Applied Stochastic Models in Business and Industry

WILEY

RESEARCH ARTICLE

Foundations of Applied Statistics

MCMC and Particle Filtering for Dynamic INAR Processes

Refik Soyer[1] | Hedibert F. Lopes[2] | Di Zhang[1]

[1]Department of Decision Sciences, George Washington University, Washington, DC, USA | [2]Insper Institute of Education and Research, São Paulo, Brazil

Correspondence: Refik Soyer (soyer@gwu.edu)

Received: 4 June 2025 | **Revised:** 2 May 2026 | **Accepted:** 11 May 2026

Keywords: Bayesian filtering and prediction | count data | MCMC | sequential Monte Carlo

ABSTRACT

In this article, we present a dynamic version of the integer autoregressive (INAR) processes for count data. The proposed Bayesian model provides a unification of the previously considered models to describe temporal correlations in univariate time series of counts. We develop Bayesian inference for the proposed class of models via MCMC and introduce a particle filtering (PF) algorithm for sequential inference. The new class of models are compared with their static counterparts using actual count series and additional insights provided by the new models are discussed.

1 | Introduction and Overview

Time series of counts arise in many areas such as business, economics, engineering, and medicine. Applications include, among others, the modeling of the number of deaths from a specific disease in a given month [1], the number of arrivals to a call center [2], the number of monthly mortgage defaults [3], time series data on crash counts in different regions [4], the number of accidents in a given time interval [5], the number of weekly shopping trips of households [6], and network flows, see Chen et al. [7]. Other applications of discrete-valued time series can be found in the volume by Davis et al. [8]. The analysis of discrete-valued time series poses both methodological and computational challenges. Davis et al. [9] provide an overview of models for time series count data and discuss methodological issues. Recent advances in Bayesian modeling of count time series are presented in Soyer and Zhang [10], and computational issues are discussed. Silva et al. [11] developed a Bayesian framework for count time series under censoring, which Silva et al. [12] recently extended to series with missing observations. In a related vein, Silva, Pereira, and McCabe [13] proposed Bayesian outlier detection methods for non-Gaussian time series, including counts.

1.1 | INAR Processes

An important class of models for time series count data is integer autoregressive (INAR) processes. They belong to the class of *observation-driven* time series models in the general framework of Cox [14]. The first order Poisson INAR process was originally introduced by McKenzie [15] and its properties and extensions were discussed in later articles by Al-Osh and Alzaid [16] and by McKenzie [17]. The recent literature on INAR and related processes is vast. For instance, Weiß [18] considers INAR(1) processes with serially dependent innovations, while Lopez et al. [19] extend the INAR processes to allow generalized Katz innovations. Monteiro et al. [20] extends INAR processes to self-exciting threshold autoregressive processes, with a Bayesian approach via Markov chain Monte Carlo (MCMC) sampling provided by Yang et al. [21].

Bayesian inference for the INAR(1) model with Poisson errors was considered by Silva et al. [22] using conjugate priors and a Gibbs sampler with a Metropolis step. Neal and Subba Rao [23] discussed Bayesian inference for integer ARMA processes with Poisson errors via MCMC. In a recent paper, Marques et al. [24] proposed a data augmentation step that allowed the implementation of a Gibbs sampler with conjugate full posterior conditionals. The authors also considered generalizations of INAR(1) by using a finite mixture model for the error process to account for potential overdispersion in time series, as well as a hierarchical extension of the Poisson error INAR(1) model for time-varying arrival rates. The latter is based on a Dirichlet process (DP) prior to the unknown distribution of dynamic arrival rates. The use of the DP prior implies exchangeability of the time-varying arrival rates and allows for clustering of homogeneous arrival rates.

1.1.1 | Contributions

In this paper, we revisit the work of Marques et al. [24] and propose a Markovian evolution for the time-varying arrival rates as an alternative to the exchangeability implied by their Dirichlet process prior. The proposed structure yields a new class of INAR processes and, by adopting a Bayesian perspective, unifies the observation-driven INAR processes with the parameter-driven state-space models of Aktekin et al. [3] for count time series; it is, to the best of our knowledge, the first state-space version of INAR processes in the sense of West and Harrison [25]. Second, the conditional conjugacy implied by the model enables full Bayesian analysis via Gibbs sampling. Third, we develop a particle filter for sequential Bayesian inference and prediction. To the best of our knowledge, this is a tool that has not been previously used for INAR processes. Finally, we extend the model and methodology to INAR(p) processes. Beyond its methodological contributions, the paper also serves as an illustration of applied Bayesian time-series analysis, with particular emphasis on probabilistic inference, sequential learning and forecasting, and dynamic model comparison.

A synopsis of our paper is as follows. In Section 2, we start with a brief introduction to the INAR(1) process with Poisson errors and present our dynamic model. In doing so, we discuss the Markov process for the time-varying arrival rates and discuss its implications on the INAR(1) model. We illustrate how the model provides a unification of the observation- and parameter-driven processes considered for time series of counts, as well as its ability to deal with overdispersed data. Section 3 develops a Bayesian analysis of the model. A Gibbs sampler using data augmentation and a particle filter for sequential analysis are presented. An extension to the pth order dynamic INAR process is discussed in Section 4 and Bayesian inference using Gibbs sampler and particle filtering is developed. We illustrate an implementation of the proposed model using actual data in Section 5. We discuss what types of additional insight can be obtained from the model and the Bayesian analysis. We conclude with some final remarks in Section 6.

2 | The Dynamic INAR(1) Model

2.1 | A Brief Review of the INAR(1) Model

For a stationary time-series of counts Y_t, the INAR(1) model is defined as

$$Y_t = \alpha \circ Y_{t-1} + \epsilon_t \tag{1}$$

where "$\circ$" is the binomial thinning operation represented by

$$\alpha \circ Y_{t-1} = \sum_{j=1}^{Y_{t-1}} B_{jt} \tag{2}$$

where B_{jt} are Bernoulli random variables with probability α. If we define $M_t = \alpha \circ Y_{t-1}$ in (2), then conditionally on Y_{t-1} and α, M_t it has a binomial distribution with parameters α and Y_{t-1}. In the above setup, it is important to note that the B_{jt}'s are independent of the ϵ_t's and that the thinning is performed at each time point t independently of the thinning at other periods. As pointed out by Kedem and Fokianos ([26], chapter 5), the INAR(1) process can be interpreted as a special case of branching processes with immigration where Y_t is the population at time t which consists of two components: M_t, those who survive from time $(t-1)$ with probability α and ϵ_t, those who arrive at the beginning of time t.

If $\{\epsilon_t\}$ is a sequence of i.i.d. Poisson random variables with rate θ in (1), then the model is referred to as an INAR(1) process with Poisson errors. Alternatively, the error distribution can be assumed as geometric or negative binomial as discussed in Weiß [27] or one can consider a mixture of Poisson and geometric as in Marques et al. [24]. In our development, we consider the INAR processes with Poisson errors.

If Y_0 is assumed to be Poisson with rate $\theta/(1-\alpha)$, then it can be shown that Y_t is a stationary Poisson series with parameter $\theta/(1-\alpha)$. In addition, the autocorrelation function of INAR(1) processes is given by $\rho_Y(k) = \alpha^k$ for $k > 0$; see McKenzie [17]. In fact, the behavior of the autocorrelation function is not limited to the Poisson errors, but it holds for all INAR(1) models.

Given Y_{t-1}, the conditional distribution of Y_t is obtained as a convolution of a binomial and a Poisson given by

$$p(Y_t|Y_{t-1}, \theta, \alpha) = \sum_{j=0}^{min(Y_{t-1}, Y_t)} \frac{e^{-\theta}\theta^{Y_t - j}}{(Y_t - j)!} \binom{Y_{t-1}}{j} \alpha^j (1-\alpha)^{Y_{t-1}-j}, \tag{3}$$

with $E[Y_t|Y_{t-1}, \theta, \alpha] = \alpha Y_{t-1} + \theta$. The forecast distribution for h-steps ahead, $p(Y_{t+h}|Y_t, \theta, \alpha)$, can be obtained as well as the conditional mean

$$E[Y_{t+h}|Y_t, \theta, \alpha] = \alpha^h \left(Y_t - \frac{\theta}{(1-\alpha)} \right) + \frac{\theta}{(1-\alpha)}. \tag{4}$$

It is important to note that the forecast distributions h-steps ahead are not available in closed form with more general error distributions, in which case numerical computations are necessary (see Weiß, [28], section 2.6).

2.2 | INAR(1) Model with Dynamic Arrival Rates

An earlier state-dependent parameter generalization of INAR models was considered in Thyregod et al. [29] who proposed a self-exciting threshold-based (SET) INAR process for analysis of tipping bucket rainfall measurements. Monteiro et al. [20] showed the stationarity of the process and investigated asymptotic properties of the maximum likelihood estimators. As previously mentioned Yang et al. [21] developed Bayesian analysis for the SET-INAR processes using MCMC methods and Lopez et al.

[19] considered an extension of SET-INAR processes using generalized Katz innovations. For modeling state-dependent parameters, the SET-INAR approach is considered by many as the standard in count time-series community.

As an alternative to the existing approaches in the literature, we follow the standard approach in state space modeling for count time-series [25] and consider a generalization of the INAR(1) process with Poisson errors in (1) by allowing the Poisson rates of ϵ_t's to be time-varying, that is, θ_t. In this case, the observation model can be written as

$$Y_t - M_t | M_t, \theta_t \sim Poisson(\theta_t)\mathcal{I}(Y_t \geq M_t), \tag{5}$$

which is a shifted Poisson for Y_t with rate θ_t. We also have

$$M_t | Y_{t-1}, \alpha \sim Bin(Y_{t-1}, \alpha).$$

We assume that θ_t has a Markov evolution characterized by

$$\theta_t | \theta_{t-1}, M^{(t-1)}, Y^{(t-1)} \sim Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma), \tag{6}$$

which is a scaled beta distribution on $(0, \theta_{t-1}/\gamma)$ and can be rewritten as a system equation

$$\theta_t = \frac{\theta_{t-1}}{\gamma} w_t \tag{7}$$

where $w_t | Y^{(t-1)}, M^{(t-1)} \sim Beta[\gamma a_{t-1}, (1-\gamma)a_{t-1}]$.

The quantity $0 < \gamma < 1$ is a discount factor in the sense of West and Harrison [25], $Y^{(t)} = (Y_t, Y^{(t-1)})$ and $M^{(t)} = (M_t, M^{(t-1)})$. In the system Equation (7), w_t can be considered as a system error term with mean γ and (7) provides a random walk-type evolution for θ_t. The Markov evolution (7) was originally introduced by Smith and Miller [30] who analyzed exponential observation models; although their approach was not fully Bayesian, Harvey and Fernandez [31] showed that the same structure can be used for Poisson arrival rates; More recently, (7) was used in Gamerman et al. [32] for parameter driven non-Gaussian time series.

