Time-varying auto-regressive models for count time-series

Arkaprava Roy

Department of Biostatistics
University of Florida
2004 Mowry Road, 5th Floor CTRB
Gainesville, FL 32611, USA
e-mail: ark007@ufl.edu

and

Sayar Karmakar

Department of Statistics
University of Florida
230 Newell Drive
Gainesville, FL 32611, USA
e-mail: sayarkarmakar@ufl.edu

Abstract: Count-valued time series data are routinely collected in many application areas. We are particularly motivated to study the count time series of daily new cases, arising from the COVID-19 spread. First, we propose a Bayesian framework to study the time-varying semiparametric AR(p) model for the count and then extend it to a more sophisticated time-varying INGARCH model. We calculate posterior contraction rates of the proposed Bayesian methods with respect to the average Hellinger metric. Our proposed structures of the models are amenable to Hamiltonian Monte Carlo (HMC) sampling for efficient computation. We substantiate our methods by simulations that show superiority compared to some of the existing methods that closely fit this setting. Finally, we analyze the daily time series data of newly confirmed cases in NYC to study the spread of COVID for three months.

Keywords and phrases: Autoregressive model, B-splines, COVID-19, count-valued time series, Hamiltonian Monte Carlo (HMC), INGARCH, posterior contraction rates, non-stationary, Poisson regression.

1. Introduction

Modeling count time series is important in many applications such as disease incidence, accident rates, integer financial datasets such as price movement, etc. This relatively new research stream was introduced in [50] and interestingly he analyzed another outbreak namely the US 1970 Polio incidence rate. This stream was furthered by [8] where Poisson generalized linear models (GLM) with an autoregressive latent process in the mean are discussed. A wide range of dependence was explored in [13] for simple autoregressive (AR) structure

and external covariates. On the other hand, a different stream explored integer-valued time series counts such as ARMA structures as in [6, 5] or INGARCH structure as done in [51, 54, 52, 53]. However, from a Bayesian perspective, the only work to the best of our knowledge is that of [45] where the authors discussed an ARMA model for different count series parameters. However, their treatment of ignoring zero-valued data or putting the MA structure by demeaned Poisson random variable remains questionable. None of these works focused on the time-varying nature of the coefficients except for a brief mention in [33].

Our goals are motivated by both the application and methodological development. To the best of our knowledge, ours is the first attempt to model possibly autoregressive count time series with time-varying coefficients which can be regarded as the time-varying analog of [19]. We consider a linear link-based GLM route instead of the traditional exponential link [20], since linear link helps in better interpretability of the coefficient functions. Linear link however requires more stringent shape restrictions on the functions. We impose those by putting constraints on the B-spline coefficients while modeling those coefficient functions. However, it is possible to extend all the computations of the current paper to an exponential link-based GLM framework. The mean function stands for the overall spread and the autoregressive coefficients stand for the effect of different lags. We are particularly motivated to study the spread of COVID-19 in New York City (NYC) from 23rd January, 2020 to 14th July, 2020 using the daily count data of new cases. In terms of our motivating data application, we wish to identify which lags are significant in our model so that it can be directly linked to the incubation period for the symptoms to show up. We find that some higher-order lags like 6, 7, and 8 are also significant. These findings are in line with several research articles discussing the incubation length for the novel coronavirus with a median of 6-7 days and 98% below 11 days. For example, see [35]. We also find that after the lockdown or stay-at-home orders it takes about 12-16 days to reach the peak and then the intercept coefficient function starts decreasing. This is also an interesting find which characterizes the fact that the number of infected but asymptomatic cases is large compared to the new cases reported. Additional to the time-varying AR model proposal, we also offer an analysis via a time-varying Bayesian integer-valued generalized autoregressive conditional heteroscedasticity (TVBINGARCH) model that assumes an additional recursive term in the conditional expectation (cf. (2.1)). This extension offers some more comprehensiveness in the modeling part as even TVBINGARCH with small orders can help us get rid of choosing an appropriate maximum lag value. Since for a Poisson model, the mean is the same as the variance, this can also be thought of as an extension to the GARCH model in the context of count data. First introduced by [17], these models were thoroughly analyzed in [54, 52, 51, 53, 1]. Our proposal for the time-varying TVBINGARCH model adapts to the general non-stationary time-series theme and also can be viewed as a new contribution from a Bayesian point of view. Finally, we contrast the time-varying AR and the GARCH for both simulations and real-data applications under different metrics of evaluation. We show that our semiparametric time-varying model provides better estimates.

Regression models with varying coefficients were introduced by [28]. They modeled the varying coefficients using cubic B-splines. Later, these models have been further explored in various directions [26, 4, 16, 21, 49]. Spline bases have been routinely used to model the time-varying coefficients within non-linear time series models [7, 30, 29, 2]. We also consider the B-spline series-based priors to model the time-varying coefficient functions. We develop efficient computational algorithms for the proposed models.

Apart from developing a computationally tractable hierarchical model, we also establish posterior contraction rates of the proposed models. [24] established posterior contraction for a general stationary Markov chain with much stricter conditions. However, they relaxed some of those conditions in Theorem 8.29 of [25]. To the best of our knowledge, the posterior contraction rate result of this paper is a first for the time-varying Markov model based on minimal assumptions under Poisson-link. We establish our results based on the approach, used to relax the conditions in Theorem 8.29 of [24]. Our posterior contraction rate is with respect to the average Hellinger metric. The primary theoretical hurdle is to construct exponentially consistent tests in a time-varying Markov setup. Our proposed test construction is inspired by [31, 37]. We construct the test relying on the Neyman-Pearson lemma with respect to negative average log affinity distance and calculate contraction rates. Then we show that the same rate holds for the average Hellinger metric as well. We also discuss a pointwise inferential tool by drawing credible intervals. Such tools are important to keep an objective perspective in terms of the evolution of the time-varying coefficients without restricting it to some specific trend models. See [33] ([32] for an earlier version) for a comprehensive discussion on time-varying models and their applications.

The rest of the paper is organized as following. Section 2 describes the proposed Bayesian models in detail. Section 3 discusses an efficient computational scheme for the proposed method. We calculate posterior contraction rates in Section 4. The performance of our proposed method in capturing true coefficient functions is studied in Section 5 and we show excellent performance over other existing methods. Section 6 deals with an application of the proposed method on COVID-19 spread for NYC. Then, we end with discussions and possible future directions in Section 7. Section 8 contains detailed theoretical proofs.

2. Modeling

Given the rapidly evolving nature of the pandemic, the patterns and number of new affected cases were changing rapidly over different geographical regions. The rapid change in the observed counts makes all earlier time-constant analysis inappropriate and builds a path where we can explore methodological and inferential development in tracking down the trajectory of this spread. Thus, we propose two novel semiparametric time-varying autoregressive models for counts to study the spread and examine the effects of these interventions in the spread based on the time-varying coefficient functions. We first consider the most general case where we model the data using a time-varying

Bayesian integer-valued generalized autoregressive conditional heteroscedasticity (TVBINGARCH) model where the conditional mean depends on the past observations as well as past conditional means. However, the relatively simpler process consisting of a time-varying mean/intercept function along with time-varying autoregressive coefficient functions up to lag-p is also important keeping in mind the scope of application to real data and its interpretation. For example, the particular lags in an AR(p) model for the COVID-19 count data can crave an interesting phenomenon in the lag dynamics of the spread. This might be lost if we model the same using a TVBINGARCH(1,1) model since typically for GARCH type models it is a standard practice to only consider smaller orders.

2.1. Time-varying generalized autoregressive conditional heteroscedasticity model for counts

The integer-valued analog of generalized autoregressive conditional heteroscedasticity model (GARCH) was proposed in [17] in the light of the observation that the variability in the number of cases of campylobacterosis infections also changes with level. We consider here a time-varying analog of such process. The conditional distribution for count-valued time-series X_t given $\mathcal{F}_{t-1} = \{X_i : i \leq (t-1)\}$ and $\mathcal{G}_{t-1} = \{\lambda_i : i \leq (t-1)\}$ is,

$$X_t | \mathcal{F}_{t-1}, \mathcal{G}_{t-1} \sim \text{Poisson}(\lambda_t),$$

$$\text{where } \lambda_t = \mu(t/T) + \sum_{i=1}^{p} a_i(t/T) X_{t-i} + \sum_{j=1}^{q} b_j(t/T) \lambda_{t-j}. \tag{2.1}$$

We call our method time-varying Bayesian Integer valued Generalized Auto Regressive Conditional Heteroscedastic (TVBINGARCH) model. We impose following constraints on the parameter space similar to [18],

$$\mathcal{P}_1 = \{\mu, a_i : 0 < \mu(x) < \infty, \sup_x \sum_{i,j}(a_i(x) + b_j(x)) < 1\}. \tag{2.2}$$

This constraint ensures a unique solution of the time-varying GARCH process as discussed in [14, 38, 18]. Now, we put priors on the functions $\mu(\cdot)$, $a_i(\cdot)$ and $b_j(\cdot)$ such that they are supported in $\mathcal{P}_1$. Using the B-spline bases, we put following hierarchical prior on the unknown functions,

$$\mu(x) = \sum_{j=1}^{K_1} \alpha_j B_j(x) \tag{2.3}$$

$$a_i(x) = \sum_{j=1}^{K_2} \theta_{ij} M_i B_j(x), \quad 0 \leq \theta_{ij} \leq 1, 1 \leq i \leq p, \tag{2.4}$$

$$b_k(x) = \sum_{j=1}^{K_3} \eta_{kj} M_{k+p} B_j(x), \quad 0 \leq \eta_{kj} \leq 1, 1 \leq k \leq q, \tag{2.5}$$

$$M_i = \frac{\tau_i}{\sum_{k=0}^{p} \tau_k}, \quad i = 1, \ldots, p+q, \tag{2.6}$$

$$\theta_{ij} \sim U(0,1) \text{ for } 1 \le i \le p, 1 \le j \le K_2, \tag{2.7}$$

$$\eta_{kj} \sim U(0,1) \text{ for } 1 \le k \le q, 1 \le j \le K_3, \tag{2.8}$$

$$\lambda_0 \sim \text{Inverse-Gamma}(d_1, d_1), \tag{2.9}$$

where λ_0 is the rate parameter for X_0. The specification for the density of X_0 is required for computation. Otherwise we need to assume λ_0 to be known which is not reasonable for a real data application. We primarily focus on the special case where $p = 1, q = 1$. Based on the constraints on the parameter space we consider following prior for α_j's and τ_i's,

$$\alpha_j \sim \text{TN}(0, c_1^2, 0, \infty), \quad \tau_i \sim U(0,1), \tag{2.10}$$

where TN stands for the truncated normal with mean 0, variance c_1^2 and truncated to $[0, \infty)$. In above construction, $\sum_{j=0}^{P} M_j = 1$. Thus $\sum_{j=1}^{p+q} M_j < 1$ if $M_0 > 0$. As $\Pi(M_0 > 0) = 1$, we have $\Pi(\sum_{j=1}^{p+q} M_j < 1) = 1$. Since $0 \le \theta_{ij} \le 1$, we have $\sup_x a_i(x) \le M_i$, and $\sup_x b_j(x) \le M_{p+j}$. Thus $\sup_x \sum_{i=1}^{p} a_i(x) + \sum_{j=1}^{q} b_j(x) \le \sum_{i=1}^{p+q} M_i < 1$. We have $\sum_{j=1}^{p+q} M_j = 1$ if and only if $\tau_0 = 0$, which has zero prior probability. On the other hand, we also have $\mu(\cdot) \ge 0$ as we have $\alpha_j \ge 0$. Thus, the induced priors, described in (2.3)− (2.9) are well supported in $\mathcal{P}_1$.

2.2. Time-varying auto-regressive model for counts

Although our previous modeling framework is more general, one may only wish to study higher-order lag dependence from the past observations. Thus we consider a simplified model in this subsection. The linear Poisson autoregressive model [50, 6] is popular in analyzing higher order lag-dependence in count valued time series. Due to the assumed non-stationary nature of the data, we propose a time-varying version of this model. The conditional distribution for count-valued time-series X_t given $\mathcal{F}_{t-1} = \{X_i : i \le (t-1)\}$ is,

$$X_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t) \text{ where } \lambda_t = \mu(t/T) + \sum_{i=1}^{p} a_i(t/T) X_{t-i}. \tag{2.11}$$

We call our method time-varying Bayesian Auto Regressive model for Counts (TVBARC). The rescaling of the time-varying parameters to the support [0,1] is usual for in-filled asymptotics. Due to the Poisson link in (2.11), both conditional mean and conditional variance depend on the past observations. The conditional expectation of X_t in the above model (2.11) is $\mathbb{E}(X_t | \mathcal{F}_{t-1}) = \mu(t/T) + \sum_{i=1}^{p} a_i(t/T) X_{t-i}$, which needs to be positive-valued. To ensure that, we impose the following constraints on parameter space for the time-varying parameters,

$$\mathcal{P}_2 = \{\mu, a_i : 0 < \mu(x) < \infty, \sup_x \sum_k a_k(x) < 1\}. \tag{2.12}$$

Note that, the conditions imposed (2.12) on the parameters are somewhat motivated by the stationarity conditions for the time-constant versions of these models. This is not uncommon in time-varying AR literature. See [12, 22, 33] for example. Even though the condition on $\mu(\cdot)$ seems restrictive in the light of what we need for invertible time-constant AR(p) process with Gaussian error, it is not unusual when it is used to model variance parameters to ensure positivity; it was unanimously imposed for all the literature mentioned above. Additionally, the above references heavily depend on local stationarity: namely, for every rescaled time $0 < t < 1$, they assume the existence of an $\tilde{X}_i$ process which is close to the observed process. One key advantage of our proposal is that it is free of any such assumption. Our assumption of only the first moment is also very mild for theoretical exploration in Section 4. Moreover, except for a very general linear model discussed in [33], to the best of our knowledge, this is the very first analysis of the time-varying parameter for count time-series modeled by Poisson regression. Thus we choose to focus on the methodological development rather than proving the optimality of these conditions. When $p = 0$, our proposed model reduces to routinely used nonparametric Poisson regression model as in [44].

