Statistics

A Journal of Theoretical and Applied Statistics

ISSN: 0233-1888 (Print) 1029-4910 (Online) Journal homepage: https://www.tandfonline.com/loi/gsta20

Analysis of Poisson varying-coefficient models with autoregression

Taeyoung Park & Seonghyun Jeong

STATISTICS, 2018
VOL. 52, NO. 1, 34–49
https://doi.org/10.1080/02331888.2017.1353514

Analysis of Poisson varying-coefficient models with autoregression

Taeyoung Park[a] and Seonghyun Jeong[b]

[a]Department of Applied Statistics, Yonsei University, Seoul, Korea; [b]Department of Statistics, North Carolina State University, Raleigh, NC, USA

ABSTRACT
In the regression analysis of time series of event counts, it is of interest to account for serial dependence that is likely to be present among such data as well as a nonlinear interaction between the expected event counts and predictors as a function of some underlying variables. We thus develop a Poisson autoregressive varying-coefficient model, which introduces autocorrelation through a latent process and allows regression coefficients to nonparametrically vary as a function of the underlying variables. The nonparametric functions for varying regression coefficients are estimated with data-driven basis selection, thereby avoiding overfitting and adapting to curvature variation. An efficient posterior sampling scheme is devised to analyse the proposed model. The proposed methodology is illustrated using simulated data and daily homicide data in Cali, Colombia.

ARTICLE HISTORY
Received 12 June 2016
Accepted 7 June 2017

KEYWORDS
Bayesian analysis; data-driven basis selection; functional relationship; partial collapse; Poisson regression

AMS SUBJECT CLASSIFICATION
62F15; 62M10

1. Introduction

A Poisson regression model has been widely used in the analysis of event count data, where it is of interest to account for the multiplicative effect of predictors on expected event counts. Let y_t be the number of events at time t and let $\mathbf{z}_t$ be a vector of predictors at time t, for $t = 1, \dots, T$. In a Poisson regression model, the log mean function is modelled by

$$\log\{E(y_t)\} = \mathbf{z}_t^\top \boldsymbol{\theta},$$

where $\boldsymbol{\theta}$ denotes the multiplicative effect of $\mathbf{z}_t$ on the mean of y_t. While Poisson assumptions can correctly deal with a non-Gaussian mechanism of data generation [1], the Poisson regression model may suffer from an overdispersion problem and fail to account for serial dependence that is likely to be present among time series of event counts [2].

In order to deal with dependence in time series of event counts, the log of expected event counts can be simply lagged in Poisson regression, i.e.

$$\log\{E(y_t)\} = \mathbf{z}_t^\top \boldsymbol{\theta} + \rho y_{t-1}.$$

Unless $\rho = 0$ or $E[y_{t-1} - y_{t-2}] = 0$, however, the resulting lagged Poisson regression model incurs a nonzero exponential growth rate and cannot be used to model stationary time series [3]. Instead, autocorrelation can be introduced to expected counts through a multiplicative latent autoregressive

CONTACT Taeyoung Park tpark@yonsei.ac.kr

process [4]. That is, given a latent process e_t, the conditional distribution of y_t is Poisson with log mean function

$$\log\{E(y_t \mid e_t)\} = \mathbf{z}_t^\top \boldsymbol{\theta} + \log e_t,$$

where $E(e_t) = 1$ and $Cov(e_t, e_{t+p}) = \sigma^2 \rho_e(p)$. Because of using generalized estimating equations, however, inference on the model based on the multiplicative latent autoregressive process has been criticized for inefficiency [5]. Alternatively, a Poisson exponentially weighted moving average model is proposed to represent persistent time series of event counts in state-space form [3]. Our work relates to the latent autoregressive model [4] but differs in that the proposed model is based on a latent autoregressive process is log-additive rather than multiplicative. For example, the Poisson regression model with an log-additive latent AR(1) process has the log mean function,

$$\begin{aligned} \log\{E(y_t \mid e_t)\} &= \mathbf{z}_t^\top \boldsymbol{\theta} + e_t, \\ e_t &= \rho e_{t-1} + \epsilon_t, \end{aligned}$$

where ϵ_t is an error term with $\epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

The aforementioned Poisson-regression-type models are, however, limited to account for a nonlinear interaction between the expected event counts and predictors as a function of some underlying variables. Because of such difficulties, there is a need for alternative and more flexible models that can directly account for the data-generation mechanism of times series of event counts and explore a dynamic feature existing in the data. With Gaussian errors, a partially linear varying-coefficient model is proposed to capture a dynamic relationship between a response variable and predictors that is confounded with some underlying variables [6,7]. The partially linear varying-coefficient model allows the effect of predictors on the response variable to nonlinearly vary with the underlying variables, without completely abandoning the existing linear regression model but with good interpretability. In the case of time series of event counts, a time-varying baseline event rate is estimated by a piecewise-constant function after adjusting for the constant effect of predictors on expected counts [8]. Other generalizations of a varying-coefficient model are also proposed for longitudinal or clustered data [9–11] and for non-Gaussian distributed data [12–14].

In this paper, we propose a Poisson autoregressive varying-coefficient model that accounts for the *nonlinear* multiplicative effects of predictors on expected event counts, while simultaneously dealing with a latent autoregressive process for time series of event counts. To our knowledge, our proposed model is the first one that allows both varying coefficients and autoregression in Poisson regression. The nonlinear structure on Poisson regression coefficients is estimated by a regression spline with data-driven basis selection [15,16]. Letting data determine the optimal set of basis terms can avoid the possibility of overfitting and adapt to inhomogeneous curvature of the nonparametric varying coefficients [10]. By imposing a Poisson-lognormal hierarchical structure on the regression coefficients, possible overdispersion is also accounted for in Poisson regression [17]. Serial dependence implied by time series of event counts is modelled with a log-additive latent autoregressive process of order p. Efficient Bayesian semiparametric inference on the proposed model is also made via the method of partial collapse [18,19].

The remainder of this paper is divided into five sections. Section 2 develops the Poisson autoregressive varying-coefficient model that allows varying regression coefficients for a log-additive latent autoregressive process in Poisson regression, and specifies prior distributions in the proposed model. In Section 3, we devise a data-driven method to estimate a nonlinear function for a Poisson regression coefficient and a latent autoregressive process via an efficient posterior sampling scheme. Section 4 validates the proposed modelling approach and methods through numerical studies. In Section 5, our proposed methodology is applied to time series of daily homicides in the city of Cali, Colombia from January 1999 to August 2008. Discussion follows in Section 6.

2. Semiparametric hierarchical model

2.1. Poisson autoregressive varying-coefficient model

To model time series of event counts, we develop the Poisson autoregressive varying-coefficient model with the log of an expected event rate,

$$\begin{aligned} \psi_t \equiv \log\{E(y_t \mid e_t)\} - \log P_t &= \xi_t + e_t, \quad t = 1, \ldots, T, \\ e_t &= \epsilon_t, \quad t = 1, \ldots, p, \\ \rho(B)e_t &= \epsilon_t, \quad t = p+1, \ldots, T, \end{aligned} \tag{1}$$

where P_t is a known offset at time t, ξ_t is a mean part in the log of an expected event rate at time t, e_t is a latent autoregressive process of order p, $\rho(B)$ is an autoregressive operator given by $\rho(B) = 1 - \rho_1 B - \rho_2 B^2 - \cdots - \rho_p B^p$ with backshift operator B such that $B^k e_t = e_{t-k}$, and ϵ_t is an error term with $\epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$, for $t = 1, \ldots, T$. In this paper, we allow the multiplicative effect of predictors on an expected event rate to vary with underlying variables, so that we have

$$\xi_t = \mathbf{x}_t^\top \boldsymbol{\beta}(\mathbf{u}_t) + \mathbf{z}_t^\top \boldsymbol{\theta}, \tag{2}$$

where $\mathbf{x}_t = (x_{t1}, \ldots, x_{tq})^\top$ and $\mathbf{z}_t = (z_{t1}, \ldots, z_{tr})^\top$ denote $q \times 1$ and $r \times 1$ vectors of predictors, respectively, $\mathbf{u}_t = (u_{t1}, \ldots, u_{tq})^\top$ is a $q \times 1$ vector of underlying variables over which the effect of $\mathbf{x}_t$ on ξ_t varies, $\boldsymbol{\beta}(\mathbf{u}_t) = (\beta_1(u_{t1}), \ldots, \beta_q(u_{tq}))^\top$ denotes a $q \times 1$ vector of varying coefficients as a function of the underlying variables $\mathbf{u}_t$, and $\boldsymbol{\theta}$ denotes a $r \times 1$ vector of fixed regression coefficients.

