simulation experiments

This section presents two simulation experiments for studying the behaviour of the MM fitting algorithm (simulation 1) and for investigating the performance of the proposed EBPs and plug-in predictors (simulation 2) under Model S1 with different values of ρ. The simulations are based on the application to real data of poverty in Galicia during 2013 (see Section 1.4.1 for more details). We use the same explanatory variables as those used in the real case, i.e. proportions of unemployed (*lab2*) and of people with university level completed (*edu3*) by counties. First simulation experiment analyses the behaviour of the MM fitting algorithm. As the MM estimate of the autocorrelation parameter ρ produces a high bias, an alternative approach using the value of the Moran's test is proposed. The second simulation experiment studies the performance of the proposed estimators (BP, EBP and plug-in) based on Model S1. In addition, we also consider the corresponding estimators under Model 1 to analyse the loss of efficiency when the spatial autocorrelation is not taken into account.

We generate independent response variables $y_d|v_d \sim \text{Poiss}(\nu_d p_d)$, where ν_d and $p_d = \exp\{\beta_0 + lab_2\beta_1 + edu_3\beta_2 + \phi v_d\}$ are the sample size and target parameter, $d = 1, \ldots, D$. The model parameters β_0, β_1, β_2, ϕ and ρ are taken from the real data case. The domain random effects, v_d, $d = 1, \ldots, D$, are generated according to a SAR(1) process with autocorrelation parameter ρ and proximity matrix $\boldsymbol{W}$ (see Section 3.2). The number of total domains is $D = 49$. It corresponds to the counties in Galicia. Actually, in Galicia there are 53 counties but in four of them no data are available. Both simulation experiments keep the number of domains D fixed and analyse the behaviour of the proposed estimators for different values of ρ. The reason to keep D fixed is due to the rigidity of the simulation study, since it is based on the real case and specially by the construction of the proximity matrix $\boldsymbol{W}$. As the estimation of the autocorrelation parameter in the application to real

data was $\hat{\rho} = 0.324$, we take $\rho = 0.1, 0.3, 0.5$. The considered Monte Carlo iterations are $K = 1000$ for the first simulation experiment and $K = 500$ for the second one.

Tables 3.6.1 and 3.6.2 present the bias and RMSE of the MM estimator for the five model parameters $\theta \in \boldsymbol{\theta} = \{\beta_0, \beta_1, \beta_2, \phi, \rho\}$. We consider two options to estimate the vector of model parameters $\boldsymbol{\theta}$. In the first option (Option 1), $\hat{\boldsymbol{\theta}}$ is given as a solution of the system of nonlinear equations (3.3.1), while in the second option (Option 2), ρ is estimated by using the Moran's I measure (3.3.3) over the Pearson residuals of Model 0 and the remaining model parameters are given as a solution of the system formed by the first $p + 1$ MM equations in (3.3.3).

Table 3.6.1: Bias of the MM fitting algorithm.

Option	ρ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}$	$\hat{\rho}$
1	0.1	-0.0167	0.1087	-0.0088	-0.0179	-0.2557
	0.3	-0.0037	0.0251	-0.0150	-0.0234	-0.4065
	0.5	-0.0122	0.0987	0.0024	-0.0153	-0.5698
2	0.1	0.0038	0.0098	-0.0401	-0.0036	-0.1088
	0.3	0.0029	0.0247	-0.0545	-0.0009	-0.2724
	0.5	0.0048	0.0282	-0.0546	0.0071	-0.4243

Table 3.6.2: RMSE of the MM fitting algorithm.

Option	ρ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\phi}$	$\hat{\rho}$
1	0.1	0.1585	1.1984	0.5409	0.0919	0.3276
	0.3	0.1623	1.3005	0.5497	0.0951	0.4536
	0.5	0.1690	1.3542	0.5388	0.0938	0.6040
2	0.1	0.1507	1.2224	0.5152	0.1020	0.1462
	0.3	0.1540	1.2229	0.5254	0.1002	0.2912
	0.5	0.1705	1.3888	0.5527	0.1038	0.4384

Table 3.6.1 suggests that the bias is lower for the fixed effects and the variance parameter ϕ, while for the autocorrelation parameter ρ is relatively high. In addition, ρ is underestimated in all cases. This behaviour usually occurs in this type of models (Cressie, 1993; Crujeiras et al., 2010; Fernández-Casal and Francisco-Fernández, 2014). It can be also boosted by the quite small number of total domains D. Regarding the comparison between Option 1 and Option 2, in general a clear reduction in bias is achieved when one uses Option 2. In addition, no pattern is observed in the behaviour of the bias for the different values of ρ. On the other hand, Table 3.6.2 reveals that the RMSE results are higher for the fixed

effects β_1 and β_2. No significant differences are observed in the comparison between the two options, but there is a general increase in the RMSE as the value of ρ increases. For the vector of fixed effects $\hat{\boldsymbol{\beta}}$ and the variance parameter $\hat{\phi}$, the variance is, by far, the most important term of MSE since bias is much smaller than the RMSE. On the other hand, for $\hat{\rho}$ the opposite situation occurs, i.e. the bias is the main part of the MSE. Then, a bias correction by bootstrap may be useful.

The target of the second simulation experiment is to investigate the behaviour of the plug-in predictor, the BP and EBP of p_d for different values of ρ. We compare these predictors based on Model S1 with the corresponding ones under Model 1 given in Chapter 2. We are interested in investigating the loss of efficiency when the spatial correlation is not taken into account. For the spatial autocorrelation parameter, we take $\rho = 0.1, 0.3, 0.5$. We run Option 2 in the MM fitting algorithm.

Tables 3.6.3 and 3.6.4 present the average across domains of the biases and the RMSEs (both $\times 10^2$) of the BP, EBP and two plug-in predictors for both area-level Poisson mixed models: Model 1 and Model S1. The two plug-in predictors are obtained by calculating the vector of random effects $\boldsymbol{v}$ by its BP (P_{BP}) and EBP (P_{EBP}). For the model with SAR(1) domain effects, Model S1, the two alternatives introduced in Section 3.5 are considered, i.e. the BP (3.4.1) and its approximation (BP^a) given in (3.4.3). The corresponding empirical versions (EBP and EBP^a) are also taken into account. For the plug-in predictors under Model S1, the random effects are calculated by using only the approximation $\hat{v}_d^a$ seen in Section 3.4.2. We run this simulation experiment with a sample size of $L = 5000$ to approximate the BP's and EBP's.

Table 3.6.3: Average across domains of the biases ($\times 10^2$) of the BP, EBP and plug-in of p_d based on the area-level Poisson mixed models with independent (Model 1) and on the SAR(1)-correlated (Model S1) random effects.

	Model 1				Model S1					
ρ	BP	P_{BP}	EBP	P_{EBP}	BP	BP^a	P^a_{BP}	EBP	EBP^a	P^a_{EBP}
0.1	0.073	0.098	0.192	0.219	0.081	0.074	0.097	0.237	0.234	0.228
0.3	0.078	0.115	0.239	0.271	0.079	0.080	0.114	0.238	0.247	0.250
0.5	0.081	0.115	0.255	0.279	0.082	0.082	0.109	0.295	0.303	0.292

Table 3.6.3 suggests a strong increase in bias when we consider empirical predictors. Re-

garding the comparison between Model 1 and Model S1, there are no substantial differences between the two models, although in general the average bias is smaller under Model 1. However, if the average across domains is ignored, the behaviour of the domain biases, B_d's, shows that predictors are not centered in many domains (see Figure 3.6.1 for more details). The approximated BP (BP^a) and EBP (EBP^a) based on Model S1 behave similar to the original predictors (without approximation) for low correlations. On the other hand, they are slightly less competitive as ρ increases, but its computational time is much lower. The plug-in predictors based on Model 1 have greater bias than the corresponding BP and EBP. When the variance components are known, the difference between the predictors BP, BP^a and P^a_{BP} based on Model S1 have the theoretical expected good behaviour with low biases. However, when we substitute the variance components by the their MM estimators, the corresponding predictors EBP, EBP^a and P^a_{EBP} based on Model S1 have much larger biases.

Table 3.6.4 presents the average across domains of the RMSEs ($\times 10^2$) of the BP, EBP and plug-in for both area-level Poisson mixed models: Model 1 and Model S1. It reveals an increase in the RMSE as the parameter ρ increases and also when one uses empirical versions instead of theoretical models. Regarding the comparisons between predictors, the plug-in predictor has, in general, a slight lower RMSE. On the other hand, approximated versions of the BP (BP^a) and EBP (EBP^a) clearly reduce the RMSE. Then, in terms of RMSE, it is preferable to use the approximate predictors under Model S1. For any estimator, the variance is the most important term of the MSE since bias is much smaller than the RMSE.

Table 3.6.4: Average across domains of the RMSEs ($\times 10^2$) of the BP, EBP and plug-in of p_d based on the area-level Poisson mixed models with independent (Model 1) and on the SAR(1)-correlated (Model S1) random effects.

	Model 1				Model S1					
ρ	BP	P_{BP}	EBP	P_{EBP}	BP	BP^a	P^a_{BP}	EBP	EBP^a	P^a_{EBP}
0.1	1.806	1.794	2.238	2.232	2.140	1.805	1.793	2.370	2.281	2.276
0.3	1.876	1.845	2.297	2.271	2.153	1.873	1.842	2.415	2.336	2.310
0.5	2.020	2.001	2.468	2.455	2.336	1.997	1.977	2.650	2.510	2.497

Figure 3.6.1 shows the boxplots of the domain biases, B_d's, (first column) and the domain root mean squares errors, RE_d's, (second column) for the predictors and the values of ρ appearing in Tables 3.6.3 and 3.6.4. In each graph, the first four boxplots refer to the

predictors based on Model 1 and the remaining six to the predictors based on Model S1. The BP's and EBP's ($\hat{p}_d(\boldsymbol{\theta})$ and $\hat{p}_d(\hat{\boldsymbol{\theta}})$) are represented in blue, their approximations based on the Model S1 ($\hat{p}^a_d(\boldsymbol{\theta})$ and $\hat{p}^a_d(\hat{\boldsymbol{\theta}})$) are plotted in green and the plug-in predictors ($\hat{p}^P_d(\boldsymbol{\theta})$ and $\hat{p}^P_d(\hat{\boldsymbol{\theta}})$) are colored in orange. They show an increase of the variability in both B_d's and RE_d's when one uses the empirical predictors. The bias of the predictors based on Model 1 has less variability, but these predictors are clearly biased (except the BP). This fact was not shown in Table 3.6.3. The predictors based on Model S1 are unbiased except the P^a_{BP} plug-in predictor. The behaviour of the RE_d's for the predictors based on Model 1 is similar to the one based on Model S1, although for $\rho = 0.3$, the RE_d's of the plug-in estimator are slightly lower. For predictors based on Model S1, the RE_d's of the approximated BP $\hat{p}^a_d(\boldsymbol{\theta})$ and EBP $\hat{p}^a_d(\hat{\boldsymbol{\theta}})$ are similar to those of P_{BP} ($\hat{p}^P_d(\boldsymbol{\theta})$) and P_{EBP} ($\hat{p}^P_d(\hat{\boldsymbol{\theta}})$) respectively, while the RE_d's of the BP $\hat{p}_d(\boldsymbol{\theta})$ and EBP $\hat{p}_d(\hat{\boldsymbol{\theta}})$ are generally higher.

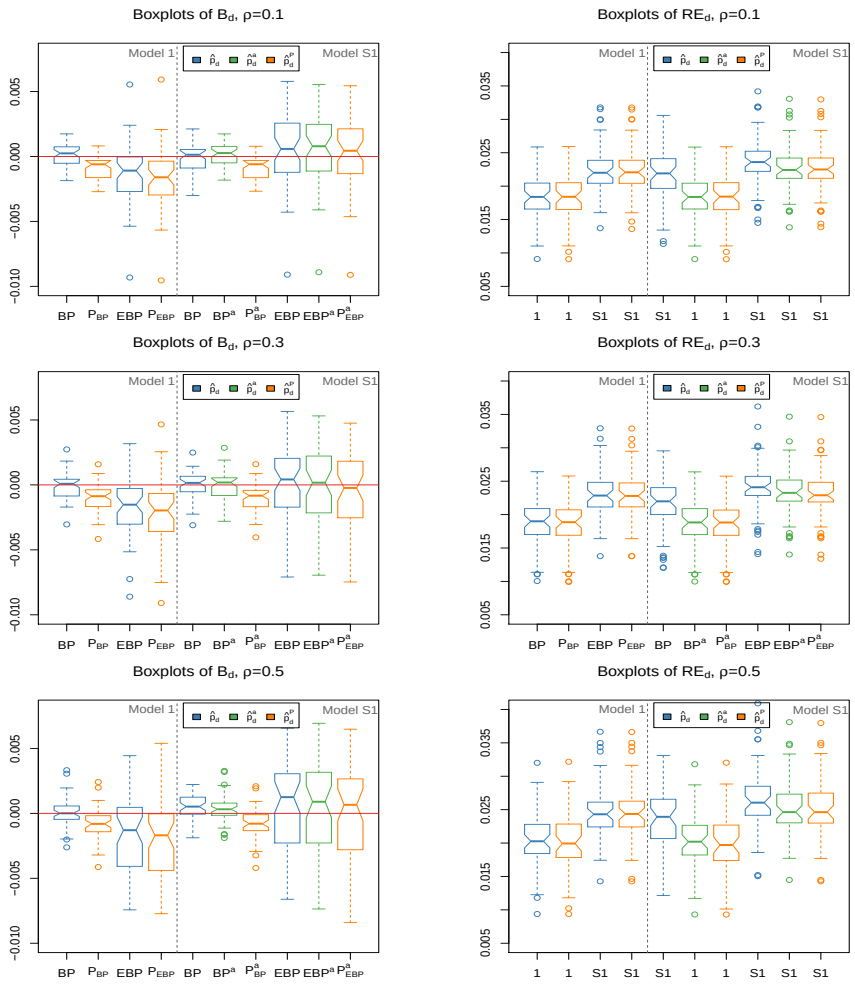

Figure 3.6.1: Boxplots of B_d's (first column) and RE_d's (second column) for the predictors of p_d and values of ρ shown in Tables 3.6.3 and 3.6.4.

From Tables 3.6.3 and 3.6.4 and Figure 3.6.1, we conclude that the approximated estimator $\hat{p}_d^a$ under Model S1 shows a competitive performance when there is an underlying spatial correlation structure in the data, since it is unbiased and its RE_d's behave similarly to those of $\hat{p}_d^P$.

The system of MM nonlinear equations (3.3.1) under Model S1 is solved by using the *nleqslv* package of R. We have also used the *mvtnorm* package to generate samples following a SAR(1) process and the package *spdep* to construct the proximity matrix $\boldsymbol{W}$ and to test the null hypothesis of no spatial autocorrelation. For $\rho = 0.3$, the average runtime of the MM fitting algorithm under Option 2 was 0.51 seconds. The computational burden of the EBP approximation, $\hat{p}_d^a$, under Model S1 is similar to that of the EBP under Model 1. The average runtimes were 0.41 and 0.39 seconds respectively. On the other hand, the EBP $\hat{p}_d$ under Model S1 has a high computational burden compared to its competitors. Its average runtime was 48.73 seconds.

3.7 Applications to real data

3.7.1 Poverty data

This section applies the developed methodology to the estimation of poverty proportions, p_d, in Galicia. The data are taken from the 2013 SLCS (see Section 1.4.1 for more details). The Galician counties are the study domains. In Galicia there are 53 counties, but in four of them there are no available data. Therefore, the number of considered domains is $D = 49$. The performance of Model S1 depends on the choice of the proximity matrix $\boldsymbol{W}$. Three different choices are tested: common border, based on distances and based on k-nearest neighbours. In the first option (common border), two domains are neighbours if they have a common delimitation. The last two options consider the Euclidean distance between the centroids of the counties. The second option sets up a proximity measure by taking the inverse of the distance between domains. The last option applies k-nearest neighbours with $k = 2$ and 3. After analysing the different possibilities, the first option is selected because it is the one giving the best results.

Figure 3.7.1 shows the proximity map that determines the proximity matrix $\boldsymbol{W}^0$, i.e. it provides for each domain, which are its neighbours. See Section 3.2 for more details on the construction of the proximity matrix $\boldsymbol{W}^0$ and $\boldsymbol{W}$.

The poverty proportion is only estimated for women since for men there is no evidence

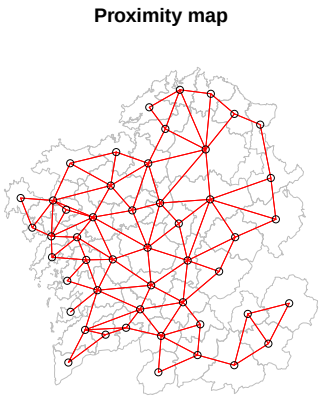

Figure 3.7.1: Proximity map for each domain d ($d = 1, \ldots, D$).

of a spatial correlation structure. The obtained p-value of the Moran's I test for men, applied on the residuals of the fixed effects model (Model 0), is 0.271 and therefore the null hypothesis of no spatial autocorrelation is accepted. As a consequence, y_d counts the number of women under the poverty line in the domain d. The p-value of Moran's I test for women is lower than 0.001 and then, it is recommended to use Model S1 to fit the data. Accordingly, we assume that the response variable y_d, $d = 1, \ldots, D$, can be explained by an area-level Poisson mixed model with SAR(1) domain effects and some auxiliary variables.

Table 3.7.1 presents the significant estimates (p-value < 0.05) of the fixed effect coefficients under Model S1 and their standard errors, z-values and p-values, using MM. The autocorrelation parameter is estimated by applying Moran's I measure (3.3.3) over the Pearson residuals of Model 0 and the remaining model parameters are given as a solution of the system formed by the first $p + 1$ MM equations in (3.3.3). Thus, Option 2 in the MM algorithm is employed.

Table 3.7.1: MM estimates of regression parameters under Model S1 using Option 2.

Variable	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	-1.8803	0.1515	-12.4086	< 0.001
*lab*2	2.9848	1.2097	2.4689	0.0136
*edu*3	-1.3809	0.5033	-2.7445	0.0061

Taking into account the signs of the estimates, the auxiliary variable *lab*2 (proportion of unemployed women), is directly related to the response variable while *edu*3 (proportion of women with university level of education), helps to decrease the women poverty rate. Each domain d, $d = 1, \ldots, D$, has a random intercept with distribution $N(0, \phi^2)$, where

$\hat{\phi} = 0.130$. The 95% percentile bootstrap confidence interval for the variance parameter is $(0.001, 0.331)$. The estimated autocorrelation parameter is $\hat{\rho} = 0.324$. To test the null hypothesis $H_0 : \phi^2 = 0$, Algorithm 5 (see Secion 2.9) is adapted to Model S1. The obtained p-value is 0.018. Then, taking $\alpha = 0.05$, the null hypothesis is rejected. To test $H_0 : \rho = 0$, the following bootstrap procedure is proposed.

Algorithm 6 A bootstrap test for $H_0 : \rho = 0$

1: Fit the Model S1 to data and calculate $\hat{\boldsymbol{\beta}}$, $\hat{\phi}$ and $\hat{\rho}$.

2: Fit the Model 1 to data and calculate $\hat{\boldsymbol{\beta}}^0$ and $\hat{\phi}^0$.

3: For $b = 1, \ldots, B$, do

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

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

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

4: Calculate the p-value

$$p = \frac{\#\left\{\left|\hat{\rho}^{*(b)}\right| > |\hat{\rho}|\right\}}{B}.$$

The obtained bootstrap p-value is 0.001. Taking $\alpha = 0.05$, the bootstrap test concludes that the autocorrelation parameter ρ is significantly different from 0, and therefore it recommends using the Model S1 to fit the data.

For comparing the performance of Model S1, a fixed-effects Poisson model (Model 0) is also fitted to data with the same auxiliary variables as those appearing in Table 3.7.1. Figure 3.7.2 plots the Pearson residuals of the synthetic estimator based on Model 0 (left), and of the EBP approximation based on Model S1 (right). Model 1 is not considered since the Moran I test suggests spatial correlation. In both cases, the distribution of the Pearson residuals is symmetrical around 0. In addition, the plots suggest a clear improvement when one uses an area-level Poisson mixed model that incorporates SAR(1) domain effects, since it is able to better capture the underlying spatial correlation structure. Therefore, the conclusion is again that Model S1 is more appropriated to fit the women poverty data in Galicia by counties in 2013.

Figure 3.7.3 (left) compares the behaviour of the EBPs based on Model S1 and the direct estimates, which are usually used in practice. Both estimators of the p_d's are sorted by

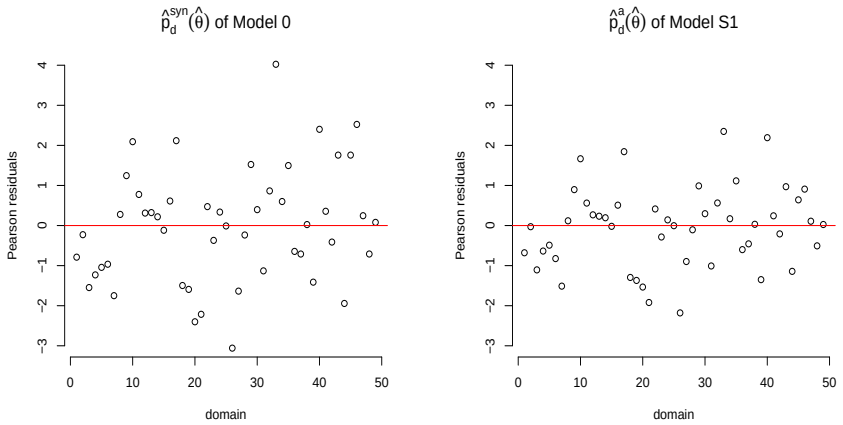

Figure 3.7.2: Pearson residuals of the synthetic estimator based on Model 0 (left) and of the EBP approximation based on Model S1 (right).

the sample sizes ν_d's. Direct estimator shows large amplitude oscillations, while the EBP presents a smoother behaviour. As the sample size increases, both estimates tend to overlap. Figure 3.7.3 (right) plots the relative root-MSEs of the EBPs based on Model S1 and the relative root-variances of the direct estimators. The direct estimates have high variability, specially for small sample sizes. As above, when ν_d increases, both accuracy measures follow the same pattern. The relative root-MSEs of the EBPs are estimated by using the bootstrap procedure of Section 3.5 with $B = 500$ replicates. The averages of the relative root-variances of the direct estimator and of the relative root-MSEs of the EBP are 0.2595 and 0.1323, respectively. According to these results, we conclude that the EBP performs better.

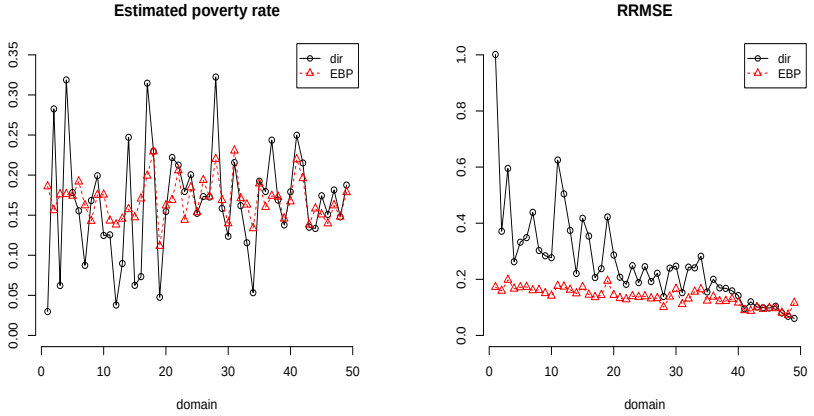

Figure 3.7.3: Direct estimates and EBPs of p_d (left) and relative root-MSEs (right) for women in 2013.

Figure 3.7.4 (left) maps the EBP approximation of p_d for women based on Model S1 in 2013. The regions where there is no data, are in white. Model S1 predicts one county with poverty proportion $p_d \leq 0.12$, 12 counties with $0.12 < p_d \leq 0.15$, 24 counties with $0.15 < p_d \leq 0.18$ and 12 counties with $p_d > 0.18$. Highest levels of poverty are found in the south and west of the community. On the other hand, the counties with the lowest estimated poverty rate are located in the north-east of the region. Figure 3.7.4 (right) maps the RRMSE estimates of the EBP of p_d by counties in 2013. We take $B = 500$ resamples. There are 8 counties with RRMSE $\leq$ 10%, 8 counties with 10% $<$ RRMSE $\leq$ 13%, 19 counties with 13% $<$ RRMSE $\leq$ 16% and 14 counties with RRMSE $>$ 16%. The highest values are found in the north-east of the region. Their minimum and maximum are 8.82% and 18.49%, respectively. As the highest RRMSE is lower than 20%, these estimates could be accepted for publication by statistical offices.

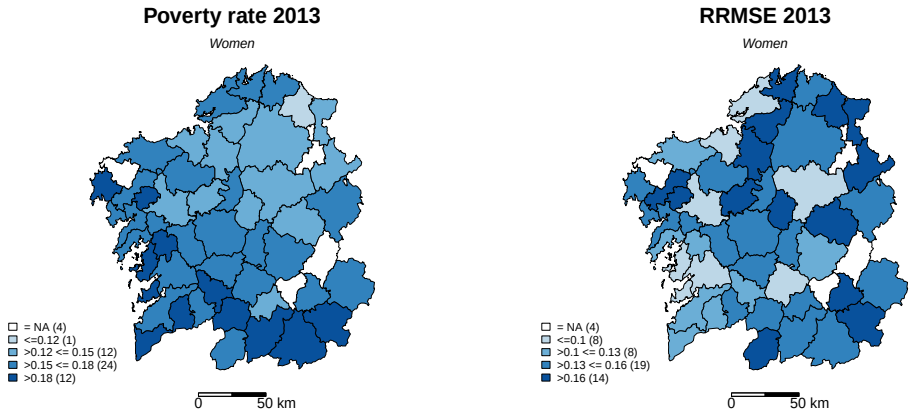

Figure 3.7.4: Poverty rate EBPs for women based on Model S1 (left) and RRMSEs (right) in Galicia during 2013.

3.7.2 Forest fires data

The modelling of the number of fires allows an improvement of the resources by the forest managers. Specially in Galicia, a region in the north-west of Spain, where wildfires produce devastating effects every year. The objective is to know the behaviour of the response variable *number of forest fires* y_d, $d = 1, \ldots, D$, by areas during the summer of 2008. The number of total forest areas in Galicia is $D = 63$ (see Section 1.4.2 for more details).

The Moran's I test is applied to the residuals of the model without random effects, Model 0. The obtained p-value is lower than 0.001. Therefore, the null hypothesis of no spatial

autocorrelation is rejected and the Poisson model with spatial correlation is selected. For the proximity matrix $\boldsymbol{W}$, the one proposed in Section 3.2 is used. Figure 3.7.5 shows the proximity map that determines the matrix $\boldsymbol{W}^0$ involved in the calculation of the final proximity matrix $\boldsymbol{W}$. That is, it provides the neighbours for each domain d.

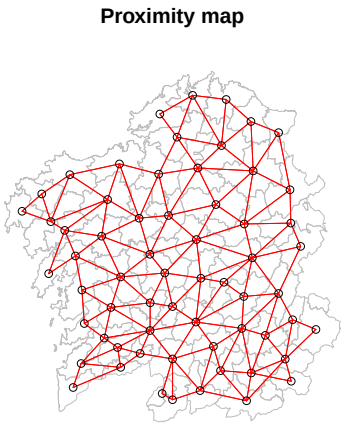

Figure 3.7.5: Proximity map for each forest area d ($d = 1, \ldots, D$).

Specifically, this section assumes that the response variable at domain d, y_d, can be explained by some auxiliary variables through an area-level Poisson mixed model with SAR(1) domain effects. The considered covariates are: population density (pop), cadastral holders ($cadHold$), percentage of forest area ($perForest$), and average measurements of meteorological stations in summer 2008 per areas, such as accumulated rain ($acumRain$), temperature ($averTemp$) and days without rain (dwr).

Table 3.7.2 compares the significant estimates (p-value < 0.05) of the fixed effects based on Model 0 (left) and on Model S1 (right). The maximum likelihood (ML) method is used to estimate the regression parameters in Model 0 and the MM algorithm in Model S1. The auxiliary variables are selected by using the Akaike information criterion (AIC) under Model 0 and the same covariates are taken for Model S1. In this way, both models are compared under the same auxiliary information. The two models suggest that $cadHold$ is directly related to the number of fires while $acumRain$ is protective, i.e. an increase in this covariate helps to reduce the number of fires if $cadHold$ is kept fixed.

Table 3.7.2: ML estimates of regression parameters under Model 0 (left) and MM estimates under Model S1 using Option 2 (right).

	Model 0				Model S1			
Variable	Est.	s.e.	z-value	$P(>\|z\|)$	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	2.455	0.038	64.415	< 0.001	2.391	0.068	34.942	< 0.001
acumRain	-0.317	0.037	-8.631	< 0.001	-0.317	0.077	-4.092	< 0.001
cadHold	0.379	0.027	14.168	< 0.001	0.378	0.058	6.552	< 0.001

The estimated variance and autocorrelation parameters are $\hat{\phi} = 0.351$ and $\hat{\rho} = 0.119$, respectively. The 95% percentile bootstrap confidence interval for the variance parameter, taking $B = 500$ bootstrap resamples, is $(0.174, 0.460)$. The obtained bootstrap p-values for testing $H_0 : \phi^2 = 0$ and $H_0 : \rho = 0$ are 0 in both cases. Then, according to Moran's I test and to the previous bootstrap confidence tests, this section selects an area-level Poisson mixed model with SAR(1) domain effects, Model S1, instead of a Poisson model with only fixed effects, Model 0. Figure 3.7.6 plots the Pearson residuals of the synthetic estimator under Model 0 (left) and of the plug-in estimator under Model S1 (right). A clear improvement is achieved when one uses a more complex model including SAR(1)-correlated spatial effects, since its Pearson residuals are closer to 0. Then, Model S1 is again preferred to fit the number of forest fires in Galicia during summer 2008.

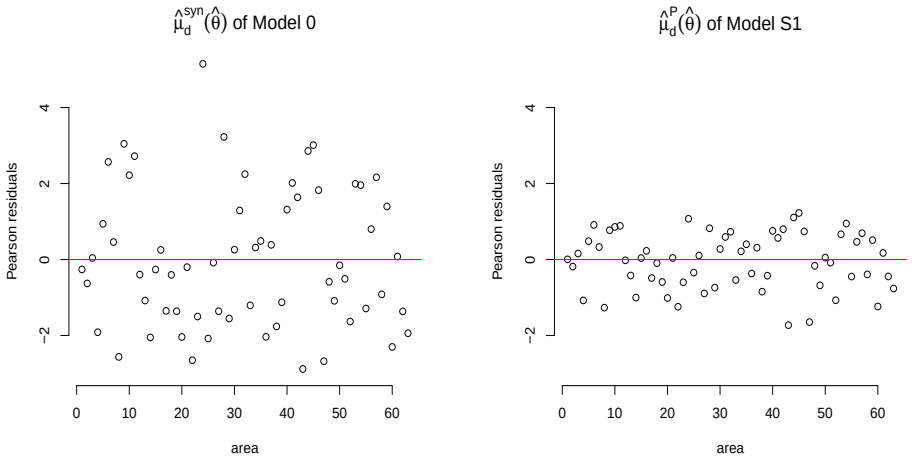

Figure 3.7.6: Pearson residuals of the synthetic estimator based on Model 0 (left) and of the plug-in predictor based on Model S1 (right).

Figure 3.7.7 (left) maps the plug-in estimates based on Model S1 by areas during the summer of 2008. The model predicts 3 areas with $\mu_d \leq 5$ forest fires, 30 with $5 < \mu_d \leq 10$, 15 with $10 < \mu_d \leq 15$ and 15 with $\mu_d > 15$. The regions with more predicted fires ($\mu_d > 15$), are located in coastal areas of the north and west of the community, and especially in southern interior areas. The average of forest fires predicted by Model S1 is 12.72. The RRMSEs of the plug-in predictor are mapped in Figure 3.7.7 (right). They are calculated as

$$RRMSEs = \frac{\sqrt{MSEs}}{\hat{\boldsymbol{y}}}, \tag{3.7.1}$$

where the root-MSEs are calculated by using the bootstrap procedure of Section 3.5 with $B = 500$ replicates and $\hat{\boldsymbol{y}}$ are the plug-in estimates of the response variable. The plug-in predictor has an error of $RRMSE < 20\%$ in 17 areas, $20\% < RRMSE \leq 25\%$ in 12 areas, $25\% < RRMSE \leq 30\%$ in 14 areas and $RRMSE > 30\%$ in 20 areas. The highest RRMSEs occur in areas with few estimated fires. Note that if $\hat{\boldsymbol{y}}$ is small, the RRMSE increases since $\hat{\boldsymbol{y}}$ is in the denominator of (3.7.1). The mean of the RRMSEs is 25.98%. The lowest errors are obtained in those areas with the highest number of fires, i.e. in coastal areas in the north and southern interior areas.

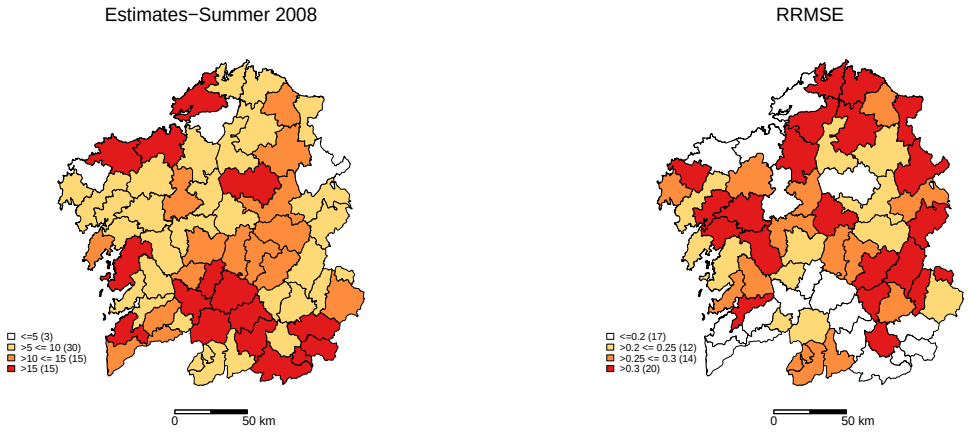

Figure 3.7.7: Estimated fires (left) and RRMSEs (right) in summer 2008 based on Model S1.

3.8 Concluding remarks

This chapter introduces an area-level Poisson mixed model with SAR(1) domain effects. It generalises the area-level Poisson mixed model proposed in Chapter 2 to the context of

spatial correlation. The MM algorithm is employed for estimating the model parameters. The empirical best predictor and a plug-in predictor of the target parameter p_d are proposed. As accuracy measure of the EBP, the MSE is considered and it is estimated by a parametric bootstrap approach.