To proceed with Bayesian computation, we put priors on the unknown functions $\mu(\cdot)$ and $a_i(\cdot)$'s such that they are supported in $\mathcal{P}_2$. The prior distributions on these functions are induced through basis expansions in B-splines. Suitable constraints on the coefficients are imposed to ensure the shape constraints as in $\mathcal{P}_2$. Detailed description of the priors are given below,

$$\mu(x) = \sum_{j=1}^{K_1} \alpha_j B_j(x) \tag{2.13}$$

$$a_i(x) = \sum_{j=1}^{K_2} \theta_{ij} M_i B_j(x), \quad 0 \leq \theta_{ij} \leq 1, \tag{2.14}$$

$$M_i = \frac{\tau_i}{\sum_{k=0}^{p} \tau_k}, \quad i = 1, \ldots, p, \tag{2.15}$$

$$\theta_{ij} \sim U(0,1) \text{ for } 1 \leq i \leq p, 1 \leq j \leq K_2. \tag{2.16}$$

Here B_j's are the B-spline basis functions. The parameters δ_j's are unbounded. Based on the constraints on the parameter space we consider following prior for α_j's and τ_i's,

$$\alpha_j \sim \text{TN}(0, c_1^2, 0, \infty), \quad \tau_i \sim U(0,1), \tag{2.17}$$

where TN stands for the truncated normal distribution with mean 0, variance c_1^2 and truncated in $[0, \infty)$. The priors induced by the above construction are $\mathcal{P}_2$-supported. The verification is very straightforward and similar to the previous subsection.

2.3. Model properties

In this paper, we only consider TVBINGARCH(1,1) which is commonly used for the GARCH class of models. One drawback of TVBARC is the proper selection of lag. To alleviate this, one may then consider the TVBINGARCH framework. As in the stationary case, TVBINGARCH(1,1) can be viewed as TVBARC with infinite order. Then the higher values in $b_1(\cdot)$'s are an indication that there might be important higher lags in TVBARC. Besides, to infer about higher lag dependence TVBARC is more suitable than TVBINGARCH. In our real data illustration, we find that the TVBARC model identifies three important higher-order lags 6,7, and 8 in COVID-19 spread. Such inference is difficult to obtain from TVBINGARCH. If the CH coefficient $b_1(\cdot)$ is uniformly zero, TVBINGARCH(1,1) reduces to TVBARC(1). However, the computational steps for TVBARC(1) do not easily follow from TVBINGARCH(1,1). Furthermore, our theoretical result of TVBINGARCH requires a lower bound for the true CH coefficient which is standard for time-varying GARCH class of models. Thus the theoretical result of TVBARC does not easily follow from TVBINGARCH.

Towards writing the likelihood, note that our proposed models are non-stationary since the coefficient functions $a_i(\cdot)$, $b_j(\cdot)$ are possibly not constant. However, we still take a simple product of individual conditional likelihoods for X_t's rather than first locally approximating it by a stationary process. The latter approach is more prominent in the frequentist framework and this phenomenon is known as 'locally stationary approximation'. This was introduced in some seminal papers by [9, 10] and was later used in many time-varying literatures. See [12, 11, 46] among many others. Towards the Bayesian approach of modeling such approximating phenomenon, interested readers can refer to [39, 40]. However, the assumption of the existence of such an approximating stationary process is somewhat stringent and is probably not required in the Bayesian paradigm. For example, see [15] where the likelihood is formed by taking the product of individual conditional likelihoods for a non-stationary time-series. Other approaches can be found in [27, 48] where the likelihoods for the proposed non-stationary processes were computed without any local stationary approximation. Moreover, note that such approximating stationary processes can be shown to exist under the general smoothness conditions as outlined in Theorem 1 in [12] (for tvARCH case) or Proposition 2.3 in [38](for tvGARCH case). These are easily extendible to the Poisson setting and for more general Holder smooth coefficient functions with probably an amended approximation rate. So in a sense, our smoothness assumption and the parameter restriction as (2.2) or (2.12) implies the existence of such stationary processes without us implicitly putting additional assumption.

3. Posterior computation

In this section, we discuss the Markov Chain Monte Carlo (MCMC) sampling method for posterior computation. Our proposed sampling is dependent on

the gradient-based Hamiltonian Monte Carlo (HMC) sampling algorithm [36]. Hence, we show the gradient computations of the likelihood with respect to different parameters for TVBARC(p) and TVBINGARCH(p, q) in the following two subsections.

We obtain the likelihood from the joint density of the data based on our Poisson error model. Since the joint density can be written as the product of conditionals, we can thus write the joint likelihood of the data as a product of conditional densities. Detailed expressions for each case are separately presented below. The likelihoods of the two models are constructed differently, thus we present them separately.

3.1. TVBINGARCH structure

We only derive the computational steps for TVBINGARCH(1,1) which is the frequent choice among GARCH-type models. While fitting this model, we assume for any $t < 0$ $X_t = 0, \lambda_t = 0$. The expression for λ_1 also involves λ_0. Thus, we need to additionally estimate the parameter λ_0, the Poisson rate parameter for X_0. Here the likelihood for TVBINGARCH(1,1) is given by $P(X_0) \prod_{t=1}^{T} P(X_t | \mathcal{F}_{t-1})$. We assume that the marginal distribution of X_0 is Poisson(λ_0) and the prior for λ_0 is Inverse-Gamma(d_1, d_2) as described in Section 2.1. The complete likelihood L_2 of the proposed Bayesian method of (2.1) is given by

$$L_2 \propto \exp\left(\sum_{t=1}^{T} \left[-\left\{ \mu(t/T) + a_1(t/T)X_{t-1} + b_1(t/T)\lambda_{t-1} \right\} + X_t \log\left\{ \mu(t/T) \right. \right.$$

$$\left. + a_1(t/T)X_{t-1} + b_1(t/T)\lambda_{t-i} \right\} \right] - \sum_{j=1}^{K_1} \alpha_j^2/(2c_1^2)$$

$$\left. - (d_1 + 1)\log \lambda_0 - d_1/\lambda_0 \right) \mathbf{1}_{0 \le \theta_{11}, \eta_{ij} \le 1, 0 \le \tau_i \le 1, \alpha_j \ge 0},$$

We calculate the gradients of negative log-likelihood ($-\log L_2$) with respect to the parameters β, θ, η and δ. The gradients are given below,

$$-\frac{d\log L_2}{\alpha_1}$$

$$= \left(1 - \sum_t \frac{B_1(t/T)X_{t-j}}{(\mu(t/T) + a_1(t/T)X_{t-j}) + b_1(t/T)\lambda_{t-1})} \right) + \alpha_j/(2c_1^2),$$

$$-\frac{d\log L_2}{\theta_{11}} = M_i\left(1 - \sum_t \frac{B_1(t/T)X_{t-j}}{(\mu(t/T) + a_j(t/T)X_{t-j}) + b_k(t/T)\lambda_{t-1})} \right),$$

$$-\frac{d\log L_2}{\eta_{kj}} = M_{p+k}\left(1 - \sum_t \frac{B_1(t/T)\lambda_{t-j}}{(\mu(t/T) + a_j(t/T)X_{t-j}) + b_k(t/T)\lambda_{t-1})} \right),$$

$$-\frac{d \log L_2}{\tau_j} = \sum_k (M_j \mathbf{1}_{\{j=k\}} - M_j M_k) \times$$

$$\left[\sum_{i \leq p} \theta_{ij} B_j(x) \left(1 - \sum_t \frac{B_j(t/T) X_{t-j}}{(\mu(t/T) + a_j(t/T) X_{t-j} + b_1(t/T) \lambda_{t-1})} \right) \mathbf{1}_{\{j \leq p\}} + \right.$$

$$\left. \sum_{1 \leq k \leq q} \eta_{kj} B_j(x) \left(1 - \sum_t \frac{B_j(t/T) \lambda_t}{(\mu(t/T) + a_j(t/T) X_{t-1} + b_1(t/T) \lambda_{t-1})} \right) \mathbf{1}_{\{j > p\}} \right].$$

The derivative of the likelihood concerning λ_0 is calculated numerically by differentiating from the first principles. Hence, it is sampled using the HMC algorithm too.

3.2. TVBARC structure

Since we do not have any information of the process for $t < 0$, our computation for TVBARC(p) is based on the likelihood $\prod_{t=p}^T P(X_t | \mathcal{F}_{t-1})$. This likelihood may thus be regarded as a quasi-likelihood as we are looking at the joint density of last $T - p + 1$ time points given the first p observations and it is similar to the likelihood from [15]. This likelihood also shares some commonality with the objective functions used for computation in [12, 22]. The complete posterior likelihood L_1 of the proposed Bayesian method in (2.11) is given by

$$L_1 \propto \exp \left(\sum_{t=p}^T \left[- \{ \mu(t/T) + \sum_{i=1}^p a_i(t/T) X_{t-i} \} + X_t \log \{ \mu(t/T) \right. \right.$$

$$\left. \left. + \sum_{i=1}^p a_i(t/T) X_{t-i} \} \right] - \sum_{j=1}^{K_1} \alpha_j^2 / (2c_1^2) \right) \mathbf{1}_{0 \leq \theta_{ij} \leq 1, 0 \leq \tau_i \leq 1, \alpha_j \geq 0},$$

where we have $\mu(x) = \sum_{j=1}^{K_1} \exp(\beta_j) B_j(x), a_i(x) = \sum_{j=1}^{K_2} \theta_{ij} M_i B_j(x)$ and $M_j = \frac{\exp(\delta_j)}{\sum_{k=0}^p \exp(\delta_k)}$. We develop efficient MCMC algorithm to sample the parameter β, θ and δ from the above likelihood. The derivatives of above likelihood with respect to the parameters are easily computable. This helps us to develop an efficient gradient-based MCMC algorithm to sample these parameters. We calculate the gradients of negative log-likelihood $(-\log L_1)$ with respect to the parameters β, θ and δ. The gradients are given below,

$$-\frac{d \log L_1}{\alpha_j} = \left(1 - \sum_t \frac{B_j(t/T) X_t}{(\mu(t/T) + \sum_j a_j(t/T) X_{t-j})} \right) + \alpha_j / (2c_1^2),$$

$$-\frac{d \log L_1}{\theta_{ij}} = M_i \left(1 - \sum_t \frac{B_j(t/T) X_t}{(\mu(t/T) + \sum_j a_j(t/T) X_{t-j})} \right),$$

$$-\frac{d \log L_1}{\tau_j} =$$

$$\sum_k (M_j \mathbf{1}_{\{j=k\}} - M_j M_k) \sum_i \theta_{ij} B_j(x) \left(1 - \sum_t \frac{B_j(t/T) X_{t-j}}{(\mu(t/T) + \sum_j a_j(t/T) X_{t-j})} \right),$$

where $\mathbf{1}_{\{j=k\}}$ stands for the indicator function which takes the value one when $j = k$.

As the parameter spaces of θ_{ij}'s and η_{kj}'s have bounded support, we map any Metropolis candidate, falling outside of the parameter space back to the nearest boundary point of the parameter space. To obtain a good acceptance rate, we tune our HMC sampler periodically. There are two tuning parameters in HMC namely the leapfrog step, and the step size parameter. The step size parameter is tuned to maintain an acceptance rate within the range of 0.6 to 0.8. The step size is reduced if the acceptance rate is less than 0.6 and increased if the rate is more than 0.8. This adjustment is done automatically after every 100 iterations. However, we choose to pre-specify the leapfrog step at 30 and obtain good results. Due to the increasing complexity of the parameter space in TVBINGARCH, we consider updating all the parameters involved in $a_i(\cdot)$'s, $b_k(\cdot)$'s, and λ_0 together.

4. Large-sample properties

In this section, we obtain posterior contraction rates for the two proposed models. Posterior contraction measures the speed at which we can recover the true parameter from the posterior distribution with increasing sample size. The notion of recovery is specified by a semi- metric d.

Definition [25]: The posterior contraction rate at the true parameter $\kappa_0 \in \mathcal{A}$ with respect to the semi-metric d on $\mathcal{A}$ is a sequence $\epsilon_T \to 0$ such that $P_{\kappa_0} \Pi(\kappa : d(\kappa, \kappa_0) > M_T \epsilon_T | X^{(T)}) \to 0$ for every $M_T \to \infty$, where $\mathcal{A}$ denotes the parameter space of θ_0. Here $X^{(T)}$ stands for the complete dataset.

Although TVBINGARCH(1,1) may reduce to TVBARC(1) assuming $b_1(x) = 0$ for all $x \in [0, 1]$, the required technical assumptions do not allow us to derive the results for TVBARC as a special case for TVBINGARCH. For clarity in presenting the assumptions under which the respective results are established, we will make the conditions in (2.12) and (2.2) more specific. Since TVBARC is a simpler model, we first develop the theoretical results for this model and then make modifications to obtain the results for TVBINGARCH.

