Check for updates

A Bayes analysis of autoregressive model having functional-coefficients and its application on exchange rate data

Praveen Kumar Tripathi[1] · Manika Agarwal[1] · Satyanshu K. Upadhyay[2]

Abstract

The paper provides a Bayes analysis, based on free-knot spline technique, of the popular autoregressive model having functional-coefficients. The model was initially proposed by Chen and Tsay (1993). The technique of polynomial splines of different orders is used to approximate the functional-coefficients. A sample based approach using the Gibbs sampler algorithm with intermediate Metropolis steps is adopted to draw the posterior estimates for the parameters involved. Additionally, the technique of reversible jump Markov chain Monte Carlo is incorporated to update the location and number of knots in the polynomial spline. The paper then proceeds with the motive of obtaining both retrospective and prospective predictions based on the selected model. The complete procedure is illustrated by both simulated and a real dataset representing the exchange rate of Indian rupees relative to the US dollars.

Keywords Autoregressive model · Exchange rate · Functional-coefficients · Gibbs sampler · Markov chain Monte Carlo · Metropolis algorithm · Polynomial spline · Retrospective and prospective predictions · Reversible jump

1 Introduction

In the last few decades, the development of non-linear models has been a milestone for researchers in the time series. The non-linear models enjoy the properties where the linear models fail to explain the nature of a time series. One of the pioneer works is done by [5]. In this paper, the authors have suggested the functional-coefficients for an autoregressive (AR) model instead of the unknown AR constants. They allow the coefficients as a function of the modelled variable at some lag apart. For a sequence of time series observations $\{y_t; t = 1, 2, \ldots, T\}$, the general form of an autoregressive model having functional-coefficients (FCAR) can be written as

✉ Manika Agarwal
   manikaagarwal16@gmail.com

1   Department of Mathematics and Statistics, Banasthali Vidyapith, Banasthali, India

2   Department of Statistics, Banaras Hindu University, Varanasi, India

∆ Springer

$$y_t = \theta_0 + \sum_{i=1}^{p} g_i(y_{t-d})y_{t-i} + \epsilon_t, \tag{1}$$

where θ_0 is the intercept, $g_i(.)$'s are unknown univariate measurable functions, d and p are suitably chosen positive integers known as the delay parameter and the AR order, respectively, and ϵ's are independently and identically distributed (i.i.d.) normal variates with mean zero and a constant variance. We shall refer to the model by FCAR(p, d) throughout the paper.

Some of the pioneering works on non-linear time series models include the threshold autoregressive (TAR) model by [20, 21] and exponential autoregressive (EXPAR) model by Haggan and Ozaki [14], etc. Besides, the FCAR model is also quite popular due to its non-linear and non-parametric aspects. The model is non-parametric due to the fact that its functional-coefficients are left unspecified. The popularity of the FCAR model can be seen because of its numerous applications and considerable number of uses in the literature over the last few decades (see, for example, Chen and Tsay [5], Cai et al. [4] and Fan and Yao [11], among others). Truly speaking, the model is quite flexible and rich to cover some of the commonly used linear as well as non-linear models. For example, if $g_i(x) = a_i$, a constant, for each $i = 1, 2, \ldots, p$, the model (1) reduces to the linear AR model. Similarly, if $g_i(x) = a_i I(x \leq c) + b_i I(x > c)$ where $I(.)$ is the usual indicator function, the model (1) changes into TAR model (see Tong [21]) and if $g_i(x) = a_i + b_i exp(-c_i x_i^2)$, it reduces to the EXPAR model proposed by Haggan and Ozaki [14]. The FCAR(p, d) model has been further extended by taking the distribution of error terms other than the normal with varying variance (see, for example, Chib and Greenberg [6] and [19]).

The functional-coefficients involved in the FCAR(p, d) model can be estimated in numerous ways. Some of the non-parametric methods discussed in the literature include local linear smoothing and spline smoothing, etc. (see, for example, Fan and Gijbels [10], DiMatteo et al. [9], Grégoire and Hamrouni [13] and Lindstrom [15]). A Bayesian approach based on free-knot spline has been discussed by [9] that requires a few well-placed knots to connect the polynomials of the spline where the number of knots and their locations can be estimated from the data. Free-knot spline technique reduces a non-parametric model into an ordinary linear form once the number and location of knots are determined. After achieving the linear form of the model, posterior summaries can be easily drawn for the parameters involved in the model just by supposing the conjugate normal priors for the spline coefficients. Normally, one can start with a combination of the preassigned number and location of knots at random and can achieve the level of smoothness comfortably (see, for example, Denison et al. [7] and Wang and Wu [28]). In the true sense, the procedure should be followed for the efficiency of the algorithm and to avoid the consideration of numerous knots.

The present work proposes to consider a complete Bayesian approach to analyze the FCAR(p, d) model with the help of free-knot spline technique. Bayesian analysis of time series models with functional-coefficients is frequently exercised by researchers in time series literature. Some of the recent works include [1, 16, 19, 27, 28], among others. We, however, follow the spline technique proposed by [28], where the authors have considered the integrated form of a time series model. The uniqueness of the proposed work lies in the fact that it allows the functional-coefficients of FCAR(p, d) model to be approximated by the polynomial splines of different orders. Consequently, the different orders provide the different levels of smoothness (see Wang and Wu [28]). The randomness involved in the knots allocation, with respect to their dimension and location, could be handled by the reversible jump Markov chain Monte Carlo (MCMC) method proposed by Green [12]. The complete analysis is performed by using the Gibbs sampler algorithm with intermediate Metropolis steps to estimate the rest of the parameters of FCAR(p, d) model (see [25]).

The proposed methodology is illustrated on a simulated and real datasets of the monthly exchange rate of Indian rupees relative to the US dollars from January 2011 to December 2020. Also, the short term retrospective and prospective predictions, for the exchange rate data, are given to see the predictive performance of the model under consideration. It may, however, be noted that to forecast smoothly the stationarity of a time series model is an essential requirement. In the present case of FCAR(p, d) model, such a requirement can be fulfilled by geometric ergodicity (see, for example, [5]). Following the Theorem 1.1 and 1.2 from [5], for the constants n_i; $i = 1, 2, \ldots, p$, if the function $g_i(.)$ is bounded such that $|g_i(.)| \leq n_i$ and if all the roots of the characteristic polynomial ($\gamma^p - n_1\gamma^{p-1} - \cdots - n_p = 0$) of (1) lie inside the unit circle, the FCAR(p, d) model is said to have the geometrical ergodicity, where γ refers the eigenvalue. It is important to mention, here, that the characteristic polynomial of (1) could possibly be written by considering the functional coefficients $g_i(.)$'s as the constant coefficients of the general linear AR model or could be replaced by the utmost constant values taken by the functional coefficients (see [5]). This, however, provides a sufficient condition of ergodicity but for some of the special cases of model (1) (as discussed above), this condition becomes both necessary and sufficient.

The paper is organized as follows. Section 2, divided into several subsections, provides the complete Bayesian model formulation of the FCAR(p, d) model starting from the chosen priors to the posterior distributions. Separate subsections are provided to describe the considered MCMC sampling strategy and its complete implementation. An algorithm to update the functional-coefficients is provided separately. Since the specification of parameters p and d is a significant step in the analysis of FCAR(p, d) model, it is suggested to use the Bayesian information criterion (BIC) as the guiding principle. The BIC is briefly reviewed in Sect. 2.3 for completeness of the work. The section finally ends with a discussion on obtaining predictive samples for the intended retrospective and prospective predictions. Section 3 illustrates the proposed methodology for a simulated as well as a real dataset on the exchange rate of Indian rupees relative to the US dollars. The compatibility of the selected model is graphically shown and the predictive ability of the model is examined by the short term retrospective prediction of the exchange rates. At the end of this section, short-term prospective predictions are also provided using the selected model. A brief conclusion is given in the last section.

2 Bayesian model formulation

Consider a time series $\underline{y} : y_1, y_2, \ldots, y_T$ from the non-linear FCAR(p, d) model (1) at equally spaced time periods $t = 1, 2, \ldots, T$. It is our supposition that the negative values of y's and ϵ's have not much effect on the conclusion drawn on the basis of model (1). Since the model (1) has a dependent characteristic on its own previous values up to the lag p, we assume that $y_t = \epsilon_t = 0$ for $t \leq 0$, which results in the conditional distribution of y_t given $y_{t-1}, \ldots, y_{t-p}$ in the form

$$f(y_t | y_{t-1}, \ldots, y_{t-p}; \theta_0, \sigma^2) \propto \frac{1}{\sigma} \times \exp\left(-\frac{1}{2\sigma^2}(y_t - \theta_0 - \sum_{i=1}^{p} g_i(y_{t-d})y_{t-i})^2\right). \quad (2)$$

The corresponding likelihood function for the model (1) from its conditional density (2) can be approximated by

$$f(\underline{y}|\theta_0, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T-p^*}{2}} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p^*+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}g_i(y_{t-d})y_{t-i})^2\right), \quad (3)$$

where $p^* = \max(p, d)$. This kind of approximation is very common in time series literature (see, for example, Box and Jenkins [3] and Wang and Wu [28], among others).

Next, we approximate, for each $i = 1, 2, \ldots, p$, the functional-coefficient $g_i(.)$ by an m_i-order polynomial spline with k_i ordered interior knots $\xi_i = (\xi_{i1}, \xi_{i2}, \ldots, \xi_{ik_i})'$, such that A' shows the transpose of any arbitrary vector A. Thus,

$$g_i(x) \approx \sum_{j=1}^{K_i} B_{ij}(x)\beta_{ij} = B_i'(x)\beta_i, \qquad x \in [a, b], \qquad (4)$$

where $K_i = m_i + k_i$, $B_i(x) = (B_{i1}(x), B_{i2}(x), \ldots, B_{iK_i}(x))'$ is $K_i \times 1$ vector of B-spline bases, $\beta_i = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{iK_i})'$ is a $K_i \times 1$ vector of spline coefficients and the boundary knots are given by

$$a = \min_{1\leq t\leq T}\{y_t\} \text{ and } b = \max_{1\leq t\leq T}\{y_t\}.$$

For each functional coefficient $g_i(.)$, the B-spline bases may be recursively obtained by the general formula

$$B_l^0(x) = I(z_l \leq x < z_{l+1}), \qquad (5)$$

and

$$B_l^q(x) = \frac{x - z_l}{z_{l+q} - z_l}B_l^{q-1}(x) + \frac{z_{l+q+1} - x}{z_{l+q+1} - z_{l+1}}B_{l+1}^{q-1}(x), \qquad q \geq 1 \qquad (6)$$

where q denotes the degree of spline and z_l represents the lth knot in a spline. We allow splines of different order and, hence, with different numbers and locations of knots in the equation (4). As stated earlier, this provides flexibility when the functional-coefficients have different levels of smoothness. Using (4), (3) can be further approximated as

$$f(\underline{y}|\theta_0, \beta, k, \xi, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{T-p^*}{2}} \times \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p^*+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p}B_i'(y_{t-d})\beta_i y_{t-i})^2\right), \qquad (7)$$

where $\beta = (\beta_1', \beta_2', \ldots, \beta_p')'$, $k = (k_1, k_2, \ldots, k_p)'$ and $\xi = (\xi_1', \xi_2', \ldots, \xi_p')'$ are the vectors of spline coefficients, knots and locations, respectively.