The behaviour of the MM fitting algorithm is empirically investigated in Simulation 1. This simulation, based on the real case with poverty data, provides the bias and the MSE for different scenarios of ρ ($\rho = 0.1, 0.3, 0.5$). In addition, a second simulation experiment has been carried out to study a plug-in predictor, the EBP and its corresponding approximation. Specifically, the simulation experiment investigates the behaviour of the BP, EBP and two plug-in predictors of p_d based on the area-level Poisson mixed model of Chapter 2 and the area-level Poisson mixed model with SAR(1) domain effects. The target is to analyse the loss of efficiency when the spatial correlation is not taken into account. For the autocorrelation parameter, we take the same scenarios as those shown in Simulation 1. This simulation experiment also studies the behaviour of the EBP approximation given in Section 3.4.1. It shows that the EBP approximation is competitive, since it is unbiased and it has a lower MSE than the original EBP. It has a similar behaviour to the plug-in predictor, which shows a clear bias. On the other hand, the MSE of the plug-in and EBP based on Model 1 is similar to the plug-in and the EBP approximation based on Model S1, but the first ones show a sharp bias. Another advantage of the EBP approximation is that it reduces the computational burden significantly.

We use the EBP approximation for estimating women poverty proportions in Galician counties. The data are taken from the 2013 SLCS. As the residuals of the model with only fixed effects, Model 0, present spatial correlation, we recommend using Model S1 to fit the data. In addition, the proposed estimator is compared against the direct estimator. The EBP estimates of the women poverty rate are smoother. As the RRMSEs of the direct estimator are too high when the sample size ν_d is small, it is preferable to use the EBP approximation. The estimates based on Model S1 suggests that the highest levels of women poverty are found in the south and west of the region. The average percentage of women poverty is 16.89% and its average error is 13.23%.

The developed methodology is also applied to predicting the number of forest fires in Galicia by areas during the summer of 2008. The performance of the plug-in predictor is compared against the synthetic estimator based on Model 0. A clear improvement is achieved when one uses a more complicated model. Although different options are explored for the proximity matrix $\boldsymbol{W}$, finally we consider a proximity matrix based on neighboring

areas. According to the plug-in predictor based on Model S1, the highest number of forest fires are located in the south and in coastal areas of the north and west of the region. In addition, we provide a bootstrap approach as accuracy measure of the plug-in predictor. The average error is 25.98%, while for the areas with highest fires it is lower than 20%.

Chapter 4

The area-level Poisson mixed model with time effects

Contents

4.1 Introduction

We are entering an era in which a large amount of information is stored, particularly over time, to study the behaviour of variables of interest. With this idea the statistical institutes have historical information. On the other hand, estimation techniques in small areas are important when the level of disaggregation is very high and the direct estimators do not work as well as it is needed. This chapter combines these two needs and provides statistical methodology for obtaining estimates of population quantities at a sufficiently disaggregated level by time periods. We stand out some works in the literature as Rao and Yu (1994), which introduce an extension of a well-known model, due to Fay and Herriot (1979), for cross-sectional data. Estimators provided by Pfeffermann and Burck (1990);

Ghosh et al. (1996); Datta et al. (2002); Saei and Chambers (2003); You et al. (2001); Esteban et al. (2012a,b) and Marhuenda et al. (2013), among others, take advantage of the two levels in linear mixed models for producing small areas estimates with good properties. In a more general context, Ugarte et al. (2009) combine a non-parametric time trend with a specific random effect for each area. López-Vizcaíno et al. (2015) introduce the multinomial logit mixed model with correlate time to estimate labour force indicators by counties. This chapter extends the idea of Boubeta et al. (2016b), that uses Poisson models for estimating counts, including the temporal effect. Four temporal area-level mixed models are considered. The first two use independent time effects and the second two assume an autoregressive process of order one. The resulting models are fitted by the method of moments introduced by Jiang (1998) for GLMM.

This chapter derives empirical best predictors, based on temporal area-level Poisson mixed models, for estimating counts and proportions. The statistical methodology is taken and adapted from Jiang and Lahiri (2001) and Jiang (2003). In addition to the EBP, a plug-in predictor is given and empirically studied in simulation experiments. For estimating the EBP mean squared error, we consider the parametric bootstrap MSE 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 present two applications of the developed methodology to data from the $2010-2013$ Spanish living conditions survey and from $2007-2008$ forest fires in Galicia. The target of the applications is to study the evolution of poverty proportions at county level by sex and the number of forest fires by areas respectively over time.

The chapter is organized as follows. Section 4.2 introduces four area-level Poisson mixed models and the employed model-based fitting algorithm. Section 4.3 presents the EBP and the plug-in predictors of functions of fixed and small area specific random effects. Section 4.4 gives an MSE estimator of the EBP based on a bootstrap approach. Section 4.5 presents two simulation experiments. The first simulation studies the behaviour of the MM fitting algorithm. The second simulation compares the performances of the EBP and the plug-in predictors. Section 4.6 applies the developed methodology to data from the 2010-2013 SLCS and from 2007-2008 forest fires of Galicia. Section 4.7 gives some conclusions.

4.2 The models and the MM algorithms

This section introduces four area-level Poisson mixed models with time effects and their fitting algorithms. They generalize the area-level Poisson mixed model introduced in Chapter 2 to the temporal context. The first two models (Models T1 and T1_2) have independent time random effects. The random effects of the second models (Models T2 and T2_2) follow an AR(1) autoregressive process within each domain. Along this chapter, D and T denote the total numbers of domains and time instants respectively. The corresponding indices are d and t, where $d = 1, \ldots, D$ and $t = 1, \ldots, T$.

4.2.1 Models with independent time effects

This section introduces two temporal models with independent time effects. Both models assume that the temporal correlation of the target variable is fully described by the auxiliary variables. Model T1 considers two independent sets of random effects such that $\{v_{1,d} : \ d = 1, \ldots, D\}$ and $\{v_{2,dt} : \ d = 1, \ldots, D, \, t = 1, \ldots, T\}$ are i.i.d. $N(0, 1)$. They denote the area and the interaction area-time effects that are not explained by the fixed part of the model. In matrix notation, we have

$$\begin{aligned} \boldsymbol{v}_1 &= \underset{1 \le d \le D}{\text{col}}(v_{1,d}) \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D), \\ \boldsymbol{v}_{2d} &= \underset{1 \le t \le T}{\text{col}}(v_{2,dt}) \sim N(\boldsymbol{0}, \boldsymbol{I}_T), \ \boldsymbol{v}_2 = \underset{1 \le d \le D}{\text{col}}(\boldsymbol{v}_{2d}) \sim N(\boldsymbol{0}, \boldsymbol{I}_{DT}) \end{aligned}$$

and $\boldsymbol{v} = (\boldsymbol{v}_1', \boldsymbol{v}_2')' \sim N(\boldsymbol{0}, \boldsymbol{I}_{D(T+1)})$. We have

$$f_v(\boldsymbol{v}_1, \boldsymbol{v}_2) = (2\pi)^{-D(T+1)/2} \exp\left\{ -\frac{1}{2} \boldsymbol{v}_1' \boldsymbol{v}_1 - \frac{1}{2} \boldsymbol{v}_2' \boldsymbol{v}_2 \right\}.$$

The distribution of the target variable y_{dt}, conditionally on the random effects $v_{1,d}$ and $v_{2,dt}$, is

$$y_{dt} | v_{1,d}, v_{2,dt} \sim \text{Poisson}(\mu_{dt}), \quad d = 1, \ldots, D, \ t = 1, \ldots, T. \tag{4.2.1}$$

Given the relationship between Poisson and binomial distributions, as in previous chapters, we take $\mu_{dt} = \nu_{dt} p_{dt}$, where ν_{dt} and p_{dt} are size and probability parameters respectively. In practice, ν_{dt} is known and equal to the sample size of domain d at time instant t. For the natural parameter, we assume that it can be expressed in terms of a set of auxiliary

variables through a regression model, i.e.

$$\text{Model T1: } \log \mu_{dt} = \log \nu_{dt} + \boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt},\ d = 1, \ldots, D,\ t = 1, \ldots, T, \quad (4.2.2)$$

where $\boldsymbol{\beta} = \underset{1\le k\le p}{\text{col}}\,(\beta_k)$ is the column vector of regression coefficients, $\boldsymbol{x}_{dt} = \underset{1\le k\le p}{\text{col}'}\,(x_{dtk})$ is the row vector of auxiliary variables and ϕ_1 and ϕ_2 are the variance component parameters. If we define $u_{1,d} = \phi_1 v_{1,d}$ and $u_{2,dt} = \phi_2 v_{2,dt}$, then ϕ_1 and ϕ_2 are the variances of $u_{1,d}$ and $u_{2,dt}$ respectively. These variances can be interpreted as the variability between domain and between time periods within each domain respectively.

Further, Model T1 assumes that the y_{dt}'s are independent conditionally on $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. It holds that

$$\mathbb{P}(y_{dt}|\boldsymbol{v}) = \mathbb{P}(y_{dt}|v_{dt}) = \frac{1}{y_{dt}!}\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}},$$

where $p_{dt} = \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}$. The probability function of the response variable $\boldsymbol{y}$ conditionally on the random effects $\boldsymbol{v}$ is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) = \prod_{d=1}^{D}\prod_{t=1}^{T}\mathbb{P}(y_{dt}|\boldsymbol{v}),$$

and

$$\mathbb{P}(\boldsymbol{y}) = \int_{\mathbb{R}^{D(T+1)}} \mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) f_v(\boldsymbol{v}_1, \boldsymbol{v}_2)\, d\boldsymbol{v}_1 d\boldsymbol{v}_2 = \int_{\mathbb{R}^{D(T+1)}} \psi(\boldsymbol{y}, \boldsymbol{v})\, d\boldsymbol{v},$$

where

$$\begin{aligned}\psi(\boldsymbol{y}, \boldsymbol{v}) &= (2\pi)^{-\frac{D(T+1)}{2}} \exp\left\{\frac{-\boldsymbol{v}_1'\boldsymbol{v}_1 - \boldsymbol{v}_2'\boldsymbol{v}_2}{2}\right\} \prod_{d=1}^{D}\prod_{t=1}^{T}\frac{\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}}}{y_{dt}!} \\ &= c(\boldsymbol{y}) \exp\left\{\frac{-\boldsymbol{v}_1'\boldsymbol{v}_1 - \boldsymbol{v}_2'\boldsymbol{v}_2}{2}\right\} \exp\left\{\sum_{d=1}^{D}\sum_{t=1}^{T}\left\{-\nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}\right\}\right\} \\ &\cdot \exp\left\{\sum_{k=1}^{p}\Big(\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}x_{dtk}\Big)\beta_k + \phi_1\sum_{d=1}^{D} y_{d\cdot}v_{1,d} + \phi_2\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}v_{2,dt}\right\},\end{aligned}$$

$c(\boldsymbol{y}) = (2\pi)^{-\frac{D(T+1)}{2}} \prod_{d=1}^{D}\prod_{t=1}^{T}\left(\nu_{dt}^{y_{dt}}/y_{dt}!\right)$ and $y_{d\cdot} = \sum_{t=1}^{T} y_{dt}$.

For fitting the area-level Poisson mixed model with independent time effects, this section uses the MM algorithm based on the method of simulated moments suggested by Jiang (1998). A natural set of equations for applying this method is

$$
\begin{aligned}
0 = f_k(\boldsymbol{\theta}) &= M_k(\boldsymbol{\theta}) - \hat{M}_k = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]x_{dtk} - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}x_{dtk}, \quad k = 1,\ldots,p,\\
0 = f_{p+1}(\boldsymbol{\theta}) &= M_{p+1}(\boldsymbol{\theta}) - \hat{M}_{p+1} = \frac{1}{DT^2}\sum_{d=1}^{D}\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] - \frac{1}{DT^2}\sum_{d=1}^{D} y_{d.}^2,\\
0 = f_{p+2}(\boldsymbol{\theta}) &= M_{p+2}(\boldsymbol{\theta}) - \hat{M}_{p+2} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}^2,
\end{aligned}
\tag{4.2.3}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi_1, \phi_2)$ is the vector of all model parameters. The MM estimator of $\boldsymbol{\theta}$ is obtained by solving the system (4.2.3) of nonlinear equations. The updating formula of the Newton-Raphson algorithm is

$$
\boldsymbol{\theta}^{(\ell+1)} = \boldsymbol{\theta}^{(\ell)} - \boldsymbol{H}^{-1}(\boldsymbol{\theta}^{(\ell)})\boldsymbol{f}(\boldsymbol{\theta}^{(\ell)}),
\tag{4.2.4}
$$

where

$$
\boldsymbol{\theta} = \underset{1\le k\le p+2}{\text{col}}(\theta_k), \quad \boldsymbol{f}(\boldsymbol{\theta}) = \underset{1\le k\le p+2}{\text{col}}(f_k(\boldsymbol{\theta})) \quad \text{and} \quad \boldsymbol{H}(\boldsymbol{\theta}) = \left(\frac{\partial f_k(\boldsymbol{\theta})}{\partial\theta_r}\right)_{k,r=1,\ldots,p+2}.
\tag{4.2.5}
$$

The expectations and partial derivatives appearing in (4.2.4) are calculated below under Model T1. For the first p MM equations, the expectation of y_{dt} is

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}]\big] = \mathbb{E}_v[\nu_{dt}p_{dt}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\nu_{dt}p_{dt}f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt}\\
&= \int_{-\infty}^{\infty}\nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \phi_1 v_{1,d}\big\}f_v(v_{1,d})\,dv_{1,d} = \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2+\phi_2^2)\big\}.
\end{aligned}
$$

Therefore

$$
f_k(\boldsymbol{\theta}) = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2+\phi_2^2)\big\}x_{dtk} - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}x_{dtk}, \quad k = 1,\ldots,p.
$$

The partial derivatives of y_{dt} are

$$
\begin{aligned}
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\beta_k} &= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2+\phi_2^2)\big\}x_{dtk},\\
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\phi_i} &= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2+\phi_2^2)\big\}\phi_i, \ \ i = 1,2.
\end{aligned}
$$

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

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] = \mathrm{var}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}] + \mathbb{E}_{\boldsymbol{\theta}}^2[y_{dt}|\boldsymbol{v}] = \nu_{dt}p_{dt} + \nu_{dt}^2 p_{dt}^2. \tag{4.2.6}$$

Therefore

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}]\big] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}p_{dt} f_v(\boldsymbol{v}_d)\, d\boldsymbol{v}_d \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}^2 p_{dt}^2 f(v_{1,d}) f(v_{2,dt})\, dv_{1,d} dv_{2,dt} = A + B,\end{aligned}$$

where

$$\begin{aligned}B &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}^2 p_{dt}^2 f(v_{2,dt}) f(v_{1,d})\, dv_{2,dt} dv_{1,d} \\ &= \nu_{dt}^2 \int_{-\infty}^{\infty} \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_2^2) + 2\phi_1 v_{1,d}\big\}(2\pi)^{-1/2} \exp\big\{-\frac{1}{2}v_{1,d}^2\big\}\, dv_{1,d} \\ &= \nu_{dt}^2 \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2)\big\}.\end{aligned}$$

As a consequence

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\big\} + \nu_{dt}^2 \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2)\big\},$$

and hence the $(p+2)$th MM equation is

$$f_{p+2}(\boldsymbol{\theta}) = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\left\{\nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\big\} + \nu_{dt}^2 \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2)\big\} - y_{dt}^2\right\}.$$

The partial derivatives of y_{dt}^2 are

$$\begin{aligned}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \beta_k} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\big\} x_{dtk} + 2\nu_{dt}^2 \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2)\big\} x_{dtk}, \\ \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \phi_i} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\big\} \phi_i + 4\nu_{dt}^2 \exp\big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2)\big\} \phi_i,\ i = 1, 2.\end{aligned}$$

The expectation of $y_{d.}^2$ is $\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2|\boldsymbol{v}]\big]$, where

$$\begin{aligned}y_{d.}^2 &= \sum_{t=1}^{T} y_{dt}^2 + \sum_{t_1 \neq t_2} y_{dt_1} y_{dt_2}, \\ \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] &= \mathrm{var}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}] + \mathbb{E}_{\boldsymbol{\theta}}^2[y_{dt}|\boldsymbol{v}] = \nu_{dt}p_{dt} + \nu_{dt}^2 p_{dt}^2.\end{aligned} \tag{4.2.7}$$

Therefore,

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2|\boldsymbol{v}] = \sum_{t=1}^{T} \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] + \sum_{t_1 \neq t_2} \mathbb{E}_{\boldsymbol{\theta}}[y_{dt_1}|\boldsymbol{v}]\mathbb{E}_{\boldsymbol{\theta}}[y_{dt_2}|\boldsymbol{v}], \tag{4.2.8}$$

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] = \sum_{t=1}^{T} \nu_{dt}\mathbb{E}_v[p_{dt}] + \sum_{t=1}^{T} \nu_{dt}^2\mathbb{E}_v[p_{dt}^2] + \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2}\mathbb{E}_v[p_{dt_1}p_{dt_2}]. \tag{4.2.9}$$

The expectation of $\mathbb{E}_v[p_{dt_1}p_{dt_2}]$ is

$$\begin{aligned}
\mathbb{E}_v[p_{dt_1}p_{dt_2}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{dt_1}p_{dt_2} f(v_{1,d}) f(v_{2,dt_1}) f(v_{2,dt_2})\, dv_{1,d} dv_{2,dt_1} dv_{2,dt_2} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \phi_2 v_{2,dt_1} \\
&+ 2\phi_1 v_{1,d}\}\, f(v_{1,d}) f(v_{2,dt_1})\, dv_{1,d} dv_{2,dt_1} \\
&= \int_{-\infty}^{\infty} \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \frac{1}{2}\phi_2^2 + 2\phi_1 v_{1,d}\} f(v_{1,d})\, dv_{1,d} \\
&= \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \phi_2^2 + 2\phi_1^2\},
\end{aligned}$$

and then

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] &= \sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\right\} + \sum_{t=1}^{T} \nu_{dt}^2 \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2(\phi_1^2 + \phi_2^2)\right\} \\
&+ \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2} \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1^2 + \phi_2^2\} \pm \sum_{t=1}^{T} \nu_{dt}^2 \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2 + \phi_2^2\right\} \\
&= \sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\right\} + \sum_{t=1}^{T} \nu_{dt}^2 \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2(\phi_1^2 + \phi_2^2)\right\} \\
&- \sum_{t=1}^{T} \nu_{dt}^2 \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2 + \phi_2^2\right\} + \left(\sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{1}{2}\phi_2^2\right\}\right)^2.
\end{aligned}$$

Therefore, the $(p+1)$th MM equation is

$$\begin{aligned}
f_{p+1}(\boldsymbol{\theta}) = \frac{1}{DT^2} \sum_{d=1}^{D} \Bigg\{ &\sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\right\} \\
&+ (e^{\phi_2^2} - 1) \sum_{t=1}^{T} \nu_{dt}^2 \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2 + \phi_2^2\right\} \\
&+ \left(\sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{1}{2}\phi_2^2\right\}\right)^2 \Bigg\} - \frac{1}{DT^2} \sum_{d=1}^{D} y_{d.}^2.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \beta_k} &= \sum_{t=1}^{T} \nu_{dt} C_{dt} x_{dtk} + 2\sum_{t=1}^{T} \nu_{dt}^2 D_{dt} x_{dtk} - 2\sum_{t=1}^{T} \nu_{dt}^2 E_{dt} x_{dtk} \\
&\quad + 2\left(\sum_{t=1}^{T} \nu_{dt} F_{dt}\right) \sum_{t=1}^{T} \nu_{dt} F_{dt} x_{dtk}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \phi_1} &= \sum_{t=1}^{T} \nu_{dt} C_{dt} \phi_1 + 4\sum_{t=1}^{T} \nu_{dt}^2 D_{dt} \phi_1 - 4\sum_{t=1}^{T} \nu_{dt}^2 E_{dt} \phi_1 \\
&\quad + 4\left(\sum_{t=1}^{T} \nu_{dt} F_{dt}\right) \sum_{t=1}^{T} \nu_{dt} F_{dt} \phi_1, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \phi_2} &= \sum_{t=1}^{T} \nu_{dt} C_{dt} \phi_2 + 4\sum_{t=1}^{T} \nu_{dt}^2 D_{dt} \phi_2 - 2\sum_{t=1}^{T} \nu_{dt}^2 E_{dt} \phi_2 \\
&\quad + 2\left(\sum_{t=1}^{T} \nu_{dt} F_{dt}\right) \sum_{t=1}^{T} \nu_{dt} F_{dt} \phi_2,
\end{aligned}$$

where

$$\begin{aligned}
&C_{dt} = \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \phi_2^2)\right\},\ D_{dt} = \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2(\phi_1^2 + \phi_2^2)\right\}, \\
&E_{dt} = \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2 + \phi_2^2\right\},\ F_{dt} = \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{1}{2}\phi_2^2\right\}.
\end{aligned}$$

The elements of the Jacobian matrix $\boldsymbol{H}$ are

$$\begin{aligned}
H_{kr} &= \frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \theta_r} x_{dtk}, \quad k = 1, \ldots, p,\ r = 1, \ldots, p+2, \\
H_{p+1r} &= \frac{\partial f_{p+1}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT^2}\sum_{d=1}^{D} \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \theta_r}, \quad r = 1, \ldots, p+2, \\
H_{p+2r} &= \frac{\partial f_{p+2}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} \frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \theta_r}, \quad r = 1, \ldots, p+2.
\end{aligned}$$

The MM algorithm under Model T1 follows the steps of Algorithm 1, replacing $\boldsymbol{\theta}$, $\boldsymbol{H}$ and $\boldsymbol{f}$ for those given in (4.2.5).

The efficiency of iterative algorithms increases if the starting value is close to the true solution. As starting value, we propose $\boldsymbol{\beta}^{(0)} = \tilde{\boldsymbol{\beta}}$, where $\tilde{\boldsymbol{\beta}}$ is the maximum likelihood estimator under the model without random effects, Model T0, defined by

$$\text{Model T0: } \log \mu_{dt} = \log \nu_{dt} + \boldsymbol{x}_{dt}\boldsymbol{\beta}, \quad d = 1, \ldots, D,\ t = 1, \ldots, T. \tag{4.2.10}$$

Regarding the variance components, we use

$$\phi_1^{(0)} = \left(\frac{1}{D} \sum_{d=1}^{D} (\tilde{\eta}_{d.} - \hat{\eta}_{d.}^{(0)})^2 \right)^{1/2}, \quad \phi_2^{(0)} = \left(\frac{1}{DT} \sum_{d=1}^{D} \sum_{t=1}^{T} (\tilde{\eta}_{dt} - \hat{\eta}_{dt}^{(0)})^2 \right)^{1/2},$$

where $\tilde{\eta}_{dt} = \log \nu_{dt} + \boldsymbol{x}_{dt}\tilde{\boldsymbol{\beta}}$, $\hat{\eta}_{dt}^{(0)} = \log \hat{p}_{dt}^{(0)}$, $\hat{p}_{dt}^{(0)} = \frac{y_{dt}+1}{\nu_{dt}+1}$, $\tilde{\eta}_{d.} = \frac{1}{T}\sum_{t=1}^{T} \tilde{\eta}_{dt}$, $\hat{\eta}_{d.}^{(0)} = \log \hat{p}_{d.}^{(0)}$ and $\hat{p}_{d.}^{(0)} = \frac{1}{T}\sum_{t=1}^{T} \frac{y_{dt}+1}{\nu_{dt}+1}$.

The simplified version of Model T1, Model T1_2, assumes (4.2.1) and incorporates only the area-time random effect $v_{2,dt}$, i.e.

$$\text{Model T1}_2\text{:} \quad \log \mu_{dt} = \log \nu_{dt} + \boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_2 v_{2,dt}, \quad d = 1, \ldots, D, \ t = 1, \ldots, T. \tag{4.2.11}$$

If the domains of Model T1_2 are the domains of Model T1 crossed by time, then Model T1_2 is equivalent to the Model 1 studied in Boubeta et al. (2016b).

The asymptotic variance of the MM estimators can be approximated by a similar bootstrap algorithm to that described in Section 2.3.1.

4.2.2 Models with AR(1)-correlated time effects

This section introduces two temporal models with correlated time effects. Model T2 considers two independent sets of random effects such that $\{v_{1,d} : d = 1, \ldots, D\}$ are i.i.d. $N(0,1)$ and $\{v_{2,dt} : d = 1, \ldots, D, t = 1, \ldots, T\}$ are correlated within each domain d and independent between domains. More concretely, it assumes that $\boldsymbol{v}_1 = \underset{1\leq d\leq D}{\text{col}}(v_{1,d}) \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D)$, $\boldsymbol{v}_{2,d} = \underset{1\leq t\leq T}{\text{col}}(v_{2,dt}) \sim N(\boldsymbol{0}, \Omega_d(\rho))$ and $\boldsymbol{v}_2 = \underset{1\leq d\leq D}{\text{col}}(\boldsymbol{v}_{2,d}) \sim N(\boldsymbol{0}, \Omega(\rho))$. The covariance matrix $\Omega(\rho)$ of $\boldsymbol{v}_2$ is a block diagonal matrix, where each block Ω_d $(d = 1, \ldots, D)$ is

$$\Omega_d = \Omega_d(\rho) = \frac{\boldsymbol{A}_d(\rho)}{1-\rho^2}, \quad \boldsymbol{A}_d(\rho) = \begin{pmatrix} 1 & \rho & \ldots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \ddots & & \rho^{T-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \rho^{T-2} & & \ddots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \ldots & \rho & 1 \end{pmatrix}.$$

For any domain d, the vector $\boldsymbol{v}_{2,d}$ has the same multivariate distribution as a section of size T of an AR(1) time series. Therefore, the components of $\boldsymbol{v}_{2,d}$ are AR(1)-correlated.

We have

$$f_v(\boldsymbol{v}_1, \boldsymbol{v}_2) = (2\pi)^{-D(T+1)/2} |\Omega_d(\rho)|^{-D/2} \exp\left\{ -\frac{1}{2}\, \boldsymbol{v}_1'\boldsymbol{v}_1 - \frac{1}{2} \sum_{d=1}^{D} \boldsymbol{v}_{2d}' \Omega_d^{-1}(\rho) \boldsymbol{v}_{2d} \right\}.$$

The distribution of the target variable y_{dt} conditional on the random effects and the natural parameter keep the expressions (4.2.1) and (4.2.2) of Model T1. Model T2 also assumes the hypothesis of independence of the response variables y_{dt}'s conditionally on the random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. The MM method is employed to fit Model T2. The system of MM nonlinear equations has the three equations (4.2.3) and the new equation associated to the temporal correlation, i.e.

$$\begin{aligned} 0 = f_{p+3}(\boldsymbol{\theta}) &= M_{p+3}(\boldsymbol{\theta}) - \hat{M}_{p+3} \\ &= \frac{1}{D(T-1)} \sum_{d=1}^{D} \sum_{t=2}^{T} \mathbb{E}_{\boldsymbol{\theta}}[y_{dt} y_{dt-1}] - \frac{1}{D(T-1)} \sum_{d=1}^{D} \sum_{t=2}^{T} y_{dt} y_{dt-1}, \end{aligned} \tag{4.2.12}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \phi_1, \phi_2, \rho)$. A Newton-Raphson algorithm can be applied to solve the system of nonlinear equations (4.2.3) and (4.2.12). The updating equation appears in (4.2.4), but the vector $\boldsymbol{f}$ and matrix $\boldsymbol{H}$ are different.

Concerning the first p MM equations, the expectation of y_{dt} is $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}]\big] = \nu_{dt}\mathbb{E}_v[p_{dt}]$. Taking into account the moment generation function of $Y \sim N\left(\mu, \sigma^2\right)$,

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

it holds that

$$\begin{aligned} \mathbb{E}_v[p_{dt}] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\right\}\, f(v_{1,d}) f(v_{2,dt})\, dv_{1,d} dv_{2,dt} \\ &= \int_{-\infty}^{\infty} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d}\right\}\, \Psi\big(\phi_2; 0, (1-\rho^2)^{-1}\big)\, f(v_{1,d})\, dv_{1,d} \\ &= \int_{-\infty}^{\infty} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\, \phi_2^2 (1-\rho^2)^{-1} + \phi_1 v_{1,d}\right\} f_v(v_{1,d})\, dv_{1,d} \\ &= \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\, \phi_2^2 (1-\rho^2)^{-1}\right\}\, \Psi(\phi_1; 0, 1) \\ &= \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\, \phi_1^2 + \frac{1}{2}\, \phi_2^2 (1-\rho^2)^{-1}\right\}. \end{aligned}$$

Then, the first $k = 1, \ldots, p$ MM equations are

$$f_k(\boldsymbol{\theta}) = \frac{1}{DT} \sum_{d=1}^{D} \sum_{t=1}^{T} \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\phi_1^2 + \frac{1}{2}\, \phi_2^2 (1-\rho^2)^{-1}\right\} x_{dtk} - \frac{1}{DT} \sum_{d=1}^{D} \sum_{t=1}^{T} y_{dt} x_{dtk}.$$

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

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \beta_k} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\,\phi_1^2 + \frac{1}{2}\,\phi_2^2(1-\rho^2)^{-1}\big\} x_{dtk},\\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \phi_1} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\,\phi_1^2 + \frac{1}{2}\,\phi_2^2(1-\rho^2)^{-1}\big\} \phi_1,\\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \phi_2} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\,\phi_1^2 + \frac{1}{2}\,\phi_2^2(1-\rho^2)^{-1}\big\} \phi_2(1-\rho^2)^{-1},\\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \rho} &= \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\,\phi_1^2 + \frac{1}{2}\,\phi_2^2(1-\rho^2)^{-1}\big\} \phi_2^2\rho(1-\rho^2)^{-2}.
\end{aligned}$$

The expectation of y_{dt}^2 is $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}]\big]$, where $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}]$ is given by (4.2.6). Therefore

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}]\big] = \nu_{dt}\mathbb{E}_v[p_{dt}] + \nu_{dt}^2\mathbb{E}_v[p_{dt}^2],$$

where

$$\begin{aligned}
\mathbb{E}_v[p_{dt}^2] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{dt}^2 f(v_{2,dt}) f(v_{1,d})\, dv_{2,dt} dv_{1,d}\\
&= \int_{-\infty}^{\infty} \exp\big\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1 v_{1,d}\big\}\, \Psi\big(2\phi_2; 0, (1-\rho^2)^{-1}\big) f(v_{1,d})\, dv_{1,d}\\
&= \int_{-\infty}^{\infty} \exp\big\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_2^2(1-\rho^2)^{-1}\big\} \exp\big\{2\phi_1 v_{1,d}\big\} f(v_{1,d})\, dv_{1,d}\\
&= \exp\big\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_2^2(1-\rho^2)^{-1}\big\} \Psi\big(2\phi_1; 0, 1\big)\\
&= \exp\big\{2\big(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2(1-\rho^2)^{-1}\big)\big\}.
\end{aligned}$$

The expectation of y_{dt}^2 is

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\big(\phi_1^2 + \phi_2^2(1-\rho^2)^{-1}\big)\big\}\\
+ \nu_{dt}^2 \exp\big\{2\big(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \phi_2^2(1-\rho^2)^{-1}\big)\big\}.
\end{aligned}$$

Then, the $(p+2)$th MM equation is

$$\begin{aligned}
f_{p+2}(\boldsymbol{\theta}) = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\Big\{\nu_{dt} \exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\big\}\\
+ \nu_{dt}^2 \exp\big\{2\big(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{\phi_2^2}{1-\rho^2}\big)\big\}\Big\} - \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}^2.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \beta_k} &= \nu_{dt} x_{dtk} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\} \\
&\quad + 2\nu_{dt}^2 x_{dtk} \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \phi_1} &= \nu_{dt}\phi_1 \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\} \\
&\quad + 4\nu_{dt}^2 \phi_1 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \phi_2} &= \frac{\nu_{dt}\phi_2}{1-\rho^2} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\} \\
&\quad + \frac{4\nu_{dt}^2\phi_2}{1-\rho^2} \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \rho} &= \frac{\nu_{dt}\phi_2^2\rho}{(1-\rho^2)^2} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\} \\
&\quad + \frac{4\nu_{dt}^2\phi_2^2\rho}{(1-\rho^2)^2} \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2 + \frac{\phi_2^2}{1-\rho^2})\right\}.
\end{aligned}$$

The expectation of $y_{d.}^2$ is $\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] = \mathbb{E}_v\left[\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2|\boldsymbol{v}]\right]$, where $y_{d.}^2$ is given in (4.2.7) and equations (4.2.8)–(4.2.9) are fulfilled. The expectation $e_{dt_1t_2} = \mathbb{E}_v[p_{dt_1}p_{dt_2}]$ is

$$\begin{aligned}
e_{dt_1t_2} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1 v_{1,d} + \phi_2 v_{2,dt_1} + \phi_2 v_{2,dt_2}\} \\
&\quad \cdot f(v_{1,d}) f(v_{2,dt_1}) f(v_{2,dt_2}|v_{2,dt_1})\, dv_{1,d} dv_{2,dt_1} dv_{2,dt_2} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1 v_{1,d} + \phi_2 v_{2,dt_1}\} \\
&\quad \cdot \Psi\Big(\phi_2; \rho^{|t_1-t_2|} v_{2,dt_1}, \frac{1-\rho^{2|t_1-t_2|}}{1-\rho^2}\Big) f(v_{1,d}) f(v_{2,dt_1})\, dv_{1,d} dv_{2,dt_1} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\Big\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1 v_{1,d} + \phi_2 v_{2,dt_1} + \phi_2 \rho^{|t_1-t_2|} v_{2,dt_1} \\
&\quad + \frac{1}{2}\phi_2^2 \frac{1-\rho^{2|t_1-t_2|}}{1-\rho^2}\Big\} f(v_{1,d}) f(v_{2,dt_1})\, dv_{1,d} dv_{2,dt_1} \\
&= \int_{-\infty}^{\infty} \exp\Big\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 \frac{1-\rho^{2|t_1-t_2|}}{1-\rho^2} + 2\phi_1 v_{1,d}\Big\} \\
&\quad \cdot \Psi\big(\phi_2(1+\rho^{|t_1-t_2|}); 0, (1-\rho^2)^{-1}\big) f(v_{1,d})\, dv_{1,d} \\
&= \int_{-\infty}^{\infty} \exp\Big\{(\boldsymbol{x}_{dt_1} + \boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{\phi_2^2}{2} \frac{1-\rho^{2|t_1-t_2|}}{1-\rho^2} + 2\phi_1 v_{1,d} \\
&\quad + \frac{\phi_2^2}{2} \frac{(1+\rho^{|t_1-t_2|})^2}{1-\rho^2}\Big\} f(v_{1,d})\, dv_{1,d}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
e_{dt_1t_2} &= \exp\Big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\frac{\phi_2^2}{2}\frac{1-\rho^{2|t_1-t_2|}}{1-\rho^2}+\frac{\phi_2^2}{2}\frac{(1+\rho^{|t_1-t_2|})^2}{1-\rho^2}\Big\}\Psi\big(2\phi_1;0,1\big)\\
&= \exp\Big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\frac{\phi_2^2}{2(1-\rho^2)}\big[1-\rho^{2|t_1-t_2|}+\rho^{2|t_1-t_2|}+2\rho^{|t_1-t_2|}+1\big]+2\phi_1^2\Big\}\\
&= \exp\Big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\frac{\phi_2^2(1+\rho^{|t_1-t_2|})}{1-\rho^2}+2\phi_1^2\Big\}\\
&= \exp\big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\phi_2^2 a_{dt_1t_2}(\rho)+2\phi_1^2\big\},
\end{aligned}$$

where

$$\begin{aligned}
a_{dt_1t_2}(\rho) &= \frac{1+\rho^{|t_1-t_2|}}{1-\rho^2},\\
a'_{dt_1t_2}(\rho) &= \frac{\partial a_{dt_1t_2}}{\partial\rho} = \frac{|t_1-t_2|\rho^{|t_1-t_2|-1}(1-\rho^2)+2\rho(1+\rho^{|t_1-t_2|})}{(1-\rho^2)^2}.
\end{aligned}$$