4.1. TVBARC structure

We start by studying large sample properties of the simpler AR model in (2.11). For simplicity, we fix order p at $p = 1$ for this section however the results are easily generalizable for any fixed order p with some additional assumptions. The posterior distribution will be the same as in the previous section. The posterior consistency is studied in the asymptotic regime of increasing sample size T. Let

$\kappa = (\mu, a_1)$ stand for the complete set of parameters. For sake of generality of the method, we put a prior on K_1 and K_2 with probability mass function given by,

$$\Pi(K_i = k) = b_{i1} \exp[-b_{i2}k(\log k)^{b_{i3}}], \tag{4.1}$$

with $b_{i1}, b_{i2} > 0$ and $0 \leq b_{i3} \leq 1$ for $i = 1, 2$. Poisson and geometric probability mass functions appear as special cases of the above prior density for $b_{i3} = 1$ or 0 respectively. These priors have not been considered while fitting the model as it would require computationally expensive reversible jump MCMC strategy. We study the posterior consistency with respect to the average Hellinger distance on the coefficient functions which is

$$d_{1,T}^2 = \frac{1}{T}d_H^2(\kappa_1, \kappa_2) = \frac{1}{T}\int(\sqrt{f_1} - \sqrt{f_2})^2,$$

where $f_1 = \prod_{t=1}^{T} P_{\kappa_1}(X_t|X_{t-1})$. Here P stands for the conditional Poisson density defined in (2.11). The contraction rate will depend on the smoothness of true coefficient functions μ and a and the parameters b_{13} and b_{23} from the prior distributions of K_1 and K_2. Let $\kappa_0 = (\mu_0, a_{10})$ be the truth of κ.

Assumptions (A): There exists constants $0 < M_\mu < M_X$ such that,

(A.1) At time $t = 0$, $\mathbb{E}_{\kappa_0}(X_0) < M_X$.
(A.2) The coefficient functions $\sup_{x \in [0,1]} \mu_0(x) < M_\mu$ and $\sup_{x \in [0,1]} a_{10}(x) < 1 - M_\mu/M_X$.
(A.3) $\inf_{x \in [0,1]} \min(\mu_0(x), a_{10}(x)) > \rho$ for some small $\rho > 0$.

Assumptions (A.1), (A.2) ensure

$$\mathbb{E}_{\kappa_0}(X_t) = \mathbb{E}_{\kappa_0}(\mathbb{E}_{\kappa_0}(X_t|X_{t-1})) < M_\mu + \left(1 - \frac{M_\mu}{M_X}\right)M_X < M_X$$

by recursion. Assumption (A.3) is imposed to ensure strict positivity of parameters and is standard in time-varying literature that deals with such constrained parameters.

Posterior consistency theory studies recovery of the 'true' parameter κ_0 with increasing sample size when the data is sampled from the distribution characterized by κ_0. Our notion of recovery is based on the average Hellinger metric $d_{1,T}^2$ defined above.

Theorem 1. *Under assumptions (A.1)-(A.3), let the true functions $\mu_0(\cdot)$ and $a_{10}(\cdot)$ be Hölder smooth functions with regularity level ι_1 and ι_2 respectively, then the posterior contraction rate with respect to the distance $d_{1,T}^2$ is*

$$\max\left\{T^{-\frac{\iota_1}{2\iota_1+1}}(\log T)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, T^{-\frac{\iota_2}{2\iota_2+1}}(\log T)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2}\right\}.$$

where b_{ij}'s are specified in (4.1).

For the proof, the first step is to calculate posterior contraction rate with respect to average log-affinity $r_T^2(f_1, f_2) = -\frac{1}{T} \log \int f_1^{1/2} f_2^{1/2}$ and then show that $r_T^2(f_1, f_2) \lesssim \epsilon_T^2$ implies $\frac{1}{T} d_H^2(f_1, f_2) \lesssim \epsilon_T^2$. The average log-affinity provides a unique advantage to construct exponentially consistent tests leveraging on the famous Neyman-Pearson Lemma as has also been used in [37] for a multivariate linear regression setup under group sparsity. The proof is postponed to Section 8. The proof is based on the general contraction rate result from [25] and some results on B-splines-based finite random series.

4.2. TVBINGARCH structure

Next, we discuss the more comprehensive tvBINGARCH model (2.1). To maintain simplicity in the proof, we again assume $p = 1, q = 1$. Similar to the previous subsection, we put a prior on the number of B-spline bases, K_i with probability mass function given by,

$$\Pi(K_i = k) = b_{i1} \exp[-b_{i2} k (\log k)^{b_{i3}}],$$

with $b_{i1}, b_{i2} > 0$ and $0 \le b_{i3} \le 1$ for $i = 1, 2, 3$. Let us assume that $\psi = (\mu, a_1, b_1)$ be the complete set of parameters. We study the posterior consistency with respect to the Hellinger distance on the coefficient functions which is

$$d_{2,T}^2 = \frac{1}{T} d_H^2(\psi_1, \psi_2) = \frac{1}{T} \int (\sqrt{f_1} - \sqrt{f_2})^2,$$

where $f_1 = P_{\phi_1}(X_0) \prod_{t=1}^T P_{\psi_1}(X_t | X_{t-1}, \lambda_{t-1})$. Here P stands for the conditional Poisson density defined in (1) and the marginal density of X_0, $P_{\phi_1}(X_0)$ is Poisson(λ_{10}) as described in our computational steps.

For this structure, we modify the assumptions as

Assumptions(B): There exists constants $0 < M_\mu < M_X$ such that,

(B.1) At time $t = 0$, $\mathbb{E}_{\psi_0}(X_0), \lambda_0 < M_X$.
(B.2) The coefficient functions $\sup_{x \in [0,1]} \mu_0(x) < M_\mu$ and $\sup_{x \in [0,1]}(a_{10}(x) + b_{10}(x)) < 1 - M_\mu/M_X$.
(B.3) $\inf_{x \in [0,1]} \min(\mu_0(x), a_{10}(x), b_{10}(x)) > \rho$ for some small $\rho > 0$.

Assumptions (B.1), (B.2) ensure

$$\mathbb{E}_{\psi_0}(X_t) = \mathbb{E}_{\psi_0}(\mathbb{E}_{\psi_0}(X_t | X_{t-1}, \lambda_{t-1})) < M_\mu + \left(1 - \frac{M_\mu}{M_X}\right) M_X < M_X$$

by recursion. Thus we have, by Assumption (B.1-B.2)

$$\mathbb{E}_{\psi_0}(X_t) < M_X, \quad \mathbb{E}_{\psi_0}(\lambda_t) = \mathbb{E}_{\psi_0}(X_t | X_{t-1}, \lambda_{t-1}) = \mathbb{E}_{\psi_0}(X_t) < M_X.$$

Assumption (B.3) is imposed to ensure strict positivity of parameters and is standard in time-varying literature that deals with such constrained parameters. Now we present our posterior contraction rate theorem below. The definition of the contraction rate is the same as before.

Theorem 2. *Under assumptions (B.1)-(B.3), let the true functions $\mu_0(\cdot)$, $a_{10}(\cdot)$ and $b_{10}(\cdot)$ be Hölder smooth functions with regularity level ι_1, ι_2 and ι_3 respectively, then the posterior contraction rate with respect to the distance $d_{2,T}^2$ is*

$$\max\left\{T^{-\frac{\iota_1}{2\iota_1+1}}(\log T)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, T^{-\frac{\iota_2}{2\iota_2+1}}(\log T)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2},\right.$$

$$\left. T^{-\frac{\iota_3}{2\iota_3+1}}(\log T)^{\iota_3/(2\iota_3+1)+(1-b_{33})/2}\right\}.$$

The proof follows from a similar strategy as in Theorem 1. An outline of the proof can be found in the Section 8.

5. Simulation studies

In this section, we study the performance of our proposed Bayesian method in capturing the true coefficient functions. We compare both TVBARC and TVBINGARCH methods with some other competing models. It is important to note that, this is to the best of our knowledge first work in Poisson autoregression with a time-varying link. Thus, we compare our method with the existing time-series models with time-constant coefficients for count data and time-varying AR with Gaussian error. We also examine the estimation accuracy of the coefficient functions for estimating the truth.

The hyperparameter c_1 of the truncated normal prior is set to 10 to ensure weak informativeness. The hyperparameters for Inverse-Gamma prior $d_1 = 0.1$, which is also weakly informative. We consider 6 equidistant knots for the B-splines based on comparing the AMSE scores. We choose the knot number after which the AMSE score does not change significantly. We collect 10000 MCMC samples and consider the last 5000 as post-burn-in samples for inferences. In absence of any alternative method for the time-varying $AR(p)$ model of count-valued data, we compare the estimated functions with the true functions in terms of the posterior estimates of functions along with its 95% pointwise credible bands. The credible bands are calculated from the MCMC samples at each point $t = 1/T, 2/T, \ldots, 1$. We also compare different competing methods in terms of average MSE (AMSE) score using the INGARCH method of `tsglm` from R package `tscount`, GARMA using `tscount` as well, `tvAR` and our proposed Bayesian methods. The AMSE is defined as $\frac{1}{T}\sum_t(X_t - \hat{\lambda}_t)^2$. We estimate this in terms of the posterior mean of AMSEs across MCMC as

$$AMSE = \frac{1}{5000}\sum_{S=1}^{5000}\frac{1}{T}\sum_t(X_t - \hat{\lambda}_t^S)^2,$$

where $\hat{\lambda}_t^S$ is the posterior estimate of λ_t at S-*th* postburn sample.

5.1. Case 1: TVBINGARCH structure

For the tvBINGARCH case, we only consider one simulation settings $p = 1, q = 1$; $X_t \sim \text{Poisson}(\mu(t/T) + a_1(t/T)X_{t-1} + b_1(t/T)\lambda_{t-1})$. Two different choices

for T have been considered, $T = 100$ and 200 and for $x \in [0,1]$ the coefficient functions are,

$$\mu_0(x) = 25 \exp\left(-(x - 0.5)^2/0.1\right),$$
$$a_1(x) = 0.3(x - 1)^2 + 0.1,$$
$$b_1(x) = 0.1x^{1.5} + 0.1$$

Figure 1 compares the estimated functions with the truth for sample size 200 for the model in (2.1) with $p = 1, q = 1$. The performance of our method is compared to other competing methods in Tables 1.
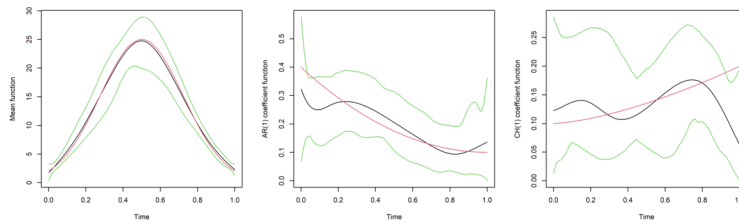

FIG 1. *Estimated coefficient functions for the TVBINGARCH(1,1) and sample size 1000. Red is the true function, black is the estimated curve along with the 95% pointwise credible bands in green.*

TABLE 1
Average MSE comparison for different sample sizes across different methods when the true model is (2.1) with $p = 1, q = 1$.

	INGARCH(1,1)	GARMA(1,1)	tvAR(10)	TVBINGARCH(1,1)
$T = 100$	27.38	27.60	24.50	**22.83**
$T = 500$	24.02	24.07	22.90	**21.23**
$T = 1000$	23.23	23.32	22.93	**21.19**

5.2. *Case 2: TVBARC structure*

Here, we consider two model settings $p = 1; X_t \sim \text{Poisson}(\mu(t/T) + a_1(t/T)X_{t-1})$ and $p = 2; X_t \sim \text{Poisson}(\mu(t/T) + a_1(t/T)X_{t-1} + a_2(t/T)X_{t-2})$ for $t = 1, \ldots, T$. Three different choices for T have been considered, $T = 100, 500$ and 1000. The true functions are for $x \in [0,1]$,

$$\mu_0(x) = 10 \exp\left(-(x - 0.5)^2/0.1\right),$$
$$a_{10}(x) = 0.3(x - 1)^2 + 0.1,$$
$$a_{02}(x) = 0.4x^2 + 0.1.$$

We compare the estimated functions with the truth for sample size 1000 in Figures 2 and Figure 3 for the models $p = 1$ and $p = 2$ respectively. Tables 2

TABLE 2
AMSE comparison for different sample sizes across different methods when the true model is (2.11) with p = 1.

	INGARCH(1,0)	GARMA(1,0)	TVAR(1)	TVBARC(1)
T = 100	11.60	11.18	11.41	**8.65**
T = 500	11.35	11.04	11.24	**8.12**
T = 1000	11.05	10.73	10.94	**7.02**

TABLE 3
AMSE comparison for different sample sizes across different methods when the true model is (2.11) with p = 2.

