

## Article

# Computationally Efficient Poisson Time-Varying Autoregressive Models through Bayesian Lattice Filters

Yuelel Sui <sup>1,\*</sup>, Scott H. Holan <sup>2,3</sup> and Wen-Hsi Yang <sup>4,5</sup><sup>1</sup> Amazon.com, Inc., New York, NY 10001, USA<sup>2</sup> Department of Statistics, University of Missouri, Columbia, MO 65211-6100, USA; holans@missouri.edu<sup>3</sup> Office of the Associate Director for Research and Methodology, Research and Methodology Directorate, U.S. Census Bureau, Washington, DC 20233-9100, USA<sup>4</sup> School of Agriculture and Food Sciences, The University of Queensland, St Lucia, QLD 4072, Australia; w.yang@uq.edu.au<sup>5</sup> School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

\* Correspondence: mkj1911@gmail.com; Tel.: +1-816-885-2368

**Abstract:** Estimation of time-varying autoregressive models for count-valued time series can be computationally challenging. In this direction, we propose a time-varying Poisson autoregressive (TV-Pois-AR) model that accounts for the changing intensity of the Poisson process. Our approach can capture the latent dynamics of the time series and therefore make superior forecasts. To speed up the estimation of the TV-AR process, our approach uses the Bayesian Lattice Filter. In addition, the No-U-Turn Sampler (NUTS) is used, instead of a random walk Metropolis–Hastings algorithm, to sample intensity-related parameters without a closed-form full conditional distribution. The effectiveness of our approach is evaluated through model-based and empirical simulation studies. Finally, we demonstrate the utility of the proposed model through an example of COVID-19 spread in New York State and an example of US COVID-19 hospitalization data.

**Keywords:** Bayesian hierarchical model; nonstationary time series; partial autocorrelation; time-varying spectral density; vector autoregressive model



**Citation:** Sui, Y.; Holan, S.H.; Yang, W.-H. Computationally Efficient Poisson Time-Varying Autoregressive Models through Bayesian Lattice Filters. *Stats* **2023**, *6*, 1037–1052. <https://doi.org/10.3390/stats6040065>

Academic Editor: Wei Zhu

Received: 16 August 2023

Revised: 26 September 2023

Accepted: 4 October 2023

Published: 9 October 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Modeling count time series is essential in many applications such as pandemic incidences, insurance claims, and integer financial data such as transactions. Different types of count time series models have been broadly developed within the two classes of observation-driven and parameter-driven models [1,2], where in an observation-driven model, current parameters are deterministic functions of lagged dependent variables as well as contemporaneous and lagged exogenous variables, while in parameter-driven models, parameters vary over time as dynamic processes with idiosyncratic innovations. The observation-driven models include the integer-valued generalized autoregressive conditional heteroskedasticity (INGARCH) model [3,4], integer-valued autoregressive model, also called Poisson autoregressive model [5], generalized linear autoregressive moving average (GLARMA) model [6,7], and Poisson AR model [8], among others (see [9,10] for a comprehensive overview). The parameter-driven models include the Poisson state space model [11], Poisson exponentially weighted moving average (PEWMA) model [12], and dynamic count mixture model [13], among others. Some of this research proceeds under a Bayesian framework [13–15].

For nonstationary count time series with changing trends, e.g., daily new COVID cases data, the stationary methods [3,4,8,12] may not capture local trends or give good multi-step-ahead forecasts. The motivation for our study is to propose an efficient method by which to capture the time-varying pattern of the means for such nonstationary count time series and therefore make better forecasts than traditional methods. To capture the

evolutionary properties, a parameter-driven (process-driven) model with a time-varying coefficient latent process is a good choice. Moreover, a latent process with appropriately modeled innovations can address the potential over-dispersion issue, which is common in count time series modeling.

We propose a time-varying Poisson autoregressive (TV-Pois-AR) model for nonstationary count time series and utilize the efficiency of the Bayesian lattice filter (BLF, [16]) and the no-U-turn sampler (NUTS, [17,18]) together to estimate the model parameters. We use a time-varying autoregressive (TV-AR) latent process to model the nonstationary intensity of the Poisson process. This flexible model can capture the latent dynamics of the time series and therefore make superior forecasts. The estimation of such a TV-AR process is greatly sped up by using BLF. Moreover, NUTS is used to sample the intensity-related parameters which have no analytical forms of full conditional distributions. NUTS is an extension to the Hamilton Monte Carlo (HMC, [19–21]) algorithm via automatic tuning of the step size and the number of steps. The use of the HMC method inside the Gibbs sampling was investigated in [22]. According to their paper, as a self-tuned extension to HMC, NUTS should work well as a univariate sampler inside the Gibbs sampling. Benefiting from the joint use of the Bayesian lattice filter and no-U-turn sampler, the estimation of the TV-Pois-AR model is efficient and fast, especially for higher model orders or longer length time series. In short, our main contribution is computational in that we develop a methodology for the efficient estimation of Poisson time-varying autoregressive models.

The rest of the paper is organized as follows. In Section 2, we formulate the proposed Bayesian time-varying coefficient Poisson model. In Section 3, we present a simulation study that illustrates the small sample properties of the proposed Bayesian estimators. In Section 4, our proposed model is demonstrated through an example of COVID-19 spread in New York State and an example of US COVID-19 hospitalization data. In Section 5, we summarize the proposed method and discuss possible future research directions. In the Appendices, the detailed algorithms are introduced.

## 2. Methodology

### 2.1. TV-Pois-AR( $P$ ) Model

For a univariate count-valued series,  $z_t$ ,  $t = 1, \dots, T$ , we propose a TV-Pois-AR model of order  $P$  (TV-Pois-AR( $P$ )) defined as

$$\begin{aligned} z_t | \mathbf{X}_t \boldsymbol{\psi}, y_t &\sim \text{Pois}(\exp(\mathbf{X}_t \boldsymbol{\psi} + y_t)), \\ y_t &= \sum_{j=1}^P a_{j,t} y_{t-j} + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0, \sigma_t^2), \end{aligned} \quad (1)$$

where  $\exp(\cdot)$  stands for the exponential function,  $\mathbf{X}_t$  is  $1 \times K$  vector of covariates,  $\boldsymbol{\psi}$  is a  $K \times 1$  vector of coefficients, and  $y_t$  is the autoregressive component of the logarithm of the Poisson intensity, which follows a TV-AR process. Depending on the specific application case,  $\mathbf{X}_t \boldsymbol{\psi}$  can be a constant term  $\mu$  or removed from the model. Throughout the context, we use a constant term  $\mu$  and do not consider any covariate. We define  $a_{j,t}$  and  $\sigma_t^2$  to be the time-varying AR coefficients of TV-AR( $P$ ) associated with time lag  $j$  at time  $t$  and the innovation variance at time  $t$ , respectively. The innovations,  $\zeta_t$ , are defined to be independent Gaussian errors.

### 2.2. Bayesian Lattice Filter for the TV-AR Process

Under the Bayesian framework, a major part of the model inference is the estimation of parameters in the latent TV-AR process of  $y_t$ , the posterior distributions of which are not conjugate. Using Monte Carlo methods to generate converged sample chains for the time-varying parameters constitutes a large computational burden. BLF provides an efficient way to directly obtain the posterior means of these time-varying parameters. Using the posterior means as MC samples greatly accelerates the convergence of sample chains. According to the Durbin–Levinson algorithm, there exists a unique correspondence between the

partial autocorrelation (PARCOR) coefficients and the AR coefficients [16,23,24]. This lattice structure provides an efficient way to estimate the PARCOR coefficients, which are associated with AR models (see [25] and the Supplemental materials [16]). The efficient estimation of this TV-AR process can be conducted through the following  $P$ -stage lattice filter. We denote  $f_t^{(P)}$  and  $b_t^{(P)}$  to be the prediction error at time  $t$  for the forward and backward TV-AR( $P$ ) models, respectively, where

$$\begin{aligned} f_t^{(P)} &= y_t - \sum_{j=1}^P a_{j,t}^{(P)} y_{t-j}, \\ b_t^{(P)} &= y_t - \sum_{j=1}^P d_{j,t}^{(P)} y_{t+j}, \end{aligned} \tag{2}$$

and  $a_{j,t}^{(P)}$  and  $d_{j,t}^{(P)}$  are the forward and backward autoregressive coefficients of the corresponding TV-AR( $P$ ) models. Then, in the  $j$ th stage of the lattice filter for  $j = 1, \dots, P$ , the forward and the backward coefficients and the forward and backward prediction errors have the following relationship:

$$\begin{aligned} f_t^{(j)} &= f_t^{(j-1)} - \alpha_{j,t}^{(j)} b_{t-j}^{(j-1)}, \\ b_t^{(j)} &= b_t^{(j-1)} - \beta_{j,t}^{(j)} f_{t+j}^{(j-1)}, \end{aligned} \tag{3}$$

where  $\alpha_{j,t}^{(j)}$  and  $\beta_{j,t}^{(j)}$  are the lag  $j$  forward and backward PARCOR coefficients at time  $t$ , respectively. The initial condition,  $f_t^{(0)} = b_t^{(0)} = y_t$ , can be obtained from the definition in (3). This implies that the samples of  $y_t$  are plugged in as the initial values of  $f_t^{(0)}$  and  $b_t^{(0)}$  in the Gibbs sampling. At the  $j$ th stage of the lattice structure, we fit time-varying AR(1) models to estimate  $\alpha_{j,t}^{(j)}$  and  $\beta_{j,t}^{(j)}$ . The corresponding forward and backward autoregressive coefficients at time  $t$ ,  $a_{j,t}^{(j)}$  and  $d_{j,t}^{(j)}$  can be obtained according to the following equations:

$$\begin{aligned} a_{i,t}^{(j)} &= a_{i,t}^{(j-1)} - \alpha_{j,t}^{(j)} d_{j-i,t}^{(j-1)}, \\ d_{i,t}^{(j)} &= d_{i,t}^{(j-1)} - \beta_{j,t}^{(j)} a_{j-i,t}^{(j-1)}, \end{aligned} \tag{4}$$

with  $i = 1, \dots, j - 1$ ,  $a_{j,t}^{(j)} = \alpha_{j,t}^{(j)}$  and  $d_{j,t}^{(j)} = \beta_{j,t}^{(j)}$ . Finally, the distribution of  $a_{j,t}^{(P)}$  and  $\zeta_t$  for  $j = 1, \dots, P$  are obtained. These distributions are used as conditional distributions of  $a_{j,t}$  and  $\zeta_t$  in the Gibbs sampling.

### 2.3. Model Specification and Bayesian Inference

We assume that each coefficient in  $\beta$  has a conjugate normal prior distribution, i.e.,  $\beta_k \stackrel{i.i.d.}{\sim} N(\mu_0, \tau_0^2)$ , and the initial state of the latent variable  $y_0$  follows a normal distribution, s.t.,  $y_0 \sim N(m_0, s_0^2)$ . In Gibbs sampling,  $\mu$  is sampled efficiently by NUTS, and this speeds up the mixing of the sample chains. Compared to the Metropolis–Hastings algorithm, which uses a Gaussian random walk as proposal distribution, NUTS generates samples converging to the target distribution. The target distribution of  $y_t$ , for  $t = 1, \dots, T$ , is its conditional distribution with the density function

$$\begin{aligned} p(y_t | z_t, \mathbf{y}_{-t}, \boldsymbol{\theta}) &= p(z_t | \mu, y_t) p(y_t | \mathbf{y}_{-t}, \boldsymbol{\theta}) \\ &\propto p(z_t | \mu + y_t) p(y_t | \mathbf{y}_{t-1}, \boldsymbol{\theta}) p(y_{t+1} | \mathbf{y}_t, \boldsymbol{\theta}), \end{aligned}$$

where  $\mathbf{y}_t = (y_t, \dots, y_{t-P+1})'$ ,  $\boldsymbol{\theta}$  denotes  $a_{j,t}$ , and  $\sigma_t^2$  for all  $t$  and  $\mathbf{y}_{-t}$  denotes  $y_i$  for all  $i$  but  $t$ . According to the previous assumptions, the conditional distributions of  $z_t$  is Poisson and

the conditional distributions of  $y_t$  and  $y_{t+1}$  are Gaussian. Having the target distribution, NUTS can adaptively draw samples of  $y_t$  conditional on all other variables for all  $t$ .

To use the BLF to derive the conditional distributions of the parameters in the latent autoregressive process of  $y_t$ , we define the distribution of its coefficients  $a_{j,t}^{(P)}$  by defining the distributions of the forward and backward PARCOR coefficients in (3). To give time-varying structures to the forward and backward PARCOR coefficients, we consider random walks for the PARCOR coefficients. The PARCOR coefficients are modeled as

$$\begin{aligned} \alpha_{j,t}^{(j)} &= \alpha_{j,t-1}^{(j)} + \epsilon_{\alpha,j,t}, \quad \epsilon_{\alpha,j,t} \sim N(0, \omega_{\alpha,j,t}), \\ \beta_{j,t}^{(j)} &= \beta_{j,t-1}^{(j)} + \epsilon_{\beta,j,t}, \quad \epsilon_{\beta,j,t} \sim N(0, \omega_{\beta,j,t}), \end{aligned}$$

where  $\omega_{\alpha,j,t}$  and  $\omega_{\beta,j,t}$  are time dependent evolution variance. These evolution variances are defined via the standard discount method in terms of the discount factors  $\gamma_{f,j}$  and  $\gamma_{b,j}$  within the range  $(0, 1)$ , respectively (see Appendices A–D and [26] for details). The discount factor  $\gamma$  controls the smoothness of PARCOR coefficients. Here, we assume  $\gamma_{f,j} = \gamma_{b,j} = \gamma_j$  at each stage  $j$ . Similarly, the innovation variances are assumed to follow multiplicative random walks and modeled as

$$\begin{aligned} \sigma_{f,j,t}^2 &= \sigma_{f,j,t-1}^2 (\delta_{f,j} / \eta_{f,j,t}), \quad \eta_{f,j,t} \sim \beta(g_{f,j,t}, h_{f,j,t}), \\ \sigma_{b,j,t}^2 &= \sigma_{b,j,t-1}^2 (\delta_{b,j} / \eta_{b,j,t}), \quad \eta_{b,j,t} \sim \beta(g_{b,j,t}, h_{b,j,t}), \end{aligned}$$

where  $\delta_{f,j}$  and  $\delta_{b,j}$  are also discount factors in the range  $(0, 1)$ , and the multiplicative innovations,  $\eta_{f,j,t}$  and  $\eta_{b,j,t}$ , follow beta distributions with hyperparameters  $(g_{f,j,t}, h_{f,j,t})$  and  $(g_{b,j,t}, h_{b,j,t})$  (see Appendices A–D and [26] for details). The smoothness of innovation variance is controlled by both  $\gamma$  and  $\delta$ . Similar to the PARCOR coefficients, we assume  $\delta_{f,j} = \delta_{b,j} = \delta_j$  at each stage. Note that  $\epsilon_{\alpha,j,t}$ ,  $\epsilon_{\beta,j,t}$ ,  $\eta_{f,j,t}$ , and  $\eta_{b,j,t}$  are mutually independent and are also independent of any other variables in the model. The discount factors  $\gamma$  and  $\delta$  are selected adaptively through a grid-search based on the likelihood (see the Appendices A–D for details) in each iteration of MCMC.

We specify conjugate initial priors for the forward and backward PARCOR coefficients, so that

$$\begin{aligned} p(\alpha_{j,0} | D_{f,j,0}, \sigma_{f,j,0}) &\sim N(\mu_{f,j,0}, C_{f,j,0}), \\ p(\beta_{j,0} | D_{b,j,0}, \sigma_{b,j,0}) &\sim N(\mu_{b,j,0}, C_{b,j,0}), \end{aligned}$$

where  $p = 1, \dots, P$ ,  $D_{f,j,0}$ ,  $D_{b,j,0}$  denotes the information available at the initial time  $t = 0$ , and  $\mu_{f,j,0}$  and  $C_{f,j,0}$  are the mean and the variance of the normal prior distribution. We also specify conjugate initial priors for the forward and backward innovation variance, so that

$$\begin{aligned} p(\sigma_{f,j,0}^2 | D_{f,j,0}) &\sim G(v_{f,j,0}/2, \kappa_{f,j,0}/2), \\ p(\sigma_{b,j,0}^2 | D_{b,j,0}) &\sim G(v_{b,j,0}/2, \kappa_{b,j,0}/2), \end{aligned}$$

where  $G(\cdot, \cdot)$  is the gamma distribution, and  $v_{f,j,0}/2$  and  $\kappa_{f,j,0}/2$  are the shape and rate parameters for the gamma prior distribution. Usually, we treat these starting values as constants over all stages. In order to reduce the effect of the prior distribution, we choose  $\mu_{f,j,0}/2$  and  $C_{f,j,0}$  to be zero and one, respectively, and fixed  $v_{f,j,0} = 1$  and set  $\kappa_{f,j,0}$  equal to  $v_{f,j,0}$  divided by the sample variance of the initial part of each series according to the formula for the expectation of the gamma distribution. The conjugate initial priors for  $\beta_{j,0}$  and  $\sigma_{b,j,0}^2$  are specified in manner analogous to those of  $\alpha_{j,0}$  and  $\sigma_{f,j,0}^2$ . A sensitivity analysis was conducted and showed that the simulation studies in Section 3 and the case studies in Section 4 are not sensitive to the choice of the priors and the hyperparameters. In such prior settings, we can use the DLM sequential filtering and smoothing algorithms [26] to derive the joint conditional posterior distributions of the forward and backward PARCOR

coefficients and innovation variances in (3). Conditional on the other variables and the data, the full conditional distribution of the latent variable  $y_t$  can easily be obtained individually. To efficiently draw samples from the individual full conditional distribution for the  $y_t$ s, we use the NUTS algorithm [17] instead of a traditional random walk Metropolis. The detailed algorithms for the Gibbs sampling, the BLF and the sequential filtering and smoothing are given in Appendices A–C.