2.1 Prior and posterior distributions

To perform a Bayesian analysis, it is essential to choose suitable prior distributions for the parameters. Such prior distributions can be informative or non-informative based on the available *a priori* information. It may be noted that prior elicitation is not the main objective of the paper and, therefore, we consider some standard priors for the parameters under consideration. To begin with, let us consider for each $i = 1, 2, \ldots, p$, the following conjugate priors for the spline coefficients β, numbers k and locations ξ of the knots (see also Wang and Wu [28]).

$$\pi_1(k_i) = \frac{\lambda_i^{k_i}}{k_i!}e^{-\lambda_i}; \qquad \lambda_i \geq 1, \qquad (8)$$

$$\pi_2(\xi_i|k_i) = \frac{k_i!}{(b-a)^{k_i}} I\{a = \xi_{i0} < \xi_{i1} < \cdots < \xi_{ik_i} < \xi_{ik_i+1} = b\}, \tag{9}$$

and

$$\beta_i|k_i, \xi_i, \sigma^2, \tau_\beta \sim N_{K_i}(0, \tau_\beta \sigma^2 I_{K_i}), \tag{10}$$

where $I\{.\}$ denotes the indicator function that takes value either zero or one, τ_β is a hyperparameter and I_{K_i} is the $K_i \times K_i$ identity matrix. We specify the inverse-gamma hyper-prior for τ_β with the two pre-specified hyperparameters r and s_τ^2 given by,

$$\pi_3(\tau_\beta) \propto \tau_\beta^{-\frac{r}{2}-1} \exp\left(\frac{-s_\tau^2}{2\tau_\beta}\right). \tag{11}$$

We have also considered the non-informative priors for the parameters θ_0 and σ^2 similar to [23] and the same are given by,

$$\pi_4(\theta_0) \propto U[-M, M]; \qquad M > 0, \tag{12}$$

and

$$\pi_5(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{13}$$

The set of hyperparameters, for each i, is $(\lambda_i, r, s_\tau^2, M)$. The prior distribution can be jointly represented as

$$\pi(\theta_0, \beta, \xi, k, \tau_\beta, \sigma^2) \propto \left\{\prod_{i=1}^p \pi_6(\beta_i|k_i, \xi_i, \sigma^2, \tau_\beta)\pi_2(\xi_i|k_i)\pi_1(k_i)\right\} \pi_3(\tau_\beta)\pi_4(\theta_0)\pi_5(\sigma^2), \tag{14}$$

where $\pi_6(\beta_i|k_i, \xi_i, \sigma^2, \tau_\beta)$ is used to denote the prior in (10). Now, the Bayes' theorem enables us to write the posterior distribution, up to proportionality, by multiplying the likelihood function (7) and the prior distributions (8) to (13) as

$$p(\Theta|\underline{y}) \propto f(\underline{y}|\Theta) \times \pi(\Theta), \tag{15}$$

where $\Theta = (\theta_0, \beta, \xi, k, \tau_\beta, \sigma^2)$. Obviously, this posterior distribution forms the basis for drawing the desired Bayesian inferences.

2.2 MCMC based sampling scheme

It can be seen that the posterior distribution (15) is not analytically tractable and, therefore, we shall propose to consider a MCMC based sampling scheme to draw the samples from the joint posterior density (15). For this purpose, we shall first calculate the conditional posterior density from (15) and then discuss the corresponding sampling scheme one by one. It may be noted that the scheme is actually the Gibbs sampler but some of the full conditionals are generated using the Metropolis algorithm or even using the updating steps offered by the reversible jump MCMC strategy. As such, the scheme can also be referred to as the hybrid Gibbs sampler with hybridization being introduced by the use of the Metropolis algorithm.

2.2.1 Sampling from the full conditionals of (β_i, k_i, ξ_i)

Combining the likelihood function (7) and the prior distributions from (8) to (10), the joint posterior of (β_i, k_i, ξ_i), $i = 1, 2, \ldots, p$, for the given remaining parameters, can be written as

$$p_1(\beta_i, k_i, \xi_i | \underline{y}, \theta_0, \beta_{-i}, k_{-i}, \xi_{-i}, \tau_\beta, \sigma^2) \propto \left(\frac{\tau_\beta^{-1/2} \lambda_i}{b-a}\right)^{k_i} \times (\sigma^2)^{-k_i/2} \exp\left(\frac{-S_i}{2\sigma^2}\right)$$
$$\times \exp\left\{\frac{-1}{2\sigma^2}(\beta_i - \hat{\beta}_i)' \Sigma_i^{-1}(\beta_i - \hat{\beta}_i)\right\},$$
(16)

where β_{-i}, k_{-i} and ξ_{-i} are obtained after removing β_i, k_i and ξ_i from the vectors β, k and ξ, respectively. We took Z_i as a $(T - p^*) \times 1$ vector with its tth component $y_{p^*+t} - \theta_0$ and X_i as a $(T - p^*) \times K_i$ matrix with its tth row $B_i'(y_{p^*+t-d}) y_{p^*+t-i}$ such that,

$$\Sigma_i = (\tau_\beta^{-1} I_{K_i} + X_i' X_i)^{-1},$$
$$\hat{\beta}_i = \Sigma_i X_i' Z_i, \tag{17}$$
$$\text{and} \quad S_i = Z_i' Z_i - \hat{\beta}_i' \Sigma_i^{-1} \hat{\beta}_i.$$

Appendix A provides details regarding the derivation of (16). Hence, the full conditional distribution of β_i can be easily obtained from the joint posterior (16) as

$$p_2(\beta_i | \underline{y}, \theta_0, \beta_{-i}, k, \xi, \tau_\beta, \sigma^2) \propto (\sigma^2)^{-K_i/2} |\Sigma_i|^{-1/2} \exp\left\{\frac{-1}{2\sigma^2}(\beta_i - \hat{\beta}_i)' \Sigma_i^{-1}(\beta_i - \hat{\beta}_i)\right\}.$$
(18)

Now, the posterior samples for β_i can be easily obtained from (18) as it follows a multivariate normal density with mean vector $\hat{\beta}_i$ and covariance matrix $\sigma^2 \Sigma_i$. Similarly, the full conditional of (k_i, ξ_i) can be jointly obtained as,

$$p_3(k_i, \xi_i | \underline{y}, \theta_0, \beta_i, k_{-i}, \xi_{-i}, \tau_\beta, \sigma^2) \propto \left(\frac{\tau_\beta^{-1/2} \lambda_i}{b-a}\right)^{k_i} \times |\Sigma_i|^{1/2} \exp\left(\frac{-S_i}{2\sigma^2}\right). \tag{19}$$

Here, the numbers k_i and ξ_i can be updated using a sample based reversible jump MCMC approach. This approach is very popular among researchers working in the Bayesian curve fitting and free-knot splines technique (see, for example, Denison et al. [7], Biller [2], DiMatteo et al. [9] and Wang and Wu [28]). The method efficiently works on the basis of a three-step procedure, which are birth step (addition), death step (deletion) and move step (relocation) (see, for example, Green [12] and Denison et al. [7]). Movements of variables under the three steps are independent of each other and can be chosen randomly with probabilities b_{k_i}, d_{k_i} and η_{k_i}, respectively, where

$$b_{k_i} = C. \min\left\{1, \frac{p(k_i + 1)}{p(k_i)}\right\}, \quad d_{k_i} = C. \min\left\{1, \frac{p(k_i - 1)}{p(k_i)}\right\}, \quad \eta_{k_i} = 1 - b_{k_i} - d_{k_i}$$
(20)

and the constant C is a tuning parameter that controls the rate at which the variables move within the three different move-steps. The recommended range for C is generally $[0, 0.5]$. For the present case, we choose it as 0.4. This value of C, also suggested in the literature (see, for example, [7]), is seen to provide a good acceptance probability. One should ensure that the

probabilities given in (20) must satisfy the following relationships: $b_{k_i} p(k_i) = d_{k_i+1} p(k_i+1)$ and in case of no change of knots, that is $k_i = 0$, we have $b_0 = 1$ and $d_0 = 0 = \eta_0$ (see, for example, [28]). Below, we provide the necessary details of three move-types one by one.

Birth step: In this move-type, we add a newly generated candidate knot from a randomly chosen sub-interval $(\xi_{ij}, \xi_{i,j+1})$. The key idea is to divide the interval (a, b) into $(k_i + 1)$ sub-intervals by means of the quantity k_i and to choose one of them at random. Next, we draw candidate value, say ϕ, uniformly from $(\xi_{ij}, \xi_{i,j+1})$ as the location of newly added knot with a jump probability

$$\frac{b_{k_i}}{k_i + 1} \sum_{j=0}^{k_i} \frac{1}{\xi_{i,j+1} - \xi_{ij}} I\{\xi_{ij} < \phi < \xi_{i,j+1}\}.$$

Death step: In this move-type, a candidate knot is randomly chosen from the set of existing knots and then deleted with a jump probability,

$$\frac{d_{k_i}}{k_i}.$$

Move step: In the move step, a candidate knot ξ_{ij} is selected uniformly from the set of existing k_i knots and a candidate location ξ_{ij}^* is generated from a distribution with mean ξ_{ij} and variance σ_m^2. A truncated normal distribution on the interval $(\xi_{i,j-1}, \xi_{i,j+1})$ seems to be a good choice for the proposal distribution (see, for example, Wang and Wu [28]). Since, for the large value of σ_m^2, the proposal density will tend to a uniform distribution over the interval $(\xi_{i,j-1}, \xi_{i,j+1})$ and, therefore, the algorithm will be similar to a standard Metropolis algorithm. Hence, the jump probability is given by

$$\frac{\eta_{k_i}}{k_i} \frac{1}{\xi_{i,j+1} - \xi_{i,j-1}}.$$

Following Green [12], the acceptance probability in each of the three steps can be defined as

$$\min(1, \text{posterior ratio} \times \text{proposal ratio}),$$

where the posterior and the proposal ratios can be obtained by the method proposed by Denison et al. [7] and DiMatteo et al. [9]. In our case, this acceptance probability can be written as

$$\min\left\{1, A\left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right)\right\}, \tag{21}$$

where Σ_i and S_i can be obtained from (17) and the components of A are given as under

$$A = \begin{cases} \frac{k_i+1}{\tau_\beta^{1/2}} \frac{\xi_{i,j+1} - \xi_{ij}}{b-a}, & k_i^* = k_i + 1 \quad \text{(Birth step)}, \\ \frac{\tau_\beta^{1/2}}{k_i} \frac{b-a}{\xi_{i,j+1} - \xi_{i,j-1}}, & k_i^* = k_i - 1 \quad \text{(Death step)}, \\ 1, & k_i^* = k_i \quad \text{(Move step)}. \end{cases} \tag{22}$$

In the above expression, k_i^* represents the number of knots in the candidate posterior distribution. It is important to mention that in the move step above, the posterior and the proposal ratios are both unity because by moving one knot to another knot any collection of the same number of knots has similar posterior probabilities and proposal distributions.

