



Restricted Empirical Likelihood Estimation for Time Series Autoregressive Models

Mahdiah Bayati¹, S.K. Ghoreishi^{2,*}, Jingjing Wu³

¹Department of Statistics, Science and Research Branch, Azad University, I. R. Iran

²Department of Statistics, Faculty of Sciences, University of Qom, I. R. Iran

³Department of Mathematics and Statistics, University of Calgary, Canada

ARTICLE INFO

Article History

Received 03 June 2020

Accepted 18 Jan 2021

Keywords

Autoregressive models

EM algorithm

Empirical likelihood

Estimating equations

2000 Mathematics Subject

Classification: 62F10, 62F15

ABSTRACT

In this paper, we first illustrate the restricted empirical likelihood function, as an alternative to the usual empirical likelihood. Then, we use this quasi-empirical likelihood function as a basis for Bayesian analysis of $AR(r)$ time series models. The efficiency of both the posterior computation algorithm, when the estimating equations are linear functions of the parameters, and the EM algorithm for estimating hyper-parameters is an appealing property of our proposed approach. Moreover, the competitive finite-sample performance of this proposed method is illustrated via both simulation study and analysis of a real dataset.

© 2021 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

The empirical likelihood method is one of the most useful statistical tools that allows us to refer to under some regularity conditions. Initially, this statistical method was introduced by Thomas and Grunkemeier [1] and subsequently further developed by Owen [2,3].

Nowadays, many statisticians use this statistical method in different fields of applications. Owen [2,3] have used the empirical likelihood ratio statistic to test the assumptions assumed by nonparametric statistics. They have shown that this statistic has an asymptotic chi-square distribution. Moreover, they have used this statistic to construct a confidence region and to perform a hypothesis testing of parameters. Asymptotic properties and some necessary corrections of this statistic have been studied by DiCiccio and Romano [4] and Hall and Scala [5]. Furthermore, Qin and Lawless [6] has demonstrated that in the nonparametric scheme, the empirical likelihood, along with estimating equations, fits the data well. For more details, see Newey and Smith [7], Chen and Cui, [8,9]), and Qin and Lowless [6].

In many statistical methods, estimating equations are commonly used to estimate the parameters of assumed model. Among them, the most well-known ones include the maximum likelihood (ML) method, the least square method, and the moment method. In this regard, the empirical likelihood approach, by applying unbiasedness constraint on estimating functions, leads to the maximum empirical likelihood (MEL) estimation, which is obtained by maximizing the empirical likelihood function, under the unbiasedness imposed on estimating equations involving the unknown probability p_1, p_2, \dots, p_n . The works from Chen *et al.* [9], Hjort *et al.* [10], Tang and Leng [11], Leng and Tang [12], and Chang *et al.* [13] have shown that empirical likelihood estimators perform well when the dimension p of parameters and the number r of estimating equations grow slower than the sample size n , i.e. $p = o(n)$ and $r = o(n)$.

Tang and Leng [11], Leng and Tang [12], and Chang and Chen [13] have investigated the properties of empirical maximal likelihood estimators for high-dimensional data with sparse parameter vector. Practically, this is done by incorporating some suitable penalty functions. However, there are some challenges in employing empirical likelihood for large-scale data. Tsao [14] has shown that for a large enough n and fixed p , the probability of the corresponding confidence region containing true parameter values can be substantially smaller than its nominal value, which leads to an under-coverage problem.

In the meantime, the use of empirical likelihood method for time series data analysis has also been a topic of interest for researchers. For more details, interested readers are referred to Mykland [15], Kitamura [16], Chan and Ling [17], and Chan and Liu [18]. Monti [19]

*Corresponding author. Email: atty_ghoreishi@yahoo.com

has developed the empirical likelihood method for the analysis of stationary time series. The author has used the M-estimator method, introduced by Whittle [21], for periodogram ordinates in time series models. Later, Yau [21] has extended the results in Monti [19] to long memory time series. Nordman and Lahiri [22] has argued that Monti's results were based on the assumption of normal distribution for the error term, and thus has extended the classical empirical likelihood to the one based on spectral distribution via Fourier transforms. Their results apply to both short and long memory dependencies; see Nordman and Lahiri [24] for details.