	INGARCH(2,0)	GARMA(2,0)	TVAR(2)	TVBARC(2)
T = 100	18.02	17.28	13.04	**11.01**
T = 500	16.42	15.86	12.61	**10.79**
T = 1000	15.79	15.25	12.75	**10.61**

(a) $\mu()$ (b) $a_1()$
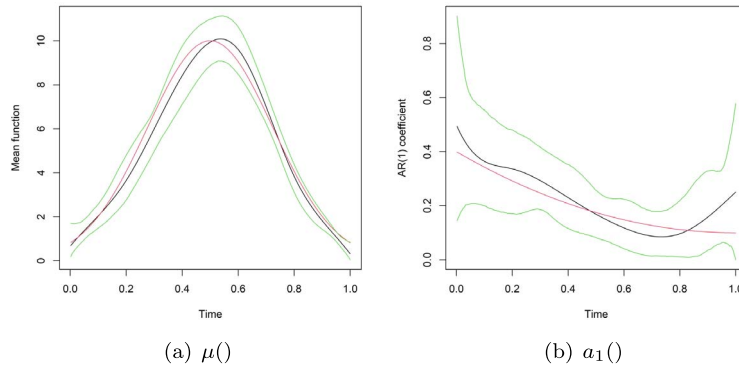
FIG 2. *Estimated mean function in 1st column and estimated AR(1) coefficient function in the 2nd column for the case p = 1 and sample size 1000. Red is the true function, black is the estimated curve along with the 95% pointwise credible bands in green.*
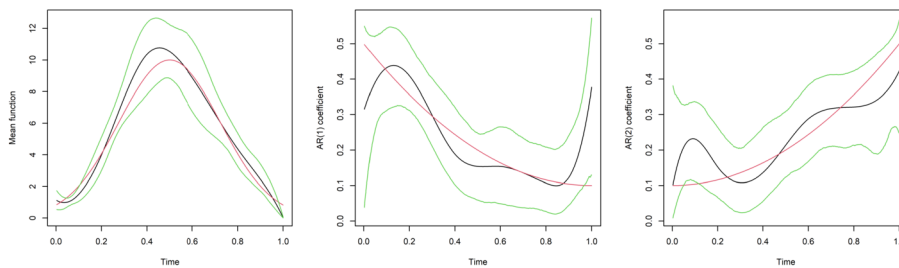

FIG 3. *Estimated coefficient functions for the simulation case p = 2 and sample size 1000. Red is the true function, black is the estimated curve along with the 95% pointwise credible bands in green.*

and 3 illustrate the performance of our method with respect to other competing methods.

Figure 1 and 2 show that our proposed Bayesian method captures the true functions quite well for both of the two simulation experiments. We find that the estimation accuracy improves as the sample size increases. As the sample size grows, the 95% credible bands are also getting tighter, implying lower uncertainty in estimation. This gives empirical evidence in favor of the estimation consistency which has also been verified theoretically in Section 4. The average mean square error (AMSE) is always the lowest for our method in Tables 2, 3 and 1.

6. COVID-19 spread at NYC

We collect the data on daily newly affected cases from 23rd January, 2020 to 14th July, 2020 from an open-source platform {https://www.kaggle.com/ sudalairajkumar/novel-corona-virus-2019-dataset}. The end date 14th July, 2020 is chosen as around that time NYC started the process of re-opening. The data on daily new cases are illustrated in Figure 4. We were particularly interested in NYC data as this city remained an epicenter in the US for about a month. With the help of government interventions and sustained lock-down, the recovery was significant in about 3 months. Such a time-varying nature of the data motivated us to retrospect as to how the mean trend and AR trend behave which can also shed some insight about effects of a lockdown or the contagious spread.

Based on the findings on the incubation of the virus in [35] and others, it is understood that the symptoms often take some time after the virus affects through contagion. Our idea is to consider different models with a varying number of lags for this. We consider TVBARC(1), TVBARC(10), and TVBINGARCH(1,1) here. The results for the TVBARC(1) are illustrated in Figure 5. We see that during the spike in daily new cases the function $a_1(\cdot)$ is the highest. Figure 6 depicts the estimated mean and coefficient functions from a TVBARC(10) model. We find that the estimated $a_1(\cdot)$ functions show a similar trend. On top of that, we see that $a_6(\cdot), a_7(\cdot)$ and $a_8(\cdot)$ also have some effect. Finally, we fit our TVBINGARCH(1,1) which might be considered TVBARC with infinite order. Figure 7 depicts the estimated functions, the mean $\{\mu(\cdot)\}$, AR(1) $\{a_1(\cdot)\}$ and CH(1) $\{b_1(\cdot)\}$ coefficient functions. In Table 4, we compare the AMSE scores across different models. For all the models, we consider 12 equidistant knots based on the AMSE scores as discussed in Section 5.

Figure 6 suggests that even lag 6, 7, and 8 have some significant contribution. The effect of this lag is suppressed in Figure 5 and is expressed in terms of $b_1(\cdot)$ of Figure 7. The estimated mean functions also behave similarly for all three cases. It shows a spike during the rise of daily new cases. After that, it decreases which can talk about successful containment strategies in NYC. More specifically it decreases after around 15 days since the strict implementation of statewide lockdown on 20th March, 2020. This is consistent with what was
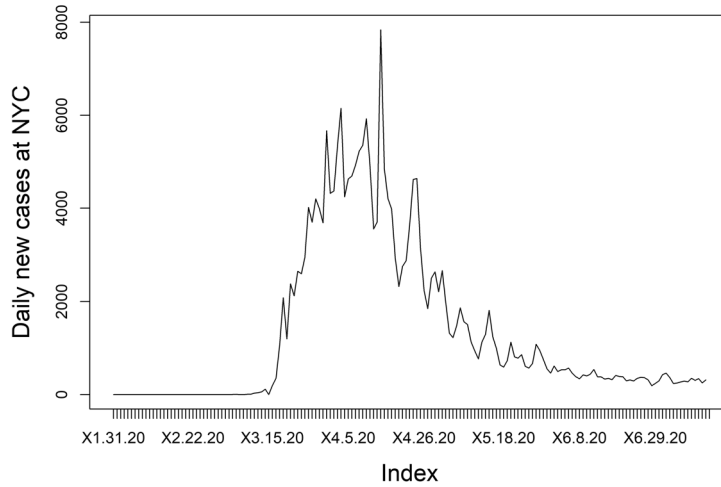
FIG 4. *Daily new COVID-19 cases from 31st January, 2020 to 14th July, 2020 recorded at NYC.*

found in our unsubmitted preprint [43] through an empirical early-stage analysis of the spread in different cities and countries.

The effect of lags 6, 7, and 8 can be attributed to the incubation period of the virus. It can also lead to the finding that there was a weekly periodicity which is probably due to shorter testing/administrative facilities being available during the weekend. Note that our choice of fitting a TVBARC(10) model is more general than separately fitting a seasonal/periodic time-series model. Another important finding is coming from the overall trend of $a_1(\cdot)$. It starts to decrease when the number of cases starts going down. However later on it varies around 0.6 which can be attributed to the fact that the number of new cases did not vary much and remained around the same level from the middle of May. The credible bands look very small around the mean function which is probably due to the large magnitude of the estimated function.

TABLE 4
Average MSE comparison for different methods on NYC data.

Method	AMSE	Method	AMSE	Method	AMSE
INGARCH(1,1)	318056.3	GARMA(10,0)	1682976.1	TVBARC(1)	210258.9
GARMA(1,1)	329610.1	tvAR(1)	338970.6	TVBARC(10)	185777.9
AR(10,0)	1376133.7	tvAR(10)	274913.7	TVBINGARCH(1,1)	212168.1

7. Discussion

We propose a time-varying Bayesian autoregressive model for counts (TVBARC) and a time-varying Bayesian integer-valued generalized autoregressive conditional heteroskedastic model (TVBINGARCH) with linear link function within
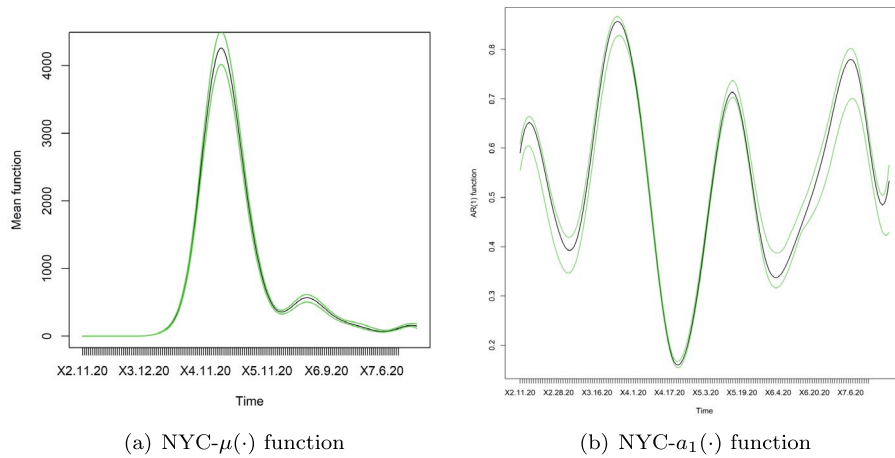
(a) NYC-$\mu(\cdot)$ function

(b) NYC-$a_1(\cdot)$ function

FIG 5. *Estimated mean functions in 1st column and estimated AR coefficient functions in the 2nd column for NYC using TVBARC(1). Black is the estimated curve along with the 95% pointwise credible bands in green for the mean and AR(1) function.*
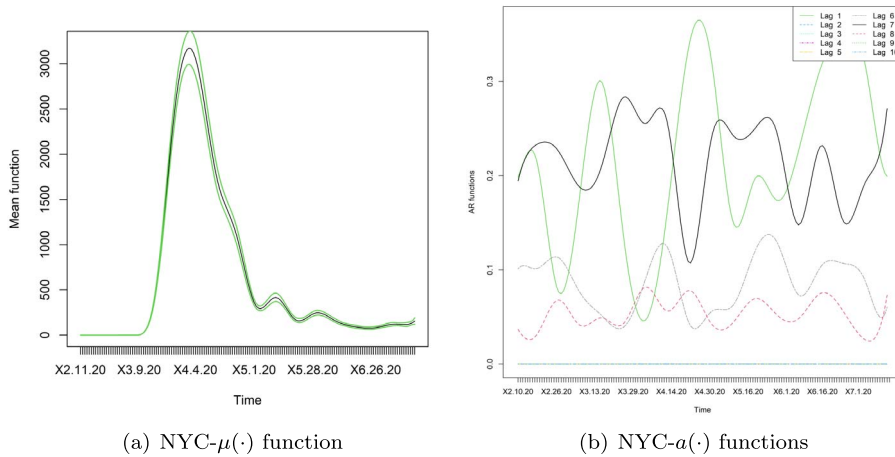

(a) NYC-$\mu(\cdot)$ function

(b) NYC-$a(\cdot)$ functions

FIG 6. *Estimated mean functions in 1st column and estimated AR coefficient functions in the 2nd column for NYC using TVBARC(10). Black is the estimated curve along with the 95% pointwise credible bands in green for the mean function.*

Poisson error to study the time series of daily new confirmed cases of COVID-19. We develop a novel hierarchical Bayesian model that satisfies the stability condition for the respective time-varying models and propose an HMC algorithm-based MCMC sampling scheme. We also establish posterior contraction rate results of the proposed Bayesian methods. The 'R' function with an example code can be found at https://github.com/royarkaprava/TVBARC. Relying on the proposed hierarchical Bayesian model, one can develop a time-varying Bayesian
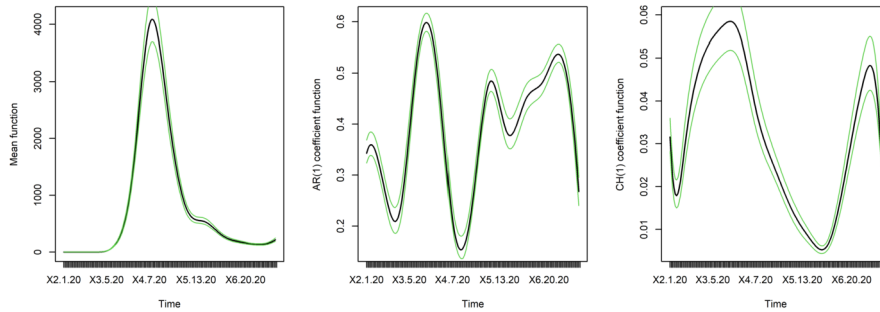
FIG 7. *Estimated coefficient functions for the TVBINGARCH(1,1) on NYC data. Black is the estimated curve along with the 95% pointwise credible bands in green.*

model for positive-valued time-series data too. Our analysis of NYC data shows that there is a time-varying effect of lags 6, 7 and 8. Some preliminary analyses on COVID data using our model based on the data until April 24 are archived in our unpublished pre-print [43]. There are some more interesting findings related to significant lags for different countries.

The definition of the posterior contraction rate we followed involves a diverging sequence $M_T \to \infty$. If it is possible to replace M_T with a large constant M without changing ϵ_T, the contraction rate then holds in a slightly stronger sense (see Chapter 8, [25]). Local stationary approximation of the proposed non-stationary process is expected to help to establish such a result. Establishing the Bernstein von-Mises type theorem to ensure asymptotic normality of the posterior distribution will also be interesting. However, such results are not yet available for the corresponding stationary cases. Nevertheless, it is also an important direction of future research.

As future work, it will be interesting to include some country-specific information such as demographic information, geographical area, the effect of environmental time-series, etc in the model. These are usually important factors for the spread of any infectious disease. We can also categorize the different types of government intervention effects to elaborate more on the specific impacts of the same. In the future we wish to analyze the number of deaths, number of recovered cases, number of severe/critical cases, etc. for these diseases as those will hopefully have different dynamics than the one considered here and can provide useful insights about the spread and measures required. For computational ease, we have considered the same level of smoothness for all the coefficient functions. Fitting this model with different levels of smoothness might be able to provide more insights. Lag selection is a difficult task for time-varying auto-regressive models. One potential future direction would be to put sparsity inducing prior to the time-varying coefficient functions in TVBARC for automatic lag detection. Other than building time-varying autoregressive models for count-valued data using the hierarchical structure from this article, one interesting future direction is to extend this model for vector-valued count data. In general, it is difficult

to model multivariate count data. There are only a limited number of methods to deal with multivariate count data [3, 47, 41]. Building on these multivariate count data models, one can extend our time-varying univariate $AR(p)$ to a time-varying vector-valued $AR(p)$. On the same note, even though we imposed the Poisson assumption for increased model interpretation, in the light of the upper bounds for the KL distance, it is not a necessary criterion and can be applied to a general multiple non-stationary count time-series. Extending some of the continuous time-series invariance results for nonlinear non-stationary and multiple series from [34] to a count series regime will be an interesting challenge. Finally, we wish to undertake an autoregressive estimation of the basic reproduction number with the time-varying version of compartmental models in epidemiology.