The expectation $\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]$ is

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] &= \sum_{t=1}^{T}\nu_{dt}\mathbb{E}_v[p_{dt}]+\sum_{t=1}^{T}\nu_{dt}^2\mathbb{E}_v[p_{dt}^2]+\sum_{t_1\neq t_2}\nu_{dt_1}\nu_{dt_2}\mathbb{E}_v[p_{dt_1}p_{dt_2}]\\
&= \sum_{t=1}^{T}\nu_{dt}P_{dt}+\sum_{t=1}^{T}\nu_{dt}^2Q_{dt}+\sum_{t_1\neq t_2}\nu_{dt_1}\nu_{dt_2}R_{dt_1t_2},
\end{aligned}$$

where

$$\begin{aligned}
P_{dt} &= \exp\Big\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\frac{1}{2}(\phi_1^2+\frac{\phi_2^2}{1-\rho^2})\Big\},\\
Q_{dt} &= \exp\Big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1^2+\frac{\phi_2^2}{1-\rho^2})\Big\},\\
R_{dt_1t_2} &= \exp\big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\phi_2^2 a_{dt_1t_2}(\rho)+2\phi_1^2\big\}.
\end{aligned}$$

Hence, the $(p+1)$th MM equation is

$$\begin{aligned}
f_{p+1}(\boldsymbol{\theta}) = \frac{1}{DT^2}\sum_{d=1}^{D}\Bigg\{&\sum_{t=1}^{T}\nu_{dt}\exp\Big\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\frac{1}{2}(\phi_1^2+\frac{\phi_2^2}{1-\rho^2})\Big\}\\
&+\sum_{t=1}^{T}\nu_{dt}^2\exp\Big\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1^2+\frac{\phi_2^2}{1-\rho^2})\Big\}\\
&+\sum_{t_1\neq t_2}\nu_{dt_1}\nu_{dt_2}\exp\big\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta}+\phi_2^2 a_{dt_1t_2}(\rho)+2\phi_1^2\big\}\Bigg\}-\frac{1}{DT^2}\sum_{d=1}^{D}y_{d.}^2.
\end{aligned}$$

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

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \beta_k} = \sum_{t=1}^{T} \nu_{dt} P_{dt} x_{dtk} + 2\sum_{t=1}^{T} \nu_{dt}^2 Q_{dt} x_{dtk} + \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2} R_{dt_1t_2}(x_{dt_1k} + x_{dt_2k}),$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \phi_1} = \phi_1 \sum_{t=1}^{T} \nu_{dt} P_{dt} + 4\phi_1 \sum_{t=1}^{T} \nu_{dt}^2 Q_{dt} + 4\phi_1 \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2} R_{dt_1t_2},$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \phi_2} = \frac{\phi_2}{1-\rho^2} \sum_{t=1}^{T} \nu_{dt} P_{dt} + \frac{4\phi_2}{1-\rho^2} \sum_{t=1}^{T} \nu_{dt}^2 Q_{dt} + 2\phi_2 a_{dt_1t_2}(\rho) \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2} R_{dt_1t_2},$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \rho} = \frac{\phi_2^2 \rho}{(1-\rho^2)^2} \sum_{t=1}^{T} \nu_{dt} P_{dt} + \frac{4\phi_2^2 \rho}{(1-\rho^2)^2} \sum_{t=1}^{T} \nu_{dt}^2 Q_{dt} + \phi_2^2 a'_{dt_1t_2}(\rho) \sum_{t_1 \neq t_2} \nu_{dt_1}\nu_{dt_2} R_{dt_1t_2}.$$

The expectation of $y_{dt}y_{dt-1}$ is $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}|\boldsymbol{v}]\big]$. It holds

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}|\boldsymbol{v}] = \mathbb{E}[y_{dt}|\boldsymbol{v}]\mathbb{E}[y_{dt-1}|\boldsymbol{v}] = \nu_{dt}\nu_{dt-1}p_{dt}p_{dt-1},$$

and therefore

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}] &= \nu_{dt}\nu_{dt-1}\mathbb{E}_v[p_{dt}p_{dt-1}] \\
&= \nu_{dt}\nu_{dt-1}\exp\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \phi_2^2 a_{dt(t-1)}(\rho) + 2\phi_1^2\} \\
&= \nu_{dt}\nu_{dt-1}\exp\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\}.
\end{aligned}$$

Then, the $(p+3)$th MM equation is

$$\begin{aligned}
f_{p+3}(\boldsymbol{\theta}) = &\frac{1}{D(T-1)} \sum_{d=1}^{D}\sum_{t=2}^{T} \nu_{dt}\nu_{dt-1}\exp\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\} \\
&- \frac{1}{D(T-1)} \sum_{d=1}^{D}\sum_{t=2}^{T} y_{dt}y_{dt-1}.
\end{aligned}$$

The derivatives of $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]$ are

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]}{\partial \beta_k} = \nu_{dt}\nu_{dt-1}\exp\big\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\big\}(x_{dtk} + x_{dt-1k}),$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]}{\partial \phi_1} = 4\nu_{dt}\nu_{dt-1}\exp\big\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\big\}\phi_1,$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]}{\partial \phi_2} = 2\nu_{dt}\nu_{dt-1}\exp\big\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\big\}\frac{\phi_2}{1-\rho},$$

$$\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]}{\partial \rho} = \nu_{dt}\nu_{dt-1}\exp\big\{(\boldsymbol{x}_{dt} + \boldsymbol{x}_{dt-1})\boldsymbol{\beta} + \frac{\phi_2^2}{1-\rho} + 2\phi_1^2\big\}\frac{\phi_2^2}{(1-\rho)^2}.$$

Finally, the elements of the Jacobian matrix are

$$H_{kr} = \frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \theta_r}\, x_{dtk}, \quad k = 1,\ldots,p,\ r = 1,\ldots,p+3,$$

$$H_{p+1r} = \frac{\partial f_{p+1}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT^2}\sum_{d=1}^{D}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \theta_r}, \quad r = 1,\ldots,p+3,$$

$$H_{p+2r} = \frac{\partial f_{p+2}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \theta_r}, \quad r = 1,\ldots,p+3,$$

$$H_{p+3r} = \frac{\partial f_{p+3}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D(T-1)}\sum_{d=1}^{D}\sum_{t=2}^{T}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}y_{dt-1}]}{\partial \theta_r}, \quad r = 1,\ldots,p+3.$$

Algorithm seeds for initiating the Newton-Raphson algorithm are $\boldsymbol{\beta}^{(0)}$, $\phi_1^{(0)}$ and $\phi_2^{(0)}$, as for Model T1. Concerning the temporal correlation, one can take $\rho^{(0)} = \phi_{12}^{(0)}/\phi_2^{(0)2}$, where

$$\phi_{12}^{(0)} = \frac{1}{D(T-1)}\sum_{d=1}^{D}\sum_{t=2}^{T}(\tilde{\eta}_{dt} - \hat{\eta}_{dt}^{(0)})(\tilde{\eta}_{dt-1} - \hat{\eta}_{dt-1}^{(0)}).$$

The simplified version of Model T2, Model T2_2, assumes (4.2.1) and incorporates only the area-time random effect $v_{2,dt}$, i.e. the natural parameter fulfills the equation (4.2.11). The natural set of equations for applying the MM in Model T2_2 is the same as for Model T2 taking $\phi_1 = 0$ and deleting the equation f_{p+1}.

4.3 The predictors

This section gives the EBP and a plug-in predictor of p_{dt} under Model T1, Model T2 and their simplified versions (Model T1_2 and Model T2_2). The best predictor, $\hat{p}_{dt}(\boldsymbol{\theta})$, of p_{dt} minimizes the mean squared error in the set of unbiased predictors. The EBP is obtained from the BP by substituting parameters and random effects by the corresponding MM estimators and EBPs. Under regularity conditions, the EBPs have asymptotically the properties of the BPs. Nevertheless, D and T are not large enough in small area estimation problems. This is why some simulation experiments are needed for empirically studying the behaviour of the EBPs.

We define $\boldsymbol{y}_d = \underset{1\le t\le T}{\text{col}}(y_{dt})$ and $\boldsymbol{y} = \underset{1\le d\le D}{\text{col}}(\boldsymbol{y}_d)$. Under models Model T1 and Model T2,

the conditional distribution of $\boldsymbol{y}$, given $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}_1,\boldsymbol{v}_2)=\prod_{d=1}^{D}\mathbb{P}(\boldsymbol{y}_d|v_{1,d},\boldsymbol{v}_{2,d}),\quad \mathbb{P}(\boldsymbol{y}_d|v_{1,d},\boldsymbol{v}_{2,d})=\prod_{t=1}^{T}\mathbb{P}(y_{dt}|v_{1,d},v_{2,dt}), \tag{4.3.1}$$

where

$$\begin{aligned}\mathbb{P}(y_{dt}|v_{1,d},v_{2,dt}) &= \frac{1}{y_{dt}!}\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}}\\ &= c_{dt}\exp\left\{y_{dt}(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt})-\nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\right\}.\end{aligned}$$

and the p.d.f. of the random effects $\boldsymbol{v}=(\boldsymbol{v}_1,\boldsymbol{v}_2)$ is

$$f(\boldsymbol{v}_1,\boldsymbol{v}_2)=f(\boldsymbol{v}_1)f(\boldsymbol{v}_2),$$

where

$$f(\boldsymbol{v}_1)=\prod_{d=1}^{D}f(v_{1,d}),\quad f(\boldsymbol{v}_2)=\prod_{d=1}^{D}f(\boldsymbol{v}_{2,d}).$$

Under Model T1$_2$ and Model T2$_2$, the conditional distribution of $\boldsymbol{y}$, given $\boldsymbol{v}_2$, is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}_2)=\prod_{d=1}^{D}\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d}),\quad \mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d})=\prod_{t=1}^{T}\mathbb{P}(y_{dt}|v_{2,dt}), \tag{4.3.2}$$

where

$$\begin{aligned}\mathbb{P}(y_{dt}|v_{2,dt}) &= \frac{1}{y_{dt}!}\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}}\\ &= c_{dt}\exp\left\{y_{dt}(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_2 v_{2,dt})-\nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_2 v_{2,dt}\}\right\}.\end{aligned}$$

4.3.1 The EBP under Model T1

Under Model T1, the p.d.f. of $\boldsymbol{v}_{2,d}$ is

$$f(\boldsymbol{v}_{2,d})=\prod_{t=1}^{T}f(v_{2,dt}),$$

and the p.d.f. of the random effects $v_{1,d}$ and $v_{2,dt}$ are

$$f(v_{1,d}) = (2\pi)^{-1/2} \exp\Big\{ -\frac{1}{2} v_{1,d}^2 \Big\},$$
$$f(v_{2,dt}) = (2\pi)^{-1/2} \exp\Big\{ -\frac{1}{2} v_{2,dt}^2 \Big\}.$$

The BP is the conditional mean $\hat{p}_{dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}]$. We have that $\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}] = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d]$ and

$$\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^{T+1}} p_{dt}\mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^{T+1}} \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}} = \frac{N_{dt}}{D_d}.$$

Using (4.3.1) and the hypothesis of temporal independence, we have

$$\begin{aligned}
N_{dt} &= \int_{\mathbb{R}^{T+1}} \prod_{\tau=1}^{T} \exp\{(y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d} \\
&= \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Bigg[\int_{\mathbb{R}} \exp\{(y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{2,d\tau})\, dv_{2,d\tau} \Bigg] f(v_{1,d})\, dv_{1,d}, \\
D_d &= \int_{\mathbb{R}^{T+1}} \prod_{\tau=1}^{T} \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d} \\
&= \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Bigg[\int_{\mathbb{R}} \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{2,d\tau})\, dv_{2,d\tau} \Bigg] f(v_{1,d})\, dv_{1,d},
\end{aligned}$$

where $\delta_{t\tau}$ denotes the Kronecker delta, i.e. $\delta_{t\tau} = 1$ if $t = \tau$ and $\delta_{t\tau} = 0$ otherwise.

Remark 4.1. The component $N_{dt}(\boldsymbol{y}_d, \boldsymbol{\theta})$ can be expressed in terms of $D_d(\boldsymbol{y}_d, \boldsymbol{\theta})$ as $N_{dt}(\boldsymbol{y}_d, \boldsymbol{\theta}) = D_d(\boldsymbol{y}_d + \boldsymbol{e}_t, \boldsymbol{\theta})$, where $\boldsymbol{e}_t = (\delta_{t1}, \ldots, \delta_{tT})'$.

The EBP of p_{dt} is obtained replacing $\boldsymbol{\theta}$ by a consistent estimator $\hat{\boldsymbol{\theta}}$, i.e. the EBP of p_{dt} is $\hat{p}_{dt}(\hat{\boldsymbol{\theta}})$. Since the EBP calculation involve high-dimensional integrals, it can be approximated by running an antithetic Monte Carlo algorithm. The algorithm steps are

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2)$ by a consistent estimator (for example, using MM).
2. For $s_1 = 1, \ldots, S_1$, $s_2 = 1, \ldots, S_2$, $\tau = 1, \ldots, T$, generate $v_{1,d}^{(s_1)}$, $v_{2,d\tau}^{(s_2)}$ i.i.d. $N(0,1)$

and calculate their antithetical $v_{1,d}^{(S_1+s_1)} = -v_{1,d}^{(s_1)}$ and $v_{2,d\tau}^{(S_2+s_2)} = -v_{2,d\tau}^{(s_2)}$.

3. Approximate the EBP of p_{dt} by $\hat{p}_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{dt}/\hat{D}_d$, where

$$\begin{aligned}
\hat{N}_{dt} &= \sum_{s_1=1}^{2S_1}\prod_{\tau=1}^{T}\sum_{s_2=1}^{2S_2}\exp\Big\{(y_{d\tau}+\delta_{t\tau})(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}}+\hat{\phi}_1 v_{1,d}^{(s_1)}+\hat{\phi}_2 v_{2,d\tau}^{(s_2)}) \\
&- \nu_{d\tau}\exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}}+\hat{\phi}_1 v_{1,d}^{(s_1)}+\hat{\phi}_2 v_{2,d\tau}^{(s_2)}\}\Big\}, \\
\hat{D}_d &= \sum_{s_1=1}^{2S_1}\prod_{\tau=1}^{T}\sum_{s_2=1}^{2S_2}\exp\Big\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}}+\hat{\phi}_1 v_{1,d}^{(s_1)}+\hat{\phi}_2 v_{2,d\tau}^{(s_2)}) \\
&- \nu_{d\tau}\exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}}+\hat{\phi}_1 v_{1,d}^{(s_1)}+\hat{\phi}_2 v_{2,d\tau}^{(s_2)}\}\Big\}.
\end{aligned}$$

As the MM algorithm does not give predictions of $v_{1,d}$ and $v_{2,dt}$, we use their EBPs. The BP of $v_{1,d}$, $d = 1,\ldots,D$, is

$$\hat{v}_{1,d}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{1,d}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^{T+1}} v_{1,d}\mathbb{P}(\boldsymbol{y}_d|v_{1,d},\boldsymbol{v}_{2,d})f(v_{1,d})f(\boldsymbol{v}_{2,d})\,dv_{1,d}d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^{T+1}}\mathbb{P}(\boldsymbol{y}_d|v_{1,d},\boldsymbol{v}_{2,d})f(v_{1,d})f(\boldsymbol{v}_{2,d})\,dv_{1,d}d\boldsymbol{v}_{2,d}} = \frac{N_{1,d}}{D_d},$$

where

$$\begin{aligned}
N_{1,d} = \int_{\mathbb{R}}\prod_{\tau=1}^{T}\Big[\int_{\mathbb{R}}&\exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,d\tau}) \\
&- \nu_{d\tau}\exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,d\tau}\}\}\,f(v_{2,d\tau})\,dv_{2,d\tau}\Big]v_{1,d}f(v_{1,d})\,dv_{1,d},
\end{aligned}$$

and the BP of $v_{2,dt}$, $d = 1,\ldots,D$, $t = 1,\ldots,T$, is

$$\hat{v}_{2,dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{2,dt}|y_{dt}] = \frac{\int_{\mathbb{R}^2} v_{2,dt}\mathbb{P}(y_{dt}|v_{1,d},v_{2,dt})f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt}}{\int_{\mathbb{R}^2}\mathbb{P}(y_{dt}|v_{1,d},v_{2,dt})f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt}} = \frac{N_{2,dt}}{D_{dt}},$$

where

$$\begin{aligned}
N_{2,dt} &= \int_{\mathbb{R}^2}\exp\{y_{dt}(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}) \\
&- \nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\}\,v_{2,dt}f(v_{2,dt})f(v_{1,d})\,dv_{2,dt}dv_{1,d}, \\
D_{dt} &= \int_{\mathbb{R}^2}\exp\{y_{dt}(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}) \\
&- \nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\}\,f(v_{2,dt})f(v_{1,d})\,dv_{2,dt}dv_{1,d}.
\end{aligned}$$

The EBPs of $v_{1,d}$ and $v_{2,dt}$ are $\hat{v}_{1,d}(\hat{\boldsymbol{\theta}})$ and $\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}})$ respectively and they can be approximated by an analogous Monte Carlo algorithm to the one used for the EBP of p_{dt}.

The EBPs of p_{dt} and $v_{2,dt}$ under Model T1$_2$ are equivalent to the respective EBPs of the model studied in Chapter 2 (Boubeta et al., 2016b), considering as domains the interaction domain-time.

4.3.2 The EBP under Model T2

Under Model T2, the p.d.f. of the random effects $v_{1,d}$ and $\boldsymbol{v}_{2,d}$ are

$$f(v_{1,d}) = (2\pi)^{-1/2} \exp\Big\{ -\frac{1}{2}\, v_{1,d}^2 \Big\},$$
$$f(\boldsymbol{v}_{2,d}) = (2\pi)^{-T/2} |\Omega_d(\rho)|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{v}_{2d}' \Omega_d^{-1}(\rho) \boldsymbol{v}_{2d} \right\}.$$

The BP of p_{dt} is $\hat{p}_{dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}] = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d]$, where

$$\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^{T+1}} p_{dt} \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^{T+1}} \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}} = \frac{N_{dt}}{D_d},$$
$$N_{dt} = \int_{\mathbb{R}^{T+1}} \prod_{\tau=1}^{T} \exp\big\{ (y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau})$$
$$- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\} \big\}\, f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d},$$
$$D_d = \int_{\mathbb{R}^{T+1}} \prod_{\tau=1}^{T} \exp\big\{ y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau})$$
$$- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\} \big\}\, f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d},$$

and $\delta_{t\tau}$ is the Kronecker delta. For the Monte Carlo approximation of the EBP, we generate random variables $v_{2,dt}$, $t = 1, \ldots, T$, with an AR(1)-correlation structure within each domain d. The EBP of p_{dt} can be approximated as follows.

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2, \hat{\rho})$.
2. For $s_1 = 1, \ldots, S_1$, generate $v_{1,d}^{(s_1)}$ i.i.d. $N(0,1)$ and calculate $v_{1,d}^{(S_1+s_1)} = -v_{1,d}^{(s_1)}$. For $s_2 = 1, \ldots, S_2$, generate $(v_{2,d1}^{(s_2)}, \ldots, v_{2,dT}^{(s_2)}) \sim N_T\big(\boldsymbol{0}, \Omega_d(\hat{\rho})\big)$ and calculate the corresponding antithetics $(v_{2,d1}^{(S_2+s_2)}, \ldots, v_{2,dT}^{(S_2+s_2)}) = -(v_{2,d1}^{(s_2)}, \ldots, v_{2,dT}^{(s_2)})$.
3. Approximate the EBP of p_{dt} as $\hat{p}_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{dt}/\hat{D}_d$, where

$$\hat{N}_{dt} = \sum_{s_1=1}^{2S_1} \sum_{s_2=1}^{2S_2} \prod_{\tau=1}^{T} \exp\Big\{ (y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)})$$
$$- \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\} \Big\},$$

$$\hat{D}_d = \sum_{s_1=1}^{2S_1} \sum_{s_2=1}^{2S_2} \prod_{\tau=1}^{T} \exp\Big\{ y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\} \Big\}.$$

The BPs of the random effects $v_{1,d}$ and $v_{2,dt}$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, are

$$\hat{v}_{1,d}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{1,d}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^{T+1}} v_{1,d}\mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^{T+1}} \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}} = \frac{N_{1,d}}{D_d},$$

$$\hat{v}_{2,dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{2,dt}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^{T+1}} v_{2,dt}\mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^{T+1}} P(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}} = \frac{N_{2,dt}}{D_d},$$

where

$$\begin{aligned} N_{1,d} = \int_{\mathbb{R}^{T+1}} \prod_{\tau=1}^{T} & \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\ & - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} v_{1,d} f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}, \\ N_{2,dt} = \int_{R^{T+1}} \prod_{\tau=1}^{T} & I_{2,dt}(\tau) \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\ & - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d} d\boldsymbol{v}_{2,d}, \end{aligned}$$

$I_{2,dt}(\tau) = v_{2,dt}$ if $t = \tau$ and $I_{2,dt}(\tau) = 1$ otherwise. The EBPs of $v_{1,d}$ and $v_{2,dt}$ are obtained replacing $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$. They are denoted by $\hat{v}_{1,d}(\hat{\boldsymbol{\theta}})$ and $\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}})$ respectively. Similarly as above, they can be approximated by a Monte Carlo algorithm.

Under Model T2$_2$, the best predictor of p_{dt} is $\hat{p}_{dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}]$, where $p_{dt} = \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_2 v_{2,dt}\}$. In this case, it holds that $\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}] = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d]$ and

$$\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^T} \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_2 v_{2,dt}\}\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d}) f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^T} \mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d}) f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}} = \frac{N_{dt}}{D_d},$$

where using (4.3.2),

$$\begin{aligned} N_{dt} &= \int_{\mathbb{R}^T} \prod_{\tau=1}^{T} \exp\{(y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}) \; - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}\}\} f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}, \\ D_d &= \int_{\mathbb{R}^T} \prod_{\tau=1}^{T} \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}\}\} f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}, \end{aligned}$$

and $\delta_{t\tau}$ is the Kronecker delta, i.e. $\delta_{t\tau} = 1$ if $t = \tau$ and $\delta_{t\tau} = 0$ otherwise.

The EBP of p_{dt} is $\hat{p}_{dt}(\hat{\boldsymbol{\theta}})$ and it can be approximated as follows.

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_2, \hat{\rho})$.

2. For $s_2 = 1, \ldots, S_2$, generate $(v_{2,d1}^{(s_2)}, \ldots, v_{2,dT}^{(s_2)}) \sim N_T\big(\mathbf{0}, \Omega_d(\hat{\rho})\big)$ and calculate the corresponding antithetics $(v_{2,d1}^{(S_2+s_2)}, \ldots, v_{2,dT}^{(S_2+s_2)}) = -(v_{2,d1}^{(S_2+s_2)}, \ldots, v_{2,dT}^{(S_2+s_2)})$.

3. Approximate the EBP of p_{dt} as $\hat{p}_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{dt}/\hat{D}_d$, where

$$\hat{N}_{dt} = \sum_{s_2=1}^{2S_2} \prod_{\tau=1}^{T} \exp\left\{(y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\}\right\},$$
$$\hat{D}_d = \sum_{s_2=1}^{2S_2} \prod_{\tau=1}^{T} \exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\}\right\}.$$

Finally, the BP of the random effects $v_{2,dt}$ $(d = 1, \ldots, D,\ t = 1, \ldots, T)$ is

$$\hat{v}_{2,dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{2,dt}|\boldsymbol{y}_d] = \frac{\int_{\mathbb{R}^T} v_{2,dt}\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d}) f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}}{\int_{\mathbb{R}^T} \mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_{2,d}) f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}} = \frac{N_{2,dt}}{D_d},$$

where

$$N_{2,dt} = \int_{\mathbb{R}^T} \prod_{\tau=1}^{T} I_{2,dt}(\tau) \exp\left\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_2 v_{2,d\tau}\}\right\} f(\boldsymbol{v}_{2,d})\, d\boldsymbol{v}_{2,d}.$$

The EBP of $v_{2,dt}$ under Model T2$_2$ is $\hat{v}_{2,dt} = \hat{v}_{2,dt}(\hat{\boldsymbol{\theta}})$ and it can be approximated by an algorithm similar to the previous one.

4.3.3 The plug-in predictors

Under Model T1 and Model T2, the plug-in predictor of p_{dt} is

$$\hat{p}_{dt}^P = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 \hat{v}_{1,d} + \hat{\phi}_2 \hat{v}_{2,dt}\},$$

where $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_1$ and $\hat{\phi}_2$ are consistent estimators of $\boldsymbol{\beta}$, ϕ_1 and ϕ_2 respectively (for example, the MM estimators) and $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ are the EBPs of $v_{1,d}$ and $v_{2,dt}$ respectively under the assumed model. Similarly, the plug-in predictor of p_{dt} takes the form

$$\hat{p}_{dt}^P = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{\phi}_2 \hat{v}_{2,dt}\}$$

under Model T1$_2$ and Model T2$_2$, where $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_2$ and $\hat{v}_{2,dt}$ are the corresponding estimators and predictors.

The plug-in predictors are not unbiased. As they are defined by using the same formula as the parameter to be estimated and they plug consistent estimators and EBPs in the place of parameters and random effects, they are consistent under regularity conditions. The problem is that D and overall T are not large enough in small area estimation problems. This is why Section 4.5 reports some empirical investigations about their behaviour.

4.4 MSE estimation

The mean squared error (MSE) of the EBP of p_{dt} is considered as accuracy measure under models Model T1 and Model T2. Due to the computational complexity of the corresponding analytical versions, we propose to estimate the MSE of $\hat{p}_{dt}$ by using the following parametric bootstrap algorithm based on the bootstrap procedure given in González-Manteiga et al. (2007). The steps are:

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}}$. Note that $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2)$ in Model T1 and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2, \hat{\rho})$ in Model T2.

2. For each domain d, $d = 1, \ldots, D$, and time instant t, $t = 1, \ldots, T$, repeat B times, $b = 1, \ldots, B$:

   i) Generate the boostrap random effects $v_{1,d}^{*(b)}$ and $v_{2,dt}^{*(b)}$. The domain random effects $v_{1,d}^{*(b)}$ are i.i.d. $N(0,1)$ in both models. The domain-time random effects $v_{2,dt}^{*(b)}$ are i.i.d. $N(0,1)$ in Model T1 and AR(1)-correlated within each domain d in Model T2.

   ii) Calculate the theoretical bootstrap EBP estimator $p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{*(b)} + \hat{\phi}_2 v_{2,dt}^{*(b)}\}$.

   iii) Generate the responses variables $y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)})$.

   iv) Calculate $\hat{\boldsymbol{\theta}}^{*(b)}$ and the corresponding EBP estimator $\hat{p}_{dt}^{*(b)} = \hat{p}_{dt}^{*(b)}(\hat{\boldsymbol{\theta}}^{*(b)}, \hat{v}_{1,d}^{*(b)}, \hat{v}_{2,dt}^{*(b)})$ under Model T1 or Model T2.

3. Output:

$$mse^*(\hat{p}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{p}_{dt}^{*(b)} - p_{dt}^{*(b)}\right)^2. \tag{4.4.1}$$

A similar bootstrap procedure estimates the MSE of $\hat{p}_{dt}$ under the models Model $T1_2$ and Model $T2_2$.

4.5 Simulation experiments

This section presents two simulation experiments for studying the behaviour of the MM fitting algorithm and the two introduced predictors under models T1 and T2. The experiments generate independent response variables $y_{dt}|v_{1,d}, v_{2,dt} \sim \text{Poisson}(\nu_{dt}p_{dt})$, where $p_{dt} = \exp\{\beta_0 + x_{dt}\beta_1 + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}$, $x_{dt} = (d + t/T)/D$, $d = 1, \ldots, D$, $t = 1, \ldots, T$. The domain random effects $v_{1,d}$, $d = 1, \ldots, D$, are i.i.d. $N(0, 1)$ and the domain-time random effects $v_{2,dt}$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, are generated under the assumptions stated in Sections 4.2.1 or 4.2.2.

The model parameters are $\nu_{dt} = 100$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, $\beta_0 = -3$, $\beta_1 = 0.8$ and $\phi_1 = \phi_2 = 0.5$. The time correlation parameter is $\rho = 0.4$ for the model with AR(1)-correlated time effects. The Monte Carlo simulation experiments are carried out with $K = 1000$ iterations under the scenarios $D = 50, 100, 150$ and $T = 5, 9, 12$. The target is studying the influence of the number of areas and time periods in the inference procedures.

4.5.1 Simulation 1

The first simulation studies the behaviour of the MM fitting algorithm introduced in Section 4.2. Tables 4.5.1 and 4.5.2 present the bias (BIAS) and the root mean squared error (RMSE) of the model parameter estimators under Model T1 and Model T2 respectively. The performance measures are

$$BIAS(\theta) = \frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta), \quad RMSE(\theta) = \left(\frac{1}{K}\sum_{k=1}^{K}(\hat{\theta}^{(k)} - \theta)^2\right)^{1/2},$$

where $\theta \in \{\beta_0, \beta_1, \phi_1, \phi_2, \rho\}$. The RMSE of all parameter estimators and the bias of the variance components estimators tend to decrease as the number of domains or time instants increases. However, the bias of the regression parameters estimators is quite stable and does not seem to decrease with D or T. Under Model T2, we recommend $T \geq 5$ for estimating ρ.

Table 4.5.1: BIAS and RMSE for Model T1.

		$T=5$		$T=9$		$T=12$	
D		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
50	$\hat{\beta}_0$	0.0063	0.1752	0.0099	0.1610	0.0214	0.1699
	$\hat{\beta}_1$	-0.0009	0.2861	-0.0082	0.2661	-0.0221	0.2737
	$\hat{\phi}_1$	-0.0408	0.0862	-0.0367	0.0768	-0.0360	0.0757
	$\hat{\phi}_2$	-0.0091	0.0515	-0.0084	0.0396	-0.0072	0.0333
100	$\hat{\beta}_0$	0.0114	0.1258	0.0110	0.1177	0.0192	0.1155
	$\hat{\beta}_1$	-0.0128	0.2035	-0.0197	0.2019	-0.0235	0.2003
	$\hat{\phi}_1$	-0.0231	0.0605	-0.0226	0.0540	-0.0282	0.0556
	$\hat{\phi}_2$	-0.0088	0.0393	-0.0078	0.0288	-0.0077	0.0252
150	$\hat{\beta}_0$	0.0132	0.1087	0.0215	0.1062	0.0098	0.0839
	$\hat{\beta}_1$	-0.0107	0.1888	-0.0257	0.1735	-0.0105	0.1490
	$\hat{\phi}_1$	-0.0196	0.0494	-0.0208	0.0456	-0.0230	0.0495
	$\hat{\phi}_2$	-0.0078	0.0314	-0.0104	0.0226	-0.0054	0.0190

Table 4.5.2: BIAS and RMSE for Model T2.

		$T=5$		$T=9$		$T=12$	
D		BIAS	RMSE	BIAS	RMSE	BIAS	RMSE
50	$\hat{\beta}_0$	0.0355	0.1997	0.0342	0.1812	0.0336	0.1621
	$\hat{\beta}_1$	-0.0619	0.3179	-0.0286	0.2977	0.0571	0.2433
	$\hat{\phi}_1$	-0.1088	0.2023	-0.0626	0.1149	-0.1471	0.1747
	$\hat{\phi}_2$	-0.0256	0.0619	-0.0136	0.0426	-0.0148	0.0295
	$\hat{\rho}$	-0.0708	0.3067	-0.0296	0.1667	0.0452	0.0891
100	$\hat{\beta}_0$	0.0322	0.1465	0.0244	0.1343	-0.0021	0.1270
	$\hat{\beta}_1$	-0.0393	0.2508	-0.0282	0.2164	0.0056	0.2071
	$\hat{\phi}_1$	-0.0871	0.1649	-0.0401	0.0749	-0.0359	0.0684
	$\hat{\phi}_2$	-0.0149	0.0435	-0.0117	0.0317	-0.0112	0.0277
	$\hat{\rho}$	-0.0307	0.2221	-0.0122	0.1200	-0.0234	0.0918
150	$\hat{\beta}_0$	0.0191	0.1169	0.0185	0.1025	0.0274	0.1043
	$\hat{\beta}_1$	-0.0231	0.1938	-0.0226	0.1743	-0.0360	0.1718
	$\hat{\phi}_1$	-0.0680	0.1443	-0.0342	0.0631	-0.0322	0.0554
	$\hat{\phi}_2$	-0.0157	0.0371	-0.0107	0.0273	-0.0102	0.0241
	$\hat{\rho}$	-0.0173	0.2006	-0.0106	0.1073	0.0019	0.0673

In both cases, the variance is the most important part of the MSE.

4.5.2 Simulation 2

The second simulation analyses the performance of the EBP and plug-in predictors. For each d and t, the EBPs are approximated by generating $S1 = S2 = 500$ random variables $v_{1,d}^{(s_1)}$ and $v_{2,dt}^{(s_2)}$ in Step 2 of the algorithms given in Sections 4.3.1 and 4.3.2 respectively. For the two predictors (EBP and plug-in) and models (Model T1 and Model T2), Tables 4.5.3 and 4.5.4 present the average across domains and time periods of biases and RMSEs, i.e.

$$B = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} B_{dt}, \quad RE = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} RE_{dt},$$

where

$$B_{dt} = \frac{1}{K}\sum_{k=1}^{K}(\hat{p}_{dt}^{(k)} - p_{dt}^{(k)}),\ RE_{dt} = \left(\frac{1}{K}\sum_{k=1}^{K}(\hat{p}_{dt}^{(k)} - p_{dt}^{(k)})^2\right)^{1/2},\ d = 1,\dots,D,\ t = 1,\dots,T.$$

Table 4.5.3 presents the simulation results obtained for $D = 50, 100, 150$ and $T = 5$, since $T = 5$ is close to the number of time instants ($T = 4$) of the first application to real data presented in Section 4.6.1. Table 4.5.4 takes $D = 100$ and $T = 5, 9, 12$. The tables are divided in two parts. The first part, containing the first two rows, gives the numerical results of simulations under Model T1. The second part, containing the second two rows, gives the numerical results of simulations under Model T2.