A detailed derivation of the jump probabilities as well as acceptance probabilities, in each move-type, is provided in Appendix 2. The values of hyperparameters λ_i, in Poisson priors (8), and the order of splines m_i are so chosen that the rates of acceptance for the above reversible jump MCMC sampler would be enough to avoid too many rejections and thereby to increase the efficiency of the sampler.

2.2.2 Sampling from the full conditional of τ_β

Combining the prior distribution in (10) with the distribution of hyperparameter τ_β in (11), the full conditional of τ_β can be written up to proportionality as

$$p_4(\tau_\beta|\underline{y}, \theta_0, \beta, k, \xi, \sigma^2) \propto \tau_\beta^{-\frac{r+\sum_{i=1}^{p} K_i}{2}-1} \exp\left[-\frac{1}{2\tau_\beta}\left(s_\tau^2 + \frac{\beta'\beta}{\sigma^2}\right)\right]. \tag{23}$$

Obviously, the full conditional (23) represents the well known density of inverse-gamma distribution from which the sampling can be done routinely using any inverse-gamma generating routine (see, for example, [8]).

2.2.3 Sampling from the full conditional of σ^2

The full conditional of σ^2 can be obtained by considering the likelihood in (7) and the prior in (13). The same can be written as

$$p_5(\sigma^2|\underline{y}, \theta_0, \beta, k, \xi, \tau_\beta) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{\left(T-p^*-\sum_{i=1}^{p} K_i+2\right)}{2}}$$

$$\times \exp\left(-\frac{1}{2\sigma^2}\left[\frac{\beta'\beta}{\tau_\beta} + \sum_{t=p^*+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p} B_i'(y_{t-d})\beta_i\, y_{t-i})^2\right]\right). \tag{24}$$

Using a simple transformation $\lambda = \frac{1}{\sigma^2}$, it can be seen that (24) represents the probability density function of a gamma distribution with the shape parameter $\frac{(T-p^*-\sum_{i=1}^{p} K_i)}{2}$ and the scale parameter $\frac{\left(\frac{\beta'\beta}{\tau_\beta}+\sum_{t=p^*+1}^{T}(y_t-\theta_0-\sum_{i=1}^{p} B_i'(y_{t-d})\beta_i\, y_{t-i})^2\right)}{2}$. Thus, samples of σ^2 can be easily obtained by using a gamma generating routine.

2.2.4 Sampling from the full conditional of θ_0

Next, considering the likelihood function in (7) and the prior in (12), the full conditional of θ_0 can be obtained as

$$p_6(\theta_0|\underline{y}, \beta, k, \xi, \tau_\beta, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}\sum_{t=p^*+1}^{T}(y_t - \theta_0 - \sum_{i=1}^{p} B_i'(y_{t-d})\beta_i\, y_{t-i})^2\right)$$

$$\times I_{[-M,M]}(\theta_0), \tag{25}$$

where $I(.)$ is the indicator function that takes value unity if θ_0 lies in the interval $[-M, M]$ and zero otherwise. It is to be noted that the full conditional (25) is not available in a nice closed form from the viewpoint of sample generation and, therefore, we propose the use of

Metropolis algorithm to obtain the required samples of θ_0 from (25). For this purpose, one can consider, among other choices, a univariate normal kernel as a proposal density with mean as the maximum likelihood (ML) estimate of θ_0 (say) and the standard deviation as c times the Hessian based approximation at ML estimate of θ_0. Both mean and standard deviation of normal kernel can be successively changed by using the current realization of θ_0. This is expected to improve the acceptance rate of the Metropolis chain. Here, c is a scaling constant and its value is recommended between 0.5 and 1.0 (see also [22, 26]).

Once the full conditionals are made available from the viewpoint of sample generation, one can implement the proposed MCMC based hybrid algorithm on the posterior (15) by proceeding with a single long run of the chain, among many other possibilities. Results are available which ensure that after a sufficiently large number of iterations, the generated sequence converges in distribution to a random sample from the corresponding posterior distribution and the ergodic means converge almost surely to the corresponding posterior expectations. Once the convergence is monitored, the generated values can be picked up at a fixed gap to form a random sample of the desired size. It may be noted that the gaps so chosen make the serial correlation among the generating variates negligibly small and thereby giving independent samples. The selected samples can then be used to estimate any posterior characteristics of interest. For further details of the algorithm, one may refer to [25, 26], among others.

2.2.5 Algorithm to update the functional-coefficients

For each $i = 1, 2, \ldots, p$, the functional-coefficient $g_i(.)$ can be updated as follows:

Step 1: Choose m_i and λ_i, to initialize the configuration, uniformly along the range $[a, b]$ at least $(m_i + 1)$ points away from each other.
Step 2: Set k_i equal to the number of interior knots.
Step 3: Generate random variable u from $U[0, 1]$, and choose the move type in the following manner.

(i)   if $u \leq b_{k_i}$, go to the birth step;
(ii)   if $b_{k_i} < u \leq b_{k_i} + d_{k_i}$, go to the death step;
(iii)   otherwise go to the move step.

Step 4: Sample β_i from the full conditional (18).
Step 5: Simulate τ_β from the full conditional (23).

2.3 Specification of the parameters *p* and *d*

Specification of the parameters p and d is an important step to exactly specify the proposed FCAR(p, d) model before implementing the MCMC algorithm. This is obviously equivalent to the model selection task that can be achieved by BIC, among others. The BIC is a Bayesian criterion of model selection that penalizes the model for its inherent complexity. Based on the likelihood function of an estimated model, the BIC can be defined as

$$\text{BIC} = -2log \ (\text{ML}_d) + nlog(T - p^*), \tag{26}$$

where n is the number of estimated parameters in the entertained model and ML_d is the maximized likelihood. The above criterion, proposed by [18], is based on the information

theory and formulated in the Bayesian context. In order to specify the most appropriate FCAR model, one can calculate the BIC values for different candidate models of FCAR given in (1) by choosing different values of p and d. The model corresponding to the least value of BIC is finally taken as the best candidate model among others for further analysis.

It is important to mention here that the likelihood function (3) is not easy to obtain the corresponding ML estimates in order to evaluate ML_d due to the involvement of the non-linear function $g(.)$. To evaluate the same, one can approximate the non-linear function $g(.)$ by an unknown constant, say ψ, and then calculate the ML estimates by using, say, a non-linear function minimization routine available in R software. It may be further noted that the approximation of the non-linear function $g(.)$ by an unknown constant ψ provides a likelihood similar to that of a general linear AR model that finally leads to easy evaluation of the approximate ML estimates and hence the value of ML_d. Alternatively, one can use the posterior modes as an approximation to the ML estimates provided the considered priors are not strong enough to affect the posterior distribution significantly. This latter suggestion is given in the literature by a number of authors (see, for example, [17]) and it is likely to be in the Bayesian spirit as well.

2.4 Predictive samples

For the given set of observed data $\underline{y} : y_1, y_2, \ldots, y_T$, one often wishes to obtain the next observed value, that is, y_{T+1}. This can be obtained from the model (1) once the estimated values of the parameters are made available. Truly speaking, if the estimated posterior density is symmetric, one can use the posterior mean, median or mode of the corresponding parameter as the most logical estimate. Among these estimates, the posterior mode is unconditionally used even if the estimated posterior density is non-symmetrical. Accordingly, the functional-coefficients $g_i(.)$'s can be estimated after getting the desired posterior samples of β, ξ and k as discussed in Sects. 2.2.1 and 2.2.5. The future observation y_{T+1}, given the observed informative data, can then be obtained using a normal distribution with mean

$$\mu_{T+1} = \theta_0 + \sum_{i=1}^{p} g_i(y_{t+1-d})y_{t+1-i} \tag{27}$$

and variance σ^2. Obviously, the next observation corresponding to the error term, ϵ_{T+1}, can be simulated from a normal distribution with mean zero and variance equal to the corresponding posterior estimate of σ^2 (see, for example, [24]).

3 Numerical illustration

3.1 Simulation study

To examine the empirical performance of the proposed methodology, let us proceed with a simulation study on the two simple forms of the general FCAR model, that is, FCAR(1,1) and FCAR(2,1). These models can be expressed, respectively, as

$$y_t = \theta_0 + g_1(y_{t-1})y_{t-1} + \epsilon_t, \tag{28}$$

$$y_t = \theta_0 + g_1(y_{t-1})y_{t-1} + g_2(y_{t-1})y_{t-2} + \epsilon_t. \tag{29}$$

In order to perform the simulation in the above two cases, one has to consider some arbitrary choices for the model parameters such as θ_0=0.05, $g_1(y_{t-1}) = (y_{t-1})exp(-y_{t-1}^2/2)$, $g_2(y_{t-1})$ $= -cos(1.5y_{t-1})/(y_{t-1}^2 + 1)$. Besides, the assumed error terms ϵ's are i.i.d. $N(0, 0.4^2)$. We, however, considered the other choices for σ^2 such as 0.01, 0.81 and 4; and obtained the estimated posterior densities of σ^2 corresponding to these values in Fig. 10 and 11 for the two considered models FCAR(1,1) and FCAR(2,1) respectively (see Appendix 3). In each of the two cases, we have considered replicating the simulation 100 times for a random sample of size 500 each. In each replication, the hyperparameters' values, that is, $\lambda_1 = 1, r = 1, s_\tau^2 = 1$ and $\sigma_m^2 = 10^5$ are chosen arbitrarily but approved by simulation experience and provide a good convergence rate in all the 100 replications. Such choices of hyperparameters are not completely arbitrary, rather guided by a literature survey (see, for example, [28]). Referring to Sect. 2.2.1, the value of the tuning parameter $\sigma_m^2 = 10^5$ ensures the uniformity to select the candidate knot from the interval $(\xi_{i,j-1}, \xi_{i,j+1})$ as for the higher values (more than 10^5), the posterior samples remain unaltered. It is important to note that in each replication, the ergodicity of $g_1(y_{t-1})$ and $g_2(y_{t-1})$ has been ensured numerically by imposing a bounded condition on these functions (see Appendix 4).

To estimate the functional-coefficients, which are estimated by the quadratic splines, the posterior samples are drawn from the full conditionals of (k_i, ξ_i) and β_i's using the reversible jump MCMC sampler as discussed in Sects. 2.2.1 and 2.2.2. Further, for the intercept θ_0, the hyperparameter M and the scaling constant c are assumed to be 100 and 0.6, respectively, in each of the two cases (see Sect. 2.2.4). Also, the posterior samples of σ^2 can be easily obtained from (24) using a gamma generating routine in each case.

