10.1071/WF23162
International Journal of Wildland Fire

Supplementary Material

Pattern recognition and modelling of virulent wildfires in Spain

María Bugallo[A,*], *María Dolores Esteban*[A], *Manuel Francisco Marey-Pérez*[B] and *Domingo Morales*[A]

[A]Center of Operations Research, Miguel Hernández University of Elche, Elche, Alicante, Spain

[B]Higher Polytechnic Engineering School, University of Santiago de Compostela, Santiago de Compostela, A Coruña, Spain

*Correspondence to: Email: mbugallo@umh.es

Appendix A Clustering algorithm: K-means

Pattern recognition deals with the construction of mechanisms capable of extracting relevant information and key patterns from sample observations. That is, the identification of regularities in the data, to impose a set of identity (classification, clustering, association, etc.) or dependency (regression) relationships. Cluster analysis, or simply clustering, is the task of grouping a set of observations in such a way that observations in the same group (cluster) are more similar (in a certain sense) to each other than those in other groups. The aim of these techniques is to form groups in order to recognise patterns or structures within the general population. Clustering itself is not a specific algorithm, but the general task to be solved. This appendix presents the K-means algorithm, the clustering method used in the data analysis and which was first proposed by Hartigan and Wong (1979).

The K-means algorithm finds $k \in \mathbb{Z}^{\geqslant 1}$ clusters (fixed value), around a given set of centres $\{\boldsymbol{m}_1^{(1)}, ..., \boldsymbol{m}_k^{(1)}\}$ which define the initial clusters $S_1^{(1)}, ..., S_k^{(1)}$, by iterating the following steps:

1. Assign each observation $\boldsymbol{\mathcal{X}}_p$ to a single cluster, being the one with the closest mean:

$$S_\ell^{(i)} = \left\{ \boldsymbol{\mathcal{X}}_p \ : \| \ \boldsymbol{\mathcal{X}}_p - \boldsymbol{m}_\ell^{(i)} \ \|_2^2 \leqslant \| \ \boldsymbol{\mathcal{X}}_p - \boldsymbol{m}_h^{(i)} \ \|_2^2, \ h = 1, ..., k \right\}, \quad \ell = 1, ..., k.$$

   It is imposed that $\boldsymbol{\mathcal{X}}_p$ is assigned to exactly one $S_\ell^{(i)}$, $\ell = 1, ..., k$, although it could be in two or more.

2. For each cluster, calculate the means that will be used as centres of the new clusters:

$$\boldsymbol{m}_\ell^{(i+1)} = \frac{1}{|S_\ell^{(i)}|} \sum_{\boldsymbol{\mathcal{X}}_h \in S_\ell^{(i)}} \boldsymbol{\mathcal{X}}_h, \quad \ell = 1, ..., k.$$

3. Update $i \leftarrow i + 1$.

The algorithm converges when the assignments no longer change. However, the iterative refinement process ends when the maximum number of iterations allowed is reached.

Appendix B Area-level zero-inflated Gamma mixed model

This appendix describes the area-level zero-inflated Gamma (aZIG) mixed model used in the data analysis. All mathematical steps are detailed, justifying the soundness of what is presented. The formulation of the model is given in an orderly fashion, followed by the description of the Laplace approximation algorithm. Subsequently, the expression of the plug-in predictor of the target quantities is provided. Given that the focus of our research is of an applied nature and this predictor achieves good results when applied to real data, more complex predictors with more sophisticated theoretical properties, such as asymptotic unbiasedness, were not investigated. Finally, boostrap inference techniques are included to calculate confidence intervals (CI) of the model parameters and estimate the mean squared error (MSE) of the predictors.