Table 4.5.3: Bias (B) and RMSE (RE) of EBPs and plug-in predictors for $T = 5$ (both $\times 10^3$).

		$D = 50$		$D = 100$		$D = 150$	
Model	Predictor	B	RE	B	RE	B	RE
T1	EBP	0.0909	27.4315	0.0731	27.3042	0.0922	27.3539
	Plug-in	0.1866	38.8997	0.3698	39.8498	0.4268	40.3961
T2	EBP	-0.8154	30.1909	-0.4611	29.1808	-0.2707	28.8227
	Plug-in	-3.6317	33.8777	-3.2122	33.5982	-2.8751	33.6778

Under the two models, the EBP performs clearly better than the plug-in predictor, as both B and RE are lower. In addition, the estimators are quite stables and they are not too affected by D or T. In the case of Model T2 with time dependency, the differences in bias

between both estimators are even bigger. The main conclusions of Simulation 2 are that the EBPs are preferable to the plug-in predictors and also that variance is the main part of the MSE since bias is much smaller than RMSE.

Table 4.5.4: Bias (B) and RMSE (RE) of EBPs and plug-in predictors for $D = 100$ (both $\times 10^3$).

		$T = 5$		$T = 9$		$T = 12$	
Model	Predictor	B	RE	B	RE	B	RE
T1	EBP	0.0731	27.3042	-0.0621	27.0282	-0.0612	27.0727
	Plug-in	0.3698	39.8498	1.2672	41.7297	1.3923	40.1531
T2	EBP	-0.4611	29.1808	-0.1696	31.8237	-0.5151	31.4158
	Plug-in	-3.2122	33.5982	-1.5382	40.6226	-2.3145	37.8867

The simulation programmes solve the systems of MM nonlinear equations (4.2.4) for Model T1 and (4.2.4) and (4.2.12) for Model T2 by using *nleqslv* package of R. In addition, they also use the *mvtnorm* package to generate samples of a multivariate normal distribution. The computational burden of the MM algorithm is much higher for the correlated model. Taking $D = 100$, the runtimes of the MM fitting algorithm for $T = 5, 9, 12$ under Model T1 are $0.06, 0.07, 0.08$ seconds respectively and $2.75, 4.59, 6.73$ under Model T2. Regarding the p_{dt} estimators, the computational difference between the EBP and the plug-in has been increased compared to the previous chapters. The reason is that now the plug-in predictor requires the calculation of two EBPs ($\hat{\boldsymbol{v}}_1$ and $\hat{\boldsymbol{v}}_2$). The runtimes of the plug-in predictor for $D = 100$ and $T = 5, 9, 12$ under Model T1 are $215, 375, 472$ seconds respectively, and $1433, 2292, 3160$ seconds under Model T2. Under the same specifications, the runtimes of the EBPs are $168, 332, 415$ seconds for Model T1 and $1058, 1890, 2704$ seconds for Model T2. These computational times reinforce the EBP as a good alternative.

4.6 Application to real data

4.6.1 Poverty data

This section presents an application to the estimation of poverty proportions, p_{dt}, in small areas and time instants. The data are taken annually from the 2010-2013 SLCS of Galicia, which is a Spanish autonomous community with 53 counties. Therefore, the number of time periods in the application is $T = 4$. The target domains are the counties crossed by

sex and, in consequence, the number of total domains is $D = 106$.

At domain d and time period t, the response variable y_{dt} is the number of sampled people under the poverty line. We assume that y_{dt} can be explained by some auxiliary variables through an area-level Poisson mixed model with time effects.

Table 4.6.1 presents the estimates of the fixed effect coefficients by using the MM algorithm, for Model T2 (with AR(1)-correlated time effects) on the left and for Model $T2_2$ (the submodel with only the area-time random effect $v_{2,dt}$) on the right. We select a significant set of auxiliary variables with p-values lower than 0.05. For the sake of comparisons, we take the same covariates in both models. By analyzing the signs of the model coefficients, we conclude that *lab2* (proportion of unemployed people) and *age4* (proportion of people over 64 years) contribute to increase the poverty, while *edu23* (proportion of people with high level of education) is protective.

Table 4.6.1: MM estimates of regression parameters under Model T2 (left) and Model $T2_2$ (right).

	Model T2				Model $T2_2$			
Variable	Est.	s.e.	z-value	$P(>\|z\|)$	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	-1.615	0.015	-107.684	< 0.001	-1.586	0.138	-11.523	< 0.001
lab2	4.933	0.033	151.368	< 0.001	4.696	0.361	13.015	< 0.001
edu23	-1.328	0.014	-91.725	< 0.001	-1.663	0.145	-11.502	< 0.001
age4	0.731	0.015	48.543	< 0.001	1.485	0.178	8.339	< 0.001

The variance component estimates of Model T2 are $\hat{\phi}_1 = 0.0255$ and $\hat{\phi}_2 = 0.0396$. Their 95% percentile bootstrap confidence interval are $[0, 0.3443)$ and $(0.0347, 0.2196)$, respectively. The estimated correlation parameter is $\hat{\rho} = -0.9770$ and its 95% percentile bootstrap confidence interval is $(-0.9970, -0.1443)$. For studying the significance of the variance parameter ϕ_1, the hypothesis $H_0 : \phi_1 = 0$ is tested by adapting Algorithm 5 (see Section 2.9) to Model T2. As the obtained p-value is 0.758, the random effects related to the domains are not significant. Therefore, the simplified version of Model T2, Model $T2_2$, with only $v_{2,dt}$ random effects is considered. The estimate of the variance component of Model $T2_2$ is $\hat{\phi}_2 = 0.1007$ and its 95% percentile bootstrap confidence interval is $(0.0240, 0.1987)$. The estimated correlation parameter is $\hat{\rho} = -0.5199$ and its 95% percentile bootstrap confidence interval is $(-0.9225, 0.7131)$. To test the significance of the correlation parameter ρ under

Model $T2_2$, the following bootstrap procedure is proposed.

Algorithm 7 A bootstrap test for $H_0 : \rho = 0$

1: Fit the Model $T2_2$ to data and calculate $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_2$ and $\hat{\rho}$.

2: Fit the Model $T1_2$ to data and calculate $\hat{\boldsymbol{\beta}}^0$ and $\hat{\phi}_2^0$.

3: For $b = 1, \ldots, B$, do

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

$$v_{dt}^{*(b)} \sim N(0,1),\ p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}}^0 + \hat{\phi}_2^0 v_{dt}^{*(b)}\},$$
$$y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)}),\ d = 1, \ldots, D,\ t = 1, \ldots, T.$$

ii) Fit the Model $T2_2$ to the bootstrap data $(y_{dt}^{*(b)}, \boldsymbol{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, and calculate $\hat{\boldsymbol{\beta}}^{*(b)}$, $\hat{\phi}_2^{*(b)}$ and $\hat{\rho}^{*(b)}$.

4: Calculate the p-value

$$p = \frac{\#\left\{\left|\hat{\rho}^{*(b)}\right| > |\hat{\rho}|\right\}}{B}.$$

The bootstrap p-value is 0.414. This suggests that instead of using a complex model (Model T2 or Model $T2_2$), in this study it is more appropriated to consider a model with independent time effects (Model T1 or Model $T1_2$).

Table 4.6.2 shows the estimates of the fixed effect coefficients for Model T1 on the left and for Model $T1_2$ on the right. They can be interpreted analogously to those given in Table 4.6.1. On the other hand, the estimates of the variance components of Model T1 are $\hat{\phi}_1 = 0$ and $\hat{\phi}_2 = 0.1117$, and the 95% percentile bootstrap confidence interval for ϕ_2 is $(0.0583, 0.1590)$. As $\hat{\phi}_1 = 0$, simplified Model $T1_2$ is considered. Model $T1_2$ is defined by (4.2.1) and (4.2.11), and incorporates only the area-time random effect $v_{2,dt}$. The estimated variance parameter of Model $T1_2$ is $\hat{\phi}_2 = 0.180$ and its 95% percentile bootstrap confidence interval is $(0.044, 0.575)$. The bootstrap p-value for testing the hypothesis $H_0 : \phi_2 = 0$ (by using Algorithm 5 in Section 2.9) is 0.006. The conclusion is that the domain-time vector of random effects is significant at the level $1 - \alpha = 0.95$. Therefore, Model $T1_2$ is finally selected. Starting from Model T2 or from Model T1, the employed model selection procedure leads also to Model $T1_2$.

Table 4.6.2: MM estimates of regression parameters under Model T1 (left) and Model $T1_2$ (right).

	Model T1				Model $T1_2$			
Variable	Est.	s.e.	z-value	$P(>\|z\|)$	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	-1.614	0.172	-9.372	< 0.001	-1.617	0.186	-8.682	< 0.001
*lab*2	4.933	0.375	13.153	< 0.001	4.933	0.465	10.618	< 0.001
*edu*23	-1.328	0.176	-7.538	< 0.001	-1.328	0.185	-7.192	< 0.001
*age*4	0.731	0.183	3.988	< 0.001	0.731	0.216	3.388	< 0.001

For the sake of comparisons, we fit a fixed-effects Poisson regression model (Model T0, cf. (4.2.10)) to the set of auxiliary data appearing in Table 4.6.1 and 4.6.2. Figure 4.6.1 plots the Pearson residuals of Model T0 without random effects (left) and of Model $T1_2$ with independent random effects (right). A clear improvement is achieved when using the model with random time effects since its residuals are loser to 0.

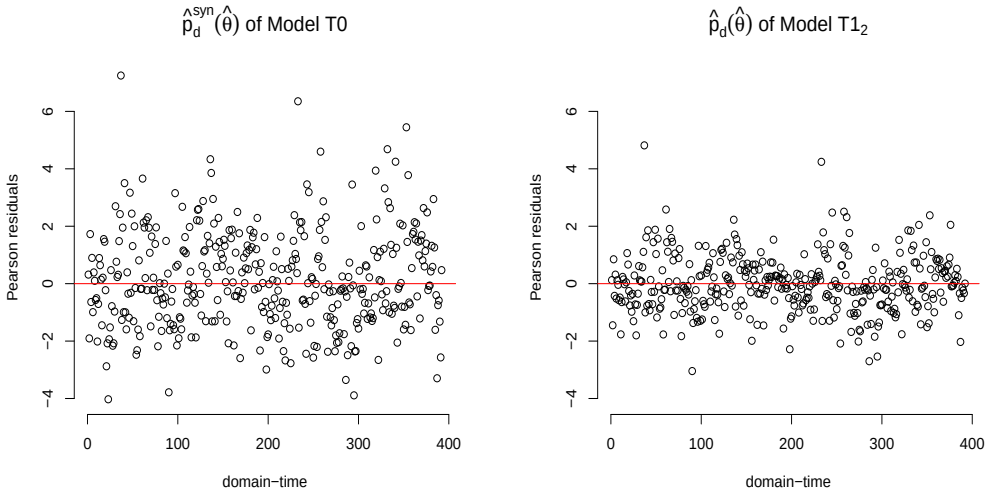

Figure 4.6.1: Pearson residuals of Model T0 (left) and Model $T1_2$ (right).

Figure 4.6.2 (left) plots the EBPs under Model $T1_2$ and the direct estimates (dir) of the p_{dt}'s sorted by the sample sizes ν_{dt}'s. Figure 4.6.2 (right) plots the relative root-MSEs of the EBPs and the relative root-variances of the direct estimators. For small sample sizes, ν_{dt}, the direct estimates have high variability while for large values of ν_{dt} they follow the same pattern as the EBPs. The relative root-MSEs of the EBPs are estimated by using a

parametric bootstrap based on the method given by González-Manteiga et al. (2007). The relative root-variances of the direct estimators are too high for small sample sizes and their accuracies improve when the sample size increases. The average across domains of relative root-variances of the direct estimators is 0.253 and the corresponding average of relative root-MSEs of the EBPs is 0.149. Their maximum values are 1.005 and 0.232, respectively. From these results, we conclude that the EBP performs better than the direct estimator.

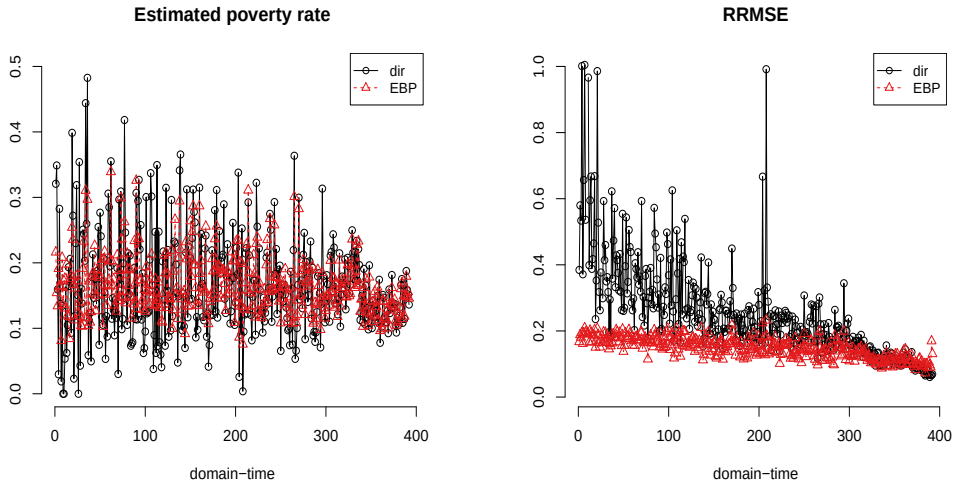

Figure 4.6.2: Direct estimates and EBPs of p_{dt} (left) and their relative root-MSEs (right).

Figure 4.6.3 maps the EBPs of p_{dt}, by county and sex for the period $2010 - 2013$, under Model $\mathrm{T1}_2$. Highest levels of poverty tend to be concentrated in the south and west of the region. We also observe a clear difference between genders, being the poverty proportion higher for women, 32 counties for men vs. 37 for women with $\hat{p}_{dt} > 0.15$ in 2013. Nevertheless, between 2010 and 2013, the poverty rate has increased in average by 1.7% for men and by 1% for women.

Figure 4.6.4 maps the bootstrap relative root-MSE (RRMSE) estimates of the EBP of p_{dt}, by county and sex for the period $2010 - 2013$, under Model $\mathrm{T1}_2$. We take $B = 500$ resamples. The average across domains of the estimated RRMSEs in 2013 is 0.147 for men and 0.148 for women. The maximum values are 0.209 and 0.210, for men and women respectively.

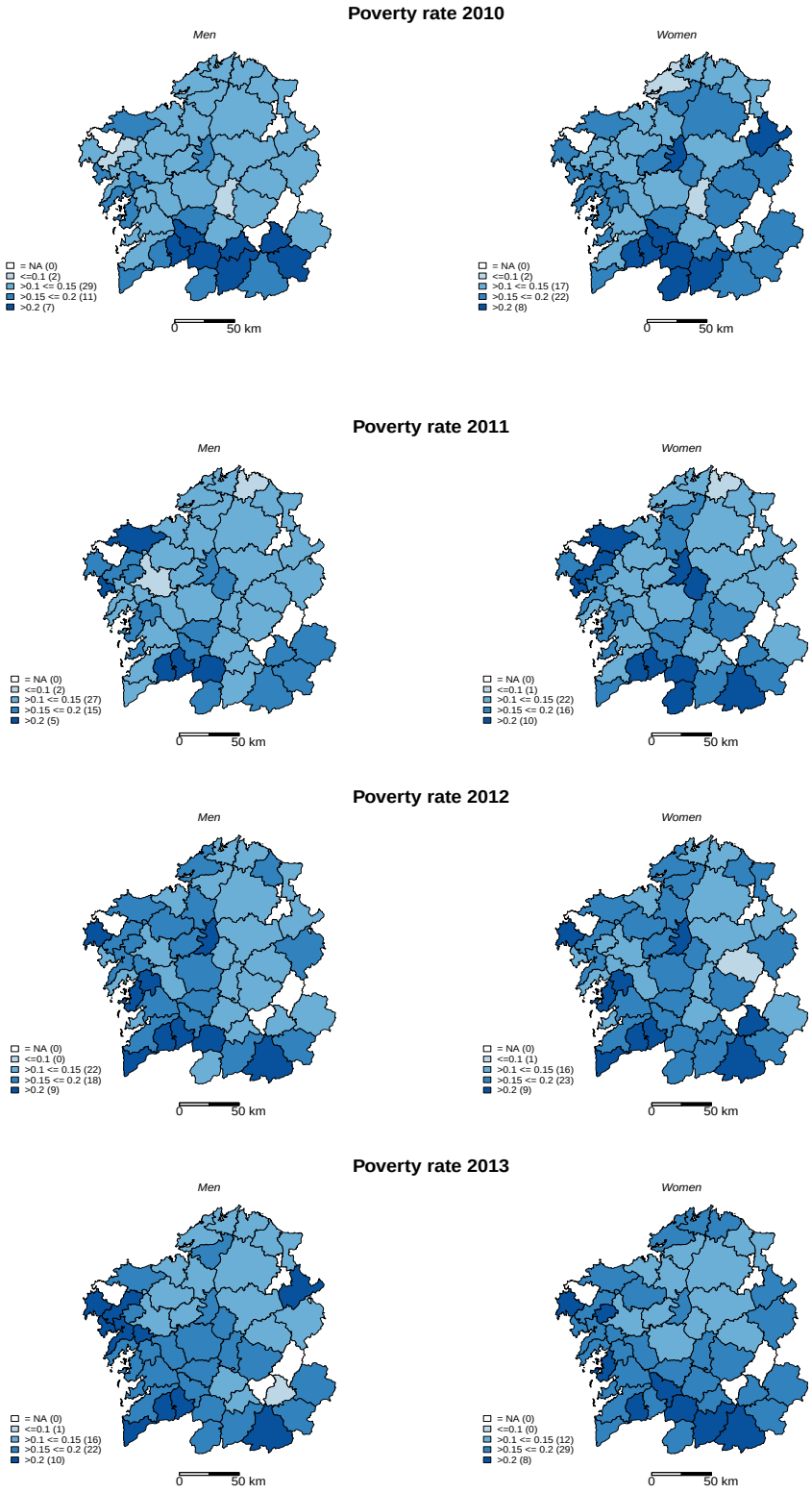

Figure 4.6.3: EBPs of poverty rates for men (left) and women (right) in 2010 – 2013.

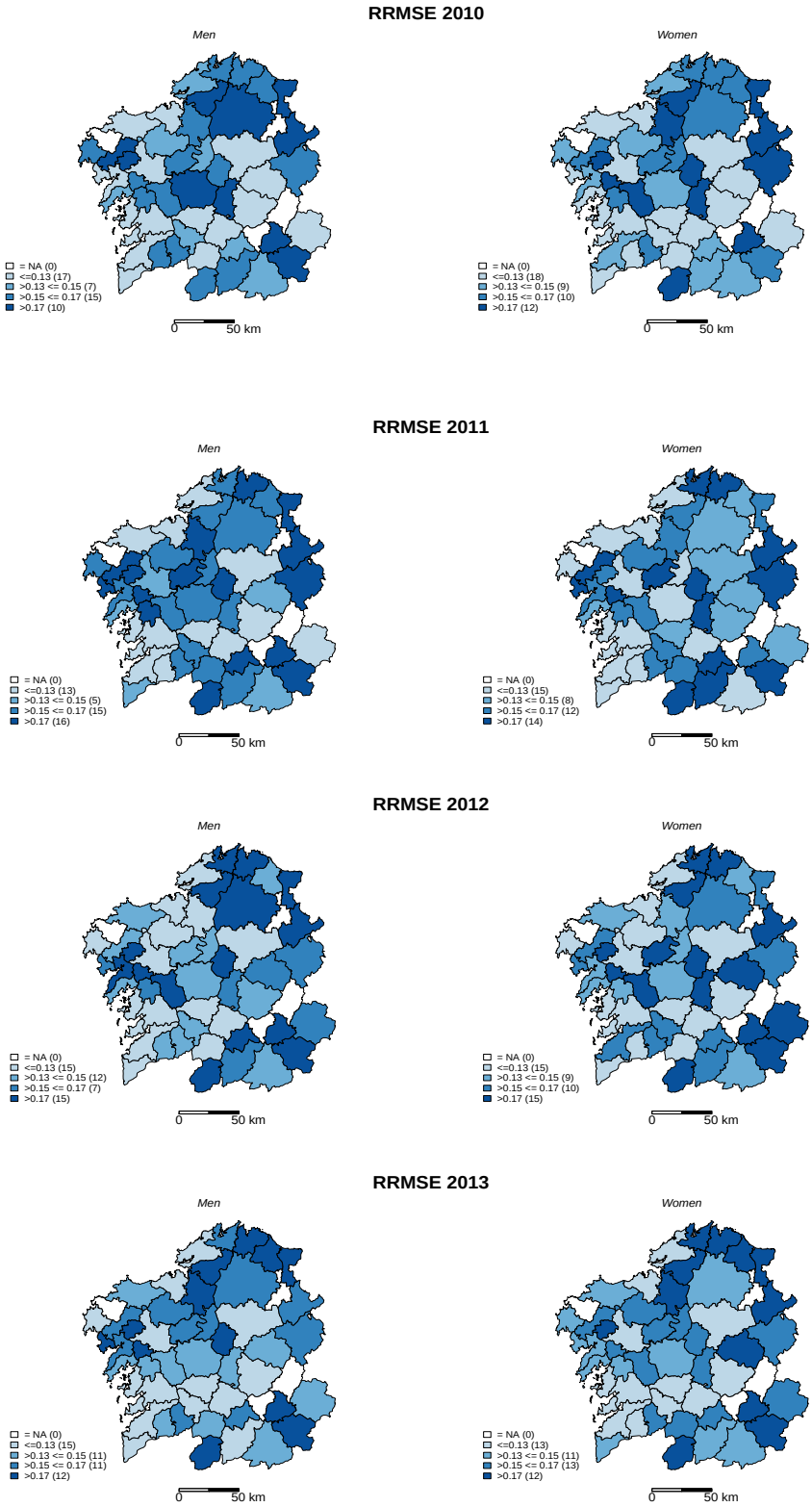

Figure 4.6.4: Bootstrap RRMSE estimates of the EBPs for men (left) and women (right) in 2010 – 2013.

4.6.2 Forest fires data

The modelling of the target variable *number of forest fires* is extremely valuable because it allows making preventive policies. This study concerns Galicia, a region in the northwest of Spain, because it has a high number of fires with devastating effects that occur systematically every summer (see Chapter 1 for more details). The interest is to know what happens by forest areas. Galicia is divided into $D = 63$ forest areas. For each domain, d, $d = 1, \ldots, D$, the evolution of the response variable is observed between the months of April to October from 2007 to 2008. The reason of this choice is because these months have the largest number of forest fires. Therefore, the number of time periods is $T = 14$.

The response variable at domain d and time t, y_{dt}, is explained by some auxiliary variables through an area-level Poisson model with time effects. Depending on their structure, two sources of auxiliary information are considered. The first source contains auxiliary variables that are constant over time. They are population size (*pop*) and cadastral holders (*cadHold*) per forest areas that only depends on the domains. The second source considers average measurements of meteorological stations per month and forest areas. It specifically examines accumulated rain (*acumRain*), average temperature (*averTemp*) and days without rain (*dwr*). See more details about the data set in Section 1.4.2.

Table 4.6.3 presents the significant MM estimates (p-value < 0.05) of the fixed effect coefficients for the two models with correlated time effects (Model T2 and Model T2_2). They are selected by using the AIC under the residuals of Model T0. In order to make a comparison under the same auxiliary information, the application selects the same set of covariates for both models. It takes the level $\alpha = 5\%$ for selecting the variables in the final model. Regression estimates suggest that *averTemp* and *cadHold* are directly related to the response variable given that an increase in those variables causes an increase in the response variable if *acumRain* remains fixed. By contrast, the relationship between *acumRain* and y_{dt} is inverse since an increase in this variable causes a decrease in the target variable. The variance parameter estimates of Model T2 are $\hat{\phi}_1 = 0.005$ and $\hat{\phi}_2 = 0.626$. Their 95% percentile bootstrap confidence intervals are $[0, 0.374)$ and $(0.503, 0.829)$, respectively. See Shao and Tu (1995) for the mathematical details on the construction of this bootstrap confidence intervals. The estimated correlation parameter is $\hat{\rho} = 0.643$ and its 95% percentile bootstrap confidence interval is $(0.400, 0.760)$. The variance parameter ϕ_1 in Model T2 is not significant since the bootstrap p-value of the hypothesis test $H_0 : \phi_1 = 0$ is 0.918. Therefore, the application considers the simplified version of Model T2 with only domain-time effects,

Model $T2_2$.

Table 4.6.3: MM estimates of regression parameters under Model T2 (left) and Model $T2_2$ (right).

	Model T2				Model $T2_2$			
Variable	Est.	s.e.	z-value	$P(>\|z\|)$	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	0.7650	0.1092	7.0038	< 0.001	0.7739	0.0753	10.2717	< 0.001
acumRain	-0.5207	0.0673	-7.7406	< 0.001	-0.5207	0.0627	-8.3073	< 0.001
averTemp	0.1505	0.0633	2.3769	0.0175	0.1505	0.0761	1.9781	0.0479
cadHold	0.2930	0.0670	4.3727	< 0.001	0.2930	0.0680	4.3065	< 0.001

Table 4.6.3 (right) presents the fixed effect estimates for this model. They can be interpreted analogously to Model T2. The estimate of the variance parameter is $\hat{\phi}_2 = 0.5981$ and its 95% bootstrap confidence interval is $(0.4484, 0.7056)$. The estimated correlation parameter is 0.6387 and its 95% bootstrap confidence interval is $(0.4241, 0.7960)$. Both parameters, ϕ_2 and ρ, are significant since the bootstrap p-values of the respective hypothesis tests $H_0 : \phi_2 = 0$ and $H_0 : \rho = 0$ are 0. Therefore, taken into account such results, the conclusion is that Model $T2_2$ is appropriated for modelling the Galician forest fires since all its components are significant.

As the proposed plug-in estimator is easy to interpret and its performance is similar to that of the EBP when T is high, this application to real data illustrates its behaviour. Figure 4.6.5 plots the Pearson residuals of the model with only fixed effects, Model T0, on the left and of the plug-in estimator, $\hat{\mu}_{dt}^P$, under Model $T2_2$ on the right. The Pearson residuals of the model with random time effects show a clear improvement since its residuals are closer to 0.

Figure 4.6.6 maps the obtained plug-in estimates by areas using the simplified area-level Poisson mixed model with AR(1)-correlated time effects, Model $T2_2$. As time instants, the application takes the months between August and October for being the month more dangerous for fires. Figure 4.6.6 suggests that forest fires tend to be concentrated in coastal areas and in the south of the region. In addition, there is a sharp decrease in the number of fires estimated in 2008.

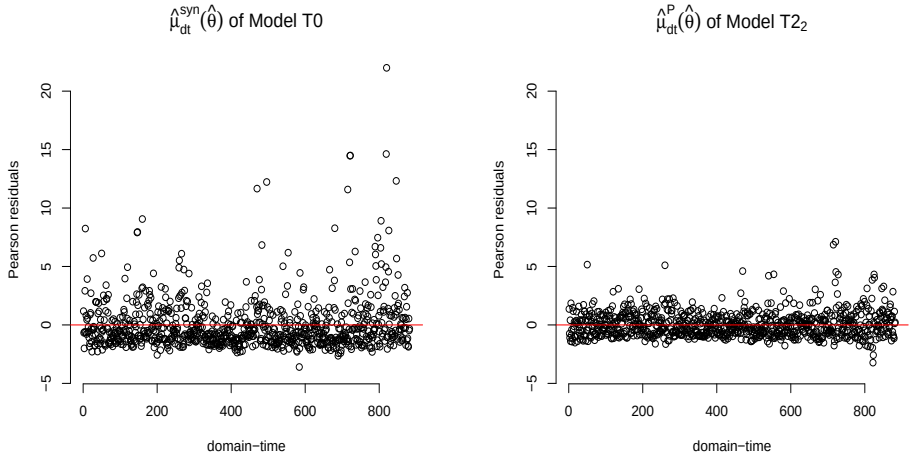

Figure 4.6.5: Pearson residuals of Model T0 with only fixed effects (left) and Model $T2_2$ (right).

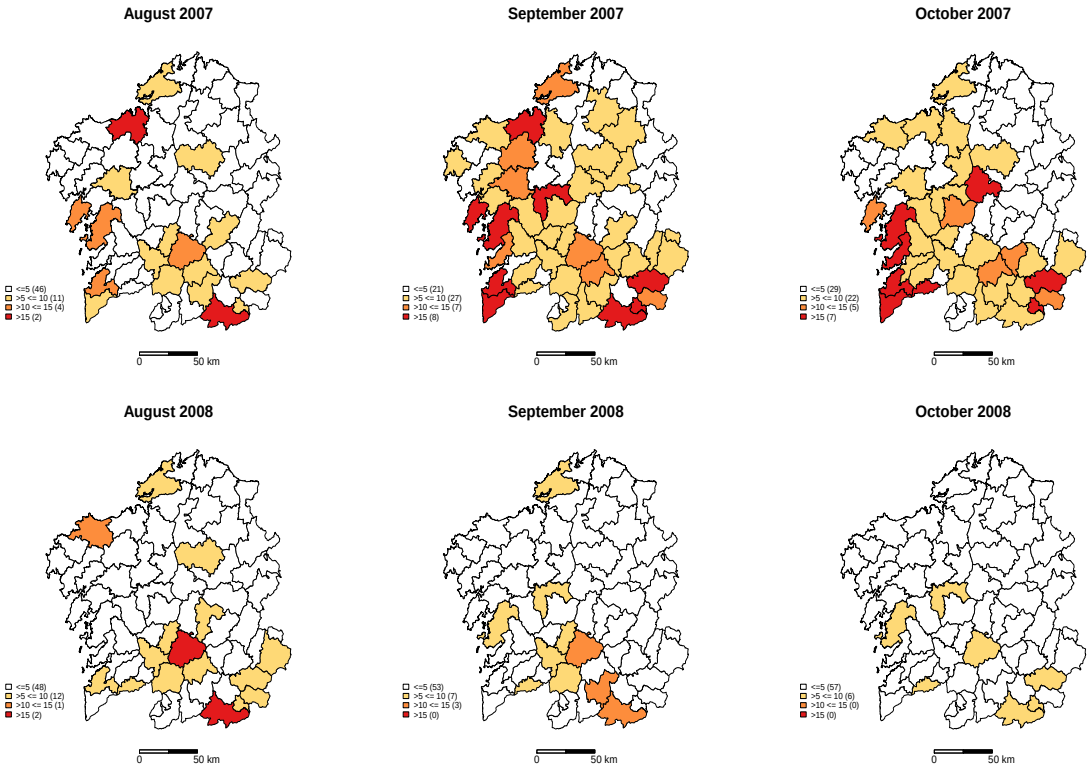

Figure 4.6.6: Estimated fires from August to October in 2007-2008.

Figure 4.6.7 plots the evolution of the MSEs, defined in (4.4.1) and adapted to the plug-in

estimator, over time for Model T0 and Model $T2_2$. The MSE is represented for the three areas with highest number of fires: Viana 1 (total fires 311), Terra de Tribes (total fires 329) and Viana 2 (total fires 347). We take $B = 500$ bootstrap resamples. The mean of the MSEs for the three areas is 8.035 in Model T0 and 2.413 in Model $T2_2$. A clear increase in accuracy is achieved when one uses the Model $T2_2$ since its MSE is much lower.

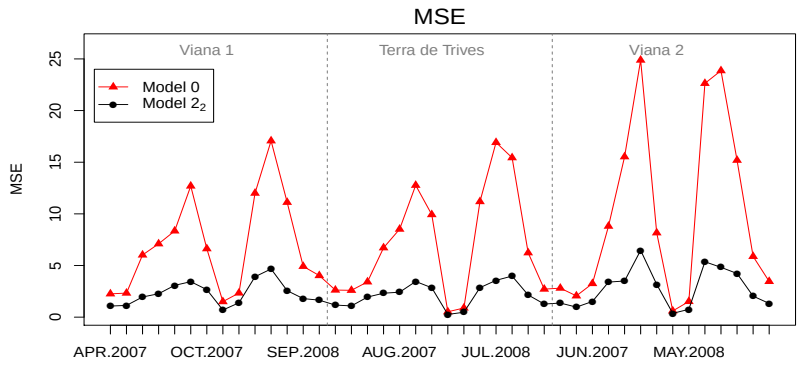

Figure 4.6.7: Bootstrap MSE estimates for the three areas with highest number of fires.

4.7 Concluding remarks

This chapter introduces four area-level Poisson mixed models with time random effects. Model T1 and Model $T1_2$ use independent time effects and Model T2 and Model $T2_2$ assume that they are AR(1)-correlated. The MM is employed for estimating the model parameters. Model-based empirical best and plug-in predictors of domain-time counts or proportions are proposed and their MSEs are estimated by a parametric bootstrap.

More complex correlation structures might also be considered for dealing with survey samples in small area estimation problems. However, most official surveys have long time periods between repetitions and sometimes their methodologies change after some few periods. In particular, the SLCS is an annual survey. This is why the number of time periods T is usually small in survey sampling problems (in our case, $T = 4$) and therefore there is no place for complex correlation structures. Based on the papers of Esteban et al. (2012a,b), Marhuenda et al. (2013, 2014) and Morales et al. (2015), deriving poverty proportion estimators for the SLCS based on area-level temporal linear mixed models, we select the AR(1) correlation structure, instead of MA(1), as an alternative to the hypothesis of independence of the time random effects.

The MM algorithm for fitting GLMMs is easy to implement. It was proposed by Jiang (1998) and later applied and studied by Jiang and Lahiri (2001) and Jiang (2003). These authors showed that the MM estimators of GLMM parameters are competitive with respect to the estimators maximizing the Laplace approximation of the log-likelihood. Simulation 1 empirically investigates the behaviour of the MM estimators and shows that their biases and mean squared errors decreases as D or T increases. Based on the simulation results and the theoretical properties and simulations given by the above cited authors, the conclusion is that the MM algorithm is a good alternative for fitting the introduced area-level temporal Poisson mixed models.

