Taylor & Francis
Taylor & Francis Group

Journal of Applied Statistics

ISSN: 0266-4763 (Print) 1360-0532 (Online) Journal homepage: http://www.tandfonline.com/loi/cjas20

A generalised linear space–time autoregressive model with space–time autoregressive disturbances

Oscar O. Melo, Jorge Mateu & Carlos E. Melo

To cite this article: Oscar O. Melo, Jorge Mateu & Carlos E. Melo (2015): A generalised linear space–time autoregressive model with space–time autoregressive disturbances, Journal of Applied Statistics, DOI: 10.1080/02664763.2015.1092506

To link to this article: http://dx.doi.org/10.1080/02664763.2015.1092506

Published online: 12 Oct 2015.

Journal of Applied Statistics, 2015
http://dx.doi.org/10.1080/02664763.2015.1092506

A generalised linear space–time autoregressive model with space–time autoregressive disturbances

Oscar O. Melo[a*], Jorge Mateu[b] and Carlos E. Melo[c]

[a]*Department of Statistics, Faculty of Sciences, Universidad Nacional de Colombia, Bogotá, Colombia;* [b]*Department of Mathematics, University Jaume I, Castellón, Spain;* [c]*Faculty of Engineering, Universidad Distrital Francisco José de Caldas, Bogotá, Colombia*

(*Received 30 December 2014; accepted 7 September 2015*)

We present a solution to problems where the response variable is a count, a rate or binary using a generalised linear space–time autoregressive model with space–time autoregressive disturbances (GLSTARAR). The possibility to test the fixed effect specification against the random effect specification of the panel data model is extended to include space–time error autocorrelation or a space–time lagged dependent variable. Space-time generalised estimating equations are used to estimate the spatio-temporal parameters in the model. We also present a measure of goodness of fit, and show the pseudo-best linear unbiased predictor for prediction purposes. Additionally, we propose a joint space–time modelling of mean and dispersion to give a solution when the variance is not constant. In the application, we use social, economic, geographic and state presence variables for 32 Colombian departments in order to analyse the relationship between the number of armed actions (AAs) per 1000 km^2 committed by the guerrillas of the FARC-EP and ELN during the years 2003–2009, and a set of covariates given by attention rate to victims of violence, forced displacement-households expelled, forced displacement-households received, total armed confrontations per year, number of AAs by military forces and percentage of people living in urban area.

Keywords: armed actions; generalised linear space–time autoregressive model; space–time autoregressive disturbances model; space–time generalised estimating equations; fixedeffect parameters; random effect parameters

1. Introduction

The idea of combining data for cross-sections and time series backs to a suggestion of Marschak [25], and has received considerable attention in the econometric, social-science, biostatistical

*Corresponding author. Email: oomelom@unal.edu.co

and geographic-science literature. Other overviews of the salient issues and suggested solutions can be found in Dielman [15], Chamberlain [11], Hsiao [21] and Anselin [3]. In recent years, the space–time literature has exhibited a growing interest in the specification and estimation of econometric and social-science relationships based on spatial panels. This interest can be explained by the fact that panel data offer researchers wider modelling possibilities as compared to the single equation cross-sectional setting, which was the primary focus of the spatial econometrics literature for a long time [22].

Space-time data deal with space–time interaction (space–time autocorrelation) and space–time structure (space–time heterogeneity) in regression models for cross-sectional and panel data [3,36]. Such a focus on location and space–time interaction has recently gained a more central place not only in applied but also in theoretical econometric and social-sciences. In the past, models that explicitly incorporated space or geography were primarily found in specialised fields such as regional science, urban, real estate economics and economic geography [4–6,35].

More recently, spatial and space–time methods have increasingly been applied in a wide range of empirical investigations in more traditional fields of economics and social-sciences as well, including, among others, studies in demand analysis, economic growth, international economics, labour market, employment indices, agricultural production and environmental pollution. Many of these studies have a continuous response variable; however, when the response variable is a count, a rate or binary, there is not much literature solving the problem of space–time autoregressive interactions including space–time autoregressive disturbances of stationary state variables. Therefore, these applications not only have led to new insights, developments and extensions, but also to new questions. Regional scientists have shown that spatial dependence data may alter, and even reverse, the results of standard time-series models.

Many of the works and studies presented above establish the importance of integrating the space–time structure into panel data analysis when the response variable is non-normal. The literature on models with single spatial and/or temporal dynamics has seen some progress in dealing with this kind of response variable, but in many cases, these two components have been treated in a separate, parallel way.

Therefore, in this paper we present a solution to problems where the response variable is a count, a rate or binary (dichotomous) using a refined generalised linear space–time autoregressive model with space–time autoregressive disturbances (coined GLSTARAR for short). This model may also contain additional spatial exogenous variables as well as time exogenous variables. The estimation of fixed effects and the determination of their significance levels are also developed. The possibility to test the fixed effect specification against the random effect specification of the panel data model is extended to include space–time error autocorrelation or a space–time lagged dependent variable using particular specification tests. Then, the determination of the variance–covariance matrix of the estimated parameters of these extended models is shown. Space-time generalised estimating equations (GEE) are used to estimate the spatio-temporal parameters in the model. We also present a measure of goodness of fit, and show the pseudo-best linear unbiased predictor for prediction purposes.

On the other hand, when mean and variance are related as in the GLSTARAR model, we proposed a methodology to joint space–time modelling of mean and dispersion. This joint modelling gives also a solution to the problem when the variance is not constant. Then, the proposed methodology can be used in variable response framed within the generalised linear models that have both spatial and temporal correlation.

In particular, we study the number of armed actions (AAs) per 1000 km^2 committed by the guerrillas of the FARC-EP and ELN in Colombia from 2003 to 2009. This is a case of space–time data with a non-normal response variable. We use social, economic, geographic and state presence variables for 32 Colombian departments in order to analyse the relationship between the number of AAs with the following exogenous variables: (i) attention rate to victims of violence (ARVV), standardised per 1000 km^2 and given in terms of 40 monthly legal minimum wage; this variable is associated with the number of households who suffer damages in their lives or serious deterioration in their personal integrity or property by reason of terrorist attacks, fighting, kidnappings, attacks and massacres, among others; (ii) forced displacement-households expelled (FDHE), as the number of households expelled per 1000 km^2; (iii) forced displacement-households received (FDHR), as the number of households received per 1000 km^2; (iv) total armed confrontations (TAC) per year per 1000 km^2; (v) number of AAs by military forces (AAMF) per 1000 km^2 and (vi) percentage of people living in urban area (PPLUA), which is given with respect to the total population of the department.

All this information is taken in an official area of 1000 km^2. This area is considered with respect to the areas where these groups operate, and it does not correspond to the area of the department. In addition, all exogenous variables show information from 2003 to 2009, that is, each of these variables is yearly based. The sources of information come from the Administrative Department for Social Prosperity (for ARVV, FDHE and FDHR), from the Vice President of the Republic (for AAs, TAC and AAMF) and from the National Administrative Department of Statistics (for PPLUA). All these are governmental entities of Colombia, and the website where the interested readers can find this information is http://sigotn.igac.gov.co/sigotn/.

Furthermore, values of Moran's I index [12,30] are assessed by a test statistic (Moran's I standard deviate) which indicates the statistical significance of spatio-temporal autocorrelation, for example amongst model residuals. Additionally, model residuals may be plotted as a map that more explicitly reveals particular patterns of spatio-temporal autocorrelation (e.g. anisotropy or non-stationarity).

The plan of paper is the following. Section 2 develops the methodological approach based on space–time dynamic generalised linear models (GLMs). Section 3 presents the space–time parameter estimation method through generalised estimating equations (GEEs). Section 4 presents the selection, validation and prediction of the fitted model using GEE in a space–time context; in this section, a goodness-of-fit measure, residual analysis and space–time prediction of new subjects are also discussed. Section 5 analyses an application that illustrates the proposed methodology, and Section 6 presents the joint space–time modelling of mean and dispersion to give a solution when the variance is not constant. Finally, the paper ends with some conclusions.

2. Space-time dynamic generalised linear models

Let $\{y(\boldsymbol{s}, t), \boldsymbol{s} \in D, t \in \mathrm{T}\}$ be a stochastic spatio-temporal process. Here, the index set D is either a continuous surface or a finite set of discrete locations, and $\mathrm{T} \subseteq \mathbb{Z}$, so that the model developed is suitable for discrete time. A distribution belongs to the exponential family if it has a density function given by

$$f(y(\boldsymbol{s}, t); \alpha_{\mathrm{st}}) = h_1(y(\boldsymbol{s}, t)) \exp\{\eta(\alpha_{\mathrm{st}}) h_2(y(\boldsymbol{s}, t)) - b(\alpha_{\mathrm{st}})\},$$

where $\eta(\alpha_{\mathrm{st}})$, $b(\alpha_{\mathrm{st}})$, $h_1(y(\boldsymbol{s}, t))$ and $h_2(y(\boldsymbol{s}, t))$ are functions that take values in the real line.

The proposed interpolation is built for a non-Gaussian random space–time dynamic model by specifically considering categorical, continuous and indicator variables in the trend model. The data-generating mechanism conditional on the signal of the model follows a classical GLM as described by McCullagh and Nelder [26]. Specifically, we focus on spatial lag (SL) and error models, in which we use both a time-wise and a spatially lagged dependent variable, and also, the dependence pertains to neighbouring locations in a different period. So, our starting point is the following model:

$$\begin{aligned}\eta_{it} &= \eta(\boldsymbol{s}_i, t) = g(\mu_{it}) = \boldsymbol{x}_{1i}^t\boldsymbol{\beta}_0 + \boldsymbol{x}_{2it}^t\boldsymbol{\beta}_t + \pi_t\sum_{i'=1}^{n} w_{ii'}^{(1)}\eta_{i't} + \varepsilon_{it},\\ \varepsilon_{it} &= \varepsilon(\boldsymbol{s}_i, t) = \psi_t\sum_{i'=1}^{n} w_{ii'}^{(2)}\varepsilon_{i't} + e_{it},\end{aligned} \tag{1}$$

with $i = 1, \dots, n$, $t = 1, \dots, T$, $|\pi_t| < 1$ and $|\psi_t| < 1$, where $\mu_{it} = \mu(\boldsymbol{s}_i, t) = \mathrm{E}[y(\boldsymbol{s}_i, t)|\boldsymbol{x}_{1i}, \boldsymbol{x}_{2it}, \varepsilon_{it}]$, $g(\cdot)$ is coined a link function which is invertible and continuous, $\boldsymbol{x}_{1i}^t\boldsymbol{\beta}_0 + \boldsymbol{x}_{2it}^t\boldsymbol{\beta}_t$ is the trend, $\boldsymbol{x}_{1i}^t = \boldsymbol{x}_1^t(\boldsymbol{s}_i) = (1, x_{i1}, \dots, x_{ip_1})$ is a vector containing explanatory variables associated with the spatial at location $\boldsymbol{s}_i$th, $\boldsymbol{\beta}_0 = (\beta_0, \beta_1, \dots, \beta_{p_1})^t$ is a vector of unknown spatial regression parameters, $\boldsymbol{x}_{2it}^t = \boldsymbol{x}_2^t(\boldsymbol{s}_i, t) = (x_{it1}, \dots, x_{itp_{2_t}})$ is a vector containing explanatory variables associated with the space–time at $\boldsymbol{s}_i$th location and tth time, and $\boldsymbol{\beta}_t = (\beta_{t1}, \dots, \beta_{tp_{2_t}})^t$ is a vector of unknown space–time regression parameters. Additionally, π_t is the spatial autoregressive coefficient in the tth time period, ε_{it} reflects the spatio-temporal autocorrelated error term at the $\boldsymbol{s}_i$th location in the tth time, ψ_t is called the spatial autocorrelation coefficient in the tth time period, and $e_{it} = e(\boldsymbol{s}_i, t)$ is an i.i.d. Gaussian random error term at the ith individual in the tth time with zero mean and covariance $\mathrm{E}(e_{it}, e_{it'}) = \sigma_{tt'}$ for $t, t' = 1, \dots, T$, with $\mathrm{E}(e_{it}, e_{i't}) = 0$ for $i, i' = 1, \dots, n$.

For simplicity, we can assume that $w_{ii'}^{(1)} = w_{ii'}^{(2)} = w_{ii'}$, with $w_{ii'}$ an element of a spatial weights matrix $\boldsymbol{W}$ describing the spatial arrangement of the units in the sample. An initial step towards weighting based on locality might be to exclude from the model calibration observations that are further than some distance d_1 from the regression point [19]. This would be equivalent to setting their weights to zero, giving a weighting function of the form

$$w_{ii'} = \begin{cases} 1 & \text{if } i' \in N(\boldsymbol{s}_i),\\ 0 & \text{otherwise,}\end{cases}$$

where $N(\boldsymbol{s}_i)$ is the set of all neighbours at location $\boldsymbol{s}_i$, which is built through a distance $(d_{ii'})$ between two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_{i'}$. These two locations are said to be neighbours if their distance is less than a threshold value (d_1).

The function distance reads points coordinates (w_x, w_y) and generates a matrix of weights $\boldsymbol{W}$. Some distance measures that can be used are: Euclidean $(d_{ii'} = \sqrt{(w_{x_i} - w_{x_{i'}})^2 + (w_{y_i} - w_{y_{i'}})^2})$, Chebyshev $(d_{ii'} = \max\{|w_{x_i} - w_{x_{i'}}|, |w_{y_i} - w_{y_{i'}}|\})$, Bray-Curtis $(d_{ii'} = (|w_{x_i} - w_{x_{i'}}| + |w_{y_i} - w_{y_{i'}}|)/(|w_{x_i} + w_{x_{i'}}| + |w_{y_i} + w_{y_{i'}}|))$ and Canberra $(d_{ii'} = (|w_{x_i} - w_{x_{i'}}| + |w_{y_i} - w_{y_{i'}}|)/(|w_{x_i}| + |w_{x_{i'}}| + |w_{y_i}| + |w_{y_{i'}}|))$, among others.

The link function $g(\cdot)$ given in Equation (1) is strictly monotonic and twice differentiable. Some possible choices for the link function $g(\mu_{it})$ are: the logit, $\eta_{it} = g(\mu_{it}) = \log\{\mu_{it}/(1 - \mu_{it})\}$; the probit, $\eta_{it} = g(\mu_{it}) = \Phi^{-1}(\mu_{it})$, where $\Phi(\cdot)$ is the cumulative distribution function of a standard normal random variable; the complementary loglog (cloglog), $\eta_{it} = g(\mu_{it}) = \log\{-\log(1 - \mu_{it})\}$; and the loglog, $\eta_{it} = g(\mu_{it}) = -\log\{-\log(\mu_{it})\}$. A rich discussion of link functions is presented in Atkinson [8] and McCullagh and Nelder [26].

Several of the above cases can be considered in a general class of link functions. Aranda-Ordaz [7] proposed a family of link functions to analyse data in the form of proportions, given by

$$\eta_{it} = g_\nu(\mu(\boldsymbol{s}_i, t)) = \log\left[\frac{(1 - \mu(\boldsymbol{s}_i, t))^{-\nu} - 1}{\nu}\right],$$

where ν is an unknown constant, which have as particular cases the logistic model when $\nu = 1$, and the complementary loglog when $\nu \to 0$.