Model

The model is proposed below in a general form. However, it is particularised for application to aggregated fire data in weeks and provinces where appropriate. Let us consider a continuous random variable y_{ijk} taking values on $[0, \infty)$, where $i \in \mathbb{I} = \{1, \ldots, I\}$, $j \in \mathbb{J} = \{1, \ldots, J\}$ and $k \in \mathbb{K} = \{1, \ldots, K\}$. Let $D = IJK$ be the total possible y-values. For instance, y_{ijk} could be the total burned area (in Ha) of a territory during a time period, or its value averaged over the number of reported forest fires. The indexes i, j and k might represent the year, week and province, so D would be the sum of domains defined by the crosses of these categories. As explained before, the target variable is posed for $K = 41$ Spanish provinces, during $J = 18$ weeks (between the 27 and 44th weeks of the year) and $I = 9$ years. Therefore, we deal with $D = IJK = 6642$ domains and work at area-level to model and predict y_{ijk}. Let z_{ijk}, $\boldsymbol{x}_{1,ijk} = (x_{1,ijk1}, \ldots, x_{1,ijkq_1})$ and $\boldsymbol{x}_{2,ijk} = (x_{2,ijk1}, \ldots, x_{2,ijkq_2})$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, be latent (non observable) variables, and $1 \times q_1$ and $1 \times q_2$ row vectors of explanatory variables, respectively. Let us define the vectors and matrices

$$\boldsymbol{y}_{jk} = \underset{1\leqslant i\leqslant I}{\text{col}}(y_{ijk}), \;\; \boldsymbol{z}_{jk} = \underset{1\leqslant i\leqslant I}{\text{col}}(z_{ijk}), \;\; \boldsymbol{X}_{1,jk} = \underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{x}_{1,ijk}), \;\; \boldsymbol{X}_{2,jk} = \underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{x}_{2,ijk}),$$

$$\boldsymbol{y} = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{y}_{jk})), \quad \boldsymbol{z} = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{z}_{jk})),$$

$$\boldsymbol{X}_1 = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{X}_{1,jk})), \quad \boldsymbol{X}_2 = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(\boldsymbol{X}_{2,jk})).$$

Let be $\boldsymbol{u}_{jk} = (u_{1,jk}, u_{2,jk})'$, with $u_{1,jk}$, $u_{2,jk}$ independent $N(0,1)$ random effects, and

$$\boldsymbol{u}_1 = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(u_{1,jk})) \sim N_{JK}(\boldsymbol{0}, \boldsymbol{I}), \, \boldsymbol{u}_2 = \underset{1\leqslant j\leqslant J}{\text{col}}(\underset{1\leqslant k\leqslant K}{\text{col}}(u_{2,jk})) \sim N_{JK}(\boldsymbol{0}, \boldsymbol{I}), \, \boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2')'.$$

The vectors (y_{ijk}, z_{ijk}), $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, follow an area-level zero-inflated Gamma (aZIG) mixed model with random intercepts on jk crossings (week per province) if

$$\begin{aligned} & z_{ijk} \sim \text{BE}(p_{ijk}), \;\; P(y_{ijk} = 0/z_{ijk} = 1) = 1, \\ & f(y_{ijk} = t/z_{ijk} = 0) = \exp\left\{-\nu\mu_{ijk}^{-1} y_{ijk} - \nu \log \mu_{ijk} + (\nu - 1)\log y_{ijk} + \nu \log \nu - \log \gamma(\nu)\right\}, \end{aligned}$$

where $t > 0$, $0 < p_{ijk} < 1$, $\nu > 0$, $\mu_{ijk} > 0$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, and p_{ijk} and μ_{ijk} depend on the explanatory variables $\boldsymbol{x}_{1,ijk}$ and $\boldsymbol{x}_{2,ijk}$, on the regression parameters $\boldsymbol{\beta}_1 = (\beta_{11}, \ldots, \beta_{1q_1})'$ and $\boldsymbol{\beta}_2 = (\beta_{21}, \ldots, \beta_{2q_2})'$ and on the standard deviations ϕ_1, $\phi_2 > 0$ by means of the link functions

$$\begin{aligned} \text{logit}(p_{ijk}) &= \log \frac{p_{ijk}}{1 - p_{ijk}} = \boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk} = \sum_{\ell=1}^{q_1} x_{1,ijk\ell}\beta_{1\ell} + \phi_1 u_{1,jk}, \\ \log(\mu_{ijk}) &= \boldsymbol{x}_{2,ijk}\boldsymbol{\beta}_2 + \phi_2 u_{2,jk} = \sum_{\ell=1}^{q_2} x_{2,ijk\ell}\beta_{2\ell} + \phi_2 u_{2,jk}, \; i \in \mathbb{I}, j \in \mathbb{J}, k \in \mathbb{K}. \end{aligned}$$