To the best of our knowledge, all the studies on empirical likelihood for time series rely on the unbiasedness property of estimating equations. But it seems that the unbiasedness condition alone, when data contains outliers, will produce misleading estimates for the parameters of interest (Bayati et al. [24]). In other words, in the usual empirical likelihood approach, features such as the dispersion of estimator functional, which depends on the dispersion of data itself, are not included in the statistical inference process. Therefore, it is necessary to use a pseudo-empirical likelihood that is a squaring function of the estimating equations. In this paper, for the first time we introduce the *restricted empirical likelihood* (REL) for the AR(r) time series models and present its Bayesian analysis. To our best knowledge, Bayati et al. [24] is the only one that has also used the idea of REL. Nevertheless, Bayati et al. [24] and this paper are very different in the sense that Bayati et al. [24] has studied the REL for independent observations while this paper investigates the REL for the more sophisticated case of dependent observations specifically from autoregressive models. In this paper, we not only prove the asymptotic normality of REL estimator for general models, but also give the detailed Bayesian analysis of AR(r) time series models using the REL and as well develop efficient algorithm for computing posterior and estimating hyper-parameters.

This article is organized as follows. After a glance at the empirical likelihood approach, we will illustrate the REL methodology in Section 2. The Bayesian analysis using the REL is discussed in Section 3. Bayesian analysis of AR(r) time series models, using the REL, are given in Section 4. A simulation study and an application to a real-world dataset are embedded in Section 5.

2. PRELIMINARIES

In this section, we first review the concept of empirical likelihood and then introduce the REL which is the main method proposed in this paper. The notations and definitions are provided along with other content.

2.1. Empirical Likelihood

Suppose $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in \mathbf{R}^n$ are n independent observations obtained from the unknown distribution function f and they relate to the vector of parameters $\boldsymbol{\phi} = (\phi_1, \dots, \phi_r)'$. Moreover, we assume that $g_j(x; \boldsymbol{\phi})$, $j = 1, \dots, r$, are some estimating functions to estimate $\boldsymbol{\phi}$ and they satisfy the unbiasedness conditions

$$\mathbb{E}_{\mathcal{P}}\{g_j(\mathbf{x}; \boldsymbol{\phi})\} = \sum_{i=1}^n p_i g_j(x_i; \boldsymbol{\phi}) = 0, \quad j = 1, \dots, r, \quad (1)$$

where $\mathcal{P} = \{p_1, \dots, p_n\}$. Moreover, let $\mathbf{g}(\mathbf{x}; \boldsymbol{\phi}) = (g_1(\mathbf{x}; \boldsymbol{\phi}), \dots, g_r(\mathbf{x}; \boldsymbol{\phi}))'$.

In the usual empirical likelihood approach, the estimates of probabilities p_i 's are obtained by maximizing the empirical likelihood function

$$L(\boldsymbol{\phi}) \propto \prod_{i=1}^n p_i \quad (2)$$

under conditions (1) and $\sum_{i=1}^n p_i = 1$. Equivalently,

$$L(\boldsymbol{\phi}) = \sup_{\mathcal{P}} \left\{ \prod_{i=1}^n p_i : p_i > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \mathbf{g}(x_i; \boldsymbol{\phi}) = \mathbf{0} \right\}. \quad (3)$$

So, it is straightforward to see that

$$p_i \propto \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\phi})' \mathbf{g}(x_i; \boldsymbol{\phi})},$$

where $\boldsymbol{\lambda}(\boldsymbol{\phi}) = (\lambda_1(\boldsymbol{\phi}), \lambda_2(\boldsymbol{\phi}), \dots, \lambda_r(\boldsymbol{\phi}))'$ is the vector of Lagrange multipliers (see Owen [2] and Qin and Lawless [6]). So, the empirical likelihood function is given as

$$L(\boldsymbol{\phi}) \propto \prod_{i=1}^n \frac{1}{1 + \boldsymbol{\lambda}(\boldsymbol{\phi})' \mathbf{g}(x_i; \boldsymbol{\phi})}. \quad (4)$$

By using (4), the MEL estimator of $\boldsymbol{\phi}$ is given by

$$\hat{\boldsymbol{\phi}}^{EL} = \arg \min_{\boldsymbol{\phi}} \max_{\lambda} \left[\sum_{i=1}^n \log \{ 1 + \lambda(\boldsymbol{\phi})' \mathbf{g}(x_i; \boldsymbol{\phi}) \} \right].$$

There is a modified two-layer coordinate descent algorithm to compute $\hat{\boldsymbol{\phi}}^{EL}$ (Tang and Wu [25]).

2.2. Restricted Empirical Likelihood

In previous subsection, the empirical likelihood function is given by (4). Now, by inequality $\mathbf{a}'\mathbf{b} \leq \frac{\mathbf{a}'\mathbf{a}+\mathbf{b}'\mathbf{b}}{2} = \frac{\|\mathbf{a}\|^2+\|\mathbf{b}\|^2}{2}$ for any arbitrary vectors \mathbf{a} and \mathbf{b} , we have

$$\prod_{i=1}^n \frac{1}{1 + \frac{\|\lambda(\boldsymbol{\phi})\|^2 + \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2}{2}} \leq \prod_{i=1}^n \frac{1}{1 + \lambda(\boldsymbol{\phi})' \mathbf{g}(x_i; \boldsymbol{\phi})}. \tag{5}$$