Another general form of link functions, proposed by Box and Cox [10] and used mainly for data with positive mean, is the Box–Cox transformation given by

$$\eta_{it} = g_\nu(\mu(\boldsymbol{s}_i, t)) = \begin{cases} \dfrac{(\mu^\nu(\boldsymbol{s}_i, t))}{\nu} & \text{if } \nu > 0, \\ \log(\mu(\boldsymbol{s}_i, t)) & \text{if } \nu = 0. \end{cases}$$

Within the field of linear models, it is usual to work with the model in its canonical form, $\eta_{it}(\alpha(\boldsymbol{s}_i, t)) = \alpha(\boldsymbol{s}_i, t) = \alpha_{it}$, $h_2(y(\boldsymbol{s}_i, t)) = y(\boldsymbol{s}_i, t)$, which includes a dispersion parameter $\phi > 0$. Specifically, conditional on explanatory variables $(\boldsymbol{x}_{it})$ and on an unobserved spatial–temporal error ε_{it}, $y(\boldsymbol{s}_i, t)$ follows a distribution of the exponential family, that is,

$$y(\boldsymbol{s}_i, t) \mid \boldsymbol{x}_1(\boldsymbol{s}_i), \boldsymbol{x}_2(\boldsymbol{s}_i, t), \varepsilon_{it} \overset{\text{ind}}{\sim} f(y(\boldsymbol{s}_i, t) \mid \boldsymbol{x}_1(\boldsymbol{s}_i), \boldsymbol{x}_2(\boldsymbol{s}_i, t), \varepsilon_{it}),$$
$$f(y(\boldsymbol{s}_i, t) \mid \boldsymbol{x}_1(\boldsymbol{s}_i), \boldsymbol{x}_2(\boldsymbol{s}_i, t), \varepsilon_{it}) = \exp\left\{\frac{1}{\phi}[y(\boldsymbol{s}_i, t)\alpha_{it} - b(\alpha_{it})] + c(y(\boldsymbol{s}_i, t), \phi)\right\}, \tag{2}$$

where ϕ is an extra-variation parameter and $c(\cdot)$ is a specific function. The conditional mean, μ_{it}, is related to α_{it} through the identity $\mu_{it} = \partial b(\alpha_{it})/\partial\alpha_{it}$, and is modelled, after a proper transformation, as the GLM given by Equation (1) in terms of both space–time fixed and random effects.

In Equation (1), the coefficients $\boldsymbol{\beta}_t$ are constant across space but vary for each time. The error terms are temporally correlated, that is, there is a constant covariance between errors for different periods for the same spatial unit. In matrix form, Equation (1) for each period t $(t = 1, \ldots, T)$ becomes

$$\begin{aligned} \boldsymbol{\eta}_t &= \boldsymbol{X}_1\boldsymbol{\beta}_0 + \boldsymbol{X}_{2t}\boldsymbol{\beta}_t + \pi_t\boldsymbol{W}_1\boldsymbol{\eta}_t + \boldsymbol{\varepsilon}_t, \\ \boldsymbol{\varepsilon}_t &= \psi_t\boldsymbol{W}_2\boldsymbol{\varepsilon}_t + \boldsymbol{e}_t, \end{aligned} \tag{3}$$

where $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{nt})^t$ is an $n \times 1$ vector, $\boldsymbol{X}_1 = (\boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{1n})^t$ is an $n \times (p_1 + 1)$ matrix of spatial explanatory variables, $\boldsymbol{X}_{2t} = (\boldsymbol{x}_{21t}, \ldots, \boldsymbol{x}_{2nt})^t$ is an $n \times p_{2_t}$ matrix of space–time explanatory variables, $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{nt})^t$ is an $n \times 1$ vector, and $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are $n \times n$ matrices that describe the SL and spatial error arrangement of the units in the sample, respectively.

Note that in Equation (3) the number of explanatory variables for $\boldsymbol{X}_{2t}$, p_{2_t}, can be different for each equation (time period). Additionally, the space–time GLM can be made operational only when more observations are available in the spatial dimension than in the time dimension $(n > T)$. In the more typical case where $T > n$, the usual regression using GLM applies.

Therefore, we can reformulate model (1) in a compact vectorial form as

$$\begin{aligned} \boldsymbol{\eta}_{\text{st}} &= g(\boldsymbol{\mu}_{\text{st}}) = g(\text{E}(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{X}_{\text{st}}, \boldsymbol{\varepsilon}_{\text{st}})) = \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + \boldsymbol{\varepsilon}_{\text{st}}, \\ \boldsymbol{\varepsilon}_{\text{st}} &= (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{\varepsilon}_{\text{st}} + \boldsymbol{e}_{\text{st}}, \end{aligned} \tag{4}$$

where $\otimes$ denotes the Kronecker product, and

$$\boldsymbol{X}_{\text{st}} = \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_{21} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{X}_1 & \boldsymbol{0} & \boldsymbol{X}_{22} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{X}_1 & \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{X}_{2T} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_0 \\ \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_T \end{bmatrix}_{nT\times 1}, \quad \boldsymbol{\Pi} = \begin{bmatrix} \pi_1 & 0 & \cdots & 0 \\ 0 & \pi_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_T \end{bmatrix}_{T\times T},$$

$$\boldsymbol{\Psi} = \begin{bmatrix} \psi_1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_T \end{bmatrix}_{T\times T}, \quad \boldsymbol{\varepsilon}_{\text{st}} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_T \end{bmatrix}_{nT\times 1}, \quad \boldsymbol{\eta}_{\text{st}} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \\ \vdots \\ \boldsymbol{\eta}_T \end{bmatrix}_{nT\times 1},$$

with $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{nt})^t$ being an $n \times 1$ vector $(t = 1, \ldots, T)$, $\boldsymbol{\mu}_{\text{st}} = \mathrm{E}(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{X}_{\text{st}}, \boldsymbol{\varepsilon}_{\text{st}})$, $\boldsymbol{\mu}_{\text{st}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_T)^t$ is an $nT \times 1$ vector with $\boldsymbol{\mu}_t = (\mu_{1t}, \ldots, \mu_{nt})^t$ an $n \times 1$ vector, and $\boldsymbol{y}_{\text{st}} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)^t$ is an $nT \times 1$ vector with $\boldsymbol{y}_t = (y(\boldsymbol{s}_1, t), \ldots, y(\boldsymbol{s}_n, t))^t$ an $n \times 1$ vector, and $\boldsymbol{X}_{\text{st}}$ a matrix of original explanatory variables, which can involve continuous, categorical and binary variables, or even a mixture of them. Furthermore, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{nt})^t$ is an $n \times 1$ vector, and $\boldsymbol{e}_{\text{st}} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_T)^t \sim MN(\boldsymbol{0}, \boldsymbol{\Sigma}_T \otimes \boldsymbol{I}_n)$ and it is an $nT \times 1$ vector with $\boldsymbol{e}_t = (e_{1t}, \ldots, e_{nt})^t$ an $n \times 1$ vector.

The model given in Equation (4) defines a GLSTARAR, which includes exogenous regressors. Space-time interactions are modelled through space–time lags and space–time errors. Moreover, the model allows for space–time interactions in the dependent variable, the explanatory variables, and the disturbances.

The expressions in Equation (4) can be rewritten as

$$
\begin{aligned}
\boldsymbol{\eta}_{\text{st}} &= \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + \boldsymbol{\varepsilon}_{\text{st}} = \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t\right)\boldsymbol{\pi} + \boldsymbol{\varepsilon}_{\text{st}} \\
&= \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}_{\text{st}}, \qquad (5) \\
\boldsymbol{\varepsilon}_{\text{st}} &= (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{\varepsilon}_{\text{st}} + \boldsymbol{e}_{\text{st}} = \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{\varepsilon}_t\right)\boldsymbol{\psi} + \boldsymbol{e}_{\text{st}} \\
&= \boldsymbol{\varepsilon}^*_{\text{st}}\boldsymbol{\psi} + \boldsymbol{e}_{\text{st}}, \qquad (6)
\end{aligned}
$$

where $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_T)^t$, $\mathbf{X} = (\boldsymbol{X}_{\text{st}}, \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t)$ with $\oplus$ denoting the direct sum operator, $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^t, \boldsymbol{\pi}^t)$, $\boldsymbol{\varepsilon}^*_{\text{st}} = (\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_T)^t$.

Models (5) and (6) can be expressed as

$$
\begin{aligned}
\boldsymbol{\eta}_{\text{st}} &= \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + \boldsymbol{\varepsilon}_{\text{st}} \\
&= \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}\boldsymbol{e}_{\text{st}}, \qquad (7)
\end{aligned}
$$

where $\boldsymbol{\varepsilon}_{\text{st}} \sim MN(\boldsymbol{0}, \mathrm{Var}(\boldsymbol{\varepsilon}_{\text{st}}))$ with $\mathrm{Var}(\boldsymbol{\varepsilon}_{\text{st}}) = [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}(\boldsymbol{\Sigma}_T \otimes \boldsymbol{I}_n)\{[\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}\}^t$.

The model given (7) can be written equivalently as

$$
\begin{aligned}
\boldsymbol{\eta}_{\text{st}} &= \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{X}_{\text{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{\eta}_{\text{st}} - (\boldsymbol{\Psi}\boldsymbol{\Pi} \otimes \boldsymbol{W}_2\boldsymbol{W}_1)\boldsymbol{\eta}_{\text{st}} + \boldsymbol{e}_{\text{st}} \\
&= \boldsymbol{X}_{\text{st}}\boldsymbol{\beta} - \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{X}_{2t}\right)(\boldsymbol{\psi} \otimes \boldsymbol{\beta}) + \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t\right)\boldsymbol{\pi} + \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{\eta}_t\right)\boldsymbol{\psi} \\
&\quad - \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{\eta}_t\right)(\boldsymbol{\psi} \odot \boldsymbol{\pi}) + \boldsymbol{e}_{\text{st}} \\
&= \boldsymbol{Z}_1\boldsymbol{\alpha}_1 + \boldsymbol{Z}_2\boldsymbol{\alpha}_2 + \boldsymbol{Z}_3\boldsymbol{\alpha}_3 + \boldsymbol{Z}_4\boldsymbol{\alpha}_4 + \boldsymbol{Z}_5\boldsymbol{\alpha}_5 + \boldsymbol{e}_{\text{st}} = \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{e}_{\text{st}}, \qquad (8)
\end{aligned}
$$

where $\boldsymbol{Z}_1 = \boldsymbol{X}_{\text{st}}$, $\boldsymbol{Z}_2 = \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{X}_{2t}$, $\boldsymbol{Z}_3 = \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t$, $\boldsymbol{Z}_4 = \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{\eta}_t$, $\boldsymbol{Z}_5 = \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{W}_1\boldsymbol{\eta}_t$, $\boldsymbol{\alpha}_1 = \boldsymbol{\beta}$, $\boldsymbol{\alpha}_2 = -\boldsymbol{\psi} \otimes \boldsymbol{\beta}$, $\boldsymbol{\alpha}_3 = \boldsymbol{\pi}$, $\boldsymbol{\alpha}_4 = \boldsymbol{\psi}$, $\boldsymbol{\alpha}_5 = -\boldsymbol{\psi} \odot \boldsymbol{\pi}$ with $\odot$ denoting the Hadamard product, $\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_5)$, and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^t, \ldots, \boldsymbol{\alpha}_5^t)$. The constraints over the parameters in Equation (8) are $\boldsymbol{\alpha}_2 = -\boldsymbol{\alpha}_4 \otimes \boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_5 = -\boldsymbol{\alpha}_4 \odot \boldsymbol{\alpha}_3$.

Example 2.1 Consider the generalised linear space–time autoregressive model without space–time autoregressive disturbances. This model is given by

$$\begin{aligned}\boldsymbol{\eta}_{\mathrm{st}} &= \boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} + (\boldsymbol{\Pi} \otimes \boldsymbol{W}_1)\boldsymbol{\eta}_{\mathrm{st}} + \boldsymbol{e}_{\mathrm{st}} \\ &= \boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} + \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t\right)\boldsymbol{\pi} + \boldsymbol{e}_{\mathrm{st}},\end{aligned}$$

where $\boldsymbol{X}_{\mathrm{st}}$, $\boldsymbol{\beta}$, $\boldsymbol{\Pi} \otimes \boldsymbol{W}_1$, $\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t$ and $\boldsymbol{\pi}$ are defined as in Equation (8), and $\boldsymbol{e}_{\mathrm{st}} \sim MN(\mathbf{0}, \boldsymbol{\Sigma}_T \otimes \boldsymbol{I}_n)$.

Example 2.2 The generalised linear space–time autoregressive disturbances model is given by

$$\begin{aligned}\boldsymbol{\eta}_{\mathrm{st}} &= \boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} + [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}\boldsymbol{e}_{\mathrm{st}}, \\ [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]\boldsymbol{\eta}_{\mathrm{st}} &= [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]\boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} + \boldsymbol{e}_{\mathrm{st}}, \\ \boldsymbol{\eta}_{\mathrm{st}} &= \boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} + (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)\boldsymbol{\eta}_{\mathrm{st}} + \boldsymbol{e}_{\mathrm{st}} \\ &= \boldsymbol{X}_{\mathrm{st}}\boldsymbol{\beta} - \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{X}_{2t}\right)(\boldsymbol{\psi} \otimes \boldsymbol{\beta}) + \left(\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{\eta}_t\right)\boldsymbol{\psi} + \boldsymbol{e}_{\mathrm{st}},\end{aligned}$$

where $\boldsymbol{X}_{\mathrm{st}}$, $\boldsymbol{\beta}$, $\boldsymbol{\Psi} \otimes \boldsymbol{W}_2$, $\bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_2\boldsymbol{\eta}_t$ and $\boldsymbol{\psi}$ are defined as in Equation (8).

3. Parameter estimation methods

3.1 *Estimation by maximum likelihood*

Assuming that each $Y(\boldsymbol{s}_i, t)$ in model (5) has a distribution of the exponential family and by independence of $Y(\boldsymbol{s}_1, t), \ldots, Y(\boldsymbol{s}_n, t)$ given $\mathbf{X}$, $\boldsymbol{\varepsilon}_{\mathrm{st}}$ and $g^{-1}(\cdot)$, the conditional density function of $\boldsymbol{Y}_{\mathrm{st}} = \boldsymbol{y}_{\mathrm{st}}$ given the observed covariates $\mathbf{X}$ and $\boldsymbol{\varepsilon}_{\mathrm{st}}$ is given by

$$f(\boldsymbol{y}_{\mathrm{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\mathrm{st}}; \boldsymbol{\beta}^*) = \prod_{i=1}^{n}\prod_{t=1}^{\mathrm{T}} f(y(\boldsymbol{s}_i, t); g^{-1}(\boldsymbol{x}(\boldsymbol{s}_i, t); \boldsymbol{\beta}^*)).$$

From a classical perspective, the likelihood function based on the observed random variables $\boldsymbol{y}_{\mathrm{st}}$ is obtained by marginalising with respect to the unobserved random variables $\boldsymbol{\varepsilon}_{\mathrm{st}}$, leading to the mixed-model likelihood. Then, the likelihood function for a space–time generalised autoregression model cannot be written in closed form, but only as a high-dimensional integral

$$\begin{aligned}L(\boldsymbol{\beta}^*, \boldsymbol{\theta}) &= f(\boldsymbol{y}_{\mathrm{st}} \mid \boldsymbol{\beta}^*, \boldsymbol{\theta}) = \int_{\mathbb{R}^{nT}} f(\boldsymbol{y}_{\mathrm{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\mathrm{st}}; \boldsymbol{\beta}^*) f(\boldsymbol{\varepsilon}_{\mathrm{st}} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\varepsilon}_{\mathrm{st}} \\ &= \int_{\mathbb{R}^{nT}} \prod_{i=1}^{n}\prod_{t=1}^{\mathrm{T}} f(y(\boldsymbol{s}_i, t); g^{-1}(\boldsymbol{x}(\boldsymbol{s}_i, t), \varepsilon(\boldsymbol{s}_i, t); \boldsymbol{\beta}^*)) f(\boldsymbol{\varepsilon}_{\mathrm{st}} \mid \boldsymbol{\theta})\, \mathrm{d}\varepsilon_{11} \cdots \mathrm{d}\varepsilon_{nT},\end{aligned} \tag{9}$$

where $f(\boldsymbol{\varepsilon}_{\mathrm{st}} \mid \boldsymbol{\theta})$ denotes the multivariate normal distribution of $\boldsymbol{\varepsilon}_{\mathrm{st}}$ given the observed covariates $\mathbf{X}$, with $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\Sigma}_T)$ the set of parameters associated with $\boldsymbol{\varepsilon}_{\mathrm{st}}$.