8. Proof of Theorems

We study the frequentist property of the posterior distribution in increasing T regime assuming that the observations are coming from a true density f_0 characterized by the parameter κ_0. We follow the general theory of [23] to study the posterior contraction rate for our problem. In the Bayesian framework, the density f is itself a random measure and has distribution Π which is the prior distribution induced by the assumed prior distribution on κ. The posterior distribution of a neighborhood $U_T = \{f : d(f, f_0) < \epsilon_T\}$ around f_0 given the observation $X^{(T)} = \{X_0, X_1, \ldots, X_T\}$ is

$$\Pi_T(U_T^c | X^{(T)}) = \frac{\int_{U_T^c} f(X^{(T)}) d\Pi(\kappa)}{\int f(X^{(T)}) d\Pi(\kappa)}$$

8.1. General proof strategy

The posterior consistency would hold if above posterior probability almost surely goes to zero in $F_{\kappa_0}^{(T)}$ probability as T goes to ∞, where $F_{\kappa_0}^{(T)}$ is the true distribution of $X^{(T)}$. Recall the definition of posterior contraction rate; for a sequence ϵ_T if $\Pi_T(d(f, f_0)|X^{(T)} \geq M_T \epsilon_T | X^{(T)}) \to 0$ in $F_{\kappa_0}^{(T)}$-probability for every sequence $M_T \to \infty$, then the sequence ϵ_T is called the posterior contraction rate. If the assertion is true for a constant $M_T = M$, then the corresponding contraction rate becomes slightly stronger.

Note that for two densities f_0, f characterized by κ_0 and κ respectively, the Kullback-Leibler divergences are given by

$$KL(\kappa_0, \kappa) = \int f_0 \log \frac{f_0}{f} = E_{\kappa_0} \left[\log \frac{\mathbb{P}_{Q_{\kappa_0}}(X_0) \prod_{t=1}^T \mathbb{P}_{\kappa_0}(X_t | \mathcal{F}_{t-1}, \lambda_0)}{\mathbb{P}_{Q_\kappa}(X_0) \prod_{t=1}^T \mathbb{P}_\kappa(X_t | \mathcal{F}_{t-1}, \lambda_0)} \right].$$

Assume that there exists a sieve in parameter space such that we have $\Pi(W_T^c) \leq \exp(-(C_T + 2)T\epsilon_T^2)$ and tests χ_T such that

$$\mathbb{E}_{\kappa_0}(\chi_T) \leq e^{-L_T T \epsilon_T^2/2} \qquad \sup_{\kappa \in W_T : d^2(f, f_0) > L_T \epsilon_T^2} \mathbb{E}_\kappa(1 - \chi_T) \lesssim e^{-L_T T \epsilon_T^2}$$

for some $L_T > C_T + 2$.

Say $U_T = \{f : d^2(f, f_0) \leq L_T \epsilon_T^2\}$ and also define $S_T = \{\int \frac{f(X^T)}{f_0(X^T)} d\Pi(\kappa) \geq \Pi_T(\frac{1}{T} KL(\kappa_0, \kappa) < \epsilon_T) \exp(-C_T T \epsilon_T^2)\}$. We can bound the posterior probability from above by,

$$\Pi_T(d(f, f_0) \geq M_T \epsilon_T | X^{(T)}) \tag{8.1}$$

$$\leq \chi_T + (1 - \chi_T) \frac{\int_{U_T^c} f(X^T) d\Pi(\kappa)}{\int f(X^{(T)}) d\Pi(\kappa)}$$

$$= \chi_T + (1 - \chi_T) \frac{\int_{U_T^c} \frac{f(X^{(T)})}{f_0(X^{(T)})} d\Pi(\kappa)}{\int \frac{f(X^{(T)})}{f_0(X^{(T)})} d\Pi(\kappa)}$$

$$\leq \chi_T + \mathbb{K}\{S_T^c\} + (1 - \chi_T) \frac{\int_{U_T^c} \frac{f(X^{(T)})}{f_0(X^{(T)})} d\Pi(\kappa)}{\exp(-C_T T \epsilon_T^2) \Pi_T\{\frac{1}{T} KL(\kappa_0, \kappa) < \epsilon_T\}}$$

$$\leq \chi_T + \mathbb{K}\{S_T^c\} + \frac{\exp(C_T T \epsilon_T^2)}{\Pi_T\{\frac{1}{T} KL(\kappa_0, \kappa) < \epsilon_T\}} (1 - \chi_T) \int_{U_T^c} \frac{f(X^{(T)})}{f_0(X^{(T)})} d\Pi(\kappa) \tag{8.2}$$

Taking expectation with respect to κ_0, first term goes to zero by construction of χ_T. The second term $\mathbb{E}_{\kappa_0} \mathbb{K}\{S_T^c\}$ goes to zero due to Lemma 8.21 of [25] for any sequence $C_T \to \infty$. We would require that $\Pi_T\{\frac{1}{T} KL(\kappa_0, \kappa) < \epsilon_T\} \geq \exp(-T \epsilon_T^2)$. Then for the third term,

$$\mathbb{E}_{\kappa_0} \exp((C_T + 1) T \epsilon_T^2)(1 - \chi_T) \frac{\int_{U_T^c} f(X^{(T)})}{f_0(X^{(T)})} d\Pi(\kappa)$$

$$= \exp((C_T + 1) T \epsilon_T^2) \int_{U_T^c} f(X^{(T)})(1 - \chi_T) d\Pi(\kappa)$$

$$\leq \exp(C_T + 1) T \epsilon_T^2) \left[\int_{U_T^c \cap W_T} f(X^{(T)})(1 - \chi_T) d\Pi(\kappa) + \Pi(W_T^c) \right]$$

$$= \exp((C_T + 1) T \epsilon_T^2) \left[\sup_{\kappa \in W_T : d^2(f, f_0) > L_T \epsilon_T^2} \mathbb{E}_\kappa (1 - \chi_T) + \Pi(W_T^c) \right]$$

$$\lesssim \exp(-T \epsilon_T^2). \tag{8.3}$$

Thus we need three things to calculate posterior contraction rate.

(i) (Prior mass Condition) We need $\Pi_T\{\frac{1}{T} KL(\kappa_0, \kappa) < \epsilon_T\} \geq \exp(-T \epsilon_T^2)$,
(ii) (Sieve) construct the sieve W_T such that $\Pi(W_T^c) \leq \exp(-(C_T + 2) T \epsilon_T^2)$ and
(iii) (Test construction) exponentially consistent tests χ_T.

We first study the contraction properties with respect to $d^2(f, f_0) = r_T^2(f, f_0) = -\frac{1}{T} \log \int \sqrt{f f_0}$ and then show that the same rate holds for average Hellinger

$\frac{1}{T}d_H^2(f, f_0)$. Note that L_T can be taken as $L_T = M_T^2$. With the above general structure, we now proceed to prove individual theorems focusing on the TVBARC and the TVINGARCH cases.

8.2. Proof of Theorem 1

For the sake of technical convenience we show our proof for time-varying AR model with 1 lag only. All the proofs go through for higher lags with the same technical tools.

8.2.1. KL Support

The likelihood based on the parameter space κ is given, $\mathbb{P}_\kappa(X_0)\prod_{t=1}^T \mathbb{P}_\kappa(X_t|X_{t-1})$. Let $Q_{\kappa,t}(X_t)$ be the distribution of X_t with parameter space κ.

We have

$$
\begin{aligned}
R &= \log \frac{\prod_{t=1}^T \mathbb{P}_{\kappa_0}(X_t|\mathcal{F}_{t-1}, \lambda_0)}{\prod_{t=1}^T \mathbb{P}_\kappa(X_t|\mathcal{F}_{t-1}, \lambda_0)} \\
&= \sum_{t=1}^T [-\{\mu_0(t/T) - \mu(t/T)\} - \{a_{01}(t/T) - a_1(t/T)\}X_{t-1} \\
&\quad + X_t\{\log(\mu_0(t/T) + a_{01}(t/T)X_{t-1}) \\
&\quad - \log(\mu(t/T) + a_1(t/T)X_{t-1})\}]
\end{aligned}
\tag{8.4}
$$

Then $KL(\kappa_0, \kappa) = \mathbb{E}_{\kappa_0}(R)$. We have in light of MVT,

$$
|R| \le \sum_{t=1}^T [|\mu(t/T) - \mu_0(t/T)| + |a_1(t/T) - a_{01}(t/T)|X_{t-1}
\tag{8.5}
$$

$$
\begin{aligned}
&+ \frac{X_t}{\mu_*(t/T) + a_{1*}(t/T)X_{t-1}}\{|\mu(t/T) - \mu_0(t/T)| + |a_1(t/T) - a_{01}(t/T)|X_{t-1}\}] \\
&\le T\|\mu - \mu_0\|_\infty + \|a_1 - a_{01}\|_\infty \sum_t X_{t-1} \\
&+ \|\mu - \mu_0\|_\infty/\rho \sum_t X_t + \|a_1 - a_{01}\|_\infty/\rho \sum_t X_t,
\end{aligned}
\tag{8.6}
$$

under the assumption that $\kappa(\cdot) = (\mu(\cdot), a_1(\cdot), b_1(\cdot))$ and $\kappa_0(\cdot) = (\mu_0(\cdot), a_{10}(\cdot), b_{10}(\cdot))$ are close and also κ_* is close to both and also in conjunction with Assumption (A.3) to imply $\inf_t a_{1*}(t/T) > \rho$ and Assumption (A.2) which implies $\mathbb{E}(X_t) < M_X$. Then for the first term we use the bound $\mu_*(t/T) + a_{1*}(t/T)X_{t-1} > \rho$ and for the second term the bound $\mu_*(t/T) + a_{1*}(t/T)X_{t-1} > \rho X_{t-1}$ is used to have $\frac{|\mu(t/T) - \mu_0(t/T)| + |a(t/T) - a(t/T)|X_{t-1}}{\mu_*(t/T) + a_{1*}(t/T)X_{t-1}} \le \|\mu - \mu_0\|_\infty/\rho + \|a_1 - a_{01}\|_\infty/\rho$ for all t. Thus,

$$
\frac{1}{T}\mathbb{E}(R) \lesssim \|\mu - \mu_0\|_\infty + \|a_1 - a_{01}\|_\infty.
\tag{8.7}
$$

8.2.2. *Posterior contraction in terms of average negative log-affinity*

In this section, we focus on the requirements to calculate posterior contraction rate as in Section 8.1.We first show posterior consistency in terms of average negative log-affinity which is defined as $r_T^2(f_1, f_2) = -\frac{1}{T} \log \int f_1^{1/2} f_2^{1/2}$ between f_1 and f_2. Here, we have $f_1 = \prod_{i=1}^{T} P_{\kappa_1}(X_i|X_{i-1})$. Then we show that, having $r_T^2(f_1, f_0) \lesssim \epsilon_n^2$ implies that our distance metric $d_{2,T}^2(f_1, f_0) \lesssim \epsilon_n^2$.

Proceeding with the rest of the proof of Theorem 1, we use the results of B-Splines, $\|\mu - \mu_0\|_\infty \leq \|\alpha - \alpha_0\|_\infty$, where $\alpha = \{\alpha_j\}$ and $\|a_1 - a_{10}\|_\infty \leq \|\gamma - \gamma_0\|_\infty$, where $\gamma_j = \theta_{1j}M_1$, such that $\gamma_j < 1$. The Hölder smooth functions with regularity ι can be approximated uniformly up to order $K^{-\iota}$ with K many B-splines. Thus we have $\epsilon_T \gtrsim \max\{K_{1T}^{-\iota_1}, K_{2T}^{-\iota_2}\}$.

We need to provide a lower bound to the prior probability as required by (i). We have the result (8.7) and the prior probabilities $\Pi(\|\alpha - \alpha_0\|_\infty \lesssim \epsilon_T, \|\gamma - \gamma_0\|_\infty \lesssim \epsilon_T) \gtrsim \epsilon_T^{K_{1T}+K_{2T}}$ based on the discussion of A2 from [44]. The rate of contraction cannot be better than the parametric rate $T^{-1/2}$, and so $\log(1/\epsilon_T) \lesssim \log T$. Thus (i) requires that in terms of pre-rate $\bar{\epsilon}_T$, we need $(K_{1T}+K_{2T})\log T \lesssim T\bar{\epsilon}_T^2$.