### 2.4. Model Selection

In order to determine the model order, we set a maximal order  $P_{max}$  and fit TV-Pois-AR( $P$ ) for  $P = 1, \dots, P_{max}$ . The model selection criteria are computed one by one for any specified order. By comparing model selection criteria, we can select the best model order. Since Bayesian inference for the TV-Pois-AR model is conducted through MCMC simulations, we choose the deviance information criterion (DIC) [27,28] and the widely applicable information criterion (WAIC) [29,30].

### 2.5. Forecasting

Having estimated all parameters, we consider 1-step-ahead forecasts of the TV-Pois-AR( $P$ ) model. Then, the 1-step-ahead predictive posterior distribution of the PARCOR coefficients and innovation variance can be obtained according to [26]. The samples of the PARCOR coefficients and innovation variance can be drawn from their predictive distribution. The samples of the 1-step-ahead prediction of the parameters  $\mathbf{a}_{T+1} = (a_{1,T+1}, \dots, a_{p,T+1})'$  can be obtained through the Durbin–Levinson algorithm from the samples of the PARCOR coefficients. After drawing the samples of innovation variance  $\sigma_{T+1}^2$  from its predictive distribution, the samples of  $y_{T+1}$  are drawn from its predictive distribution, such that

$$y_{T+1}^{(j)} | y_{1:T}, \mathbf{a}_{T+1}^{(j)}, \sigma_{T+1}^{2(j)} \sim N\left(\sum_{p=1}^P a_{p,T+1}^{(j)} y_{T+1-p}, \sigma_{T+1}^{2(j)}\right), \quad j = 1, \dots, J.$$

with the samples of  $\mu$  from its posterior distribution, the samples of the 1-step-ahead forecast are given as

$$z_{T+1}^{(j)} | y_{T+1}^{(j)}, \mu^{(j)} \sim \text{Pois}(\exp(y_{T+1}^{(j)} + \mu^{(j)})), \quad j = 1, \dots, J. \tag{5}$$

We use the posterior median of  $z_{T+1}$  obtained through the samples in (5) as the 1-step-ahead forecast. This forecast can be easily extended to  $h$ -steps ahead. The details of forecasting up to  $h$ -steps ahead can be found in the Appendices A–D.

## 3. Simulation Study

In this section, first, we simulate the nonstationary Poisson time series from the exact TV-Pois-AR( $P$ ) model to evaluate the parameter estimation of the latent TV-AR process. Second, we generate a nonstationary Poisson time series based on a known time-dependent intensity parameter in order to compare our TV-Pois-AR model with other models. This constitutes an empirical simulation.

### 3.1. Simulation 1

We simulated 100 time series for each of the lengths  $T = 200, 300, 400$  from the following Poisson TV-AR(6) model, for  $t = 1, \dots, T$ ,

$$z_t | \mu, y_t \sim \text{Pois}(\exp(\mu + y_t)),$$

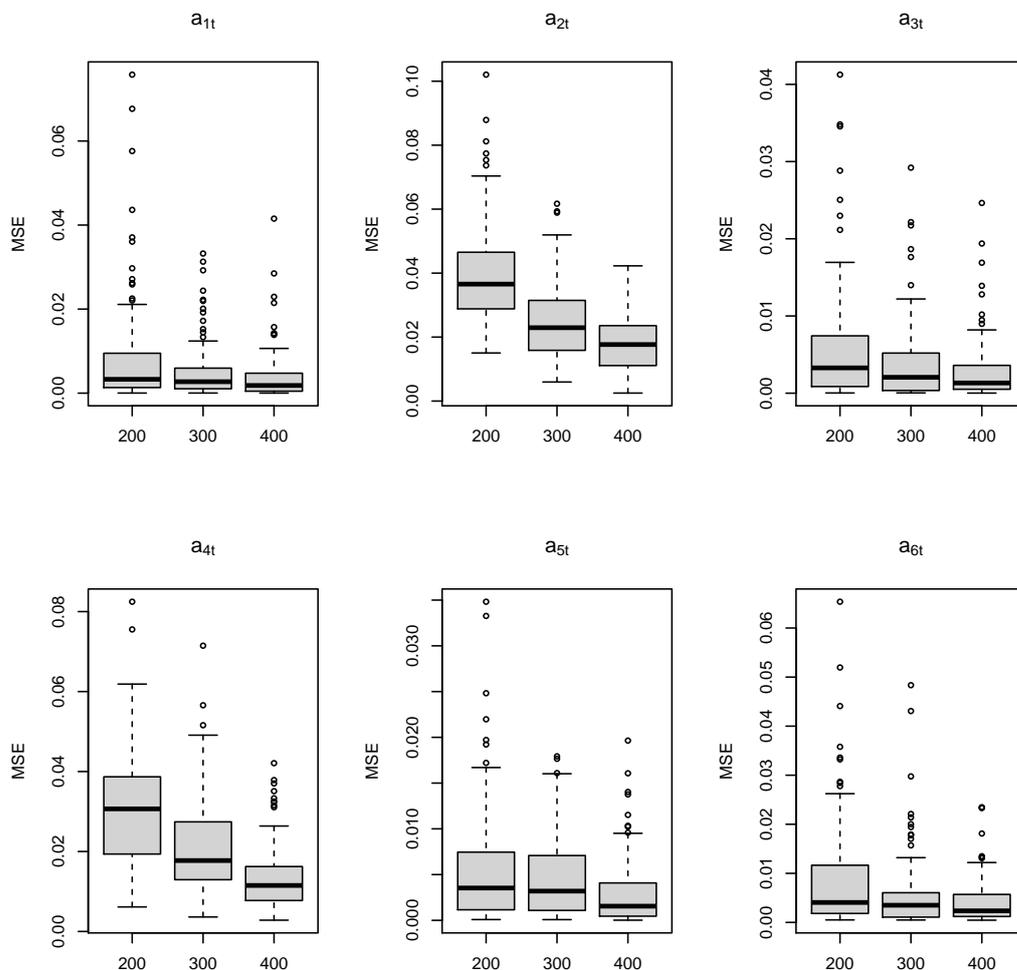
$$y_t = \sum_{j=1}^6 \phi_{j,t} y_{t-j} + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0, 1),$$

where  $\mu = 3$ , which gives a constant mean level to the intensity. The latent process of  $y_t$  is the same time-varying TV-AR(6) process as in [31]. This TV-AR(6) process can be defined