We hope that the value of $\boldsymbol{\phi}$ that maximizes the left side of (5) approaches to the maximizing value of $\boldsymbol{\phi}$ for the right side. The left side of (5) can be rewritten as

$$\prod_{i=1}^n \frac{1}{1 + \frac{\|\lambda(\boldsymbol{\phi})\|^2 + \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2}{2}} = \prod_{i=1}^n 2w(\boldsymbol{\phi}) \frac{1}{1 + w(\boldsymbol{\phi}) \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2} = [2w(\boldsymbol{\phi})]^n \prod_{i=1}^n \frac{1}{1 + w(\boldsymbol{\phi}) \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2}, \tag{6}$$

where $w(\boldsymbol{\phi}) = \frac{1}{2 + \|\lambda(\boldsymbol{\phi})\|^2}$. Therefore, the pseudo-empirical maximum likelihood estimator (MLE) of $\boldsymbol{\phi}$ is given by

$$\hat{\boldsymbol{\phi}} = \arg \min_{\boldsymbol{\phi} \in \Phi} \max_{\mathbf{w} \in \mathbb{R}^{+n}} \left[\sum_{i=1}^n \log \{ 1 + w(\boldsymbol{\phi}) \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2 \} - n \log w(\boldsymbol{\phi}) \right].$$

Indeed, the above equation is equivalent to the case that the function $\sum_{i=1}^n \log [1 + w(\boldsymbol{\phi}) \|\mathbf{g}(x_i; \boldsymbol{\phi})\|^2]$ is optimized under the penalty function $-\log w(\boldsymbol{\phi})$. It gives us the idea that we can extend the expression (6) to a general form given by

$$e^{-nH(\mathbf{w}, \boldsymbol{\phi})} \prod_{i=1}^n \frac{1}{1 + \mathbf{w}' \mathbf{g}^*(x_i; \boldsymbol{\phi})}, \tag{7}$$

where $\mathbf{w} = (w_1, \dots, w_r)'$ and $\mathbf{g}^*(x; \boldsymbol{\phi}) = (g_1^2(x; \boldsymbol{\phi}), \dots, g_r^2(x; \boldsymbol{\phi}))'$. Here, we assume $H(\mathbf{w}, \boldsymbol{\phi})$ plays the role of a penalty function and define our REL, an alternative to the usual empirical likelihood, as

$$L_R(\mathbf{w}, \boldsymbol{\phi}) = \prod_{i=1}^n \frac{1}{1 + \mathbf{w}' \mathbf{g}^*(x_i; \boldsymbol{\phi})}. \tag{8}$$

Practically, the REL $L_R(\mathbf{w}, \boldsymbol{\phi})$ has two appealing properties in contrast to $L(\boldsymbol{\phi})$ given in (4). First, we can assume in $L_R(\mathbf{w}, \boldsymbol{\phi})$ that the elements of vector \mathbf{w} are independent of $\boldsymbol{\phi}$, while in $L(\boldsymbol{\phi})$ dependency of λ on $\boldsymbol{\theta}$ is a basic assumption. This means that $L_R(\mathbf{w}, \boldsymbol{\phi})$ can be considered as a semi-parametric version of the usual EL function. Second, by assuming sparsity of vectors $\boldsymbol{\phi}$ and \mathbf{w} (to reduce the number of parameters and the number of the estimating equations respectively) and $-\log REL$ being convex, we can use the REL method for high-dimensional data. For more details on the appealing properties of the REL $L_R(\mathbf{w}, \boldsymbol{\phi})$, readers are referred to Bayati et al. [24]. With these appealing properties, we propose the REL estimator of $\boldsymbol{\phi}$ as given by

$$(\hat{\boldsymbol{\phi}}^{REL}, \hat{\mathbf{w}}^{REL}) = \arg \min_{\boldsymbol{\phi} \in \mathbb{R}^r, \mathbf{w} \in \mathbb{R}^{+r}} \left[\sum_{i=1}^n \log \{ 1 + \mathbf{w}' \mathbf{g}^*(x_i; \boldsymbol{\phi}) \} + nH(\mathbf{w}, \boldsymbol{\phi}) \right].$$

Under some regularity conditions on \mathbf{g}^* and H , the REL estimator $\hat{\boldsymbol{\phi}}^{REL}$ has an asymptotically multivariate normal distribution. This result is presented in following theorem with details. Its proof is straightforward and thus omitted to save space.