In our proposed dynamic INAR(1) process, Equations (5) and (7) will play the roles of the observation equation and the system equation, respectively. We complete the definition of our model by assuming that, at time $t-1$, the rate θ_{t-1} follows a gamma distribution

$$\theta_{t-1} | M^{(t-1)}, Y^{(t-1)} \sim Gamma(a_{t-1}, b_{t-1}). \tag{8}$$

It follows from (6) and (8) that the forecast distribution of θ_t at $t-1$ is given by

$$\theta_t | M^{(t-1)}, Y^{(t-1)} \sim Gamma(\gamma a_{t-1}, \gamma b_{t-1}). \tag{9}$$

Using the Bayes' rule

$$p(\theta_t | M^{(t)}, Y^{(t)}) \propto p(\theta_t | M^{(t-1)}, Y^{(t-1)}) p(Y_t - M_t | M_t, \theta_t),$$

it can be shown that the posterior (or filtering) distribution of θ_t can be obtained as

$$\theta_t | M^{(t)}, Y^{(t)} \sim Gamma(a_t, b_t), \tag{10}$$

for $a_t = \gamma a_{t-1} + (Y_t - M_t)$ and $b_t = \gamma b_{t-1} + 1$.

We note that starting at time 0 with prior $\theta_0 \sim Gamma(a_0, b_0)$, the proposed model provides us with a conjugate update of the dynamic rates which is attractive in developing posterior and predictive Bayesian inferences, as will be discussed in Section 3. Other attractive features of the model are its ability to deal with overdispersed time series of counts and the inclusion of some of the previously considered models as special cases. We will elaborate on this in Section 2.3.

2.2.1 | INAR(1) with Dynamic Thinning and Covariates

An alternative dynamic INAR(1) model can be considered by assuming a dynamic structure for the thinning probabilities in (1). More specifically, we can write a dynamic version of (1) as

$$Y_t = \alpha_t \circ Y_{t-1} + \epsilon_t$$

where ϵ_t's are Poisson error terms with static rate θ and the binomial thinning is now given by

$$M_t \equiv \alpha_t \circ Y_{t-1} = \sum_{j=1}^{Y_{t-1}} B_{jt}$$

where B_{jt}'s are conditionally independent Bernoulli random variables with probability α_t. Time evolution of α_t' can be described by using a logit type transformation and a Markovian structure such as

$$\phi_t = \log\left(\frac{\alpha_t}{1-\alpha_t}\right) = \phi_{t-1} + w_t, \tag{11}$$

where w_t's are independent zero mean Gaussian innovation terms with variance σ_w^2. Bayesian analysis of the model using a Gibbs sampler will require use of some data augmentation type approach within the Gibbs as in Polson et al. [33]. This is currently under investigation.

It is also possible to incorporate covariates into the dynamic INAR(1) models. For example, we can define λ_t as the dynamic arrival rate of the error term ϵ_t in (5) via

$$\lambda_t = \theta_t \exp\left\{\beta' z_t\right\},$$

where z_t is a vector of covariates with unknown parameters β and θ_t is the baseline arrival rate following a Markov evolution as in (6). As noted by Aktekin et al. [3], Bayesian analysis of the model will require use of a Metropolis step within the Gibbs sampler discussed in Section 3.1.

2.3 | Other Properties of the Dynamic INAR(1) Model

The Markov evolution given by (6) enables us to deal with overdispersed data. Using the forecast (or prior) distribution of θ_t given by (9), we can obtain the predictive distribution of $\epsilon_t = Y_t - M_t$ via

$$p(\epsilon_t | \gamma, M^{(t)}, Y^{(t-1)}) = \int_0^\infty p(\epsilon_t | \theta_t, M_t) p(\theta_t | \gamma, M^{(t)}, Y^{(t-1)}) d\theta_t,$$

which can be reduced to

$$p(\epsilon_t|\gamma, M^{(t)}, Y^{(t-1)}) = \frac{\Gamma(\gamma a_{t-1} + \epsilon_t)}{\Gamma(\epsilon_t + 1)\Gamma(\gamma a_{t-1})} \left(1 - \frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\epsilon_t} \left(\frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\gamma a_{t-1}}, \tag{12}$$

which is a negative binomial for integer values of γa_{t-1}. As a result, the Markov structure provides with a negative binomial type dynamic error distribution which can handle over-dispersed counts. As will be discussed in Section 3, a Bayesian analysis using this model can also be developed.

2.3.1 | Unifying the Count Time Series Models

The dynamic INAR(1) model has two static parameters: the thinning probability α and the discount factor γ. These two parameters play an important role in assessing the suitability of different classes of models for analyzing time series count data. We note that as $\alpha \to 0$ in the dynamic INAR(1) model, the autocorrelation of Y_t's disappears and in the limit $\alpha = 0$ the model reduces to a parameter-driven model as in Harvey and Fernandes [31] and Aktekin and Soyer [2] where, given θ_t's, Y_t's are conditionally independent Poisson counts. On the other hand, as we can see from the Markov model for θ_t and the system Equation (7), the arrival rates become smoother over time as $\gamma \to 1$. In the limit $\gamma = 1$, $\theta_t = \theta$ for all t and our model reduces to the static INAR(1) process of Equation (1). In other words, these two cases can be obtained in our dynamic INAR(1) by putting a degenerate prior for α at 0 or for γ at 1. The trivial case of a Poisson time series of white noise can be obtained by placing a joint degenerate prior for (α, γ) at $(0, 1)$. Therefore, the proposed dynamic INAR(1) process provides a unification of these models.

3 | Monte Carlo-Based Posterior Inference

In this section, we propose an MCMC scheme for joint posterior inference of all unknowns of the model. We also derive a particle filter for online inference. In our development, we assume independent priors for model parameters α and γ. Specifically, we use a beta prior for α as $\alpha \sim Beta(a_\alpha, b_\alpha)$ and denote the prior for the discount parameter by $p(\gamma)$. Furthermore, α and γ are a priori assumed to be independent of the initial arrival rate θ_0.

3.1 | Gibbs Sampler

We can develop Bayesian inference using a Gibbs sampler given counts $Y^{(t)}$ from T time periods. In doing so, we need the posterior full conditionals of α, γ, $\theta^{(T)} = (\theta_1, \dots, \theta_T)$, and $M^{(T)} = (M_1, \dots, M_T)$. Among these, the trickiest is the full conditional distribution of $M^{(T)}$, $p(M^{(T)}|\alpha, \gamma, \theta^{(T)}, Y^{(T)})$.

3.1.1 | Learning $M^{(T)}$

Given α and γ, suppressed below, assuming $M_1 = 0$, we can write the joint distribution $p(\theta_1, Y_1, \theta_2, Y_2, M_2, \dots, \theta_T, Y_T, M_T)$ as

$$p(\theta_1|D_0)p(Y_1|\theta_1)\prod_{t=2}^{T} p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}) p(M_t|Y_{t-1})p(Y_t|M_t, \theta_t), \tag{13}$$

so the full conditional for each M_t can be obtained as

$$p(M_t|M^{(-t)}, \theta^{(T)}, Y^{(T)}) \propto \left\{ \prod_{s=t+1}^{T} p(\theta_s|\theta_{s-1}, M^{(s-1)}, Y^{(s-1)}) \right\} \times p(M_t|Y_{t-1})p(Y_t|M_t, \theta_t), \tag{14}$$

for $t = 2, \dots, T-1$, where $M^{(-t)} = \{M_s; s \neq t\}$ and D_0 denotes the prior information. It is important to note that in the above product

$$(\theta_s|\theta_{s-1}, M^{(s-1)}, Y^{(s-1)}) \sim Beta[\gamma a_{s-1}, (1-\gamma)a_{s-1}]\mathcal{I}(\theta_s < \theta_{s-1}/\gamma). \tag{15}$$

This suggests that for $s = (t+1), (t+2), \dots, T$, the a_{s-1} values should be evaluated at the given value of M_t. For example, when $s = (t+1)$, we have $a_t = \gamma a_{t-1} + (Y_t - M_t)$. Therefore, for each value of M_t, we need to evaluate $a_t, a_{t+1}, \dots, a_T$ and these values will be different than the values in updating of θ_t's. For $t = T$, we can write

$$p(M_T|M^{(-T)}, \theta^{(T)}, Y^{(T)}) \propto p(M_T|Y_{T-1})p(Y_T|M_T, \theta_T).$$

3.1.2 | Learning α

The full conditional of α can be obtained as

$$\alpha|M^{(T)}, Y^{(T)} \sim Beta\left(a_\alpha + \sum_{t=2}^{T} M_t, b_\alpha + \sum_{t=2}^{T} (Y_{t-1} - M_t)\right). \tag{16}$$

3.1.3 | Learning $\theta^{(T)}$

In order to implement a Gibbs sampler, the full conditional of the vector $\theta^{(T)}$, $p(\theta_1, \dots, \theta_T|M^{(T)}, \gamma, Y^{(T)})$, can be obtained by using the forward filtering backward sampling (FFBS) algorithm of Frühwirth-Schnatter [34]. We can write the joint full conditional as

$$p(\theta_T|M^{(T)}, Y^{(T)}, \gamma)p(\theta_{T-1}|\theta_T, M^{(T-1)}, Y^{(T-1)}, \gamma) \cdots p(\theta_1|\theta_2, Y_1, \gamma), \tag{17}$$

where $p(\theta_{t-1}|\theta_t, M^{(t-1)}, Y^{(t-1)}, \gamma)$ is a shifted-gamma density over $(\gamma\theta_t, \infty)$, denoted as

$$Gamma((1-\gamma)a_{t-1}, b_{t-1})\mathcal{I}(\theta_{t-1} > \gamma\theta_t).$$

3.1.4 | Learning γ

Note that to obtain the full conditional of γ we can take the joint distribution in (13) and note that $p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$ is conditional on γ. Thus, the full conditional of γ can be obtained as

$$p(\gamma|\theta^{(T)}, M^{(T)}, Y^{(T)}) \propto \left\{ \prod_{t=2}^{T} p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}, \gamma) \right\} p(\gamma), \tag{18}$$

ALGORITHM 1 | Gibbs sampler algorithm.