To obtain the desired posterior estimates of the parameters of the two models, (28) and (29), the proposed MCMC scheme (see Sect. 2.2) is implemented. Under the discussed initial setup, we have considered a long run of 5K iterations of the proposed MCMC scheme after observing a smooth convergence at about 2K iterations in its each replication. For each of the two models, the final posterior estimates are obtained by considering a random sample of size 1K, in each of the MCMC replications, after ignoring the initial transient behaviour and by maintaining a gap of 3. It was noted that a gap of 3 was sufficient to provide serial correlation negligibly small. The final posterior estimates are obtained as the "average" of 100 values of 'posterior means', 'posterior medians', 'posterior modes' and the 'highest posterior density intervals' with coverage probability 0.95 (0.95 HPD). Table 1 and Table 2 are providing the final posterior estimates (an average estimate of 100 replications) separately for FCAR(1,1) and FCAR(2,1) models respectively. Figure 2 demonstrates an estimate of functional-coefficient $g_1(.)$ of the FCAR(1,1) model along with 0.95 HPD region and the true function based on the average posterior estimate, for 100 replications, corresponding to each of the data point. A similar plot for the estimates of functional-coefficients $g_1(.)$ and $g_2(.)$ of the FCAR(2,1) model along with the corresponding 0.95 HPD regions and the true functions are provided in Fig. 3. The overall posterior estimates are quite satisfactory from the viewpoint of non-linear characteristic of FCAR model and their proximity with the true values of the parameters. In a single replication of FCAR(1,1) model, the trace plots of 10 equidistant grid points, which are picked up randomly from 500 observations, are demonstrated by Fig. 4a–j. Figure 4k–l show the ergodic plot and autocorrelation function (ACF) plot, respectively, for the FCAR(1,1) model. The ergodic plot advocates the convergence of the chain at about 2K iterations whereas, the ACF plot conveys the continuous decay of the autocorrelations of sample values at lag 100. It is to be noticed that Fig. 4k–l are obtained for a randomly chosen grid point (corresponding to 7th observation in our case). The plots obtained in Fig. 4 are quite subjective and demonstrate the whole simulation procedure at just one sight. Moreover,

Table 1 Posterior summaries for the parameters of FCAR(1, 1) model corresponding to simulated dataset

Parameter	Posterior mean	Posterior median	Posterior mode	0.95 HPD	
θ_0	−0.0032	−0.0026	−0.0058	−0.0849	0.0905
σ^2	0.2462	0.2444	0.2414	0.1533	0.3400

Table 2 Posterior summaries for the parameters of FCAR(2, 1) model corresponding to simulated dataset

Parameter	Posterior mean	Posterior median	Posterior mode	0.95 HPD	
θ_0	0.0006	0.0006	0.0073	−0.0868	0.0905
σ^2	0.5133	0.5109	0.5112	0.1056	0.7837
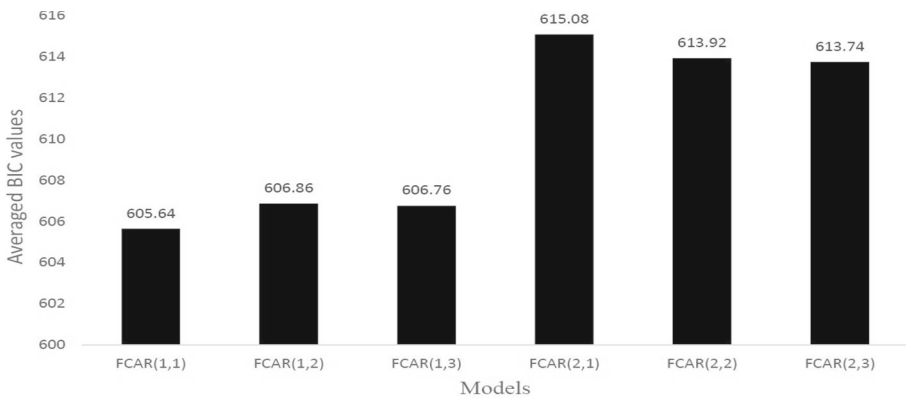
Fig. 1 Average BIC values based on the simulated dataset

in all other replications and at any arbitrarily chosen grid point, the behaviour is not going to be changed, in general. One may draw similar plots for FCAR(2,1) model as well, although the same are not provided due to paucity of space and left to the part of readers because of the ease of understanding of the proposed method.

The appropriateness of the model selection criterion is demonstrated by considering the simulated dataset of size 500 from the FCAR(1,1) model. Obviously, the BIC value should be least for a model from which the dataset is actually simulated. In order to investigate the same, let us obtain the BIC values for FCAR models with different nearby choices of p and d in each of the 100 replications. Figure 1 demonstrates the average of BIC values, for 100 replications, for different nearby choices of p and d of a general FCAR model. The height of each bar indicates the strength of BIC value for each considered model. Obviously, the BIC value is least corresponding to FCAR(1,1) model from which the dataset is actually simulated. This finding clearly indicates the appropriateness of the considered model selection criterion.

3.2 Real data example

Let us now begin with a real data example on monthly exchange rate of Indian rupees relative to the US dollars from January 2011 to December 2020. The dataset has been taken from the

Fig. 2 Estimated functional-coefficient (solid line) with 0.95 HPD interval (shaded region) of the functional coefficient $g_1(.)$ corresponding to FCAR(1,1) model. The corresponding true value of $g_1(.)$ is shown by means of dotted line
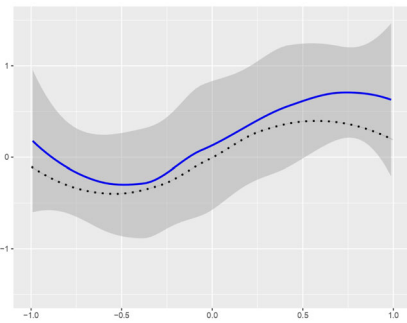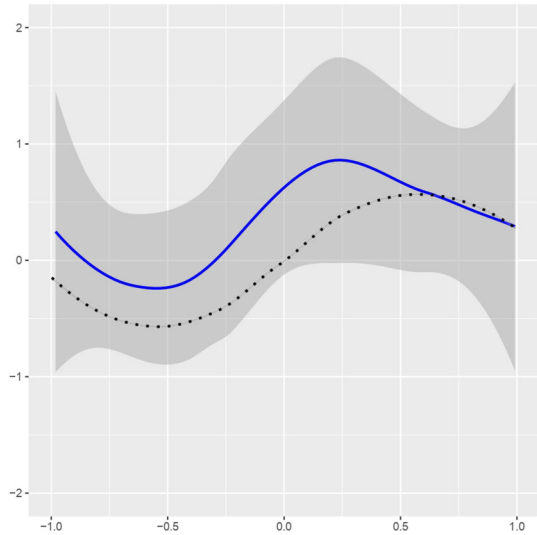
(a) Estimated functional-coefficient of the functional-coefficient $g_1(.)$
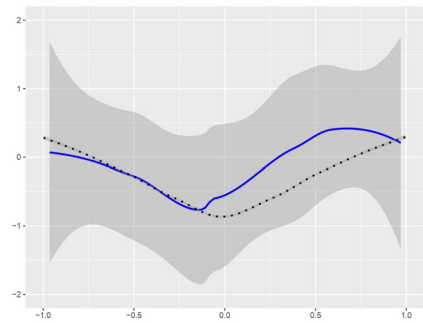
(b) Estimated functional-coefficient of the functional-coefficient $g_2(.)$

Fig. 3 Estimated functional-coefficients (solid lines) with 0.95 HPD intervals (shaded regions) of the functional-coefficients $g_1(.)$ and $g_2(.)$ corresponding to FCAR(2,1) model. The corresponding true values of $g_1(.)$ and $g_2(.)$ are shown by means of dotted lines

website of Fusion Media Limited group of the British Virgin Islands (see http://in.investing.com/currencies/usd-inr-historical-data) and is reported in Table 3.

The time series plot for the dataset reported in Table 3 is shown in Fig. 5. One can observe that the time series exhibits an increasing pattern, which clearly advocates the non-stationary behaviour of the original dataset (see Fig. 5). To remove non-stationarity from the data, we considered the first difference from the original data and plotted the same as time series in Fig. 6. It can be seen that the first difference plot exhibits nearly a stable pattern at least with regard to its mean value and, therefore, considering the first difference appears logical for further analyses of the data. Moreover, since the stability in the stationarity pattern is examined only graphically (Fig. 6), it is pertinent to consider numerical evidence as well to examine stationarity before proceeding further. For this purpose, we considered the augmented Dickey-Fuller (ADF) test on the data and noted that the test was significant at 5% level (p-value close to 0.01). Obviously, the first differenced data ensures stationarity on the basis of objective
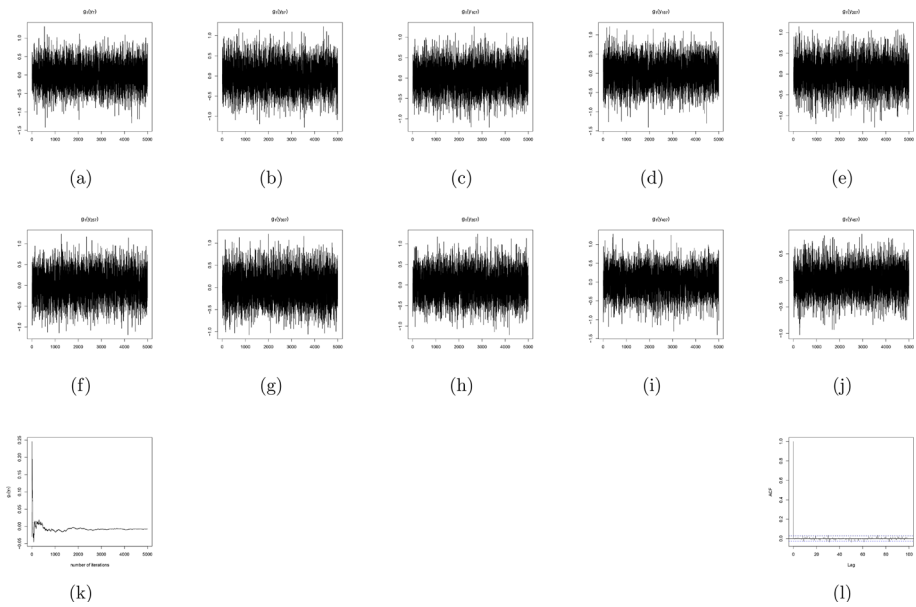
Fig. 4 (a)-(j) Trace plots of the estimated functional-coefficient $g_1(.)$ corresponding to FCAR(1,1) model for 5k iterations at 10 random grid points. (k) Ergodic plot of the estimated functional-coefficient $g_1(.)$ corresponding to FCAR(1,1) for 5K iterations. (l) ACF plot of the estimated functional-coefficient $g_1(.)$ corresponding to FCAR(1,1) model

consideration as well. For further details on the ADF test, one may refer to [23, 24], among others.