3.1.1 *Fixed effect parameters case.*

Even though the likelihood equations are numerically difficult, we can write them in a simpler form. From Equation (9), the log likelihood is given by

$$\ell = \log \int_{\mathbb{R}^{nT}} f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}} = \log f(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{\beta}^*, \boldsymbol{\theta}), \tag{10}$$

so that

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\beta}^*} &= \frac{\partial}{\partial \boldsymbol{\beta}^*} \int_{\mathbb{R}^{nT}} \frac{f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})}{f(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{\beta}^*, \boldsymbol{\theta})} \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}} \\ &= \int_{\mathbb{R}^{nT}} \left[\frac{\partial}{\partial \boldsymbol{\beta}^*} f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) \right] \frac{f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})}{f(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{\beta}^*, \boldsymbol{\theta})} \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}}, \end{aligned} \tag{11}$$

since $f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})$ does not involve $\boldsymbol{\beta}^*$. Noting that

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}^*} f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) &= \left(\frac{1}{f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)} \frac{\partial f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} \right) f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) \\ &= \frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*), \end{aligned} \tag{12}$$

we can rewrite Equation (11) as

$$\begin{aligned} \frac{\partial \ell}{\partial \boldsymbol{\beta}^*} &= \int_{\mathbb{R}^{nT}} \frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*) \frac{f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})}{f(\boldsymbol{y}_{\text{st}} \mid \boldsymbol{\beta}^*, \boldsymbol{\theta})} \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}} \\ &= \int_{\mathbb{R}^{nT}} \frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{y}_{\text{st}}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}}, \end{aligned} \tag{13}$$

where

$$\frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} = \frac{\partial}{\partial \boldsymbol{\beta}^*} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \left\{ \frac{1}{\phi} [y_{i,t}\alpha_{it} - b(\alpha_{it})] + c(y_{it}, \phi) \right\}.$$

Then, the score equations from a likelihood analysis have the form

$$\begin{aligned} \frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} &= \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \left(y_{it} \frac{\partial \alpha_{it}}{\partial \boldsymbol{\beta}^*} - \frac{\partial b(\alpha_{it})}{\partial \alpha_{it}} \frac{\partial \alpha_{it}}{\partial \boldsymbol{\beta}^*} \right) = \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} (y_{it} - \mu_{it}) \frac{\partial \alpha_{it}}{\partial \boldsymbol{\beta}^*} \\ &= \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} (y_{it} - \mu_{it}) \frac{\partial \alpha_{it}}{\partial \mu_{it}} \frac{\partial \mu_{it}}{\partial \boldsymbol{\beta}^*} = \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \frac{(y_{it} - \mu_{it})}{v(\mu_{it})} \frac{\partial \mu_{it}}{\partial \eta_{it}} \frac{\partial \eta_{it}}{\partial \boldsymbol{\beta}^*} \\ &= \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \frac{(y_{it} - \mu_{it})}{v(\mu_{it}) g'(\mu_{it})} \mathbf{x}_{it}^t = \frac{1}{\phi} \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} (y_{it} - \mu_{it}) \xi_{it} g'(\mu_{it}) \mathbf{x}_{it}^t, \end{aligned} \tag{14}$$

since $\mathrm{E}(y_{it}) = \mu_{it} = \partial b(\alpha_{it})/\partial \alpha_{it}$, $\mathrm{Var}(y_{it}) = \phi \partial^2 b(\alpha_{it})/\partial \alpha_{it}^2 = \phi v(\mu_{it})$ with $v(\mu_{it})$ being a variance function, $\partial \alpha_{it}/\partial \mu_{it} = (\partial \mu_{it}/\partial \alpha_{it})^{-1} = (\partial^2 b(\alpha_{it})/\partial \alpha_{it}^2)^{-1} = 1/v(\mu_{it})$, $\partial \mu_{it}/\partial \eta_{it} = \partial \mu_{it}/\partial g(\mu_{it}) = (\partial g(\mu_{it})/\partial \mu_{it})^{-1} = (g'(\mu_{it}))^{-1}$ and $\partial \eta_{it}/\partial \boldsymbol{\beta}^* = \partial g(\mu_{it})/\partial \mu_{it} = \mathbf{x}_{it}^t$, where $\mathbf{x}_{it}^t$ the itth row of $\mathbf{X}$. Additionally, $\xi_{it} = \{v(\mu_{it})[g'(\mu_{it})]^2\}^{-1}$.

We can write Equation (14) in matrix notation as

$$\frac{\partial \log f(\boldsymbol{y}_{\mathrm{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\mathrm{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} = \frac{1}{\phi}\mathbf{X}^t\, \boldsymbol{\Xi}\, \boldsymbol{\Delta}(\boldsymbol{y}_{\mathrm{st}} - \boldsymbol{\mu}_{\mathrm{st}}), \tag{15}$$

where $\boldsymbol{\Xi} = \mathrm{diag}(\xi_{it})$ and $\boldsymbol{\Delta} = \mathrm{diag}(g'(\mu_{it}))$, $i = 1, \ldots, n$ and $j = 1, \ldots, T$.

Then, replacing Equation (15) in Equation (13), we obtain

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\beta}^*} &= \int_{\mathbb{R}^{nT}} \frac{1}{\phi}\mathbf{X}^t\, \boldsymbol{\Xi}\, \boldsymbol{\Delta}(\boldsymbol{y}_{\mathrm{st}} - \boldsymbol{\mu}_{\mathrm{st}}) f(\boldsymbol{\varepsilon}_{\mathrm{st}} \mid \boldsymbol{y}_{\mathrm{st}}, \boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\varepsilon}_{\mathrm{st}} \\ &= \mathbf{X}^t \mathrm{E}(\boldsymbol{\Xi}^* \mid \boldsymbol{y}_{\mathrm{st}}) - \mathbf{X}^t \mathrm{E}(\boldsymbol{\Xi}^* \boldsymbol{\mu}_{\mathrm{st}} \mid \boldsymbol{y}_{\mathrm{st}}),\end{aligned} \tag{16}$$

where $\boldsymbol{\Xi}^* = \mathrm{diag}(\{\phi v(\mu_{it}) g'(\mu_{it})\}^{-1})$ and $\mathrm{E}(\cdot \mid \boldsymbol{y}_{\mathrm{st}})$ is the conditional expected value given $\boldsymbol{y}_{\mathrm{st}}$.

The likelihood equation for $\boldsymbol{\beta}^*$ is therefore

$$\mathbf{X}^t \mathrm{E}(\boldsymbol{\Xi}^* \mid \boldsymbol{y}_{\mathrm{st}}) = \mathbf{X}^t \mathrm{E}(\boldsymbol{\Xi}^* \boldsymbol{\mu}_{\mathrm{st}} \mid \boldsymbol{y}_{\mathrm{st}}). \tag{17}$$

Typically, these equations are non-linear functions of $\boldsymbol{\beta}^*$, and so, Equation (17) cannot be solved analytically.

The solution of the maximum likelihood (ML) equation in Equation (17) for $\boldsymbol{\beta}^*$ is usually performed by an iterative weighted least squares method. This can be derived as an example of the use of Fisher scoring [40, p. 295]. Fisher scoring is an iterative method for maximising a likelihood, and takes the form

$$\begin{aligned}\boldsymbol{\beta}^{*(m+1)} &= \boldsymbol{\beta}^{*(m)} - \left\{ \mathrm{E}\left[\left. \frac{\partial^2 \ln f(\boldsymbol{y}_{\mathrm{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\mathrm{st}}; \boldsymbol{\beta}^{*(m)})}{\partial \boldsymbol{\beta}^* \partial \boldsymbol{\beta}^{*t}} \right| \boldsymbol{y}_{\mathrm{st}} \right] \right\}^{-1} \mathrm{E}\left[\left. \frac{\partial \ln f(\boldsymbol{y}_{\mathrm{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\mathrm{st}}; \boldsymbol{\beta}^{*(m)})}{\partial \boldsymbol{\beta}^*} \right| \boldsymbol{y}_{\mathrm{st}} \right] \\ &= \boldsymbol{\beta}^{*(m)} + (\mathbf{X}^t \boldsymbol{\Xi}^* \mathbf{X})^{-1} \mathbf{X}^t [\mathrm{E}(\boldsymbol{\Xi}^* \mid \boldsymbol{y}_{\mathrm{st}}) - \mathrm{E}(\boldsymbol{\Xi}^* \boldsymbol{\mu}_{\mathrm{st}} \mid \boldsymbol{y}_{\mathrm{st}})].\end{aligned} \tag{18}$$

3.1.2 *Random effect parameters case.*

Once estimated the fixed parameters in model (5), given by $\boldsymbol{\beta}^*$, we are ready to estimate the parameters $\boldsymbol{\theta}$. From Equation (5), we find that $\hat{\boldsymbol{\varepsilon}}_{\mathrm{st}} = \boldsymbol{\eta}_{\mathrm{st}} - \mathbf{X}\hat{\boldsymbol{\beta}}^*$. Then, we use the model presented in Equation (6) to obtain

$$\hat{\boldsymbol{\varepsilon}}_{\mathrm{st}} = \hat{\boldsymbol{\varepsilon}}^*_{\mathrm{st}} \boldsymbol{\psi} + \boldsymbol{e}_{\mathrm{st}}, \tag{19}$$

where $\boldsymbol{e}_{\mathrm{st}} \sim MN(\mathbf{0}, \boldsymbol{\Sigma}_T)$ with $\boldsymbol{\Sigma}_T = \sigma^2 \boldsymbol{R}(\boldsymbol{\vartheta})$ being a $T \times T$ symmetric covariance matrix. $\boldsymbol{R}(\boldsymbol{\vartheta})$ is a $T \times T$ symmetric correlation matrix, and $\boldsymbol{\vartheta}$ is a $q \times 1$ vector which fully characterises $\boldsymbol{R}(\boldsymbol{\vartheta})$. Then, for model (19), the estimators obtained with the maximum likelihood method are given by

$$\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\varepsilon}}^{*t}_{\mathrm{st}} \hat{\boldsymbol{\varepsilon}}^*_{\mathrm{st}})^{-1} \hat{\boldsymbol{\varepsilon}}^{*t}_{\mathrm{st}} \hat{\boldsymbol{\varepsilon}}_{\mathrm{st}}, \tag{20}$$

$$\hat{\boldsymbol{\Sigma}}_T = \frac{1}{n}(\hat{\mathbf{E}} - \hat{\mathbf{E}}_{\hat{\psi}})^t (\hat{\mathbf{E}} - \hat{\mathbf{E}}_{\hat{\psi}}), \tag{21}$$

where $\hat{\mathbf{E}} = (\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_T)$ is an $n \times T$ matrix, and $\hat{\mathbf{E}}_{\hat{\psi}} = (\psi_1 \boldsymbol{W}_2 \boldsymbol{\varepsilon}_1, \ldots, \psi_T \boldsymbol{W}_2 \boldsymbol{\varepsilon}_T)$ is an $n \times T$ matrix.

Since the estimator given in Equation (21) is biased, we can use the following unbiased estimator of $\boldsymbol{\Sigma}_T$:

$$\hat{\boldsymbol{\Sigma}}_T = \frac{1}{n - T}(\hat{\mathbf{E}} - \hat{\mathbf{E}}_{\hat{\psi}})^t (\hat{\mathbf{E}} - \hat{\mathbf{E}}_{\hat{\psi}}). \tag{22}$$

However, we can choose any other specific $\boldsymbol{\Sigma}_T$ different from Equation (22). To define the structure that follows the covariance matrix, it is necessary to take into account the variation between times. We also need to specify the type of relationship between the observations (the sample covariance or correlation matrix). Several covariance and correlation matrices can be found in Diggle *et al.* [16], Molenberghs and Verbeke [29], and Davidian [14].

A result similar to Equation (13) can be derived for the ML equations for the parameters $\boldsymbol{\theta}$ in the distribution of $f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{y}_{\text{st}}, \boldsymbol{\theta})$, so that

$$\begin{aligned}\frac{\partial \ell}{\partial \boldsymbol{\theta}} &= \int_{\mathbb{R}^{nT}} \frac{\partial \log f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{y}_{\text{st}}, \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\varepsilon}_{\text{st}} \\ &= \mathrm{E}\left[\frac{\partial \log f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{y}_{\text{st}} \right]. \end{aligned} \tag{23}$$

Since $f(\boldsymbol{\varepsilon}_{\text{st}} \mid \boldsymbol{\theta})$ follows a multivariate normal distribution, some simplifications are possible. However, we prefer using Equation (23) because of its simplicity.

3.2 *Space-time generalised estimating equations*

The difficulty in evaluating the likelihood for a model such as Equation (5), and since the numerical maximisation is not so simple, we must turn to alternative approaches, as those presented above, to investigate ways to maximise the likelihoods. The GEEs approach begins by defining a marginal GLM for the mean of $\boldsymbol{y}_{\text{st}}$ as a function of the predictors.

Liang and Zeger [24] developed the GEE approach as an extension of GLMs. When responses are measured repeatedly through time or space, which happens in our GLSTARAR model, we need to estimate the correlation between times into the same location and the correlation between regions in a same time. The GEE method takes correlations within clusters of sampling units into account by means of a parameterised correlation matrix, while correlations between clusters are assumed to be zero. In a space–time context, such clusters can be interpreted as geographical regions measured along time, if distances between different regions measured along time are large enough [2]. We modified the approach of Liang and Zeger [24] to use these GEE models in a space–time context. Fortunately, estimates of regression parameters are fairly robust against miss-specification of the correlation matrix [17]. The GEE approach is especially suited for parameter estimation rather than prediction [9], but it can be also used for prediction purposes.

Firstly, we present an estimator $\hat{\boldsymbol{\beta}}_I^*$ of $\boldsymbol{\beta}^*$ in model (5) which arises under the working assumption that observations along time in different locations are independent. Under the independence working assumption, let $\boldsymbol{Y}_{\text{st}}$ be the response variable vector whose distribution can be written in the form of Equation (2), with expected value $\boldsymbol{\mu}_{\text{st}}$ linked to the linear predictor $\boldsymbol{\eta}_{\text{st}} = \mathbf{X}\boldsymbol{\beta}^*$ by the expression $g(\boldsymbol{\mu}_{\text{st}}) = \boldsymbol{\eta}_{\text{st}}$. Then, the log-likelihood function given in Equation (15) can be written as

$$\begin{aligned}\boldsymbol{U}_I &= \frac{\partial \log f(\boldsymbol{y}_{\text{st}} \mid \mathbf{X}, \boldsymbol{\varepsilon}_{\text{st}}; \boldsymbol{\beta}^*)}{\partial \boldsymbol{\beta}^*} = \frac{1}{\phi} \mathbf{X}^t \, \boldsymbol{\Xi} \, \boldsymbol{\Delta} (\boldsymbol{y}_{\text{st}} - \boldsymbol{\mu}_{\text{st}}) \\ &= \frac{1}{\phi} \mathbf{D}^t \mathbf{A}^{-1} (\boldsymbol{y}_{\text{st}} - \boldsymbol{\mu}_{\text{st}}), \end{aligned} \tag{24}$$

where $\mathbf{D} = \partial \boldsymbol{\mu}_{\text{st}} / \partial \boldsymbol{\beta}^*$ and $\mathbf{A} = \operatorname{diag}(v(\mu_{it}))$ is an $nT \times nT$ diagonal matrix with itth element $v(\mu_{it})$, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. The estimator $\hat{\boldsymbol{\beta}}_I$ is defined as the solution of Equation (24) using the Fisher-scoring method.