To complete the definition, and conditioned to $\boldsymbol{u}$, it is assumed that the vectors $(y_{ijk}, z_{ijk})'$,

$i \in \mathbb{I},\ j \in \mathbb{J},\ k \in \mathbb{K}$, are independent. Inverting the above functions, it follows that

$$p_{ijk} = \frac{\exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\}}{1 + \exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\}},\ \mu_{ijk} = \exp\{\boldsymbol{x}_{2,ijk}\boldsymbol{\beta}_2 + \phi_2 u_{2,jk}\},\ i \in \mathbb{I},\ j \in \mathbb{J},\ k \in \mathbb{K}.$$

In short, the model aZIG is a mixture model of two mixed submodels. The BE-submodel drives the mixture and incorporates the information derived from the excess of zeros. The GA-submodel deals with strictly positive target values using the Gamma distribution with means $\mu_{ijk} > 0$ and constant shape $\nu > 0,\ i \in \mathbb{I},\ j \in \mathbb{J},\ k \in \mathbb{K}$.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \phi_1, \phi_2)'$ be the vector of model parameters and define $\xi_{ijk} = I_{\{0\}}(y_{ijk})$. The components of the marginal distribution are

$$\begin{aligned}
& g(y_{ijk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}) = \xi_{ijk} p_{ijk} \\
+ \quad & (1-\xi_{ijk})\Big[(1-p_{ijk})\exp\Big\{-\nu\mu_{ijk}^{-1}y_{ijk} - \nu\log\mu_{ijk} + (\nu-1)\log y_{ijk} + \nu\log\nu - \log\gamma(\nu)\Big\}\Big] \\
= \quad & \big(1+\exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\}\big)^{-1}\Big\{\xi_{ijk}\exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\} \\
+ \quad & (1-\xi_{ijk})\exp\Big\{-\nu y_{ijk}\exp\{-\boldsymbol{x}_{2,ijk}\boldsymbol{\beta}_2 - \phi_2 u_{2,jk}\} - \nu(\boldsymbol{x}_{2,ijk}\boldsymbol{\beta}_2 + \phi_2 u_{2,jk}) \\
+ \quad & (\nu-1)\log y_{ijk} + \nu\log\nu - \log\gamma(\nu)\Big\}\Big\},\ i \in \mathbb{I},\ j \in \mathbb{J},\ k \in \mathbb{K}. \qquad \text{(B.1)}
\end{aligned}$$

By the independence assumptions, it follows that

$$g(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta}) = \prod_{j=1}^{J}\prod_{k=1}^{K} g(\boldsymbol{y}_{jk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}), \quad g(\boldsymbol{y}_{jk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}) = \prod_{i=1}^{I} g(y_{ijk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}).$$

The likelihood and log-likelihood functions of model aZIG are, respectively,

$$g(\boldsymbol{y};\boldsymbol{\theta}) = \int_{\mathbb{R}^{2JK}} g(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta}) f_u(\boldsymbol{u})\, d\boldsymbol{u} = \prod_{j=1}^{J}\prod_{k=1}^{K}\int_{\mathbb{R}^2}\prod_{i=1}^{I} g(y_{ijk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}) f_{N_2(0,I)}(\boldsymbol{u}_{jk})\, d\boldsymbol{u}_{jk}, \qquad \text{(B.2)}$$

$$\ell(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{j=1}^{J}\sum_{k=1}^{K}\log\int_{\mathbb{R}^2}\prod_{i=1}^{I} g(y_{ijk}|\boldsymbol{u}_{jk};\boldsymbol{\theta}) f_{N_2(0,I)}(\boldsymbol{u}_{jk})\, d\boldsymbol{u}_{jk}. \qquad \text{(B.3)}$$

Given $\boldsymbol{y}$, the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$ can be calculated as follows

$$\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta} \in \Theta}\, \ell(\boldsymbol{\theta}; \boldsymbol{y}), \quad \Theta = \mathbb{R}^{q_1+q_2} \times \mathbb{R}_+^2, \quad \mathbb{R}_+ = (0, \infty).$$

By (B.3), the maximisation of $\ell(\boldsymbol{\theta}; \boldsymbol{y})$ involves integrals in $\mathbb{R}^2$. One way to solve it is to apply two functions sequentially. First, the integral on $\boldsymbol{u}_{jk}$ needs to be calculated and then, the maximisation on $\boldsymbol{\theta}$ could be performed. A maximisation method is described below.