Under the considered models, the simulation experiments suggest that the EBP is a good alternative since its bias and mean squared error are generally lower than the corresponding ones of the plug-in predictor. The simulation experiment shows that Model T2 needs a sufficiently large number of time instants ($T \geq 5$) to capture the correlation structure. Given that the MM only provides estimates for the fixed effects and variance parameters, the calculation of the plug-in predictor requires EBP predictors of the random effects. This fact increases the computational burden of the plug-in predictor regarding the EBP.

In the application to 2010-2013 SLCS data, the EBP approach is used for estimating poverty proportions of Galician counties by sex (see Boubeta et al. (2017b) for more details). The application compares the behaviour of the EBP under the area-level Poisson mixed models with independent time effects and the synthetic estimator under Model T0. The model with independent time effects has a better performance. The application finally selects the simpler Model T1_2 with only independent time-domain random effects. From the data analysis, the conclusion is that the poverty proportions increased in Galicia during the period $2010 - 2013$. Highest levels of poverty are found in the west and south of the autonomous community. Additionally, the poverty proportion is in general higher for women. The average across domains of the bootstrap RRMSEs of the given poverty estimates is around 14% in both sexes during 2013.

Regarding the application to forest fires in Galicia, the developed methodology is applied for estimating the number of forest fires by area and month during $2007-2008$ (Boubeta et al., 2017a). The performance of the plug-in estimators in the area-level Poisson mixed models with time effects is studied and compared against the corresponding estimator obtained from the fixed effects model. A clear improvement is achieved when one uses a mixed model. A temporal correlation structure is uncovered by the auxiliary data and therefore Model T2 or Model T2_2 are more appropriate in this context. As the domain effects are

not significant, it is recommended to use the simplified version, i.e. the Model $T2_2$. From the data analysis, the conclusion is that the forest fires tend to be concentrated in coastal areas and in the south of the region. As accuracy measure of the proposed estimator, a bootstrap MSE based on a parametric bootstrap is considered. A clear improvement is observed when using the proposed model against the traditional Poisson model. So this methodology can be an important tool to make a preventive policy in the context of forest fires.

Chapter 5

The area-level Poisson mixed model with SAR(1) and time effects

Contents

5.1 Introduction

Area-level Poisson mixed models are good tools for modelling count data at area level. However, the basic area-level Poisson mixed model (Boubeta et al., 2016b) has several limitations. It does not take into account temporal components or complex spatial structures. This manuscript introduces several extensions of the basic area-level Poisson mixed model for better fitting the needs of real data, given rise to increasingly complex models. Chapter 3 generalizes the basic Poisson mixed model by incorporating a SAR(1) spatial structure.

Chapter 4 gives several extensions to the temporal framework. This chapter incorporates both extensions in a single model and introduces an area-level spatio-temporal Poisson mixed model. The new model can explain both sources of variability and correlation structures.

A spatio-temporal extension of the Fay-Herriot model was proposed by Singh et al. (2005) using the Kalman filtering approach. Under this model, they obtain a second order approximation to the MSE of the EBLUP. Later, Pereira and Coelho (2012) and Marhuenda et al. (2013) derive empirical best linear unbiased predictors under the above model. Specifically, they consider a SAR(1) spatial correlation structure and an AR(1) process for the temporal component. They also propose bootstrap procedures for estimating the mean squared error and they analyse the behaviour of the proposed model against other simpler models through several simulation experiments. Esteban et al. (2016) present a new spatio-temporal model by assuming MA(1)-correlated random effects.

In the Bayesian framework, Choi et al. (2011) examine several spatio-temporal mixed models in small area health data applications and develop new accuracy measures to assess the recovery of true relative risks. They apply the spatio temporal models to study chronic obstructive pulmonary disease at county level in Georgia.

All the cited authors apply spatio-temporal LMMs to the small area estimation setup. However, this manuscript deals with GLMMs instead of with LMM. Chapter 5 introduces and study the applicability of EBPs, based on a spatio-temporal area-level Poisson mixed model, to the estimation of domain counts and proportions.

Chapter 5 is organized as follows. Section 5.2 introduces the area-level Poisson mixed model with SAR(1) and independent time effects. The MM fitting algorithm is obtained in Section 5.3. Three bootstrap algorithms are presented in Section 5.4 to test the significance of the variance and autocorrelation parameters. Section 5.5 gives the empirical best predictor of the Poisson parameter and of the domain and domain-time random effects. Section 5.6 proposes a parametric bootstrap procedure as accuracy measure of the EBP. Section 5.7 investigates the behaviour of the proposed fitting algorithm and empirically compares the performance of the plug-in and EBP by means of simulation experiments. Section 5.8 illustrates the developed methodology in a environmental field. Finally, Section 5.9 collects the main conclusions of Chapter 5.

5.2 The model

This section extends the area-level Poisson mixed model (Model 1), introduced in Chapter 2 (Boubeta et al., 2016b), to the spatio-temporal context. In particular, the section introduces a model with SAR(1)-correlated domain random effects and with independent domain-time random effects (Model ST1). It can also be seen as the generalization of Model S1 (see Chapter 3) and Model T1 (see Chapter 4) in a single model. The total number of domains and time instants are denoted by D and T respectively, and the corresponding indices by d, $d = 1, \ldots, D$, and t, $t = 1, \ldots, T$. Model ST1 considers two independent sets of random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, such that $\boldsymbol{v}_1 = \{v_{1,d} : \ d = 1, \ldots, D\}$ are the domain effects and $\boldsymbol{v}_2 = \{v_{2,dt} : \ d = 1, \ldots, D, \ t = 1, \ldots, T\}$ are the domain-time effects.

The model assumes that the vector of domain random effects $\boldsymbol{v}_1$ is spatially correlated, following a SAR(1) process with unknown autoregression parameter ρ and known proximity matrix $\boldsymbol{W}$, i.e.

$$\boldsymbol{v}_1 = \rho \boldsymbol{W} \boldsymbol{v}_1 + \boldsymbol{u}_1, \tag{5.2.1}$$

where $\boldsymbol{u}_1 \sim N_D(\boldsymbol{0}, \boldsymbol{I}_D)$, $\boldsymbol{0}$ is the $D \times 1$ zero vector and $\boldsymbol{I}_D$ denotes the $D \times D$ identity matrix. It also assumes that the matrix $(\boldsymbol{I}_D - \rho \boldsymbol{W})$ is non-singular. Then, $\boldsymbol{v}_1$ can be expressed as

$$\boldsymbol{v}_1 = (\boldsymbol{I}_D - \rho \boldsymbol{W})^{-1} \boldsymbol{u}_1. \tag{5.2.2}$$

The proximity matrix $\boldsymbol{W}$ is obtained as it was explained in Section 3.2. Then, the autoregression parameter ρ is a correlation, $\rho \in (-1, 1)$, and is called spatial autocorrelation parameter. Equation (5.2.2) implies that $\boldsymbol{v}_1 = \underset{1 \leq d \leq D}{\text{col}}(v_{1,d}) \sim N_D(\boldsymbol{0}, \boldsymbol{\Gamma}(\rho))$, where $\boldsymbol{\Gamma}(\rho)$ is given in (3.2.3). Therefore, the density function of the domain random effects $\boldsymbol{v}_1$ is

$$f_v(\boldsymbol{v}_1) = (2\pi)^{-D/2} |\boldsymbol{\Gamma}(\rho)|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{v}_1' \boldsymbol{\Gamma}^{-1}(\rho) \boldsymbol{v}_1 \right\}.$$

Further, it holds that $v_{1,d} \sim N\big(0, \gamma_{dd}(\rho)\big)$ and $v_{1,d_2} | v_{1,d_1} \sim N\big(\mu_{d_2|d_1}, \sigma^2_{d_2|d_1}\big)$, where

$$\mu_{d_2|d_1} = \frac{\gamma_{d_1 d_2}(\rho)}{\gamma_{d_1 d_1}(\rho)} v_{d_1}, \quad \sigma^2_{d_2|d_1} = \gamma_{d_2 d_2}(\rho) - \frac{\gamma^2_{d_1 d_2}(\rho)}{\gamma_{d_1 d_1}(\rho)}.$$

The interaction domain-time random effects, $\boldsymbol{v}_2$, are assumed to be independent over time, i.e.

$$\boldsymbol{v}_{2d} = \underset{1 \leq t \leq T}{\text{col}}(v_{2,dt}) \sim N(\boldsymbol{0}, \boldsymbol{I}_T), \quad \boldsymbol{v}_2 = \underset{1 \leq d \leq D}{\text{col}}(\boldsymbol{v}_{2d}) \sim N(\boldsymbol{0}, \boldsymbol{I}_{DT}).$$

Then, the join density function of the random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ is

$$f_v(\boldsymbol{v}_1, \boldsymbol{v}_2) = (2\pi)^{-D(T+1)/2} |\boldsymbol{\Gamma}(\rho)|^{-1/2} \exp\left\{ -\frac{1}{2}\, \boldsymbol{v}_1' \boldsymbol{\Gamma}^{-1}(\rho) \boldsymbol{v}_1 - \frac{1}{2}\, \boldsymbol{v}_2' \boldsymbol{v}_2 \right\}.$$

The distribution of the target variable y_{dt}, conditionally on the random effects $v_{1,d}$ and $v_{2,dt}$, is

$$y_{dt}|v_{1,d}, v_{2,dt} \sim \text{Poisson}(\mu_{dt}), \quad d = 1, \ldots, D,\ t = 1, \ldots, T,$$

where μ_{dt} denotes the mean of the Poisson distribution. Given the relationship between Poisson and binomial distributions, the Poisson parameter is expressed as $\nu_{dt} p_{dt}$, where ν_{dt} and p_{dt} are size and probability parameters respectively. The natural parameter, $log(\mu_{dt})$, is expressed in terms of a set of p auxiliary variables $\boldsymbol{x}_{dt}$ by a regression model, i.e.

$$\begin{aligned} \text{Model ST1:}\ \log \mu_{dt} &= \log \nu_{dt} + \log p_{dt} \\ &= \log \nu_{dt} + \boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}, \quad d = 1, \ldots, D,\ t = 1, \ldots, T, \end{aligned}$$

where $\mu_{dt} = \mathbb{E}[y_{dt}|v_{1,d}, v_{2,dt}]$, $\boldsymbol{x}_{dt} = \underset{1 \le k \le p}{\text{col}'}(x_{dtk})$ is the row vector of auxiliary variables, $\boldsymbol{\beta} = \underset{1 \le k \le p}{\text{col}}(\beta_k)$ is the vector of regression coefficients and ϕ_1 and ϕ_2 are the variance parameters. Defining $u_{1,d} = \phi_1 v_{1,d}$ and $u_{2,dt} = \phi_2 v_{2,dt}$, then ϕ_1 and ϕ_2 are the variances of $u_{1,d}$ and $u_{2,dt}$ respectively. These variances can be interpreted as the variability between domains and between time periods within each domain respectively.

Further, Model ST1 assumes that the y_{dt}'s are independent conditionally on the random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$. It holds that

$$\mathbb{P}(y_{dt}|\boldsymbol{v}) = \mathbb{P}(y_{dt}|v_{dt}) = \frac{1}{y_{dt}!} \exp\{-\nu_{dt} p_{dt}\} \nu_{dt}^{y_{dt}} p_{dt}^{y_{dt}},$$

where $p_{dt} = \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}$ represents the target parameter. The probability function of the response variable $\boldsymbol{y} = \{y_{dt}, d = 1, \ldots, D,\ t = 1, \ldots, T\}$ conditionally on the random effects $\boldsymbol{v} = (\boldsymbol{v}_1, \boldsymbol{v}_2)$ is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) = \prod_{d=1}^{D} \prod_{t=1}^{T} \mathbb{P}(y_{dt}|\boldsymbol{v}).$$

As a consequence, the probability function of the response variable $\boldsymbol{y}$ is

$$\mathbb{P}(\boldsymbol{y}) = \int_{\mathbb{R}^{D(T+1)}} \mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) f_v(\boldsymbol{v}_1, \boldsymbol{v}_2)\, d\boldsymbol{v}_1 d\boldsymbol{v}_2 = \int_{\mathbb{R}^{D(T+1)}} \psi(\boldsymbol{y}, \boldsymbol{v})\, d\boldsymbol{v},$$

where

$$\begin{aligned}\psi(\boldsymbol{y},\boldsymbol{v}) &= f_v(\boldsymbol{v}_1,\boldsymbol{v}_2)\prod_{d=1}^{D}\prod_{t=1}^{T}\frac{\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}}}{y_{dt}!}\\ &= c(\boldsymbol{y})|\boldsymbol{\Gamma}(\rho)|^{-1/2}\exp\left\{-\frac{1}{2}\,\boldsymbol{v}_1'\boldsymbol{\Gamma}^{-1}(\rho)\boldsymbol{v}_1-\frac{1}{2}\,\boldsymbol{v}_2'\boldsymbol{v}_2\right\}\\ &\cdot\exp\left\{-\sum_{d=1}^{D}\sum_{t=1}^{T}\nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\right\}\\ &\cdot\exp\left\{\sum_{k=1}^{p}\Big(\sum_{d=1}^{D}\sum_{t=1}^{T}y_{dt}x_{dtk}\Big)\beta_k+\phi_1\sum_{d=1}^{D}y_{d.}v_{1,d}+\phi_2\sum_{d=1}^{D}\sum_{t=1}^{T}y_{dt}v_{2,dt}\right\},\end{aligned}$$

$c(\boldsymbol{y}) = (2\pi)^{-\frac{D(T+1)}{2}}\prod_{d=1}^{D}\prod_{t=1}^{T}\left(\nu_{dt}^{y_{dt}}/y_{dt}!\right)$ and $y_{d.} = \sum_{t=1}^{T}y_{dt}$.

5.3 The MM algorithm

This section derives the MM algorithm, based on the method of simulated moments proposed by Jiang (1998), to fit the area-level Poisson mixed model with SAR(1) domain and independent time effects. A set of natural equations for applying the MM algorithm is

$$\begin{aligned}0 = f_k(\boldsymbol{\theta}) &= \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]x_{dtk}-\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}y_{dt}x_{dtk},\quad k=1,\ldots,p,\\ 0 = f_{p+1}(\boldsymbol{\theta}) &= \frac{1}{D}\sum_{d=1}^{D}\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]-\frac{1}{D}\sum_{d=1}^{D}y_{d.}^2,\\ 0 = f_{p+2}(\boldsymbol{\theta}) &= \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]-\frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}y_{dt}^2,\\ 0 = f_{p+3}(\boldsymbol{\theta}) &= \frac{1}{D(D-1)}\sum_{d_1\neq d_2}^{D}\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1.}y_{d_2.}]-\frac{1}{D(D-1)}\sum_{d_1\neq d_2}^{D}y_{d_1.}y_{d_2.},\end{aligned}\tag{5.3.1}$$

where the vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta},\phi_1,\phi_2,\rho)$. The MM estimator of $\boldsymbol{\theta}$ is obtained by solving the system (5.3.1) of nonlinear equations. The updating formula of the Newton-Raphson algorithm is (2.3.3), where

$$\boldsymbol{\theta} = \underset{1\leq k\leq p+3}{\text{col}}(\theta_k),\quad \boldsymbol{f}(\boldsymbol{\theta}) = \underset{1\leq k\leq p+3}{\text{col}}(f_k(\boldsymbol{\theta})),\quad \boldsymbol{H}(\boldsymbol{\theta}) = \left(\frac{\partial f_k(\boldsymbol{\theta})}{\partial\theta_\ell}\right)_{k,\ell=1,\ldots,p+3}.\tag{5.3.2}$$

In what follows, the expectations appearing in $\boldsymbol{f}(\boldsymbol{\theta})$ and its partial derivatives under Model ST1 are calculated. For ease of exposition, the elements of $\boldsymbol{\Gamma}$ and its derivatives are denoted

by $\gamma_{d_1d_2} = \gamma_{d_1d_2}(\rho)$ and $\dot{\gamma}_{d_1d_2} = \dot{\gamma}_{d_1d_2}(\rho)$, respectively. The calculations start with the first p MM equations. The expectation of y_{dt} is

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}] &= \mathbb{E}_v\big[E_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}]\big] = \mathbb{E}_v[\nu_{dt}p_{dt}] = \mathbb{E}_v\left[\nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}\right] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}\; f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt} \\
&= \int_{-\infty}^{\infty} \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \phi_1 v_{1,d}\big\} f_v(v_{1,d})\,dv_{1,d} \\
&= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\big(\phi_1^2\gamma_{dd} + \phi_2^2\big)\big\}.
\end{aligned}$$

Therefore, the first p MM equations are

$$\begin{aligned}
f_k(\boldsymbol{\theta}) = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}\big(\phi_1^2\gamma_{dd} + \phi_2^2\big)\big\}x_{dtk} \\
- \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}x_{dtk},\;\; k = 1,\ldots,p.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\beta_k} &= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\big\}x_{dtk}, \\
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\phi_1} &= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\big\}\phi_1\gamma_{dd}, \\
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\phi_2} &= \nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\big\}\phi_2, \\
\frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial\rho} &= \frac{1}{2}\nu_{dt}\exp\big\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\big\}\phi_1^2\dot{\gamma}_{dd}.
\end{aligned}$$

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

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] = \mathrm{var}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}] + \mathbb{E}_{\boldsymbol{\theta}}^2[y_{dt}|\boldsymbol{v}] = \nu_{dt}p_{dt} + \nu_{dt}^2p_{dt}^2.$$

Therefore

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}]\big] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}p_{dt}f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt} \\
&\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}^2p_{dt}^2f(v_{1,d})f(v_{2,dt})\,dv_{1,d}dv_{2,dt} = S_1 + S_2,
\end{aligned}$$

where

$$\begin{aligned}
S_2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \nu_{dt}^2 p_{dt}^2 f(v_{2,dt}) f(v_{1,d})\, dv_{2,dt} dv_{1,d} \\
&= \nu_{dt}^2 \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1 v_{1,d} + 2\phi_2 v_{2,dt}\right\} f(v_{2,dt})\, dv_{2,dt} \right] f(v_{1,d})\, dv_{1,d} \\
&= \nu_{dt}^2 \int_{-\infty}^{\infty} \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_2^2) + 2\phi_1 v_{1,d}\right\} f(v_{1,d})\, dv_{1,d} \\
&= \nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\}.
\end{aligned}$$

Then, the expectation $\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]$ is

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2] = \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} + \nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\},$$

and as a consequence, the $(p+2)$-th MM equation is

$$\begin{aligned}
f_{p+2}(\boldsymbol{\theta}) = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} &\left\{ \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} + \nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\} \right\} \\
&- \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T} y_{dt}^2.
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \beta_k} &= \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} x_{dtk} + 2\nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\} x_{dtk}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \phi_1} &= \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_1\gamma_{dd} + 4\nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_1\gamma_{dd}, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \phi_2} &= \nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_2 + 4\nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_2, \\
\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \rho} &= \frac{1}{2}\nu_{dt} \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_1^2\dot{\gamma}_{dd} + 2\nu_{dt}^2 \exp\left\{2(\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \phi_2^2)\right\} \phi_1^2\dot{\gamma}_{dd}.
\end{aligned}$$

The expectation of $y_{d.}^2$ is $\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] = \mathbb{E}_v\left[\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2|\boldsymbol{v}]\right]$, where

$$y_{d.}^2 = \sum_{t=1}^{T} y_{dt}^2 + \sum_{t_1\neq t_2} y_{dt_1} y_{dt_2}, \quad \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] = \mathrm{var}_{\boldsymbol{\theta}}[y_{dt}|\boldsymbol{v}] + \mathbb{E}_{\boldsymbol{\theta}}^2[y_{dt}|\boldsymbol{v}] = \nu_{dt}p_{dt} + \nu_{dt}^2 p_{dt}^2.$$

The expectation of $y_{d.}^2$, conditionally on the random effects $\boldsymbol{v}$, is

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2|\boldsymbol{v}] &= \sum_{t=1}^{T}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2|\boldsymbol{v}] + \sum_{t_1\neq t_2}\mathbb{E}_{\boldsymbol{\theta}}[y_{dt_1}|\boldsymbol{v}]\mathbb{E}_{\boldsymbol{\theta}}[y_{dt_2}|\boldsymbol{v}] \\ &= \sum_{t=1}^{T}\left\{\nu_{dt}p_{dt} + \nu_{dt}^2 p_{dt}^2\right\} + \sum_{t_1\neq t_2}\nu_{dt_1}p_{dt_1}\nu_{dt_2}p_{dt_2}.\end{aligned}$$

Therefore

$$\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] = \sum_{t=1}^{T}\nu_{dt}\mathbb{E}_v[p_{dt}] + \sum_{t=1}^{T}\nu_{dt}^2\mathbb{E}_v[p_{dt}^2] + \sum_{t_1\neq t_2}\nu_{dt_1}\nu_{dt_2}\mathbb{E}_v[p_{dt_1}p_{dt_2}],$$

where the expectation of $p_{dt_1}p_{dt_2}$ is

$$\begin{aligned}\mathbb{E}_v[p_{dt_1}p_{dt_2}] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1 v_{1,d} + \phi_2 v_{2,dt_1} + \phi_2 v_{2,dt_2}\} \\ &\quad\cdot f(v_{2,dt_1})f(v_{2,dt_2})f(v_{1,d})\,dv_{2,dt_1}dv_{2,dt_2}dv_{1,d} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \phi_2 v_{2,dt_1} + 2\phi_1 v_{1,d}\} \\ &\quad\cdot f(v_{2,dt_1})f(v_{1,d})\,dv_{2,dt_1}dv_{1,d} \\ &= \int_{-\infty}^{\infty}\exp\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \frac{1}{2}\phi_2^2 + \frac{1}{2}\phi_2^2 + 2\phi_1 v_{1,d}\}f(v_{1,d})\,dv_{1,d} \\ &= \exp\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta} + \phi_2^2 + 2\phi_1^2\gamma_{dd}\}.\end{aligned}$$

Then, the expectation of $y_{d.}^2$ is

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2] &= \sum_{t=1}^{T}\nu_{dt}\exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} + \sum_{t=1}^{T}\nu_{dt}^2\exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} \\ &\quad + \sum_{t_1\neq t_2}\nu_{dt_1}\nu_{dt_2}\exp\{(\boldsymbol{x}_{dt_1}+\boldsymbol{x}_{dt_2})\boldsymbol{\beta} + 2\phi_1^2\gamma_{dd} + \phi_2^2\} \\ &\quad \pm \sum_{t=1}^{T}\nu_{dt}^2\exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2\gamma_{dd} + \phi_2^2\right\} \\ &= \sum_{t=1}^{T}\nu_{dt}\exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \frac{1}{2}(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} + \sum_{t=1}^{T}\nu_{dt}^2\exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2(\phi_1^2\gamma_{dd} + \phi_2^2)\right\} \\ &\quad - \sum_{t=1}^{T}\nu_{dt}^2\exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta} + 2\phi_1^2\gamma_{dd} + \phi_2^2\right\} + \left(\sum_{t=1}^{T}\nu_{dt}\exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1^2\gamma_{dd} + \frac{1}{2}\phi_2^2\right\}\right)^2,\end{aligned}$$

and as a consequence, the $(p+1)$-th MM equation is

$$\begin{aligned} f_{p+1}(\boldsymbol{\theta}) = \frac{1}{D}\sum_{d=1}^{D}\Bigg\{ & \sum_{t=1}^{T}\nu_{dt}\exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\frac{1}{2}(\phi_1^2\gamma_{dd}+\phi_2^2)\right\} \\ & + \left(e^{\phi_2^2}-1\right)\sum_{t=1}^{T}\nu_{dt}^2\exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta}+2\phi_1^2\gamma_{dd}+\phi_2^2)\right\} \\ & + \left(\sum_{t=1}^{T}\nu_{dt}\exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1^2\gamma_{dd}+\frac{1}{2}\phi_2^2\right\}\right)^2\Bigg\} - \frac{1}{D}\sum_{d=1}^{D}y_{d.}^2. \end{aligned}$$

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

$$\begin{aligned} \frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial\beta_k} &= \sum_{t=1}^{T}\nu_{dt}P_{dt}x_{dtk} + 2\sum_{t=1}^{T}\nu_{dt}^2Q_{dt}x_{dtk} - 2\sum_{t=1}^{T}\nu_{dt}^2R_{dt}x_{dtk} \\ &\quad + 2\left(\sum_{t=1}^{T}\nu_{dt}S_{dt}\right)\sum_{t=1}^{T}\nu_{dt}S_{dt}x_{dtk}, \\ \frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial\phi_1} &= \sum_{t=1}^{T}\nu_{dt}P_{dt}\phi_1\gamma_{dd} + 4\sum_{t=1}^{T}\nu_{dt}^2Q_{dt}\phi_1\gamma_{dd} - 4\sum_{t=1}^{T}\nu_{dt}^2R_{dt}\phi_1\gamma_{dd} \\ &\quad + 4\left(\sum_{t=1}^{T}\nu_{dt}S_{dt}\right)\sum_{t=1}^{T}\nu_{dt}S_{dt}\phi_1\gamma_{dd}, \\ \frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial\phi_2} &= \sum_{t=1}^{T}\nu_{dt}P_{dt}\phi_2 + 4\sum_{t=1}^{T}\nu_{dt}^2Q_{dt}\phi_2 - 2\sum_{t=1}^{T}\nu_{dt}^2R_{dt}\phi_2 \\ &\quad + 2\left(\sum_{t=1}^{T}\nu_{dt}S_{dt}\right)\sum_{t=1}^{T}\nu_{dt}S_{dt}\phi_2, \\ \frac{\partial\mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial\rho} &= \sum_{t=1}^{T}\frac{1}{2}\nu_{dt}P_{dt}\phi_1^2\dot{\gamma}_{dd} + 2\sum_{t=1}^{T}\nu_{dt}^2Q_{dt}\phi_1^2\dot{\gamma}_{dd} - 2\sum_{t=1}^{T}\nu_{dt}^2R_{dt}\phi_1^2\dot{\gamma}_{dd} \\ &\quad + 2\left(\sum_{t=1}^{T}\nu_{dt}S_{dt}\right)\sum_{t=1}^{T}\nu_{dt}S_{dt}\phi_1^2\dot{\gamma}_{dd}, \end{aligned}$$

where

$$\begin{aligned} P_{dt} &= \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\frac{1}{2}(\phi_1^2\gamma_{dd}+\phi_2^2)\right\}, \\ Q_{dt} &= \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta}+2(\phi_1^2\gamma_{dd}+\phi_2^2)\right\}, \\ R_{dt} &= \exp\left\{2\boldsymbol{x}_{dt}\boldsymbol{\beta}+2\phi_1^2\gamma_{dd}+\phi_2^2\right\}, \\ S_{dt} &= \exp\left\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1^2\gamma_{dd}+\frac{1}{2}\phi_2^2\right\}. \end{aligned}$$

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

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1.}y_{d_2.}] &= \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1.}y_{d_2.}|\boldsymbol{v}]\big] = \mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1.}|v_{1,d_1},\boldsymbol{v}_{2,d_1}]\mathbb{E}_{\boldsymbol{\theta}}[y_{d_2.}|v_{1,d_2},\boldsymbol{v}_{2,d_2}]\big]\\ &= \sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathbb{E}_v\big[\mathbb{E}_{\boldsymbol{\theta}}[y_{d_1t_1}|v_{1,d_1},v_{2,d_1t_1}]\mathbb{E}_{\boldsymbol{\theta}}[y_{d_2t_2}|v_{1,d_2},v_{2,d_2t_2}]\big]\\ &= \sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\nu_{d_1t_1}\nu_{d_2t_2}\mathbb{E}_v\big[p_{d_1t_1}p_{d_2t_2}\big].\end{aligned}$$

By defining $\varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta}) = \mathbb{E}_v\big[p_{d_1t_1}p_{d_2t_2}\big]$, it holds that

$$\begin{aligned}\varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta}) &= \int_{\mathbb{R}^4}\exp\big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\phi_1(v_{1,d_1}+v_{1,d_2})+\phi_2(v_{2,d_1t_1}+v_{2,d_2t_2})\big\}\\ &\quad\cdot f(v_{2,d_2t_2})dv_{2,d_2t_2}\,f(v_{2,d_1t_1})dv_{2,d_1t_1}f(v_{1,d_2}|v_{1,d_1})dv_{1,d_2}\,f(v_{1,d_1})\,dv_{1,d_1}\\ &= \int_{\mathbb{R}^3}\exp\big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\phi_1(v_{1,d_1}+v_{1,d_2})+\phi_2v_{2,d_1t_1}+\frac{1}{2}\phi_2^2\big\}\\ &\quad\cdot f(v_{2,d_1t_1})dv_{2,d_1t_1}\,f(v_{1,d_2}|v_{1,d_1})dv_{1,d_2}\,f(v_{1,d_1})\,dv_{1,d_1}\\ &= \int_{\mathbb{R}^2}\exp\big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\phi_1(v_{1,d_1}+v_{1,d_2})+\frac{1}{2}\phi_2^2+\frac{1}{2}\phi_2^2\big\}\\ &\quad\cdot f(v_{1,d_2}|v_{1,d_1})dv_{1,d_2}\,f(v_{1,d_1})\,dv_{1,d_1}\\ &= \int_{\mathbb{R}}\exp\Big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\phi_1v_{1,d_1}+\frac{\gamma_{d_1d_2}}{\gamma_{d_1d_1}}v_{1,d_1}\phi_1\\ &\quad+\frac{1}{2}\Big(\gamma_{d_2d_2}-\frac{\gamma_{d_1d_2}^2}{\gamma_{d_1d_1}}\Big)\phi_1^2+\phi_2^2\Big\}f(v_{1,d_1})\,dv_{1,d_1}\\ &= \exp\Big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\frac{1}{2}\Big(1+\frac{\gamma_{d_1d_2}}{\gamma_{d_1d_1}}\Big)^2\gamma_{d_1d_1}\phi_1^2+\frac{1}{2}\Big(\gamma_{d_2d_2}-\frac{\gamma_{d_1d_2}^2}{\gamma_{d_1d_1}}\Big)\phi_1^2+\phi_2^2\Big\}\\ &= \exp\Big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}+\frac{1}{2}\,\phi_1^2(\gamma_{d_1d_1}+2\gamma_{d_1d_2}+\gamma_{d_2d_2})+\phi_2^2\Big\}.\end{aligned}$$

Therefore, the $(p+3)$-th MM equation is

$$\begin{aligned}f_{p+3}(\boldsymbol{\theta}) &= \frac{1}{D(D-1)}\sum_{d_1\neq d_2}^{D}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\nu_{d_1t_1}\nu_{d_2t_2}\exp\Big\{(\boldsymbol{x}_{d_1t_1}+\boldsymbol{x}_{d_2t_2})\boldsymbol{\beta}\\ &\quad+\frac{1}{2}\,\phi_1^2(\gamma_{d_1d_1}+2\gamma_{d_1d_2}+\gamma_{d_2d_2})+\phi_2^2\Big\}-\frac{1}{D(D-1)}\sum_{d_1\neq d_2}^{D}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\nu_{d_1t_1}y_{d_1.}y_{d_2.}\end{aligned}$$

The derivatives of $\varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})$ are

$$\begin{aligned}
\frac{\partial \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})}{\partial \beta_k} &= \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})(x_{d_1t_1k} + x_{d_2t_2k}),\\
\frac{\partial \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})}{\partial \phi_1} &= \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})\phi_1(\gamma_{d_1d_1} + \gamma_{d_2d_2} + 2\gamma_{d_1d_2}),\\
\frac{\partial \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})}{\partial \phi_2} &= 2\varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})\phi_2,\\
\frac{\partial \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})}{\partial \rho} &= \frac{1}{2}\varphi_{d_1,d_2}(\boldsymbol{\theta})\phi_1^2(\dot{\gamma}_{d_1d_1} + \dot{\gamma}_{d_2d_2} + 2\dot{\gamma}_{d_1d_2}).
\end{aligned}$$

The elements of the Jacobian matrix are

$$\begin{aligned}
H_{kr} &= \frac{\partial f_k(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}]}{\partial \theta_r}\, x_{dtk}, \quad k = 1,\ldots,p,\ r = 1,\ldots,p+3,\\
H_{p+1r} &= \frac{\partial f_{p+1}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D}\sum_{d=1}^{D}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{d.}^2]}{\partial \theta_r}, \quad r = 1,\ldots,p+3,\\
H_{p+2r} &= \frac{\partial f_{p+2}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{DT}\sum_{d=1}^{D}\sum_{t=1}^{T}\frac{\partial \mathbb{E}_{\boldsymbol{\theta}}[y_{dt}^2]}{\partial \theta_r}, \quad r = 1,\ldots,p+3,\\
H_{p+3r} &= \frac{\partial f_{p+3}(\boldsymbol{\theta})}{\partial \theta_r} = \frac{1}{D(D-1)}\sum_{d_1\neq d_2}^{D}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\nu_{d_1t_1}\nu_{d_2t_2}\frac{\partial \varphi_{d_1d_2}^{t_1t_2}(\boldsymbol{\theta})}{\partial \theta_r}, \quad r = 1,\ldots,p+3.
\end{aligned}$$

The MM algorithm under Model ST1 keeps the steps of Algorithm 1 (see Section 2.3.1), replacing $\boldsymbol{\theta}$, $\boldsymbol{H}$ and $\boldsymbol{f}$ for those given in (5.3.2).

As algorithm seeds for $\boldsymbol{\beta}$, ϕ_1 and ϕ_2, the algorithm may take the MM estimator under the model with no spatial correlation ($\rho = 0$), i.e. under Model T1. For ρ, it may the Moran's I measure of spatial autocorrelation

$$I = \frac{D}{\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1d_2}}\frac{\sum_{d_1=1}^{D}\sum_{d_2=1}^{D} w_{d_1d_2}(\tilde{v}_{1,d_1} - \tilde{v}_1)(\tilde{v}_{1,d_2} - \tilde{v}_1)}{\sum_{d=1}^{D}(\tilde{v}_{1,d} - \tilde{v})^2},$$

where $\tilde{v}_{1,d}$, $d = 1,\ldots,D$, are the predicted random effects under Model T1, $\tilde{v} = \frac{1}{D}\sum_{d=1}^{D}\tilde{v}_{1,d}$ and the $w_{d_1d_2}$'s are the elements of the proximity matrix $\boldsymbol{W}$ (see Section 3.2 for more details).

The asymptotic variance of the MM estimator under Model ST1 can be approximated by a similar bootstrap algorithm to that described in Section 2.3.1.

5.4 Hypothesis tests for the model parameters

This section presents three bootstrap algorithms for testing the significance of the variance parameters, ϕ_1 and ϕ_2, and of the autocorrelation parameter ρ. Algorithm 8 gives a bootstrap procedure to test the hypothesis $H_0 : \phi_1 = 0$. In this case, we study Model T1_2 (see Section 4.2.1) against Model ST1.

Algorithm 8 A bootstrap test for $H_0 : \phi_1 = 0$

1: Fit the Model ST1 to data and calculate $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\rho}$.