In order to see the correlation between observations for the same individual and among individuals, the correlation structure is incorporated by selecting a correlation matrix $\boldsymbol{R}(\boldsymbol{\theta})$ using the following expression:

$$\boldsymbol{\Gamma} = \mathbf{A}^{1/2}\boldsymbol{R}(\boldsymbol{\theta})\mathbf{A}^{1/2},$$

where $\boldsymbol{R}(\boldsymbol{\theta}) = [\mathrm{diag}(\mathrm{Var}(\varepsilon_{it}))]^{-1/2}\mathrm{Var}(\boldsymbol{\varepsilon}_{\mathrm{st}})[\mathrm{diag}(\mathrm{Var}(\varepsilon_{it}))]^{-1/2}$ with $\mathrm{Var}(\boldsymbol{\varepsilon}_{\mathrm{st}}) = [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}(\boldsymbol{\Sigma}_T \otimes \boldsymbol{I}_n)\{[\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}\}^t$, $\mathrm{diag}(\mathrm{Var}(\varepsilon_{it}))$ being an $nT \times nT$ diagonal matrix, and $\boldsymbol{\Sigma}_T = \sigma^2\boldsymbol{R}(\boldsymbol{\vartheta})$.

Then, the GEE for the vector parameter $\boldsymbol{\beta}^*$ takes the following form:

$$\boldsymbol{U}(\boldsymbol{\beta}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi) = \frac{1}{\phi}\mathbf{D}^t\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}_{\mathrm{st}} - \boldsymbol{\mu}_{\mathrm{st}}). \tag{25}$$

Liang and Zeger [24] showed that the vector solution of Equation (25), $\hat{\boldsymbol{\beta}}^*$, follows a multivariate normal distribution with mean $\boldsymbol{\beta}^*$ and variance–covariance matrix given by

$$\boldsymbol{H}_1^{-1}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi) = \phi(\hat{\mathbf{D}}^t\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{D}})^{-1}. \tag{26}$$

The consistency of the estimates given in Equation (26) depends on the correct specification of the link function used in Equation (5). To remedy this problem, it is often used as variance–covariance matrix of $\hat{\boldsymbol{\beta}}^*$ the following expression:

$$\begin{aligned}\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta}))) &= nT[\boldsymbol{H}_1(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)]^{-1}\boldsymbol{H}_2(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)\\ &\quad\times [\boldsymbol{H}_1(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)]^{-1},\end{aligned} \tag{27}$$

where

$$\boldsymbol{H}_2(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi) = \frac{1}{\phi}[\hat{\mathbf{D}}^t\hat{\boldsymbol{\Gamma}}^{-1}(\boldsymbol{y}_{\mathrm{st}} - \hat{\boldsymbol{\mu}}_{\mathrm{st}})(\boldsymbol{y}_{\mathrm{st}} - \hat{\boldsymbol{\mu}}_{\mathrm{st}})^t\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{D}}].$$

For solving the system of equations given in Equation (25), we use the Fisher-scoring method with the matrix $H_1^{-1}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)$ and the vector $\boldsymbol{U}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)$. The mth iteration of the process is given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}^{*(m+1)} &= \hat{\boldsymbol{\beta}}^{*(m)} + [\boldsymbol{H}_1^{(m)}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)]^{-1}\boldsymbol{U}^{(m)}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\boldsymbol{\theta})), \phi)\\ &= \hat{\boldsymbol{\beta}}^{*(m)} + (\hat{\mathbf{D}}^t\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{D}})^{-1}\hat{\mathbf{D}}^t\boldsymbol{\Gamma}^{-1}(\boldsymbol{y}_{\mathrm{st}} - \hat{\boldsymbol{\mu}}_{\mathrm{st}}^{(m)})\end{aligned} \tag{28}$$

or equivalently, we can write

$$\hat{\boldsymbol{\beta}}^{*(m+1)} = (\hat{\mathbf{D}}^t\hat{\boldsymbol{\Gamma}}^{-1}\hat{\mathbf{D}})^{-1}\hat{\mathbf{D}}^t\boldsymbol{\Gamma}^{-1}[\hat{\mathbf{D}}\hat{\boldsymbol{\beta}}^{*(m)} + (\boldsymbol{y}_{\mathrm{st}} - \hat{\boldsymbol{\mu}}_{\mathrm{st}}^{(m)})]. \tag{29}$$

This iterative procedure for calculating $\boldsymbol{\beta}^*$ is equivalent to performing an iteratively reweighted linear regression of $[\hat{\mathbf{D}}\hat{\boldsymbol{\beta}}^{*(m)} + (\boldsymbol{y}_{\mathrm{st}} - \hat{\boldsymbol{\mu}}_{\mathrm{st}}^{(m)})]$ on $\hat{\mathbf{D}}$ with weight matrix $\boldsymbol{\Gamma}^{-1}$.

Once estimated the fixed parameters in model (5), $\boldsymbol{\beta}^*$, we are ready to estimate the parameters $\boldsymbol{\theta}$. The random effect parameters are estimated using the same procedure presented in Section 3.1.2. In Section 3.3, we presented several structures that $\boldsymbol{R}(\boldsymbol{\vartheta})$ can take, where $\boldsymbol{\vartheta}$ is involved in $\boldsymbol{\theta}$. Additionally, the scale parameter ϕ can be estimated by

$$\hat{\phi} = \frac{1}{nT - p^*}\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}} r_{it}^2, \tag{30}$$

where

$$r_{it} = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{v(\hat{\mu}_{it})}}.$$

3.3 *Specific choices of $\boldsymbol{R}(\boldsymbol{\vartheta})$*

Here, several specific choices of $\boldsymbol{R}(\boldsymbol{\vartheta})$ are discussed. The number of nuisance parameters and the estimator of $\boldsymbol{\vartheta}$ vary from case to case. We only present some of the most used correlation structures in the practice problems, but other structures like the proposed by Mukherjee [31] can be used.

Example 3.1 Independence between times. When $\boldsymbol{R}(\boldsymbol{\vartheta}) = \boldsymbol{R}_0 = \boldsymbol{I}_T$, the identity matrix, we obtain the independence estimating equation along time but not in the space. Note that for any specified $\boldsymbol{R}_0$, no knowledge on ϕ is required in estimating $\boldsymbol{\beta}^*$ and $\mathrm{Var}(\boldsymbol{\beta}^*)$.

Example 3.2 l-Dependence. Let $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{T-1})^t$, where $\vartheta_t = \mathrm{Corr}(y_{it}, y_{i(t+1)})$ for $t = 1, \ldots, T-1$. A natural estimator of ϑ_t, given $\boldsymbol{\beta}^*$ and ϕ, is

$$\hat{\vartheta}_t = \frac{1}{\phi(n-p^*)} \sum_{i=1}^{n} \hat{r}_{it}\hat{r}_{i(t+1)}, \quad t = 1, \ldots, T-1.$$

As a special case, we can set $q = 1$ and $\vartheta_t = \vartheta$ $(t = 1, \ldots, T-1)$. Then, the common ϑ can be estimated by

$$\hat{\vartheta} = \frac{1}{T-1} \sum_{t=1}^{T-1} \hat{\vartheta}_t.$$

An extension to l-dependence is straightforward.

Example 3.3 Compound symmetry. Let $q = 1$ and assume that $\mathrm{Corr}(y_{it}, y_{it'}) = \vartheta$ for all $t \neq t'$. This is the exchangeable correlation structure obtained from a random effects model with a random level for each subject [23]. Given ϕ, ϑ can be estimated by

$$\hat{\vartheta} = \frac{1}{\phi[(1/2)T(T-1) - p^*]} \sum_{i=1}^{n} \sum_{t>t'} \hat{r}_{it}\hat{r}_{it'}.$$

Note that an arbitrary number of observations and observation times for each subject are possible with this assumption.

Example 3.4 First-order autoregressive. Let $\mathrm{Corr}(y_{it}, y_{it'}) = \vartheta^{|t-t'|}$, then $\mathrm{E}(\hat{r}_{it}\hat{r}_{it'}) \approx \vartheta^{|t-t'|}$, and ϑ can be estimated by the slope from the regression of $\log(\hat{r}_{it}\hat{r}_{it'})$ onto $\log|t-t'|$. Note that an arbitrary number and spacing of observations can be accommodated within this working structure.

Example 3.5 Unstructured. Let $\boldsymbol{R}(\boldsymbol{\vartheta})$ be totally unspecified, that is $q = \frac{1}{2}T(T-1)$. Now $\boldsymbol{R}(\boldsymbol{\vartheta})$ can be estimated by

$$\frac{1}{n\phi} \sum_{i=1}^{n} \mathbf{A}_i^{-1/2}(\boldsymbol{\mu}_i - \boldsymbol{y}_i)(\boldsymbol{\mu}_i - \boldsymbol{y}_i)^t \mathbf{A}_i^{-1/2},$$

where $\mathbf{A}_i = \mathrm{diag}(v(\bar{\mu}_i))$ is a $T \times T$ diagonal matrix with $\bar{\mu}_i = (1/T)\sum_{t=1}^{\mathrm{T}} \mu_{it}$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT})^t$, and $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iT})^t$.

Since the GEEs depend of the parameters $\boldsymbol{\beta}^*$ and $\boldsymbol{\theta}$, several steps are required in the estimation process:

(1) Obtain an initial estimation of $\boldsymbol{\beta}^*$, $\hat{\boldsymbol{\beta}}^{*(0)}$, using GLM assuming independence between observations and times, or equivalently, using Equation (29) with $\boldsymbol{R}(\boldsymbol{\theta}) = \boldsymbol{I}_{nT}$ and $\phi = 1$.
(2) Use $\hat{\boldsymbol{\beta}}^{*(0)}$ obtained in the previous step to calculate $\hat{\boldsymbol{\varepsilon}}_{\text{st}}$ using Equation (5).
(3) Use Equation (19) to find the parameter vector $\boldsymbol{\psi}$ and the appropriate correlation structure of $\boldsymbol{\Sigma}_T$ (see Examples 3.1– 3.5). We can use in this step Equations (20) and (21).
(4) Estimate the scale parameter using Equation (30).
(5) Obtain $\boldsymbol{R}(\boldsymbol{\theta})$ and estimate $\boldsymbol{\beta}^*$ using Equation (29).
(6) Repeat the steps 3–5 until to get convergence in the parameters $\boldsymbol{\beta}^*$ and $\boldsymbol{\theta}$.

4. Selection, validation and prediction of the fitted model using the space–time GEE

We should note that by construction GEE is a method that is not based on the use of the likelihood function. This makes that many tools designed for the construction of models within the field of likelihood cannot be used in the context of GEE. Thus, the well-known Akaike Information Criterion (AIC) cannot be directly applied, since AIC is based on maximum likelihood estimation, while GEE is non-likelihood based. Pan [37] presented an approach that is a modification of AIC, where the likelihood is replaced by the quasi-likelihood and a proper adjustment is made for the penalty term. This modification is given by

$$\text{QSTIC}(\boldsymbol{R}(\hat{\boldsymbol{\theta}}), \phi) \equiv -2\text{Q}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{I}_{nT}) + 2\text{trace}(\hat{\boldsymbol{\Omega}}_I \hat{\boldsymbol{\Gamma}}), \tag{31}$$

where

$$\text{Q}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{R}(\hat{\boldsymbol{\theta}}) = \boldsymbol{I}_{nT}) = \sum_{i=1}^{n}\sum_{t=1}^{\text{T}} \int_{y_{it}}^{\mu_{it}} \frac{y_{it} - \tau}{\hat{\phi} v(\tau)} \text{d}\tau,$$

$$\hat{\boldsymbol{\Omega}}_I = -\left.\frac{\partial^2 \text{Q}(\boldsymbol{\beta}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}; \boldsymbol{I}_{nT})}{\partial \boldsymbol{\beta}^* \partial \boldsymbol{\beta}^{*t}}\right|_{\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}^*}, \tag{32}$$

and $\hat{\boldsymbol{\Gamma}}$ was given above. This is the spatio-temporal quasi-likelihood expression under the independence model criterion (QSTIC) for GEE. Also, this criterion can be used to select the best structure of the matrix $\boldsymbol{R}(\boldsymbol{\vartheta})$ according to the data, or to select explanatory variables to be considered in the linear predictor of the model.

Another important point is hypothesis testing that can be raised in the form of the general linear hypothesis, given by

$$H_0 : \boldsymbol{L}\boldsymbol{\beta}^* = \boldsymbol{0} \quad \text{vs} \quad H_1 : \boldsymbol{L}\boldsymbol{\beta}^* \neq \boldsymbol{0},$$

where $\boldsymbol{L}$ is an $l \times p^*$ known matrix. One of the criteria mostly used and implemented in different data analysis Softwares is the Wald statistic [39,43], which in this case is given by

$$\text{W} = (\boldsymbol{L}\hat{\boldsymbol{\beta}}^*)^t [\boldsymbol{L}\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})))\boldsymbol{L}^t]^{-1} (\boldsymbol{L}\hat{\boldsymbol{\beta}}^*), \tag{33}$$

where $\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})))$ is given in Equation (27). This statistic follows an asymptotic χ^2 distribution with $\text{rank}(\boldsymbol{L}) = l$ degrees of freedom.

After fitting the GLSTARAR model, it is important to carry out a diagnostic analysis to verify the goodness of fit of the estimated model.

4.1 *Goodness-of-fit measure*

A global measure of explained variation is obtained by computing the pseudo R^2 defined as

$$R_k^2 = r^2(\hat{\boldsymbol{\eta}}_{\mathrm{st}}, g(\boldsymbol{y}_{\mathrm{st}})), \quad 0 \leq R_k^2 \leq 1, \tag{34}$$

where $r(\hat{\boldsymbol{\eta}}_{\mathrm{st}}, g(\boldsymbol{y}_{\mathrm{st}}))$ is the sample correlation coefficient between $\hat{\boldsymbol{\eta}}_{\mathrm{st}}$ and $g(\boldsymbol{y}_{\mathrm{st}})$. When $R_k^2 = 1$, there is a perfect agreement between $\hat{\boldsymbol{\eta}}_{\mathrm{st}}$ and $g(\boldsymbol{y}_{\mathrm{st}})$, hence between $\hat{\boldsymbol{\mu}}_{\mathrm{st}}$ and $\boldsymbol{y}_{\mathrm{st}}$.

4.2 *Residual analysis*

As it is known, the residual analysis aims at identifying violations of assumptions as nonhomogeneous variance, atypical observations and/or model misspecification. It can be based on ordinary residuals, $y_{it} - \hat{\mu}_{it}$, but in the GLSTARAR model this residual technically is not appropriate, since $\mathrm{Var}(y_{it})$ is not constant. So, the type of residual that is most intuitive is the Pearson residual [26,33,34], which is given by

$$r_{P_{it}} = r_P(\boldsymbol{s}_i, t) = \frac{y_{it} - \hat{\mu}_{it}}{\sqrt{\phi v(\hat{\mu}_{it})}}, \tag{35}$$

where $v(\hat{\mu}_{it})$ is a variance function.