Laplace approximation algorithm

This section describes the Laplace approximation of the loglikelihood function of model aZIG and the algorithm to calculate the ML estimators of the model parameters and to obtain modal predictors of the random effects (Kristensen et al, 2016; Brooks et al, 2017; Morales et al., 2021). First, the likelihood function of model aZIG is

$$g(\boldsymbol{y}; \boldsymbol{\theta}) = \int_{\mathbb{R}^{2JK}} g(\boldsymbol{y}|\boldsymbol{u}; \boldsymbol{\theta}) f_u(\boldsymbol{u})\, d\boldsymbol{u} = \int_{\mathbb{R}^{2JK}} \exp\left\{h(\boldsymbol{u}; \boldsymbol{y}, \boldsymbol{\theta})\right\} d\boldsymbol{u}, \tag{B.4}$$

where

$$h(\boldsymbol{u}; \boldsymbol{y}, \boldsymbol{\theta}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} \log g(y_{ijk}|\boldsymbol{u}_{jk}; \boldsymbol{\theta}) - \frac{2JK}{2}\log 2\pi - \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K}(u_{1,jk}^2 + u_{2,jk}^2).$$

To apply the Laplace approximation to the integral in (B.4), we have to maximize $h(\boldsymbol{u}; \boldsymbol{y}, \boldsymbol{\theta})$ in $\boldsymbol{u}$, given $\boldsymbol{y}$ and $\boldsymbol{\theta}$. For simplicity, we write $h(\boldsymbol{u})$. We can carry out the maximization by applying an R function of optimization. Alternatively, we can implement a Newton-Raphson algorithm after calculating the first and second partial derivatives of h with respect to $u_{1,jk}$ and $u_{2,jk}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, given $\boldsymbol{y}$ and $\boldsymbol{\theta}$. Let $\dot{h}$ and $\ddot{h}$ denote the $2JK \times 1$ vector and the $2JK \times 2JK$ matrix of first and second order partial derivatives of $h(\boldsymbol{u})$ with respect to $\boldsymbol{u}$, respectively. The Newton-Raphson updating equation is

$$\boldsymbol{u}^{(i+1)} = \boldsymbol{u}^{(i)} - \ddot{h}^{-1}(\boldsymbol{u}^{(i)})\,\dot{h}(\boldsymbol{u}^{(i)}). \tag{B.5}$$

Let us denote by $\boldsymbol{u}^{\circ}$ the argument of maxima of $h(\boldsymbol{u})$. It holds $\dot{\boldsymbol{h}}(\boldsymbol{u}^{\circ}) = \mathbf{0}$ and the matrix $\ddot{h}(\boldsymbol{u}^{\circ})$ is negative definite. The loglikelihood of model aZIG can be approximated by

$$\log P(\boldsymbol{y};\boldsymbol{\theta},) \approx 2JK\log 2\pi + h(\boldsymbol{u}^{\circ}) - \frac{1}{2}\log|-\ddot{h}(\boldsymbol{u}^{\circ})| \triangleq \psi(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}^{\circ}).$$

The following step is to maximize $\psi(\boldsymbol{\theta};\boldsymbol{y},\boldsymbol{u}^{\circ})$ in $\boldsymbol{\theta} \in \Theta$. For simplicity, we write $\psi(\boldsymbol{\theta})$. Once again, a suitable option is to apply a Newton-Raphson algorithm after calculating the first and second partial derivatives of g with respect to the components of $\boldsymbol{\theta}$, given $\boldsymbol{y}$ and $\boldsymbol{u}^{\circ}$. Let us define $M = \dim(\Theta) = q_1 + q_2 + 2$. Let $\dot{\psi}$ and $\ddot{\psi}$ denote the $M \times 1$ vector and the $M \times M$ matrix of first and second order partial derivatives of $g(\boldsymbol{\theta})$, respectively. The Newton-Raphson updating equation is