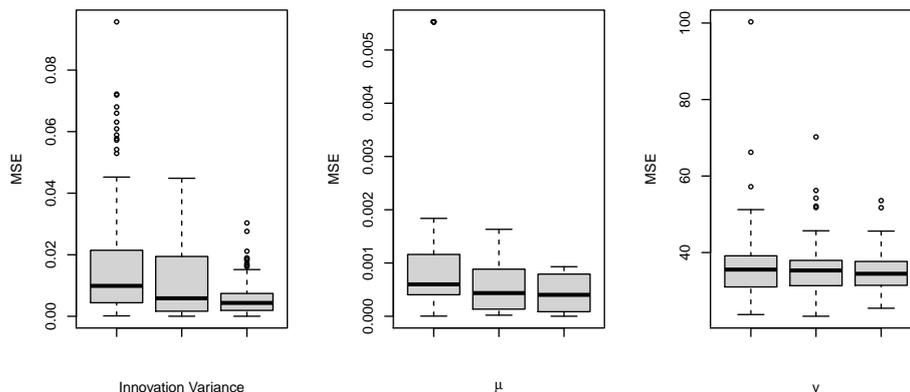
as  $\phi_t(B)y_t = \zeta_t, t = 1, \dots, T$ , through a characteristic polynomial function  $\phi_t(B)$ , with  $B$  as the backshift operator (i.e.,  $B^p y_t = y_{t-p}$ ). In this TV-AR(6) process, the characteristic polynomial function is factorized as

$$\phi_t(B) = (1 - \phi_{t,1}B)(1 - \phi_{t,1}^*B)(1 - \phi_{t,2}B)(1 - \phi_{t,2}^*B)(1 - \phi_{t,3}B)(1 - \phi_{t,3}^*B),$$

where the superscript  $*$  denotes the complex conjugate of a complex number. Moreover, let  $\phi_{t,j}^{-1} = A_p \exp(2\pi i d_{t,j})$  for  $p = 1, 2, 3$ , where the  $d_{t,j}$ s are defined by  $d_{t,1} = 0.05 + (0.1/(T - 1))t$ ,  $d_{t,2} = 0.25$ , and  $d_{t,3} = 0.45 - (0.1/(T - 1))t$ , and the values of  $A_1, A_2$ , and  $A_3$  are equal to 1.1, 1.12, and 1.1, respectively. Here, we take  $T = 200, 300$ , and  $400$  to be of a similar order to our case study in Section 4.1. According to DIC and WAIC, 98% of the simulated datasets are identified to follow an order-6 model (TV-Pois-AR). To evaluate the parameter estimation of time-varying parameters, we use the mean squared error (MSE); that is, the average of the squared difference between the estimated parameter value and its true value at each observed time point. Table 1 and Figures 1 and 2 show the MSEs of 6 time-varying autoregressive coefficients, the time-varying innovation variance, the mean level  $\mu$ , and the latent variable  $y_t$  over 100 simulated datasets. As expected, when the series length increases, the TV-Pois-AR model gives a more accurate estimation of each parameter.



**Figure 1.** Boxplots of the MSEs of each of the six time-varying coefficients  $a_{1,t}$  through  $a_{6,t}$  for 100 simulated datasets of different lengths: 200, 300, 400.



**Figure 2.** Boxplots of the MSEs of the intensity and the parameters in the latent process for 100 simulated datasets of different lengths. For each of them, three boxplots of length 200, 300, 400 are put side by side from left to right. The left plot shows the MSEs of the innovation variance. The middle plot shows the MSEs of the mean level  $\mu$ . The right plot shows the MSEs of the latent variable  $y$ .

**Table 1.** Average and standard deviation (s.d.) of MSEs of each of the six time-varying coefficients  $a_{1,t}$  through  $a_{6,t}$  for 100 simulated datasets of different length: 200, 300, 400.s.

	Average of MSEs (s.d. of MSEs)		
	200	300	400
$a_{1t}$	0.0086 (0.0136)	0.0055 (0.0070)	0.0040 (0.0063)
$a_{2t}$	0.0404 (0.0165)	0.0254 (0.0119)	0.0183 (0.0083)
$a_{3t}$	0.0061 (0.0080)	0.0039 (0.0052)	0.0030 (0.0042)
$a_{4t}$	0.0307 (0.0139)	0.0212 (0.0121)	0.0136 (0.0086)
$a_{5t}$	0.0058 (0.0068)	0.0047 (0.0047)	0.0031 (0.0039)
$a_{6t}$	0.0091 (0.0117)	0.0058 (0.0081)	0.0044 (0.0049)

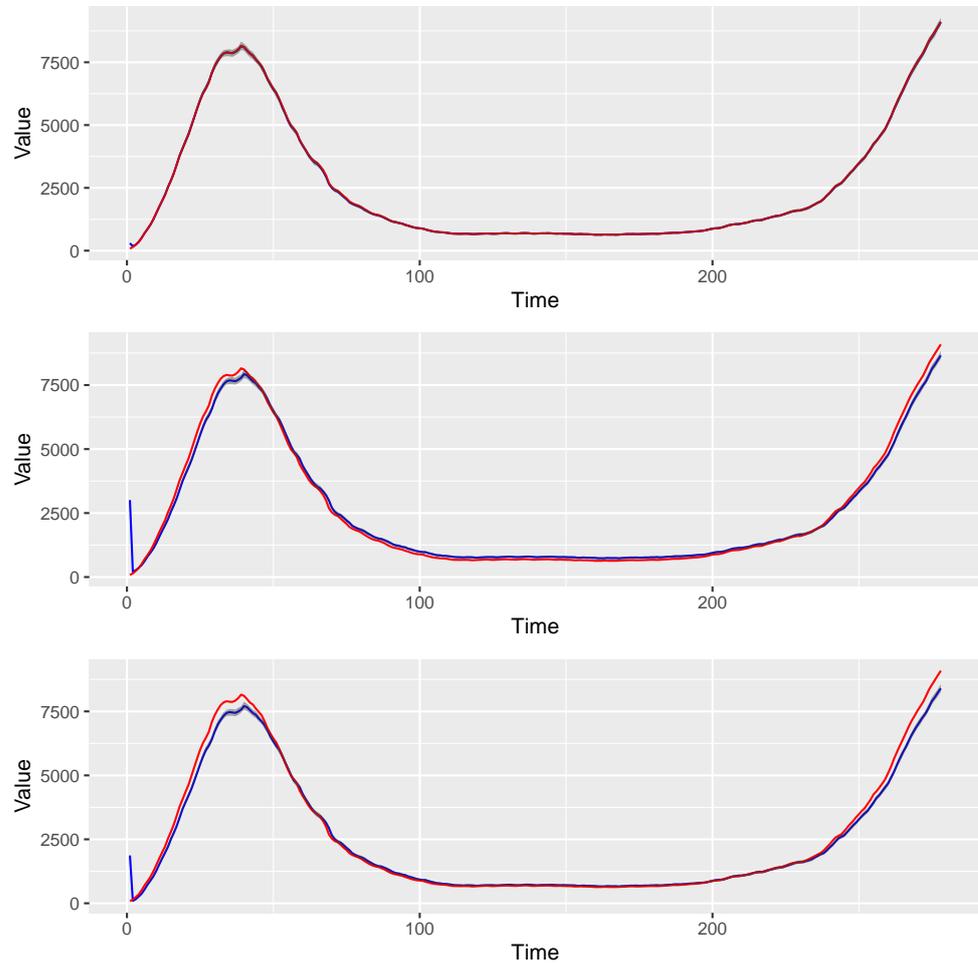
### 3.2. Simulation 2—An Empirical Simulation

In this study, we simulated the signals based on the COVID-19 data in New York State (see Section 3.1 for a complete discussion) so that they exhibit similar properties. We generated 100 time series of length  $T = 278$  from a Poisson process:

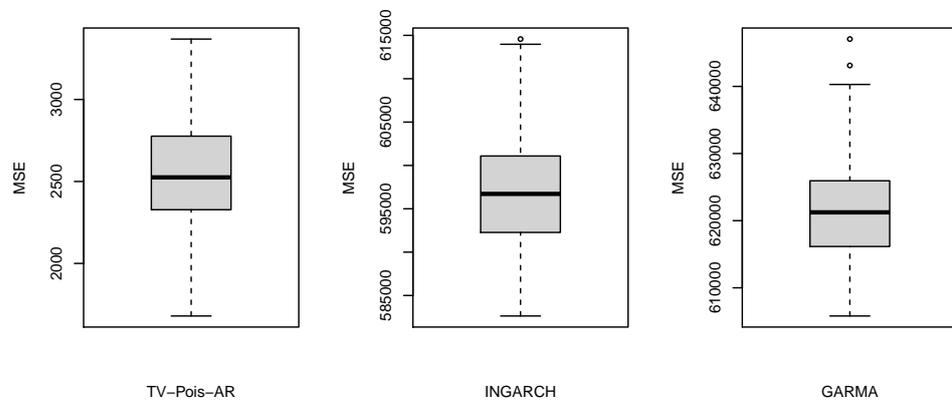
$$z_t | \lambda_t \sim \text{Pois}(\lambda_t), \quad t = 1, \dots, T,$$

where  $\lambda_t$  was the 7-day moving average of the estimated intensity of daily new COVID-19 cases in New York State from 3/3/2020 to 12/5/2020. With this type of nonstationary signal, different models are compared by the estimation of the known time-varying intensity parameter  $\lambda_t$ , including the INGARCH and GLARMA model. The INGARCH and GLARMA models are conducted via *tsglm* from R package *tscount*. Although these models have different underlying assumptions, they are sometimes used in practices as they can still provide reasonable forecasts. Using the Akaike information criterion (AIC) and the quasi information criterion (QIC) [32], INGARCH(1,0) and GLARMA(1,0) are selected. Both DIC and WAIC indicate that TV-Pois-AR(1) is the best model for these simulated datasets (see details in Section 2.4). To compare the estimation from the frequentist and Bayesian models, the average MSE (AMSE) of the Poisson intensity parameter is computed and shown in Table 2 and Figure 3, where  $\text{AMSE} = \frac{1}{100T} \sum_{s=1}^{100} \sum_{t=1}^T (\lambda_t - \hat{\lambda}_t)^2$ . The boxplots in Figure 4 summarize the MSEs of 100 simulated datasets. The estimated intensity figure shows the mean and 90% coverage interval by the three models. As shown, TV-Pois-AR makes better forecasts on these simulated datasets. We expect the TV-Pois-AR model to

give a better performance than INGARCH and GLARMA on similar pandemic data and other nonstationary count time series that show similar characteristics.