Initial values $(\theta^{(T)}, M^{(T)}, \alpha, \gamma)^{(0)}$ and repeat the following 4 steps iteratively.
for $g = 1 \dots G$ **do**
    Draw $\alpha^g \sim Beta\left(a_\alpha + \sum_{t=2}^{T} M_t, b_\alpha + \sum_{t=2}^{T}(Y_{t-1} - M_t)\right)$.
    Draw $\theta_T^g \sim Gamma(a_T, b_T)$
    for $t = T-1 \dots 1$ **do**

$$\text{Draw } \theta_t^g \sim Gamma((1-\gamma)a_t, b_t)\mathcal{I}(\theta_t > \gamma\theta_{t+1}^g)$$

    end for
    for $t = 2 \dots T-1$ **do**
    Draw $M_t^g \in \{0, 1, \dots, min(Y_t, Y_{t-1})\}$ from the discrete distribution proportional to

$$\left\{\prod_{s=t+1}^{T} p(\theta_s^g|\theta_{s-1}^g, M^{(s-1)}Y^{(s-1)})\right\} p(M_t|Y_{t-1})p(Y_t|M_t, \theta_t^g).$$

    end for
end for

where $p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}, \gamma)$ is the scaled beta distribution given by (6). We can use a Metropolis step or use a discrete prior to drawing samples from the full conditional of γ.

Alternatively, we can learn about γ outside the Gibbs sampler by obtaining its marginal likelihood. The marginal likelihood for the discount parameter γ, conditional on $M^{(T)}$, is

$$p(\epsilon_1, \dots, \epsilon_T|M^{(T)}, Y^{(T)}, \gamma) = \prod_{s=1}^{T} p(\epsilon_s|M^{(s)}, Y^{(s-1)}, \gamma), \qquad (19)$$

where $p(\epsilon_s|M^{(s)}, Y^{(s-1)}, \gamma)$ is given by the predictive distribution (12). This can also be used to obtain the posterior distribution of γ, that is,

$$p(\gamma|Y^{(T)}) \propto p(\epsilon_1, \dots, \epsilon_T|\gamma)p(\gamma),$$

where $p(\epsilon_1, \dots, \epsilon_T|\gamma)$ can be obtained using samples generated from the distribution of $p(M^{(T)}|\gamma, Y^{(T)})$ through the Gibbs sampler. One can also use a discrete prior in $p(\gamma)$ to evaluate the posterior distribution of γ. In Algorithm 1, we show the schematic representation of the Gibbs sampler for our proposed dynamic INAR(1) model.

3.2 | Sequential Monte Carlo and Particle Filtering

It is well known that MCMC methods are not computationally efficient for sequential Bayesian learning and forecasting, as they require rerunning of the Markov chains with each additional observation to obtain the posterior distributions of time-varying parameters, such as Poisson rates $\theta_1, \dots, \theta_T$ in our case. Sequential Monte Carlo (SMC) methods have been proposed to alleviate such computational inefficiencies. An important class of SMC methods are particle filters (PF), that was originally proposed by Gordon et al. [35]; see Lopes and Tsay [36] for a review of PF and Singpurwalla et al. [37] for a historical perspective.

As pointed out by Carvalho et al. [38], and extensively discussed by Lopes et al. [39], learning about static parameters in PF is not trivial, but if the conditional posterior distributions of static parameters are available with known (conditional) sufficient statistics, one can develop an efficient recursive updating scheme which they referred to as *particle learning* (PL). We can develop a PL algorithm in the sense of Lopes et al. [39] for the dynamic INAR(1) model.

Assume that in the time period $t-1$, given the data $Y^{(t-1)}$, we have particles $\{\theta_{t-1}^i, M_{t-1}^i, \alpha^i\}$, for $i = 1, \dots, N$. Note that given this set of particles, we can obtain a sample from the prior $p(M_t|Y^{(t-1)})$ by sampling M_t^i from the binomial distribution $Bin(Y_{t-1}, \alpha^i)$, for $i = 1, \dots, N$. Our objective is to resample θ_{t-1} based on the observed value in time t, Y_t. For this, we need resampling weights that are proportional to the predictive $p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$. We can obtain the distribution as follows.

$$\int p(Y_t|M_t, \theta_{t-1}, \theta_t, M^{(t-1)}, Y^{(t-1)})p(\theta_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})d\theta_t,$$

which reduces to

$$p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)}) = \int p(Y_t|M_t, \theta_t)p(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)})d\theta_t$$

where $(Y_t|M_t, \theta_t) \sim Poi(\theta_t)\mathcal{I}(Y_t \geq M_t)$ and $(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$ is a scaled beta distribution denoted as

$$(\theta_t|\theta_{t-1}, M^{(t-1)}, Y^{(t-1)}) \sim Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma).$$

It can be shown that $p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$ is a confluent hypergeometric negative binomial distribution that we can evaluate numerically. Specifically, we have $p(Y_t|M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$ is given by

$$\int_0^{\theta_{t-1}/\gamma} \frac{\theta_t^{(Y_t - M_t)} e^{-\theta_t}}{(Y_t - M_t)!} \xi(\gamma, a_{t-1})(\gamma/\theta_{t-1})^{a_{t-1}-1}\theta_t^{\gamma a_{t-1}-1} \times \left(\frac{\theta_{t-1}}{\gamma} - \theta_t\right)^{(1-\gamma)a_{t-1}-1} d\theta_t,$$

where $\xi(\gamma, a_{t-1}) = Beta\big(\gamma a_{t-1}, (1-\gamma)a_{t-1}\big)$. Recall that the Beta function is defined as $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

Also, by using change of variable $u_t = (\gamma/\theta_{t-1})\theta_t$, we can write the above integral as

$$\kappa(Y_t, M_t, \theta_{t-1}, \gamma, a_{t-1}) \int_0^1 e^{-\frac{\theta_{t-1}}{\gamma}u_t} u_t^{\gamma a_{t-1} + (Y_t - M_t) - 1} \times \left(1 - u_t\right)^{(1-\gamma)a_{t-1}-1} du_t,$$

where

$$\kappa(Y_t, M_t, \theta_{t-1}, \gamma, a_{t-1}) = \frac{(\theta_{t-1}/\gamma)^{(Y_t - M_t)}}{(Y_t - M_t)!}\xi(\gamma, a_{t-1}),$$

with the integral in the above expression represented as

$$Beta(Y_t - M_t + \gamma a_{t-1}, (1-\gamma)a_{t-1})CHF(a, a+b, -c),$$

where $CHF(a, a+b, -c)$ is the confluent hyper-geometric function of Abramowitz and Stegun [40] with $a = (Y_t - M_t) + \gamma a_{t-1}$, $b = (1-\gamma)a_{t-1}$ and $c = \theta_{t-1}/\gamma$. To evaluate the CHF function, we can use the R package `gsl` of Hankin [41]. Thus, the weights $p(Y_t | M_t^i, \theta_{t-1}^i, M^{(t-1)}, Y^{(t-1)})$ are proportional to

$$\begin{aligned} \omega^i \propto\ & \kappa(Y_t, M_t^i, \theta_{t-1}^i, \gamma, a_{t-1}) \\ & \times Beta(Y_t - M_t^i + \gamma a_{t-1}, (1-\gamma)a_{t-1}) \\ & \times CHF((Y_t - M_t^i) + \gamma a_{t-1}, (Y_t - M_t^i) + a_{t-1}, -\theta_{t-1}^i/\gamma), \end{aligned} \quad (20)$$

which are used to obtain the resampled values of $(M_t^{k(i)}, \theta_{t-1}^{k(i)})$. Note that we have not resampled α values, that is, their samples are based on the data from $(t-1)$, since these are not needed to resample the values of $(M_t^{k(i)}, \theta_{t-1}^{k(i)})$. However, given $M_t^{k(i)}$, we can have resampling values of α by updating enough statistics of the beta distribution.

Next, we propogate to θ_t using $(M_t^{k(i)}, \theta_{t-1}^{k(i)})$, and for this we need the density $p(\theta_t | \theta_{t-1}, M_t, M^{(t-1)}, Y^{(t)})$,

$$p(\theta_t | \theta_{t-1}, M_t, Y^{(t)}, M^{(t-1)}) \propto p(Y_t | M_t, \theta_t) p(\theta_t | M_t, \theta_{t-1}, M^{(t-1)}, Y^{(t-1)})$$

which can be written as

$$\begin{aligned} p(\theta_t | \theta_{t-1}, M_t, Y^{(t)}, M^{(t-1)}) &\propto e^{-\theta_t} \\ \times \left(\frac{\theta_{t-1}}{\gamma} - \theta_t \right)^{(1-\gamma)a_{t-1}-1} & \theta^{\gamma a_{t-1} + (Y_t - M_t) - 1}. \end{aligned} \quad (21)$$

The above is proportional to a scaled hypergeometric beta density as discussed by Gordy [42]. We can sample from (21) using Metropolis-Hastings or some form of rejection sampling.

Alternatively, in propogating θ_t, if generating from the hypergeometric beta density in (21) is not computationally efficient, we can use a sequential importance sampling type step as suggested in Aktekin et al. [6]. More specifically, θ_t can be (blind) propagated via

$$p(\theta_t | \theta_{t-1}, M^{(t-1)}, Y^{(t-1)}),$$

which is a scaled Beta distribution, $Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma)$. The particles are then resampled with weights

$$p(Y_t | \theta_t, M_t),$$

which is a shifted-$Poisson(\theta_t)$, with $Y_t \geq M_t$. Finally, we can update γ offline or alternatively using a discrete prior with the marginal likelihood given by (19). Algorithm 2 provides a schematic representation of this particle filter algorithm.

3.2.1 | Collapsed Particle Filtering

An alternative to particle filtering that avoids the evaluation of CHF is to integrate out θ_t and using the predictive distribution of $\epsilon_t = Y_t - M_t$ given $M^{(t-1)}$, $Y^{(t-1)}$ for resampling. we refer to this approach as "collapsed particle filtering."

Assume that in the time period $t-1$ given $Y^{(t-1)}$, we have particles of (M_{t-1}^i, α^i) $i = 1, \ldots, N$. Given these particles, we can obtain particles from M_t^i using the binomial distribution $Bin(Y_{t-1}, \alpha^i)$ for $i = 1, \ldots, N$. These can be considered as "prior" particles for M_t and at time t we need to update these given Y_t. In so doing, we can resample M_t's using the predictive distribution

$$p(\epsilon_t | \gamma, M^{(t-1)}, Y^{(t-1)}) = \frac{\Gamma(\gamma a_{t-1} + \epsilon_t)}{\Gamma(\epsilon_t + 1)\Gamma(\gamma a_{t-1})} \left(1 - \frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\epsilon_t} \left(\frac{\gamma b_{t-1}}{\gamma b_{t-1} + 1}\right)^{\gamma a_{t-1}}.$$

In other words, we use the above predictive distribution as the resampling weight to obtain the posterior particles $M_t^{k(i)}$ and update α using the particles $M_t^{k(i)}$ in the beta distribution. Note that once we have the posterior particles of we can update the sufficient statistics in the posterior distribution of $(\theta_t | \gamma, M_t, M^{(t-1)}, Y^{(t-1)}) \sim Gamma(a_t, b_t)$ where $a_t = \gamma a_{t-1} + (Y_t - M_t^{k(i)})$ and $b_t = \gamma b_{t-1} + 1$. In Algorithm 3, we present a schematic representation of the collapsed particle filtering algorithm.