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \ddot{\psi}^{-1}(\boldsymbol{\theta}^{(i)})\,\dot{\psi}(\boldsymbol{\theta}^{(i)}). \tag{B.6}$$

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

1. Set the initial values $i = 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $\varepsilon_3 > 0$, $\varepsilon_4 > 0$, $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\theta}^{(-1)} = \boldsymbol{\theta}^{(0)} + \mathbf{1}$, $\boldsymbol{u}^{(0)} = \mathbf{0}$, $\boldsymbol{u}^{(-1)} = \mathbf{1}$, where $\mathbf{0}$ and $\mathbf{1}$ are column vectors of zeros and ones, respectively.

2. Until $\|\boldsymbol{\theta}^{(i)} - \boldsymbol{\theta}^{(i-1)}\|_2 < \varepsilon_1$, $\|\boldsymbol{u}^{(i)} - \boldsymbol{u}^{(i-1)}\|_2 < \varepsilon_2$, do

   (a) Apply algorithm (B.5) with seeds $\boldsymbol{u}^{(i)}$, convergence tolerance ε_3 and $\boldsymbol{\theta} = \boldsymbol{\theta}^{(i)}$ fixed. Output: $\boldsymbol{u}^{(i+1)}$.

   (b) Apply algorithm (B.6) with seed $\boldsymbol{\theta}^{(i)}$, convergence tolerance ε_4 and $\boldsymbol{u} = \boldsymbol{u}^{(i+1)}$ fixed. Output: $\boldsymbol{\theta}^{(i+1)}$.

(c) Update $i \leftarrow i+1$.

3. Output: $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{(i)}$ and $\hat{\boldsymbol{u}} = \boldsymbol{u}^{(i)}$.

The output includes the ML estimators of the model parameters, $\hat{\boldsymbol{\theta}}$, and the modal predictors of the random effects, $\hat{\boldsymbol{u}}$. Taking into account the consistency and asymptotic normality of the ML estimators, $\hat{\boldsymbol{\theta}} \sim N_M(\boldsymbol{\theta}, \boldsymbol{Q}(\boldsymbol{\theta}))$, it is possible to approximate the asymptotic covariance matrix. It should be remembered that it is the inverse of the Fisher information matrix. In practice, we use the Hessian matrix. That is, the asymptotic variance matrix of $\hat{\boldsymbol{\theta}}$, $\boldsymbol{Q}(\boldsymbol{\theta})$, can be approximated as $\boldsymbol{Q}(\boldsymbol{\theta}) \approx -\ddot{\psi}^{-1}(\hat{\boldsymbol{\theta}})$. This allows the calculation of Wald statistics to test hypotheses about the model parameters. Further, an asymptotic CI at the $1-\alpha$ level for a component θ_ℓ of $\boldsymbol{\theta}$ is $\hat{\theta}_\ell \pm z_{1-\alpha/2}\, q_{\ell\ell}^{1/2}$, $\ell = 1, \ldots, M$, where $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^{\kappa}$, $\boldsymbol{Q}(\boldsymbol{\theta}^{\kappa}) = (q_{ab})_{a,b=1,\ldots,M}$, κ is the last iteration of the Laplace algorithm and z_α is the α-quantile of the $N(0,1)$ distribution. For the regression parameters $\beta_{a\ell}$, $a = 1, 2$, $\ell = 1, \ldots, q_a$, we can give asymptotic p-values to test significance. For example, if $\hat{\beta}_{1\ell} = \beta_0$, the p-value to test $H_0: \beta_{1\ell} = 0$ is

$$p\text{-value} = 2P_{H_0}(\hat{\beta}_{1\ell} > |\beta_0|) = 2P(N(0,1) > |\beta_0|/\sqrt{q_{\ell\ell}}), \quad \ell = 1, \ldots, q_1.$$

To test $H_0: \beta_{2\ell} = 0$, we apply the same procedure but using $q_{q_1+\ell\, q_1+\ell}$ instead of $q_{\ell\ell}$.