On the other hand, based on the quasi-likelihood function presented in Equation (32), we can write

$$\mathrm{Q}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{I}_{nT}) = \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \mathrm{Q}_{it}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{I}_{nT}), \tag{36}$$

where $\mathrm{Q}_{it}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{I}_{nT}) = \int_{y_{it}}^{\mu_{it}} ((y_{it} - \tau)/\hat{\phi} v(\tau))\,\mathrm{d}\tau$. Analogous to the GLM, we can construct a deviance function based on a quasi-likelihood function. For a single observation, this function is defined as

$$r_{D_{it}} = r_D(\boldsymbol{s}_i, t) = -2\hat{\phi}\mathrm{Q}_{it}(\hat{\boldsymbol{\beta}}^*(\boldsymbol{R}(\hat{\boldsymbol{\theta}})), \hat{\phi}, \boldsymbol{I}_{nT}) = 2\int_{\mu_{it}}^{y_{it}} \frac{y_{it} - \tau}{v(\tau)}\mathrm{d}\tau, \tag{37}$$

which is used as a discrepancy measured between the fitted value, μ_{it}, and the observed value, y_{it}.

4.3 *Space-time prediction of new individuals*

Specifically, for the value $\boldsymbol{y}_0 = (y(\boldsymbol{s}_{n+1}, t_1), \ldots, y(\boldsymbol{s}_{n+n'}, t_{n'}))^t$ of a random field $\boldsymbol{Y}_0$ at l $(1 \leq t_l \leq T$ with $l = 1, \ldots, n')$ pre-specified space–time points from observations $y(\boldsymbol{s}_i, t)$, $i = 1, \ldots, n$ and $t = 1, \ldots, T$, we focus on interpolation of random effects over a space–time area when the observations are non-Gaussian. Thus, let $\boldsymbol{\eta}_{\mathrm{st}}^0 = (\eta(\boldsymbol{s}_{n+1}, t_1), \ldots, \eta(\boldsymbol{s}_{n+n'}, t_{n'}))^t$ be the functional prediction, and let $f(\boldsymbol{\eta}_{\mathrm{st}}^0, \boldsymbol{\eta}_{\mathrm{st}})$ be the joint density function of $\boldsymbol{\eta}_{\mathrm{st}}$ and a vector $\boldsymbol{\eta}_{\mathrm{st}}^0$. Let us confine our interest to pseudo-unbiased linear predictors of the form

$$\tilde{\boldsymbol{\eta}}_{\mathrm{st}} = \boldsymbol{p} + \boldsymbol{Q}\boldsymbol{\eta}_{\mathrm{st}}, \tag{38}$$

for some conformable vector $\boldsymbol{p}$ and matrix $\boldsymbol{Q}$ [28]; thus, minimising the mean-squared-error of the prediction, we find the pseudo-best linear unbiased predictor given by

$$\tilde{\boldsymbol{\eta}}_{\mathrm{st}} = \mathbf{X}^0\boldsymbol{\beta}^* + \mathrm{Cov}^t(\boldsymbol{\eta}_{\mathrm{st}}, \boldsymbol{\eta}_{\mathrm{st}}^0)\boldsymbol{\Sigma}_{\eta_{\mathrm{st}}}^{-1}[\boldsymbol{\eta}_{\mathrm{st}} - \mathbf{X}\boldsymbol{\beta}^*],$$

where $\boldsymbol{\Sigma}_{\eta_{\mathrm{st}}} = [\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}(\boldsymbol{\Sigma}_T \otimes \boldsymbol{I}_n)\{[\boldsymbol{I}_{nT} - (\boldsymbol{\Psi} \otimes \boldsymbol{W}_2)]^{-1}\}^t$, $\mathbf{X}^0 = (\boldsymbol{X}_{\mathrm{st}}^0, \bigoplus_{t=1}^{\mathrm{T}} \boldsymbol{W}_1\boldsymbol{\eta}_t^0)$, with $\boldsymbol{X}_{\mathrm{st}}^0$ being a matrix of $p_1 + p_{2_t} + 1$ explanatory variables for the n' new space–time subjects including a vector of 1's, that is, $\mathbf{1}_{n'}$ is of size $n' \times 1$, and $\boldsymbol{\eta}_t^0 = (\eta_{1t}^0, \ldots, \eta_{n't}^0)$.

The covariance matrix for the prediction has the following general form:

$$\mathrm{Var}(\tilde{\boldsymbol{\eta}}_{\mathrm{st}} \mid \boldsymbol{y}_{\mathrm{st}}) \approx \boldsymbol{\Sigma}_0 + \mathrm{Cov}^t(\boldsymbol{\eta}_{\mathrm{st}}, \boldsymbol{\eta}^0)\boldsymbol{\Sigma}_{\eta_{\mathrm{st}}}^{-1}\mathrm{Cov}(\boldsymbol{\eta}_{\mathrm{st}}, \boldsymbol{\eta}_{\mathrm{st}}^0),$$

where $\boldsymbol{\Sigma}_0 = \mathrm{Var}(\boldsymbol{\eta}_{\mathrm{st}}^0) - \mathrm{Cov}^t(\boldsymbol{\eta}_{\mathrm{st}}, \boldsymbol{\eta}_{\mathrm{st}}^0)\boldsymbol{\Sigma}_{\eta_{\mathrm{st}}}^{-1}\mathrm{Cov}(\boldsymbol{\eta}_{\mathrm{st}}, \boldsymbol{\eta}_{\mathrm{st}}^0)$.

5. Real data analysis

Recalling the data set description given in the Introduction, we have a clear panel structure. In particular, we study the number of AAs per 1000 km^2 committed by the guerrillas of the FARC-EP and ELN in Colombia from 2003 to 2009. We use a GLSTARAR model to fit the number of AAs by department throughout the study period, and analyse the causes of the presence and expansion of these armed groups for the years 2003–2009. Our model is of a spatio-temporal type since we introduce variables that capture the effects generated by the space–time autocorrelation, and the space–time heterogeneity.

The FARC-EP and the ELN are guerrilla groups that operate in Colombia and in the border regions of Brazil, Ecuador, Panama, Peru and Venezuela. They are participants of the armed conflicts since their formation. Their actions include drug trafficking, guerrilla warfare and terrorist techniques as the implementation of anti-personnel mines, the killing of civilians, government officials, police and military, kidnapping for extortion or political, attacks with bombings and unconventional weapons (gas cylinders and bomb animals), and a number of acts that have caused forced displacement of civilians [18].

Colombia invested in military spending 11,128 million dollars in 2009, which corresponded to 3.9% of the gross domestic product (GDP), one of the greatest in Latin America dedicated to this area. However, this expenditure was well below of Brazil, which in 2009 spent 34,334 millions [41]. In 2012, the Colombian military spending remained as of 3.9% of the GDP. Thus, in these last years, Colombia has spent a good portion of its GDP to combat guerrillas.

5.1 *Exploring the spatio-temporal structure*

Much of the literature on exploratory spatio-temporal data analysis has focussed on the exploration of areal-time data with respect to the spatio-temporal association. In this section, we look at local indicators of spatio-temporal association within this line, but will also consider how larger scale regularities may be revealed by using the Moran mapping. A topical area that has not been given enough attention is that of regression diagnostics for fitted spatio-temporal regression models [20]. While users appear to want heteroscedasticity-corrected standard errors, few seem to realise that the miss-specification could arguably be better handled if diagnostic methods were used (see also Mur and Lauridsen [32]).

We first perform a descriptive analysis to analyse the data behaviour, and to identify possible weight matrices $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ involved in the GLSTARAR model. The department of the islands of San Andres and Providencia were not considered in the analysis because the number of AAs was null throughout the period of study, and these islands are far from the land of the Atlantic coast. Therefore, this department was not considered in our analysis because it could generate unnecessary noise.

Figure 1 shows the number of AAs in quintiles by department. The regions with similar values in the number of AAs correspond to the same quintile; this behaviour can be seen throughout time. Therefore, dark colors indicate high values for the number of AAs, while light colors low values. In general, low values are found throughout the Amazon region (departments of Putumayo, Caqueta, Amazonas, Vaupes and Guainia), located at the bottom right of Figure 1. At the same time, high values are located in the Andes Mountains; specifically, for 2009 in

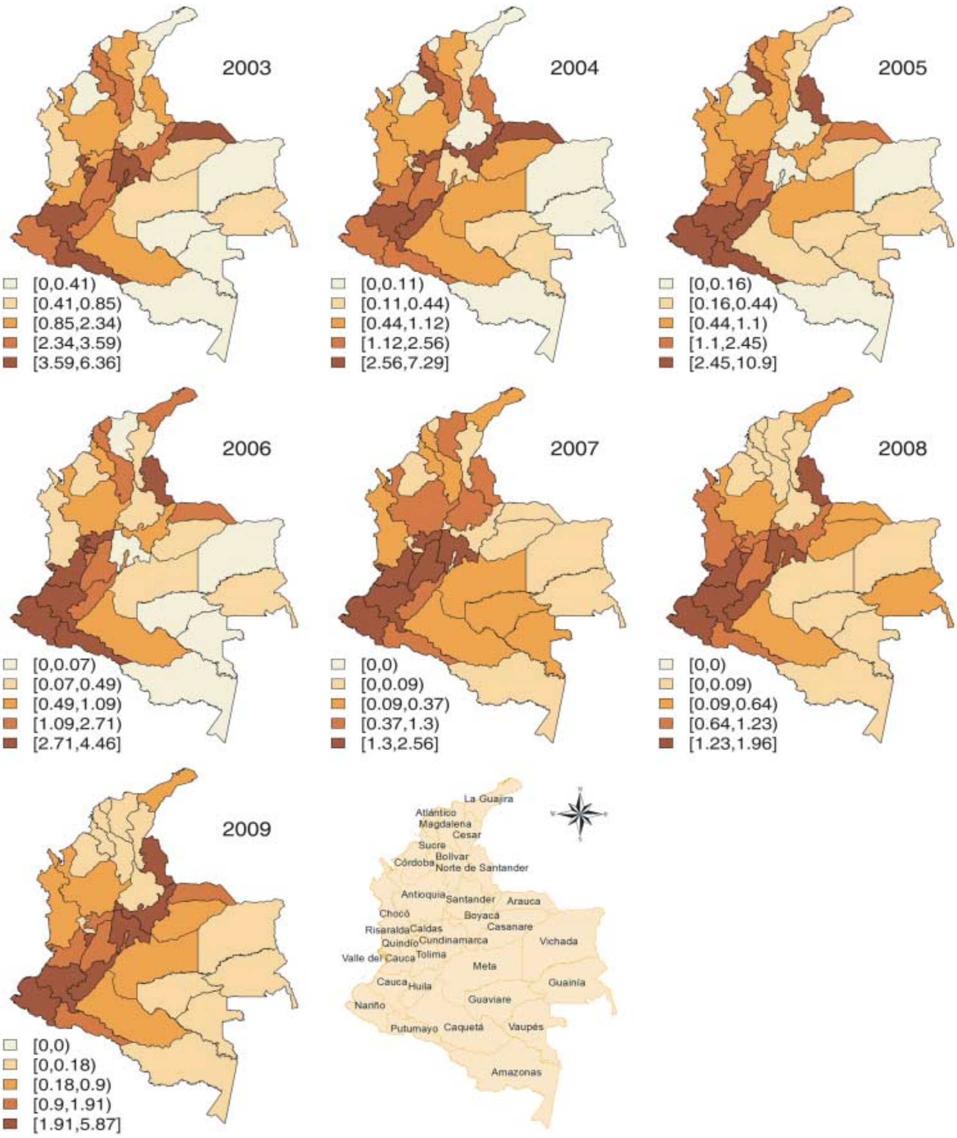

Figure 1. Maps of the number of AAs by department, in quintiles, for the period 2003–2009, and map of the departmental administrative political division of Colombia (last map).

the Eastern Cordillera. On the Atlantic coast, we find high values for the departments of Sucre and Bolivar in 2003 and 2004. With regard to the Pacific region, there are higher values in the period 2006–2009 for the departments of Nariño, Cauca and Valle del Cauca, together with the departments of Huila, Tolima and Cundinamarca in the Andean region (see Figure 1). The number of AAs clearly decreases over time on the Atlantic coast, while the numbers increase on the Eastern Cordillera.

On the other hand, the spatial dependence structures can be generated from the spatial relationships between departments, and these relationships are used to find the spatial weights matrix. There are three common ways to establish these structures: the queen contiguity effect of orders 1, 2 or 3, the *k*-nearest neighbours to each department considering 1, 2 or 4 nearest neighbours, and the threshold distance which is constructed from the thresholds 100, 200 and 300 km.

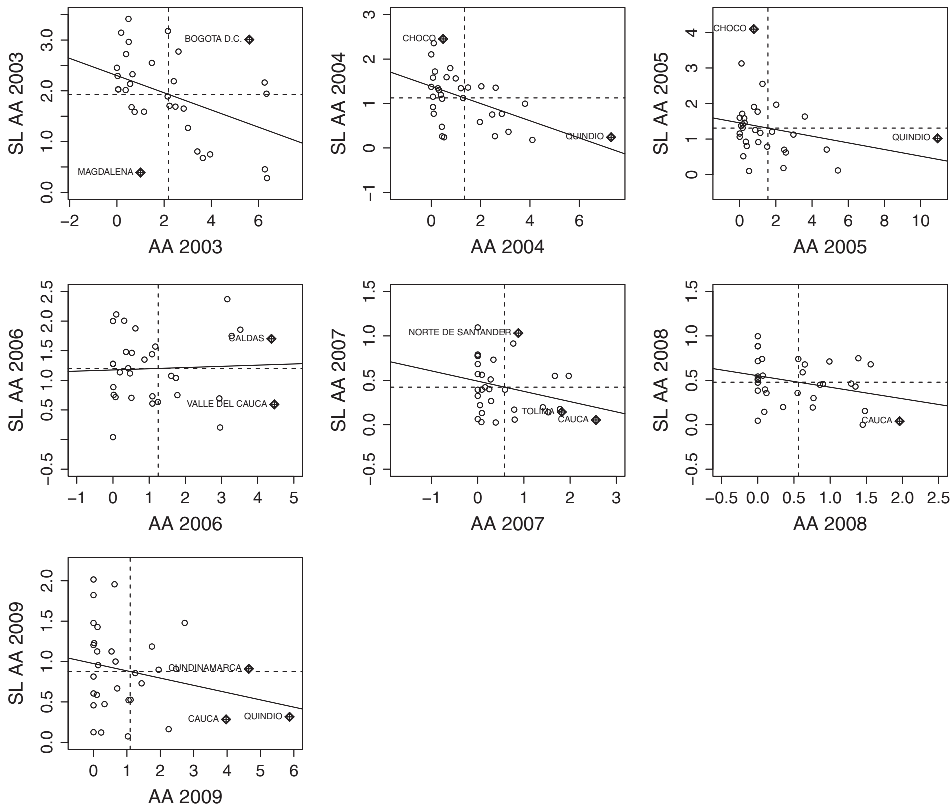

Figure 2. Moran scatter plots for the number of AAs by department in the period 2003–2009.

Figure 2 displays the Moran scatter plots for the number of AAs in the period 2003–2009. The different values of Moran's index I_t are obtained from the following equation:

$$I_t = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}[y(\boldsymbol{s}_i, t) - \bar{y}_{\bullet t}][y(\boldsymbol{s}_j, t) - \bar{y}_{\bullet t}]}{S_0 \sum_{i=1}^{n} [y(\boldsymbol{s}_i, t) - \bar{y}_{\bullet t}]^2}, \quad t = 1, \ldots, T,$$

where $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$ and $\bar{y}_{\bullet t} = (1/n) \sum_{i=1}^{n} y(\boldsymbol{s}_i, t)$.