Selection of an appropriate model for further analysis is of utmost importance in order to retrieve the reliable results. We, therefore, implemented the methodology discussed in Sect. 2.3 to select an appropriate FCAR model by choosing the most appropriate values of its order p and delay parameter d. For this purpose, we begin by considering a few combinations of the AR order p and delay parameter d in the FCAR model (1) and then obtain the BIC value corresponding to each such model. We have considered 25 combinations of the FCAR model by using different values of $p(= 1, 2, 3, 4, 5)$ and $d(= 1, 2, 3, 4, 5)$. The BIC values for different d are then plotted against the variation of p in Fig. 7. The plot actually summarizes the numerical values of BIC for different considered combinations of p and d and the line combining similar plotting symbols is used for clarity of presentation and, of course, to display a trend. One may interpret that the BIC values show an increasing pattern as the order p increases for each value of the delay parameter d. Obviously, based on the values of BIC shown in Fig. 7, $p = 1$ appears to be an appropriate choice as the corresponding BIC value is least.

The choice of d, however, seems difficult on the basis of pictorial representation, since the BIC values corresponding to all the considered FCAR$(1, d)$ models for different d appear to originate from the same point (see Fig. 7). Therefore, in order to provide a clear-cut conclusion, the numerical values of BIC corresponding to FCAR$(1, d)$ models with different $d(= 1, 2, 3, 4, 5)$ are shown in Table 4. Obviously, the BIC corresponding to FCAR$(1, 1)$ model is least recommending for the use of FCAR$(1, 1)$ for the considered dataset. It may, however, be noted that other BIC values (for $d(= 2, 3, 4, 5)$) are not too far away from the value corresponding to FCAR$(1, 1)$ and, therefore, one can proclaim why not to consider

Table 3 Monthly exchange rate of Indian rupees relative to the US dollars from January 2011 to December 2020

Month	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
January	45.800	49.500	53.260	62.680	62.010	67.868	67.490	63.540	70.950	71.540
February	45.260	49.100	54.365	61.790	61.649	68.198	66.700	65.200	70.830	72.534
March	44.520	50.870	54.280	60.005	62.281	66.250	64.850	65.110	69.180	75.333
April	44.240	52.650	53.680	60.340	63.520	66.415	64.270	66.450	69.636	75.077
May	45.055	56.030	56.570	59.180	63.733	67.199	64.505	67.420	69.570	75.590
June	44.690	55.500	59.523	60.045	63.594	67.493	64.615	68.450	68.940	75.540
July	44.200	55.550	60.840	60.550	63.978	66.630	64.170	68.450	68.860	74.916
August	45.790	55.520	65.700	60.505	66.402	66.963	63.930	71.000	71.451	73.254
September	49.010	52.840	62.575	61.925	65.507	66.546	65.305	72.500	70.640	73.560
October	48.690	53.800	61.614	61.390	65.413	66.675	64.740	73.950	70.978	74.554
November	52.120	54.260	62.389	62.195	66.452	68.588	64.480	69.640	71.746	73.990
December	53.010	54.990	61.800	63.030	66.198	67.945	63.830	69.560	71.350	73.036

Fig. 5 Time series plot of exchange rate data of Indian rupees relative to the US dollars
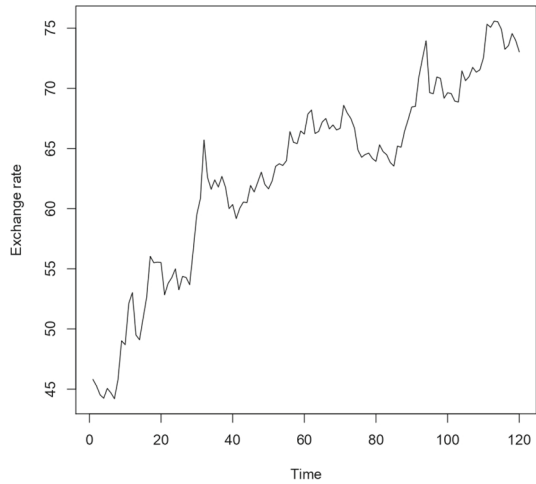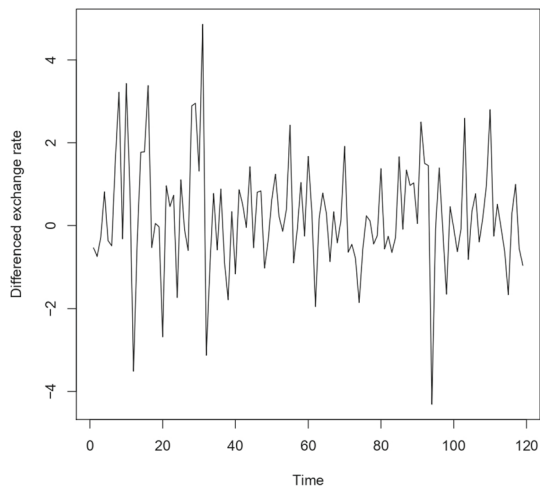

Fig. 6 First differenced time series plot of exchange rate data of Indian rupees relative to the US dollars

recommending other values of d. The answer is obvious. First, the value corresponding to FCAR(1, 1) is the least and, second, the parsimony principle never allows us to go for complicated models unless there is a real requirement. Moreover, the second term in (26) is almost unaffected by a variation in d and it is only the first term that provides a minor variation. So our final conclusion certainly supports FCAR(1, 1) model for the considered dataset.

Let us now represent the selected FCAR(1, 1) model mathematically, for the differenced dataset, as

$$\Delta y_t = \theta_0 + g_1(\Delta y_{t-1})\Delta y_{t-1} + \epsilon_t, \tag{30}$$

where Δy_t denotes the first difference data at time t. To perform the Bayesian analysis, for the selected FCAR (1, 1) model (30), we choose the quadratic spline for $g_1(.)$. It is needless to mention that all the forthcoming analyses will be performed on the first difference data where we have noticed a stationarity pattern in the time series. To begin with the Bayesian analysis, we assign numerical values to the hyperparameters defined in the Sect. 2.1 as
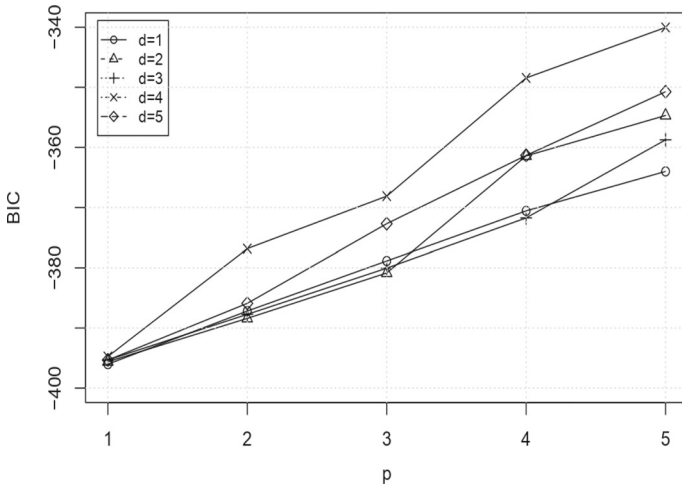
Fig. 7 Plot of BIC values for different choices of parameters p and d

Table 4 BIC values for FCAR$(1, d)$ models with different choices of d $(= 1, 2, 3, 4, 5)$

Model	FCAR(1,1)	FCAR(1,2)	FCAR(1,3)	FCAR(1,4)	FCAR(1,5)
BIC value	-396.02	-395.60	-395.32	-395.13	-394.75

$\lambda_1 = 1$, $r = 1$ and $s_\tau^2 = 1$ and $M = 100$. Of course, there is no basis in the selection of these hyperparameters except using the suggestions given by [28] and [23, 24], etc. It is to be noted that these choices of hyperparameters approximately result into weakly informative priors in the sense that the resulting priors are more or less flat in an appreciable range and, as such, most of the inferences can be regarded as driven by the likelihood function only.

With these choices of the prior hyperparameters, the MCMC implementation was done as detailed in Sect. 2.2. The value of the tuning parameter, in the move step, can be taken large enough to make the proposal uniform over the random interval (see Sect. 2.2.1) and it is noted that $\sigma_m^2 = 10^5$ is an appropriate choice for this. The posterior samples of β_i's, on the other hand, can be obtained easily by using the multivariate normal routine (see (18)). Also, the posterior samples of σ^2 can be effortlessly obtained from (24) using the gamma generating routine for the current values of shape and scale parameters. Now, following the three move types, defined in Sect. 2.2, the posterior samples for the number of knots k_i's and the locations ξ_i's can be easily simulated and finally updated to get the functional-coefficient $g_i(.)$ (see also Sect. 2.2.5). A reversible jump MCMC based estimates of the functional-coefficient is shown in Fig. 8. These estimates are obtained exactly in a similar way as described for the simulated dataset. Figure 8 indicates that the estimated quadratic spline, corresponding to the functional coefficient $g_1(.)$, is continuous at point 0 and has a range approximately from -0.0175 to 0.025 with a coverage probability of 0.95. To be specific, the estimated spline meets the zero line at two different location points, which further justifies the consideration of quadratic spline as mentioned in the above paragraph. The trace plots at several grid points and ergodic plots can also be obtained similar to Fig. 4, though we skip such plots due to the paucity of space.
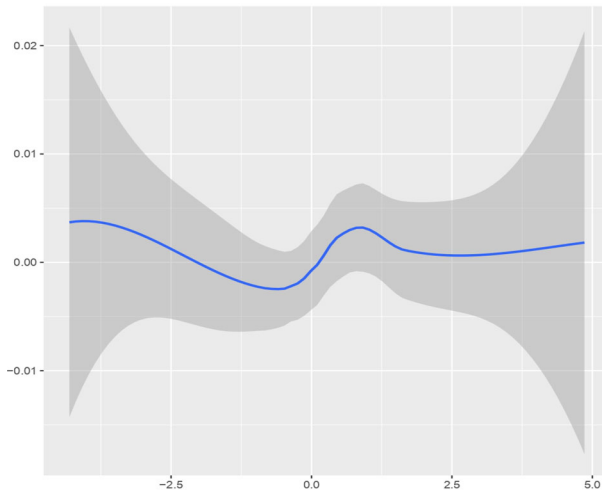
Fig. 8 Estimated functional-coefficient $g_1(.)$ (solid line) with 0.95 HPD (shaded region) for the differenced data

Table 5 Posterior summaries for the parameters of FCAR(1, 1) model corresponding to first difference data

Parameter	Posterior mean	Posterior median	Posterior mode	0.95 HPD	
θ_0	0.229	0.228	0.248	0.061	0.425
σ^2	1.925	1.908	1.861	1.456	2.390

Next, the posterior samples of the intercept parameter θ_0 can be obtained by implementing the Metropolis algorithm as discussed in Sect. 2.2.4. As mentioned, we fixed the value of the hyperparameter $M = 100$ and chose the value of the scaling constant as $c = 0.6$. This choice of $M = 100$ can certainly be regarded as providing a vague choice of the prior and hence allowing the inferences to be data driven. Also, the choice of scaling constant, as mentioned above, was well within the permissible range and resulted in the maximum acceptance probability (see also [24]).