2: Fit the Model T1_2 to data and calculate $\hat{\boldsymbol{\beta}}^0$ and $\hat{\phi}_2^0$.

3: For $b = 1, \ldots, B$, do

i) Generate a bootstrap resample under $H_0 : \phi_1 = 0$, i.e.

$$v_{2,dt}^{*(b)} \sim N(0,1),\ p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}}^0 + \hat{\phi}_2^0 v_{2,dt}^{*(b)}\},$$
$$y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)}),\ d = 1, \ldots, D,\ t = 1, \ldots, T.$$

ii) Fit the Model ST1 to the bootstrap data $(y_{dt}^{*(b)}, \boldsymbol{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, and calculate $\hat{\boldsymbol{\beta}}^{*(b)}$, $\hat{\phi}_1^{*(b)}$, $\hat{\phi}_2^{*(b)}$ and $\hat{\rho}^{*(b)}$.

4: Calculate the p-value

$$p = \frac{\#\left\{\hat{\phi}_1^{*(b)} > \hat{\phi}_1\right\}}{B}.$$

If the null hypothesis $H_0 : \phi_1 = 0$ is rejected, the significance of the autocorrelation parameter can be tested. Algorithm 9 presents a bootstrap procedure for testing $H_0 : \rho = 0$.

Algorithm 9 A bootstrap test for $H_0 : \rho = 0$

1: Fit the Model ST1 to data and calculate $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\rho}$.

2: Fit the Model T1 to data and calculate $\hat{\boldsymbol{\beta}}^0$, $\hat{\phi}_1^0$ and $\hat{\phi}_2^0$.

3: For $b = 1, \ldots, B$, do

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

$$v_{1,d}^{*(b)} \sim N(0,1),\ v_{2,dt}^{*(b)} \sim N(0,1),\ p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}}^0 + \hat{\phi}_1^0 v_{1,d}^{*(b)} + \hat{\phi}_2^0 v_{2,dt}^{*(b)}\},$$
$$y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)}),\ d = 1, \ldots, D,\ t = 1, \ldots, T.$$

ii) Fit the Model ST1 to the bootstrap data $(y_{dt}^{*(b)}, \boldsymbol{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, and calculate $\hat{\boldsymbol{\beta}}^{*(b)}$, $\hat{\phi}_1^{*(b)}$, $\hat{\phi}_2^{*(b)}$ and $\hat{\rho}^{*(b)}$.

4: Calculate the p-value

$$p = \frac{\#\left\{\left|\hat{\rho}^{*(b)}\right| > |\hat{\rho}|\right\}}{B}.$$

Finally, Algorithm 10 gives a bootstrap procedure for testing the null hypothesis $H_0 : \phi_2 = 0$. If it is accepted, the working model would be the Model ST1$_1$ defined below.

Algorithm 10 A bootstrap test for $H_0 : \phi_2 = 0$

1: Fit the Model ST1 to data and calculate $\hat{\boldsymbol{\beta}}$, $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\rho}$.

2: Fit the model

$$\text{Model ST1}_1\text{: } \log \mu_{dt} = \log \nu_{dt} + \boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d},\ d = 1, \ldots, D,\ t = 1, \ldots, T,$$

to data and calculate $\hat{\boldsymbol{\beta}}^0$, $\hat{\phi}_1^0$ and $\hat{\rho}^0$.

3: For $b = 1, \ldots, B$, do

i) Generate a bootstrap resample under $H_0 : \phi_2 = 0$, i.e.

$$\begin{aligned} &\boldsymbol{v}_1^{*(b)} \sim N_D(\boldsymbol{0}, \boldsymbol{\Gamma}(\hat{\rho}^0)),\ p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}}^0 + \hat{\phi}_1^0 v_{1,d}^{*(b)}\}, \\ &y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)}),\ d = 1, \ldots, D,\ t = 1, \ldots, T. \end{aligned}$$

ii) Fit the Model ST1 to the bootstrap data $(y_{dt}^{*(b)}, \boldsymbol{x}_{dt})$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, and calculate $\hat{\boldsymbol{\beta}}^{*(b)}$, $\hat{\phi}_1^{*(b)}$, $\hat{\phi}_2^{*(b)}$ and $\hat{\rho}^{*(b)}$.

4: Calculate the p-value

$$p = \frac{\#\left\{\hat{\phi}_2^{*(b)} > \hat{\phi}_2\right\}}{B}.$$

5.5 The predictors

This section derive the EBP and proposes a plug-in predictor of p_{dt} under Model ST1. The EBP of p_{dt} is obtained from the corresponding BP replacing the vector of model parameters $\boldsymbol{\theta}$ by an estimator $\hat{\boldsymbol{\theta}}$. As the MM estimators are consistent, they are employed for calculating the EBPs. To avoid overflow numerical problems in the calculation of the exact EBP, this section proposes two alternative approximations.

Let $\boldsymbol{y}_d$ be the response vector within the domain d, i.e. $\boldsymbol{y}_d = \underset{1 \le t \le T}{\text{col}}(y_{dt})$. The conditional distribution of the response variable $\boldsymbol{y}$, given the random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}_1, \boldsymbol{v}_2) = \prod_{d=1}^{D} \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}), \quad \mathbb{P}(\boldsymbol{y}_d|v_{1,d}, \boldsymbol{v}_{2,d}) = \prod_{t=1}^{T} \mathbb{P}(y_{dt}|v_{1,d}, v_{2,dt}),$$

where

$$\begin{aligned}\mathbb{P}(y_{dt}|v_{1,d}, v_{2,dt}) &= \frac{1}{y_{dt}!}\exp\{-\nu_{dt}p_{dt}\}\nu_{dt}^{y_{dt}}p_{dt}^{y_{dt}} \\ &= c_{dt}\exp\left\{y_{dt}(\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}) - \nu_{dt}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\right\}.\end{aligned}$$

5.5.1 The empirical best predictor

This section derives the EBP of the probability and mean parameters, p_{dt} and μ_{dt}, $d = 1,\ldots,D$, $t = 1,\ldots,T$. They are obtained from the corresponding BPs, replacing the theoretical vector of model parameters $\boldsymbol{\theta}$ (unknown in practice), by an estimator $\hat{\boldsymbol{\theta}}$.

The BP of p_{dt} is the unbiased predictor minimizing the MSE. It is obtained as the conditional expectation $\hat{p}_{dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}]$. It holds that

$$\mathbb{E}_{\boldsymbol{\theta}}[p_{dt}|\boldsymbol{y}] = \frac{\int_{\mathbb{R}^{D(T+1)}} p_{dt}\mathbb{P}\left(\boldsymbol{y}|\boldsymbol{v}_1,\boldsymbol{v}_2\right) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)}{\int_{\mathbb{R}^{D(T+1)}} \mathbb{P}\left(\boldsymbol{y}|\boldsymbol{v}_1,\boldsymbol{v}_2\right) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)} = \frac{N_{dt}}{B}, \tag{5.5.1}$$

where $N_{dt} = N_{dt}(\boldsymbol{y},\boldsymbol{\theta})$ and $B = B(\boldsymbol{y},\boldsymbol{\theta})$ are given by

$$\begin{aligned}N_{dt} &= \int_{\mathbb{R}^{D(T+1)}} \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta}+\phi_1 v_{1,d}+\phi_2 v_{2,dt}\}\left(\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\mathbb{P}(y_{\ell\tau}|v_{1,\ell},v_{2,\ell\tau})\right) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\ &= \int_{\mathbb{R}^{D(T+1)}} \prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\exp\Big\{(y_{\ell\tau}+\delta_{d\ell}\delta_{t\tau})(\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta}+\phi_1 v_{1,\ell}+\phi_2 v_{2,d\tau}) \\ &\quad - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta}+\phi_1 v_{1,\ell}+\phi_2 v_{2,\ell\tau}\}\Big\} f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2, \\ B &= \int_{\mathbb{R}^{D(T+1)}} \left(\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\mathbb{P}(y_{\ell\tau}|v_{1,\ell},v_{2,\ell\tau})\right) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\ &= \int_{\mathbb{R}^{D(T+1)}} \prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\exp\Big\{y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta}+\phi_1 v_{1,\ell}+\phi_2 v_{2,d\tau}) \\ &\quad - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta}+\phi_1 v_{1,\ell}+\phi_2 v_{2,\ell\tau}\}\Big\} f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2,\end{aligned} \tag{5.5.2}$$

and $\delta_{d\ell}$ and $\delta_{t\tau}$ denote the Kronecker delta, i.e. $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. In this chapter, as the denominator of the BP does not depend on d and t, the notation is changed and it is not the same as in the previous chapters (now it is denoted by B). This is done to avoid confusion with the total number of domains D.

Remark 5.1. The component $N_{dt}(\boldsymbol{y},\boldsymbol{\theta})$ can be expressed in terms of $B(\boldsymbol{y},\boldsymbol{\theta})$ as $N_{dt}(\boldsymbol{y},\boldsymbol{\theta}) = B(\boldsymbol{y}+\boldsymbol{e}_{dt},\boldsymbol{\theta})$, where $\boldsymbol{e}_{dt} = \{\delta_{d\ell}\delta_{t\tau}\}$, $d = 1,\ldots,D$, $t = 1,\ldots,T$.

The EBP of p_{dt} is $\hat{p}_{dt} = \hat{p}_{dt}(\hat{\boldsymbol{\theta}})$. The EBP calculation involves complex integrals in a high-dimensional space. The integrals are approximated by using an antithetic Monte Carlo algorithm. The steps are:

1. Generate $\boldsymbol{v}_1^{(s_1)} \sim N_D\big(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho})\big)$, $v_{2,\ell\tau}^{(s_2)}$ i.i.d. $N(0,1)$ and calculate $\boldsymbol{v}_1^{(S_1+s_1)} = -\boldsymbol{v}_1^{(s_1)}$, $v_{2,\ell\tau}^{(S_2+s_2)} = -v_{2,\ell\tau}^{(s_2)}$, $s_1 = 1, \ldots, S_1$, $s_2 = 1, \ldots, S_2$, $\ell = 1, \ldots, D$, $\tau = 1, \ldots, T$.

2. Approximate the EBP of p_{dt} as $\hat{p}_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{dt}/\hat{B}$, where

$$\begin{aligned}
\hat{N}_{dt} = \sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T} \exp\Big\{&(y_{\ell\tau} + \delta_{d\ell}\delta_{t\tau})(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}) \\
&- \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}\}\Big\}, \\
\hat{B} = \sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T} \exp\Big\{&y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}) \\
&- \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}\}\Big\}. \qquad (5.5.3)
\end{aligned}$$

One might find overflow numerical problems when running the above algorithm. In what follows, an alternative way running Step 2, especially oriented to its programming, is presented. For all s_1, s_2, ℓ, τ, let

$$\begin{aligned}
\eta_{s_1 s_2 \ell\tau} &= \delta_{d\ell}\delta_{t\tau}(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}), \\
\xi_{s_1 s_2 \ell\tau} &= y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}) - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}\}, \\
\bar{\xi} &= \frac{1}{4S_1S_2DT}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\sum_{\ell=1}^{D}\sum_{\tau=1}^{T}\xi_{s_1 s_2 \ell\tau}, \\
\sigma_\xi^2 &= \frac{1}{4S_1S_2DT-1}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\sum_{\ell=1}^{D}\sum_{\tau=1}^{T}(\xi_{s_1 s_2 \ell\tau} - \bar{\xi})^2.
\end{aligned}$$

Note that

$$e^x = e^{x-\mu+\mu} = e^\mu e^{x-\mu} = e^\mu\left(\exp\left\{\frac{x-\mu}{\sigma}\right\}\right)^\sigma.$$

Therefore, $\hat{N}_{dt}$ and $\hat{B}$ are substituted in Step 2 by

$$\begin{aligned}
\hat{N}_{dt} &= e^{DT\bar{\xi}}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\left(\exp\left\{\frac{\xi_{s_1 s_2 \ell\tau} + \eta_{s_1 s_2 \ell\tau} - \bar{\xi}}{\sigma_\xi}\right\}\right)^{\sigma_\xi}, \\
\hat{B} &= e^{DT\bar{\xi}}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\left(\exp\left\{\frac{\xi_{s_1 s_2 \ell\tau} - \bar{\xi}}{\sigma_\xi}\right\}\right)^{\sigma_\xi}, \qquad (5.5.4)
\end{aligned}$$

and $e^{DT\bar{\xi}}$ is cancelled when substituting into $\hat{p}_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{dt}/\hat{B}$.

As the relationship between the mean and probability parameters is $\mu_{dt} = \nu_{dt}p_{dt}$, and ν_{dt} is known (size parameter), the EBP of μ_{dt} is obtained as an immediate consequence of the EBP of p_{dt}. That is to say, the EBP of μ_{dt} is $\hat{\mu}_{dt}(\hat{\boldsymbol{\theta}}) = \nu_{dt}\hat{p}_{dt}(\hat{\boldsymbol{\theta}})$.

The two previous alternatives are computationally demanding. For that, this section proposes an approximation to the BP of p_{dt} (5.5.1) under the area-level spatio-temporal Poisson mixed model. Divide $\boldsymbol{y}$ and $\boldsymbol{v} = (\boldsymbol{v}_1', \boldsymbol{v}_2')'$ into two parts $(\boldsymbol{y}_d', \boldsymbol{y}_{d-}')'$ and $(\boldsymbol{v}_d', \boldsymbol{v}_{d-}')'$, where $\boldsymbol{y}_d = \underset{1\leq t\leq T}{\text{col}}(y_{dt})$ $\boldsymbol{y}_{d-} = \underset{1\leq i\leq D, i\neq d}{\text{col}}(\boldsymbol{y}_i)$, $\boldsymbol{v}_d = (v_{1,d}, \boldsymbol{v}_{2,d}')'$ and $\boldsymbol{v}_{d-} = \underset{1\leq i\leq D, i\neq d}{\text{col}}(\boldsymbol{v}_i)$. The conditional distribution of $\boldsymbol{y}$, given $\boldsymbol{v}$, is

$$\mathbb{P}(\boldsymbol{y}|\boldsymbol{v}) = \prod_{i=1}^{D}\mathbb{P}(\boldsymbol{y}_i|\boldsymbol{v}_i) = \mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_d)\prod_{i=1,i\neq d}^{D}\mathbb{P}(\boldsymbol{y}_i|\boldsymbol{v}_i) = \mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_d)\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}). \tag{5.5.5}$$

The p.d.f. of $\boldsymbol{v}_d$ is

$$f(\boldsymbol{v}_d) = f(v_{1,d})f(\boldsymbol{v}_{2,d}),$$

where $v_{1,d} \sim N(0, \gamma_{dd}(\rho))$ and $\boldsymbol{v}_{2,d} \sim N(\boldsymbol{0}, \boldsymbol{I}_T)$. The component B of the BP of p_{dt} (5.5.1) can be rewritten by using the decomposition of the conditional probability given in (5.5.5), i.e.

$$B = \int_{\mathbb{R}^{T+1}}\Big[\int_{\mathbb{R}^{(D-1)(T+1)}}\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-})f(\boldsymbol{v}_{d-}|\boldsymbol{v}_d)\,d\boldsymbol{v}_{d-}\Big]\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_d)f(\boldsymbol{v}_d)\,d\boldsymbol{v}_d.$$

As $\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-})f(\boldsymbol{v}_{d-}|\boldsymbol{v}_d) = \mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}, \boldsymbol{v}_d)f(\boldsymbol{v}_{d-}|\boldsymbol{v}_d)$, the inner integral is

$$\int_{\mathbb{R}^{(D-1)(T+1)}}\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_{d-}, \boldsymbol{v}_d)f(\boldsymbol{v}_{d-}|\boldsymbol{v}_d)\,d\boldsymbol{v}_{d-} = \mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_d),$$

and as a consequence, it holds that

$$B(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}^{T+1}}\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_d)\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_d)f(\boldsymbol{v}_d)\,d\boldsymbol{v}_d.$$

By applying similar developments as for the component $N_{dt}(\boldsymbol{y}, \boldsymbol{\theta})$ of (5.5.1), and by taking into account Remark 5.1, it holds that

$$N_{dt}(\boldsymbol{y}, \boldsymbol{\theta}) = \int_{\mathbb{R}^{T+1}}\exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\}\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_d)\mathbb{P}(\boldsymbol{y}_d|\boldsymbol{v}_d)f(\boldsymbol{v}_d)\,d\boldsymbol{v}_d,$$

Under the assumption

$$\mathbb{P}(\boldsymbol{y}_{d-}|\boldsymbol{v}_d) \approx \mathbb{P}(\boldsymbol{y}_{d-}), \tag{5.5.6}$$

the BP of p_{dt}, $\hat{p}_{dt}(\boldsymbol{\theta})$, can be approximated by

$$\hat{p}^a_{dt}(\boldsymbol{\theta}) = N^a_{dt}(\boldsymbol{y}_d, \boldsymbol{\theta})/B^a_d(\boldsymbol{y}_d, \boldsymbol{\theta}),$$

where

$$\begin{aligned}
N^a_{dt} &= \int_{\mathbb{R}^{T+1}} \exp\{\boldsymbol{x}_{dt}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\} \prod_{\tau=1}^{T} \mathbb{P}(y_{d\tau}|v_{1,d}, v_{2,d\tau}) f(v_{1,d}) f(\boldsymbol{v}_{2,d})\, dv_{1,d}\, d\boldsymbol{v}_{2,d} \\
&= \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Big[\int_{\mathbb{R}} \exp\Big\{ (y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&\quad - \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\} \Big\} f(v_{2,d\tau}) dv_{2,d\tau} \Big] f(v_{1,d}) dv_{1,d}, \\
B^a_d &= \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Big[\int_{\mathbb{R}} \exp\Big\{ y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) \\
&\quad - \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\} \Big\} f(v_{2,d\tau}) dv_{2,d\tau} \Big] f(v_{1,d}) dv_{1,d}. \qquad (5.5.7)
\end{aligned}$$

Then, the EBP of p_{dt}, $\hat{p}_{dt}(\hat{\boldsymbol{\theta}})$, can be approximated as follows.

1. Estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\phi}_1, \hat{\phi}_2, \hat{\rho})$.

2. For $s_1 = 1, \ldots, S_1$, $s_2 = 1, \ldots, S_2$, $\ell = 1, \ldots, D$, $\tau = 1, \ldots, T$, generate $\boldsymbol{v}_1^{(s_1)} \sim N_D\big(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho})\big)$, $v_{2,\ell\tau}^{(s_2)}$ i.i.d. $N(0,1)$ and calculate $\boldsymbol{v}_1^{(S_1+s_1)} = -\boldsymbol{v}_1^{(s_1)}$, $v_{2,\ell\tau}^{(S_2+s_2)} = -v_{2,\ell\tau}^{(s_2)}$.

3. Calculate $\hat{p}^a_{dt}(\hat{\boldsymbol{\theta}}) = \hat{N}^a_{dt}/\hat{B}^a_d$, $d = 1, \ldots, D$, $t = 1, \ldots, T$, where

$$\begin{aligned}
\hat{N}^a_{dt} &= \sum_{s_1=1}^{2S_1} \prod_{\tau=1}^{T} \sum_{s_2=1}^{2S_2} \exp\Big\{ (y_{d\tau} + \delta_{t\tau})(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}) \\
&\quad - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\} \Big\}, \\
\hat{B}^a_d &= \sum_{s_1=1}^{2S_1} \prod_{\tau=1}^{T} \sum_{s_2=1}^{2S_2} \exp\Big\{ y_{d\tau}(\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}) \\
&\quad - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{(s_1)} + \hat{\phi}_2 v_{2,d\tau}^{(s_2)}\} \Big\}.
\end{aligned}$$

The EBP approximation, $\hat{p}^a_{dt}(\hat{\boldsymbol{\theta}})$, maintains the expression of the EBP of p_{dt} under the area-level Poisson mixed model with independent time effects given in Section 4.3.1, but unlike then, now the domain random effects, $v_{1,d}$, are generated according to a SAR(1) process.

5.5.2 The plug-in predictor

This section gives a plug-in predictor of the target parameter p_{dt}, where the random effects $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are predicted by their EBPs. The plug-in predictor of p_{dt} is obtained by replacing, in the theoretical expression of p_{dt}, the model parameters by their estimates and the random effects by their predictions, i.e.

$$\hat{p}_{dt}^{P} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{\phi}_1\hat{v}_{1,d} + \hat{\phi}_2\hat{v}_{2,dt}\}. \tag{5.5.8}$$

As above, the EBPs of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are obtained from the respective BPs. The BP of $v_{1,d}$ is

$$\hat{v}_{1,d}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{1,d}|\boldsymbol{y}] = \frac{N_{1,d}(\boldsymbol{y},\boldsymbol{\theta})}{B(\boldsymbol{y},\boldsymbol{\theta})} = \frac{N_{1,d}}{B},$$

where B was defined in (5.5.2) and

$$\begin{aligned} N_{1,d} &= \int_{\mathbb{R}^{D(T+1)}} v_{1,d}\left(\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\mathbb{P}(y_{\ell\tau}|v_{1,\ell},\boldsymbol{v}_{2,\ell})\right) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\ &= \int_{\mathbb{R}^{D(T+1)}} \prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\exp\Big\{y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta} + \phi_1 v_{1,\ell} + \phi_2 v_{2,\ell\tau}) \\ &\quad - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta} + \phi_1 v_{1,\ell} + \phi_2 v_{2,\ell\tau}\}\Big\} v_{1,d} f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2. \end{aligned}$$

The EBP of $v_{1,d}$ is $\hat{v}_{1,d} = \hat{v}_{1,d}(\hat{\boldsymbol{\theta}})$ and it can be approximated as follows.

1. Generate $\boldsymbol{v}_1^{(s_1)} \sim N_D\big(\boldsymbol{0}, \boldsymbol{\Gamma}(\hat{\rho})\big)$, $v_{2,\ell\tau}^{(s_2)}$ i.i.d. $N(0,1)$ and calculate $\boldsymbol{v}_1^{(S_1+s_1)} = -\boldsymbol{v}_1^{(s_1)}$, $v_{2,\ell\tau}^{(S_2+s_2)} = -v_{2,\ell\tau}^{(s_2)}$, $s_1 = 1,\ldots,S_1$, $s_2 = 1,\ldots,S_2$, $\ell = 1,\ldots,D$, $\tau = 1,\ldots,T$.

2. Calculate $\hat{v}_{1,d}(\hat{\boldsymbol{\theta}}) = \hat{N}_{1,d}/\hat{B}$, where $\hat{B}$ was defined in (5.5.3) and

$$\begin{aligned} \hat{N}_{1,d} = \sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\{1 + \delta_{\ell d}(v_{1,\ell}^{(s_1)} - 1)\}\prod_{\tau=1}^{T}\exp\Big\{y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}) \\ - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}\}\Big\}. \end{aligned}$$

An alternative way of calculating $\hat{N}_{1,d}$ in Step 2 is

$$\hat{N}_{1,d} = e^{DT\bar{\xi}}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\{1 + \delta_{\ell d}(v_{1,\ell}^{(s_1)} - 1)\}\prod_{\tau=1}^{T}\left(\exp\left\{\frac{\xi_{s_1 s_2 \ell\tau} - \bar{\xi}}{\sigma_\xi}\right\}\right)^{\sigma_\xi},$$

and $e^{DT\bar{\xi}}$ is cancelled when substituting $\hat{N}_{1,d}$ and $\hat{B}$ (given in eq. (5.5.4)) into $\hat{v}_{1,d}(\hat{\boldsymbol{\theta}}) = \hat{N}_{1,d}/\hat{B}$.

The BP of the domain-time random effects $v_{2,dt}$ is

$$\hat{v}_{2,dt}(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\theta}}[v_{2,dt}|\boldsymbol{y}] = \frac{N_{2,dt}(\boldsymbol{y},\boldsymbol{\theta})}{B(\boldsymbol{y},\boldsymbol{\theta})} = \frac{N_{2,dt}}{B},$$

where B was defined in (5.5.2) and

$$\begin{aligned} N_{2,dt} &= \int_{\mathbb{R}^{D(T+1)}} v_{2,dt}\Big(\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\mathbb{P}(y_{\ell\tau}|v_{1,\ell},\boldsymbol{v}_{2,\ell})\Big) f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\ &= \int_{\mathbb{R}^{D(T+1)}} \prod_{\ell=1}^{D}\prod_{\tau=1}^{T} \exp\Big\{y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta} + \phi_1 v_{1,\ell} + \phi_2 v_{2,d\tau}) \\ &\quad - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\boldsymbol{\beta} + \phi_1 v_{1,\ell} + \phi_2 v_{2,\ell\tau}\}\Big\} v_{2,dt} f(\boldsymbol{v}_1)f(\boldsymbol{v}_2)\,d\boldsymbol{v}_1 d\boldsymbol{v}_2. \end{aligned}$$

The EBP of $v_{2,dt}$ is $\hat{v}_{2,dt} = \hat{v}_{2,dt}(\hat{\boldsymbol{\theta}})$ and it can be approximated as follows.

1. Generate $\boldsymbol{v}_1^{(s_1)} \sim N_D\big(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho})\big)$, $v_{2,\ell\tau}^{(s_2)}$ i.i.d. $N(0,1)$ and calculate $\boldsymbol{v}_1^{(S_1+s_1)} = -\boldsymbol{v}_1^{(s_1)}$, $v_{2,\ell\tau}^{(S_2+s_2)} = -v_{2,\ell\tau}^{(s_2)}$, $s_1 = 1,\ldots,S_1$, $s_2 = 1,\ldots,S_2$, $\ell = 1,\ldots,D$, $\tau = 1,\ldots,T$.

2. Calculate $\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{2,dt}/\hat{B}$, where $\hat{B}$ was defined in (5.5.3) and

$$\begin{aligned} \hat{N}_{2,dt} = \sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\big[1 + \delta_{d\ell}\delta_{t\tau}(v_{2,\ell\tau}^{(s_2)} - 1)\big] \exp\Big\{y_{\ell\tau}(\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}) \\ - \nu_{\ell\tau}\exp\{\boldsymbol{x}_{\ell\tau}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,\ell}^{(s_1)} + \hat{\phi}_2 v_{2,\ell\tau}^{(s_2)}\}\Big\}. \end{aligned}$$

An alternative way of calculating $\hat{N}_{2,dt}$ in Step 2 is

$$\hat{N}_{2,dt} = e^{DT\bar{\xi}}\sum_{s_1=1}^{2S_1}\sum_{s_2=1}^{2S_2}\prod_{\ell=1}^{D}\prod_{\tau=1}^{T}\big[1 + \delta_{d\ell}\delta_{t\tau}(v_{2,\ell\tau}^{(s_2)} - 1)\big]\Big(\exp\Big\{\frac{\xi_{s_1 s_2 \ell\tau} - \bar{\xi}}{\sigma_\xi}\Big\}\Big)^{\sigma_\xi},$$

and $e^{DT\bar{\xi}}$ is cancelled when substituting $\hat{N}_{2,dt}$ and $\hat{B}$ (given in eq. (5.5.4)) into $\hat{v}_{2,dt}(\hat{\boldsymbol{\theta}}) = \hat{N}_{2,dt}/\hat{B}$.

Again, the EBPs of $v_{1,d}$ and $v_{2,dt}$ are computationally demanding and therefore this section proposes two approximations, $\hat{v}_{1,d}^a$ and $\hat{v}_{2,dt}^a$, for the EBPs $\hat{v}_{1,d}$ and $\hat{v}_{2,dt}$ respectively. Under the assumption (5.5.6), the BP of the domain random effects, $\hat{v}_{1,d}(\boldsymbol{\theta})$, can be approximated

by

$$\hat{v}^a_{1,d}(\boldsymbol{\theta}) = \frac{N^a_{1,d}(\boldsymbol{y}_d, \boldsymbol{\theta})}{B^a_d(\boldsymbol{y}_d, \boldsymbol{\theta})},$$

where $B^a_d(\boldsymbol{y}_d, \boldsymbol{\theta})$ is given in (5.5.7) and

$$N^a_{1,d}(\boldsymbol{y}_d, \boldsymbol{\theta}) = \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Bigg[\int_{\mathbb{R}} \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{2,d\tau})\, dv_{2,d\tau} \Bigg] v_{1,d} f(v_{1,d})\, dv_{1,d}.$$

Again, the difference between the approximated BP, $\hat{v}^a_{1,d}(\boldsymbol{\theta})$, and the corresponding one under Model T1 (see Section 4.3.1) is that now, the distribution of the domain random effects is SAR(1).

The BP of $v_{2,dt}$, $\hat{v}_{2,dt}(\boldsymbol{\theta})$, can be approximated by

$$\hat{v}^a_{2,dt}(\boldsymbol{\theta}) = \frac{N^a_{2,dt}(\boldsymbol{y}_d, \boldsymbol{\theta})}{B^a_d(\boldsymbol{y}_d, \boldsymbol{\theta})},$$

where

$$N^a_{2,dt}(\boldsymbol{y}_d, \boldsymbol{\theta}) = \int_{\mathbb{R}} \prod_{\tau=1}^{T} \Bigg[\int_{\mathbb{R}} (1 + \delta_{t\tau}(v_{2,d\tau} - 1)) \exp\{y_{d\tau}(\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}) - \nu_{d\tau} \exp\{\boldsymbol{x}_{d\tau}\boldsymbol{\beta} + \phi_1 v_{1,d} + \phi_2 v_{2,d\tau}\}\} f(v_{2,d\tau})\, dv_{2,d\tau} \Bigg] f(v_{1,d})\, dv_{1,d}.$$

In this case, the underlying spatial correlation structure slightly complicates the expression of the approximated BP, $\hat{v}^a_{2,dt}(\boldsymbol{\theta})$, with respect to that obtained under Model T1.

5.6 MSE estimation

As accuracy measure of the EBP of p_{dt} under Model ST1, this section considers the MSE. It proposes estimating the MSE of the EBP of p_{dt} by using a parametric bootstrap algorithm based on the bootstrap procedure given in González-Manteiga et al. (2007), since the analytical estimation is not feasible computationally. The steps are:

1. Fit the model to the sample and calculate the estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\phi}_1, \hat{\phi}_2, \hat{\rho})$ under Model ST1.

2. For each domain d, $d = 1, \ldots, D$, and time instant t, $t = 1, \ldots, T$, repeat B times,

$b = 1, \ldots, B$.

i) Generate the bootstrap random effects $\boldsymbol{v}_1^{*(b)} \sim N_D\big(\mathbf{0}, \boldsymbol{\Gamma}(\hat{\rho})\big)$ and $\{v_{2,dt}^{*(b)}\}$ i.i.d. $N(0,1)$.

ii) Calculate the theoretical bootstrap EBP estimator $p_{dt}^{*(b)} = \exp\{\boldsymbol{x}_{dt}\hat{\boldsymbol{\beta}} + \hat{\phi}_1 v_{1,d}^{*(b)} + \hat{\phi}_2 v_{2,dt}^{*(b)}\}$.

iii) Generate the response variables $y_{dt}^{*(b)} \sim \text{Poiss}(\nu_{dt} p_{dt}^{*(b)})$.

iv) For each bootstrap resample b, calculate the estimator $\hat{\boldsymbol{\theta}}^{*(b)}$ and the EBP $\hat{p}_{dt}^{*(b)} = \hat{p}_{dt}^{*(b)}(\hat{\boldsymbol{\theta}}^{*(b)})$.

3. Output:

$$mse^*(\hat{p}_{dt}) = \frac{1}{B}\sum_{b=1}^{B}\big(\hat{p}_{dt}^{*(b)} - p_{dt}^{*(b)}\big)^2.$$

5.7 Simulation experiments

This section presents two model-based simulation experiments. The first one studies the the behaviour of the MM fitting algorithm while the second one compares the performance of the two introduced predictors, i.e. the EBP and the plug-in. The response variables are generated independently as $y_{dt}|v_{1,d}, v_{2,dt} \sim \text{Poisson}(\nu_{dt} p_{dt})$, where

$$p_{dt} = \exp\{\beta_0 + x_{dt}\beta_1 + \phi_1 v_{1,d} + \phi_2 v_{2,dt}\},\ x_{dt} = \frac{d + t/T}{D},\ d = 1, \ldots, D,\ t = 1, \ldots, T.$$

The domain random effects, $v_{1,d}\,(d = 1, \ldots, D)$, are generated according to a SAR(1) process, i.e.

$$\boldsymbol{v}_1 = \underset{1 \le d \le D}{\text{col}}(v_{1,d}) = (\boldsymbol{I}_D - \rho \boldsymbol{W})^{-1}\boldsymbol{u}_1,$$

where $\boldsymbol{I}_D$ denotes the $D \times D$ identity matrix, ρ is the autocorrelation parameter, $\boldsymbol{W} = (\omega_{ij})_{i,j=1,\ldots,D}$ is a proximity matrix and $\boldsymbol{u}_1 \sim N(\mathbf{0}, \boldsymbol{I}_D)$. For the $D \times D$ proximity matrix $\boldsymbol{W}$, a 7-diagonal matrix is considered. Let k be the number of diagonals of $\boldsymbol{W}$, then the number of upper and lower diagonals is $m = \lfloor k/2 \rfloor$, where $\lfloor k/2 \rfloor$ denotes integer part of $k/2$. The diagonals are denoted by 1 (main) and j (upper and lower), $j = 2, \ldots, m+1$. The diagonals are constructed in the following way.

- Diagonal 1 (main): if $|i - j| = 0$, then $\omega_{ij} = 0$.
- Diagonal 2 (upper and lower): if $|i - j| = 1$, then $\omega_{ij} = \frac{1}{2} - \big(\frac{1}{2^3} + \frac{1}{2^4}\big)$.

- Diagonals $3-(m+1)$ (upper and lower): if $|i-j| \in \{2,\ldots,m\}$, then $\omega_{ij} = 1/2^{|i-j|+1}$.

Then, the elements of diagonals 1, 2, 3 and 4 (upper and lower) are 0, 5/16, 2/16 and 1/16 respectively. This rule does not apply to the first and the last m rows ($m = 3$). For those rows, the numerators of the diagonal elements are kept fixed (0, 5, 2 and 1) and the denominators are recalculated so that the sum of the row is 1. Using this criterion, the 9×9 7-diagonal matrix $\boldsymbol{W}$ is

$$\boldsymbol{W} = \left(\begin{array}{ccccccccc}
0 & 5/8 & 2/8 & 1/8 & 0 & 0 & 0 & 0 & 0 \\
5/13 & 0 & 5/13 & 2/13 & 1/13 & 0 & 0 & 0 & 0 \\
2/15 & 5/15 & 0 & 5/15 & 2/15 & 1/15 & 0 & 0 & 0 \\
\hline
1/16 & 2/16 & 5/16 & 0 & 5/16 & 2/16 & 1/16 & 0 & 0 \\
0 & 1/16 & 2/16 & 5/16 & 0 & 5/16 & 2/16 & 1/16 & 0 \\
0 & 0 & 1/16 & 2/16 & 5/16 & 0 & 5/16 & 2/16 & 1/16 \\
\hline
0 & 0 & 0 & 1/15 & 2/15 & 5/15 & 0 & 5/15 & 2/15 \\
0 & 0 & 0 & 0 & 1/13 & 2/13 & 5/13 & 0 & 5/13 \\
0 & 0 & 0 & 0 & 0 & 1/8 & 2/8 & 5/8 & 0
\end{array}\right).$$

That 9×9 7-diagonal matrix $\boldsymbol{W}$ can be generalized to a $D \times D$ matrix by repeating $D-6$ times the weights of the central rows (i.e., 0, 5/16, 2/16 and 1/16).