ALGORITHM 2 | Particle filter algorithm.

1: Let $\{(\theta_{t-1}, M_{t-1}, \alpha)^i\}_{i=1}^N$ be the particle set at time $t-1$.
2: Draw $M_t^i \sim Binomial(Y_{t-1}, \alpha^i)$.
3: Compute resampling weights from Equation (18)
4: Resampling particles $\{(\theta_{t-1}, M_t)^{k(i)}\}_{i=1}^N$ with weights from 2.
5: Update (a_t, b_t):$a_t^i = \gamma a_{t-1}^{k(i)} + (Y_t - M_t^{k(i)})$ and $b_t^i = \gamma b_{t-1}^{k(i)} + 1$.
6: Draw thinning: $\alpha | M^t, Y^t \sim Beta(S_{1t}^i, S_{2t}^i)$, where$S_{1t}^i = S_{1,t-1}^{k(i)} + M_t^{k(i)}$ and$S_{2t}^i = S_{2,t-1}^{k(i)} + (Y_{t-1} - M_t^{k(i)})$, where $S_{10} = a_\alpha$ and $S_{20} = b_\alpha$.
7: Propagate θ_t from$\theta_t \sim Beta(\gamma a_{t-1}, (1-\gamma)a_{t-1}, \theta_{t-1}/\gamma)$, a scaled Beta distribution.
8: Resample with weights$p(Y_t | \theta_t, M_t)$, a shifted $Poisson(\theta_t)$, with $Y_t \geq M_t$.

4 | Dynamic INAR(p) Processes

Consider the p-th order INAR setup presented in Du and Li [43] as

$$Y_t = \alpha_1 \circ Y_{t-1} + \alpha_2 \circ Y_{t-2} + \ldots + \alpha_p \circ Y_{t-p} + \epsilon_t,$$

where

$$\alpha_j \circ Y_{t-j} = \sum_{i=1}^{Y_{t-j}} B_{ijt}$$

and B_{ijt}'s are Bernoulli random variables with success probability α_j, $j = 1, 2$. For stationarity of the INAR(p) model we need $0 < \sum_{j=1}^p \alpha_j < 1$. Given Y_{t-j}, $\alpha_j \circ Y_{j-1}$ has a binomial distribution with parameters α_j and Y_{t-j}.

In the above setup it is assumed that $\{\epsilon_t\}$ is a sequence of independent Poisson random variables with parameter θ_t and B_{ijt}'s are independent of ϵ_t's. It is important to note that thinning is performed at each time point t and is independent of thinning at other periods. We define latent variables $M_{jt} = \alpha_j \circ Y_{t-j}$ where

$$M_{jt} | Y_{t-j}, \alpha_j \sim Bin(Y_{t-j}, \alpha_j)$$

ALGORITHM 3 | Collapsed particle filter algorithm.

1: Let $\{(M_{t-1},\alpha)^i\}_{i=1}^N$ be the particle set at time $t-1$.
2: Draw $M_t^i \sim Binomial(Y_{t-1},\alpha^i)$.
3: Compute resampling weights from equation (12):

$$\omega^i = \frac{\Gamma(\gamma a_{t-1}+\epsilon_t^i)}{\Gamma(\epsilon_t^i+1)\Gamma(\gamma a_{t-1})}\left(1-\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\epsilon_t^i}\left(\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\gamma a_{t-1}},$$

where $\epsilon_t^i = Y_t - M_t^i$.
4: $M_t^{k(i)}$ are the resample draws.
5: Update sufficient statistics and sample thinning probability α^i

$$S_{1t}^i = S_{1,t-1}^{k(i)} + M_t^{k(i)}$$
$$S_{2t}^i = S_{2,t-1}^{k(i)} + (Y_{t-1} - M_t^{k(i)})$$
$$\alpha^i \sim Beta(S_{1t}^i, S_{2t}^i),$$

where $S_{10}=a_\alpha$ and $S_{20}=b_\alpha$.
6: Update a_t and b_t and sample θ_t from $(\theta_t|\gamma, M^{(t)}, Y^{(t)})$

$$a_t^i = \gamma a_{t-1}^{k(i)} + (Y_t - M_t^{k(i)})$$
$$b_t^i = \gamma b_{t-1}^{k(i)} + 1$$
$$\theta_t^i \sim Gamma(a_t^i, b_t^i).$$

and $Y_t|M_{1t},\ldots,M_{pt},\theta_t$ is a shifted Poisson, that is,

$$p(Y_t|M_{1t},M_{2t},\ldots,M_{pt},\theta_t) = \frac{e^{-\theta_t}\theta_t^{Y_t-\sum_{j=1}^p M_{jt}}}{(Y_t-\sum_{j=1}^p M_{jt})!}\mathcal{I}(Y_t \geq \sum_{j=1}^p M_{jt}).$$

Note that based on the above $0 \leq M_{jt} \leq min(Y_t - \sum_{i\neq j} M_{it}, Y_{t-j})$ and if $(Y_t - \sum_{i\neq j} M_{it}) = 0$ or $Y_{t-j}=0$ then $M_{jt}=0$ in the model. Also, if $Y_t = 0 \Rightarrow M_{1t}=M_{2t}=\cdots=M_{pt}=0$.

Using the Markov evolution of θ_t given by (6), we can develop Bayesian inference for the dynamic INAR(p) model. In this case, we define $M_t=(M_{1t},\ldots,M_{pt})$ as a vector. It can be shown that the filtering distribution of θ_t given by (9) will have parameters $a_t=\gamma a_{t-1}+(Y_t-\sum_{j=1}^p M_{jt})$ and $b_t=\gamma b_{t-1}+1$. Also, for stationarity of the model we assume a Dirichlet prior for vector $\alpha = (\alpha_1,\ldots,\alpha_p)$ as

$$p(\alpha) \propto \prod_{j=1}^{p+1}\alpha_j^{a_{\alpha_j}-1}, \tag{22}$$

where $a_{\alpha_j}>0$, for $j=1,\ldots p+1$, and $\alpha_{p+1}=1-\sum_{j=1}^p\alpha_j$.

4.1 | Gibss Sampler

As in Section 3.1, given the above structure, we can obtain the full conditionals of all the quantities in the dynamic INAR(p) model. For the maturation terms, M_{jt}'s we can show that

$$p(M_{jt}|M^{(-jt)},\theta^{(T)},Y^{(T)}) \propto \left\{\prod_{s=t+1}^T p(\theta_s|\theta_{s-1},M^{(s-1)},Y^{(s-1)})\right\}$$
$$p(M_{jt}|Y_{t-1})p(Y_t|M_t,\theta_t)$$

where $M_{jt}\in\{0,1,\ldots,min(Y_t-\sum_{i\neq j}M_{it},Y_{t-j})\}$ for $t=(p+1),\ldots,T-1$ and $j=1,\ldots,p$. For $t=T$ and $j=1,\ldots,p$

$$p(M_{jT}|M^{(-jT)},\theta^{(T)},Y^{(T)})) \propto p(M_{jT}|Y_{T-j})p(Y_T|M_T,\theta_T).$$

For components of α, using (20), we have conditional priors of α_j's as scaled beta densities

$$p(\alpha_j|\alpha^{(-j)}) \propto \alpha_j^{a_{\alpha_j-1}}(1-\sum_{j=1}\alpha_j)^{a_{\alpha_{p+1}-1}},$$

where $\alpha_j\in(0,1-\sum_{i\neq j}\alpha_i)$. It can be shown that the posterior full conditionals of α_j's can be obtained as scaled beta densities as

$$\alpha_j|\alpha^{(-j)},M^{(T)},Y^{(T)} \sim Beta\left(a_{\alpha_j}+\sum_{t=p+1}^T M_{jt}, a_{\alpha_{p+1}} + \sum_{t=p+1}^T (Y_{t-j}-M_{jt})\right)\mathcal{I}(\alpha_j<1-\sum_{i\neq j}\alpha_i). \tag{23}$$

Learning of $\theta^{(T)}$ and the discount parameter γ follow along the same lines as discussed in Section 3.1.

4.2 | Particle Filtering

The sequential Monte Carlo approach of Section 3.2 can be modified for the INAR(p) extension. Specifically, in Algorithm 2, in step 1, we need to draw $M_{jt}^i \sim Binomial(Y_{t-j},\alpha_j^i)$ for $j=1.2,\ldots,p$. The resampling weights of Equation (18) are now computed by replacing (Y_t-M_t) with $(Y_t-\sum_{j=1}^p M_{jt})$. This also applies to the resampling particles of step 3 and to the updatings of (a_t,b_t) in Step 4. Propogation of θ_t follows the step 6 of the Algorithm 2, and finally θ_t's can be resampled using weights based on a shifted Poisson with $Y_t \geq \sum_{j=1}^p M_{jt})$. However, since there are no sufficient statistics for $(\alpha_1,\ldots,\alpha_p)$, in step 5, a mixture-based approach as in Liu and West [44] can be used to draw the thinning probabilities.

Alternatively, by not enforcing stationarity, independent beta priors can be specified for $(\alpha_1,\ldots,\alpha_p)$ and sufficient statistics can be updated in a similar manner as in step 5 of Algorithm 3. In what follows, we present a collapsed PF by assuming that α_j's have independent beta priors with parameters $a_{\alpha_j}, b_{\alpha_j}$ for $j=1,\ldots,p$.

Without loss of generality, we consider the case of $p=2$ where we have the following collapsed PF algorithm:

Let $\{(M_{1,t-1},M_{2,t-1},\alpha_1,\alpha_2)^i\}_{i=1}^N$ be the particle set at time $t-1$.