Predictors

After introducing model aZIG and a sound fitting algorithm, we are going to provide predictors of the target quantities. This is done by predicting the expected value of a non-negative response variable, accounting for excess zeros and area-level aggregation. In this sense, it can be used to model the target variable and to study the dependence relationships with a set of auxiliary variables, but it is also a forecasting tool. In mathematical terms, the inference is focused on the expected values $\mu_{yijk} \triangleq E[y_{ijk}|\boldsymbol{u}_{jk}] = (1-p_{ijk}(u_{1,jk}))\mu_{ijk}(u_{2,jk})$, where

$$\begin{aligned} p_{ijk}(u_{1,jk}) &= \frac{\exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\}}{1+\exp\{\boldsymbol{x}_{1,ijk}\boldsymbol{\beta}_1 + \phi_1 u_{1,jk}\}}, \\ \mu_{ijk}(u_{2,jk}) &= \exp\{\boldsymbol{x}_{2,ijk}\boldsymbol{\beta}_2 + \phi_2 u_{2,jk}\}, \quad i \in \mathbb{I}, j \in \mathbb{J}, k \in \mathbb{K}. \end{aligned}$$

By plugging ML estimators and modal predictors, the plug-in predictor of μ_{yijk} is

$$\hat{\mu}_{yijk}^{in} = \left(1+\exp\{\boldsymbol{x}_{1,ijk}\hat{\boldsymbol{\beta}}_1 + \hat{\phi}_1\hat{u}_{1,jk}\}\right)^{-1}\exp\{\boldsymbol{x}_{2,ijk}\hat{\boldsymbol{\beta}}_2 + \hat{\phi}_2\hat{u}_{2,jk}\}, \quad i \in \mathbb{I}, j \in \mathbb{J}, k \in \mathbb{K}.$$

According to the applied cut-off of the current research, the plug-in predictor will be adequate to achieve our goals. Its ease of interpretation and calculation, as well as its computational performance and execution times, are unsurpassed. Moreover, it provides successful results in the application to the Spanish provincial data for weeks 27-44 and years 2007-2015. Years 2007-2014 are used for model fitting and 2015 is reserved for prediction.

Bootstrap inference

In this section we formalise how to compute bootstrap-based CIs for the model parameters and bootstrap estimates of the MSE of the predictors, and of the quantiles of the bootstrap distribution of the predictions. In all cases, we rely on bootstrap resampling methods.

Confidence intervals for model parameters

Let θ_ℓ be a component of the vector of model parameters $\boldsymbol{\theta}$. Let $\alpha \in (0,1)$. The following procedure calculates a $(1-\alpha)\%$ percentile bootstrap CI for θ_ℓ. Algorithm A has the following steps:

1. Fit the model to the sample and calculate the ML estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \hat{\phi}_1, \hat{\phi}_2)'$.
2. Repeat B times $(b = 1, \dots, B)$:

(a) For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, generate $u_{1,jk}^{*(b)} \sim N(0,1)$, $u_{2,jk}^{*(b)} \sim N(0,1)$ and calculate

$$\begin{aligned} p_{ijk}^{*(b)} &= \exp\left\{\boldsymbol{x}_{1,ijk}\hat{\boldsymbol{\beta}}_1 + \hat{\phi}_1 u_{1,jk}^{*(b)}\right\}\left(1+\exp\left\{\boldsymbol{x}_{1,ijk}\hat{\boldsymbol{\beta}}_1 + \hat{\phi}_1 u_{1,jk}^{*(b)}\right\}\right)^{-1}, \\ \mu_{ijk}^{*(b)} &= \exp\left\{\boldsymbol{x}_{2,ijk}\hat{\boldsymbol{\beta}}_2 + \hat{\phi}_2 u_{2,jk}^{*(b)}\right\}. \end{aligned}$$

(b) Generate $z_{ijk}^{*(b)} \sim \text{BE}(p_{ijk}^{*(b)})$. If $z_{ijk}^{*(b)} = 1$, do $y_{ijk}^{*(b)} = 0$. If $z_{ijk}^{*(b)} = 0$, generate $y_{ijk}^{*(b)} \sim GA(\mu_{ijk}^{*(b)}, \nu)$.