To flexibly model a functional relationship between the predictors and expected event rate, we approximate varying coefficients as linear combinations of spline basis functions. Because shape and smoothness of a spline curve severely depend on knot-placements, we propose a data-driven method to determine the number and location of knots. For the jth varying coefficient $\beta_j(u_{tj})$, we thus consider the $(m_j + 2) \times 1$ vector of potential radial basis functions,

$$\mathbf{b}_j(u_{tj}) = \left\{1, u_{tj}, \left|\frac{u_{tj} - \tau_{j1}}{c_j}\right|^2 \log\left|\frac{u_{tj} - \tau_{j1}}{c_j}\right|, \ldots, \left|\frac{u_{tj} - \tau_{jm_j}}{c_j}\right|^2 \log\left|\frac{u_{tj} - \tau_{jm_j}}{c_j}\right|\right\}^\top, \tag{3}$$

where $\boldsymbol{\tau}_j = (\tau_{j1}, \ldots, \tau_{jm_j})^\top$ is an $m_j \times 1$ vector of knot candidates that are typically chosen as equally spaced interior points or observed order statistics that lie in the range of the corresponding underlying variable u_{tj} and c_j denotes a predetermined scale factor that is set to a sample standard deviation of u_{tj} in our study. To circumvent an potential overfitting problem, we consider data-driven selection of the spline basis functions by introducing an $(m_j + 1) \times 1$ vector of latent basis selection indicator variables $\boldsymbol{\gamma}_j = (\gamma_{j0}, \ldots, \gamma_{jm_j})^\top$, where $\gamma_{jl} = 1$ if the $(l+2)$th term in (3) is used as a basis function and 0 otherwise, for $l = 0, 1, \ldots, m_j$. Note that the first term in (3) corresponding to a constant basis function is always included in the model, while the other terms may not be included in the model to circumvent the possibility of overfitting. For notational simplicity, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_q^\top)^\top$ is defined as a $(\sum_{j=1}^q m_j) \times 1$ vector of all latent knot indicator variables with total sum $J(\boldsymbol{\gamma}) = \sum_{j=1}^q J_j(\boldsymbol{\gamma}_j)$ and partial sums $J_j(\boldsymbol{\gamma}_j) = \sum_{l=0}^{m_j} \gamma_{jl}$, for $j = 1, \ldots, q$. Letting $\mathbf{b}_{\boldsymbol{\gamma}_j}(u_{tj})$ be a $(J_j(\boldsymbol{\gamma}_j) + 1) \times 1$ subvector of $\mathbf{b}_j(u_{tj})$, which is determined by basis selection $\boldsymbol{\gamma}_j$, and letting $\boldsymbol{\phi}_{\boldsymbol{\gamma}_j}$ be a $(J_j(\boldsymbol{\gamma}_j) + 1) \times 1$ vector of the corresponding coefficients, the jth varying coefficient $\beta_j(u_{tj})$ is approximated as follows:

$$\beta_j(u_{tj}) \approx \mathbf{b}_{\boldsymbol{\gamma}_j}(u_{tj})^\top \boldsymbol{\phi}_{\boldsymbol{\gamma}_j}.$$

Because we have $\rho(B)e_t = \epsilon_t$ for $t = p+1\ldots,T$, i.e. $e_t = (1-\rho(B))e_t + \epsilon_t$, the model in (1) and (2) is rewritten as follows:

$$\psi_t = \mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}} + e_t$$
$$= \begin{cases} \mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}} + \epsilon_t, & t = 1,\ldots,p, \\ (1-\rho(B))\psi_t + \rho(B)\mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}} + \epsilon_t, & t = p+1,\ldots,T, \end{cases}$$

where $\mathbf{v}_{\boldsymbol{\gamma},t} = (\mathbf{w}_{\boldsymbol{\gamma},t}^\top, \mathbf{z}_t^\top)^\top$, $\boldsymbol{\alpha}_{\boldsymbol{\gamma}} = (\boldsymbol{\phi}_{\boldsymbol{\gamma}}^\top, \boldsymbol{\theta}^\top)^\top$, $\mathbf{w}_{\boldsymbol{\gamma},t} = (x_{t1}\mathbf{b}_{\boldsymbol{\gamma}_1}(u_{t1})^\top, \ldots, x_{tq}\mathbf{b}_{\boldsymbol{\gamma}_q}(u_{tq})^\top)^\top$, and $\boldsymbol{\phi}_{\boldsymbol{\gamma}} = (\boldsymbol{\phi}_{\boldsymbol{\gamma}_1}^\top, \ldots, \boldsymbol{\phi}_{\boldsymbol{\gamma}_q}^\top)^\top$. Note that the dimension of $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ is not fixed but varying as a function of the latent basis selection, so that a varying-dimensional analysis is devised in Section 3.

2.2. Prior specification

To impose a hierarchical structure on the Poisson autoregressive varying-coefficient model, we begin with assigning conjugate prior distributions on the basis selection indicator variables $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_q^\top)^\top$. That is, each basis selection indicator variable is assumed to a priori follow an independent Bernoulli distribution with hyperparameter π_j having a beta distribution,

$$\gamma_{jl}|\pi_j \overset{\text{ind}}{\sim} \text{Bernoulli}(\pi_j), \quad j = 1,\ldots,q, \;\; l = 0,\ldots,m_j$$
$$\pi_j \sim \text{Beta}(a_{\pi j}, b_{\pi j}),$$

where $a_{\pi j}$ and $b_{\pi j}$ are fixed in advance. In this way, data-driven basis selection can be implemented through posterior computation. To streamline posterior computation through conjugacy, we devise a multivariate normal prior distribution on the varying-dimensional regression coefficients in the Poisson autoregressive varying-coefficient model, i.e.

$$\boldsymbol{\alpha}_{\boldsymbol{\gamma}} \mid (\boldsymbol{\rho}, \kappa, \sigma^2) \sim \mathrm{N}_{J(\boldsymbol{\gamma})+q+r}\left(\mathbf{0}, \kappa\sigma^2\left[\sum_{t=1}^{p} \mathbf{v}_{\boldsymbol{\gamma},t}\mathbf{v}_{\boldsymbol{\gamma},t}^\top + \sum_{t=p+1}^{\mathrm{T}} \mathbf{v}_{\boldsymbol{\gamma},t}^{*}\mathbf{v}_{\boldsymbol{\gamma},t}^{*\top}\right]^{-1}\right), \tag{4}$$

where N_k denotes a k-dimensional multivariate normal distribution, κ represents a dispersion factor, and $\mathbf{v}_{\boldsymbol{\gamma},t}^* = \rho(B)\mathbf{v}_{\boldsymbol{\gamma},t}$. While the prior distribution in (4) facilitates posterior computation with invariance property [16,20], vague prior information cannot be achieved in the limit of $\kappa \to \infty$ [10,21,22]. Thus we impose a hierarchy on the hyperparameter κ to let data determine its value, i.e.

$$\kappa \sim \mathrm{IG}(a_\kappa/2, b_\kappa/2),$$

where IG denotes an inverse gamma distribution, and a_κ and b_κ are fixed in advance. For the remaining parameters associated with a latent autoregressive process, our prior can be expressed as follows:

$$\boldsymbol{\rho} \sim \mathrm{TN}_{S_\rho}(\mathbf{a}_\rho, \mathbf{B}_\rho^{-1}),$$
$$\sigma^2 \sim \mathrm{IG}(a_\sigma/2, b_\sigma/2),$$

where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_p)^\top$ represents a $p \times 1$ vector of autoregressive coefficients, TN_{S_ρ} denotes a multivariate normal distribution truncated to the region $S_{\boldsymbol{\rho}}$ that implies a stationary error process, and $\mathbf{a}_\rho$, $\mathbf{B}_\rho$, a_σ, and b_σ are chosen for vague prior information.