Each scatter plot is built considering the number of AAs and its SL. In addition, Figure 2 allows us to identify influential points in the space–time domain. These points are obtained after calculating the statistical Dffits and covariance rate (cov.r) of the residuals obtained from the linear regression model between the number of AAs and its SL. The Moran scattergram shows the interrelationships between each of the polygons and their neighbours; the slope of the regression model mentioned is usually called Moran's I_t. Then, if the slope or Moran's I_t is positive indicates a predominance of spatial concentrations of similar values for the number of AAs, either high values surrounded by high values or low values surrounded low values. A negative slope indicates a concentration of the dissimilar values in the number of AAs; thus, we have low values surrounded by high values or vice versa. For example, in 2003, Figure 2 indicates a negative spatial autocorrelation since a spatial concentration with dissimilar values in the number of AAs dominates. In addition, Figure 2 shows that Bogotá and Magdalena are influential departments since their concentrations are associated with similar values, with high values surrounded by high values and low values surrounded by low values, respectively. In the analysed periods, only in 2006, we do not observe a slope significantly different from 0, which indicates the absence

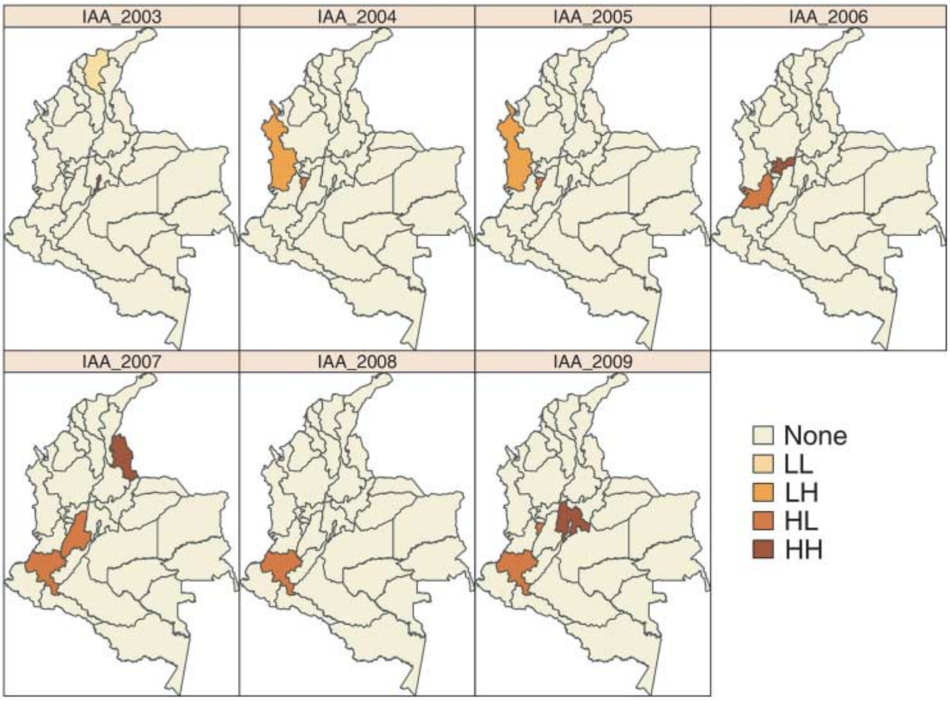

Figure 3. Influential maps using the Moran scatter quadrants for the number of AAs by department in the period 2003–2009.

of spatial autocorrelation for that particular year. For the other years, the slopes are significantly different from 0, and Chocó, Cauca and Quindio are highlighted as influential departments.

Once established the influential departments with respect to the spatial structure, we determine the higher (H) and lower (L) values for each of the polygons throughout time. Subsequently, we cross the SLs associated with their neighbourhoods. If a region has a high value, and it is surrounded by high average values for neighbouring regions, then this region is denoted by 'HH'. If a region has a low value, and it is surrounded by low average values for neighbouring regions, then this region is denoted by 'LL'. If a region has a high value, and it is surrounded by low average values for neighbouring regions, then this region is denoted by 'HL'. Finally, if a region has a low value and it is surrounded by high average values for neighbouring regions, then this region is denoted by 'LH'. Therefore, we note that in the years 2003 and 2009, there are two influential departments indicated by Moran's scatter quadrant (Bogotá and Magdalena, see Figure 3). Moreover, in the years 2004 and 2005, Chocó and Quindio are influential departments. From 2004 to 2009, the influential departments correspond to dissimilar values, that is, 'HL' or 'LH'. These results are shown in Figure 3.

Figure 4 shows the behaviour of the number of AAs for the period 2003–2009. In this period, there exists evidence of a decrease in the levels of AAs, indicating greater control by the public forces (police and military forces) and better policies in the control of insurgent groups.

5.2 *The GLSTARAR model*

Spatio-temporal residual dependence was handled as follows. Let $y(\boldsymbol{s}_i, t)$ be the number of AAs at the $\boldsymbol{s}_i$th department in the tth time ($\boldsymbol{s}_1 = (w_{x_1}, w_{y_1}), \ldots, \boldsymbol{s}_{32} = (w_{x_{32}}, w_{y_{32}})$ and $t = 1, \ldots, 7$). The model assumes that the $y(\boldsymbol{s}_i, t)$'s are independent Poisson random variables given the unobserved

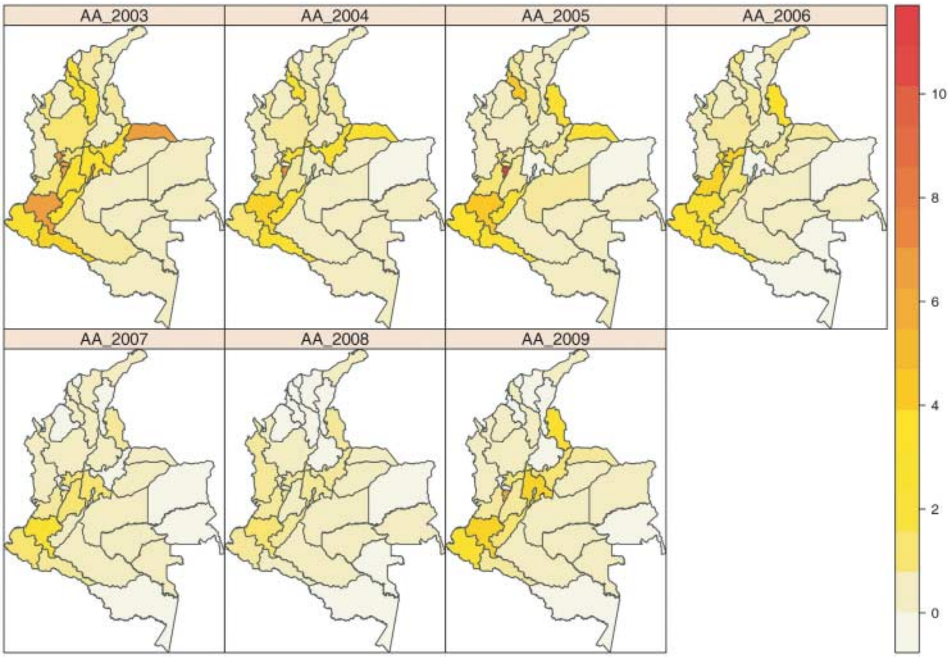

Figure 4. Maps for the number of AAs by department in the period 2003–2009.

space–time stochastic process ε_{it}. Also, the mean response in $(\boldsymbol{s}_i, t)$ depends on the observed explanatory variables at the location $\boldsymbol{s}_i$ and time t. Specifically, we can write the GLSTARAR model as

$$\log y(\boldsymbol{s}_i, t) = \beta_0 + \boldsymbol{x}_{2it}^t \boldsymbol{\beta}_t + \pi_t \sum_{i'=1}^{32} w_{ii'}^{(1)} \log y(\boldsymbol{s}_{i'}, t) + \varepsilon_{it},$$
$$\varepsilon_{it} = \varepsilon(\boldsymbol{s}_i, t) = \psi_t \sum_{i'=1}^{32} w_{ii'}^{(2)} \varepsilon_{i't} + e_{it}, \tag{39}$$

with $i = 1, \dots, 32$, $t = 1, \dots, 7$, β_0 is an unknown parameter associated with the intercept, $\boldsymbol{x}_{2it}^t = \boldsymbol{x}_2^t(\boldsymbol{s}_i, t) = (AVV2003_{i1}, FM2003_{i1}, \dots, DFHR2009_{i7})$ is a vector containing explanatory variables associated with the space–time location $\boldsymbol{s}_i$ and $\boldsymbol{\beta}_t = (\beta_{11}, \dots, \beta_{7p_{2_7}})^t$ is a vector of unknown space–time regression parameters. Also, $w_{ii'}^{(1)} = w_{ii'}^{(2)}$ are the weights using a SL with first-order queen contiguity (see details in [3]), π_t is the spatial autoregressive coefficient in the tth time, ε_{it} is the spatio-temporal autocorrelated error term at the $\boldsymbol{s}_i$th location in the tth time, ψ_t is the coefficient of spatial autocorrelation in the tth time, and $e_{it} = e(\boldsymbol{s}_i, t)$ is an i.i.d. normal random error associated with the $\boldsymbol{s}_i$th location in the tth time with zero mean and covariance $\mathrm{E}(e_{it}, e_{it'}) = \sigma_{tt'}$ for $t, t' = 1, \dots, 7$, with $\mathrm{E}(e_{it}, e_{i't}) = 0$ for $i, i' = 1, \dots, 32$.

We used the GEE method as commented in Section 3 to obtain the estimated parameters associated with the GLSTARAR model given in Equation (39). The results are shown in Table 2. In this model, all correlation structures of errors e_{it}, as shown in Section 3.3, were considered. However, none of the parameters of the correlation matrices were significant to 5%, and thus these errors were considered independent. The pseudo $R^2 = 69.85\%$ indicates a satisfactory goodness of fit for the GLSTARAR model.

Table 1. Estimated parameters using the GLSTARAR model with a space–time error, obtained by means of the GEE method.

Weight lag	Coefficients Estimate	Std. err	Wald	Pr(> |W|)
$w\varepsilon_{2003}$	0.1932	0.0227	72.4	$< 2e-16$
$w\varepsilon_{2004}$	0.1839	0.0214	74.0	$< 2e-16$
$w\varepsilon_{2005}$	0.1897	0.0268	50.2	$1.4e-12$
$w\varepsilon_{2006}$	0.1917	0.0171	125.1	$< 2e-16$
$w\varepsilon_{2007}$	0.1659	0.0316	27.6	$1.5e-07$
$w\varepsilon_{2008}$	0.1980	0.0158	156.2	$< 2e-16$
$w\varepsilon_{2009}$	0.1737	0.0279	38.6	$5.2e-10$
Estimated scale parameters (intercept)	0.837	0.434		

According to Table 1, all SLs of the errors were significant at the 5% ($w\varepsilon_{2003}, \ldots, w\varepsilon_{2009}$). Since the SL coefficients through time were different, we note an interaction between space and time in the space–time error model. However, these coefficients looked quite similar, and thus a test of equal slopes might be considered. In this paper, we do not develop such a test, but a future work could develop a test to judge asymptotic equality among the parameters. The signs of all the estimated parameters were positive. This positive sign can be interpreted as a department with high/low error surrounded by departments with high/low errors. The scale parameter was $\hat{\sigma}^2 = 0.837$, meaning that the dispersion was quite low.

Table 2 shows that all the SLs of the log(AA) were significant at the 5%. Contrary to the coefficients of the SLs of the errors being all positive (as shown in Table 1), now the signs of the coefficients vary over time indicating an interaction between space and time in the SL model. Thus, in the year 2009, for example, $w \log y(\boldsymbol{s}, 2009)$ has a positive sign indicating that a certain department has a high/low number of AAs on the FARC and ELN and their neighbours again have a high/low number of AAs. However, in 2008 $w \log y(\boldsymbol{s}, 2008)$ has a negative sign which means that a department with a high number of AAs of the FARC and ELN is surrounded by neighbours with a low number of AAs or vice versa. The scale parameter is $\hat{\phi} = 0.615$, a low value indicating that the problem with overdispersion is not important. Note also that all explanatory variables in different years were significant at the 5%. Since the coefficients of the explanatory variables change along time, we underline an interaction between space and time for the different explanatory variables. A variable that shows a positive sign along time is the total armed confrontations per year, indicating that a larger amount of armed confrontations corresponds to a larger number of AAs per 1000 km^2 of the FARC and ELN guerrilla groups.

5.3 *Validation of GLSTARAR model assumptions*

We note in Figure 4 that the value of the number of AAs for the Cundinamarca department is around 5, while the prediction using the GLSTARAR model is approximately 2 (see Figure 5), that is, there is an approximate error of 3. This error occurs because Cundinamarca is an influential data, as shown in Moran's scatter plot (see Figure 2). Given the spatial contiguity characteristics of the GLSTARAR model, the influential or outlier data cannot be removed; however, this is an interesting issue to consider in future works.

In general, the predictions of the number of AAs generated by the GLSTARAR model show good results, as shown in Figure 5 (prediction maps), Figure 6 (Pearson residuals maps) and Figure 7 (deviance residuals maps). The Pearson residuals lie mostly within the interval $(-3, 3)$, and the deviance residuals within the interval $(-2, 2)$. Again, the Cundinamarca department for

Table 2. Estimated parameters using the GLSTARAR model with a spatio-temporal lag, obtained by means of the GEE method.