**Figure 3.** The mean and 90% pointwise credible intervals of estimated intensity over 100 simulated datasets using different methods. The blue line denotes the estimated values, and the band denotes 90% pointwise credible intervals. The red line denotes the true values of intensity. Note that “pointwise credible intervals” are calculated from 100 point estimates independently at each time.



**Figure 4.** Boxplots of MSE of the estimated intensity over 100 simulated datasets using different methods. Note that the scales of the three boxplots are different.

**Table 2.** Average mean squared error (AMSE) of the estimated intensity over the 100 simulated datasets using different methods.

Model	INGARCH	GLARMA	TV-Pois-AR
AMSE	597,041.3	621,732.5	2568.6

#### 4. Case Studies

To illustrate our proposed methodology, we provide two case studies. The first case study considers COVID-19 case in the New York State; whereas the second case study considers COVID-19 hospitalizations in the U.S. Both case studies are meant to be an illustration of the methodology and thus do not represent a substantive analysis of the COVID-19 pandemic.

##### 4.1. Case Study 1: COVID-19 in New York State

We obtained the 278 daily numbers of confirmed COVID-19 cases in New York State from 3/3/2020 to 12/5/2020 from The COVID Tracking Project (<https://covidtracking.com> (accessed on 2/7/2021)). We picked New York State data as New York city remained an epicenter in the U.S. for about a month. Our research is motivated by the time-varying nature of the COVID-19 data. Inferences on the trend of the data may give us some insight into the spread of COVID-19 and, possibly, insight into the effect of government interventions.

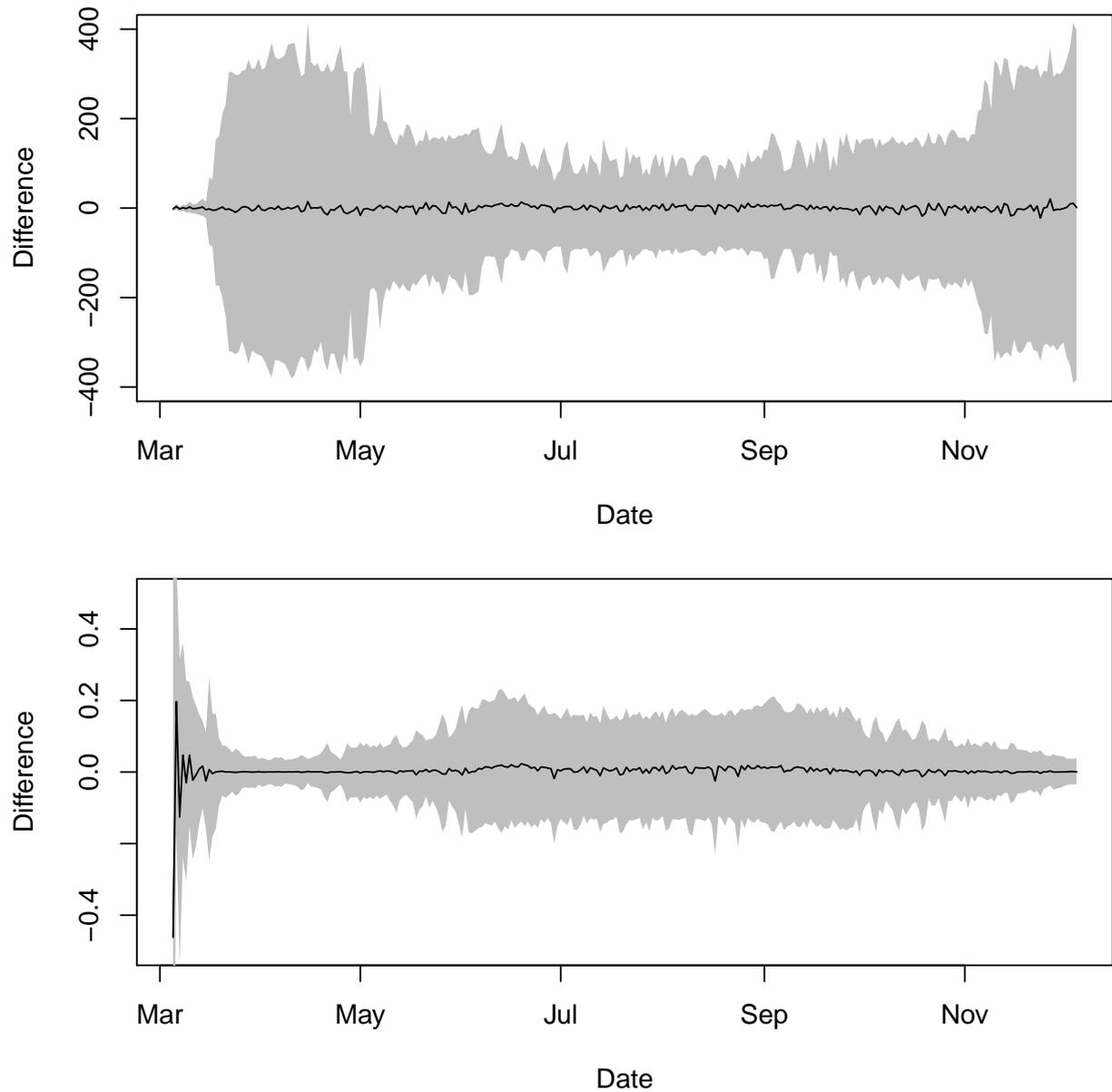
A TV-Pois-AR model was applied without the fix effect,  $X\beta$ , because we do not have scientific information about any potential covariates. By setting a maximum order of 5, order 2 was considered as the best based on DIC and WAIC. The difference between the estimated exponential of the intensity parameter,  $\exp(\lambda_t)$ , and the observed series is shown in Figure 5. Figure 6 shows the estimated parameters. Table 3 shows the model selection results. A series of restrictions in New York State began on 3/12/2020, and a state-wide stay-at-home order was declared on 3/20/2020. The number of new cases reached its peak about two weeks after the lockdown. In Figure 6, the first dashed line in the first two plots denotes 3/20/2020, the time when the state-wide stay-at-home order was declared. We can see that the estimated autoregressive coefficients keep changing significantly after this date. This change in the autoregressive coefficients coincides with the lockdown process. The second dashed line in the first two plots denotes 9/26/2020. On that day, the number showed an uptick in cases, with more than 1000 daily COVID-19 cases, which was the first time since early June. About two weeks before this date, the coefficients show some evidence of pattern change. This may be an indication that the lockdown affected the spread of COVID-19. The innovation variance of the intensity is becoming smaller and smaller, probably due to the improvement in testing and reporting. The dashed line in the third plot denotes the date when the peak number of new cases occurred. Since then, the innovation variance has stabilized at a low level.

To evaluate the performance of forecasting, we conducted a rolling one-day-ahead prediction and compared the mean squared prediction errors (MSPE). We picked a starting date and made a one-day-ahead prediction based on the data up to this date. Then, we moved to the next day and made a one-day-ahead prediction based on the data up to the new date. By repeating this until one day before the last day, we obtained the rolling one-day-ahead prediction. Additionally, we conducted a 20-day prediction to evaluate the performance of the long-term forecast.