Under the above setup, the joint posterior density (15) was managed for MCMC implementation to get the desired posterior samples for drawing the sample based inferences. To draw such inferences, we considered a single long run of the chain up to 20K iterations although the convergence was noticed at about 8K iterations. Next, after ignoring the outcomes of the first 10K iterations, we picked up posterior samples of size 1K from the last 10K iterations by maintaining a gap of 10 in order to minimize serial correlation among the generating variates. The extracted posterior summaries for the relevant parameters of the selected FCAR(1, 1) model are given below in Table 5.

A statistical interpretation of Table 5 can be easily made and is mostly fact-driven. It is obvious from the above posterior summaries that the estimated marginal posterior densities of θ_0 and σ^2 are quite close to symmetry. Since the above estimates are solely based on the simulated posterior samples from their respective full conditionals, the accuracy of the estimates conveys a strong message that the FCAR(1, 1) model is having an error term with (an almost) constant value of σ^2 and, hence, the model is homoscedastic in nature. Moreover,
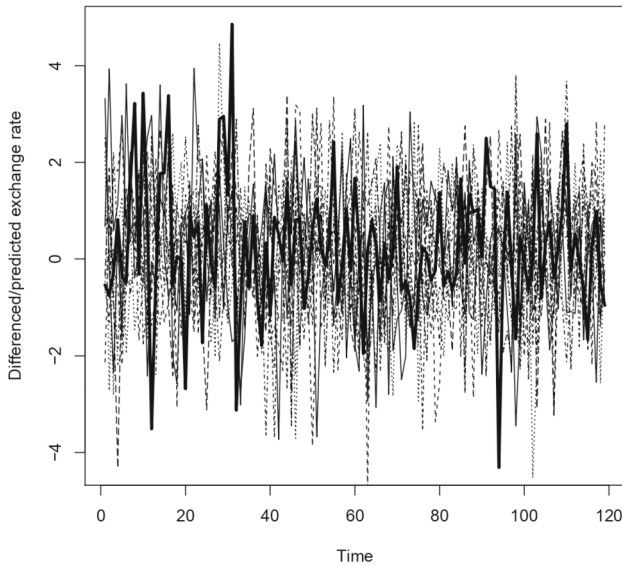
Fig. 9 Time series plots for the predictive datasets with superimposed differenced dataset (the dark line corresponds to the differenced data)

approximately normal shape of the densities was also assessed, which is not given here due to the paucity of space.

To show the model compatibility with the data in hand, we relied on graphical tools which appeared to be more striking than other statistical tools. We, therefore, plotted the differenced time series (using solid line) along with the predictive samples of time series (using dotted lines). For such an assessment, we considered 10 predictive samples, each of size equal to that of the differenced data, and superimposed them on to a plot of differenced data time series. It may be noted that the 10 predictive samples were obtained on the basis of 10 posterior samples of the concerned parameters, which were picked up randomly from the converged set of simulated posterior samples (see Sect. 2.4). The corresponding plot is shown in Fig. 9. It can be easily confirmed that the predicted time series and the observed (differenced) time series exhibit similar pattern (see Fig. 9), which advocates the adequacy of the proposed model with the data in hand.

Let us come to the final objective of our study to investigate the predictive ability of the proposed FCAR(1, 1) model. For this purpose, we performed the retrospective predictions of the exchange rate for the period from July 2020 to December 2020 by considering only the data from January 2011 to June 2020 as an informative dataset. Obviously, we did not use the observed dataset from July 2020 to December 2020, rather kept them for a comparison with the values obtained from the retrospective predictions. It is important to mention here that the entire posterior analysis was repeated for the updated informative set of data. By updating the informative dataset, we mean to include the predicted observations one after another in the informative dataset until the last value is predicted. Moreover, at each stage of prediction, we used posterior mode to predict 1K predictive samples. The result of retrospective predictions is given in the form of predictive modes and predictive intervals with coverage probability 0.95 (0.95 PI) (see Table 6). These estimates are everywhere based on the entire 1K predictive samples.

Table 6 The retrospective predictions of exchange rate of Indian rupees relative to the US dollars from July 2020 to December 2020 based on FCAR(1, 1) model

Month	True value	Predictive mode	0.95 PI	
July 2020	74.916	76.039	73.293	78.535
August 2020	73.254	75.770	73.031	79.033
September 2020	73.560	75.748	73.517	78.447
October 2020	74.554	75.955	72.788	78.613
Novembe 2020	73.990	76.284	73.686	78.916
December 2020	73.036	75.521	72.215	78.012

Table 7 The prospective predictions of exchange rate of Indian rupees relative to the US dollars from January 2021 to June 2021 based on FCAR(1, 1) model

Month	Predictive mode	0.95 PI	
January 2021	72.791	70.486	75.805
February 2021	73.392	70.682	76.098
March 2021	73.451	70.686	76.327
April 2021	73.752	71.307	75.933
May 2021	73.629	71.261	77.051
June 2021	74.177	70.787	76.929

Clearly, it can be seen that the predictive point estimates are not far away from their corresponding true values and they lie well within the estimated predictive intervals (see Table 6). Obviously, our analysis supports the FCAR (1, 1) model for the considered exchange rate dataset as the model based on retrospective predictions is in complete agreement with the true values of informative data.

Once the retrospective prediction is successfully observed with the proposed FCAR(1, 1) model, it is pertinent to look for the prospective prediction. Our objective was to predict the values from January, 2021 to June, 2021. Thus, for the prospective prediction, we considered the complete posterior analysis discussed earlier for the dataset reported in Table 3 with the posterior estimates given in Table 5. Using estimated posterior modes as given in Table 5, we obtained 1K predictive samples for predicting the corresponding value for January, 2021. The results are given in Table 7 in the form of predictive modes and 0.95 PI for January, 2021. Now adding the predictive mode for January, 2021 in the informative data, we repeated the entire procedure and obtained the corresponding predictive mode and 0.95 PI for February, 2021. This procedure was repeated until we obtained the predictive mode and 0.95 PI for the next four months, that is, up to June, 2021 (see Table 7). We do not provide a comparison of the prospective predictions with the actual values, but our belief conveys that the predicted results are close to the reality and the true values are reflected almost at the middle of our estimated 0.95 PIs although there can be slight differences in the point estimates.

4 Conclusion

This paper discusses the Bayesian analysis of FCAR(p, d) model using a free-knot spline technique. The significance of the work can be realized from the fact that it not only provides the complete Bayesian analysis using a hybrid MCMC based strategy but also looks

at several important aspects such as model compatibility, specification of both p and d of FCAR(p, d) model and, most importantly, both retrospective and prospective predictions. The first part obviously employs the Gibbs sampler algorithm with the intermediate Metropolis steps supported by the reversible jump MCMC technique. Numerical illustrations based on the simulated data as well as on a real dataset on the exchange rate of Indian rupees relative to the US dollars not only convey the ease of implementation of our methodology but also suggest that the proposed procedure has enough potential with regard to the prediction as well. It is expected that such an analysis will help the business analysts, investors and policy makers to come across an appropriate planning.

Appendix 1

Derivation of the conditional joint posterior of (β_i, k_i, ξ_i)

Let us write the joint posterior of (β_i, k_i, ξ_i), $i = 1, 2, \ldots, p$, after combining the likelihood function (7) and the prior distributions ((8) to (10)), up to proportionality as

$$\pi(\beta_i, k_i, \xi_i | \underline{y}, \theta_0, \beta_{-i}, k_{-i}, \xi_{-i}, \tau_\beta, \sigma^2)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2} \sum_{t=p^*+1}^{T} (y_t - \theta_0 - \sum_{i=1}^{p} B_i'(y_{t-d})\beta_i y_{t-i})^2\right) \times \left(\frac{\lambda_i \tau_\beta^{-1/2}}{b-a}\right)^{k_i}$$

$$\times (\sigma^2)^{-k_i/2} \times \exp\left(-\frac{1}{2\sigma^2 \tau_\beta}\beta_i'\beta_i\right),$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(Z_i - \beta_i X_i)'(Z_i - \beta_i X_i)\right) \times \left(\frac{\lambda_i \tau_\beta^{-1/2}}{b-a}\right)^{k_i}$$

$$\times (\sigma^2)^{-k_i/2} \times \exp\left(-\frac{1}{2\sigma^2 \tau_\beta}\beta_i'\beta_i\right),$$

where Z_i and X_i are defined in Sect. 2.2.1. Now expanding the first exponential term, in above expression, a more simplified version can be obtained as

$$\pi(\beta_i, k_i, \xi_i | \underline{y}, \theta_0, \beta_{-i}, k_{-i}, \xi_{-i}, \tau_\beta, \sigma^2)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2 \tau_\beta}\{\beta_i'\beta_i - Z_i'X_i\beta_i\tau_\beta - X_i'\beta_i'Z_i\tau_\beta + X_i'X_i\beta_i'\beta_i\tau_\beta)\}\right)$$

$$\times \left(\frac{\lambda_i \tau_\beta^{-1/2}}{b-a}\right)^{k_i} \times (\sigma^2)^{-k_i/2} \times \exp\left(-\frac{1}{2\sigma^2}Z_i'Z_i\right),$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\Sigma_i^{-1}\{\beta_i'\beta_i - \beta_i\hat{\beta}_i' - \beta_i'\hat{\beta}_i + \hat{\beta}_i'\hat{\beta}_i)\}\right) \times \left(\frac{\lambda_i \tau_\beta^{-1/2}}{b-a}\right)^{k_i}$$

$$\times (\sigma^2)^{-k_i/2} \times \exp\left(-\frac{1}{2\sigma^2}\{Z_i'Z_i - \hat{\beta}_i'\Sigma_i^{-1}\hat{\beta}_i\}\right),$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}(\beta_i - \hat{\beta}_i)'\Sigma_i^{-1}(\beta_i - \hat{\beta}_i)\right) \times \left(\frac{\lambda_i \tau_\beta^{-1/2}}{b-a}\right)^{k_i} \times (\sigma^2)^{-k_i/2} \times \exp\left(-\frac{1}{2\sigma^2}S_i\right),$$

where the mathematical form of Σ_i, $\hat{\beta}_i$ and S_i are defined in equation (17) of Sect. 2.2.1.