(c) Based on the sample $(y_{ijk}^{*(b)}, \boldsymbol{x}_{ijk})$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate the ML estimate $\hat{\theta}_\ell^{*(b)}$.

3. Sort the values $\hat{\theta}_\ell^{*(b)}$, $b = 1, \ldots, B$, from smallest to largest. They are $\hat{\theta}_{\ell(1)}^* \leqslant \ldots \leqslant \hat{\theta}_{\ell(B)}^*$. A $(1-\alpha)\%$ percentile bootstrap CI for θ_ℓ is $\left(\hat{\theta}_{\ell(\lfloor(\alpha/2)B\rfloor)}^*, \hat{\theta}_{\ell(\lfloor(1-\alpha/2)B\rfloor)}^*\right)$.

Mean squared error estimation

We can estimate the MSE of a predictor $\hat{\mu}_{yijk}$ by using a resampling method. The following procedure calculates a parametric bootstrap estimator of $MSE(\hat{\mu}_{yijk})$. It also provides bootstrap estimates for the quantiles of the distribution of the predictions. Algorithm B has the following steps:

1. Fit the model to the sample and calculate the ML estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \hat{\phi}_1, \hat{\phi}_2)'$.
2. Repeat B times $(b = 1, \ldots, B)$:

(a) For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, generate $u_{1,jk}^{*(b)} \sim N(0,1)$, $u_{2,jk}^{*(b)} \sim N(0,1)$ and calculate

$$\begin{aligned} p_{ijk}^{*(b)} &= \exp\left\{\boldsymbol{x}_{1,ijk}\hat{\boldsymbol{\beta}}_1 + \hat{\phi}_1 u_{1,jk}^{*(b)}\right\}\left(1+\exp\left\{\boldsymbol{x}_{1,ijk}\hat{\boldsymbol{\beta}}_1 + \hat{\phi}_1 u_{1,jk}^{*(b)}\right\}\right)^{-1}, \\ \mu_{ijk}^{*(b)} &= \exp\left\{\boldsymbol{x}_{2,ijk}\hat{\boldsymbol{\beta}}_2 + \hat{\phi}_2 u_{2,jk}^{*(b)}\right\}. \end{aligned}$$

(b) Generate $z_{ijk}^{*(b)} \sim \text{BE}(p_{ijk}^{*(b)})$. If $z_{ijk}^{*(b)} = 1$, do $y_{ijk}^{*(b)} = 0$. If $z_{ijk}^{*(b)} = 0$, generate $y_{ijk}^{*(b)} \sim GA(\mu_{ijk}^{*(b)}, \nu)$.

(c) For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate $\mu_{yijk}^{*(b)} = (1 - p_{ijk}^{*(b)})\mu_{ijk}^{*(b)}$.

(d) Based on the sample $(y_{ijk}^{*(b)}, \boldsymbol{x}_{ijk})$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate the ML estimate $\hat{\boldsymbol{\theta}}^{*(b)}$ and the predictor $\hat{\mu}_{yijk}^{*(b)}$, $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$.

3. For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, calculate $mse^*(\hat{\mu}_{yijk}) = \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mu}_{yijk}^{*(b)} - \mu_{yijk}^{*(b)}\right)^2$,

$$rmse^*(\hat{\mu}_{yijk}) = (mse^*(\hat{\mu}_{yijk}))^{\frac{1}{2}}, \; rrmse^*(\hat{\mu}_{yijk}) = \frac{rmse^*(\hat{\mu}_{yijk})}{\hat{\mu}_{yijk}}.$$

4. For $i \in \mathbb{I}$, $j \in \mathbb{J}$, $k \in \mathbb{K}$, sort the values $\hat{\mu}_{yijk}^{*(b)}$, $b = 1, \ldots, B$, from smallest to largest. They are $\hat{\mu}_{yijk(1)} \leqslant \ldots \leqslant \hat{\mu}_{yijk(B)}$. Let $\alpha \in (0, 1)$. The bootstrap quantile of the distribution of the predictor $\hat{\mu}_{yijk}$ that leaves its left-hand probability α is $\hat{q}_{ijk,\alpha} := \hat{\mu}_{yijk(\lfloor \alpha B \rceil)}$.