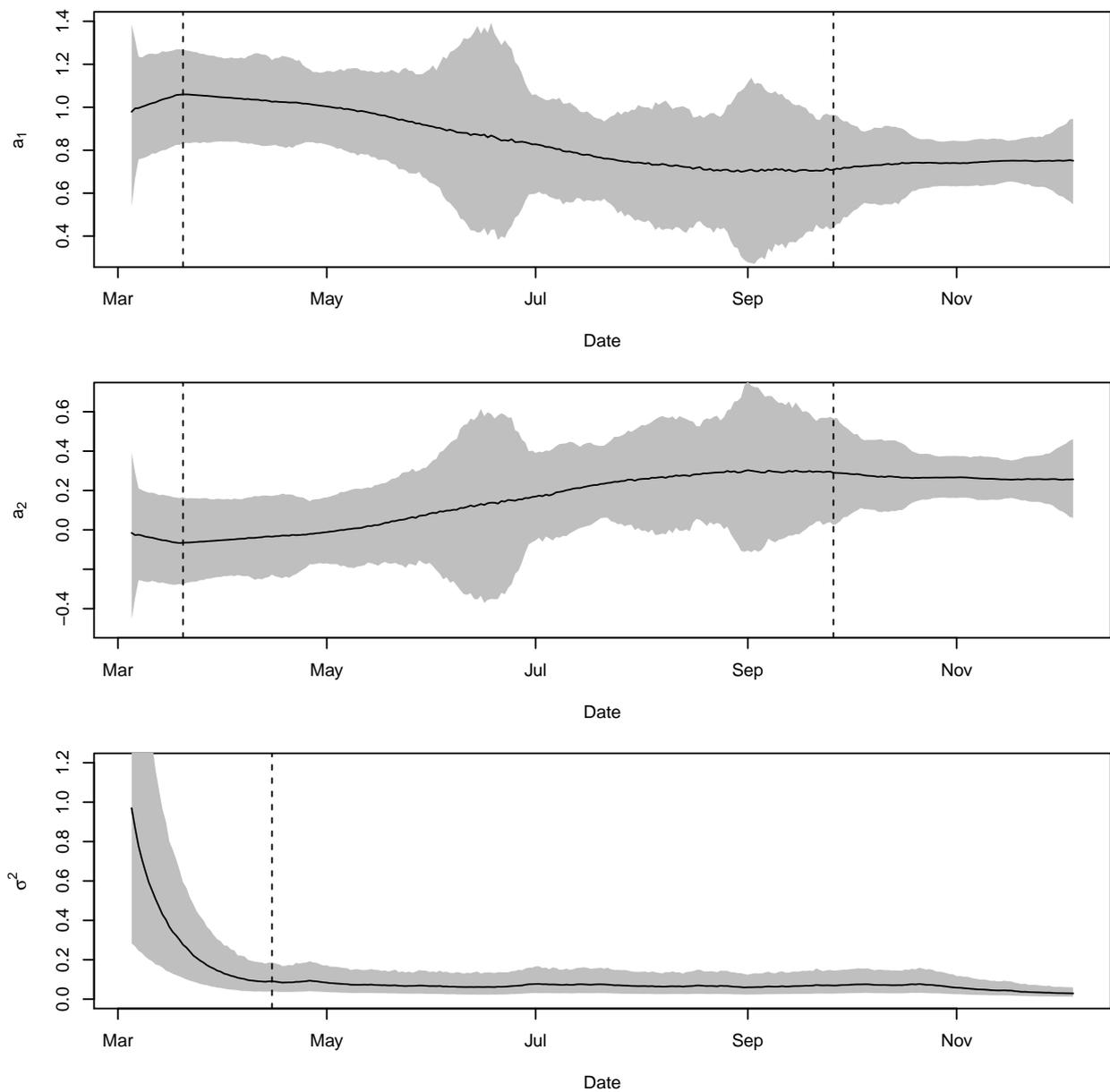
**Table 3.** Model selection of Poisson TV-AR model for the daily new COVID-19 cases in New York State. Each column gives the model order  $P$  and the value of the model selection criterion.

P	1	2	3	4	5
DIC	1738.540	1735.748	1783.484	1779.946	1796.255
WAIC	303.004	295.531	317.326	316.626	320.476

We compare the forecast performance of four methods, where Naive denotes the naive forecast; that is, using the previous period to forecast for the next period (carry-forward). The average MSPE over these days is used for comparison. Table 4 presents the performance of the rolling one-day-ahead prediction. The TV-Pois-AR forecasting outperforms the two existing models and the Naive forecasting. Figure 7 shows an example of a 20-day forecast of the daily new COVID-19 cases from 10/18/2020 to 11/6/2020 in New York State. The example demonstrates how the TV-Pois-AR model captures the time-varying trend.



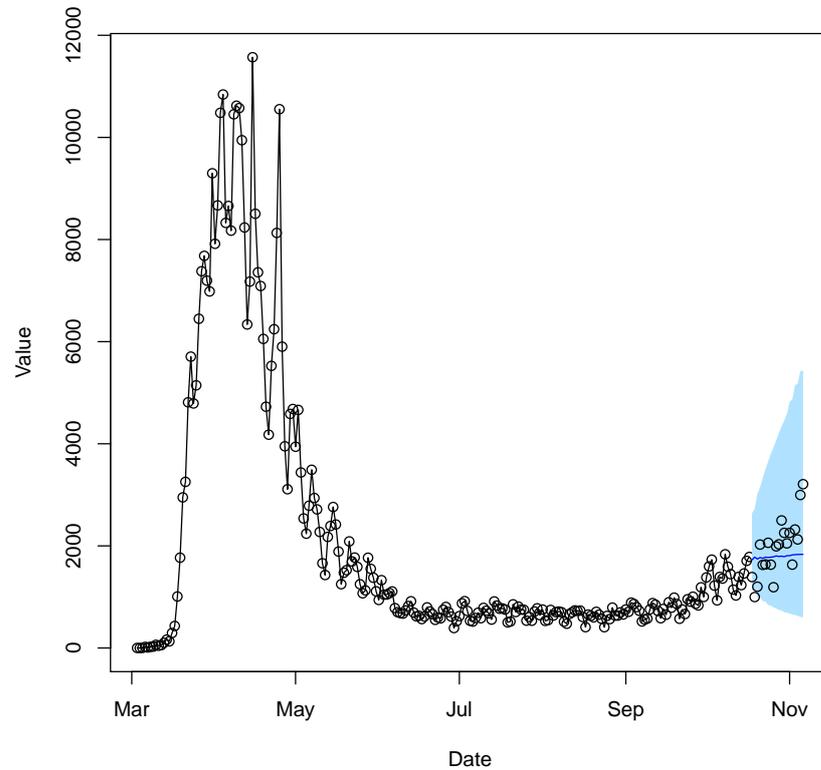
**Figure 5.** The difference between daily new COVID-19 cases in New York State and the estimated expected values. The black line is the difference and the grey region shows the corresponding 90% credible intervals. The top plot shows the difference in the original scale and the bottom plot shows the difference in the log scale.



**Figure 6.** The estimated  $a_{1,t}$ ,  $a_{2,t}$ , and  $\sigma_t^2$  of the Poisson TV-VAR(2) model applied to daily new COVID-19 cases in New York State from top to bottom, respectively. The grey region shows the corresponding 90% credible intervals.

**Table 4.** One-step-ahead predictive performance of TV-Pois-AR(2) and other models on COVID-19 data in New York State from 7/19/2020. There are two start dates for the rolling predictions.

Model	MPSE
TV-Pois-AR(2)	$2.277 \times 10^5$
GLARMA(6,2)	$2.363 \times 10^5$
INGARCH(1,0)	$2.675 \times 10^5$
Naive	$2.286 \times 10^5$



**Figure 7.** The 20-day forecast of the daily new COVID-19 cases of the last 20 days in New York State. The black overplotted points and lines are the observed daily new cases used for model fitting from 3/3/2020 to 10/17/2020. The black dots are the true daily new cases in the forecast region from 10/18/2020 to 11/6/2020. The blue line shows the 20-day forecast. The light blue region is the 90% prediction interval.