Appendix 2

Derivations of jump probabilities and acceptance probabilities in the three move-types

Jump probability:
Let M_{k_i} represents a model having k_i interior knots and M_{k_i+1} represents a model after adding an additional knot. The jump probability in **birth step** can be written as

$$
\begin{aligned}
q(M_{k_i+1}|M_{k_i}) &= b_{k_i} \times \text{probability of selecting a subinterval out of} (k_i+1) \text{subintervals} \\
&\quad \times \text{probability of drawing a candidate knot, say,} \\
&\quad \times \phi \text{uniformly from} (\xi_{ij}, \xi_{i,j+1}) \\
&= \frac{b_{k_i}}{k_i+1} \sum_{j=0}^{k_i} \frac{1}{\xi_{i,j+1}-\xi_{ij}} I\{\xi_{ij} < \phi < \xi_{i,j+1}\}.
\end{aligned}
$$

The probability defined above is actually the mixture of uniform densities as ϕ can be drawn from any of k_i different uniform distributions. Similarly, in the deletion process of **death step**, the model moves from M_{k_i} to M_{k_i-1} and hence the jump probability is given by

$$
\begin{aligned}
q(M_{k_i-1}|M_{k_i}) &= d_{k_i} \times \text{probability of randomly chosing a knot out of} k_i \text{existing knots} \\
&= \frac{d_{k_i}}{k_i}.
\end{aligned}
$$

The jump probability in **move step** can be defined as the probability of movement of the current model M_{k_i} to the candidate model $M_{k_i}^*$ with the same number of knots k_i and it can be given as

$$
\begin{aligned}
q(M_{k_i}^*|M_{k_i}) &= \eta_{k_i} \times \text{probability of chosing a candidate knot} \xi_{ij} \text{randomly out of} k_i \text{existing} \\
&\quad \times \text{knots probability of drawing a new location of candidate knot, say,} \\
&\quad \xi_{ij}^* \text{uniformly from} (\xi_{i,j-1}, \xi_{i,j+1}) \\
&= \frac{\eta_{k_i}}{k_i} \frac{1}{\xi_{i,j+1}-\xi_{i,j-1}}.
\end{aligned}
$$

Acceptance probability: The acceptance probability in all the three steps is defined earlier as

$$
\min(1, \text{posterior ratio} \times \text{proposal ratio}).
$$

In **birth step**, the posterior ratio is obtained from the full conditional (19) as

$$
\begin{aligned}
\frac{\pi(M_{k_i+1})}{\pi(M_{k_i})} &= \frac{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i+1} \times |\Sigma_i^*|^{1/2} \exp\left(\frac{-S_i^*}{2\sigma^2}\right)}{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i} \times |\Sigma_i|^{1/2} \exp\left(\frac{-S_i}{2\sigma^2}\right)} \\
&= \left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right) \times \left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right)
\end{aligned}
$$

and the proposal ratio is given by

$$\frac{\pi^*(M_{k_i+1})}{\pi^*(M_{k_i})} = \frac{d_{k_i+1}/(k_i+1)}{b_{k_i}/(k_i+1) \times \frac{1}{\xi_{i,j+1}-\xi_{ij}}}$$

$$= \frac{d_{k_i+1}}{b_{k_i}} \times (\xi_{i,j+1} - \xi_{ij})$$

$$= \frac{p(k_i)}{p(k_i+1)} \times (\xi_{i,j+1} - \xi_{ij})$$

$$= \frac{(k_i+1)}{\lambda_i} \times (\xi_{i,j+1} - \xi_{ij}).$$

The acceptance probability is, then, given by

$$\frac{\pi(M_{k_i+1})}{\pi(M_{k_i})} \times \frac{\pi^*(M_{k_i+1})}{\pi^*(M_{k_i})} = \frac{k_i+1}{\tau_\beta^{1/2}} \frac{\xi_{i,j+1}-\xi_{ij}}{b-a} \times \left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right).$$

In **death step**, the posterior ratio is obtained from the full conditional (19) as

$$\frac{\pi(M_{k_i-1})}{\pi(M_{k_i})} = \frac{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i-1} \times |\Sigma_i^*|^{1/2} \exp\left(\frac{-S_i^*}{2\sigma^2}\right)}{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i} \times |\Sigma_i|^{1/2} \exp\left(\frac{-S_i}{2\sigma^2}\right)}$$

$$= \left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{-1} \times \left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right)$$

and the proposal ratio is given by

$$\frac{\pi^*(M_{k_i-1})}{\pi^*(M_{k_i})} = \frac{b_{k_i-1}/k_i}{d_{k_i}/k_i} \times \frac{1}{\xi_{i,j+1}-\xi_{i,j-1}}$$

$$= \frac{p(k_i)}{p(k_i-1)} \times \frac{1}{\xi_{i,j+1}-\xi_{i,j-1}}$$

$$= \frac{\lambda_i}{k_i} \times \frac{1}{\xi_{i,j+1}-\xi_{i,j-1}}.$$

The acceptance probability is, therefore, given by

$$\frac{\pi(M_{k_i-1})}{\pi(M_{k_i})} \times \frac{\pi^*(M_{k_i-1})}{\pi^*(M_{k_i})} = \frac{\tau_\beta^{1/2}}{k_i} \frac{b-a}{\xi_{i,j+1}-\xi_{i,j-1}} \times \left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right).$$

Finally, in **move step**, the number of knots remain same and the only change is the relocation of a selected knot by the candidate knot. The posterior ratio is obtained from the full conditional (19) as
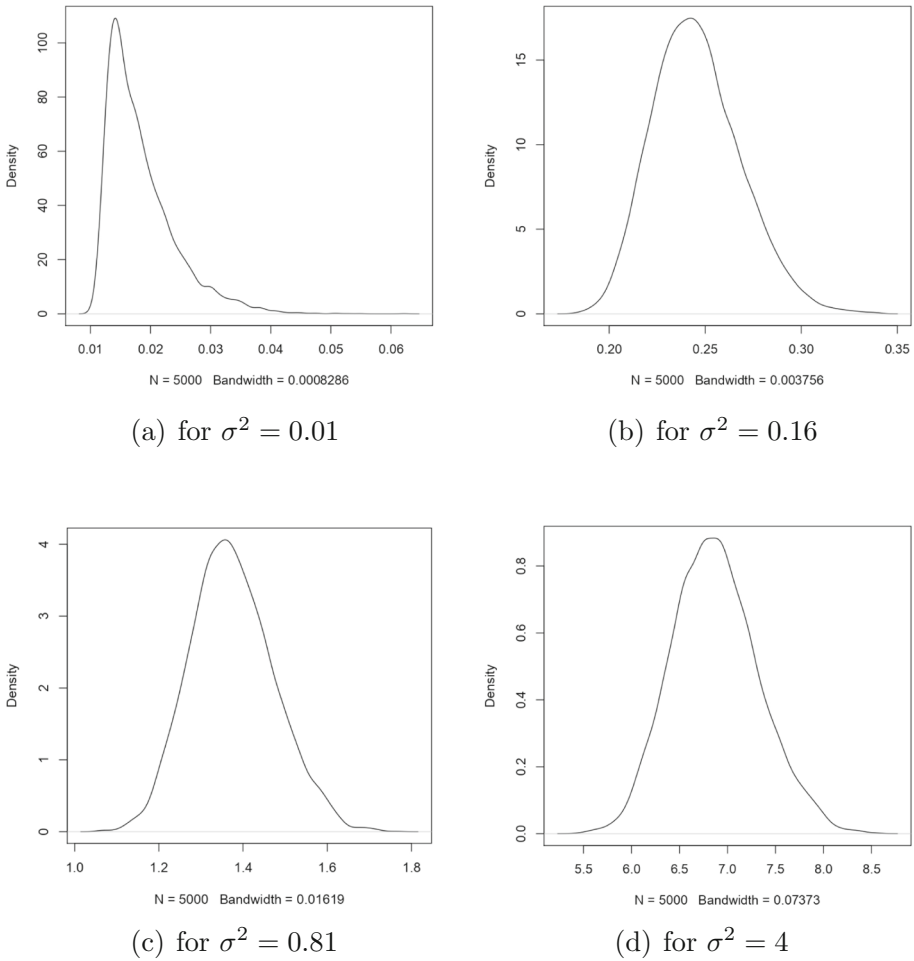
(a) for $\sigma^2 = 0.01$

(b) for $\sigma^2 = 0.16$

(c) for $\sigma^2 = 0.81$

(d) for $\sigma^2 = 4$

Fig. 10 Estimated posterior density plots of σ^2, at different values, corresponding to FCAR(1,1) model

$$
\frac{\pi(M_{k_i}^*)}{\pi(M_{k_i})} = \frac{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i} \times |\Sigma_i^*|^{1/2} \exp\left(\frac{-S_i^*}{2\sigma^2}\right)}{\left(\frac{\tau_\beta^{-1/2}\lambda_i}{b-a}\right)^{k_i} \times |\Sigma_i|^{1/2} \exp\left(\frac{-S_i}{2\sigma^2}\right)}
$$

$$
= \left(\frac{|\Sigma_i^*|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S_i^*}{2\sigma^2}\right).
$$

The proposal ratio in this case is given by

$$
\frac{\pi^*(M_{k_i}^*)}{\pi^*(M_{k_i})} = 1.
$$

Hence the acceptance probability is given by

$$\frac{\pi(M^*_{k_i})}{\pi(M_{k_i})} \times \frac{\pi^*(M^*_{k_i})}{\pi^*(M_{k_i})} = \left(\frac{|\Sigma^*_i|}{|\Sigma_i|}\right)^{1/2} \exp\left(\frac{S_i - S^*_i}{2\sigma^2}\right).$$

Appendix 3

As discussed in Sect. 3.1, we have checked the performance of two considered models for different choices of error variance. We have performed the whole analysis for the values 0.01, 0.16, 0.81 and 4 of σ^2 and have plotted the estimated posterior densities in Fig. 10 and Fig. 11 for FCAR(1,1) and FCAR(2,1) models respectively.