In our problem, we consider following sieve as required by (ii)

$$W_T = \{K_1, K_2, \alpha, \gamma : K_1 \leq K_{1T}, K_2 \leq K_{2T}, \|\alpha\|_\infty \leq A_T, \min(\alpha, \gamma) > \rho_T,$$
$$\gamma \leq 1 - A_T/B_T, \lambda_0 \leq B_T, A_T < B_T\}, \tag{8.8}$$

where A_T, B_T are at least polynomial in T and λ_0 is the mean of X_0 and $K_T = \max\{K_{1T}, K_{2T}\}$. We take $\rho_T \asymp T^{-a}$ with $a < 1$, $A_T \asymp T^{a_1}, B_T \asymp T^{a_2}$ with $a_2 > a_1$ for technical need. Note that, for $\kappa \in W_T$, we have $\mathbb{E}_\kappa(X_t) < B_T$. We need to choose these bounds carefully so that we have $\Pi(W_T^c) \leq \exp(-(1 + C_1)T\epsilon_T^2)$, which depend on tail properties of the prior. We have, $\Pi(W_T^c) = \Pi[K_1 > K_{1T}, K_2 > K_{2T}, \alpha_{K_{1T}} \in \{x : \inf x > \rho_T, \sup x < A_T, \gamma_{K_{2T}} \notin \{x : \inf x > \rho_T, \sup x < 1 - \frac{A_T}{B_T}\}, \lambda_0 > B_T]$.

Hence we have, $\Pi(W_T^c) \leq \Pi(K_1 > K_{1T}) + \Pi(K_2 > K_{2T}) + \Pi\{\alpha_{K_{1T}} \notin [\rho_T, A_T]^{K_{1T}}\} + \Pi\{\gamma_{K_{2T}} \notin \left[\rho_T, 1 - \frac{A_T}{B_T}\right]^{K_{2T}}\} + \Pi\{\lambda_0 > B_T\}$ where $\alpha_{K_{1T}}$ is the vector of full set of coefficients of length K_{1T} and $\gamma_{K_{2T}}$ is the vector of coefficients of length K_{2T}. The quantity $\Pi[\alpha_{K_{1T}} \notin [\rho_T, A_T]^{K_{1T}}$ can be further upper bounded by $K_{1T}\Pi(\alpha_1 \notin [\rho_T, A_T])) \leq K_{1T} \exp\{-R_1 T^{a_3}\}$, for some constant $R_1, a_3 > 0$ which can be verified from the discussion of the assumption A.2 of [44] for our choice of prior which exponential. On the other hand, $\Pi\{\gamma_{K_{2T}} \notin \left[\rho_T, 1 - \frac{A_T}{B_T}\right]^{K_{2T}} \leq K_{2T}\Pi(\gamma_1 \notin [\rho_T, 1 - \frac{A_T}{B_T}]) \leq K_{2T} \exp\{-R_2 T^{a_4}\}$ for some constant $R_2, a_4 > 0$ which can be verified from the proof of [42]. The inverse-gamma prior of λ_0 has exponential tail similar to α_1 and thus can be ignored as K_{1T} grows with T. Since $B_T > A_T$, the tail of λ_0 can be upper bounded by tail of α_1

Hence, $\Pi(W_T^c) \lesssim F_1(K_{1T}) + F_2(K_{2T}) + (K_{1T} + K_{2T}) \exp\{-RT^{a_5}\}$. The two functions F_1 and F_2 in the last expression stand for the tail probabilities of

the prior of K_1 and K_2. We can calculate their asymptotic order as, $F_1(x) = \Pi(K_1 > x) \asymp \exp\{-x(\log x)^{b_{13}}\}$ and $F_2(x) = \Pi(K_2 > x) \asymp \exp\{-x(\log x)^{b_{23}}\}$. We need $\Pi(W_T^c) \lesssim \exp\{-(1 + C_T)T\epsilon_T^2\}$. Hence, we calculate pre-rate from the following equation for some sequence $H_T \to \infty$,

$$K_{1T}(\log T)^{b_{13}} + K_{2T}(\log T)^{b_{23}} \gtrsim H_T T\bar{\epsilon}_T^2, \quad \log(K_{1T} + K_{2T}) + H_T T\bar{\epsilon}_T^2 \lesssim T^{a_5}. \tag{8.9}$$

Now, we construct test χ_T such that

$$\mathbb{E}_{\kappa_0}(\chi_T) \leq e^{-L_T T\epsilon_T^2/2} \qquad \sup_{\kappa \in W_T : r_T^2(\kappa,\kappa_0) > L_T \epsilon_T^2} \mathbb{E}_{\kappa}(1 - \chi_T) \lesssim e^{-L_T T\epsilon_T^2}$$

for some $L_T > C_T + 2$.

To construct the test as required in (iii), we first construct the test for point alternative $H_0 : \kappa = \kappa_0$ vs $H_1 : \kappa = \kappa_1$. The most powerful test for such problem is Neyman-Pearson test $\phi_{1T} = \mathbb{1}\{f_1/f_0 \geq 1\}$. For $r_T^2 > L_T \epsilon_T^2$, we have

$$\mathbb{E}_{\kappa_0}\phi_{1T} = \mathbb{E}_{\kappa_0}(\sqrt{f_1/f_0} \geq 1) \leq \int \sqrt{f_1 f_0} \leq \exp(-L_T T\epsilon_T^2),$$

$$\mathbb{E}_{\kappa_1}(1 - \phi_{1T}) = \mathbb{E}_{\kappa_1}(\sqrt{f_0/f_1} \geq 1) \leq \int \sqrt{f_0 f_1} \leq \exp(-L_T T\epsilon_T^2).$$

It is natural to have a neighborhood around κ_1 such that the Type II error remains exponentially small for all the alternatives in that neighborhood under the test function ϕ_{1T}. By Cauchy-Schwarz inequality, we can write that

$$\mathbb{E}_{\kappa}(1 - \phi_{1T}) \leq \{\mathbb{E}_{\kappa_1}(1 - \phi_{1T})\}^{1/2}\{\mathbb{E}_{\kappa_1}(f/f_1)^2\}^{1/2}.$$

In the above expression, the first factor is already exponentially decaying. The second factor can be allowed to grow at most of order $e^{cT\epsilon_T^2}$ for some positive small constant c. We show that $\mathbb{E}_{\kappa_1}(f/f_1)^2$ is bounded for every κ such that

$$\|\mu - \mu_1\|_\infty \leq \frac{\sqrt{\rho_T}}{\sqrt{T}}, \|a - a_1\|_\infty \leq \frac{\sqrt{\rho_T}}{\sqrt{TB_T}}.$$

We have, in the light of AM-GM inequality,

$$\mathbb{E}_{\kappa_1}(f/f_1)^2 = \int \frac{f^2}{f_1^2}f_1 = \int \frac{f}{f_1}f = \mathbb{E}_\kappa \frac{f}{f_1} = \mathbb{E}_\kappa \prod_{t=1}^T \frac{f(X_t|X_{t-1})}{f_1(X_t|X_{t-1})}$$

$$\leq \frac{1}{T}\sum_{t=1}^T \mathbb{E}_\kappa \left(\frac{f(X_t|X_{t-1})}{f_1(X_t|X_{t-1})}\right)^T$$

Towards uniformly bounding the summand in the above display, we write

$$\mathbb{E}_\kappa \left(\frac{f(X_t|X_{t-1})}{f_1(X_t|X_{t-1})}\right)^T$$

$$= \mathbb{E}_{X_{t-1},\kappa} \sum_{X_t=0}^{\infty} \frac{\{f(X_t|X_{t-1})\}^T}{\{f_1(X_t|X_{t-1})\}^T} f(X_t|X_{t-1})$$

$$= \mathbb{E}_{X_{t-1},\kappa} \exp[-T(\lambda-\lambda_1)-\lambda] \sum_{X_t=0}^{\infty} \left(\frac{\lambda^{T+1}}{\lambda_1^T}\right)^{X_t} /X_t!$$

$$= \mathbb{E}_{X_{t-1},\kappa} \exp[-T(\lambda-\lambda_1)-\lambda+\frac{\lambda^{T+1}}{\lambda_1^T}]$$

$$= \mathbb{E}_{X_{t-1},\kappa} \exp\Bigg[-T\{\mu(t/T)-\mu_1(t/T)\}-T\{a_1(t/T)-a_{11}(t/T)\}X_{t-1}$$

$$-\mu(t/T)-a_1(t/T)X_{t-1}+\frac{(\mu(t/T)+a_1(t/T)X_{t-1})^{T+1}}{(\mu_1(t/T)+a_{11}(t/T)X_{t-1})^T}\Bigg]. \quad (8.10)$$

where, $\lambda=\mu(t/T)+a_1(t/T)X_{t-1}$, $\lambda_1=\mu_1(t/T)+a_{11}(t/T)X_{t-1}$ and $\mathbb{E}_{X_{t-1},\kappa}$ denotes unconditional expectation over X_{t-1} under the density f with parameter κ. Let us define $r_1=\|\mu-\mu_1\|_\infty$ and $r_2=\|a_1-a_{11}\|_\infty$

Assuming $\mu(t/T)-\mu_1(t/T)$ and $a_1(t/T)-a_{11}(t/T)$ very small, we can write

$$\left[\frac{(\mu(t/T)+a_1(t/T)X_{t-1})^{T+1}}{(\mu_1(t/T)+a_{11}(t/T)X_{t-1})^T}\right]$$

$$= \left\{1+\frac{\mu(t/T)-\mu_1(t/T)+(a_1(t/T)-a_{11}(t/T))X_{t-1}}{\mu_1(t/T)+a_{11}(t/T)X_{t-1}}\right\}^T \quad (8.11)$$

$$\times(\mu(t/T)+a_1(t/T)X_{t-1})$$

$$\approx \left\{1+T\frac{\mu(t/T)-\mu_1(t/T)+(a_1(t/T)-a_{11}(t/T))X_{t-1}}{\mu_1(t/T)+a_{11}(t/T)X_{t-1}}\right\} \quad (8.12)$$

$$\times(\mu(t/T)+a_1(t/T)X_{t-1}) \quad (8.13)$$

For the above approximation to hold, we need

$$\frac{\mu(t/T)-\mu_1(t/T)+(a_1(t/T)-a_{11}(t/T))X_{t-1}}{\mu_1(t/T)+a_{11}(t/T)X_{t-1}}$$

to be small. To verify that, observe that

$$\left|\frac{\mu(t/T)-\mu_1(t/T)+(a_1(t/T)-a_{11}(t/T))X_{t-1}}{\mu_1(t/T)+a_{11}(t/T)X_{t-1}}\right| \le \frac{r_1}{\rho_T}+\frac{r_2}{\rho_T}$$

$$= \frac{1}{\sqrt{T}\rho_T}(1+\frac{1}{\sqrt{B_T}}).$$

As we have $\rho_T=T^{-a}$ with $a<1$, it follows directly. Thus (8.10) before $\mathbb{E}_{X_{t-1},\kappa}$ applying on (8.13) becomes

$$\exp\left[\frac{[T\{\mu(\frac{t}{T})-\mu_1(\frac{t}{T})\}+T\{a_1(\frac{t}{T})-a_{11}(\frac{t}{T})\}X_{t-1}]}{\mu_1(\frac{t}{T})+a_{11}(\frac{t}{T})X_{t-1}}\right] \quad (8.14)$$

$$\times \exp\left[\frac{[\{\mu(\frac{t}{T}) - \mu_1(\frac{t}{T})\} + \{a_1(\frac{t}{T}) - a_{11}(\frac{t}{T})\}X_{t-1}]}{\mu_1(\frac{t}{T}) + a_{11}(\frac{t}{T})X_{t-1}}\right]$$

$$\le \exp[Tr_1^2/\rho_T + 2Tr_1r_2/\rho_T + Tr_2^2 X_{t-1}/\rho_T] \tag{8.15}$$

The bound in (8.15) is obtained by applying a combination of the following inequalities $\mu(t/T) + a_1(t/T)X_{t-1} > \rho_T$ or $> \rho_T X_{t-1}$, $|\mu(t/T) - \mu_1(t/T)| < r_1$ and $|a_1(t/T) - a_{11}(t/T)| < r_2$. Taking $q = Tr_2^2/\rho_T$, last part becomes $\mathbb{E}(e^{qX_{t-1}})$ after taking exectation over (8.15). We have $\mathbb{E}(e^{qX_0}) = e^{\lambda_0(e^q-1)} < e^{B_T(e^q-1)} = e^Q$ for $Q = B_T(e^q - 1) \implies (e^q - 1) = Q/B_T$, B_T is the upper bound for λ_0 in the sieve). We will show $\mathbb{E}(e^{qX_1}) < Q$ under the above choice of r_1 and r_2. Then by recursion it holds for all t. We use the result $e^q - 1 \le 2q$ for $q < 1$. With $\lambda_1(X_0) = \mu(1) + a_1(1)X_0$, we have

$$\mathbb{E}(e^{qX_1}) = \mathbb{E}(\mathbb{E}(e^{qX_1}|X_0)) = E(e^{\lambda_1(X_0)(e^q-1)}) = e^{(e^q-1)\mu(1)}e^{\lambda_0(e^{(e^q-1)a_1(1)}-1)}$$

Then choose sieve parameters such that $Qa_1/B_T = a_1(1)(e^q - 1) \le 2a_1(1)q$ is very small which is ensured as q is very small. Then $\mu(1)Q/B_T + \lambda_0(e^{Qa_1(1)/B_T} - 1) \approx Q\mu(1)/B_T + \lambda_0(\frac{Qa_1(1)}{B_T}) \le Q\{\mu(1)/B_T + a_1(1)\} < Q$ as within the sieve $\mu(1)/B_T + a_1(1) < A_T/B_T + (1 - A_T/B_T) = 1$. Hence, $\mathbb{E}(e^{qX_1}) < e^Q$. Recursively, for all t, we can show $\mathbb{E}(e^{qX_t}) < e^Q$.