In both simulation experiments, we take $\beta_0 = -3$, $\beta_1 = 0.8$, $\phi_1 = 0.5$, $\phi_2 = 0.5$ and $\nu_{dt} = 100$, $d = 1,\ldots,D$, $t = 1,\ldots,T$. The simulation considers the scenarios $D = 100$ and $T = 4, 8$ for studying the influence of the time periods. For each scenario, it takes $\rho = 0.1, 0.3, 0.5$. This section runs the Monte Carlo simulation experiments with $K = 1000$ iterations.

5.7.1 Simulation 1

The target of Simulation 1 is to check the behaviour of the MM fitting algorithm introduced in Section 5.3. Table 5.7.1 presents the bias and Table 5.7.2 the root mean squared error (RMSE) for the model parameters $\theta \in \boldsymbol{\theta} = \{\beta_0, \beta_1, \phi_1, \phi_2, \rho\}$. As in Chapter 3, this section considers two options for estimating the vector of all model parameters $\boldsymbol{\theta}$. In the first

option (Opt. 1), the vector $\hat{\boldsymbol{\theta}}$ is obtained as a solution of the system of $p+3$ nonlinear equations (5.3.1). In the second option (Opt. 2), $\hat{\rho}$ is given by calculating the Moran's I measure (3.3.3) over the predicted domain random effects under Model T1.

Table 5.7.1: Bias of the MM fitting algorithm under Model ST1.

		$T=4$		$T=8$	
ρ		Opt. 1	Opt. 2	Opt. 1	Opt. 2
0.1	$\hat{\beta}_0$	0.0115	0.0115	0.0171	0.0171
	$\hat{\beta}_1$	-0.0145	-0.0145	-0.0233	-0.0234
	$\hat{\phi}_1$	-0.0219	-0.0230	-0.0240	-0.0250
	$\hat{\phi}_2$	-0.0098	-0.0098	-0.0066	-0.0066
	$\hat{\rho}$	-0.1689	-0.0848	-0.1593	-0.0820
0.3	$\hat{\beta}_0$	0.0196	0.0197	0.0150	0.0151
	$\hat{\beta}_1$	-0.0225	-0.0226	-0.0168	-0.0169
	$\hat{\phi}_1$	-0.0144	-0.0186	-0.0129	-0.0181
	$\hat{\phi}_2$	-0.0099	-0.0099	-0.0087	-0.0087
	$\hat{\rho}$	-0.3236	-0.1986	-0.3441	-0.1865
0.5	$\hat{\beta}_0$	0.0285	0.0285	0.0394	0.0394
	$\hat{\beta}_1$	-0.0387	-0.0387	-0.0666	-0.0664
	$\hat{\phi}_1$	0.0101	0.0059	0.0062	0.0033
	$\hat{\phi}_2$	-0.0123	-0.0123	-0.0111	-0.0111
	$\hat{\rho}$	-0.7090	-0.2856	-0.8400	-0.2601

Both options behave similarly for the fixed effects and the variance parameters. For these parameters, the variance is the most important term of the MSE since bias is much smaller than RMSE. On the other hand, Opt. 2 produces more competitive estimates for the autocorrelation parameter ρ, since it drastically reduces both bias and RMSE. For ρ, bias is the main part of the MSE since it takes similar absolute values to the RMSE. Then, a bias correction by bootstrap might be useful.

Table 5.7.2: Root mean squared error of the MM fitting algorithm under Model ST1.

		$T=4$		$T=8$	
ρ		Opt. 1	Opt. 2	Opt. 1	Opt. 2
0.1	$\hat{\beta}_0$	0.1388	0.1387	0.1291	0.1291
	$\hat{\beta}_1$	0.2337	0.2337	0.2162	0.2161
	$\hat{\phi}_1$	0.0658	0.0659	0.0561	0.0563
	$\hat{\phi}_2$	0.0449	0.0449	0.0289	0.0289
	$\hat{\rho}$	0.1904	0.1129	0.1796	0.1090
0.3	$\hat{\beta}_0$	0.1682	0.1682	0.1554	0.1554
	$\hat{\beta}_1$	0.2824	0.2823	0.2609	0.2609
	$\hat{\phi}_1$	0.0649	0.0647	0.0569	0.0545
	$\hat{\phi}_2$	0.0463	0.0463	0.0309	0.0309
	$\hat{\rho}$	0.3537	0.2144	0.3872	0.2039
0.5	$\hat{\beta}_0$	0.2189	0.2189	0.2100	0.2100
	$\hat{\beta}_1$	0.3736	0.3736	0.3509	0.3509
	$\hat{\phi}_1$	0.0742	0.0650	0.0696	0.0530
	$\hat{\phi}_2$	0.0451	0.0451	0.0307	0.0307
	$\hat{\rho}$	0.7766	0.2998	0.9038	0.2755

5.7.2 Simulation 2

The second simulation experiment investigates the behaviour of the considered p_{dt} predictors for different time instants, T, and autocorrelation parameters, ρ. Specifically, it calculates BP-plug-in, BP, plug-in and EBP. Given the computational burden presented by the BPs (and EBPs) of the target parameter and of the two random effects under Model ST1, the simulation considers their approximated versions (see Section 5.5). The first predictor, BP-plug-in, is obtained from (5.5.8) by using the theoretical vector of model parameters $\boldsymbol{\theta}$. The BPs and EBPs are approximated by generating $S1 = 500$ random variables $\boldsymbol{v}_1^{(s_1)}$ and $S2 = 700$ random variables $v_{2,\ell_T}^{(s2)}$. For the empirical predictors (plug-in and EBP), the model parameters are estimated by using the second option in MM, since it has presented

better results in the previous simulation experiment.

Table 5.7.3 presents the average across domains and time instants of the biases and the RMSEs (both $\times 10^2$) for BP-plug-in, BP, plug-in and EBP. BP and EBP are more competitive than the respective plug-in. Specially in bias, where a significant improvement is achieved. The obtained results also suggest that T does not affect too much the results and that variance is the most important term of the MSE since bias is much smaller than RMSE.

Table 5.7.3: Bias (B) and root mean squared error (RMSE) of the BP-plug-in, BP, plug-in and EBP of p_{dt} under Model ST1 (both $\times 10^2$).

		$T=4$		$T=8$	
ρ	Predictors	B	RMSE	B	RMSE
0.1	BP-plug-in	0.3350	2.8325	0.3318	2.7739
	BP	0.0701	2.7581	0.0689	2.7165
	plug-in	0.3251	2.8248	0.3261	2.7710
	EBP	0.0699	2.7645	0.0687	2.7203
0.3	BP-plug-in	0.3423	2.8445	0.3217	2.7840
	BP	0.0711	2.7655	0.0674	2.7306
	plug-in	0.3406	2.8354	0.3194	2.7851
	EBP	0.0717	2.7723	0.0677	2.7342
0.5	BP-plug-in	0.3363	2.8939	0.3311	2.8125
	BP	0.0692	2.8129	0.0829	2.7562
	plug-in	0.3290	2.8830	0.3282	2.8169
	EBP	0.0699	2.8200	0.0824	2.7595

The system of MM nonlinear equations (5.3.1) is solved by using the R *nleqslv* package. The *mvtnorm* package is also used to generate samples of a SAR(1) process. The computational burden of the first option in MM is much higher. Taking $T = 4$, the average runtime of the first option is 60.4 seconds, while for the second option is 0.1 seconds. On the other hand, regarding the computational burden of the p_{dt} predictors, the EBP is faster than the plug-in. The reason is because the proposed plug-in predictor requires the calculation of two EBPs ($\hat{\boldsymbol{v}}_1$ and $\hat{\boldsymbol{v}}_2$). The average runtimes are 210.7 seconds for the EBP and 320.5 seconds for the plug-in.

5.8 Application to real data

This section presents only an application to real data of wildfires in Galicia during 2007 – 2008, since in the application to poverty data, the domain variance parameter is not significant. The objective of this study is to analyse the target variable *number of wildfires* by forest areas and months. The domains are the forest areas. For each domain, the data set collects the number of wildfires by month. The section takes the months between April to October for being the months with the greatest number of fires. The total number of domains and time instants are $D = 63$ and $T = 14$, respectively.

The response variable at domain d, $d = 1, \ldots, D$, and time t, $t = 1, \ldots, T$, y_{dt}, is explained by some auxiliary variables through an area-level spatio-temporal Poisson mixed model with SAR(1)-correlated domain and independent domain-time effects. Table 5.8.1 presents the MM estimates of the fixed effects under Model ST1. They are obtained by using the second option (see Section 5.7). The same set of auxiliary variables as those used in Section 4.6.2 are selected. This table suggests that *acumRain* is protective, since it causes a decrease in the response variable if it increases and the other variables remain fixed. On the other hand, *averTemp* and *cadHold* are directly related to the response variable since their signs are positive. The three covariates are significant taking $\alpha = 5\%$.

Table 5.8.1: MM estimates of regression parameters under Model ST1.

Variable	Est.	s.e.	z-value	$P(>\|z\|)$
Intercept	0.7736	0.0870	8.8910	< 0.001
acumRain	-0.5207	0.0682	-7.6318	< 0.001
averTemp	0.1507	0.0661	2.2820	0.0225
cadHold	0.2931	0.0605	4.8466	< 0.001

The MM estimates of the variance parameters are $\hat{\phi}_1 = 0.331$ and $\hat{\phi}_2 = 0.696$. Their 95% percentile bootstrap confidence intervals are $(0.130, 0.467)$ and $(0.503, 0.829)$ respectively. The estimate of the autocorrelation parameter is $\hat{\rho} = 0.327$. The Algorithms 8 and 10 are used to test the significance of the variance parameters ϕ_1 and ϕ_2. The obtained bootstrap p-values are 0.01 and 0.00 respectively. The conclusion is that both variance parameters are significantly different from 0. In addition, as the hypothesis test $H_0 : \phi_1 = 0$ is rejected, the Algorithm 9 is applied to test $H_0 : \rho = 0$. The obtained bootstrap p-value is 0.00. On the other hand, the Moran's test yielded p-values lower than 0.05 in 7 of 14 months.

Then, based on the bootstrap p-values and on the Moran's test, it is recommended to use a spatio-temporal Poisson mixed model to analyse wildfires in Galicia by forest areas and months during $2007 - 2008$.

Figure 5.8.1 plots the Pearson residuals of the synthetic estimator, $\hat{\mu}_{dt}^{syn}$, under Model T0 (left) and of the EBP $\hat{\mu}_{dt}$ under Model ST1 (right). A clear improvement is achieved when one uses a more complex model, since the Pearson residuals are closer to 0. In this case, the EBP of μ_{dt} under Model ST1 is more competitive than the plug-in predictor under Model $T2_2$ (see Section 4.6.2), since the empirical MSEs are 0.454 and 1.247 respectively.

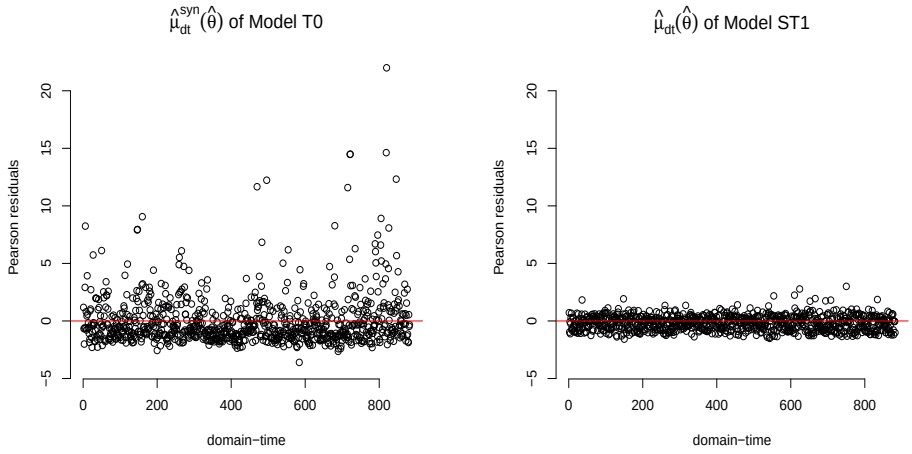

Figure 5.8.1: Pearson residuals of the synthetic estimator under Model T0 (left) and of the EBP of μ_{dt} under Model ST1 (right).

Figure 5.8.2 maps the EBP estimates by forest areas under the area-level Poisson mixed model with SAR(1) domain effects and independent domain-time effects. The same months as those showed in Section 4.6.2 are taken, i.e. the months between August and October. The figure suggests that the highest number of wildfires are concentrated in western coastal areas and in the south of the region. On the other hand, regarding the temporal behaviour, the highest number of fires is found in the months of 2007 (specially in September and October), while in 2008 there was an impressive decrease.

Figure 5.8.3 plots the bootstrap MSE seen in Section 5.6 for the three areas with highest number of fires, i.e. the same forest areas as those shown in Section 4.6.2 (Viana 1, Terra de Tribes and Viana 2). The number of bootstrap replicates is $B = 500$. The mean of the MSEs for the three areas is 3.978 in Model T0 and 1.219 in Model ST1. Then, a clear improvement is achieved when one uses the area-level spatio-temporal Poisson mixed model.

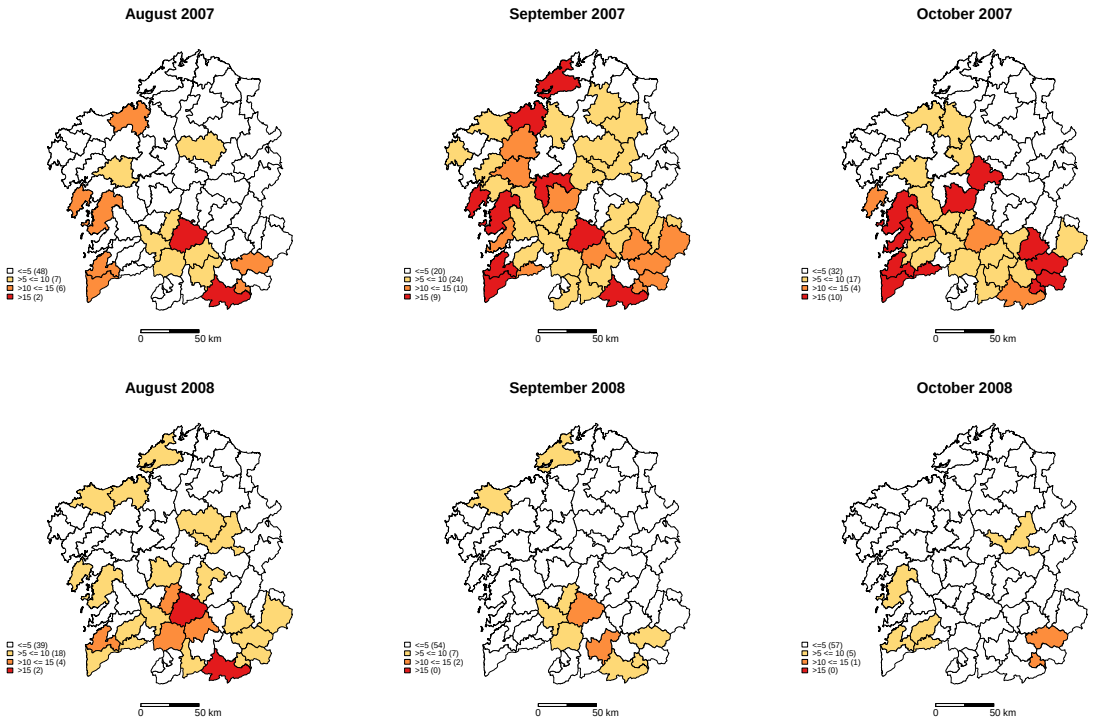

Figure 5.8.2: Estimated fires from August to October in 2007-2008.

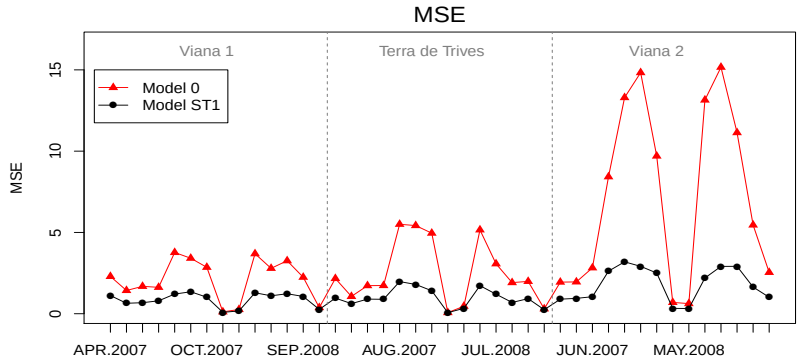

Figure 5.8.3: Bootstrap MSE estimates for the three areas with highest number of fires.

5.9 Concluding remarks

This chapter presents the area-level spatio-temporal Poisson mixed model with SAR(1)-correlated domain and independent domain-time random effects. It incorporates the area-

level Poisson mixed model with SAR(1) domain effects (see Chapter 3) and the area-level Poisson mixed model with independent time effects (see Chapter 4) in a single model.

The MM algorithm is obtained to fit the model parameters. It is based on the method of simulated moments proposed by Jiang (1998). The first simulation experiment is carried out for empirically investigating the behaviour of the MM estimator. Specifically, two options are considered. In the first option, all model parameters are obtained as a solution of the system of $p+3$ nonlinear equations (5.3.1), while in the second option $\hat{\rho}$ is obtained by the Moran's I measure and the remaining parameters as a solution of the system formed by the first $p+2$ equations in (5.3.1). Second option is computationally faster and reduces the bias and the RMSE of the autocorrelation parameter.

Two predictors of the target parameter p_{dt} are proposed and empirically investigated through a simulation experiment. They are the EBP and a plug-in predictor. The simulation experiment concludes that EBP performance better in both bias and RMSE. In addition, the EBP is computationally faster since the plug-in requires calculating two EBPs ($\hat{\boldsymbol{v}}_1$ and $\hat{\boldsymbol{v}}_2$). Given the computational burden of the EBPs under Model ST1, approximated versions are proposed.

The developed methodology is applied to the data set of wildfires in Galicia by forest areas and month during $2007-2008$. Based on the good results obtained in the simulation experiment, the EBP is employed. This estimator is also compared against the synthetic estimator obtained under the model with only fixed effects. A clear improvement is achieved when one uses a more complex model incorporating random effects. In addition, this chapter recommends using an area-level SAR(1) Poisson mixed model with independent time effects to analyse wildfires in Galicia since the Moran's test yielded p-values lower than 0.05 in 7 of 14 months. For these data, the behaviour of the EBP under Model ST1 is also compared against the plug-in under Model $\mathrm{T2}_2$ shown in the application to forest fires in Chapter 4. The EBP under Model ST1 is more competitive for these data, since the obtained empirical MSE (0.454) is lower than the one obtained by the plug-in predictor under Model $\mathrm{T2}_2$ (1.247). The conclusion is that forest fires tend to be concentrated in coastal areas and in the south of the region. An important increase in wildfires is observed in September and October 2007. The introduced bootstrap MSE estimator is considered as accuracy measure of the proposed EBP. It is achieved a clear improvement when using the proposed model and estimator against the classical Poisson regression model.

Chapter 6

Conclusions

Generalized linear mixed models are a useful tools for modelling response variables whose distributions belong to the exponential family. They extend the classical GLMs by incorporating random effects that capture the variability between groups or clusters not explained by the fixed effects. GLMMs include as particular cases logistic, probit and Poisson mixed models. The first two models assume that the response variable is dichotomic, i.e. it takes only the values 0 and 1, while the last model assumes that the response is a count variable. Given their flexibility and ability to combine different sources of information, GLMMs constitute a good tool for treating small area estimation problems. Specifically, this manuscript deals with a particular family of Poisson mixed models in which the response variable is available at the area level. These models are known as area-level Poisson mixed models and they can be seen as the natural extension of the Fay-Herriot model to the context of counting variables. In addition, they are really useful for analysing target variables measured at high levels of disaggregation, since they reduce the error with respect to their conventional competitors.

Several extensions of the basic area-level Poisson mixed model are considered, given rise to increasingly flexible and complex models but, on the other hand, more difficult to estimate. For each area-level Poisson mixed model, the mean parameter μ_d associated with each area d, $d = 1, \ldots, D$, can be expressed as $\nu_d p_d$ because of the relationship between the Poisson and the binomial distributions, where ν_d and p_d are size and probability parameters respectively. This fact provides greater generality to the model, since it allows to obtain both the estimation of the response (count) variable and its proportion. Given that the size parameter ν_d is assumed to be known, this manuscript focusses on the estimation of p_d, which is called target parameter. As an immediate consequence, the estimate $\hat{\mu}_d$ is

obtained as $\nu_d \hat{p}_d$, where $\hat{p}_d$ denotes an estimator of the target parameter.

The considered models are fitted by the MM approach. The MM estimators are based on the method of simulated moments estimator proposed by Jiang (1998) under GLMMs. Unlike ML-based methods that may lead to biased estimators, the MM method gives an alternative consistent estimator. On the other hand, the EBP and a plug-in predictor are employed for estimating the target parameter. The first one is obtained from the corresponding BP by replacing the vector of unknown theoretical model parameters by an estimator. As the BP minimizes the MSE in the set of unbiased predictors, its corresponding empirical version, the EBP, constitutes a competitive alternative for estimating the target parameter p_d. Under regularity conditions, the EBPs have asymptotically the same properties as the BPs. The second one, the plug-in predictor, is obtained by replacing in the theoretical expression of p_d, the unknown parameters by their estimates. As the MM algorithm does not provide predictions of the random effects, they are predicted by their EBPs. To measure the accuracy of the proposed EBPs, this manuscript provides MSE bootstrap algorithms based on the parametric bootstrap procedure given in González-Manteiga et al. (2007). The developed methodology and software are applied in two fields of practical interest: poverty and forest fires.

For the most basic area-level Poisson mixed model (see Chapter 2), the PQL fitting algorithm is also considered. This is a well-known method that maximizes the joint likelihood of the target data and the random effects. Both fitting algorithms (MM and PQL) are empirically investigated through some simulation experiments. Despite the inconsistency, in the developed simulation experiment, PQL performs better for the fixed effects although MM captures the variance parameter more precisely. As PQL algorithm provides estimates of model parameters and predictions of random effects, we consider under this model another plug-in estimator of p_d, where the parameters are estimated by the corresponding PQL estimates (Saei and Chambers, 2003). In addition to the MSE bootstrap, under this model two analytical estimators are also calculated. Both are plug-in estimators without and with bias correction of the second order. The three MSE estimators are analysed in a simulation study. The bias correction term is computationally intensive and the two analytical estimators are quite similar. On the other hand, bootstrap procedure is easy to implement and gives similar results. For applying the developed methodology, two applications to real data are considered. The first one studies the poverty in Spain by provinces and sex (Boubeta et al., 2016b), and the second one analyses the number of fires in Galicia by forest areas (Boubeta et al., 2015). In both cases, the proposed estimators increase the predictive capacity against their classic competitors.

The above model in extended in two ways. First, a SAR(1) correlation structure between domains is allowed and second a temporal component is incorporated to the model. In the first case (see Chapter 3), unlike Model 1 in which the random effects are assumed to be independent, under Model S1 the small area estimates that are spatially close may be alike than estimates for areas that are further apart. The performance of Model S1 depends on the choice of the proximity matrix $\boldsymbol{W}$, which is assumed to be known. Different options are tested: common border, based on distances and based on k-nearest neighbours. In this manuscript, the first option is the one that gives better results. This proximity matrix considers that two areas are neighbours if they share a common delimitation. As the EBP calculation involves complex integrals in high dimension, we propose an approximation. A simulation experiment is carried out for investigating the loss of efficiency when the spatial correlation is not taken into account. The simulation study reveals that estimators under Model 1 are clearly biased and they have little variability, while estimators under Model S1 are unbiased but its variability is greater. The conclusion is that the EBP approximation is competitive since it is unbiased and the MSE is similar to that of Model 1. Two applications to Galician data are considered. The first one uses the EBP approximation for estimating women poverty proportions by counties in 2013 and the second one uses plug-in approximation for estimating the number of forest fires by areas during 2008. In both applications, a clear improvement is achieved when one uses more complex models. In the second case, a temporal component is incorporated (see Chapter 4). Specifically, the chapter provides four area-level Poisson mixed models with time effects and their fitting algorithms. The first two, Model T1 and Model T1_2, use independent time effects and the remaining two, Model T2 and Model T2_2, assume AR(1)-correlated time effects. Simulation experiments show that for capturing the correlation structure under Model T2 it is desirable a sufficiently large number of time instants (it is recommended $T \geq 5$). The simulations also reveal that the EBP is a good alternative to estimate the target parameter p_{dt}, since its bias and MSE are generally lower than the corresponding ones of the plug-in predictor. Two applications to real data focussed on Galicia are used to illustrate the developed methodology. The first one estimates the poverty proportions by counties and gender during $2010-2013$ by using the EBP under a model with independent time effects (Boubeta et al., 2017b) and the second one estimates the number of forest fires by area and month during $2007-2008$ by using a plug-in predictor under a model with AR(1) temporal correlation (Boubeta et al., 2017a). This methodology can be used as an important tool to make preventive policy in the context of forest fires.

Finally, Model ST1 (see Chapter 5) incorporates in a single model the two previous ex-

tensions, i.e. it incorporates both spatial correlation and temporal component. It is the most flexible model but also the most complex. Under this model, the chapter gives the MM fitting algorithm, the EBP and a plug-in predictor of the target parameter, and the EBPs of the two vectors of random effects. Two simulation experiments are carried out to analyse the behaviour of the proposed estimators. They conclude that EBP performs better than the plug-in predictor, since both bias and RMSE are lower in general. The application to real data of wildfires in Galicia shows a clear improvement over the classical Poisson regression model and also over the area-level Poisson mixed with AR(1)-correlated time effects seen in Chapter 4.

For every considered model, all the programme codes have been implemented in R, both the MM fitting algorithm and the two proposed estimators (EBP and plug-in). The computational burden for both estimators under the area-level Poisson mixed models with independent and SAR(1) domain effects is similar. Although the calculation of the estimators is more expensive for Model S1, the proposed approximate version equates the computational times against Model 1. On the other hand, under the temporal models (Model T1 and Model T2), the calculation of the EBP is much faster than the plug-in estimator. The reason is because the plug-in predictor requires the calculation of two EBPs for the random effects.

Appendix A

Resumen en Castellano

Capítulo 1: Introducción

Los modelos lineales mixtos (LMMs) constituyen una generalización de los modelos de regresión lineal contemplando efectos aleatorios. A diferencia de los modelos clásicos en donde se asume que las observaciones son i.i.d. y pertenecen a la misma población, en los modelos mixtos se tiene una estructura multinivel de mayor complejidad. Los LMMs asumen que las observaciones están agrupadas en diferentes niveles, de tal manera que observaciones que pertenecen a diferentes niveles se consideran independientes, mientras que observaciones que pertenecen al mismo nivel se consideran dependientes ya que comparten información de la subpoblación. Un estudio contemporáneo de aplicación de los modelos mixtos es el análisis de datos longitudinales, donde cada serie de tiempo representa un nivel, pero también pueden ser usados para tratar datos con diferentes fuentes de variabilidad, medidas repetidas o problemas de reconstrucción de imágenes.

Los modelos mixtos generalizados (GLMMs) extienden a los LMMs de variable respuesta normal en dos sentidos:

- La distribución de la variable respuesta se asume que pertenece a la familia exponencial.
- Una función, no necesariamente lineal (función link), de la media de la variable respuesta se modela linealmente.

El estimador de máxima verosimilitud (ML) de los parámetros bajo los GLMMs es consistente cuando el número de niveles o clusters, D, tiende a infinito, mientras el número de

observaciones por cluster permanece uniformemente acotado. Una importante ventaja del enfoque ML es que genera estimaciones de todos los parámetros del modelo (efectos fijos y parámetros de la varianza) y predicciones de los efectos aleatorios. La maximización de la log-verosimilitud requiere el uso de algoritmos iterativos tales como Newton-Raphson, Fisher scoring o algoritmos del tipo EM. Dada la complejidad de las integrales involucradas en la función de log-verosimilitud, habitualmente los modelos iterativos se combinan con técnicas de aproximación de integrales tales como Laplace (LA), quasi-verosimilitud restringida (PQL), Gauss-Hermite, métodos Monte Carlo o métodos de integración numérica (reglas del trapecio, Simpson, etc.).

En algunos casos, los estimadores basados en ML pueden conducir a estimadores inconsistentes y sesgados. Por ello, Jiang (1998) propone el método de simulación de momentos (MSM) como enfoque alternativo para estimar los parámetros en un GLMM. Este estimador es computacionalmente atractivo y proporciona estimadores consistentes de los parámetros del modelo.

Dada su flexibilidad y habilidad para combinar diferentes fuentes de información, los modelos mixtos constituyen una herramienta idónea para la estimación en áreas pequeñas (SAE), siendo esta una rama de la estadística que involucra la estimación de parámetros en subconjuntos pequeños (llamados áreas pequeñas o dominios) de una población original. Habitualmente, el término área pequeña se refiere a áreas geográficas con un nivel de desagregación inferior al que se considera en el diseño original (tales como comarcas, municipios o divisiones censales), grupos demográficos (edad $\times$ sexo $\times$ raza), grupos demográficos en una región geográfica, etc. Usualmente los tamaños muestrales en áreas pequeñas son demasiado pequeños, ya que los diseños muestrales se desarrollan para la población original y no para los dominios. En este contexto, los estimadores de los parámetros de la población poseen la precisión deseada a nivel poblacional pero no a nivel dominio.

Los estimadores directos proporcionan estimadores insesgados respecto a la distribución del diseño muestral pero su varianza es elevada en áreas pequeñas. Por ello, uno de los principales cometidos de los investigadores en áreas pequeñas es el de encontrar estimadores más sofisticados de menor variabilidad. Las técnicas de estimación en áreas pequeñas pueden ser de tres tipos: basadas en el diseño, asistidas por modelos o basadas en el modelo. En cuanto al nivel de agregación de la variable respuesta, los modelos de áreas pequeñas pueden ser clasificados en dos grupos: (i) modelos de área y (ii) modelos de individuo. La metodología desarrollada en esta tesis se centra en el primer grupo de modelos.

En la práctica hay un gran número de aplicaciones en modelos de área. Una de las prin-

cipales razones se debe al secreto de confidencialidad, ya que las oficinas de estadística habitualmente no pueden proporcionar información a nivel de individuo pero sí agregada por regiones. Estos modelos han ido evolucionado con el transcurso de los años, adaptándose a las necesidades de los datos, dando lugar a modelos cada vez más flexibles, pero por otro lado, más complejos de estimar. El modelo de área básico es el modelo de Fay-Herriot (FH). Este modelo viene definido en dos etapas. En la primera se asume que la variable respuesta, y_d, asociada al dominio d, $d = 1, \ldots, D$, se puede expresar a través de su media más un término de error,

$$y_d = \mu_d + \varepsilon_d,\ d = 1, \ldots, D,$$

mientras que en la segunda se asume que la media viene dada como suma del término de regresión y un efecto aleatorio u_d, es decir

$$\mu_d = \boldsymbol{x}_d\boldsymbol{\beta} + u_d,\ d = 1, \ldots, D,$$

donde D denota el número total de dominios o áreas, $\boldsymbol{x}_d$ el vector de p variables auxiliares y $\boldsymbol{\beta}$ el vector de coeficientes fijos. Los efectos aleatorios u_d's y los términos de error ε_d's se asumen independientes con $u_d \overset{i.i.d.}{\sim} N(0, \sigma_u^2)$ y $\varepsilon_d \overset{ind}{\sim} N(0, \sigma_d^2)$. Un estimador popular de μ_d bajo el modelo FH es el predictor lineal insesgado óptimo (EBLUP). Este predictor se define como aquel que minimiza el error cuadrático medio (MSE) en el conjunto de predictores insesgados de μ_d. Otros predictores comúnmente utilizados en la práctica son el predictor plug-in, definido por

$$\hat{\boldsymbol{\mu}}^P = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{u}},$$

donde $\hat{\boldsymbol{u}}$ denota un predictor del vector de efectos aleatorios $\boldsymbol{u}$. Una versión simplificada del predictor plug-in es el estimador sintético. Este se diferencia del anterior en que no incorpora una predicción de los efectos aleatorios. Es decir,

$$\hat{\boldsymbol{\mu}}^{syn} = \boldsymbol{X}\hat{\boldsymbol{\beta}}.$$

El modelo FH se puede extender al contexto de los GLMMs asumiendo que la distribución de la variable respuesta, y_d, pertenece a la familia exponencial y que su esperanza, transformada previamente por la función link, g, puede ser modelada linealmente. El trabajo desarrollado a lo largo del manuscrito se centra en el cálculo del predictor óptimo empírico (EBP) bajo los modelos mixtos de Poisson de área. Este estimador constituye la extensión natural del EBLUP al contexto GLMM. Bajo el modelo de Poisson, la función enlace es la función logaritmo, y la variable objetivo es un conteo de eventos de interés por dominios.

A lo largo del manuscrito se llevan paralelamente dos aplicaciones a datos reales. La primera de carácter socio-económico y la segunda de naturaleza medioambiental.

En la primera aplicación se usan dos conjuntos de datos que recogen información de la *Encuesta de Condiciones de Vida*. Esta encuesta proporciona información de los ingresos del hogar en el año previo. Su diseño permite que el tamaño muestral sea lo suficientemente grande para que el estimador directo alcance un cierto grado de precisión a nivel comunidad autónoma, pero no a nivel provincia o comarca. La estimación en áreas pequeñas trata este tipo de problemas introduciendo estimadores indirectos. El primer conjunto de datos viene dado a nivel nacional por provincias durante el año 2008, mientras que el segundo se centra en las comarcas de la comunidad autónoma de Galicia durante el periodo $2010 - 2013$. Las bases de datos a nivel individuo han sido proporcionadas por el Instituto Nacional de Estadística y el Instituto Galego de Estatística respectivamente, y la agregación a nivel de área es de elaboración propia. Las áreas pequeñas (dominios) en ambas bases de datos son las 50 provincias españolas en el primer caso, y las 53 comarcas gallegas en el segundo caso, o el cruce de las provincias y comarcas por sexo. La variable respuesta en ambos casos es el número de personas bajo el umbral de pobreza por dominio. La información auxiliar disponible hace referencia a la proporción de individuos (por sexo) en cada categoría de las siguientes variables.