We can easily verify that the nature of plots are almost unchanged (close to symmetry) except for the value very close to zero (for σ^2=0.01); which ultimately putting negligible effects on the predicted values and non-linearity of the model. We, however, considered
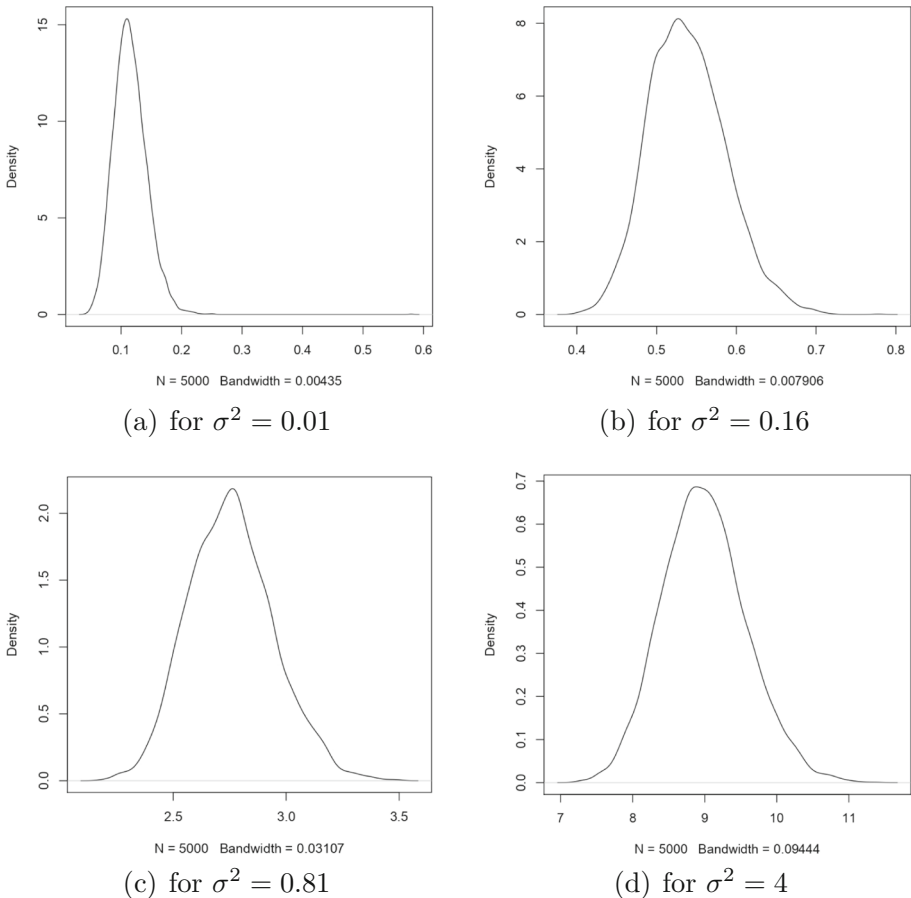

(a) for $\sigma^2 = 0.01$

(b) for $\sigma^2 = 0.16$

(c) for $\sigma^2 = 0.81$

(d) for $\sigma^2 = 4$

Fig. 11 Estimated posterior density plots of σ^2, at different values, corresponding to FCAR(2,1) model

σ^2=0.16 in order to get the final posterior estimates of parameters and functional coefficients, for the two models, in the simulation study.

Appendix 4

Proof of boundedness for the proposed functions in simulation study

1) Proof of boundedness for the functional-coefficient $g_1(y_{t-1})$:

We have considered the following function for the non-linear functional-coefficient $g_1(y_{t-1})$:

$$g_1(x) = x \times exp(-x^2/2); \qquad \forall \qquad x \in \mathbb{R}.$$

We know that any function $f : D \to R$, where D and R represent the possible domain and range of the function respectively, will be differentiable at any point $c \in D$, if and only if both the left-hand and the right-hand derivatives of the function are finite and equal, such that-

$$\lim_{h \to 0} \frac{f(c-h) - f(c)}{-h} = \lim_{h \to 0} \frac{f(c+h) - f(c)}{h}.$$

It is important to mention that, for the considered function $g_1(.)$, domain (D) and range (R) are the set of real numbers $\mathbb{R}$. In order to show that the function is differentiable, we check the left-hand and the right-hand derivative of the function $g_1(.)$ at any arbitrary point, say, c. We can write the left-hand derivative as,

$$
\begin{aligned}
g_1'(c_-) &= \lim_{h \to 0} \frac{g_1(c-h) - g_1(c)}{-h} \\
&= \lim_{h \to 0} \left[\frac{(c-h)e^{-(c-h)^2/2} - ce^{-c^2/2}}{-h} \right] \\
&= \lim_{h \to 0} \left[\frac{(c-h)^2 e^{-(c-h)^2/2} - e^{-(c-h)^2/2}}{-1} \right] \\
&= -e^{-c^2/2}(c^2 - 1).
\end{aligned}
$$

Similarly, the right-hand derivative can be written as,

$$
\begin{aligned}
g_1'(c_+) &= \lim_{h \to 0} \frac{g_1(c+h) - g_1(c)}{h} \\
&= \lim_{h \to 0} \left[\frac{(c+h)e^{-(c+h)^2/2} - ce^{-c^2/2}}{h} \right] \\
&= \lim_{h \to 0} \left[\frac{-(c+h)^2 e^{-(c+h)^2/2} + e^{-(c+h)^2/2}}{1} \right] \\
&= -e^{-c^2/2}(c^2 - 1).
\end{aligned}
$$

One can see that both the left-hand and the right-hand derivatives, for the considered function, are finite and equal. Hence, the function $g_1(x)$ is differentiable. Moreover, the optimum points

for the function $g_1(x)$ can also be obtained by calculating the first derivative and putting it equals to zero.

$$g_1'(x) = e^{-x^2/2} - x^2 e^{-x^2/2}$$
$$= e^{-x^2/2}\left(1 - x^2\right).$$

In above expression, the optimum points can be obtained by putting

$$1 - x^2 = 0; \; (\because e^{-x^2/2} \neq 0)$$
$$\Rightarrow x = \pm 1.$$

Now, the value of the function can easily be obtained at the optimum points as,

$$g_1(-1) = -e^{-1/2} \text{and}$$
$$g_1(1) = e^{-1/2}$$
$$\Rightarrow -e^{-1/2} \leq g_1(x) \leq e^{-1/2}.$$

Hence, the function $g_1(x)$ is bounded on a closed interval $[-e^{-1/2}, e^{-1/2}]$ and, hence,

$$|g_1(x)| \leq e^{-1/2} \quad \text{or} \quad (|g_1(x)| \leq 0.6065).$$

2) Proof of boundedness for the functional-coefficient $g_2(y_{t-1})$:

We have considered the following function for the non-linear functional-coefficient $g_2(y_{t-1})$:

$$g_2(x) = -\cos(1.5x)/(x^2 + 1); \qquad \forall \qquad x \in \mathbb{R}.$$

We know that any real-valued function $f(x)$ defined on a subset S of the real numbers is said to be bounded if there exist a constant M, such that for all $x \in$ S, the inequality $|f(x)| \leq M$ holds, where M belongs to some positive real number ($\mathbb{R}^+$).

Now, we know that the range of the cos function lies in the interval $[-1, 1]$. So, the range of $cos(1.5x)$ also lies within the range $[-1, 1]$ such that

$$|-\cos(1.5x)| \leq 1$$

Also, for all $x \in \mathbb{R}$,

$$\frac{1}{1 + x^2} \leq 1$$
$$\Rightarrow \left|\frac{-\cos(1.5x)}{1 + x^2}\right| \leq 1$$
$$\Rightarrow -1 \leq \frac{-\cos(1.5x)}{1 + x^2} \leq 1$$

Hence, the function $g_2(x)$ is a bounded function.

Data availability statement The dataset is provided in the manuscript along with itssource link.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Araveeporn, A.: Comparing random coefficient autoregressive model with and without autocorrelated errors by Bayesian analysis. Stat. J. IAOS **33**(2), 537–545 (2017)
2. Biller, C.: Adaptive Bayesian regression splines in semi-parametric generalized linear models. J. Comput. Graph. Stat. **9**(1), 122–140 (2000)
3. Box, G.E.P., Jenkins, G.M.: Time series analysis: forecasting and control, revised Holden-Day (1976)
4. Cai, Z., Fan, J., Yao, Q.: Functional-coefficient regression models for non-linear time series. J. Am. Stat. Assoc. **95**(451), 941–956 (2000)
5. Chen, R., Tsay, R.S.: Functional-coefficient autoregressive models. J. Am. Stat. Assoc. **88**(421), 298–308 (1993)
6. Chib, S., Greenberg, E.: Bayes inference in regression models with ARMA (p, q) errors. J. Econ. **64**(1), 183–206 (1994)
7. Denison, D.G.T., Mallick, B.K., Smith, A.F.M.: Automatic Bayesian curve fitting. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **60**(2), 333–350 (1998)
8. Devroye, L.: Non-uniform random variate generations. Springer-Verlag, New York (1986)
9. DiMatteo, I., Genovese, C.R., Kass, R.E.: Bayesian curve-fitting with free-knot splines. Biometrika **2**, 1055–1071 (2001)
10. Fan, J., Gijbels, I.: Variable bandwidth and local linear regression smoothers. Ann. Stat. **2**, 2008–2036 (1992)
11. Fan, J., Yao, Q.: Non-linear time series: non-parametric and parametric methods. Springer-Verlag, New York (2003)
12. Green, P.J.: Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika **2**, 711–732 (1995)
13. Grégoire, G., Hamrouni, Z.: Change point estimation by local linear smoothing. J. Multivar. Anal. **83**(1), 56–83 (2002)
14. Haggan, V., Ozaki, T.: Modelling non-linear random vibrations using an amplitude-dependent autoregressive time series model. Biometrika **68**(1), 189–196 (1981)
15. Lindstrom, M.J.: Bayesian estimation of free-knot splines using reversible jumps. Comput. Stat. Data Anal. **41**(2), 255–269 (2002)
16. Liu, L.-M., Tiao, G.C.: Random coefficient first-order autoregressive models. J. Econ. **13**(3), 305–325 (1980)
17. Mukherjee, B., Gupta, A., Upadhyay, S.K.: A Bayesian study for the comparison of generalized gamma model with its components. Sankhya B **72**, 154–174 (2010)
18. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**(2), 461–464 (1978)
19. Song, X.Y., Cai, J.H., Feng, X.N., Jiang, X.J.: Bayesian analysis of the functional-coefficient autoregressive heteroscedastic model. Bayesian Anal. **9**(2), 371–396 (2014)
20. Tong, H.: On a threshold model. In: Chen, C.H. (ed.) Pattern recognition and signal processing, pp. 575–586. Sijthoff and Noordhoff, Amsterdam (1978)
21. Tong, H.: Non-linear time series: a dynamical system approach. Oxford University Press, Oxford (1990)
22. Tripathi, P.K., Agarwal, M.: Bayesian prediction of monthly gold prices using an EARSV model and its competitive component models. Int. J. Math. Stat. **22**(3), 1–17 (2021)
23. Tripathi, P.K., Ranjan, R., Pant, R., Upadhyay, S.K.: An approximate Bayes analysis of ARMA model for Indian GDP growth rate data. J. Stat. Manag. Syst. **20**(3), 399–419 (2017)
24. Tripathi, P.K., Sen, R., Upadhyay, S.K.: A Bayes algorithm for model compatibility and comparison of ARMA(p, q) models. Stat. Trans. New Ser. **22**(2), 95–123 (2021)

25. Tripathi, P.K., Upadhyay, S.K.: Bayesian analysis of extended auto regressive model with stochastic volatility. J. Indian Soc. Prob. Stat. **20**(1), 1–29 (2019)
26. Upadhyay, S. K., Vasishta, N., and Smith, A. F. M. (2001). Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002), 63(1):15–40
27. Wang, D., Ghosh, S.K.: Bayesian analysis of random coefficient autoregressive models. Technical report, North Carolina State University, Dept. of Statistics (2004)
28. Wang, H.B., Wu, P.: Bayesian inference of autoregressive and functional-coefficient moving average models. Commun. Stat. Theory Methods **44**(3), 453–467 (2015)