Our primary goal of showing $\mathbb{E}_{\kappa_1}(f/f_1)^2 < \infty$ can be fulfilled if Q is a constant, independent of T. To ensure Q is independent of T we need $B_T(e^q-1)$ is constant. It suffices to make qB_T constant as $qB_T < B_T(e^q - 1) < 2qB_T$. Thus, for $r_2 \le \frac{\sqrt{\rho_T}}{\sqrt{TB_T}}$ and in the light of (8.15) $r_1 \le \frac{\sqrt{\rho_T}}{\sqrt{T}}$ we have $\mathbb{E}_{\kappa_1}\left(\frac{f}{f_1}\right)^2$ bounded.

The test function χ_T satisfying exponentially decaying Type I and Type II probabilities is then obtained by taking maximum over all tests ϕ_{jT}'s for each ball, having above radius. Thus $\chi_T = \max_j \phi_{jT}$. Type I and Type II probabilities are given by

$$P_0(\chi_T) \le \sum_j P_0\phi_{jT} \le D_T P_0\phi_{jT},$$

$$\sup_{\kappa \in W_T : r_T^2(\kappa,\kappa_0) > L_T\epsilon_T^2} P(1 - \chi_T) \le \exp(-TL_T\epsilon_T^2).$$

Hence, we need to show that $\log D_T \lesssim T\epsilon_T^2$, where D_T is the required number of balls of above radius needed to cover our sieve W_T. We have

$$\log D_T \le \log D(r_1, \|\alpha\|_\infty \le A_T, \min(\alpha) > \rho_T, \|\cdot\|_\infty)$$

$$+ \log D(r_2, \|\gamma\|_\infty \le 1 - \frac{A_T}{B_T}, \min(\gamma) > \rho_T, \|\cdot\|_\infty)$$

$$\le K_{1T}\log(3K_{1T}A_T/r_1) + K_{2T}\log(3K_{2T}/r_2) \tag{8.16}$$

Given our choices of A_T, B_T and ρ_T, the two radii r_1 and r_2 are some fractional polynomials in T. Thus $\log D_T \lesssim (K_{1T} + K_{2T})\log T$, which is required to be $\lesssim T\epsilon_T^2$ as in the prior mass condition due to (i).

Based on (8.9), we thus have $\bar{K}_{1T} \asymp T^{1/(2\iota_1+1)}(\log T)^{-1/(2\iota_1+1)}$, $K_{2T} \asymp T^{1/(2\iota_2+1)}(\log T)^{-1/(2\iota_2+1)}$. A pre-rate may thus be expressed as

$$\bar{\epsilon}_T = \max\left\{ T^{-\iota_1/(2\iota_1+1)}(\log T)^{\iota_1/(2\iota_1+1)}, T^{-\iota_2/(2\iota_2+1)}(\log T)^{\iota_2/(2\iota_2+1)} \right\}.$$

The actual rate will be slower that pre-rate. Now, the covering number condition, prior mass conditions and basis approximation result give us $(K_{1T} + K_{2T})\log T \lesssim T\epsilon_T^2$ and $\epsilon_T \gtrsim \max\{K_{1T}^{-\iota_1}, K_{2T}^{-\iota_2}\}$. Combining all these conditions, we would thus require that $K_{1T} \asymp T^{1/(2\iota_1+1)}(\log T)^{2\iota_1/(2\iota_1+1)-b_{13}}$, $K_{2T} \asymp T^{1/(2\iota_2+1)}(\log T)^{2\iota_2/(2\iota_2+1)-b_{23}}$. Hence we calculate the posterior contraction rate as ϵ_T equal to

$$\max\left\{ T^{-\frac{\iota_1}{2\iota_1+1)}}(\log T)^{\iota_1/(2\iota_1+1)+(1-b_{13})/2}, T^{-\frac{\iota_2}{2\iota_2+1)}}(\log T)^{\iota_2/(2\iota_2+1)+(1-b_{23})/2} \right\}.$$

8.2.3. Posterior contraction in terms of average Hellinger

We can write Reyni divergence as $r_T^2 = -\frac{1}{T}\log \int \sqrt{f_0 f_1} = -\frac{1}{T}\log \mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}$. We need to show $r_T^2 \lesssim \epsilon_T^2$ implies that $d_{2,T}^2(\kappa_0, \kappa) \lesssim \epsilon_T^2$ as ϵ_T goes to zero.

If $r_T^2 \leq \epsilon_T^2$, we have $\left(\mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}\right)^{-1/T} \leq \exp(\epsilon_T^2)$ which implies for small ϵ_T^2, we have $\left(\mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}\right)^{1/T} \geq 1 - \epsilon_T^2$. By Cauchy-Squarz inequality $\left(\int \sqrt{f_0 f_1}\right)^2 \leq \int f_0 \int f = 1$. Thus we have,

$$1 - \epsilon_T^2 \leq \left(\mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}\right)^{1/T} \leq 1,$$

Since $d_H^2(f_1, f_0) = 2(1 - \mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}})$

$$\left(\mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}\right)^{1/T} = \left\{1 - \left(1 - \mathbb{E}_{\kappa_0}\sqrt{\frac{f_1}{f_0}}\right)\right\}^{1/T} \approx 1 - \frac{1}{2T}d_H^2(f_1, f_0).$$

Thus $\frac{1}{T}d_H^2(f_1, f_0) \lesssim \epsilon_T^2$. Thus it is consistent under average Hellinger distance.

8.3. Proof of Theorem 2

The proof will follow similar path as in the previous section. Thus we just specifically touch upon the parts that require different treatment. We can rewrite history of the INGARCH process as $\{\mathcal{F}_{t-1}, \mathcal{G}_{t-1}\} = \{\mathcal{F}_{t-1}, \lambda_0\}$. For the INGARCH case, the likelihood based on the parameter space κ is different from above and is given by, $\mathbb{P}_{\psi_0}(X_0, \lambda_0)\prod_{t=1}^{T}\mathbb{P}_{\psi}(X_t|\mathcal{F}_{t-1}, \lambda_0)$. Since all the steps are similar for the proof of Theorem 2, we only provide an outline. First to bound

KL by the sup-norm distances among functions, we need to tackle $|b_{11}(t/T)\lambda_{1t} - b_{01}(t/T)\lambda_{0t}|$. For this term we have

$$|b_{11}(t/T)\lambda_{1t} - b_{01}(t/T)\lambda_{0t}| \leq \lambda_{0t}\|b_{11} - b_{01}\|_\infty + \max_t b_{11}(t)|\lambda_{1t} - \lambda_{0t}|. \quad (8.17)$$

When ψ_1 is near ψ_0, we have for all t

$$|\lambda_{1t} - \lambda_{0t}| \leq \|\mu_1 - \mu_0\|_\infty + X_{t-1}\|a_{11} - a_{01}\|_\infty + (1 - \frac{M_\mu}{M_X})|\lambda_{1,t-1} - \lambda_{0,t-1}|$$
$$+ \lambda_{0,t-1}|b_{11} - b_{01}|_\infty$$

as we can upper bound $\max_t b_{11}(t)$ by $(1 - \frac{M_\mu}{M_X})$ since ψ_1 is close to ψ_0. We have

$$\sum_{t=1}^{T-1} \frac{M_\mu}{M_X}|\lambda_{1t} - \lambda_{0t}| + |\lambda_{1T} - \lambda_{0T}|$$
$$\leq T\|\mu_1 - \mu_0\|_\infty + \sum_t X_{t-1}\|a_{11} - a_{01}\|_\infty + (1 - \frac{M_\mu}{M_X})|\lambda_{10} - \lambda_{00}|$$
$$+ \sum_t \lambda_{0,t-1}|b_{11} - b_{01}|_\infty$$

As $M_\mu < M_X$,

$$\sum_{t=1}^{T} |\lambda_{1t} - \lambda_{0t}| \leq \frac{M_X}{M_\mu}\{T\|\mu_1 - \mu_0\|_\infty + \sum_t X_{t-1}\|a_{11} - a_{01}\|_\infty$$
$$+ (1 - \frac{M_\mu}{M_X})|\lambda_{10} - \lambda_{00}| + \sum_t \lambda_{0,t-1}|b_{11} - b_{01}|_\infty\}.$$

which implies,

$$\mathbb{E}\sum_{t=1}^{T} |\lambda_{1t} - \lambda_{0t}| \leq \frac{M_X}{M_\mu}\{T\|\mu_1 - \mu_0\|_\infty + TM_X\|a_{11} - a_{01}\|_\infty$$
$$+ (1 - \frac{M_\mu}{M_X})|\lambda_{10} - \lambda_{00}| + TM_X|b_{11} - b_{01}|_\infty\}. \quad (8.18)$$

Using the definition of R as in (8.4), we have

$$|R| \leq \sum_{t=1}^{T} \left[|\lambda_{1t} - \lambda_{0t}| + \frac{X_t}{\mu_*(t/T) + a_{1*}(t/T)X_{t-1}}|\lambda_{1t} - \lambda_{0t}|\right] \quad (8.19)$$

The first part follows directly. For the second part as ψ_1 and ψ_0 are close

$$\sum_t \mathbb{E}\left(\mathbb{E}\left(\frac{X_t}{\lambda_{*t}}|\lambda_{1t} - \lambda_{0t}| \mid \mathcal{F}_t\right)\right) \leq \sum_t \frac{M_X}{\rho}\mathbb{E}(|\lambda_{1t} - \lambda_{0t}|)$$
$$= \frac{M_X}{\rho}\mathbb{E}(\sum_t |\lambda_{1t} - \lambda_{0t}|).$$

Thus $\mathbb{E}(\frac{R}{T})$ can again be bounded by sup-norm differences in functions as before and $|\lambda_{10} - \lambda_{00}|$ using (8.18). Next, we need to construct a sieve and construct tests. We consider similar sieve

$$W_T = \{K_1, K_2, K_3\alpha, \gamma_1, \gamma_2 : K_1 \le K_{1T}, K_2 \le K_{2T}, K_3 \le K_{3T}, \|\alpha\|_\infty \le A_T,$$
$$\min(\alpha, \gamma_1, \gamma_2) > \rho_T, \max \gamma_1 + \max \gamma_2 \le 1 - A_T/B_T, \lambda_0 \le B_T\},$$
(8.20)

as in the previous problem. Within the sieve, we have $\mathbb{E}_{\mathbb{E}_{t-1}}(\max(X_t, \lambda_t)) < B_T$. Here the extra terms such as K_3 stand for number of basis in $b_1(t)$ and the vectors γ_1 and γ_2 correspond to the B-spline coefficients of the functions $a_1(t)$ and $b_1(t)$ respectively. Also note that we now have a lower bound for A_T for technical need. We take $\rho_T \approx T^{-a}$ with $a < 1$, $A_T = B_T(1 - \exp(\log T/T)\rho_T)$, $B_T \approx T^{a_2}$ for sufficiently large T such that $\exp(\log T/T)\rho_T < 1$. Within the sieve again we use a variant of above inequality. Note that within the sieve $\mathbb{E}(X_t) \le B_T$ and $\mathbb{E}(\lambda_t) \le B_T$.

We have that,

$$|\lambda_{1t} - \lambda_t| \le \|\mu_1 - \mu\|_\infty + X_{t-1}\|a_{11} - a_1\|_\infty + (1 - \frac{A_T}{B_T})|\lambda_{1,t-1} - \lambda_{t-1}|$$
$$+ \lambda_{t-1}|b_{11} - b_{01}|_\infty$$
(8.21)

and also,

$$\frac{|\lambda_t - \lambda_{1t}|}{\lambda_t} \le \frac{1}{\rho_T}\|\mu - \mu_1\|_\infty + \frac{1}{\rho_T}\|a_1 - a_{11}\|_\infty + \frac{1 - A_T/B_T}{\rho_T}\frac{|\lambda_{t-1} - \lambda_{1,t-1}|}{\lambda_{t-1}}$$
$$+ \frac{1}{\rho_T}\|b_1 - b_{11}\|_\infty$$

By recursion,

$$\frac{|\lambda_t - \lambda_{1t}|}{\lambda_t} \le \frac{G_T^t - 1}{(G_T - 1)\rho_T}[\|\mu - \mu_1\|_\infty + \|a_1 - a_{11}\|_\infty + \|b_1 - b_{11}\|_\infty]$$
$$+ \frac{G_T^{t-1}}{\rho_T}|\lambda_0 - \lambda_{01}|,$$
(8.22)

where $G_T = \frac{1 - A_T/B_T}{\rho_T} > 1$. Since RHS is increasing in t, we only need to find a bound for $t = T$. If A_T, B_T and ρ_T are chosen in such a way that $G_T \asymp \exp(\log T/T)$. This entails $G_T^T \asymp T$. Based on that r_1, r_2, r_3 and r_4 can be chosen. For sufficiently large $T(> 1/a)$ we have $(1 - \exp(\log T/T)\rho_T) < 1$. Let us assume that $\|\mu - \mu_1\|_\infty = r_1, \|a - a_1\|_\infty = r_2, \|b - b_1\|_\infty = r_3, |\lambda_0 - \lambda_{01}| = r_4$. Then for $r_i \le \frac{\rho_T}{T^{1+a_3}}$, we have that $\frac{|\lambda_t - \lambda_{1t}|}{\lambda_t} \le 1/T^{a_3}$ for all t with $a_3 > 0$. The choice of a_3 is shown later. Next goal is to find the radii for which $\mathbb{E}_\psi\left(\frac{f}{f_1}\right)^2$ is bounded.