Theorem 2.1. Under some well-known regularity conditions on g^* and H , i.e., $\left| \frac{\partial^2 g^*(x; \phi)}{\partial \phi_i \partial \phi_j} \right| < \infty$ and $\left| \frac{\partial^2 H(\mathbf{w}, \phi)}{\partial \phi_i \partial \phi_j} \right| < \infty$, when x_1, x_2, \dots, x_n are i.i.d., it holds that

$$\sqrt{n}(\hat{\phi} - \phi) \rightarrow N_r(\mathbf{0}, I^{-1}(\phi)),$$

where

$$I(\phi) = \left[E \left(\frac{\partial^2 [\log \{1 + \mathbf{w}^T \mathbf{g}^*(x; \phi)\} + H(\mathbf{w}, \phi)]}{\partial \phi_i \partial \phi_j} \right) \right]_{i,j=1, \dots, r}$$

and \mathbf{w} is replaced with its estimate $\hat{\mathbf{w}}^{REL}$.

Proof. Since $\hat{\phi}$ is the maximizer of $\sum_{i=1}^n \log \{1 + \mathbf{w}^T \mathbf{g}^*(x_i; \phi)\} + nH(\mathbf{w}, \phi)$, so we define

$$\begin{aligned} T_n(\phi) &= \frac{1}{n} \left[\sum_{i=1}^n \log \{1 + \mathbf{w}^T \mathbf{g}^*(x_i; \phi)\} + nH(\mathbf{w}, \phi) \right] \\ &= \frac{1}{n} \sum_{i=1}^n [\log \{1 + \mathbf{w}^T \mathbf{g}^*(x_i; \phi)\} + H(\mathbf{w}, \phi)]. \end{aligned}$$

It is obvious that $\hat{\phi}$ is the MLE when $T_n(\phi)$ is log-likelihood function normalized by $\frac{1}{n}$. Therefore, the proof is the same as asymptotic normality of MLEs. Moreover, under the given regularity conditions, MLE $\hat{\phi}$ is constant, i.e., $\hat{\phi} \rightarrow \phi$ as $n \rightarrow \infty$. □

In the next section, we focus on Bayesian analysis of ϕ , considering the REL as the usual likelihood function.

3. BAYESIAN ANALYSIS WITH REL

Our main purpose in this paper is the Bayesian analysis of ϕ , viewing the REL as playing the role of a regular likelihood function. Based on this idea, we hope that the Bayesian estimation of ϕ can provide an acceptable approximation of the true values. For simplicity, we assume that the penalty function $H(\mathbf{w}, \phi)$ is an additive function, i.e., $H(\mathbf{w}, \phi) = H_1(\mathbf{w}) + H_2(\phi)$, and moreover assume $e^{-nH(\mathbf{w}, \phi)}$ is proportional to a multivariate density. Consequently, we assume that

$$\pi_1(\mathbf{w}) \propto e^{-nH_1(\mathbf{w})} \text{ and } \pi_2(\phi) \propto e^{-nH_2(\phi)}, \tag{9}$$

are the corresponding prior densities. We have the following two points to make on the densities.

1. We assume in (9) that the two vectors \mathbf{w} and ϕ are independent. This assumption is often considered for simplicity and computational purposes.
2. Unlike the usual way, the priors π_1 and π_2 are dependent of n . This is often assumed in Bayesian statistical analysis; for more details see Bhattacharya *et al.* [26] and Ghoreishi [27]. However, the functions H_1 and H_2 can also be chosen in such a way that the effect of sample size is ignorable.

Based on the above assumptions, using (8) and (9) along with observations $D = \{x_1, \dots, x_n\}$, the joint posterior density of (\mathbf{w}, ϕ) is

$$\pi(\mathbf{w}, \phi | D) \propto L_R(\mathbf{w}, \phi) \pi_1(\mathbf{w}) \pi_2(\phi). \tag{10}$$

Given the estimating functions g^* and the priors π_1 and π_2 , one can estimate the desired parameters using the Bayesian approach. In practice, the marginal posterior distribution of parameters often do not have a closed form, therefore any inference is derived using samples via MCMC methods. However, in many statistical models in which the estimating functions are usually linear of the parameters, the conditional posterior distributions have the closed form and they are easy to sample using Gibbs sampling method. In the next section, we discuss this issue with more details for autoregressive models of order r , i.e., AR(r).

4. REL FOR AR(r) MODELS

4.1. Basic Concepts

Autoregressive models show a random process where the output variable is a linear function of the same variable. Specifically, the AR(r) model is defined as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_r X_{t-r} + \epsilon_t, \quad t = r + 1, \dots, n, \quad (11)$$

where $\phi_1, \phi_2, \dots, \phi_r$ are the model parameters, and ϵ_t 's are white noises with mean 0 and constant variance σ^2 . In our empirical approach, we assume the distribution of ϵ_t 's is totally unknown. In practice, there are several methods to estimate the model coefficients. A popular approach is Yule-Walker's system of equations, i.e.,

$$\sum_{t=r+1}^n X_{t-j} (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_r X_{t-r}) = 0, \quad j = 1, 2, \dots, r,$$

where

$$g_{jt}(\mathbf{X}, \boldsymbol{\phi}) = X_{t-j} (X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_r X_{t-r}), \quad j = 1, 2, \dots, r \quad (12)$$

are the estimating equations. Later in our Bayesian approach, we will use the weighed Yule-Walker's system of equations as our estimating equations.