1. Draw $M_{1t}^i \sim Binomial(Y_{t-1},\alpha_1^i)$ and $M_{2t}^i \sim Binomial(Y_{t-2},\alpha_2^i)$
2. Compute resampling weights from Equation (10):

$$\omega^i = \frac{\Gamma(\gamma a_{t-1}+\epsilon_t^i)}{\Gamma(\epsilon_t^i+1)\Gamma(\gamma a_{t-1})}\left(1-\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\epsilon_t^i}\left(\frac{\gamma b_{t-1}}{\gamma b_{t-1}+1}\right)^{\gamma a_{t-1}},$$

where $\epsilon_t^i = Y_t - M_{1t}^i - M_{2t}^i$. We use the above predictive distribution as the resampling weight to obtain the posterior particles $M_{1t}^{k(i)}, M_{2t}^{k(i)}$ and update α_1,α_2 using these particles in the beta distribution

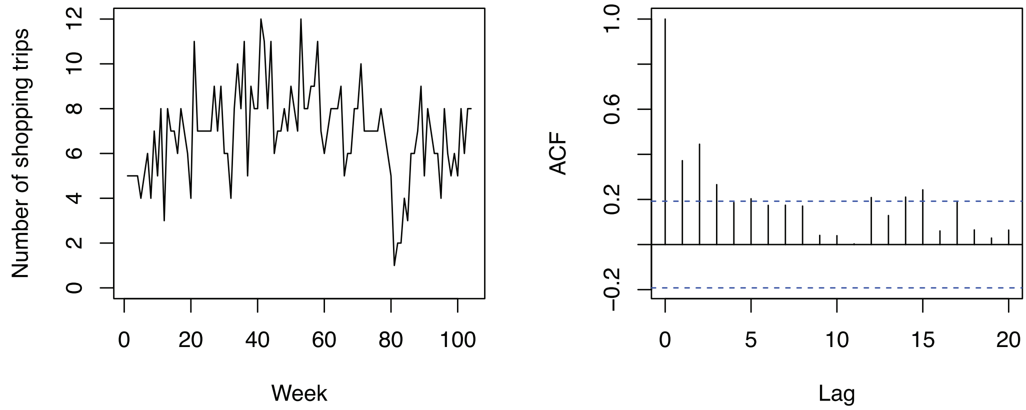

FIGURE 1 | Number of weekly shopping trips time series and sample ACF.

3. Update sufficient statistics and sample thinning probability α^i

$$S_{11t}^{i} = S_{11,t-1}^{k(i)} + M_{1t}^{k(i)}$$
$$S_{12t}^{i} = S_{12,t-1}^{k(i)} + (Y_{t-1} - M_{1t}^{k(i)})$$
$$\alpha_1^i \sim Beta(S_{11t}^{i}, S_{12t}^{i}),$$

and similarly

$$S_{21t}^{i} = S_{21,t-1}^{k(i)} + M_{2t}^{k(i)}$$
$$S_{22t}^{i} = S_{22,t-1}^{k(i)} + (Y_{t-2} - M_{2t}^{k(i)})$$
$$\alpha_2^i \sim Beta(S_{21t}^{i}, S_{22t}^{i}),$$

where $S_{110} = a_{\alpha_1}, S_{120} = b_{\alpha_1}$ and $S_{210} = a_{\alpha_2}, S_{220} = b_{\alpha_2}$.

4. Update a_t and b_t and sample θ_t from $(\theta_t|\gamma, M^{(t)}, Y^{(t)})$

$$a_t^i = \gamma a_{t-1}^{k(i)} + (Y_t - M_{1t}^{k(i)} - M_{2t}^{k(i)})$$
$$b_t^i = \gamma b_{t-1}^{k(i)} + 1$$
$$\theta_t^i \sim Gamma(a_t^i, b_t^i).$$

5 | Numerical Illustrations

To illustrate the implementation of the Gibbs sampler and particle filtering algorithms, we consider actual data on the number of weekly shopping trips for a household to the supermarket over 104 weeks. The data is a subset of a large set used in Kim [45].

In Figure 1 we show the weekly time series and its sample autocorrelation function (ACF). We note that the time series exhibits moderate-level autocorrelations at lags 1 and 2.

Overall sample mean and variance for 104 weeks are 6.92 and 4.26, respectively. However, as can be seen from Table 1, there is evidence suggesting that mean is not constant over the 104 week period.

5.1 | Static INAR(1)

Next, we consider the Bayesian analysis of the static INAR(1) model where $\theta_t = \theta$ for all t. We will use the static model as a reference in our analysis. The posterior distributions of α and θ from the static INAR(1) model are shown in Figure 2. We can obtain the posterior means of α and θ as 0.622 and 2.565, respectively.

TABLE 1 | Mean and variance of weekly shopping trips over different time periods.

Weeks	Sample mean	Sample variance
1–26	6.19	2.88
27–52	8.04	3.72
53–78	7.81	2.40
79–104	5.65	4.24

As previously mentioned, the static model arises as a special case of the dynamic INAR(1) model as the value of the discount parameter γ approaches to 1. In other words, these results can be replicated by using a Gibbs sampler or the PF algorithm when we set γ to a large value such as 0.999.

5.2 | Dynamic INAR(1)

The use of the Gibbs sampler is not computationally efficient for sequential updating, as it requires the sampler to be rerun for every data point observed. Therefore, in what follows, we will focus on the collapsed particle filtering results. The choice of the discount parameter γ plays an important role in the Bayesian analysis of the dynamic INAR(1) model. As discussed in Section 3.2, we can learn about γ using the marginal likelihood in (19) sequentially using the Monte Carlo (MC) average of the updated samples of $M^{(t)}$ at each time point t. Alternatively, one can use the marginal likelihood (19) to specify a fixed value of the discount factor, which is the most supported by the data as in West and Harrison [25]. In other words, using the posterior samples of $M^{(t)}$, we can obtain a MC average of (19) for fixed values of γ and find the γ that maximizes the MC average. This can be done sequentially over time, in the sense of Dawid [46], to see which value of γ is supported dynamically by the data.

It is important to note that γ is a tuning parameter in dynamic models. It represents the change in uncertainty about θ values

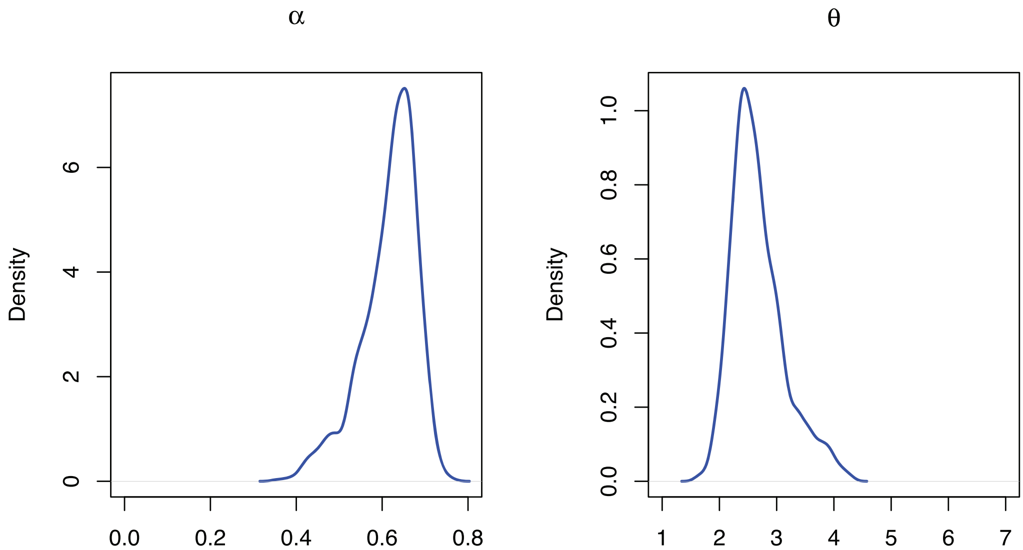

FIGURE 2 | Marginal posterior distributions for α and θ based on the static INAR(1) model.

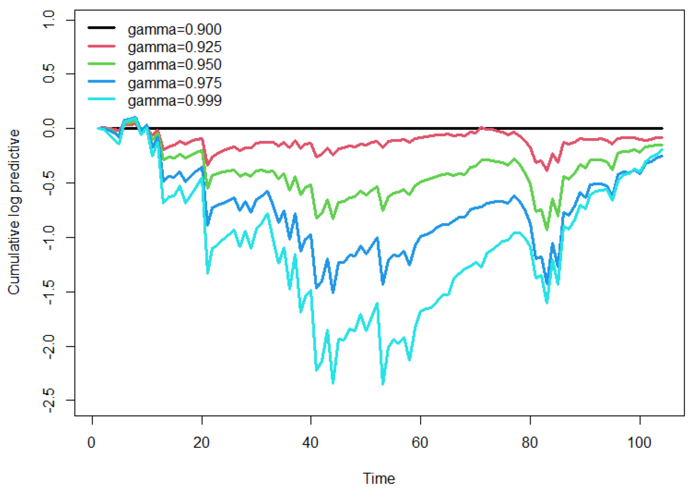

FIGURE 3 | Cumulative marginal likelihood for different values γ from INAR(1), relative to $\gamma = 0.9$.

neighboring each other. Based on the Markovian model presented in Section 2.2, we can show that the percent increase in uncertainty (in terms of variance) from time $(t-1)$ to t is given by $(1-\gamma)/\gamma$ for all t. As noted by West and Harrison ([25], page 51), it is common practice to select $\gamma \in (0.8, 1)$.

We consider the values of $\gamma = 0.900, 0.925, 0.950, 0.975$ and 0.999 and evaluate the MC average of the marginal likelihood (19) based on the particles of $M^{(t)}$ at each time point $t = 2, \ldots, T$. Based on the results of the collapsed particle filtering, in Figure 3, we illustrate the log of cumulative marginal likelihoods for different values γ relative to $\gamma = 0.9$. Figure 3 suggests that the static INAR(1) model (which is represented by $\gamma = 0.999$) is not supported by the data. Based on the figure, we will use $\gamma = 0.9$ in our analysis of the dynamic INAR(1) model.

The posterior distribution of α obtained using the collapsed particle filter is shown in Figure 4, suggesting that the median values of the probability of thinning stabilize around the values of 0.55–0.60.

The posterior (filtering) distributions of the θ_t's, when $\gamma = 0.9$ are shown in Figure 5. As expected, the filtering distributions of θ_t's in this case exhibit more variability over time compared to their counterparts from models with higher values of γ, which are not shown here.

In Figure 6, we present the plot of $E[M_t + \theta_t | Y^t]$ versus actual values of Y_t's for $t = 2, \ldots, T$. The figure illustrates the fit of the dynamic INAR(1) model to the actual values of number of shopping trips. We can look at the predictive means of Y_t's at time $t-1$ to assess the forecast performance of the model. Specifically, we can use

$$E[Y_t | Y^{(t-1)}] = E[\alpha | Y^{(t-1)}] Y_{t-1} + E[\theta_t | Y^{(t-1)}]$$

which can be obtained as a MC average for time periods $t = 2, \ldots, T$.