3. Efficient Bayesian semiparametric inference

With diffuse prior distributions on $(\boldsymbol{\rho}, \boldsymbol{\gamma}, \{\pi_j\}_{j=1}^q, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2)$, a target posterior distribution is computed as $p(\boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \{\pi_j\}_{j=1}^q, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2 \mid \mathbf{Y})$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_T)^\top$ and $\mathbf{Y} = (y_1, \ldots, y_T)^\top$. For efficient Bayesian semiparametric inference on the proposed model, we propose to use the method of (partial) collapse that greatly improves the convergence of an iterative sampling scheme [19,23]. In particular, after $\{\pi_j\}_{j=1}^q$ is completely collapsed, $(\boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \sigma^2)$ are partially collapsed out of the resulting target posterior distribution given by $p(\boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2 \mid \mathbf{Y})$. An iterative sampling scheme resulting from the method of partial collapse may consist of a functionally incompatible set of conditional distributions, so that the sampling order of the conditional distributions must be determined by three basic tools (marginalization, permutation, and trimming) to preserve a desired stationary distribution [19]. The resulting iterative sampling scheme for efficient posterior computation is thus constructed as follows. First, each component of $\boldsymbol{\gamma}$ is iteratively sampled from its reduced conditional distribution with $(\boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \sigma^2)$ collapsed:

Step 1: Draw γ_{jl} from $p(\gamma_{jl} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_{-(jl)}, \kappa, \mathbf{Y})$ that is Bernoulli with probability

$$\begin{aligned} &P(\gamma_{jl} = 1 \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}_{-(jl)}, \kappa, \mathbf{Y}) \\ &\quad = \frac{p(\gamma_{jl} = 1, \boldsymbol{\gamma}_{-(jl)} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \kappa, \mathbf{Y})}{p(\gamma_{jl} = 1, \boldsymbol{\gamma}_{-(jl)} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \kappa, \mathbf{Y}) + p(\gamma_{jl} = 0, \boldsymbol{\gamma}_{-(jl)} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \kappa, \mathbf{Y})}, \end{aligned}$$

for $j = 1, \ldots, q$ and $l = 1, \ldots, m_j$, where $\boldsymbol{\gamma}_{-(jl)} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_{j-1}^\top, \boldsymbol{\gamma}_{j\backslash l}^\top, \boldsymbol{\gamma}_{j+1}^\top, \ldots, \boldsymbol{\gamma}_q^\top)^\top$ with $\boldsymbol{\gamma}_{j\backslash l} = (\gamma_{j0}, \gamma_{j1}, \ldots, \gamma_{j,l-1}, \gamma_{j,l+1}, \ldots, \gamma_{jm_j})^\top$, and the conditional distribution of $\boldsymbol{\gamma}$ given $(\boldsymbol{\psi}, \boldsymbol{\rho}, \kappa, \mathbf{Y})$ is given by

$$p(\boldsymbol{\gamma} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \kappa, \mathbf{Y}) \propto \frac{\prod_{j=1}^q B(a_{\pi j} + J_j(\boldsymbol{\gamma}_j), b_{\pi j} + m_j + 1 - J_j(\boldsymbol{\gamma}_j))}{(1+\kappa)^{(J(\boldsymbol{\gamma})+q+r)/2}(b_\sigma + S)^{(T+a_\sigma)/2}},$$

with

$$\begin{aligned} S &= \sum_{t=1}^p \psi_t^2 + \sum_{t=p+1}^{\mathrm{T}} \psi_t^{*2} - \frac{\kappa}{\kappa+1} \tilde{\mathbf{a}}^\top \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{a}}, \\ \tilde{\mathbf{a}} &= \sum_{t=1}^p \mathbf{v}_{\boldsymbol{\gamma},t} \psi_t + \sum_{t=p+1}^{\mathrm{T}} \mathbf{v}_{\boldsymbol{\gamma},t}^* \psi_t^*, \\ \tilde{\mathbf{A}} &= \sum_{t=1}^p \mathbf{v}_{\boldsymbol{\gamma},t} \mathbf{v}_{\boldsymbol{\gamma},t}^\top + \sum_{t=p+1}^{\mathrm{T}} \mathbf{v}_{\boldsymbol{\gamma},t}^* \mathbf{v}_{\boldsymbol{\gamma},t}^{*\top}, \end{aligned}$$

where $B(\cdot,\cdot)$ is the beta function and $\psi_t^* = \rho(B)\psi_t$. Second, σ^2 is sampled from its reduced conditional distribution with $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ collapsed:

Step 2: Draw σ^2 from $p(\sigma^2 \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \kappa, \mathbf{Y})$ that is inverse gamma, i.e.

$$\sigma^2 \mid (\boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \kappa, \mathbf{Y}) \sim \mathrm{IG}\left(\frac{T + a_\sigma}{2}, \frac{S + b_\sigma}{2}\right).$$

Third, $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ is sampled from its full conditional distribution:

Step 3: Draw $\boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ from $p(\boldsymbol{\alpha}_{\boldsymbol{\gamma}} \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \kappa, \sigma^2, \mathbf{Y})$ that is multivariate normal, i.e.

$$\boldsymbol{\alpha}_{\boldsymbol{\gamma}} | (\boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \kappa, \sigma^2, \mathbf{Y}) \sim \mathrm{N}_{J(\boldsymbol{\gamma})+k+q}\left(\frac{\kappa}{\kappa+1} \tilde{\mathbf{A}}^{-1} \tilde{\mathbf{a}}, \frac{\kappa\sigma^2}{\kappa+1} \tilde{\mathbf{A}}^{-1}\right).$$

Forth, each component of $\boldsymbol{\psi}$ is iteratively sampled from its full conditional distribution:
Step 4. Draw ψ_t from $p(\psi_t \mid \boldsymbol{\psi}_{-t}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2, \mathbf{Y})$ that is nonstandard and proportional to

$$p(\psi_t \mid \boldsymbol{\psi}_{-t}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2, \mathbf{Y})$$
$$\propto \begin{cases} \exp\left(y_t\psi_t - P_t e^{\psi_t} - \dfrac{1}{2\sigma^2}\left\{ (\psi_t - \mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}})^2 + \displaystyle\sum_{k=p+1}^{p+t} (\psi_k^* - \mathbf{v}_{\boldsymbol{\gamma},k}^{*\top}\boldsymbol{\alpha}_{\boldsymbol{\gamma}})^2 \right\} \right), & t = 1, \dots, p, \\ \exp\left(y_t\psi_t - P_t e^{\psi_t} - \dfrac{1}{2\sigma^2} \displaystyle\sum_{k=t}^{(p+t)\wedge T} (\psi_k^* - \mathbf{v}_{\boldsymbol{\gamma},k}^{*\top}\boldsymbol{\alpha}_{\boldsymbol{\gamma}})^2 \right), & t = p+1, \dots, T, \end{cases}$$

where $\boldsymbol{\psi}_{-t} = (\psi_1, \dots, \psi_{t-1}, \psi_{t+1}, \dots, \psi_T)^\top$. To simulate each ψ_t, we use a random-walk Metropolis algorithm with a normal proposal distribution given by $\mathrm{N}(\psi_t, \varsigma_t^2)$, where ς_t^2 is scaled to adjust the efficiency of the resulting Metropolis algorithm [24]. Fifth, $\boldsymbol{\rho}$ is sampled from its full conditional distribution:

Step 5. Draw $\boldsymbol{\rho}$ from $p(\boldsymbol{\rho} \mid \boldsymbol{\psi}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2, \mathbf{Y})$ that is nonstandard and proportional to

$$p(\boldsymbol{\rho} \mid \boldsymbol{\psi}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \kappa, \sigma^2, \mathbf{Y})$$
$$\propto \exp\left(-\frac{1}{2}(\boldsymbol{\rho} - \tilde{\mathbf{r}})^\top \tilde{\mathbf{R}} (\boldsymbol{\rho} - \tilde{\mathbf{r}}) \right) \exp\left(-\frac{1}{2\kappa\sigma^2} \boldsymbol{\alpha}_{\boldsymbol{\gamma}}^\top \left[\sum_{t=p+1}^{\mathrm{T}} \mathbf{v}_{\boldsymbol{\gamma},t}^{*} \mathbf{v}_{\boldsymbol{\gamma},t}^{*\top} \right] \boldsymbol{\alpha}_{\boldsymbol{\gamma}} \right) I(\boldsymbol{\rho} \in S_{\boldsymbol{\rho}}),$$

where $\tilde{\mathbf{r}} = \tilde{\mathbf{R}}^{-1}(\sigma^{-2}\mathbf{E}^\top\mathbf{e} + \mathbf{B}_\rho \mathbf{a}_\rho)$, $\tilde{\mathbf{R}} = \sigma^{-2}\mathbf{E}^\top\mathbf{E} + \mathbf{B}_\rho$, $\mathbf{e} = (e_{p+1}, \dots, e_T)^\top$ is a $(T-p) \times 1$ vector of errors defined by $e_t = \psi_t - \mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ for $t = 1, \dots, T$, and $\mathbf{E}$ is a $(T-p) \times p$ matrix whose tth row is $(e_{p+t-1}, \dots, e_t)$. To implement Step 5, we use an independent Metropolized sampler with $\mathrm{TN}_{S_\rho}(\tilde{\mathbf{r}}, \tilde{\mathbf{R}}^{-1})$ as its proposal distribution. Last, κ is sampled from its full conditional distribution:

Step 6: Draw κ from $p(\kappa \mid \boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \sigma^2, \mathbf{Y})$ that is inverse gamma, i.e.

$$\kappa \mid (\boldsymbol{\psi}, \boldsymbol{\rho}, \boldsymbol{\gamma}, \boldsymbol{\alpha}_{\boldsymbol{\gamma}}, \sigma^2, \mathbf{Y}) \sim \mathrm{IG}\left(\frac{a_\kappa + J(\boldsymbol{\gamma}) + q + r}{2}, \frac{b_\kappa + \sigma^{-2}\boldsymbol{\alpha}_{\boldsymbol{\gamma}}^\top \tilde{\mathbf{A}} \boldsymbol{\alpha}_{\boldsymbol{\gamma}}}{2} \right).$$

Note that the stationary distribution of the Markov chain constructed by the iterative sampling scheme may be changed by permuting the sampling order of the conditional distributions, so that care must be taken to maintain the stationary distribution [18,25]. An iterative sampling scheme of this class is called a Metropolis–Hastings within partially collapsed Gibbs sampler and has proven useful for efficient posterior inference [8,10,25–32].

4. Numerical studies

4.1. Sensitivity analysis

In this section, we analyse simulated data to assess the performance of the proposed method. For simulated data, a mean function in the log of an expected event rate at time t, i.e. ξ_t in (2) is assumed to have two time-varying coefficients and three fixed coefficients:

$$\beta_1(t) = \sin\{8(t/T - 0.5)\} + 2\exp\{-16^2(t/T - 0.5)^2\},$$
$$\beta_2(t) = 2\exp\{-200(t/T - 0.2)^2\} + 0.5\exp\{-50(t/T - 0.6)^2\},$$

and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)^\top = (1, -2, 3)^\top$. While the predictor x_{t1} corresponding to $\beta_1(t)$ is set to 1 for $t = 1, \dots, T$, the rest of predictors x_{t2} and $\mathbf{z}_t = (z_{t1}, z_{t2}, z_{t3})^\top$ are generated from an independent and identical standard normal distribution for $t = 1, \dots, T$. In this way, $\beta_1(t)$ represents the log of an expected baseline event rate as a function of time and $\beta_2(t)$ represents a time-varying multiplicative effect of the predictor x_{t2} on an expected event rate. As for a latent autoregressive process, the true values of autoregression coefficients are set to $\boldsymbol{\rho} = (\rho_1, \rho_2)^\top = (0.5, -0.2)^\top$ with order $p = 2$, and the true value of σ^2 is set to 0.5. The values of the known offset P_t are drawn from Poisson(10) + 1. Under the above simulation setting, two independent data sets of size $T = 250$ and $T = 1000$ are generated.

Along with the correctly specified model, we fit each data set with four misspecified models: models with a correctly specified autoregressive order ($p = 2$) but with (1) an underspecified mean function ($\beta_1(t)x_{t1} + \beta_2 x_{t2} + \sum_{i=1}^{3} \theta_i z_{ti}$) and (2) an overspecified mean function ($\sum_{i=1}^{2} \beta_i(t)x_{ti} + \theta_1(t)z_{t1} + \sum_{i=2}^{3} \theta_i z_{ti}$), and models with a correctly specified mean function ($\sum_{i=1}^{2} \beta_i(t)x_{ti} + \sum_{i=1}^{3} \theta_i z_{ti}$) but with (3) an underspecified autoregressive order ($p = 1$) and (4) an overspecified autoregressive order ($p = 3$). We call these models the underspecified mean model, the overspecified mean model, the underspecified autoregressive model, and the overspecified autoregressive model, respectively.

For data-driven basis selection when estimating each varying coefficient, candidates of knot locations are chosen to be 30 equally spaced interior points which lie on the interval $(0, T)$. Based on the simulation settings, we generate time series of event counts, $\mathbf{Y} = (y_1, \dots, y_T)^\top$, from the corresponding inhomogeneous Poisson process. The simulated data sets are analysed by running the proposed iterative sampling scheme of 10,000 iterations with two over-dispersed starting values. The convergence of the iterative sampling scheme is then assessed by computing the $\hat{R}^{1/2}$ statistic [33] and checking if $\hat{R}^{1/2} < 1.2$ for all model parameters of fixed dimension, and our posterior inference is based on the second halves from the two chain runs.

Figure 1 shows the results of posterior inference about the two time-varying coefficients by using the correctly specified model. The upper plot in each panel of Figure 1 presents the estimated posterior probabilities of inclusion for the linear basis function and knot candidates. For both $T = 250$ and $T = 1000$, the true time-varying coefficients are correctly estimated by the point-wise posterior medians of the data-driven regression splines and well covered by the corresponding point-wise 95% posterior intervals. In particular, our data-driven basis selection shows that more knots are chosen to locally adapt to bumps in the true varying-coefficient functions, capturing spatial inhomogeneity.

Figure 2 shows posterior inference on varying-coefficient functions used in four misspecified models: (a) a model with underspecified mean functions, (b) a model with overspecified mean functions, (c) a model with an underspecified autoregressive order, and (d) a model with an overspecified autoregressive order. Panels (a) and (b) in Figure 2 illustrate that the correctly specified varying-coefficient functions, i.e. $\beta_1(t)$ and $\beta_2(t)$, are well estimated and the fixed coefficient θ_1 overspecified as a varying coefficient is also correctly estimated as a constant, eliminating the possibility of overfitting. As illustrated by panels (c) and (d) in Figure 2, the estimation of varying-coefficient functions is seldom affected by either slightly underspecifying or slightly overspecifying the autoregressive order p; refer to Section 4.2 for a possible problem caused by the excessive overspecification of the autoregressive order.

Figure 3 shows the marginal posterior distributions of model parameters except the varying-coefficient functions in the correctly specified model and four misspecified models. For the correctly specified model, the true values of the model parameters used to simulate our test data are well covered by the corresponding marginal posterior distributions, thereby illustrating the validity of the proposed method. In the underspecified mean model, the variance parameter σ^2 and the autoregressive coefficients, ρ_1 and ρ_2, are not properly estimated, while estimation of the varying and fixed coefficients seems reasonable. This happens because both the error term and the latent autoregressive process attempt to account for the unexplained variation and time-varying effect which are supposed to be accounted for by $\beta_2(t)$; note that posterior inference on the parameters is directly involved with $e_t = \psi_t - \mathbf{v}_{\boldsymbol{\gamma},t}^\top \boldsymbol{\alpha}_{\boldsymbol{\gamma}}$ which is sensitive to the specification of a mean function. On the other hand,