Component	Coefficients Estimate	Std.err	Wald	Pr(> |W|)
(Intercept)	− 8.48e − 01	1.93e − 07	1.94e + 13	< 2e − 16
$w \log y(\boldsymbol{s}, 2003)$	4.28e − 02	2.85e − 09	2.24e + 14	< 2e − 16
$w \log y(\boldsymbol{s}, 2004)$	1.19e − 02	3.15e − 09	1.43e + 13	< 2e − 16
$w \log y(\boldsymbol{s}, 2005)$	2.11e − 02	6.12e − 09	1.19e + 13	< 2e − 16
$w \log y(\boldsymbol{s}, 2006)$	− 9.95e − 03	5.56e − 09	3.20e + 12	< 2e − 16
$w \log y(\boldsymbol{s}, 2007)$	− 2.93e − 02	4.16e − 07	4.97e + 09	< 2e − 16
$w \log y(\boldsymbol{s}, 2008)$	− 4.49e − 02	2.18e − 08	4.23e + 12	< 2e − 16
$w \log y(\boldsymbol{s}, 2009)$	1.88e − 02	1.24e − 08	2.29e + 12	< 2e − 16
ARVV2003	1.15e − 01	5.58e − 09	4.27e + 14	< 2e − 16
ARVV2004	− 1.50e − 03	1.42e − 10	1.12e + 14	< 2e − 16
ARVV2005	− 3.93e − 02	1.32e − 09	8.87e + 14	< 2e − 16
ARVV2006	4.77e − 02	3.21e − 09	2.21e + 14	< 2e − 16
ARVV2007	1.00e − 02	6.89e − 09	2.11e + 12	< 2e − 16
ARVV2008	− 1.88e − 02	5.02e − 08	1.40e + 11	< 2e − 16
ARVV2009	− 4.59e − 02	6.41e − 09	5.13e + 13	< 2e − 16
FDHE2003	1.60e − 03	1.10e − 10	2.11e + 14	< 2e − 16
FDHE2004	4.15e − 03	3.98e − 10	1.08e + 14	< 2e − 16
FDHE2005	7.51e − 03	2.98e − 10	6.33e + 14	< 2e − 16
FDHE2006	5.08e − 03	3.16e − 10	2.59e + 14	< 2e − 16
FDHE2007	− 8.32e − 04	3.85e − 09	4.66e + 10	< 2e − 16
FDHE2008	2.86e − 03	1.37e − 09	4.35e + 12	< 2e − 16
FDHE2009	1.12e − 02	1.69e − 09	4.40e + 13	< 2e − 16
FDHR2003	− 1.73e − 04	2.73e − 11	4.05e + 13	< 2e − 16
FDHR2004	− 1.38e − 03	7.02e − 11	3.89e + 14	< 2e − 16
FDHR2005	3.06e − 04	2.23e − 10	1.89e + 12	< 2e − 16
FDHR2006	− 3.04e − 04	8.46e − 13	1.29e + 17	< 2e − 16
FDHR2007	− 1.60e − 03	4.03e − 08	1.57e + 09	< 2e − 16
FDHR2008	− 9.97e − 04	2.91e − 10	1.18e + 13	< 2e − 16
FDHR2009	− 5.82e − 04	7.49e − 10	6.04e + 11	< 2e − 16
TAC2003	2.02e − 01	4.57e − 09	1.96e + 15	< 2e − 16
TAC2004	5.87e − 01	1.84e − 08	1.02e + 15	< 2e − 16
TAC2005	9.05e − 02	2.69e − 08	1.13e + 13	< 2e − 16
TAC2006	2.91e − 01	7.26e − 10	1.61e + 17	< 2e − 16
TAC2007	9.07e − 01	1.37e − 06	4.37e + 11	< 2e − 16
TAC2008	8.41e − 01	1.46e − 07	3.34e + 13	< 2e − 16
TAC2009	1.14e − 01	1.60e − 08	5.06e + 13	< 2e − 16
NAAMF2003	− 1.57e − 01	7.07e − 09	4.93e + 14	< 2e − 16
NAAMF2004	− 6.63e − 01	2.96e − 08	5.02e + 14	< 2e − 16
NAAMF2005	1.41e − 01	4.87e − 08	8.34e + 12	< 2e − 16
NAAMF2006	− 1.95e − 01	5.39e − 10	1.31e + 17	< 2e − 16
NAAMF2007	− 6.97e − 01	1.30e − 06	2.88e + 11	< 2e − 16
NAAMF2008	− 5.07e − 01	1.78e − 07	8.09e + 12	< 2e − 16
NAAMF2009	1.13e − 01	3.19e − 08	1.25e + 13	< 2e − 16
PPLUA2003	− 6.21e − 01	1.97e − 07	9.97e + 12	< 2e − 16
PPLUA2004	− 4.65e − 01	1.99e − 07	5.44e + 12	< 2e − 16
PPLUA2005	− 1.31e + 00	8.74e − 08	2.23e + 14	< 2e − 16
PPLUA2006	− 2.25e + 00	6.23e − 08	1.30e + 15	< 2e − 16
PPLUA2007	− 2.80e + 00	7.40e − 06	1.43e + 11	< 2e − 16
PPLUA2008	− 2.38e + 00	3.86e − 07	3.79e + 13	< 2e − 16
Estimated scale parameters				
(Intercept)	0.615	3.24e − 08		

the year 2009 showed high residuals in both the Pearson (close to 3) and the deviance (close to 2) residuals.

Moreover, Figure 8 shows a good fit because there are no outliers falling outside the rank −2 to 2 according to deviance residuals. Furthermore, the standardized residuals, $\boldsymbol{r}_{\text{st}} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{K}(\boldsymbol{y}_{\text{st}} - \hat{\boldsymbol{\mu}}_{\text{st}})$, are given by using the singular value decomposition, where $\boldsymbol{\Lambda}$ and $\boldsymbol{K}$ are

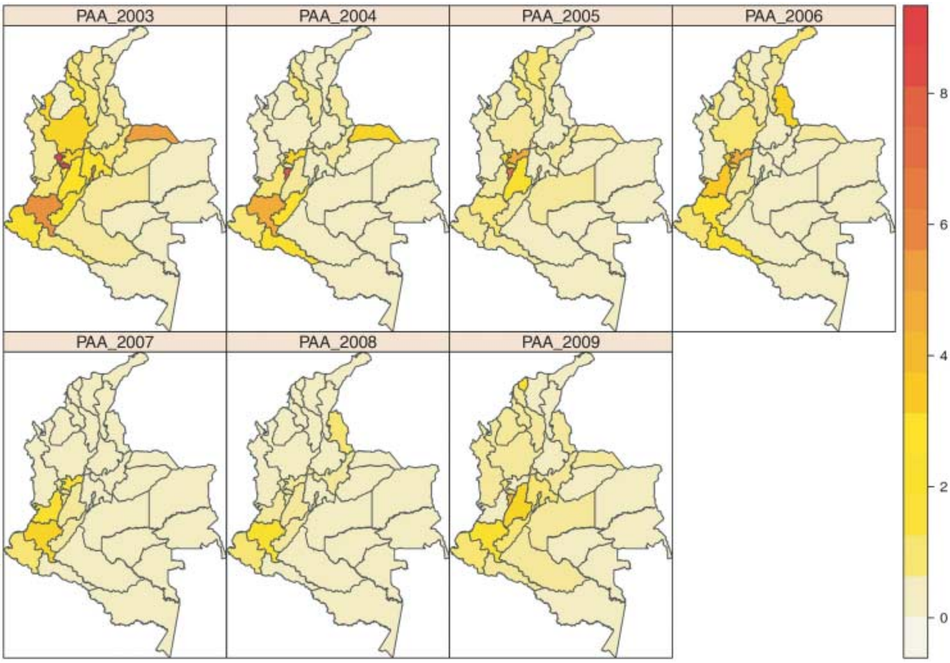

Figure 5. Prediction maps for the number of AAs by department under the GLSTARAR model in the period 2003–2009.

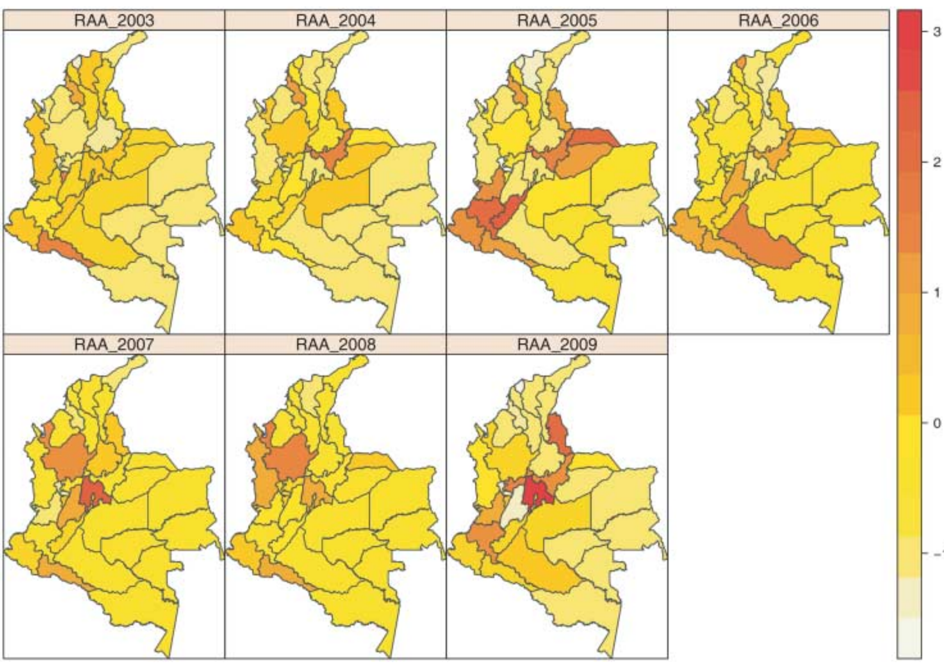

Figure 6. Pearson residual maps by department under the GLSTARAR model in the period 2003–2009.

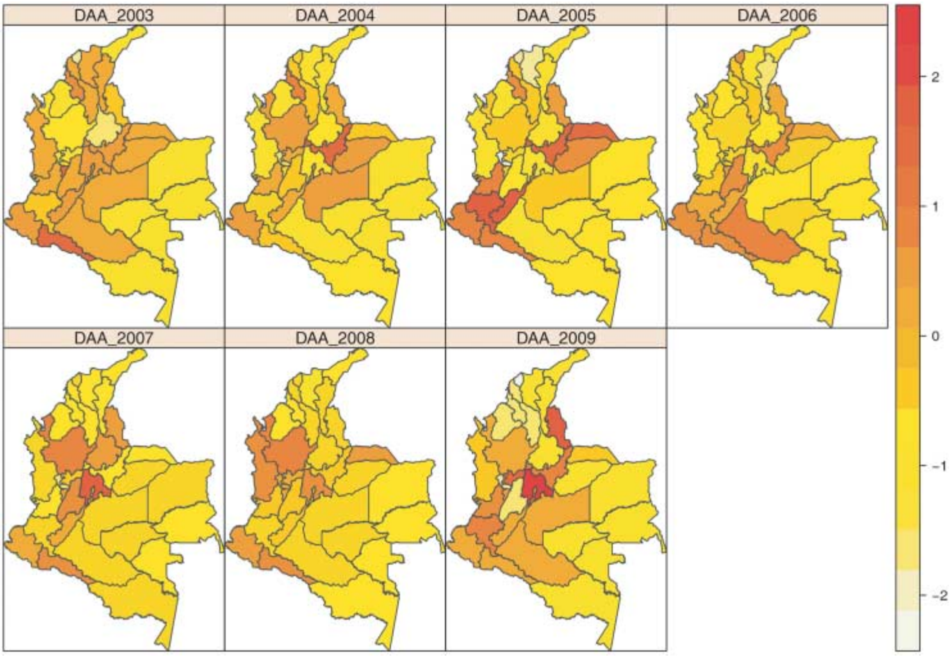

Figure 7. Deviance maps by department under the GLSTARAR model in the period 2003–2009.

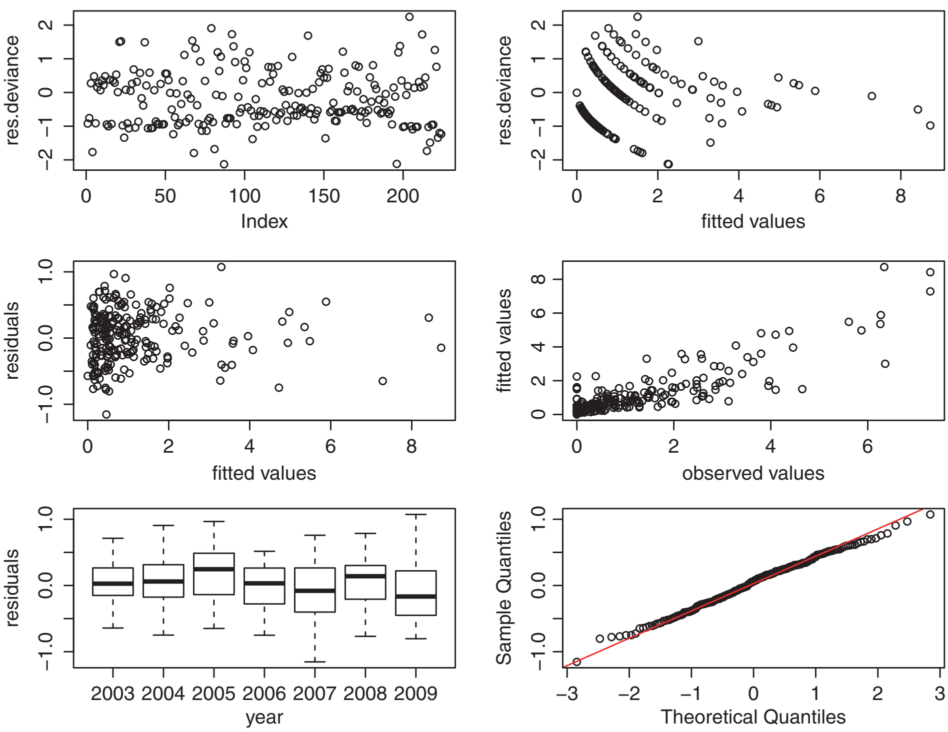

Figure 8. Diagnostics for the GLSTARAR model.

the diagonal matrix of eigenvalues and eigenvectors, respectively, of the variances–covariances $[\boldsymbol{I}_{nT} - (\hat{\boldsymbol{\Psi}} \otimes \boldsymbol{W}_2)]^{-1}(\hat{\boldsymbol{\Sigma}}_T \otimes \boldsymbol{I}_n)[\boldsymbol{I}_{nT} - (\hat{\boldsymbol{\Psi}} \otimes \boldsymbol{W}_2)]^{-1}$. Then, these residuals are exactly analogous to the common residuals used in multiple regression, and we can employ the traditional test of homogeneity and normality. Thus, according to Figure 8, we can say that there are homogeneity of variances through years (Bartlett test, p-value$=$ 0.59, and Levene's test, $p-$value $= 0.65$), and the normal probability plot does not show apparent problems of normality because residuals are within close to line (Shapiro–Wilk normality test, $p-$value $= 0.70$). Additionally, the observed data and fitted values are similar, a good indication of an adequate fit.

6. Joint space–time modelling of mean and dispersion

According to McCullagh and Nelder [26], the idea of using a linked pair of generalised linear models for the simultaneous modelling of mean and dispersion effects was first put forward by Pregibon [38]. For linear models with normal errors the idea is much older, a simple case being that of heterogeneous variances defined by a grouping factor; see Aitkin [1] for a general treatment or Cook and Weisberg [13], who discussed score tests. Smyth [42] compared different algorithms for the estimation of mean and dispersion effects.

Therefore, in the model given in Equation (1), the variance of a response variable has the form

$$\mathrm{Var}(y_{it}) = \phi v(\mu_{it}),$$

where $v(\mu_{it})$ is a known variance function. The choice of variance function determines the interpretation of ϕ; so, for example, if $v(\mu_{it}) = 1$, then ϕ is the response variance or if $v(\mu_{it}) = \mu_{it}$, then ϕ is the squared coefficient of variation of the variable response [26]. In our generalised linear space–time autoregressive model (Equation (1)), the dispersion parameter ϕ is a constant, but in circumstances where Y_{it} is the average of observations it may be appropriate to assume that ϕ_{it} is proportional to known weights. More generally, ϕ_{it} varies in a systematic way with other measured covariates. Therefore, formal models can construct and fit for the dependence of both μ_{it} and ϕ_{it} on several covariates, following suggestions such as the proposed by McCullagh and Nelder [26], McCulloch and Searle [27], and Pregibon [38].