We can alternatively use the medians of the predictive distribution Y_t's at time $t-1$ to assess the forecast performance. To

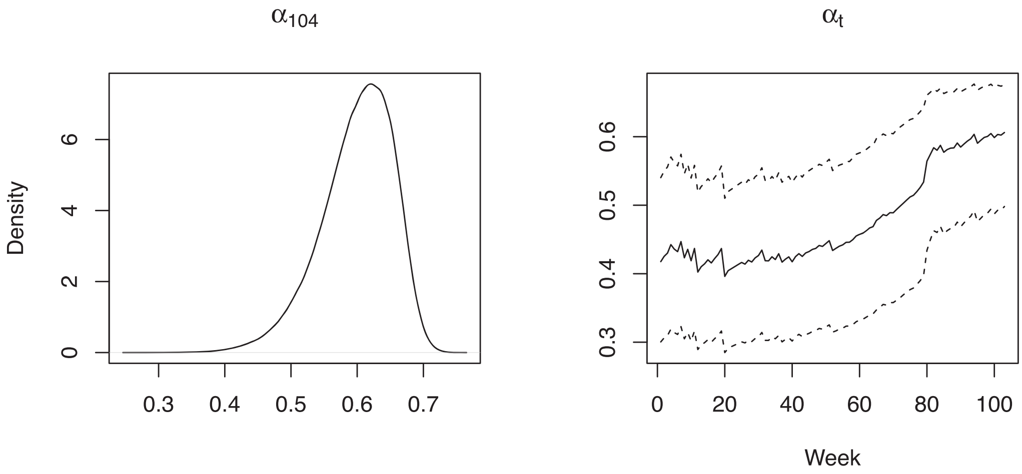

FIGURE 4 | Marginal posterior distribution of α at time $t = 104$ (left) and 5th, 50th, and 95th percentiles of $p(\alpha|Y^{(t)})$, for $t = 1, \ldots, 104$ (right), when $\gamma = 0.9$.

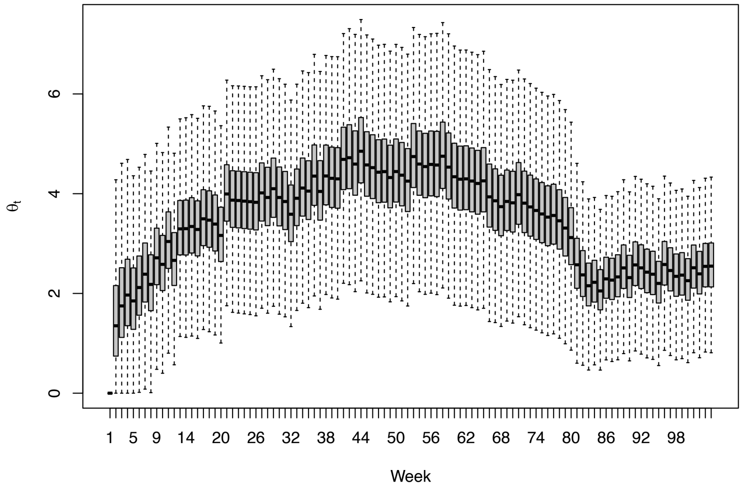

FIGURE 5 | Filtering distributions of θ_t's in INAR(1) model when $\gamma = 0.9$.

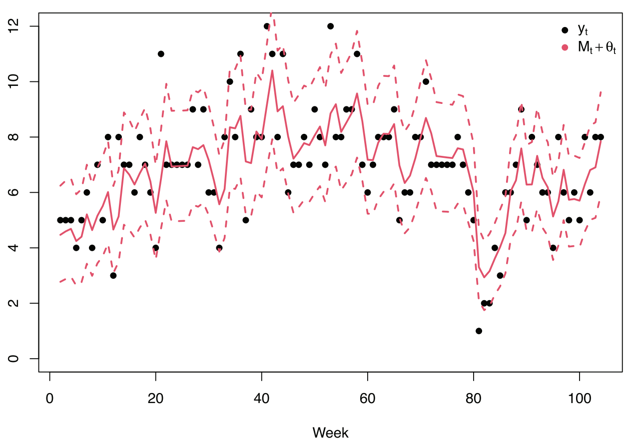

FIGURE 6 | Actual number of trips versus $E(M_t + \theta_t|Y^t)$ in INAR(1) model.

obtain the median of Y_t at time $(t-1)$ we need the draws from the distribution of Y_t. Given the N particles from the distributions $p(\alpha|Y^{(t-1)})$, $p(\theta_{t-1}|Y^{(t-1)})$, for $i = 1, \ldots, N$ we can draw from the predictive distribution of Y_t using the following.

1. Draw $M_t^i \sim Binom(\alpha^i, Y_{t-1})$ and draw $\theta_t^i \sim Gamma(\gamma a_{t-1}, \gamma b_{t-1})$
2. Draw $\epsilon_t^i \sim Pois(\theta_t^i)$
3. Obtain $Y_t^i = M_t^i + \epsilon_t^i$.

We present the plot of actual versus one-step ahead forecasts using predictive means (shown in red) and predictive medians (in blue) in Figure 7 of Y_t's at time $t-1$ along with 90 per cent Bayesian credibility intervals. At the bottom portion of the figure the posterior predictive densities $p(Y_t|Y^{(t-1)})$ are shown for $t = 82, \ldots, 85$ with the observed values of Y_t are displayed as a $\bullet$. It is important to note that these time points correspond to predictive distributions which are more skewed, but in spite of that the predictive performance is quite satisfactory.

5.3 | Comparing INAR(1) and INAR(2)

In Figure 8, we present the ACF of the residuals from the INAR(1) model based on predictive means. The partial autocorrelation function (PACF) of the data is also shown on the figure. The ACF

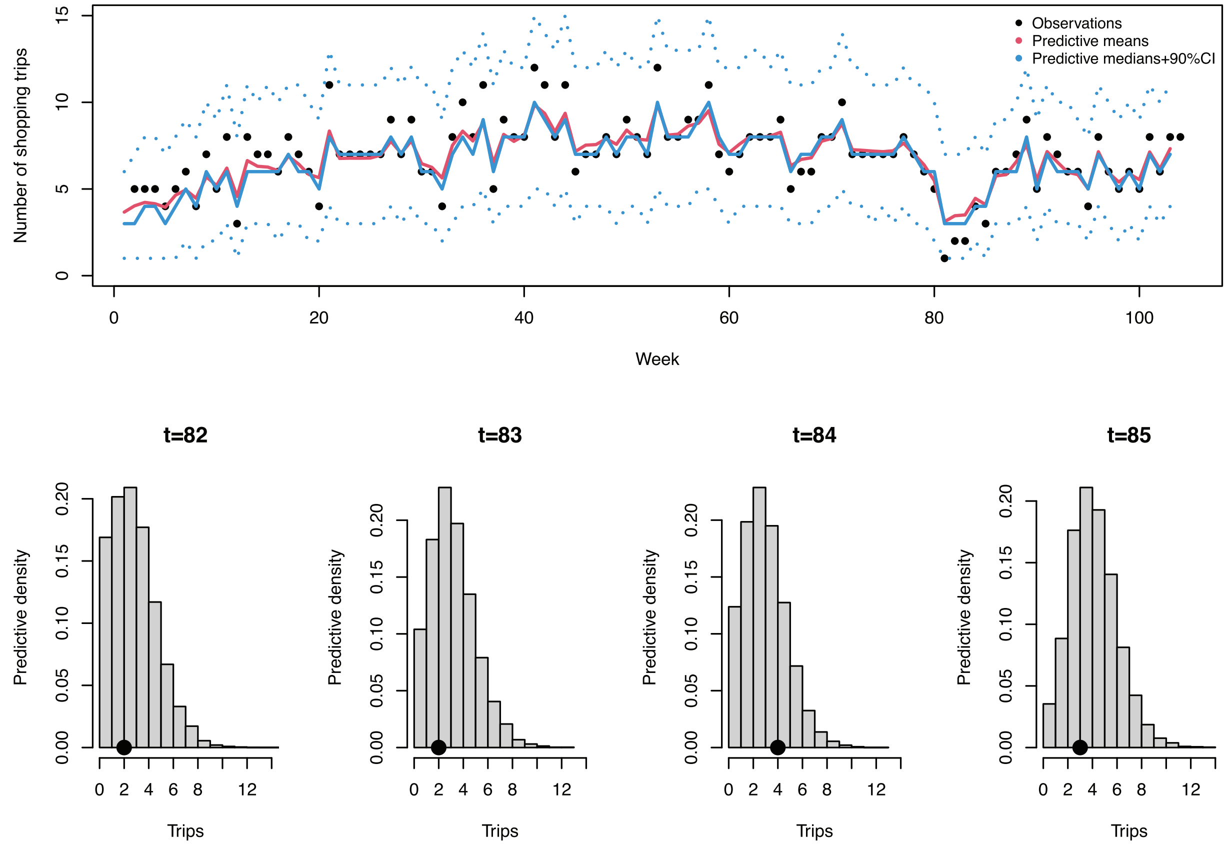

FIGURE 7 | *Top row:* Observed number of shopping trips versus predicted means, predictive medians and predictive 90% credibility intervals. *Bottom row:* Predictive densities, $p(Y_t|Y^{(t-1)})$, for $t = 82, \ldots, 85$.

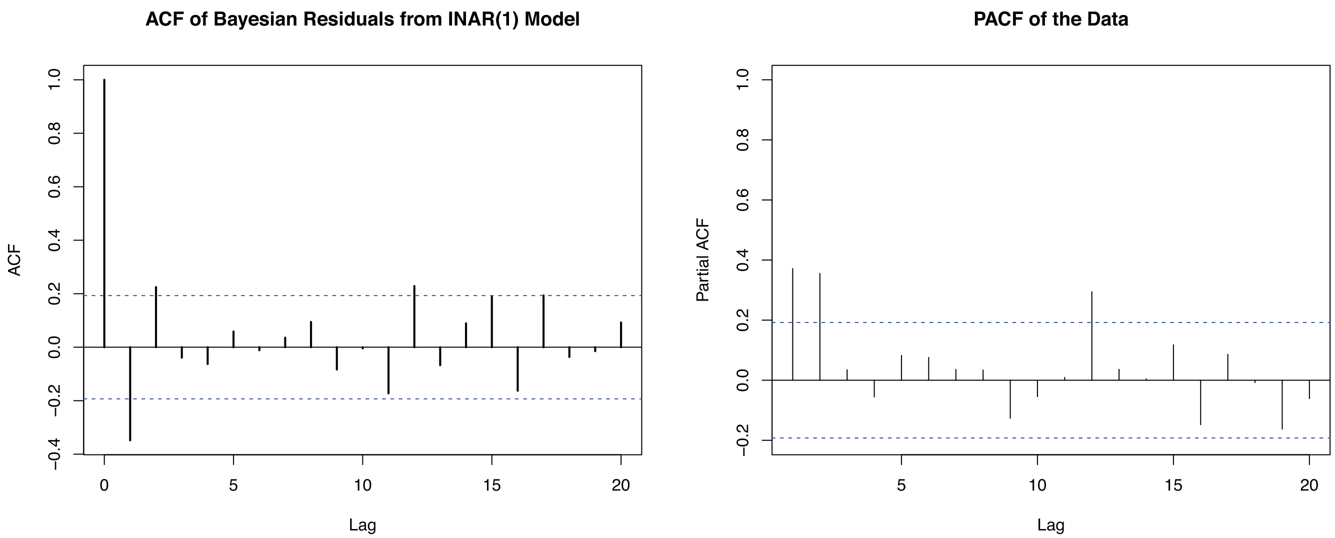

FIGURE 8 | ACF of residuals from the INAR(1) model and the PACF of the data.

of the residuals still displays autocorrelation at lag 1. Since the PACF of the original data exhibits autocorrelations both at lags 1 and 2, an INAR(2) process may be more appropriate in this case. This will be considered next.