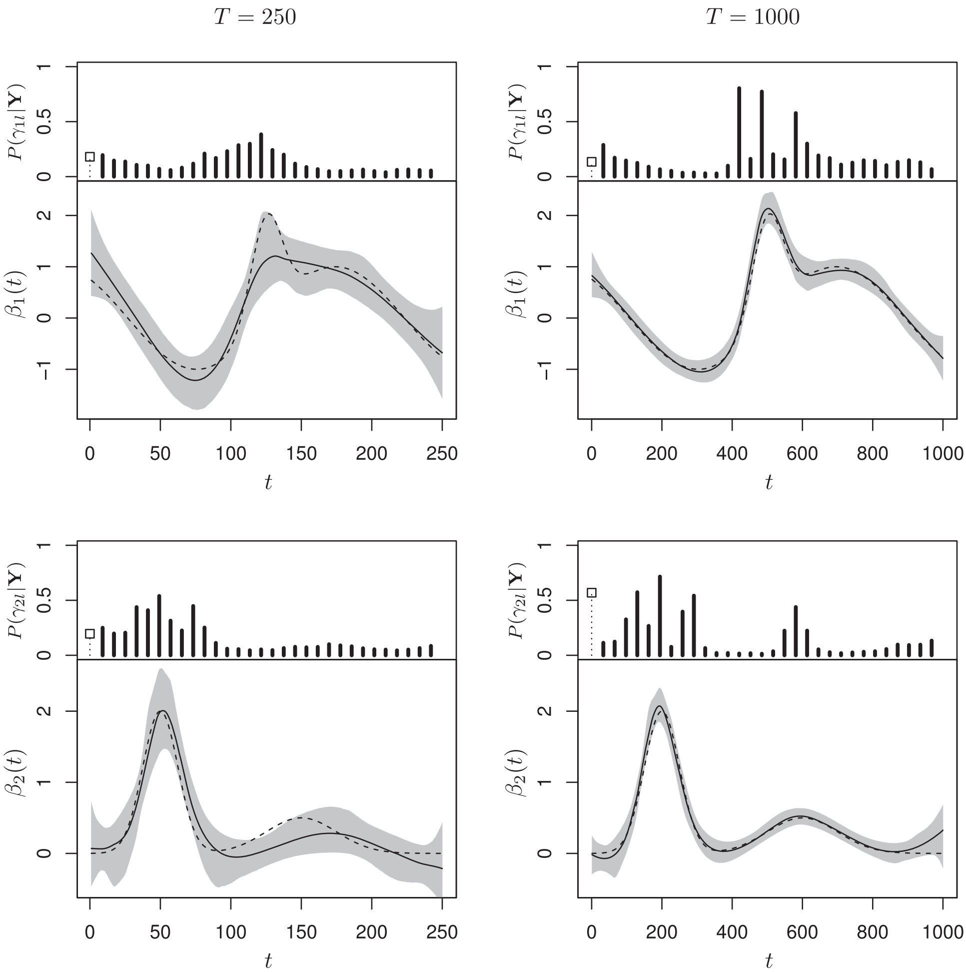

Figure 1. Posterior inclusion probabilities and estimated varying-coefficient functions with the correctly specified model. The left and right panels correspond to $T = 250$ and $T = 1000$, respectively. The upper plot in each panel shows the estimated posterior probabilities of inclusion for the linear basis function (open square with a dotted line) and knot candidates (solid lines). The bottom plot in each panel shows the true varying-coefficient functions (dashed lines), point-wise posterior medians of the estimated functions (solid lines), and point-wise 95% posterior intervals (grey regions).

the overspecified mean model provides the similar results to the correctly specified model. In effect, the data-driven basis selection allows our proposed model to be constructed as a highly overspecified mean model with only the varying coefficients, e.g. $\sum_{i=1}^{2} \beta_i(t)x_{ti} + \sum_{i=1}^{3} \theta_i(t)z_{ti}$ because of robustness against overfitting; refer to Figure 2(b). As shown in Figure 3, the slightly misspecified autoregressive models provide reliable estimates for the fixed coefficients, although estimates for the variance parameter and autoregressive coefficients are slightly off. Discussion on the overspecification of the autoregressive order follows in Section 4.2.

4.2. More on overspecification of autoregression

To illustrate the issue of overspecification for the autoregressive order, the test data generated in Section 4.1 are fitted with two misspecified models that have a correctly specified mean function but

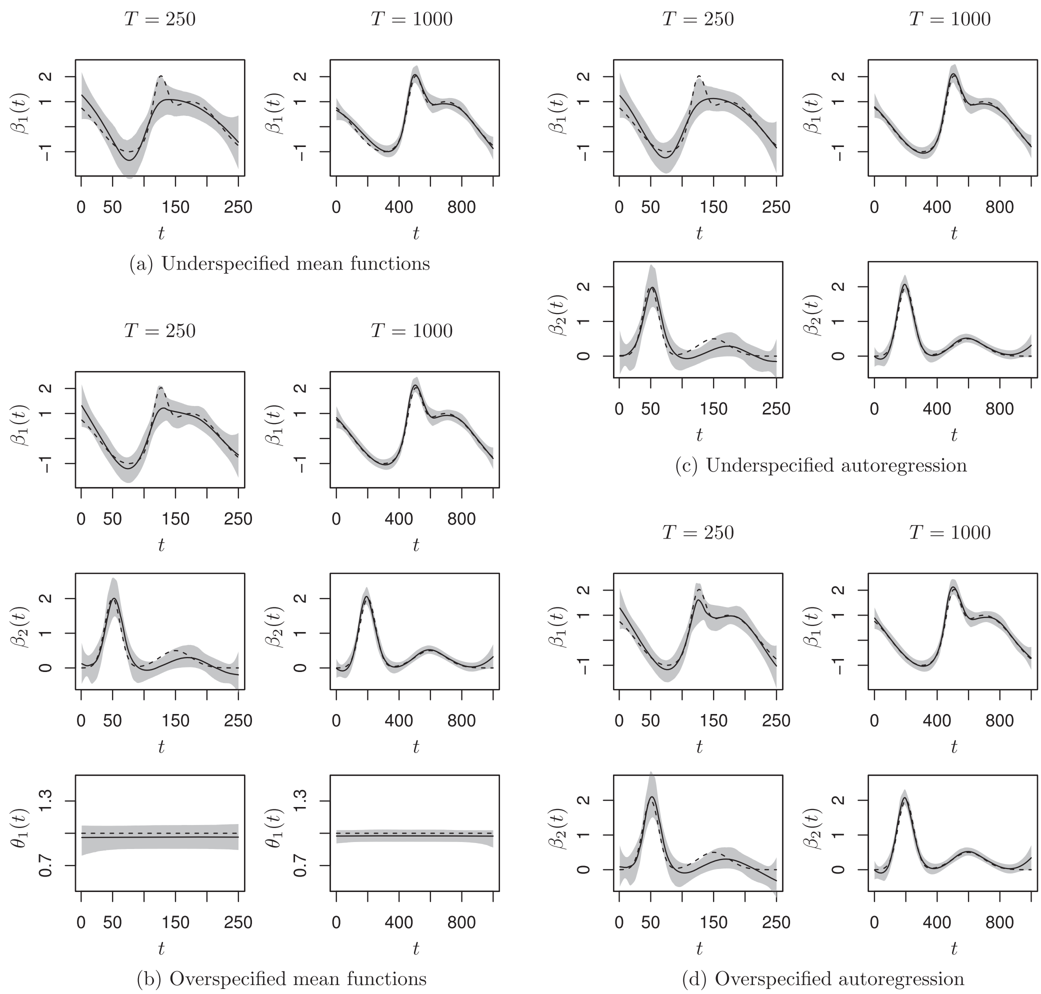

Figure 2. Posterior inference on varying-coefficient functions with four misspecified models. In each sub-figure, the left and right panels correspond to $T = 250$ and $T = 1000$, respectively. The true varying-coefficient functions, point-wise posterior medians of the estimated functions, and point-wise 95% posterior intervals are represented by dashed lines, solid lines, and grey regions, respectively.

a highly overspecified autoregressive order of $p = 4$ or $p = 8$. The convergence of the proposed sampling scheme is assessed by showing that the $\hat{R}^{1/2}$ statistic is less than 1.2 for all model parameters of fixed dimension.

Figure 4 presents posterior inference on the two varying-coefficient functions. While the moderately overspecified autoregressive model with $p = 4$ still correctly estimates the true varying coefficients, the excessively overspecified autoregressive model with $p = 8$ fails to fit a time-varying baseline effect, $\beta_1(t)$. This is because the time-varying intercept can be fully accounted for by the latent autoregressive process of order $p = 8$. Based on our data-driven method, $\beta_1(t)$ is thus reduced to a constant function. This illustrates that the excessive overspecification of the autoregressive order should be avoided to properly account for a potentially time-varying intercept.