As we see, the estimating equations (12) are linear functions of parameters $\phi_1, \phi_2, \dots, \phi_r$. Therefore, we hope that the posterior conditional distributions have a closed form when we use the REL function instead of a regular likelihood function. Moreover, in the autoregressive model (11), the conditional distribution of (x_{r+1}, \dots, x_n) given (x_1, \dots, x_r) is

$$f(x_{r+1}, \dots, x_n | \boldsymbol{\phi}, x_1, \dots, x_r) = \prod_{t=r+1}^n f(x_t | \boldsymbol{\phi}, x_{t-1}, \dots, x_{t-r}) = \prod_{t=r+1}^n p_t.$$

where $p_t := f(x_t | \boldsymbol{\phi}, x_{t-1}, x_{t-2}, \dots, x_{t-r})$. On the other hand, the corresponding likelihood function is

$$L(\boldsymbol{\phi}) = \prod_{t=r+1}^n p_t, \quad (13)$$

Since the probabilities p_t 's are practically unknown, we try to estimate them here using our REL approach. We assume that the observation x_t is weighted with its corresponding conditional density function value p_t as the weight. This means that up to a normalizing constant, the estimating equations (12) satisfy

$$\begin{aligned} \mathbb{E}_{\mathcal{D}} \{g_{jt}(x; \boldsymbol{\phi})\} &= \sum_{t=r+1}^n p_t g_j(x, \boldsymbol{\phi}) \\ &= \sum_{t=r+1}^n p_t x_{t-j} (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_r x_{t-r}) = 0, \end{aligned} \quad (14)$$

where $\mathcal{D} = \{p_{r+1}, p_{r+2}, \dots, p_n\}$. Finally, combining (8), (13) and (14) gives the corresponding REL as

$$L_R(\boldsymbol{\phi}, \mathbf{w}) = \prod_{t=r+1}^n \frac{1}{1 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_r x_{t-r})^2}. \quad (15)$$

4.2. Posterior Conditional Distributions

Without loss of generality, for Bayesian analysis using the REL (15), we assume w_j, ϕ_j and s_j are i.i.d. with the following common priors

$$\begin{aligned} w_j &\sim \pi_1 = \text{inv.gamma}(n, \beta_0), \\ \phi_j &\sim \pi_2 = N(0, \sigma_0^2), \\ s_j &\sim f(u_j) = \text{Exp}(1), \end{aligned} \quad (16)$$

where the additional variables $s_{r+1}, s_{r+2}, \dots, s_n$ (the argument u_{r+1}, \dots, u_n in their respective distribution functions) appeared in equality

$$L_R(\boldsymbol{\phi}, \mathbf{w}) = \prod_{t=r+1}^n \frac{1}{1 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_r x_{t-r})^2}$$

$$= \int_0^\infty \dots \int_0^\infty e^{-\sum_{t=r+1}^n u_t (1 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \sum_{v=1}^r \phi_v x_{t-v})^2)} du_{r+1} \dots du_n.$$

Then the conditional posterior distributions have closed forms given by

$$\left\{ \begin{array}{l} u_t \sim \text{Exp} \left(2 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \phi_1 x_{t-1} - \dots - \phi_r x_{t-r})^2 \right), \quad t = r + 1, \dots, n, \\ w_j \sim \text{gIG} \left(2 \sum_{t=r+1}^n u_t x_{t-j}^2 (x_t - \phi_1 x_{t-1} - \dots - \phi_r x_{t-r})^2 + 1, 2\beta_0, -n \right), \quad j = 1, \dots, r, \\ \phi_l \sim N(\mu_l, \sigma_l^2), \quad l = 1, \dots, r, \end{array} \right. \tag{17}$$

where gIG stands for the generalized inverse Gaussian distribution and

$$\mu_l = \frac{\sum_{t=r+1}^n \sum_{j=1}^r u_t w_j x_{t-j}^2 x_t x_{t-l} - \sum_{t=r+1}^n \sum_{j=1}^r u_t w_j x_{t-j}^2 \sum_{k \neq l} x_{t-l} x_{t-k} \phi_k}{\sum_{t=r+1}^n \sum_{j=1}^r u_t w_j x_{t-j}^2 x_{t-l}^2 + \frac{1}{2\beta_0}},$$

$$\sigma_l = \left[2 \sum_{t=r+1}^n \sum_{j=1}^r u_t w_j x_{t-j}^2 x_{t-l}^2 + \frac{1}{\beta_0} \right]^{-0.5}.$$

4.3. EM Algorithm for Estimating Hyper-parameters

One of the major challenges in sampling from the conditional posterior densities (17) is that the hyper-parameters β_0 and σ_0^2 are unknown in practice. However, there are several methods to estimate these quantities. In this subsection, we introduce an EM algorithm for this purpose.