We compare the results of dynamic INAR(1) with $\gamma = 0.9$ with the implementations of dynamic INAR(2) for various values of γ. The INAR(2) model analysis is developed using the collapsed PF results presented in Section 4. Evaluation of the marginal likelihood sequentially for different values of $\gamma \in (0.900, 0.999)$, under the INAR(2) model, suggests $\gamma = 0.98$ is favored. Figure 9 shows the quantiles of θ_t for $\gamma = 0.9, 0.98, 0.999$. As expected θ_t under $\gamma = 0.98$ exhibits a smoother behavior over time compared to the case of $\gamma = 0.9$. It is important to note that the static INAR(2) model which is corresponding to $\gamma = 0.999$ in Figure 9 (shown in green) can not adapt to the drop in the data in week 81. Figure 10

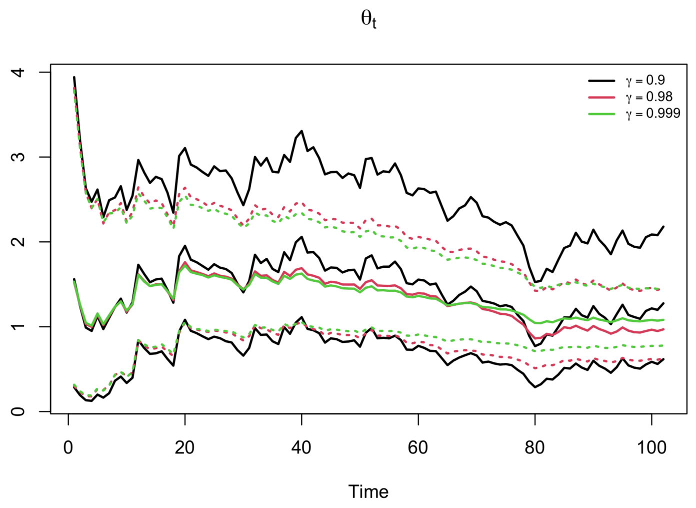

FIGURE 9 | Sequential posterior quantiles (5%, 50% and 95%) for the state space θ_t at various values of γ.

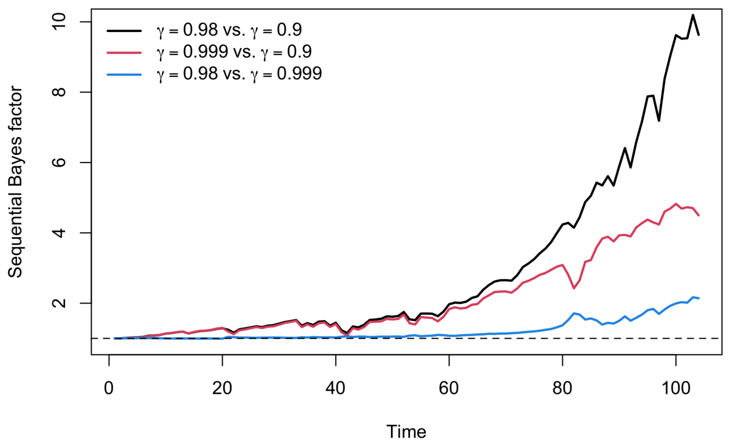

FIGURE 10 | Sequential Bayes factor for the INAR(2) model for various values of γ.

shows the sequential Bayes factor for the INAR(2) model for $\gamma \in \{0.9, 0.98, 0.999\}$. As it can be observed, $\gamma = 0.98$ is better than $\gamma = 0.999$, which is better than $\gamma = 0.9$. The evidence in favor of the dynamic INAR(2) over the static INAR(2) becomes more pronounced after week 81 where the data exhibits a drop.

Figure 11 presents the sequential posterior quantiles for α_1 and α_2 in INAR(2). As can be seen, their time-varying behaviors are quite similar, with only minor variations at the end of the sample size. In any event, α_1 seems to be around 0.3, while α_2 settles around 0.6.

Finally, Figure 12 presents the cumulative logarithmic Bayes factor for the top INAR(2) model relative to the INAR(1) model. As can be argued, after 20 observations, the INAR(2) fit becomes overly better than the INAR(1). The ACF of the residuals are presented in Figure 13. Unlike the INAR(1) residuals, the residuals from the dynamic INAR(2) model exhibit no considerable autocorrelation at any of the lags, which also confirms the appropriateness of the dynamic INAR(2) model.

In conclusion, the Bayesian viewpoint provides quantification of uncertainty via probability. As a result, all inferences, forecasts, and model comparisons are based on probabilistic statements which are derived using calculus of probability. This is attractive from an applied perspective, since additional insights can be obtained from Bayesian analysis using the posterior and/or predictive distributions. For example, in the dynamic INAR(2) model, a hypothesis such as $\alpha_1 < \alpha_2$ can be resolved simply using the joint posterior distribution of α_1 and α_2. Similarly, in our use of collapsed particle filtering for the dynamic INAR(2) model with independent beta priors for α_1 and α_2, we are able to assess if $\alpha_1 + \alpha_2 < 1$ which is the condition for stationarity of the static INAR(2).

Our proposed sequential Bayesian predictive framework–based on full predictive distributions–combined with careful model specification and selection, provides a natural approach to achieving calibration in the sense of Dawid [48]. Model comparison in our framework is conducted probabilistically via Bayes factors, which decompose into cumulative predictive likelihoods and are therefore consistent with the prequential perspective.

In addition, certain frequentist tools, such as the sample autocorrelation and partial autocorrelation functions, can be employed for model identification (see Figures 1 and 8). Further, examination of the sample autocorrelations of Bayesian

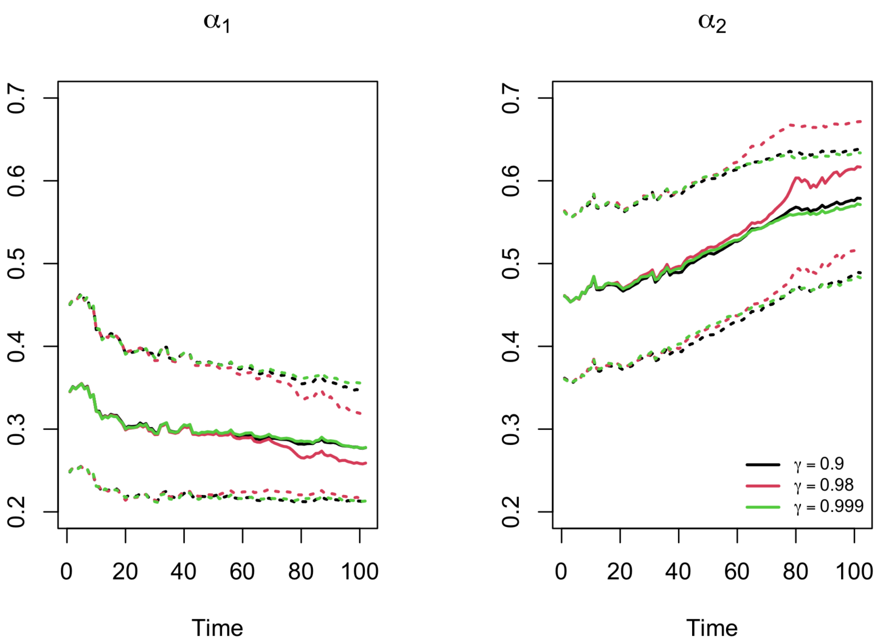

FIGURE 11 | Sequential posterior quantiles (5%, 50% and 95%) for both thinning parameters α_1 and α_2 at various values of γ.

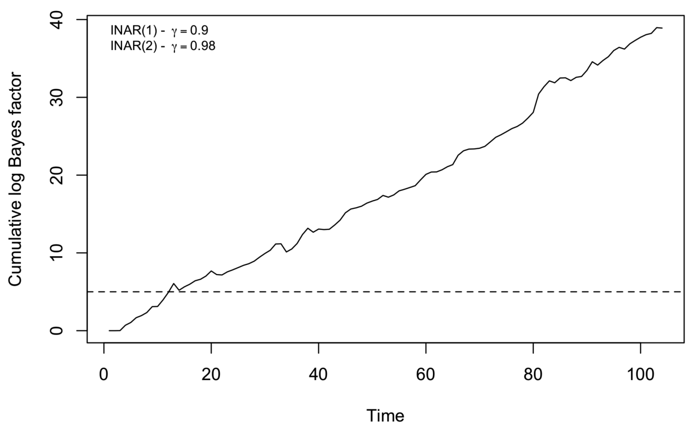

FIGURE 12 | Comparing the best INAR(2) model, where $\gamma = 0.98$, to the best INAR(1) model, where $\gamma = 0.9$, in terms of cumulative log predictive densities. The dashed line represents the threshold for *very strong evidence against* the INAR(1) model, based on the suggestions by Kass and Raftery [47].

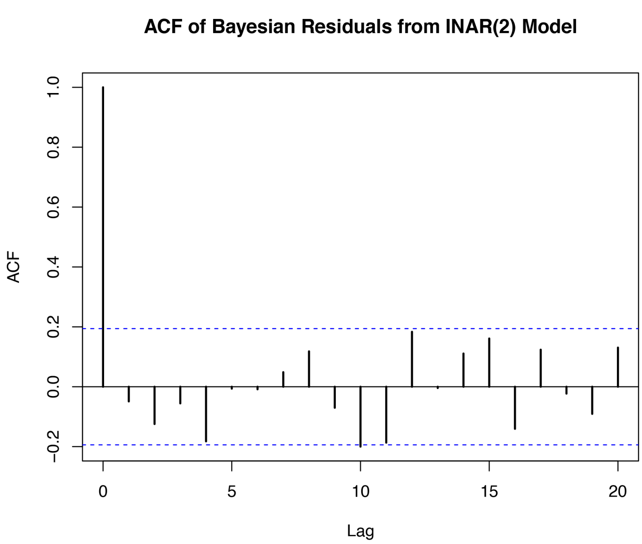

FIGURE 13 | ACF of residuals from the INAR(2) model.

residuals (Figures 8 and 13) provides a useful diagnostic for model adequacy.