Figure 5 shows the marginal posterior distributions of the fixed coefficients and autoregressive coefficients in the two misspecified models. The moderately overspecified autoregressive model with $p = 4$ correctly estimates all the model parameters including the ones related to variation and serial correlation. By contrast, the excessively overspecified autoregressive model with $p = 8$ fails to provide

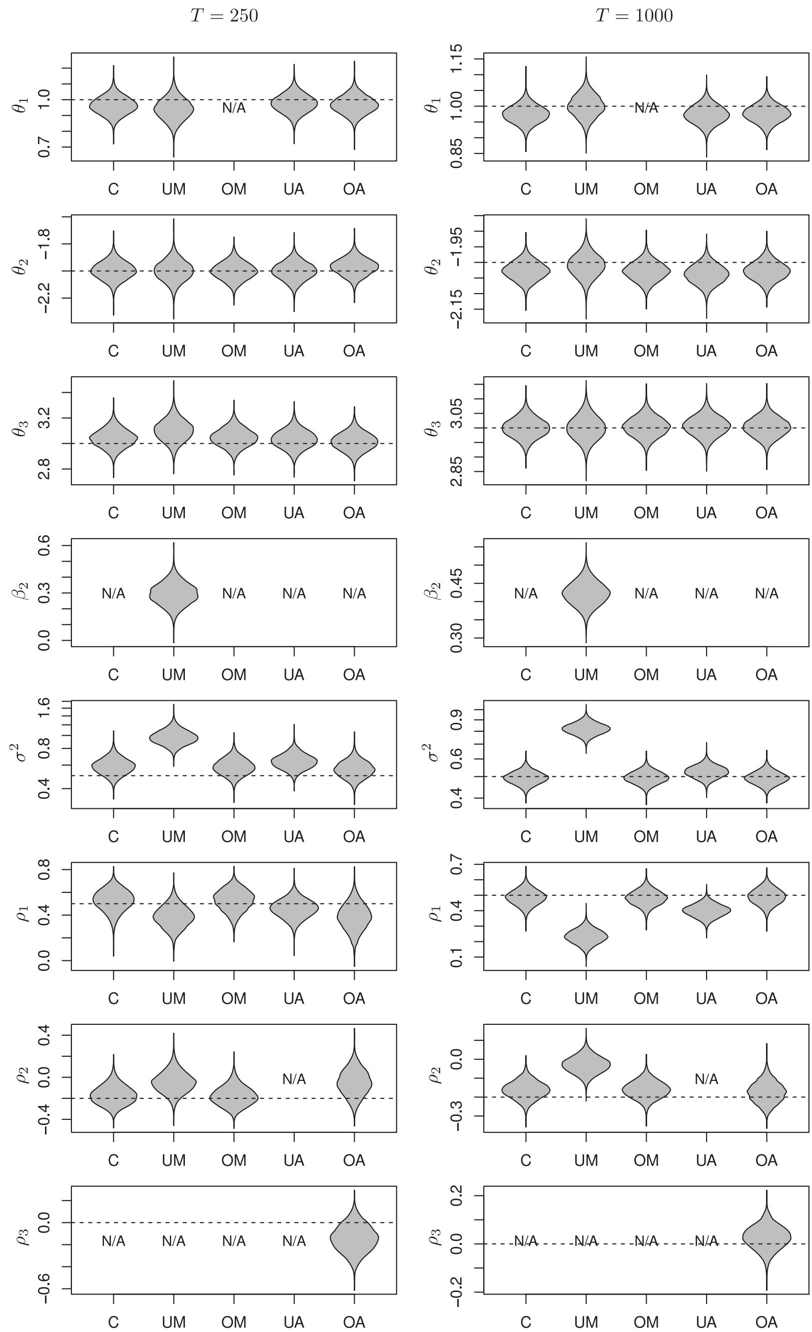

Figure 3. Beanplots for marginal posterior distributions of model parameters except the varying-coefficient functions. The left and right panels correspond to $T = 250$ and $T = 1000$, respectively. The labels C, UM, OM, UA, and OA stand for the correctly specified model, the underspecified mean model, the overspecified mean model, the underspecified autoregressive model, and the overspecified autoregressive model, respectively. The horizontal dashed lines represent the true values of the parameters. The coefficient β_2 should in fact be included as a varying coefficient, so its true value is not presented in the figure.

reasonable posterior estimates about σ^2 and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_8)^\top$, although it correctly estimates the fixed regression coefficients. As in the underspecified mean model in Section 4.1, this is because the estimation of a mean function and a latent autoregressive process can be interwoven. As illustrated

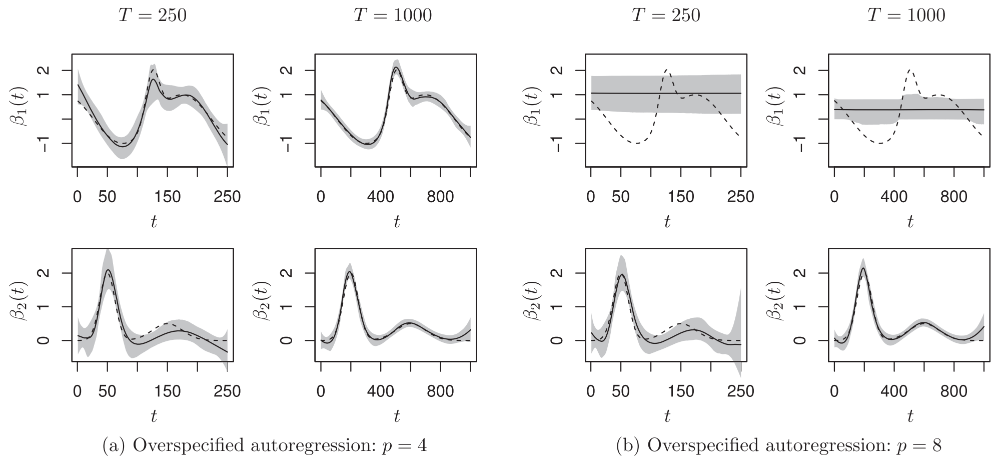

Figure 4. Posterior inference on varying-coefficient functions with highly overspecified autoregressive models ($p = 4$ and $p = 8$). In each sub-figure, the left and right panels correspond to $T = 250$ and $T = 1000$, respectively. The true varying-coefficient functions, point-wise posterior medians of the estimated functions, and point-wise 95% posterior intervals are represented by dashed lines, solid lines, and grey regions, respectively.

in Figure 5, however, the estimation of a mean function is not affected by that of a latent autoregressive process unless the corresponding autoregressive order is excessively overspecified. We thus recommend using low autoregressive orders such as $p = 1$ to 3 if a time-varying baseline effect is of main interest. If it is believed that a baseline effect is constant over time, an intercept can be modelled as a constant and a higher order p of a latent autoregressive process can be used.

5. Homicide Data in Cali, Colombia

5.1. Background and motivation

The approach is applied to time series of daily homicides that occurred in the city of Cali, Colombia from January 1999 to August 2008, which were collected by the Cali Crime Observatory's fatal injury surveillance system [8,34]. Because the consumption of alcohol is associated with an increase in homicide, the city mayor of Cali implemented policies restricting the sales of alcoholic beverages after given hours in an attempt to reduce homicide rates in Cali. The benefits of policies on alcohol sales had not been objectively assessed, so that different policies were implemented on an irregular basis but continuously enforced for as short as 3 days to as long as 346 days.

Part of the time series data was analysed using a conditional autoregressive negative binomial regression [34]. To account for a potentially time-varying baseline rate, the entire time series data were analysed using a Poisson change-point regression that flexibly models a log baseline rate as a piecewise-constant function with unknown number and location of change points [8]. In the Poisson change-point regression model, a variable selection method was developed to select a subset of significant predictors [26]. In analysing the time series data, we use the approach developed in Section 3. Our goal is to correctly evaluate the effect of each alcohol policy by flexibly and parsimoniously representing a potentially time-varying effect of predictors on an expected homicide rate. Because of data-driven basis selection in a varying-coefficient function, we can eliminate the possibility of overfitting and achieve a flexible, yet parsimonious model for varying coefficients.

5.2. Analysis and results

We focus on evaluating the effect of distinct alcohol policies on an expected homicide rate while accounting for a time-varying baseline homicide rate as well as the time-varying effect of weekend