- Edad: ≤ 15 ($age0$), $16 - 24$ ($age1$), $25 - 49$ ($age2$), $50 - 64$ ($age3$) y ≥ 65 ($age4$).
- Educación: inferior a primaria ($edu0$), primaria ($edu1$), secundaria ($edu2$) y universitaria ($edu3$).
- Nacionalidad: española ($cit0$) y no española ($cit1$).
- Situación laboral: ≤ 15 ($lab0$), empleados ($lab1$), desempleados ($lab2$) e inactivos ($lab3$).

La base de datos utilizada en la aplicación medioambiental a datos reales de incendios forestales en Galicia ha sido proporcionada por el *Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente* del Gobierno de España. La agregación a nivel dominio es de elaboración propia. En este caso los dominios son las áreas forestales. Actualmente, la comunidad autónoma de Galicia se divide en 63 áreas forestales. Para cada dominio y mes, la base de datos contiene información de la variable objetivo *número de incendios* y cierta información auxiliar durante el periodo $2006 - 2008$. Se consideran dos fuentes de información auxiliar dependiendo de su naturaleza. En el primero grupo se incluyen todas aquellas

variables que dependen únicamente del dominio (y no del instante temporal), es decir, son constantes a lo largo del tiempo. En particular, se dispone de las variables *tamaño poblacional*, *número de parcelas catastrales*, *número de titulares catastrales*, *número de unidades ganaderas*, *porcentaje de superficie de matorral*, *porcentaje de superficie húmeda* y *porcentaje de superficie arbolada* por área forestal. En el segundo grupo se incluyen variables promedio obtenidas de estaciones meteorológicas por área forestal y mes. Concretamente, se han considerado *precipitación acumulada*, *temperatura media* y *días sin lluvia*.

El presente trabajo se centra en el desarrollo de técnicas de estimación en áreas pequeñas usando GLMMs de área. Concretamente, se consideran variables respuesta de conteo y, en consecuencia, la familia de modelos se restringe a los modelos de Poisson mixtos de área. Además, se extiende el modelo de Poisson mixto de área básico incorporando correlación espacial, componentes temporales y correlación espacio-temporal.

Capítulo 2: Modelo de Poisson mixto de área

Los modelos de regresión de Poisson y binomial son modelos lineales generalizados (GLMs) que se usan para modelar variables respuesta de conteo (como por ejemplo número de personas bajo el umbral de pobreza). En ocasiones los GLMs no pueden explicar la variabilidad de la variable objetivo a través de las variables auxiliares seleccionadas. Esto puede suceder cuando las observaciones pertenecientes a diferentes dominios son independientes pero existe cierta estructura de dependencia dentro de los dominios. Los GLMMs son una extensión de los GLMs, que capturan la variabilidad entre dominios introduciendo efectos aleatorios. Habitualmente se asume que los efectos aleatorios se distribuyen según una distribución normal.

A pesar de la gran utilidad de los GLMMs, la inferencia basada en estos modelos presenta importantes dificultades, ya que la función de verosimilitud involucra integrales en alta dimensión que no pueden ser evaluadas analíticamente. Varios métodos se han propuesto para abordar este problema, la mayoría de ellos relacionados con el método de Taylor, el método de Laplace para aproximación de integrales o el algoritmo PQL. Desafortunadamente, en algunos casos el método PQL puede conducir a estimadores inconsistentes e insesgados.

En este capítulo se estudia el modelo de Poisson mixto de área básico y se consideran tres procedimientos de ajuste: el método de los momentos (MM) y los algoritmos PQL y LA. El MM obtenido está basado en el método de simulación de momentos propuesto por Jiang (1998). Este método es computacionalmente atractivo y proporciona estimadores insesga-

dos. Los dos últimos se usan únicamente en este capítulo a efectos comparativos. Para su programación, se han utilizando funciones implementadas en el software estadístico R. Sin embargo, la aplicación de estas funciones se restringe únicamente a algunos modelos mixtos más básicos y no cubre todos los procedimientos de ajuste. Para analizar el comportamiento de los diferentes algoritmos de ajuste considerados, se ha desarrollado un estudio de simulación basado en la aplicación a datos reales de pobreza en España. El análisis concluye que los métodos PQL y LA ajustan mejor los efectos fijos, mientras que el MM propuesto captura el parámetro de la varianza con mayor precisión.

Dada la estrecha relación entre las distribuciones binomial y Poisson, se asume que el parámetro μ_d de la Poisson se puede descomponer como producto de la variable exposición o tamaño, ν_d, y una función de probabilidad p_d. Como ν_d se supone conocida, el parámetro p_d determina unívocamente el parámetro de la Poisson μ_d. Por lo tanto, a lo largo de la memoria nos centramos en obtener un estimador de p_d, al que denominamos parámetro objetivo.

Para estimar el parámetro objetivo, hemos obtenido el EBP bajo el modelo de Poisson mixto de área. La metodología estadística fue tomada y adaptada de Jiang and Lahiri (2001) y Jiang (2003), donde se desarrollan EBPs en el contexto de los modelos mixtos logísticos y GLMMs respectivamente. Además de los EBPs se consideran, y analizan empíricamente a través de diversos estudios de simulación, dos estimadores plug-in (usando MM y PQL). A pesar de la inconsistencia del algoritmo PQL, el estimador plug-in de p_d obtenido usando este método de ajuste es competitivo, especialmente cuando el parámetro de la varianza es pequeño. Además, su carga computacional es menor.

Como medida de precisión del EBP propuesto, se considera el MSE. La estimación del MSE bajo los modelos mixtos no es una tarea fácil. En este trabajo hemos adaptado los cálculos del MSE dados por Jiang and Lahiri (2001) y Jiang (2003) al contexto de los modelos de Poisson mixtos de área. Concretamente, proporcionamos dos estimadores analíticos del MSE, con y sin término de corrección de sesgo. Dado que en la práctica los estimadores analíticos del MSE son computacionalmente exigentes, proporcionamos también un estimador bootstrap basado en el procedimiento bootstrap introducido por González-Manteiga et al. (2007) bajo el modelo logístico mixto. Se ha realizado un estudio de simulación para investigar empíricamente el comportamiento de los estimadores del MSE propuestos. El término de corrección de sesgo es computacionalmente intensivo y los resultados obtenidos por los estimadores sesgo corregidos son similares a los obtenidos sin corrección.

Finalmente, se aplica la metodología desarrollada a los datos de la *Encuesta de Condiciones de Vida* del año 2008 y de incendios forestales en Galicia en verano del 2007. En la primera aplicación se propone el EBP para estimar la proporción de pobreza en España, pues sus resultados son más satisfactorios que los obtenidos con el estimador directo. Se concluye que las mayores tasas de pobreza se encuentran en las provincias del sur y del oeste del país. Las estimaciones de la raíz del error cuadrático medio relativo (RRMSE) obtenidas son menores del 20.25% en todas las provincias. En cuanto a la aplicación a datos reales de incendios, se utiliza el estimador plug-in con LA, obteniendo una impactante mejora respecto al estimador sintético bajo el modelo de efectos fijos.

Capítulo 3: Modelo de Poisson mixto con efectos SAR(1)

Cuando las variables auxiliares relacionadas con la variable objetivo están disponibles a nivel de área, el modelo de Poisson mixto de área básico enlaza todos los dominios para mejorar la estimación en un área particular, es decir, toma prestada la fuerza de otras áreas. El modelo básico posee efectos aleatorios que tienen en cuenta la variabilidad entre dominios no explicada por las variables auxiliares. Este modelo asume que los efectos aleatorios del dominio son independientes. Sin embargo, en las aplicaciones socioeconómicas, ambientales y epidemiológicas, las estimaciones de las áreas más cercanas pueden ser más parecidas que las de las áreas más alejadas. De hecho, Cressie (1993) muestra que no emplear modelos espaciales puede conducir a inferencias ineficientes cuando las variables auxiliares no explican la correlación espacial de la variable de estudio.

En SAE, la modelización de la correlación espacial entre diferentes áreas permite tomar prestada la fuerza de las áreas más próximas. Sin embargo, no existen trabajos que se ocupen de cálculo del EBP bajo GLMMs espaciales. Es por ello que en este capítulo se estudia el modelo de Poisson mixto de área con efectos aleatorios correlados según un proceso SAR(1). Bajo este modelo, se obtiene el estimador MM. Su comportamiento se investiga empíricamente a través de un estudio de simulación basado en la aplicación a datos reales de pobreza en Galicia durante 2013.

Para estimar el parámetro objetivo se propone el EBP. Además, dada su complejidad, se propone también una aproximación. Ambos estimadores se analizan empíricamente a través de un estudio de simulación y se comparan frente al plug-in. Concretamente, en el estudio se analiza el BP, EBP y dos estimadores plug-in (bajo el contexto teórico y real). Estos estimadores se comparan, a su vez, frente a los obtenidos bajo el modelo de Poisson mixto de área básico. El objetivo, además de investigar el comportamiento de los

estimadores, es analizar la pérdida de eficiencia cuando no se tiene en cuenta la correlación espacial. Los resultados obtenidos muestran que la aproximación del EBP es competitiva, ya que proporciona estimaciones insesgadas y su MSE es menor que el del original. Los estimadores basados en el modelo de Poisson mixto de área básico muestran un claro sesgo.

Se ha usado la aproximación del EBP para estimar la tasa de pobreza de mujeres en Galicia por comarcas. El estimador propuesto se compara frente a otras alternativas como el estimador sintético o el estimador directo. Los resultados obtenidos sugieren que la aproximación del EBP se comporta mejor. Las estimaciones obtenidas rebelan que los mayores niveles de pobreza en mujeres se concentran en comarcas del sur y del oeste de la comunidad. El porcentaje medio de mujeres bajo el umbral pobreza en Galicia es del 16.89% y este se obtiene con un error medio del 13.23%. En cuanto a la aplicación a datos reales de incendios forestales, se usa el estimador plug-in para modelar el número de incendios forestales en Galicia por áreas forestales durante el verano de 2008. Los resultados obtenidos muestran una clara mejora respecto a las técnicas tradicionales. Las áreas con mayor número de incendios se concentran en el sur y en zonas costeras del norte y oeste de la comunidad. La tasa de error media obtenida es del 25.98%.

Capítulo 4: Modelo de Poisson mixto de área con efectos temporales

En los últimos años se almacena una gran cantidad de información, en particular a lo largo del tiempo, para estudiar el comportamiento de variables de interés. Por otro lado, las técnicas de SAE son importantes cuando el nivel de desagregación es muy alto y los estimadores directos no se comportan tan bien como se necesita. Este capítulo combina estas dos necesidades. Además, se proporciona una metodología estadística para estimar características poblacionales a un nivel suficientemente desagregado por cada instante temporal. Algunos trabajos destacados en la literatura son Rao and Yu (1994), donde se introduce una extensión del modelo FH para datos transversales. Los estimadores propuestos por Pfeffermann and Burck (1990); Ghosh et al. (1996); Datta et al. (2002); Saei and Chambers (2003); You et al. (2001); Esteban et al. (2012a,b) y Marhuenda et al. (2013), entre otros, aprovechan los dos niveles para proporcionar estimaciones en áreas pequeñas con buenas propiedades.

Este capítulo extiende el modelo propuesto en Boubeta et al. (2016b), incluyendo efectos temporales. Concretamente se consideran cuatro modelos. Los dos primeros incorporan efectos temporales independientes y los otros dos asumen un proceso autorregresivo de orden uno. Los modelos resultantes se ajustan por MM. Diversos estudios de simulación se

llevan a cabo para analizar el efecto del número de dominios e instantes temporales en el algoritmo de ajuste. Para cada uno de los modelos anteriores se obtienen los correspondientes EBPs. Además de los EBPs, se proporcionan predictores plug-in y ambos se comparan en un estudio de simulación. Los resultados obtenidos sugieren que el EBP es una buena alternativa, ya que su sesgo y MSE son, en general, menores que los obtenidos con el predictor plug-in. Además, su tiempo computacional es mucho menor que el del plug-in propuesto, que requiere calcular dos EBPs para ambos efectos aleatorios.

Para estimar el MSE del EBP, proponemos un procedimiento bootstrap basado en un bootstrap paramétrico. La metodología desarrollada se aplica a los datos de pobreza en Galicia durante $2010-2013$ y de incendios forestales en Galicia durante $2007-2008$. El objetivo de ambas aplicaciones es estudiar la evolución de la proporción de pobreza por comarca-sexo y el número de incendios forestales por áreas a lo largo del tiempo respectivamente. En ambos casos, el parámetro de la varianza asociado a los efectos aleatorios del dominio es no significativo, por lo que se consideran las versiones simplificadas del modelo temporal. En la aplicación a datos de pobreza se considera el EBP y en la aplicación a datos de incendios se considera el predictor plug-in. En el primer caso el parámetro de correlación temporal es no significativo y por lo tanto el modelo resultante es un modelo simplificado con independencia temporal, mientras que en el segundo se recomienda usar un modelo temporal simplificado con correlación AR(1) para modelar el número de incendios forestales en Galicia. En ambos casos, se obtiene una importante mejora cuando uno usa modelos más complejos incorporando efectos aleatorios.

Capítulo 5: Modelo de Poisson mixto de área con efectos espacio-temporales

Los modelos de Poisson mixtos de área constituyen una buena herramienta para modelar variables respuesta de conteo. Sin embargo, el modelo de Poisson mixto de área básico (Boubeta et al., 2016b) tiene importantes limitaciones, ya que no tiene en cuenta componentes temporales o estructuras de correlación espacial complejas. Por ello, se consideran diversas extensiones del modelo de Poisson mixto de área básico, dando lugar a modelos cada vez mas complejos que permiten adaptarse mejor a las necesidades de los datos reales. Específicamente, en el Capítulo 3 se generaliza el modelo de Poisson mixto básico incorporando una estructura espacial SAR(1) y en el Capítulo 4 se proporcionan varias extensiones al marco temporal. En este capítulo se incorporan ambas extensiones en un único modelo, un modelo de Poisson mixto espacio-temporal, que incorpora tanto correlación espacial

como componente temporal.

Extensiones espacio-temporales del modelo FH han sido propuestas, entre otros, por Singh et al. (2005); Pereira and Coelho (2012) o Marhuenda et al. (2013). Estos autores consideran una estructura de correlación espacial SAR(1) y un proceso AR(1) para la componente temporal. Esteban et al. (2016) presentan un nuevo modelo espacio-temporal suponiendo efectos aleatorios correlados según un proceso MA(1). Los autores citados aplican LMMs espacio-temporales al contexto de estimación en áreas pequeñas. Sin embargo, este manuscrito trata GLMMs en lugar de LMMs.

Para ajustar los parámetros del modelo, se propone el estimador MM y se diseña un estudio de simulación para analizar empíricamente la influencia de los instantes temporales. Además, se estudia la aplicabilidad de los EBPs a la estimación de recuentos y proporciones por dominios. Dada la complejidad computacional de los estimadores obtenidos, se proponen diferentes alternativas y aproximaciones. El estimador propuesto se compara frente a un estimador plug-in. Los resultados obtenidos sugieren que el estimador propuesto tiene un mejor rendimiento ya que arroja, en general, un menor sesgo y RMSE. Además, su tiempo de computación es sustancialmente menor.

La metodología desarrollada en este último capítulo se aplica a los datos de incendios en Galicia por áreas forestales durante el periodo $2007 - 2008$. Se observa una clara mejora cuando uno usa un modelo de mayor complejidad incorporando efectos aleatorios. Además, estos resultados se comparan con los obtenidos bajo el modelo temporal del Capítulo 4. En este caso, recomendamos usar un modelo espacio-temporal para analizar el número de incendios forestales en Galicia durante el citado periodo, ya que proporciona un menor MSE.

Para cada uno de los modelos considerados a lo largo de la memoria, se ha implementado en R tanto el algoritmo de ajuste del método de los momentos como los estimadores del parámetro objetivo propuestos, es decir el EBP y un predictor plug-in. El estimador MM es computacionalmente atractivo y proporciona estimaciones consistentes. En cuanto a los dos estimadores del parámetro objetivo considerados, el EBP tiene, en general, un mejor rendimiento. La carga computacional de ambos estimadores bajo los modelos introducidos en los Capítulos 2 y 3 es similar. Sin embargo, bajo los modelos introducidos en los Capítulos 4 y 5, el EBP es sustancialmente más rápido puesto que el plug-in propuesto requiere calcular dos EBPs (uno por cada efecto aleatorio).

References

Bajocco, S. and Ricotta, C. (2008). Evidence of selective burning in Sardinia (Italy): which land-cover classes do wildfires prefer? *Landscape Ecology*, 23:241–248.

Baldermann, C., Salvati, N., and Schmid, T. (2016). Robust small area estimation under spatial non-stationarity. *Discussion Paper, School of Business and Economics: Economics, No. 2016/5*.

Balsa-Barreiro, J. and Hermosilla, T. (2013). Socio-geographic analysis of wildland fires: causes of the 2006's wildfires in Galicia (Spain). *Forest Systems*, 22:497–509.

Barreal, J., Loureiro, M., and Picos, J. (2011). Estudo da incidencia dos incendios en Galicia: Unha perspectiva socioeconomica. *Revista Galega de Economía*, 20:227–246.

Battese, G. E., Harter, R. M., and Fuller, W. A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 80:28–36.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the flase discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57:289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29:1165–1188.

Boubeta, M., Lombardía, M. J., González-Manteiga, W., and Marey-Pérez, M. (2016a). Burned area prediction with semiparametric models. *International Journal of Wildland Fire*, 25(6):669–678.

Boubeta, M., Lombardía, M. J., Marey-Pérez, M., and Morales, D. (2015). Prediction of forest fires occurrences with area-level poisson mixed models. *Journal of Environmental Management*, 154:151–158.

Boubeta, M., Lombardía, M. J., Marey-Pérez, M., and Morales, D. (2017a). Poisson mixed models for analysing the evolution of forest fires. *(in review)*.

Boubeta, M., Lombardía, M. J., and Morales, D. (2016b). Empirical best prediction under area-level Poisson mixed models. *Test*, 25:548–569.

Boubeta, M., Lombardía, M. J., and Morales, D. (2017b). Poisson mixed models for studying the poverty in small areas. *Computational Statistics and Data analysis*, 107:32–47.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Bruña García, X. and Marey-Pérez, M. (2014). Public participation: a need of forest planning. *IForest*, 7:215–225.

Chambers, R., Chandra, H., and Tzavidis, N. (2011). On bias-robust mean squared error estimation for pseudo-linear small area estimators. *Survey Methology*, 37:153–170.

Chandra, H., Salvati, N., and Chambers, R. (2015). A spatially nonstationary Fay-Herriot model for small area estimation. *Journal of Survey Statistics and Methodology*, 3:109–135.

Chandra, H., Salvati, N., and Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. *Submitted*.

Chandra, H., Salvati, N., Chambers, R., and Tzavidis, N. (2012). Small area estimation under spatial nonstationarity. *Computational Statistics and Data Analysis*, 56:2875–2888.

Choi, J., Lawson, A. B., Cai, B., and Hossain, M. M. (2011). Evaluation of Bayesian spatiotemporal latent models in small area health data. *Environmetrics*, 22(8):1008–1022.

Cressie, N. (1993). *Statistics for spatial data*. Wiley, New York.

Crujeiras, R. M., Fernández-Casal, R., and González-Manteiga, W. (2010). Goodness-of-fit tests for the spatial spectral density. *Stochastic Environmental Research and Risk Assessment*, 24:67–69.

Datta, G. S. (2009). Chapter 32 - model-based approach to small area estimation. In Rao, C., editor, *Handbook of StatisticsSample Surveys: Inference and Analysis*, volume 29, Part B of *Handbook of Statistics*, pages 251 – 288. Elsevier.

Datta, G. S. and Ghosh, M. (1991). Bayesian prediction in linear models: applications to small area estimation. *The Annals of Statistics*, 19:1746–1770.

Datta, G. S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, 10:613–627.

Datta, G. S., Lahiri, P., and Maiti, T. (2002). Empirical Bayes estimation of median income of four-person families by state using time series and cross-sectional data. *Journal of Statistical Planning and Inference*, 102:83–97.

Davison, A. and Hinkley, D. (2007). *Bootstrap methods and their applications*. Cambridge University Press, 9th Ed.

Demidenko, E. (2004). *Mixed models: theory and applications*. Wiley.

Deville, J. C. and Sarndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87(418):376–382.

Díaz-Delgado, R., Lloret, F., and Pons, X. (2004). Spatial patterns of fire occurrence in Catalonia, NE Spain. *Landscape Ecology*, 19:731–745.

Diggle, P., Heagerty, P., Liang, H.-Y., and Zeger, S. (2002). *Analysis of longitudinal data*. Oxford statistical science series.

Esteban, M. D., Morales, D., and Pérez, A. (2016). *Area-level spatio-temporal small area estimation models*. In Analysis of poverty data by small area estimation (ed M. Pratesi). John Wiley and Sons, Ltd, Chichester, UK.

Esteban, M. D., Morales, D., Pérez, A., and Santamaría, L. (2012a). Two area-level time models for estimating small area poverty indicators. *Journal of the Indian Society of Agricultural Statistics*, 66(1):75–89.

Esteban, M. D., Morales, D., Pérez, A., and Santamaría, L. (2012b). Small area estimation of poverty proportions under area-level time models. *Computational Statistics and Data Analysis*, 56:2840–2855.

Eurostat (2016). Europe 2020 indicators - poverty and social exclusion. [Online; access 2017/01/30].

Fay, R. E. and Herriot, R. A. (1979). Estimation of income from small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74:269–277.

Fernández-Casal, R. and Francisco-Fernández, M. (2014). Nonparametric bias-corrected variogram estimation under non-constant trend. *Stochastic Environmental Research and Risk Assessment*, 28:1247–1259.

Fuentes-Santos, I., Marey-Pérez, M., and González-Manteiga, W. (2013). Forest fire spatial pattern analysis in Galicia (NW Spain). *Journal of Environmental Management*, 128:30–42.

Fuller, W. A. (1975). Regression analysis for sample surveys. *Sankhya, Series C*, 37:117–132.

Ghosh, M., Nangia, N., and Kim, D. (1996). Estimation of median income of four-person families: a Bayesian time series approach. *Journal of the American Statistical Association*, 91:1423–1431.

Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. P. (1998). Generalized linear models for small-area estimation. *Journal of the American Statistical Association*, 93:273–282.

Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal (with discussion). *Statistical Science*, 9:55–93.

Gómez-Vázquez, I., Álvarez-Álvarez, P., and Marey-Pérez, M. (2009). Conflicts as enhancers or barriers to the management of privately owned common land: A method to analyze the role of conflicts on a regional basis. *Forest Policy and Economics*, 11:617–627.

González, J., Palahí, M., Tasobares, A., and Pukkala, T. (2006). A fire probability model for forest stands in Catalonia (North-East Spain). *Annals of Forest Science*, 63:169–176.

González, J. and Pukkala, T. (2007). Characterization of forest fires in Catalonia (Northeast Spain). *European Journal of Forest Research*, 126:421–429.

González, J., Trasobares, A., Palahí, M., and Pukkala, T. (2007). Predicting stand damage and tree survival in burned forests in Catalonia (North-East Spain). *Annals of Forest Science*, 64:733–742.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2007). Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. *Computational Statistics and Data Analysis*, 51:2720–2733.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008a). Bootstrap mean squared error of small-area EBLUP. *Journal of Statistical Computation and Simulation*, 78:443–462.

González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., and Santamaría, L. (2008b). Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, 52:5242–5252.

Hall, P. and Maiti, T. (2006). On parametric bootstrap methods for small area prediction. *Journal of the Royal Statistical Society: Series B*, 68:221–238.

Hartley, H. O. and Rao, J. N. (1967). Maximum-likelihood estimation for the mixe analysis of the variance model. *Boiometrika*, 54:93–108.

Hobza, T. and Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, 32(3):661–669.

Holt, D., Smith, T. M. F., and Tomberlin, T. J. (1979). A model-based approach to estimation for small subgroups of a population. *Journal of the American Statistical Association*, 74:405–410.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

Jiang, J. (1998). Consistent estimators in generalized linear models. *Journal of the American Statistical Association*, 93:720–729.

Jiang, J. (2003). Empirical best prediction for small-area inference based on generalized linear mixed models. *Journal of statistical planning and inference*, 111:117–127.

Jiang, J. (2007). *Linear and generalized linear mixed models and their applications*. Springer Series in Statistics.

Jiang, J. and Lahiri, P. (2001). Empirical best prediction for small area inference with binary data. *Annals of the Institute of Statistical mathematics*, 53:217–243.

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

Jiang, J., Lahiri, P., and Wan, S. M. (2002). A unified jackknife theory for empirical best predicion with m-estimation. *Annals of Statistics*, 30:1782–1810.

Johnson, F. A., Chandra, H., Brown, J., and 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. *Journal of Offcial Statistics*, 26 (2):341–359.

Koutsias, N., Arianoutsou, M., Kallimanis, A., and Dimopoulos, P. (2009). Is there any special pattern of the extreme wildland fires occurred in Greece in the summer of 2007? In *52nd International Symposium of the International Association for Vegetation Science, Vegetation Processes and Human Impact in a Changing World, Chania, Greece.*

Lahiri, P. and Meza, J. L. (2002). Small-area estimation. *In A. H. El-Shaarawi and W. W. Piegorsch, Eds., Encyclopedia of Environmetric, John Wiley and Sons, Chichester*, 4:2010–2014.

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

Lehtonen, R. and Veijanen, A. (1998). Logistic generalized regression estimators. *Survey Methodology*, 24(1):51–55.

Lehtonen, R. and Veijanen, A. (2009). Chapter 31 - design-based methods of estimation for domains and small areas. In Rao, C., editor, *Handbook of Statistics Sample Surveys: Inference and Analysis*, volume 29, Part B of *Handbook of Statistics*, pages 219 – 249. Elsevier.

Lin, X. (2007). Estimation using penalized quasilikelihood and quasi-pseudo-likelihood in Poisson mixed models. *Lifetime Data Analysis*, 13:533–544.

López-Vizcaíno, E., Lombardía, M. J., and Morales, D. (2013). Multinomial-based small area estimation of labour force indicators. *Statistical Modelling*, 13(2):153–178.

López-Vizcaíno, E., Lombardía, M. J., and Morales, D. (2015). Small area estimation of labour force indicators under a multinomial model with correlated time and area effects. *Journal of the Royal Statistical Association: Series A*, 178:535–565.

MacNab, Y. C. and Lin, X. (2009). On empirical Bayes penalized quasi-likelihood inference in glmms and in Bayesian disease mapping and ecological modelling. *Computational Statistics and Data Analysis*, 53:2950–2967.

MAPAMA (2017). Ministerio de Agricultura y Pesca, Alimentación y Medio Ambiente. http://www.mapama.gob.es/es/estadistica/temas/estadisticas-ambientales/. [Online; access 2017/02/07].

Marchetti, S., Tzavidis, N., and Pratesi, M. (2012). Non-parametric bootstrap mean squared error estimation for m-quantile estimators of small area averages, quantiles and poverty indicators. *Computational Statistics and Data Analysis*, 56:2889–2902.

Marey-Pérez, M. and Gómez-Vázquez, I. (2010). Modelo para la caracterización del nivel de participación social y la conflictividad en los Montes Vecinales en Mano Común (MVMC) en Galicia. *Spanish Journal of Rural Development*, 1:85–102.

Marey-Pérez, M. and Rodríguez-Vicente, V. (2008). Forest transition in northern Spain: Local responses on large-scale programmes of field-afforestation. *Land Use Policy*, 26:139–156.

Marey-Pérez, M. and Rodríguez-Vicente, V. (2009). Forest transition in Northern Spain: local responses on large-scale programmes of field-afforestation. *Land Use and Policy*, 26:139–156.

Marey-Pérez, M., Rodríguez-Vicente, V., and Álvarez-López, C. (2012). Practical application of multivariant analysis techniques to the forest management of active farmers in the northwest of Spain. *Small scale forestry*, DOI 10.1007/s11842-012-9195-1.

Marey-Pérez, M., Rodríguez-Vicente, V., and Crecente-Maseda, R. (2006). Using GIS to measure changes in temporal and spatial dynamics of forestland: experiences from north-west Spain. *Forestry*, 79:409–423.

Marhuenda, Y., Molina, I., and Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis*, 58:308 – 325. The Third Special Issue on Statistical Signal Extraction and Filtering.

Marhuenda, Y., Morales, D., and Pardo, M. C. (2014). Information criteria for Fay-Herriot model selection. *Computational Statistics and Data Analysis*, 70:268–280.

Marshall, R. J. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society: Series A*, 154:421–441.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, linear, and mixed models*. Wiley.

Molina, I. and Rao, J. N. K. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38(3):369–385.

Molina, I., Salvati, N., and Pratesi, M. (2009). Bootstrap for estimating the MSE of the spatial EBLUP. *Computational Statistics*, 24:441–458.

Montané, F., Casals, P., Taull, M., Lambert, B., and Dale, M. (2009). Spatial patterns of shrub cover after different fire disturbances in the Pyrenees. *Annals of Forest Science*, 66:612–620.

Morales, D., Pagliarella, M. C., and Salvatore, R. (2015). Small area estimation of poverty indicators under partitioned area-level time models. *SORT-Statistics and Operations Research Transactions*, 39(1):19–34.

Moreira, F., Vaz, P., Catry, F., and Silva, J. (2009). Regional variations in wildfire susceptibility of land-cover types in Portugal: implications for landscape management to minimize fire hazard. *International Journal of Wildland Fire*, 18:563–574.

Mouillot, F., Ratte, J., Joffre, R., Mouillot, D., and Rambal, S. (2005). Long-term forest dynamic after land abandonment in a fire prone Mediterranean landscape (central Corsica, France). *Landscape Ecology*, 20:101–112.

Moura, F. A. S. and Migon, H. S. (2002). Bayesian spatial models for small area estimation of proportions. *Statistical Modelling*, 2(3):183–201.

Muckhopadhyay, P. (1998). *Small area estimation in survey sampling*. Narosa Publishing House, New Delhi.

Nunes, M., Vasconcelos, M., Pereira, J., Dasgupta, N., Alldredge, R., and Rego, F. (2005). Land cover type and fire in Portugal: do fires burn land cover selectively. *Landscape Ecology*, 20:661–673.

Opsomer, J. D., Claeskens, G., Ranalli, M. G., Kauermann, G., and Breidt, F. J. (2008). Nonparametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society: Series B*, 70:265–286.

Pereira, L. N. and Coelho, P. S. (2012). Small area estimation using a spatio-temporal linear mixed model. *REVSTAT - Statistical Journal*, 10:285–308.

Petrucci, A. and Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment. *Journal of Agricultural, Biological, and Environmental Statistics*, 11:169–172.

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

Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28:40–68.

Pfeffermann, D. and Burck, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, 16:217–237.

Pfeffermann, D. and Correa, S. (2012). Empirical bootstrap bias correction and estimation of prediction mean square error in small area estimation. *Biometrika*, 99:457–472.

Pinheiro, J. C. and Bates, D. M. (1996). Unconstrained parametrzations for variance-covariance matrices. *Statistics Computing*, 6:289–296.

Prasad, N. G. N. and Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, 85:163–171.

Pratesi, M. (2016). *Analysis of poverty data by small area estimation.* Wiley Series in Survey Methodology.

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.

Purcell, N. J. and Kish, L. (1979). Estimates for small domain. *Biometrics*, 35:365–384.

Rao, J. N. K. (1986). Synthetic estimators, SPREE and the best model based predictors. *In proceedings of the Conference on Survey Methods in Agriculture, pp 1-16. U. S. Dept. of Agricultur, Washington, D. C.*

Rao, J. N. K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25:175–186.

Rao, J. N. K. (2003). Small area estimation. *John Wiley and Sons, New York.*

Rao, J. N. K. (2008). Some methods for small area estimation. *Revista Internazionale di Siencze Sociali*, 4:387–406.

Rao, J. N. K. and Molina, I. (2015). *Small area estimation.* Second Edition, Wiley, Hoboken, New Jersey.

Rao, J. N. K. and Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 22:511–528.

Rodríguez, P., Anca, A., Diéguez, A., Arenas, S., and Amann, G. (2013). A necesidade dunha xestión eficiente do monte galego. *Grial Dialogos*, 3:26–35.

Rodríguez-Vicente, V. and Marey-Pérez, M. (2009a). Land-use and land-base patterns in non-industrial private forests: Factors affecting forest management in Northern Spain. *Forest Policy and Economics*, 7:475–490.

Rodríguez-Vicente, V. and Marey-Pérez, M. (2009b). Characterization of nonindustrial private forest owners and their influence on forest management aims and practices in Northern Spain. *Small-Scale Forestry*, 8:479–513.

Rodríguez-Vicente, V. and Marey-Pérez, M. (2010). Analysis of individual private forestry in Northern Spain according to economic factors related to management. *Journal of Forest Economics*, 16:269–295.

Rupert, D., Wand, M., and Carroll, R. (2008). *Semiparametric regression*. Cambridge series in statistical and probabilistic mathematics.

Saei, A. and 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., and Wretman, J. (1992). *Model assisted survey sampling*. Springer.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance components*. Wiley series in probability and statistics.

Sebastian-Lopez, A., Salvador-Civil, R., Gonzalo-Jimenez, J., and SanMiguel-Ayanz, J. (2008). Integration of socio-economic and environmental variables for modelling long-term fire danger in southern Europe. *European Journal of Forest Research*, 127:149–163.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer Series in Statistics.

Singh, B., Shukla, G., and Kundu, D. (2005). Spatio-temporal models in small area estimation. *Survey Methodology*, 31:183–195.

Sugasawa, S., Kawakubo, Y., and Ogasawara, K. (2015). Geographically weighted empirical Bayes estimation via natural exponential family. *Discussion Paper No. 2015-01. Tokyo Institute of Technology.*

Tzavidis, N., Salvati, N., Pratesi, M., and Chambers, R. (2008). M-quantile models with application to poverty mapping. *Statistical Methods and Applications*, 17:393–411.

Ugarte, M. D., Goicoa, T., and Militino, A. F. (2010). Spatio-temporal modeling of mortality risk using penalized splines. *Environmetrics*, 21:270–289.

Ugarte, M. D., Goicoa, T., Militino, A. F., and Durbán, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics and Data Analysis*, 53(10):3616 – 3629.

Ugarte, M. D., Ibáñez, B., and Militino, A. F. (2006). Modelling risks in disease mapping. *Statistical methods in medical research*, 15:21–35.

You, Y., Rao, J. N. K., and Gambino, J. (2001). Model-based unemployment rate estimation for the Canadian labour force survey: A hierarchical Bayes approach. Technical report, Household Survey Methods Division, Statistics Canada.

You, Y. and Zhou, Q. M. (2011). Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology*, 37:25–37.