The joint model is specified in terms of the dependence on covariates of the first two moments. For the mean, the model is given by

$$\begin{aligned} \mathrm{E}[y_{it}|\boldsymbol{x}_{1i}, \boldsymbol{x}_{2it}, \varepsilon_{it}] &= \mu_{it}, \quad \eta_{it} = g(\mu_{it}) = \boldsymbol{x}_{1i}^t\boldsymbol{\beta}_0 + \boldsymbol{x}_{2it}^t\boldsymbol{\beta}_t + \pi_t \sum_{i'=1}^{n} w_{ii'}^{(1)} \eta_{i't} + \varepsilon_{it}, \\ \mathrm{Var}(y_{it}|\boldsymbol{x}_{1i}, \boldsymbol{x}_{2it}, \varepsilon_{it}) &= \phi_{it} v(\mu_{it}), \end{aligned} \tag{40}$$

where $\varepsilon_{it} = \psi_t \sum_{i'=1}^{n} w_{ii'}^{(2)} \varepsilon_{i't} + e_{it}$, and the same assumptions of model (1). The dispersion parameter is no longer assumed constant but instead is assumed to vary in the following model:

$$\begin{aligned} \mathrm{E}[d_{it}|\boldsymbol{u}_{1i}, \boldsymbol{u}_{2it}, \epsilon_{it}] &= \phi_{it}, \quad \zeta_{it} = h(\phi_{it}) = \boldsymbol{u}_{1i}^t\boldsymbol{\gamma}_0 + \boldsymbol{u}_{2it}^t\boldsymbol{\gamma}_t + \varrho_t \sum_{i'=1}^{n} w_{ii'}^{(1)} \zeta_{i't} + \epsilon_{it}, \\ \mathrm{Var}(d_{it}|\boldsymbol{u}_{1i}, \boldsymbol{u}_{2it}, \epsilon_{it}) &= \delta v(\phi_{it}), \quad i = 1, \ldots, n, \ \ t = 1, \ldots, T, \end{aligned} \tag{41}$$

where $\epsilon_{it} = \varphi_t \sum_{i'=1}^{n} w_{ii'}^{(2)} \epsilon_{i't} + \xi_{it}$ with $|\varrho_t| < 1$ and $|\varphi_t| < 1$, $d_{it} \equiv d_{it}(y_{it}; \mu_{it})$ is a suitable statistic chosen as a measure of dispersion at $\boldsymbol{s}_i$-th location and tth time, ζ_{it} is the dispersion linear predictor at $\boldsymbol{s}_i$th location and tth time, $h(\cdot)$ is the dispersion link function which is invertible and continuous, $\boldsymbol{u}_{1i}^t\boldsymbol{\gamma}_0 + \boldsymbol{u}_{2it}^t\boldsymbol{\gamma}_t$ is the trend, $\boldsymbol{u}_{1i}^t = \boldsymbol{u}_1^t(\boldsymbol{s}_i) = (1, u_{i1}, \ldots, u_{iq_1})$ is a vector containing explanatory variables associated with the spatial at location $\boldsymbol{s}_i$th, $\boldsymbol{\gamma}_0 = (\gamma_0, \gamma_1, \ldots, \gamma_{q_1})^t$ is a vector of unknown spatial regression parameters, $\boldsymbol{u}_{2it}^t = \boldsymbol{u}_2^t(\boldsymbol{s}_i, t) = (u_{it1}, \ldots, u_{itq_{2_t}})$ is a vector containing explanatory variables associated with the space–time at $\boldsymbol{s}_i$th location and tth time, and $\boldsymbol{\gamma}_t = (\gamma_{t1}, \ldots, \gamma_{tq_{2_t}})^t$ is a vector of unknown space–time regression parameters. Additionally, ϱ_t is the spatial autoregressive coefficient in the tth time period, ϵ_{it} reflects the spatio-temporal autocorrelated error term at the $\boldsymbol{s}_i$th location in the tth time, φ_t is called the spatial autocorrelation coefficient in the tth time period, $\delta > 0$ is a dispersion parameter, and $\xi_{it} = \xi(\boldsymbol{s}_i, t)$ is an i.i.d. Gaussian random error term at the ith individual in the tth time with zero mean and covariance $\mathrm{E}(\xi_{it}, \xi_{it'}) = \sigma^*_{tt'}$ for $t, t' = 1, \ldots, T$, with $\mathrm{E}(\xi_{it}, \xi_{i't}) = 0$ for $i, i' = 1, \ldots, n$.

In this specification, $v(\phi_{it})$ is the dispersion variance function. The dispersion covariates u's are commonly, but not necessarily, a subset of the regression covariates x's. Two possible choices for the dispersion statistic d_{it} are (i) the generalised Pearson contribution, $d_{it} = r_{P_{it}}$ and (ii) the contribution to the deviance of unit (i, t): $d_{it} = r_{D_{it}}$. Note that for Normal-theory models but not otherwise, the two forms are equivalent. That is, when evaluated at the true μ_{it}, $\mathrm{E}(r^2_{P_{it}}) = \phi$ exactly, whereas $\mathrm{E}(r^2_{D_{it}}) \simeq \phi$ only approximately.

The two models (40) and (41) are interlinked, that is, the mean model requires an estimate of $1/\phi_{it}$ to be used as prior weight, while the dispersion model requires an estimate of μ_{it} in order to form the dispersion response variable d_{it}. The form of the interlinking suggests an obvious algorithm for fitting these models, whereby we alternate between fitting the model for the means for given weights $1/\hat{\phi}_{it}$ and fitting the model for the dispersion using the response variable d_{it}.

On the other hand, an estimate of the variance can be formed for each distinct point in the covariate space–time of the model for $\mathrm{E}(y_{it})$ if the data contain replicate observations for each space–time point. The correct choice of variance function for the mean is also important if distortion of the dispersion model is to be avoided. Thus, in a generalised linear space–time autoregressive model with space–time autoregressive disturbances for modelling both mean and dispersion, it is advisable to have estimates of dispersion based on pure replicates. Information from null contrasts can then be combined with the information from replicate contrasts if they prove compatible.

6.1 *Extended quasi-likelihood as a criterion*

The extended quasi-likelihood, Q^+, developed in Section 4.2, provides a possible criterion to be maximized for the estimation of μ_{it} and ϕ_{it} and for measuring the goodness of fit. Then, we can write

$$-2Q^+ = \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \frac{r_{D_{it}}}{\phi_{it}} + \sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \log(2\pi \phi_{it} v(y_{it})), \tag{42}$$

where $r_{D_{it}}$'s are the deviance components given by Equation (37) in the generalised linear space–time autoregressive model for means, and it can be replaced by the Pearson residual $r_{P_{it}}$.

Now suppose that the two parts of the model are parameterized as $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ and $\boldsymbol{\phi} = \boldsymbol{\phi}(\boldsymbol{\gamma})$. Then, from Equation (42), the estimating equations for $\boldsymbol{\beta}$ are the Wedderburn quasi-likelihood equations given by

$$\sum_{i=1}^{n} \sum_{t=1}^{\mathrm{T}} \frac{y_{it} - \mu_{it}}{\phi_{it} v(\mu_{it})} \frac{\partial \mu_{it}}{\partial \beta_j} = 0, \tag{43}$$

except that $1/\phi_{it}$ must now be included as a weight, the dispersion being non-constant.

The estimating equations for $\boldsymbol{\gamma}$ are given by

$$\sum_{i=1}^{n}\sum_{t=1}^{\mathrm{T}}\frac{r_{D_{it}}-\phi_{it}}{\phi_{it}^{2}}\frac{\partial\phi_{it}}{\partial\gamma_j}=0. \tag{44}$$

These are the Wedderburn quasi-likelihood equations for $v(\mu)=\mu^2$ with the deviance component as response variable. Thus, so far as estimation is concerned, the use of Q^+ as an optimizing criterion is equivalent to assuming that the deviance component has a variance function of the form $v_D(\phi)=\phi^2$, regardless of the variance function for the y_{it}'s. This can only be approximately correct, so the above joint space–time modelling of mean and dispersion procedures are computationally complicated because we need too much time to find a joint numerically solution, which cannot be the joint optimum estimation.

7. Conclusions

We have presented a solution to problems where the variable response is a counting or a binary variable using our GLSTARAR model. We have developed a spatio-temporal autoregressive generalised linear model with spatio-temporal autoregressive disturbances for mean model. Under our proposed model, we can consider a parameter of constant dispersion (ϕ) and correlation both spatial and temporal terms, including a noise (ε_{it}), which allows one to fit the heteroscedasticity of the response variable. In the proposed model, the estimations of fixed and random effects were developed, and their significance levels were determined. We extended the ability to judge the fixed effects specification against the specification of random effects for panel data. This was done by including the autocorrelation of the space–time error or the space–time lagged dependent variable using specification tests.

The parameters of the covariance matrix in the extended model were estimated. The estimation process of the different parameters was performed using an adaptation of the generalised estimating equations for the space–time context. A benefit of the approach is the ability to draw likelihood-based inferences from the data. Furthermore, selection, validation and prediction for the GLSTARAR model were presented using the space–time GEE method. A measure of goodness of fit is proposed, some measures for residual analysis were given, and the space–time prediction of new subjects was performed.

When mean and variance are related as in our GLSTARAR model, we proposed a methodology to joint space–time modelling of mean and dispersion. This joint modelling gives also a solution to the problem when the variance is not constant. However, in our application we did not find this problem, for that reason our recommendation is explored in more detail this proposed in future works. Then, the proposed methodology can be used in variable response framed within the generalised linear models that have both spatial and temporal correlation.

The application presented on the number of AAs by guerrillas of the FARC-EP and ELN in Colombia showed the potential of the proposed methodology. We fitted a spatio-temporal model with a good performance on prediction, where all assumptions were validated using normal probability plot, deviance residuals plot with respect to fitted values, Pearson residuals and box-plots for homogeneity of variances.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Work partially funded and supported by the Spanish Ministry of Science and Education under Grants MTM2013-43917-P and PROMETEOII/2014/062; Carolina Foundation; Applied Statistics in Experimental Research, Industry and Biotechnology (Universidad Nacional de Colombia); and Core Spatial Data Research (Faculty of Engineering, Universidad Distrital Francisco José de Caldas).

References

[1] M. Aitkin, *Modelling variance heterogeneity in Normal regression using GLIM*, Appl. Stat. 36 (1987), pp. 332–339.
[2] P.S. Albert and L.M. McShane, *A generalized estimating equations approach for spatially correlated binary data: applications to the analysis of neuroimaging data*, Biometrics 51 (1995), pp. 627–638.
[3] L. Anselin, *Spatial Econometrics: Methods and Models*, Kluwer Academic, Dordrecht, 1988.
[4] L. Anselin and R. Florax, *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin, 1995.
[5] L. Anselin, R.J.G.M. Florax, and S. Rey, *Advances in Spatial Econometrics, Methodology, Tools and Applications*, Springer, Berlin, 2004.
[6] L. Anselin and H.H. Kelejian, *Testing for spatial error autocorrelation in the presence of endogenous regressors*, Int. Regional Sci. Rev. 20 (1997), pp. 153–182.
[7] F.J. Aranda-Ordaz, *On two families of transformations to additivity for binary response data*, Biometrika 68 (1981), pp. 357–363.
[8] A. Atkinson, *Plots, Tranformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Clarendon Press, Oxford, 1985.
[9] N.H. Augustin, E. Kublin, B. Metzler, E. Meierjohann, and G. Wuhlischvon, *Analyzing the spread of beech canker*, Forest Sci. 51 (2002), pp. 438–448.
[10] G.E. Box and D.R. Cox, *An analysis of transformations*, J. R. Stat. Soc. Ser. B 26 (1964), pp. 211–246.
[11] G. Chamberlain, In *Handbook of Econometrics, Vol. II*, Z. Griliches and M. Intriligator, eds., Amsterdam, 1984, pp. 1247–1318.
[12] A.D. Cliff and J.K. Ord, *Spatial Processes: Models and Applications*, Pion, London, 1981.
[13] R.D. Cook and S. Weisberg, *Diagnostics for heteroscedasticity in regression*, Biometrika 70 (1983), pp. 1–10.
[14] M. Davidian, *Applied Longitudinal Data Analysis*, Chapman and Hall, North Carolina State University, NC, 2005.
[15] T. Dielman, *Pooled cross-sectional and time series data: A survey of current statistical methodology*, Am. Statist. 37 (1983), pp. 111–122.
[16] P. Diggle, P. Heagerty, K.Y. Liang, and S.L. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, New York, 2002.
[17] A.J. Dobson, *An Introduction to Generalized Linear Models*, 2nd ed., Chapman Hall, New York, 2002.
[18] S. Dudley, *Walking Ghosts: Murder and Guerrilla Politics in Colombia*, Routledge, New York, 2004.
[19] A.S. Fotheringham and F. Zhan, *A comparison of three exploratory methods for cluster detection in spatial point patterns*, Geograph. Anal. 28 (1996), pp. 200–218.
[20] R. Haining, *Diagnostics for regression modeling in spatial econometrics*, J. Regional Sci. 34 (1994), pp. 325–341.
[21] C. Hsiao, *Benefits and limitations of panel data*, Econometr. Rev. 4 (1985), pp. 121–174.
[22] C. Hsiao, *Analysis of Panel Data*, 2nd ed., Cambridge University Press, Cambridge, 2003.
[23] N.M. Laird and J.H. Ware, *Random-effects models for longitudinal data*, Biometrics 38 (1982), pp. 963–974.
[24] K.-Y. Liang and S.L. Zeger, *Longitudinal data analysis using generalized linear models*, Biometrika 73 (1986), pp. 13–22.
[25] J. Marschak, *On combining market and budget data in demand studies: A suggestion*, Econometrica 7 (1939), pp. 332–335.
[26] P. McCullagh and J. Nelder, *Generalized Linear Models*, Chapman Hall, London, 1989.
[27] C.E. McCulloch and S.R. Searle, *Generalized, Linear and Mixed Models*, Wiley, New York, 2001.
[28] C.E. McCulloch, S.R. Searle, and J.M. Neuhaus, *Generalized, Linear, and Mixed Models*, 2nd ed., Wiley, New Jersey, 2008.
[29] G. Molenberghs and G. Verbeke, *Models for Discrete Longitudinal Data*, Springer, New York, 2005.
[30] P.A.P. Moran, *Notes on continuous stochastic phenomena*, Biometrika 37 (1950), pp. 17–23.
[31] B.N. Mukherjee, *A Simple Approach To Testing of Hypotheses Regarding A class of Covariance Structures*, *Proceeding of the Indian Statistical Institute Golden Jubilee International Conference On Statistics: Applications and New Directions*, 16 December–19 December, Kolkata, 1981, pp. 442–465.
[32] J. Mur and J. Lauridsen, *Outliers and spatial dependence in cross-sectional regressions*, Environ. Plann. A 39 (2007), pp. 1752–1769.
[33] R.H. Myers, D.C. Montgomery and G.G. Vinning, *Generalized Linear Models with Applications in Engineering and the Sciences*, Wiley, New York, 2002.

[34] R.H. Myers, D.C. Montgomery, G.G. Vinning, and T.J. Robinson, *Generalized Linear Models with Applications in Engineering and the Sciences*, 2nd ed., Wiley, New Jersey, 2010.

[35] R.K. Pace, R. Barry, and C.F. Sirmans, *Spatial statistics and real estate*, J. Real Estate Finance Econ. 17 (1998), pp. 5–13.

[36] J.H.P. Paelinck and L.H. Klaassen, *Spatial Econometrics*, Saxon House, Farnborough, 1979.

[37] W. Pan, *Akaike's information criterion in generalized estimating equations*, Biometrics 57 (2001), pp. 120–125.

[38] D. Pregibon, *Review of generalized linear models*, Ann. Stat. 12 (1984), pp. 1589–1596.

[39] R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2015. Available at http://www.R-project.org/.

[40] S.R. Searle, G. Casella, and C.E. McCulloch, *Variance Components*, Wiley, New York, 1987.

[41] Sipri, *Military Expenditure by Country, In Constant (2011) US, 1988–2012*, 2012 Stockholm International Peace Research Institute.

[42] G.K. Smyth, *Coupled and separable iterations in nonlinear estimation*, Ph.D. thesis, Australian National University, 1985.

[43] M.E. Stokes, C.S. Davis, and G.G. Koch, *Categorical Data Analysis Using SAS*, 3rd ed., SAS Institute Inc., Cary, NC, 2012.