The use of frequentist tools for model validation was advocated by George Box [49], who emphasized Bayesian model-based inference while employing frequentist ideas for model criticism. Little ([50] and [51]) extended this perspective by emphasizing the frequentist performance of Bayesian procedures, referring to this approach as "calibrated Bayes." It is important to note that Little's notion of calibration is based on repeated sampling properties, which differs fundamentally from Dawid's [48] concept of calibration, grounded in sequential predictive performance rather than hypothetical repeated sampling.

6 | Concluding Remarks

We introduced a new class of dynamic integer autoregressive (INAR) models for count data. The proposed class of models include static INAR models as well as some of the parameter driven Bayesian time series models for counts as special cases. Bayesian analysis of the proposed model was developed using a Gibbs sampler and particle filtering algorithms were introduced for sequential analysis of the model. Extension of the dynamic model to high-order INAR processes was discussed. Using a real life time series of counts, the dynamic INAR(2) model was shown to perform better than its static counter part as well as the dynamic INAR(1) process. Alternative dynamic INAR processes and the incorporation of covariates into the models were discussed as possible extensions that are currently under investigation. Our future work also involves development of dynamic multivariate INAR processes.

It is important to note that the proposed dynamic INAR(p) models and the presented Bayesian MCMC and sequential MC approaches are novel and differ from the previously considered state-dependent Bayesian INAR models in several respects. First of all, the proposed class of Bayesian models' inherent conditional conjugacy allows for using the Gibbs sampler without any Metropolis steps. Secondly, the proposed PF algorithms exploit this conjugacy and the availability of sufficient statistics to update static parameters and provide efficient sequential inference and forecasting. Such a sequential analysis cannot be performed efficiently by previously introduced Bayesian models that relied on MCMC methods.

Funding

H.F.L. receives support from FAPESP through projects numbers 2018/04654-9 and 2023/02538-0.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

1. A. M. Schmidt and J. B. M. Pereira, "Modelling Time Series of Counts in Epidemiology," *International Statistical Review* 79 (2011): 48–69.

2. T. Aktekin and R. Soyer, "Call Center Modeling: A Bayesian State Space Approach," *Naval Research Logistics* 58 (2011): 28–42.

3. T. Aktekin, R. Soyer, and F. Xu, "Assessment of Mortgage Default Risk via Bayesian State-Space Models," *Annals of Applied Statistics* 7 (2013): 1450–1473.

4. S. Hu, N. Ravishanker, J. N. Ivan, and J. Mooradian, "Temporal Modeling of Highway Crash Counts for Senior and Non-Senior Drivers," *Accident Analysis and Prevention* 50 (2013): 1003–1013.

5. V. Serhiyenko, N. Ravishanker, and R. Venkatesan, "Approximate Bayesian Estimation for Multivariate Count Time Series Models," in *Ordered Data Analysis, Modeling and Health Research Methods*, ed. P. K. Choudhary, C. H. Nagaraja, and H. K. Tony Ng (Springer, 2015), 155–167.

6. T. Aktekin, N. G. Polson, and R. Soyer, "Sequential Bayesian Analysis of Multivariate Poisson Count Data," *Journal of Bayesian Analysis* 13 (2018): 385–409.

7. X. Chen, D. Banks, and M. West, "Bayesian Dynamic Modeling and Monitoring of Network Flows," *Network Science* 7 (2019): 292–318.

8. R. Davis, S. Holan, R. Lund, and N. Ravishanker, *Handbook of Discrete-Valued Time Series* (Chapman and Hall/CRC, 2016).

9. R. A. Davis, K. Fokianos, S. H. Holan, et al., "Count Time Series: A Methodological Review," *Journal of the American Statistical Association* 116, no. 535 (2021): 1533–1547.

10. R. Soyer and D. Zhang, "Bayesian Modeling of Multivariate Time Series of Counts," *WIREs Computational Statistics* 14, no. 6 (2021): e1559.

11. I. Silva, M. E. Silva, I. Pereira, and B. McCabe, "Time Series of Counts Under Censoring: A Bayesian Approach," *Entropy* 25, no. 4 (2023): 549.

12. I. Silva, M. E. Silva, and I. Pereira, "Bayesian Modelling of Time Series of Counts With Missing Data," in *New Frontiers in Statistics and Data Science. SPE* 2021. Springer Proceedings in Mathematics & Statistics, vol 469. Springer, (2025), https://doi.org/10.1007/978-3-031-68949-9_7.

13. M. E. Silva, I. Pereira, and B. McCabe, "Bayesian Outlier Detection in Non-Gaussian Autoregressive Time Series," *Journal of Time Series Analysis* 40, no. 5 (2019): 631–648.

14. D. R. Cox, "Statistical Analysis of Time Series: Some Recent Developments," *Scandinavian Journal of Statistics* 8 (1981): 93–115.

15. E. McKenzie, "Some Simple Models for Discrete Variate Time Series," *Water Resources Bulletin* 21 (1985): 645–650.

16. M. A. Al-Osh and A. A. Alzaid, "First-Order Integer-Valued Autoregressive (INAR(1)) Process," *Journal of Time Series Analysis* 8 (1987): 261–275.

17. E. McKenzie, "Some ARMA Models for Dependent Sequences of Poisson Counts," *Advances in Applied Probability* 20 (1988): 822–835.

18. C. H. Weiß, "A Poisson INAR(1) Model With Serially Dependent Innovations," *Metrika* 78 (2015): 829–851.

19. O. B. Lopez, F. Bassetti, G. Carallo, and R. Casarin, "First-Order Integer-Valued Autoregressive Processes With Generalized Katz Innovations," *Econometrics and Statistics* (2025) Available online 12 March 2025.

20. M. Monteiro, M. G. Scotto, and I. Pereira, "Integer-Valued Self-Exciting Threshold Autoregressive Processes," *Communications in Statistics - Theory and Methods* 41, no. 15 (2012): 2717–2737.

21. K. Yang, X. Yu, Q. Zhang, and X. Dong, "On MCMC Sampling in Self-Exciting Integer-Valued Threshold Time Series Models," *Computational Statistics and Data Analysis* 169 (2022): 107410.

22. I. Silva, M. E. Silva, I. Pereira, and N. Silva, "Replicated INAR(1) Process," *Methodology and Computing in Applied Probability* 7 (2005): 517–542.

23. P. Neal and T. Subba Rao, "MCMC for Integer-Valued ARMA Processes," *Journal of Time Series Analysis* 28 (2007): 92–110.

24. P. C. F. Marques, H. Graziadei, and H. F. Lopes, "Bayesian Generalizations of the Integer-Valued Autoregressive Model," *Journal of Applied Statistics* 49, no. 2 (2022): 336–356.

25. M. West and P. J. Harrison, *Bayesian Forecasting and Dynamic Models*, 2nd ed. (Springer-Verlag, 1997).

26. B. Kedem and K. Fokianos, *Regression Models for Time Series Analysis* (Wiley, 2002).

27. C. H. Weiß, "Thinning Operations for Modeling Time Series of Counts-a Survey," *Advances in Statistical Analysis* 92 (2008): 319–341.

28. C. H. Weiß, *An Introduction to Discrete-Valued Time Series* (John Wiley & Sons Ltd, 2018).

29. P. Thyregod, J. Carstensen, H. Madsen, and K. Arnbjerg-Nielsen, "Integer Valued Autoregressive Models for Tipping Bucket Rainfall Measurements," *Environmetrics* 10, no. 4 (1999): 395–411.

30. R. Smith and J. E. Miller, "A Non-Gaussian State Space Model and Application to Prediction of Records," *Journal of the Royal Statistical Society, Series B* 48 (1986): 79–88.

31. A. C. Harvey and C. Fernandes, "Time Series Models for Count or Qualitative Observations," *Journal of Business and Economic Statistics* 7 (1989): 407–417.

32. D. Gamerman, T. R. dos Santos, and G. C. Franco, "A Non-Gaussian Family of State-Space Models With Exact Marginal Likelihood," *Journal of Time Series Analysis* 34 (2013): 625–645.

33. N. G. Polson, J. G. Scott, and J. Windle, "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables," *Journal of the American Statistical Association* 108, no. 504 (2013): 1339–1349.

34. S. Fruhwirth-Schnatter, "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis* 15 (1994): 183–202.

35. N. Gordon, D. Salmond, and A. F. M. Smith, "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings F Radar and Signal Processing* 140 (1993): 107–113.

36. H. F. Lopes and R. S. Tsay, "Particle Filters and Bayesian Inference in Financial Econometrics," *Journal of Forecasting* 30 (2011): 168–209.

37. N. D. Singpurwalla, N. G. Polson, and R. Soyer, "From Least Squares to Signal Processing and Particle Filtering," *Technometrics* 60 (2018): 146–160.

38. C. Carvalho, M. S. Johannes, H. F. Lopes, and N. G. Polson, "Particle Learning and Smoothing," *Statistical Science* 25 (2010): 88–106.

39. H. F. Lopes, C. M. Carvalho, M. S. Johannes, and N. G. Polson, "Particle Learning for Sequential Bayesian Computation (With Discussion)," in *Bayesian Statistics 9*, ed. J. M. Bernardo, M. J. Bayarri, J. O. Berger, et al. (Oxford University Press, 2011), 317–360.

40. M. Abramowitz and I. A. Stegun, eds., *Handbook of Mathematical Functions, No. 55*. Applied Mathematics Series (National Bureau of Standards, 1968).

41. R. K. S. Hankin, "Special Functions in R: Introducing the Gsl Package," *R News* 6 (2006): 392.

42. M. B. Gordy, "Computationally Convenient Distributional Assumptions for Common-Value Auctions," *Computational Economics* 12, no. 1 (1998): 61–78.

43. J. G. Du and Y. Li, "The Integer-Valued Autoregressive INAR(p) Model," *Journal of Time Series Analysis* 12 (1991): 129–142.

44. J. Liu and M. West, "Combined Parameters and State Estimation in Simulation Based Filtering," in *Sequential Monte Carlo Methods in Practice*, ed. A. Doucet, N. de Freitas, and N. Gordon (Springer, 2001), 197–223.

45. B. Kim, "Essays in Dynamic Bayesian Models in Marketing," (Ph.D. Thesis, The George Washington University, 2013).

46. A. P. Dawid, "Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society: Series A (General)* 147, no. 2 (1984): 278–290.

47. R. Kass and A. Raftery, "Bayes Factors," *Journal of the American Statistical Association* 90 (1995): 773–795.

48. A. P. Dawid, "The Well-Calibrated Bayesian," *Journal of the American Statistical Association* 77, no. 379 (1982): 605–610.

49. G. E. Box, "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society, Series A* 143, no. 4 (1980): 383–404.

50. R. J. Little, "Calibrated Bayes," *American Statistician* 60, no. 3 (2006): 213–223.

51. R. J. Little, "Calibrated Bayes, for Statistics in General, and Missing Data in Particular," *Statistical Science* 26, no. 2 (2011): 162–174.