Similar steps as before first give us $\mathbb{E}_{\psi_1}\left(\frac{f}{f_1}\right)^2 \le \frac{1}{T}\sum_{t=1} \mathbb{E}_\psi\left(\frac{f(X_t|\mathcal{F}_{t-1}, \lambda_0)}{f_1(X_t|\mathcal{F}_{t-1}, \lambda_0)}\right)^T$

and then the following,

$$\mathbb{E}_\psi \left(\frac{f(X_t|\mathcal{F}_{t-1}, \lambda_0)}{f_1(X_t|\mathcal{F}_{t-1}, \lambda_0)} \right)^T \approx \mathbb{E}_\psi \exp \frac{T(\lambda_{1t} - \lambda_t)(\lambda_{1t} - \lambda_t)}{\lambda_t}$$

$$\leq \mathbb{E}_\psi \exp(T^{1-a^3}|\lambda_{1t} - \lambda_t|) \leq \mathbb{E}_\psi \exp \frac{\lambda_t}{T^{2a_3-1}}.$$

We have by Jensen's inequality, $\mathbb{E}_\psi \exp \left[\frac{\lambda_t}{T^{2a_3-1}} \right] \leq \mathbb{E}_\psi \exp \left[\frac{X_t}{T^{2a_3-1}} \right]$ as $\lambda_t = \mathbb{E}_\psi(X_t|\mathcal{F}_{t-1}, \lambda_0)$. We can again show by induction that within the sieve $\mathbb{E}(e^{qX_t}) < e^Q$ for some constant Q following similar argument with $q = T^{1-2a_3}$. We again need qB_T independent of T. Hence our choice for a_3 will be $a_3 = \frac{1+a_2}{2} > 1/2$. Thus q is small for sufficiently large T and hence $e^q - 1 \approx q$. We have from MGF of Poisson,

$$\mathbb{E}(e^{qX_t})$$

$$= \mathbb{E}(\exp\{\lambda_t(e^q - 1)\}) \approx \mathbb{E}(\exp(\mu(t)q + a_1(t)X_{t-1}q) + b_1(t)\lambda_{t-1}q\}$$

$$= \mathbb{E}(\mathbb{E}_{t-1}(\exp\{\mu(t)q + a_1(t)X_{t-1}q\})(\exp\{b_1(t)\mathbb{E}_{t-1}(X_{t-1})q\}))$$

$$\leq \mathbb{E}(\mathbb{E}_{t-1}(\exp\{\mu(t)q + a_1(t)X_{t-1}q\})\mathbb{E}_{t-1}(\exp\{b_1(t)X_{t-1}q\}))$$

$$\leq \mathbb{E}(\mathbb{E}_{t-1}(\exp\{\mu(t)q + a_1(t)X_{t-1}q + b_1(t)X_{t-1}q\}))$$

$$= \mathbb{E}(\exp\{\mu(t)q + (a_1(t) + b_1(t))X_{t-1}q\}), \tag{8.23}$$

by first Jensen's inequality as $\lambda_t = \mathbb{E}_\psi(X_t|\mathcal{F}_{t-1}, \lambda_0)$ and positive correlation between $\exp\{a_1(t)X_{t-1}q\}$ and $\exp\{b_1(t)X_{t-1}q\}$ under the expectation $\mathbb{E}_{t-1}$. For two positively correlated random variables Y and Z under the sample space, we have $E(YZ) > E(Y)E(Z)$. Now using this recurrence result (8.23) of $\mathbb{E}(e^{qX_t})$, we again arrive at similar type of bounds for $r_1 \leq \frac{\sqrt{\rho_T}}{\sqrt{T}}, r_2 \leq \frac{\sqrt{\rho_T}}{\sqrt{TB_T}}$ to ensure that $\mathbb{E}(e^{qX_t}) < e^Q$ for some constant Q for all t. We also need that $r_4 \asymp r_1, r_3 \asymp r_2$, where $\asymp$ means asymptotically equivalent. Finally we need $r_1 \leq \min\{\frac{\sqrt{\rho_T}}{\sqrt{T}}, \frac{\rho_T}{T^{1+a_3}}\}$ and $r_2 \leq \min\{\frac{\sqrt{\rho_T}}{\sqrt{TB_T}}, \frac{\rho_T}{T^{1+a_3}}\}$ and $r_4 \asymp r_1, r_3 \asymp r_2$. These radii are also of polynomial order in T. Rest of the pieces of the proof follow similar arguments as before.

References

[1] Ali Ahmad and Christian Francq. Poisson QMLE of count time series models. *Journal of Time Series Analysis*, 37(3):291–314, 2016. MR3512959
[2] Leila D Amorim, Jianwen Cai, Donglin Zeng, and Maurício L Barreto. Regression splines in the time-dependent coefficient rates model for recurrent event data. *Statistics in Medicine*, 27(28):5890–5906, 2008. MR2597750
[3] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2):192–225, 1974. MR0373208

[4] Clemens Biller and Ludwig Fahrmeir. Bayesian varying-coefficient models using adaptive regression splines. *Statistical Modelling*, 1(3):195–211, 2001.

[5] Atanu Biswas and Peter X-K Song. Discrete-valued ARMA processes. *Statistics & Probability Letters*, 79(17):1884–1889, 2009. MR2749936

[6] Patrick T Brandt and John T Williams. A linear Poisson autoregressive model: The Poisson AR (p) model. *Political Analysis*, 9(2):164–184, 2001.

[7] Zongwu Cai, Jianqing Fan, and Qiwei Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956, 2000. MR1804449

[8] KS Chan and Johannes Ledolter. Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, 90(429):242–252, 1995. MR1325132

[9] Rainer Dahlhaus. Fitting time series models to nonstationary processes. *Annals of Statistics*, 25(1):1–37, 1997. MR1429916

[10] Rainer Dahlhaus. A likelihood approximation for locally stationary processes. *Annals of Statistics*, 28(6):1762–1794, 2000. MR1835040

[11] Rainer Dahlhaus. Locally stationary processes. *Handbook of Statistics*, 30:351–413, 2012.

[12] Rainer Dahlhaus and Suhasini Subba Rao. Statistical inference for time-varying ARCH processes. *Annals of Statistics*, 34(3):1075–1114, 2006. MR2278352

[13] Richard A Davis, William TM Dunsmuir, and Sarah B Streett. Observation-driven models for Poisson counts. *Biometrika*, 90(4):777–790, 2003. MR2024757

[14] Richard A Davis and Thomas Mikosch. Extreme value theory for GARCH processes. In *Handbook of Financial Time Series*, pages 187–200. Springer, 2009.

[15] Maria DeYoreo and Athanasios Kottas. A Bayesian nonparametric Markovian model for non-stationary time series. *Statistics and Computing*, 27(6):1525–1538, 2017. MR3687324

[16] Jianqing Fan and Wenyang Zhang. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1(1):179, 2008. MR2425354

[17] René Ferland, Alain Latour, and Driss Oraichi. Integer-valued GARCH process. *Journal of Time Series Analysis*, 27(6):923–942, 2006. MR2328548

[18] Guillermo Ferreira, Jean P Navarrete, Francisco J Rodríguez-Cortés, and Jorge Mateu. Estimation and prediction of time-varying GARCH models through a state-space representation: a computational approach. *Journal of Statistical Computation and Simulation*, 87(12):2430–2449, 2017. MR3660364

[19] Konstantinos Fokianos, Anders Rahbek, and Dag Tjøstheim. Poisson autoregression. *Journal of the American Statistical Association*, 104(488):1430–1439, 2009. MR2596998

[20] Konstantinos Fokianos and Dag Tjøstheim. Log-linear Poisson autoregression. *Journal of Multivariate Analysis*, 102(3):563–578, 2011. MR2755016

[21] Maria Franco-Villoria, Massimo Ventrucci, and Håvard Rue. A unified view on Bayesian varying coefficient models. *Electronic Journal of Statistics*,

13(2):5334–5359, 2019. MR4047589

[22] Piotr Fryzlewicz, Theofanis Sapatinas, and Suhasini Subba Rao. Normalized least-squares estimation in time-varying ARCH models. *Annals of Statistics*, 36(2):742–786, 2008. MR2396814

[23] Subhashis Ghosal, Jayanta K Ghosh, and Aad W Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000. MR1790007

[24] Subhashis Ghosal and Aad Van Der Vaart. Convergence rates of posterior distributions for noniid observations. *Annals of Statistics*, 35(1):192–223, 2007. MR2332274

[25] Subhashis Ghosal and Aad van der Vaart. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2017. MR3587782

[26] Chong Gu and Grace Wahba. Smoothing spline ANOVA with component-wise Bayesian "confidence intervals". *Journal of Computational and Graphical Statistics*, 2(1):97–117, 1993. MR1272389

[27] Beniamino Hadj-Amar, Bärbel Finkenstädt Rand, Mark Fiecas, Francis Lévi, and Robert Huckstepp. Bayesian model search for nonstationary periodic time series. *Journal of the American Statistical Association*, 115(531):1320–1335, 2020. MR4143468

[28] Trevor Hastie and Robert Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–779, 1993. MR1229881

[29] Jianhua Z Huang and Haipeng Shen. Functional coefficient regression models for non-linear time series: a polynomial spline approach. *Scandinavian Journal of Statistics*, 31(4):515–534, 2004. MR2101537

[30] Jianhua Z Huang, Colin O Wu, and Lan Zhou. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128, 2002. MR1888349

[31] Seonghyun Jeong. Frequentist properties of Bayesian procedures for high-dimensional sparse regression. Technical report, North Carolina State University, 2019. MR4094193

[32] Sayar Karmakar. Asymptotic theory for simultaneous inference under dependence. Technical report, University of Chicago, 2018.

[33] Sayar Karmakar, Stefan Richter, and Wei Biao Wu. Simultaneous inference for time-varying models. *Journal of Econometrics*, 2021.

[34] Sayar Karmakar and Wei Biao Wu. Optimal Gaussian approximation for multiple time series. *Statistica Sinica*, 30(3):1399–1417, 2020.

[35] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of Internal Medicine*, 2020.

[36] Radford M Neal. MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2(11):2, 2011. MR2858447

[37] Bo Ning, Seonghyun Jeong, and Subhashis Ghosal. Bayesian linear regression for multivariate responses under group sparsity. *Bernoulli*, 26(3):2353–2382, 2020. MR4091112

[38] Neelabh Rohan and TV Ramanathan. Nonparametric estimation of a time-varying GARCH model. *Journal of Nonparametric Statistics*, 25(1):33–52, 2013. MR3039969

[39] Ori Rosen, David S Stoffer, and Sally Wood. Local spectral analysis via a Bayesian mixture of smoothing splines. *Journal of the American Statistical Association*, 104(485):249–262, 2009. MR2504376

[40] Ori Rosen, Sally Wood, and David S Stoffer. Adaptspec: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association*, 107(500):1575–1589, 2012. MR3036417

[41] Arkaprava Roy and David B Dunson. Nonparametric graphical model for counts. *Journal of Machine Learning Research*, 21(229):1–21, 2020. MR4209515

[42] Arkaprava Roy, Subhashis Ghosal, and Kingshuk Roy Choudhury. High dimensional single-index bayesian modeling of brain atrophy. *Bayesian Analysis*, 15(4):1229–1249, 2020. MR4171149

[43] Arkaprava Roy and Sayar Karmakar. Bayesian semiparametric time varying model for count data to study the spread of the COVID-19 cases. *arXiv preprint arXiv:2004.02281*, 2020.

[44] Weining Shen and Subhashis Ghosal. Adaptive Bayesian procedures using random series priors. *Scandinavian Journal of Statistics*, 42(4):1194–1213, 2015. MR3426318

[45] Breno Silveira de Andrade, Marinho G Andrade, and Ricardo S Ehlers. Bayesian GARMA models for count data. *Communications in Statistics: Case Studies, Data Analysis and Applications*, 1(4):192–205, 2015.

[46] Lionel Truquet. Local stationarity and time-inhomogeneous Markov chains. *Annals of Statistics*, 47(4):2023–2050, 2019. MR3953443

[47] Eunho Yang, Pradeep K Ravikumar, Genevera I Allen, and Zhandong Liu. On Poisson graphical models. In *Advances in Neural Information Processing Systems*, pages 1718–1726, 2013.

[48] Hou-Cheng Yang and Jonathan R Bradley. Bayesian inference for big spatial data using non-stationary spectral simulation. *arXiv preprint arXiv:2001.06477*, 2020. MR4253852

[49] Yu Ryan Yue, Daniel Simpson, Finn Lindgren, and Håvard Rue. Bayesian adaptive smoothing splines using stochastic differential equations. *Bayesian Analysis*, 9(2):397–424, 2014. MR3217001

[50] Scott L Zeger. A regression model for time series of counts. *Biometrika*, 75(4):621–629, 1988. MR0995107

[51] Fukang Zhu. A negative binomial integer-valued GARCH model. *Journal of Time Series Analysis*, 32(1):54–67, 2011. MR2790672

[52] Fukang Zhu. Modeling overdispersed or underdispersed count data with generalized Poisson integer-valued GARCH models. *Journal of Mathematical Analysis and Applications*, 389(1):58–71, 2012. MR2876481

[53] Fukang Zhu. Modeling time series of counts with COM-Poisson INGARCH

models. *Mathematical and Computer Modelling*, 56(9-10):191–203, 2012.
MR2950860

[54] Fukang Zhu. Zero-inflated Poisson and negative binomial integer-valued
GARCH models. *Journal of Statistical Planning and Inference*, 142(4):826–
839, 2012. MR2863870