From (15) and (16), the marginal density of the observations, \mathbf{x} , is obtained as

$$f(\mathbf{x} | \beta_0, \sigma_0^2, x_1, \dots, x_r) = \int \dots \int e^{-\sum_{t=r+1}^n s_t (1 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \sum_{v=1}^r \phi_v x_{t-v})^2)}$$

$$\times \frac{1}{\sigma_0^2} e^{-\frac{1}{2\sigma_0^2} \sum_{j=1}^r \phi_j^2} \times \prod_{j=1}^r \frac{\beta_0^n}{w_j^{n+1}} e^{-\frac{\beta_0}{w_j}} ds dw d\boldsymbol{\phi}. \tag{18}$$

Now consider the logarithm of the integrand in (18) as a function of β_0 and σ_0^2 given by

$$I(\beta_0, \sigma_0^2) = - \sum_{t=r+1}^n s_t \left(1 + \sum_{j=1}^r w_j x_{t-j}^2 \left(x_t - \sum_{v=1}^r \phi_v x_{t-v} \right)^2 \right) - \frac{r}{2} \log \sigma_0$$

$$- \frac{1}{2\sigma_0^2} \sum_{j=1}^r \phi_j^2 + nr \log \beta_0 - (n + 1) \sum_{j=1}^r \log w_j - \sum_{j=1}^r \frac{\beta_0}{w_j}.$$
(19)

Our proposed EM algorithm for estimating β_0 and σ_0 has the following steps:

1. Select the initial estimates $\beta_0^{(0)} (> 0)$ and $\sigma_0^{2(0)} (> 0)$ and start with $k = 1$.
2. Given $\beta_0^{(k)} (> 0)$ and $\sigma_0^{2(k)} (> 0)$, produce a sample of size N from the conditional distributions (17).

3. E-step: Calculate the averaged $I(\beta_0, \sigma_0^2)$ given by (19) using the generated samples in Step 2. Here, it should be noted that it is enough to calculate the mean of that expression in $I(\beta_0, \sigma_0^2)$ which only depend on β_0 and σ_0^2 .
4. M-step: Update β_0 and σ_0^2 by maximizing the following two quantities with respect to β_0 and σ_0^2 respectively:

$$\begin{cases} Q_3(\sigma_0^2) = \frac{r}{4} \log \sigma_0^2 + \frac{1}{2\sigma_0^{2(k)}} \sum_{j=1}^r E_{\sigma_0^2, \beta_0^{(k)}} [\phi_j^2], \\ Q_4(\beta_0) = nr \log \beta_0 - (n+1) \sum_{j=1}^r E_{\sigma_0^{2(k)}, \beta_0^{(k)}} \log w_j - \beta_0 \sum_{j=1}^r E_{\sigma_0^{2(k)}, \beta_0^{(k)}} \left[\frac{1}{w_j} \right]. \end{cases}$$

The updated estimates at time $(k + 1)$ are obtained as

$$\sigma_0^{2(k+1)} = \frac{1}{r} \sum_{j=1}^r E_{\sigma_0^{2(k)}, \beta_0^{(k)}} (\phi_j^2), \quad \beta_0^{(k+1)} = \frac{nc}{\sum_{j=1}^r E_{\sigma_0^{2(k)}, \beta_0^{(k)}} \left[\frac{1}{w_j} \right]}.$$

Repeat Steps 2-4 until the algorithm converges.

4.4. Model Evaluation Criteria

In practice, there are several criteria in literature for evaluating the efficiency of time series models. One of those is the mean squared prediction error defined as $MSPE = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2$. In this article, we will use a similar criterion to BIC. This criterion, for AR(r) model, is constructed based on the REL function, abbreviated as EBIC, defined as

$$EBIC = r \log n + \sum_{t=1}^n \log \left\{ 1 + \sum_{j=1}^r w_j x_{t-j}^2 (x_t - \hat{\phi}_1 x_{t-1} - \hat{\phi}_2 x_{t-2} - \dots - \hat{\phi}_r x_{t-r})^2 \right\}.$$

In practice, a small EBIC value confirms the adequacy of the model, whereas a large value indicates the inadequacy of the model.

5. SIMULATION STUDY AND APPLICATION

In this section, we first carry out a simulation study to confirm our theoretical results and then apply the proposed methods to a real dataset.