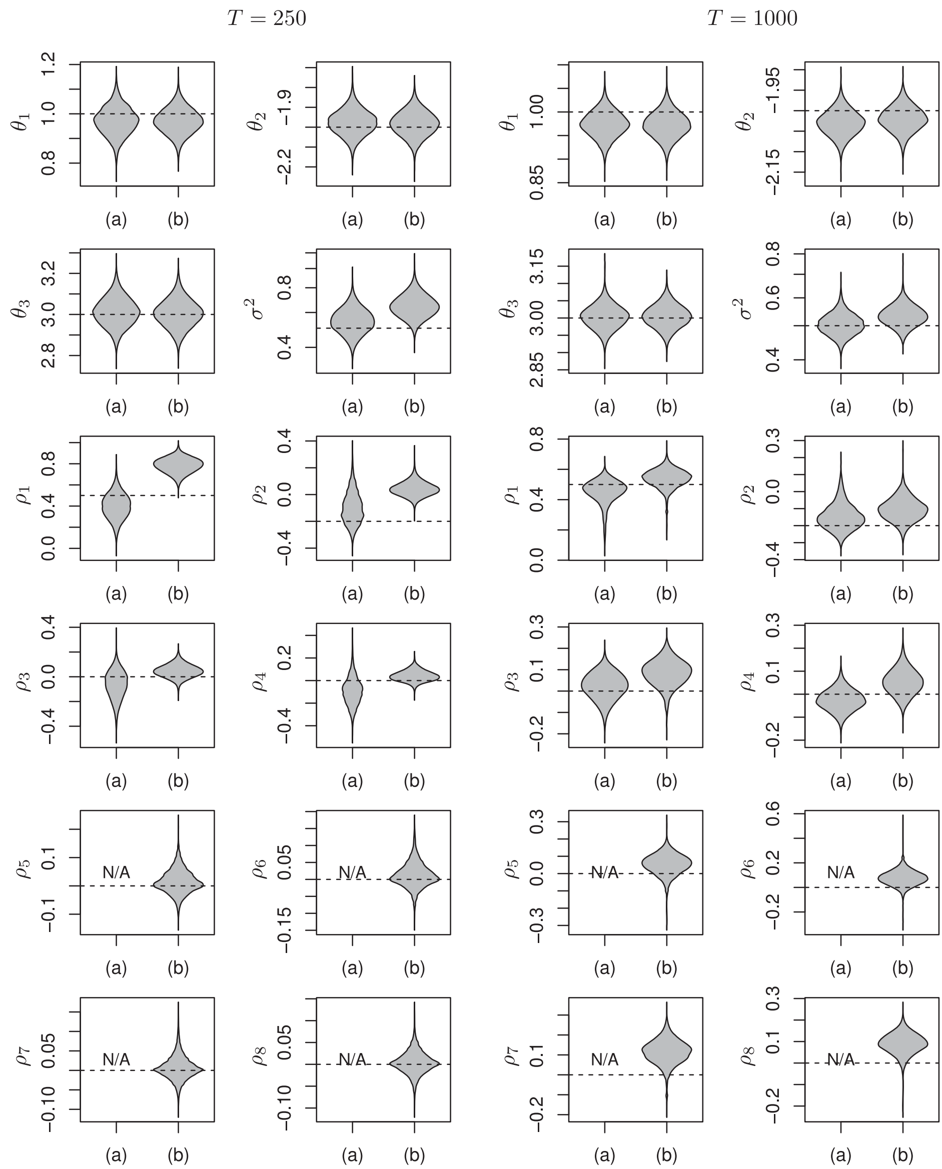

Figure 5. Beanplots for Marginal posterior distributions of model parameters except the varying coefficients for each data set. The labels (a) and (b) stand for the model with $p = 4$ and the model with $p = 8$, respectively. The horizontal dashed lines represent the true values of the parameters.

(Friday, Saturday, and Sunday) on the expected homicide rate. The Poisson autoregressive varying-coefficient model in (1) is then used to analyse time series of daily homicides in Cali, Colombia and constructed as follows:

$$\psi_t = \beta_1(t) + \beta_2(t)W_t + \theta_1 H_t + \theta_2 F_t + \theta_3 B_{1t} + \theta_4 B_{2t} + \theta_5 B_{3t}$$

$$+ \theta_6 F_t W_t + \theta_7 B_{1t} W_t + \theta_8 B_{2t} W_t + \theta_9 B_{3t} W_t + e_t,$$

$$e_t = \rho e_{t-1} + \epsilon_t,$$

Table 1. Posterior summary statistics of model parameters except the ones related to varying coefficients.

Parameter	Posterior mean	Posterior median	2.5 percentile	97.5 percentile
θ_1	1.0025	1.0009	0.8366	1.1714
θ_2	0.1466	0.1467	0.0357	0.2566
θ_3	0.1291	0.1291	0.0412	0.2158
θ_4	0.0175	0.0188	−0.1258	0.1522
θ_5	−0.1484	−0.1475	−0.4453	0.1536
θ_6	−0.1887	−0.1892	−0.3293	−0.0472
θ_7	−0.0939	−0.0940	−0.1706	−0.0157
θ_8	−0.0055	−0.0063	−0.1797	0.1746
θ_9	−0.1775	−0.1750	−0.7066	0.3461
$\theta_2 + \theta_6$	−0.0421	−0.0414	−0.1489	0.0622
$\theta_3 + \theta_7$	0.0352	0.0360	−0.0561	0.1222
$\theta_4 + \theta_8$	0.0121	0.0119	−0.1364	0.1552
$\theta_5 + \theta_9$	−0.3259	−0.3238	−0.7697	0.1065
ρ	0.1204	0.1208	0.0159	0.2229
σ^2	0.0896	0.0895	0.0768	0.1031

where ψ_t is the log of an expected homicide rate on day t, $\beta_1(t)$ represents the log of a time-varying baseline homicide rate, $\beta_2(t)$ represents a time-varying coefficient to W_t that is an indicator of whether day t is Friday, Saturday, or Sunday, H_t is an indicator of whether day t is the main holiday (Christmas, New Year's Eve, and New Year's Day) during which a relatively high number of homicides had been recorded, F_t is an indicator of whether day t is during La Feria de Cali that is the premier cultural event in Cali, the indicator variables B_{1t}, B_{2t}, and B_{3t} are equal to 1 if alcohol sales and consumption are prohibited from 3 am to 10 am, from 12 pm /1am to 10 am, and all day long, respectively, and 0 otherwise, an indicator of whether alcohol sales and consumption are prohibited from 2 to 10 am is used as a reference level, and a latent autoregressive process is used with order $p = 1$ and $\epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$.

The estimated functions of varying coefficients, $\beta_1(t)$ and $\beta_2(t)$, are shown in Figure 6. After adjusting for other measurable predictors, the estimated log baseline homicide rate function is clearly time-dependent, where a high log baseline homicide rate remains more or less constant across 1999–2003, rises slightly up to a certain level in 2004, and then gradually decreases to be constant across 2005–2007 with recent further decrease in 2008. The results are consistent with the Poisson change-point regression model that approximates a smoothly time-varying log baseline homicide rate with an unknown step function; sociopolitical events corresponding to estimated change points can be examined for implications [8]. As shown in the bottom panel of Figure 6, the varying-coefficient function $\beta_2(t)$ is estimated as constant over the entire time period. This illustrates that when a varying coefficient model is misspecified, the possibility of overfitting can be avoided with data-driven basis selection.

Table 1 shows the summary statistics for fixed coefficients and parameters associated with a latent autoregressive process. Because of economical and cultural dependencies on alcohol sales and consumption, there had been considerable debate over a policy that can balance the dependencies with public health. At the centre of the debate is whether alcohol sales and consumption are prohibited after 2 am (a more restrictive policy) or 3 am (a less restrictive policy). The estimated risk ratio of homicide for the more restrictive policy compared to the less restrictive policy is $\exp(\hat{\theta}_3) = 1.14$ with 95% posterior interval of (1.04, 1.24) during the weekdays and $\exp(\hat{\theta}_3 + \hat{\theta}_7) = 1.04$ with 95% posterior interval of (0.95, 1.13) during the weekends. At the 0.05 level, the more restrictive policy has a significantly lower expected homicide rate than the less restrictive policy during the weekdays, but there is no significant difference in an expected homicide rate between the two policies during the weekends. Therefore, along with the previous suggestion [8], a policy that is more restrictive during the weekdays and less restrictive during the weekends can balance the economical and cultural impact of alcohol sales and consumption with public health.