4.2. Case Study 2: COVID-19 hospitalization in the U.S.

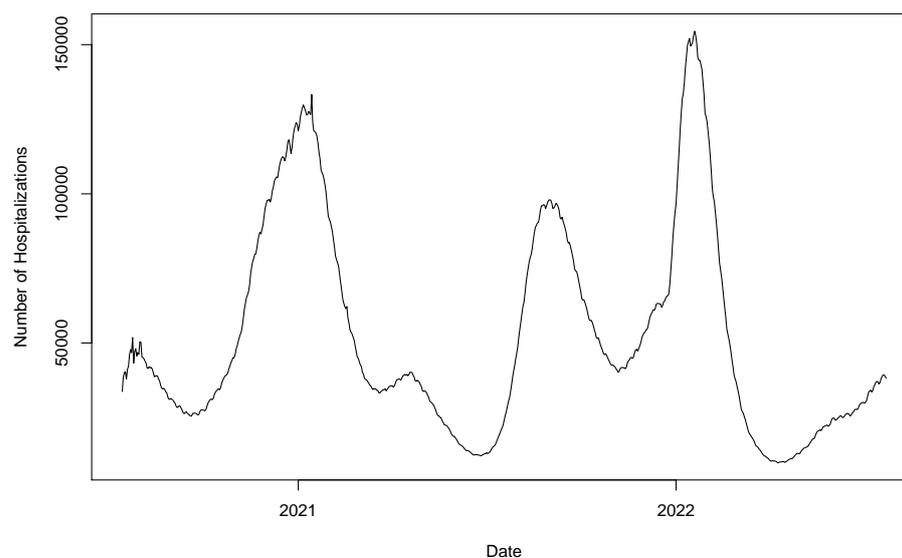
Since the number of daily new COVID-19 cases is no longer systematically collected (starting from early 2022), we use the 739 daily numbers of COVID-19 patients in hospitals in the US from 7/15/2020 to 7/23/2022 (shown in Figure 8) from *Our World in Data* (<https://ourworldindata.org/> (accessed on 11/29/2022)) as another data example. A TV-Pois-AR model is applied, with the model order selected based on DIC and WAIC. By setting a maximum order of 10, order 4 is considered the best, as shown in Table 5. To evaluate the forecasting performance, we make a rolling one-day-ahead prediction and compare the mean squared prediction error (MSPE), as in Case Study 1. The rolling prediction start dates are from 8/19/2021 to 6/14/2022. We compare the forecast performance of four methods, as in Case Study 1. Table 6 shows the rolling one-day-ahead forecast performance of each method in terms of MSPE over the rolling observed COVID data.

**Table 5.** Model selection of Poisson TV-AR model for the daily COVID-19 hospitalization in the U.S. Each column gives the model order  $P$  and the value of the model selection criterion

P	1	2	3	4	5
DIC	24,481.39	23,898.07	23,248.08	22,726.58	23,278.76
WAIC	6518.81	6190.746	5863.422	5612.337	5881.701

**Table 6.** Percentage of TV-Pois-AR(4) giving better forecasts of one-step-ahead rolling predictions on US COVID-19 hospitalization data from 11/27/2021 to 9/22/2022. The posterior medians of the future observations are used as the forecast values.

Model	MSPE
Pois-TVAR	$1.51 \times 10^6$
GLARMA(6,2)	$16.00 \times 10^6$
INGARCH(1,0)	$11.10 \times 10^6$
Naive	$3.20 \times 10^6$



**Figure 8.** The daily number of COVID-19 patients in hospital in the US.

## 5. Discussion

We develop a novel hierarchical Bayesian model to model the nonstationary count time-varying models and propose an efficient estimation approach using an MCMC sampling scheme with embedded NUTS algorithm. We also provide a model selection method by which to choose the discount factors and the optimal model order. The simulation cases show that the parameter estimates have a small mean squared error that, as expected, decreases as the sample size increases. The data example shows that the time-varying coefficients and innovation covariance can reveal the changing pattern over time. The proposed method can be applied not only to the confirmed cases of COVID-19 but also to the number of deaths, number of recovered cases, number of critical cases, and many other parameters for different diseases. Such studies may provide important insights into the spread and the measures required.

The current study is limited to univariate nonstationary count-valued time series. One subject for future research is an extension of the model to multivariate and/or spatiotemporal cases by adding some region-specific effects and jointly modeling the series in multiple regions. Modeling multivariate count-valued time series data is an important research topic in ecology and climatology. Moreover, regarding univariate applications on epidemic disease data, we can consider different government interventions as covariates, which usually have an essential impact on the spread of any infectious disease.

**Author Contributions:** Methodology, Y.S., S.H.H. and W.-H.Y.; software, Y.S. and W.-H.Y.; writing—original draft preparation, Y.S., S.H.H. and W.-H.Y.; writing—review and editing, Y.S., S.H.H. and W.-H.Y. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially supported by the U.S. National Science Foundation (NSF) under NSF grants SES-1853096 and NCSE-2215168.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank the anonymous referees for providing comments that helped improve an earlier version of this article. This article is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the NSF or U.S. Census Bureau.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A. Algorithm of Fitting Poisson TV-AR Time Series

To fit a TV-Pois-AR model, we use Gibbs sampling to generate samples from the full conditional distribution of each of the parameters and the latent variables iteratively. After burn-in, the sample distributions of these parameters and latent variables are the estimated posterior distributions. These samples are generated via Gibbs sampling with the following steps:

- Draw samples of  $\mathbf{y} = (y_1, \dots, y_T)'$  from the full conditional distribution  $p(\mathbf{y}|\mathbf{z}, \mu, \mathbf{a}, \sigma^2) \propto p(\mathbf{z}|\mathbf{y}, \mu)p(\mathbf{y}|\mathbf{a}, \sigma^2)$  using a No-U-Turn Sampler [17,18];
- Use the posterior means of  $\mathbf{a}$  and  $\sigma^2$  obtained from the BLF (see Appendix B) as samples, where  $\mathbf{a} = (a_1, \dots, a_T)'$  and  $\sigma^2 = (\sigma_1^2, \dots, \sigma_T^2)'$ ;
- Draw samples of  $\mu$  from the full conditional distribution  $p(\mu|\mathbf{z}, \mathbf{y}) \propto p(\mathbf{z}|\mathbf{y}, \mu)p(\mu)$  using a no-U-turn sampler.

### Appendix B. Bayesian Lattice Filter

- Step 1. Repeat Step 2 for stage  $p = 1, \dots, P$ ;
- Step 2. Apply the sequential filtering and smoothing algorithm (see Appendix C) to the prediction errors of last stage,  $f_t^{(p-1)}$  and  $b_t^{(p-1)}$ , to obtain  $\hat{\alpha}_t^{(p)} = \mu_t^{(p)}$  and  $\hat{\sigma}_t^{2(p)} = s_t^{(p)}$  of the forward and backward equations, and the forward and backward prediction errors,  $f_t^{(p)}$  and  $b_t^{(p)}$ , for  $t = 1, \dots, T$ ;
- Step 3. The posterior mean of  $a_t$  and  $\sigma_t^2$  are  $\hat{\alpha}_t^{(P)} = \mu_t^{(P)}$  and  $\hat{\sigma}_t^{2(P)} = s_t^{(P)}$  obtained from the  $P$ th stage Step 2.

### Appendix C. Sequential Filtering and Smoothing Algorithm

The filtering and smoothing algorithm can be obtained for the backward case in a similar manner. For any series, any stage, we denote the posterior distribution at time  $t$  as  $(\alpha_t|D_t) \sim T_{\nu_t}(\mu_t, C_t)$ , a multivariate T distribution with  $\nu_t$  df, location parameter  $\mu_t$ , and scale matrix  $C_t$ , and  $(\sigma_t^{-2}|D_t) \sim G(\nu_t/2, \kappa_t/2)$ , a gamma distribution with shape parameter  $\nu_t/2$  and scale parameter  $\kappa_t/2$ . These parameters can be computed for all  $t$  using the filtering equations below. Note that we use  $s_t = \kappa_t/\nu_t$  to denote the usual point estimate of  $\sigma_t^2$ .  $f_t$  in the equation is the forward prediction error. For  $t = 2, \dots, T$ , we have

$$\begin{aligned}\mu_t &= \mu_{t-1} + z_t e_t, \\ C_t &= (R_t - z_t z_t' q_t)(s_t/s_{t-1}),\end{aligned}$$

and

$$\begin{aligned}\nu_t &= \delta \nu_{t-1} + 1, \\ \kappa_t &= \delta \kappa_{t-1} + s_{t-1} e_t^2 / q_t,\end{aligned}$$

where

$$e_t = f_t - z'_{t-1}m_{t-1},$$

$$q_t = z'_{t-1}R_t z'_{t-1} + s_{t-1},$$

and

$$z_t = R_t f_{t-1} / q_t,$$

$$R_t = C_{t-1} + G_t,$$

$$G_t = C_t(1 - \beta) / \beta.$$

After applying the filtering equations up to  $T$ , we compute the full marginal posterior distribution  $(\alpha_t | D_T) \sim T_{\nu_t}(\mu_{t,T}, C_t)$  and  $(\sigma_t^{-2} | D_T) \sim G(\nu_{t,T}/2, \kappa_{t,T}/2)$  through the smoothing equations

$$\mu_{t,T} = (1 - \beta)\mu_t + \beta\mu_{t+1,T}$$

$$C_{t,T} = [(1 - \beta)C_t + \beta^2 C_{t+1,T}](s_{t,T}/s_t)$$

$$\nu_{t,T} = (1 - \delta)\nu_t + \delta\nu_{t+1,T}$$

$$1/s_{t,T} = (1 - \delta)/s_t + \delta s_{t+1,T}$$

and  $\kappa_{t,T} = \nu_{t,T}s_{t,T}$  for  $t = T - 1, \dots, 1$ .