5.1. Simulation Study

Our simulation study focuses on the following autoregressive model of order 2:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t, \quad t = 3, 4, \dots, n.$$

For this model, we assume several scenarios for $\phi = (\phi_1, \phi_2)$ and n . In order for the AR(2) model to look like a nonparametric model, we assume that ϵ_t 's are generated from the following a little more complex distribution (a combination of normal and Cauchy distributions)

$$\epsilon_i \sim 0.5 N(0, 2) + 0.5 C(0, 1).$$

Here, we run the corresponding Gibbs sampling scheme (12) for two sample sizes $n = 50$ and 500 . With the generated data under each scenario, we try to fit the following three models:

- Model I:

$$X_t = \phi_1 X_{t-1} + \epsilon_t$$

with estimating equation $g(X, \phi_1) = X_{t-1}(X_t - \phi_1 X_{t-1})$.

· Model II:

$$X_t = \phi_2 X_{t-2} + \epsilon_t$$

with estimating equation $g(X, \phi_2) = X_{t-2}(X_t - \phi_2 X_{t-2})$.

· Model III:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$$

with estimating equations $g_1(X, \phi_1, \phi_2) = X_{t-1}(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})$ and $g_2(X, \phi_1, \phi_2) = X_{t-2}(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})$.

Moreover, for simplicity, we assume $w := w_1 = w_2$. We ran the sampling scheme (17) in order to generate 5000 samples from the marginal distributions of ϕ_1 and ϕ_2 , and each time we fit Models I–III to the data. In addition to the estimates of (ϕ_1, ϕ_2) and their standard errors, the MSPE and EBIC criteria are also computed. The numerical results are presented in Table 1. As we can see from it, Model III fits the data much better than the other two competitor models.

Table 1 | The REL and EL estimates of (ϕ_1, ϕ_2) under various models (and their standard errors inside the parentheses) and the associated MSPE and EBIC.

	(ϕ_1, ϕ_2)	Method	$\hat{\phi}_1$	$\hat{\phi}_2$	MSPE	EBIC
Model I	$n = 50$	REL	-0.319(0.127)	-	50.941	10.693
		EL	-0.221(0.191)	-	54.627	
	$n = 500$	REL	-0.355(0.108)	-	34.713	10.261
		EL	-0.312(0.123)	-	39.101	
		REL	-0.302(0.168)	-	58.121	14.180
		EL	-0.223(0.172)	-	61.001	
Model II	$n = 50$	REL	0.368(0.143)	-	33.975	13.990
		EL	0.311(0.144)	-	42.214	
	$n = 500$	REL	-	-0.146(0.221)	55.914	8.573
		EL	-	-0.121(0.243)	67.423	
		REL	-	-0.417(0.197)	39.128	7.857
		EL	-	0.314(0.202)	44.121	
Model III	$n = 50$	REL	-	-0.586(0.200)	56.842	10.000
		EL	-	-0.372(0.213)	71.001	
	$n = 500$	REL	-	-0.222(0.209)	28.613	8.156
		EL	-	0.199(0.266)	37.813	
		REL	-0.547(0.032)	-0.767(0.047)	20.542	4.705
		EL	-0.570(0.035)	-0.726(0.052)	22.117	
$n = 50$	REL	0.546(0.041)	-0.642(0.032)	21.222	4.440	

(Continued)

Table 1 | The REL and EL estimates of (ϕ_1, ϕ_2) under various models (and their standard errors inside the parentheses) and the associated MSPE and EBIC. (Continued)

	(ϕ_1, ϕ_2)	Method	$\hat{\phi}_1$	$\hat{\phi}_2$	MSPE	EBIC
Model III $n = 500$	(-0.5, -0.8)	EL	0.561(0.043)	-0.689(0.033)	22.473	
		REL	-0.505(0.012)	-0.805(0.014)	17.124	6.805
	(0.5, -0.8)	EL	-0.515(0.013)	-0.821(0.017)	17.921	
		REL	0.491(0.011)	-0.798(0.011)	16.978	6.508
		EL	0.484(0.012)	-0.811(0.012)	17.008	

Table 2 | Total Index of Consumer Goods and Services and their fitted values in urban areas of Iran from 1990 to 2017.