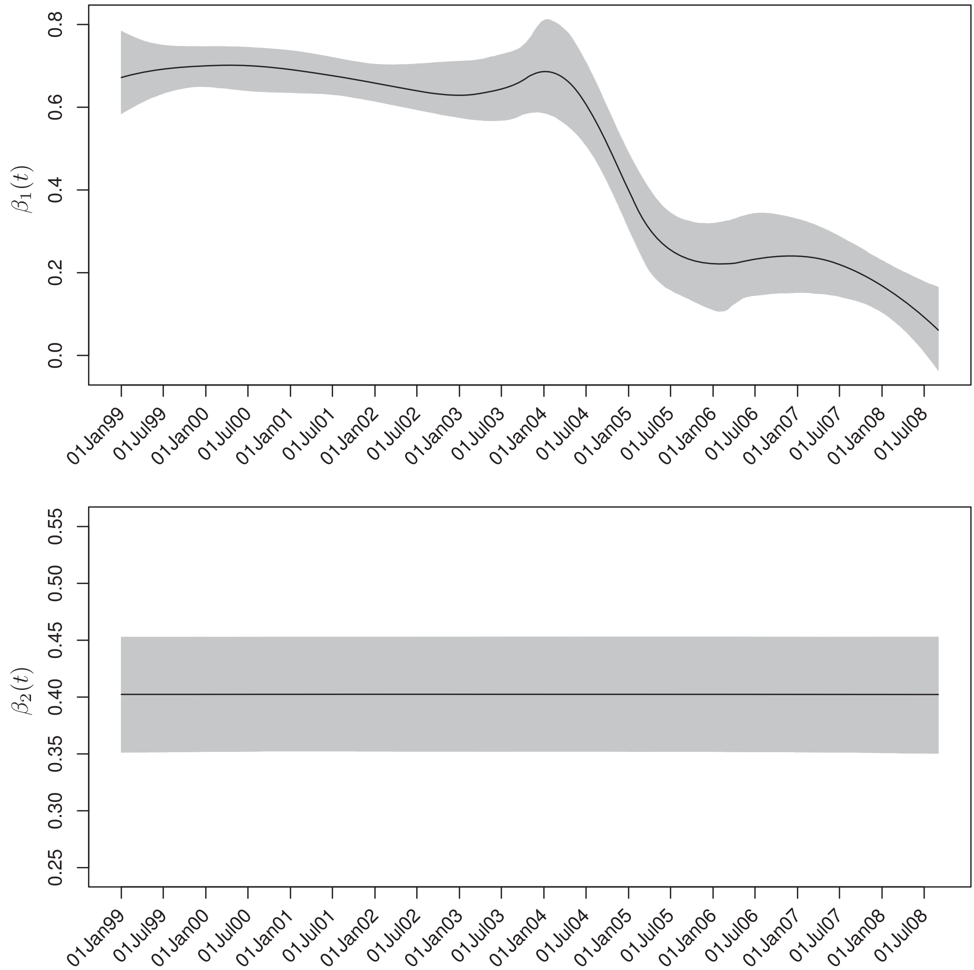

Figure 6. Estimated varying coefficient functions. Solid lines represent point-wise posterior medians of the estimated functions and grey regions represent the corresponding point-wise 95% posterior intervals.

6. Discussion

This paper proposes a Poisson autoregressive varying-coefficient model, which allows the effect of predictors in Poisson regression to nonparametrically vary with some underlying variables while time series of event counts are assumed to be generated by an inhomogeneous Poisson autoregressive process. Each function of varying coefficients is characterized as a linear combination of basis functions which are selected by an efficient data-driven method. When a varying-coefficient function is misspecified, a potential overfitting problem can thus be avoided by such data-driven basis selection, and a parsimonious but flexible model can be specified for the varying-coefficient function. For efficient posterior inference on varying coefficients and a hidden autoregressive process, an efficient posterior sampling scheme is devised by the method of partial collapse.

In this paper, we focus primarily on the analysis of time series of event counts. The binary or categorical response variable is also common in practice. Thus, future research is planned to extend the methodology developed in the paper to the analysis of binary or categorical time series. Allowing varying coefficients as well as autoregression in the generalized linear model may correct biases and improve the fit and interpretability of the model. The data-driven basis selection developed in this paper can be used to select significant predictors in varying coefficients and test for the shape of varying-coefficient functions, which is related to the variable and model selection problems in the varying-coefficient models. We thus plan on providing results on these extensions in future work.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education [grant number NRF-2014R1A1A2056049].

References

[1] Freeland RK, McCabe BPM. Analysis of low count time series data by Poisson autoregression. J Time Ser Anal. 2004;25(5):701–722.
[2] Jazi MA, Jones G, Lai CD. First-order integer valued AR processes with zero inflated Poisson innovations. J Time Ser Anal. 2012;33(6):954–963.
[3] Brandt PT, Williams JT, Fordham BO, et al. Dynamic modeling for persistent event-count time series. Am J Pol Sci. 2000;44(4):823–843.
[4] Zeger SL. A regression model for time series of counts. Biometrika. 1988;75(4):621–629.
[5] Davis RA, Dunsmuir WTM, Wang Y, Modelling time series of count data. In: Ghosh S, editor. Asymptotics, nonparametrics and time series. New York: Marcel Dekker; 1998. p. 63–114.
[6] Ahmad I, Leelahanon S, Li Q. Efficient estimation of a semiparametric partially linear varying coefficient model. Ann Stat. 2005;33(1):258–283.
[7] Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli. 2005;11(6):1031–1057.
[8] Park T, Krafty RT, Sanchez AI. Bayesian semi-parametric analysis of Poisson change-point regression models: application to policy-making in Cali, Colombia. J Appl Stat. 2012;39(10):2285–2298.
[9] Sentürk D, Müller H-G. Generalized varying coefficient models for longitudinal data. Biometrika. 2008;95(3): 653–666.
[10] Jeong S, Park T. Bayesian semiparametric inference on functional relationships in linear mixed models. Bayesian Anal. 2016;11(4):1137–1163.
[11] Jeong S, Park M, Park T. Analysis of binary longitudinal data with time-varying effects. Comput Stat Data Anal. 2017;112:145–153.
[12] Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. J Am Stat Assoc. 2000;95(451):888–902.
[13] Kuruwita CN, Kulasekera KB, Gallagher CM. Generalized varying coefficient models with unknown link function. Biometrika. 2011;98(3):701–710.
[14] Verhasselt A. Generalized varying coefficient models: a smoothing variable selection technique. Stat Sin. 2014;24(1):147–171.
[15] Smith M, Kohn R. Nonparametric regression using Bayesian variable selection. J Econom. 1996;75(2):317–343.
[16] Kohn R, Smith M, Chan D. Nonparametric regression using linear combinations of basis functions. Stat Comput. 2001;11(4):313–322.
[17] Winkelmann R. Econometric analysis of count data. 5th ed. New York: Springer Science & Business Media; 2008.
[18] Park T, van Dyk DA. Partially collapsed Gibbs samplers: illustrations and applications. J Comput Graph Stat. 2009;18(2):283–305.
[19] van Dyk DA, Park T. Partially collapsed Gibbs samplers: theory and methods. J Am Stat Assoc. 2008;193(482): 790–796.
[20] Zellner A. On assessing prior distributions and bayesian regression analysis with g-prior distributions. Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti.. 1986;6:233–243.
[21] Clyde M, George EI. Model uncertainty. Stat Sci. 2004;81–94.
[22] Liang F, Paulo R, Molina G, et al. Mixtures of g priors for Bayesian variable selection. J Am Stat Assoc. 2008;103(481).
[23] Liu JS. The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. J Am Stat Assoc. 1994;89(427):958–966.
[24] Gelman A, Roberts GO, Gilks WR, Efficient Metropolis jumping rule. In: Bernardo J, Berger J, Dawid A, A S, editors. Bayesian statistics, Vol. 5. New York: Oxford University Press; 1996. p. 599–607.

[25] van Dyk DA, Jiao X. Metropolis–Hastings within partially collapsed Gibbs samplers. J Comput Graph Stat. 2015;24(2):301–327.
[26] Min S, Park T. Bayesian variable selection in Poisson change-point regression analysis. Commun Stat – Simul Comput. 2017;46(3):2267–2282.
[27] Park T. Bayesian analysis of individual choice behavior with aggregate data. J Comput Graph Stat. 2011;20(1):158–173.
[28] Park T, Jeong JH, Lee JW. Bayesian nonparametric inference on quantile residual life function: application to breast cancer data. Stat Med. 2012;31(18):1972–1985.
[29] Park T, Jeong S. Efficient Bayesian analysis of multivariate aggregate choices. J Stat Comput Simul. 2015;85(16):3352–3366.
[30] Park T, Min S. Partially collapsed Gibbs sampling for linear mixed-effects models. Commun Stat – Simul Comput. 2016;45(1):165–180.
[31] Park T, van Dyk DA, Siemiginowska A. Searching for narrow emission lines in X-ray spectra: computation and methods. Astrophys J. 2008;688(2):807–825.
[32] van Dyk DA, Park T, Partially collapsed Gibbs sampling and path-adaptive Metropolis–Hastings in high-energy astrophysics. In: Brooks S, Gelman A, Jones GL, Meng XL, editors. Handbook of Markov chain Monte Carlo. New York: Chapman & Hall/CRC Press; 2011. p. 383–400.
[33] Gelman A, Rubin DB. Inference from iterative simulations using multiple sequences (with discussion). Statist Sci. 1992;7(4):457–472.
[34] Sanchez AI, Villaveces A, Krafty RT, et al. Policies for alcohol restriction and their association with interpersonal violence: a time-series analysis of homicides in Cali, Colombia. Int J Epidemiol. 2011;40(4):1037–1046.