### Appendix D. Forecasting

We can undertake  $h$ -step-ahead forecasting by following these steps.

- For stage  $p$ , compute the  $h$ -step-ahead predictive distribution of the PARCOR coefficients following [26]  $(\alpha_{p,T+h}^{(p)} | D_T) \sim N(\mu_T^{(p)}(h), C_T^{(p)}(h))$  where  $\mu_T(h) = \mu_T^{(p)}$  and  $C_T^{(p)}(h) = C_T^{(p)} + hG_{T+h}^{(p)}$  with  $G_{T+1}^{(p)} = C_T^{(p)}(1 - \beta) / \beta$ ;
- Draw  $J$  samples of  $\{\alpha_{p,T+h}^{(p)}, p = 1, \dots, P\}$  from their predictive distribution;
- For stage  $p$ , compute the  $h$ -step-ahead predictive distribution of innovation variance following [26]:  $(\sigma_{T+h}^{2(p)} | D_T) \sim G(\nu_T^{(p)}(h)/2, \kappa_T^{(p)}(h)/2)$ , where  $\nu_T^{(p)}(h) = \delta^h \nu_T^{(p)}$  and  $\kappa_T^{(p)}(h) = \delta^h \kappa_T^{(p)}$ ;
- For stage  $p$ , draw  $J$  samples of  $\sigma_{T+h}^{2(p)}$  from its predictive distribution  $G(\nu_{k+(T-1)K}^{(p)}(h)/2, \kappa_{k+(T-1)K}^{(p)}(h)/2)$ ;
- Compute the samples of the AR coefficients  $\{a_{p,T+h}^{(p)}, p = 1, \dots, P\}$  through the Durbin–Levinson algorithm from the samples of  $\{\alpha_{p,T+h}^{(p)}, p = 1, \dots, P\}$ ;
- The samples of  $y_{T+h}$  are generated from its predictive distribution, such that

$$y_{T+h}^{(j)} | y_{1:T}, y_{T:(T+h-1)}^{(j)}, \mathbf{a}_{T+h}^{(j)}, \sigma_{T+h}^{2(j)} \sim N\left(\sum_{p=1}^P a_{p,T+h}^{(j)} y_{T+h-p}^{(j)}, \sigma_{T+h}^{2(j)}\right), \quad j = 1, \dots, J,$$

where  $y_{T+h-p}^{(j)} = y_{T+h-p}$  if  $h - p \leq 0$ ;

- With the samples of  $\mu$  from its posterior distribution, the samples of the  $h$ -step-ahead forecast are drawn as

$$z_{T+h}^{(j)} | y_{T+h}^{(j)}, \mu^{(j)} \sim \text{Pois}(\exp(y_{T+h}^{(j)} + \mu^{(j)})), \quad j = 1, \dots, J; \tag{A1}$$

- We use the posterior median of  $z_{T+h}$  obtained through the samples in (A1) as the  $h$ -step-ahead forecast.

## References

1. Cox, D.R.; Gudmundsson, G.; Lindgren, G.; Bondesson, L.; Harsaae, E.; Laake, P.; Juselius, K.; Lauritzen, S.L. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scand. J. Stat.* **1981**, *8*, 93–115.
2. Koopman, S.J.; Lucas, A.; Scharth, M. Predicting time-varying parameters with parameter-driven and observation-driven models. *Rev. Econ. Stat.* **2016**, *98*, 97–110. [[CrossRef](#)]
3. Ferland, R.; Latour, A.; Oraichi, D. Integer-valued GARCH process. *J. Time Ser. Anal.* **2006**, *27*, 923–942. [[CrossRef](#)]
4. Fokianos, K.; Rahbek, A.; Tjøstheim, D. Poisson autoregression. *J. Am. Stat. Assoc.* **2009**, *104*, 1430–1439. [[CrossRef](#)]
5. Al-Osh, M.; Alzaid, A.A. First-order integer-valued autoregressive (INAR (1)) process. *J. Time Ser. Anal.* **1987**, *8*, 261–275. [[CrossRef](#)]
6. Zeger, S.L. A regression model for time series of counts. *Biometrika* **1988**, *75*, 621–629. [[CrossRef](#)]
7. Dunsmuir, W.T. Generalized Linear Autoregressive Moving Average Models. In *Handbook of Discrete-Valued Time Series*; Chapman & Hall/CRC: Boca Raton, FL, USA: 2016; p. 51.
8. Brandt, P.T.; Williams, J.T. A linear Poisson autoregressive model: The Poisson AR (p) model. *Political Anal.* **2001**, *9*, 164–184. [[CrossRef](#)]
9. Davis, R.A.; Holan, S.H.; Lund, R.; Ravishanker, N. (Eds.) *Handbook of Discrete-Valued Time Series*; Chapman & Hall/CRC: Boca Raton, FL, USA : 2016.
10. Davis, R.A.; Fokianos, K.; Holan, S.H.; Joe, H.; Livsey, J.; Lund, R.; Pipiras, V.; Ravishanker, N. Count time series: A methodological review. *J. Am. Stat. Assoc.* **2021**, *116*, 1533–1547. [[CrossRef](#)]
11. Smith, R.; Miller, J. A non-Gaussian state space model and application to prediction of records. *J. R. Stat. Soc. Ser. B (Methodol.)* **1986**, *48*, 79–88. [[CrossRef](#)]
12. Brandt, P.T.; Williams, J.T.; Fordham, B.O.; Pollins, B. Dynamic modeling for persistent event-count time series. *Am. J. Political Sci.* **2000**, *44*, 823–843. [[CrossRef](#)]
13. Berry, L.R.; West, M. Bayesian forecasting of many count-valued time series. *J. Bus. Econ. Stat.* **2019**, *38*, 872–887. [[CrossRef](#)]
14. Bradley, J.R.; Holan, S.H.; Wikle, C.K. Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). *Bayesian Anal.* **2018**, *13*, 253–310. [[CrossRef](#)]
15. Bradley, J.R.; Holan, S.H.; Wikle, C.K. Bayesian hierarchical models with conjugate full-conditional distributions for dependent data from the natural exponential family. *J. Am. Stat. Assoc.* **2020**, *115*, 2037–2052. [[CrossRef](#)]
16. Yang, W.H.; Holan, S.H.; Wikle, C.K. Bayesian lattice filters for time-varying autoregression and time-frequency analysis. *Bayesian Anal.* **2016**, *11*, 977–1003. [[CrossRef](#)]
17. Hoffman, M.D.; Gelman, A. The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **2014**, *15*, 1593–1623.
18. Märtens, K. HMC No-U-Turn Sampler (NUTS) Implementation in R. 2017. Available online: <https://github.com/kasparmartens/NUTS> (accessed on 27 July 2020).
19. Neal, R.M. An improved acceptance procedure for the hybrid Monte Carlo algorithm. *J. Comput. Phys.* **1994**, *111*, 194–203. [[CrossRef](#)]
20. Neal, R.M. MCMC using Hamiltonian dynamics. In *Handbook of Markov Chain Monte Carlo*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2011, Volume 2, p. 2.
21. Duane, S.; Kennedy, A.D.; Pendleton, B.J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222. [[CrossRef](#)]
22. Martino, L.; Yang, H.; Luengo, D.; Kannianen, J.; Corander, J. A fast universal self-tuned sampler within Gibbs sampling. *Digit. Signal Process.* **2015**, *47*, 68–83. [[CrossRef](#)]
23. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2006.
24. Kitagawa, G. *Introduction to Time Series Modeling*; Chapman & Hall/CRC: Boca Raton, FL, USA: 2010.
25. Hayes, M.H. *Statistical Digital Signal Processing and Modeling*; John Wiley & Sons: Hoboken, NJ, USA, 1996.
26. West, M.; Harrison, J. *Bayesian Forecasting and Dynamic Models*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 1997.
27. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **2002**, *64*, 583–639. [[CrossRef](#)]
28. Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; Chapman & Hall/CRC: Boca Raton, FL, USAC : 2013.
29. Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **2010**, *11*, 3571–3594.
30. Watanabe, S. A widely applicable Bayesian information criterion. *J. Mach. Learn. Res.* **2013**, *14*, 867–897.
31. Rosen, O.; Stoffer, D.S.; Wood, S. Local spectral analysis via a Bayesian mixture of smoothing splines. *J. Am. Stat. Assoc.* **2009**, *104*, 249–262. [[CrossRef](#)]
32. Pan, W. Akaike’s information criterion in generalized estimating equations. *Biometrics* **2001**, *57*, 120–125. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.