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
x_t	1.00	1.20	1.50	1.80	2.50	3.70	4.50	5.30	6.30	7.50	8.50	9.40	10.90	12.60
\hat{x}_t	-	-	1.14	1.51	1.87	2.79	4.22	4.92	5.85	7.09	8.55	9.70	10.86	12.92
Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
x_t	14.50	16.10	18.00	21.30	26.70	29.50	33.20	40.30	52.60	70.90	81.90	91.70	100.00	109.6
\hat{x}_t	15.04	17.39	19.31	21.85	26.19	32.76	34.96	39.65	48.56	62.91	82.94	90.53	100.02	108.18

5.2. Application

In this subsection, we analyze a real dataset on the Total Index of Consumer Goods and Services in urban areas of Iran for the period 1990–2017, extracted by The Central Bank of the Islamic Republic of Iran. The data are shown in Table 2 which demonstrates the presence of time dependency. A primary analysis reveals that an AR(2) model fits the data well after removing the trend. For this purpose, we first estimate the trend using a cubic polynomial given by

$$\begin{aligned} x_t &= \alpha + \beta t^3 + \epsilon_t, \\ \epsilon_t &= \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + u_t, \end{aligned} \tag{20}$$

where the error terms u_t 's have an unknown distribution with mean 0 and variance σ^2 . With the cubic polynomial, we obtain the estimates $\hat{\alpha} = -0.639$ and $\hat{\beta} = 0.004$, by fitting a linear regression model. Kolmogorov–Smirnov test statistic for ϵ_t is equal to 0.233 with $Sig = 0.000$. This shows that the residuals do not have a normal distribution. Since the size of the data is not large enough, there is no strong evidence that

$$\hat{\epsilon}_t = x_t - \hat{\alpha} - \hat{\beta}t^3$$

follow any particular distribution. Thus, we will use our proposed REL method to fit a time series model to $\hat{\epsilon}_t$'s. Using the posterior conditional distributions (17), the Bayesian estimates of ϕ_1 and ϕ_2 in model (20) are $\hat{\phi}_1 = 1.335$ and $\hat{\phi}_2 = -0.465$ respectively. The results indicate that the two-level model (20) fits the data well with the determination coefficient $R^2 = 0.996$, $MSPE = 3.96$, and $EBIC = 6.327$.

6. CONCLUSION

In this paper, a Bayesian method using the REL has been proposed for fitting the AR(r) time series model. An useful algorithm for computing the posterior has been developed for this approach. Moreover, the competitive performance of the proposed method has been clearly demonstrated via both simulation studies and a real data application. It is important to note that this method can be easily applied to other general linear models.

REFERENCES

1. D.R. Thomas, G.L. Grunkemeier, J. Am. Stat. Assoc. 70 (1975), 865–871.
2. A.Owen, Biometrika. 75 (1988), 237–249.
3. A. Owen, Ann. Stat. 18 (1990), 90–120.
4. T.J. DiCiccio, J.P. Romano, Biometrika. 76 (1989), 447–456.
5. P. Hall, B. La Scala, Internat. Statist. Review. 58 (1990), 109–127.
6. J. Qin, J. Lawless, Ann. Stat. 22 (1994), 300–325.

7. W.K. Newey, R.J. Smith, *Econometrica*. 72 (2004), 219–255.
8. S.X. Chen, H. Cui, *Biometrika*. 93 (2007), 215–220.
9. S.X. Chen, L. Peng, Y.L. Qin, *Biometrika*. 96 (2009), 711–722.
10. N.L. Hjort, I.W. McKeague, I.V. Keilegom, *Ann. Stat.* 37 (2009), 1079–1111.
11. C.Y. Tang, C. Leng, *Biometrika*. 97 (2010), 905–920.
12. C. Leng, C.Y. Tang, *Biometrika*. 99 (2012), 703–716.
13. V.J. Chang, X. Chen, *Econometrics*. 185 (2015), 283–304.
14. M. Tsao, *Ann. Stat.* 32 (2004), 1215–1221.
15. P.A. Mykland, *Ann. Stat.* 23 (1995), 396–421.
16. Y. Kitamura, *Ann. Stat.* 25 (1997), 2084–2102.
17. N.H. Chan, S. Ling, *Econ. Theory*. 22 (2006), 402–428.
18. N.H. Chan, L. Liu, *Electron. J. Stat.* 40 (2010), 221–238.
19. A.C. Monti, *Biometrika*. 84 (1997), 395–405.
20. P. Whittle, *Arkiv Matematik*. 2 (1953), 423–434.
21. C.Y. Yau, *J. Time Ser. Anal.* 33 (2012), 269–275.
22. D.J. Nordman, S.N. Lahiri, *Ann. Stat.* 34 (2006), 3019–3050.
23. D.J. Nordman, S.N. Lahiri, *Plan. Inference*. 155 (2014), 1–18.
24. M. Bayati, S.K. Ghoreishi, J. Wu, *Comput. Stat.* (2020).
25. C.Y. Tang, T.T. Wu, *J. Stat. Comput. Simul.* 84 (2014), 1917–1930.
26. A. Bhattacharya, D. Pati, N.S. Pillai, D.B. Dunson, Dirichlet-Laplace priors for optimal shrinkage, arXiv:1401.5398v1, 2014.
27. S.K. Ghoreishi, *J. Stat. Theory Appl.* 16 (2017), 53–64.