Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p123

A generalized time series model based on Kumaraswamy distribution to predict double-bounded relative humidity data

By Shad, Sharma, Singh

Published: 20 May 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 123-144
DOI: 10.1285/i20705948v15n1p123

A generalized time series model based on Kumaraswamy distribution to predict double-bounded relative humidity data

Mohammad Shad*, Y. D. Sharma, and Abhishek Singh

National Institute of Technology, Hamirpur (H.P.), India

Published: 20 May 2022

In this study, Kumaraswamy seasonal autoregressive moving average (KSARMA) model was developed to predict double-bounded relative humidity time-series data. In the proposed model, we used the conditional maximum-likelihood method to estimate parameters of the model. For the conditional score vector and conditional Fisher information matrix, the closed type expression were derived. This paper conjointly discusses interval estimation, hypothesis testing, model selection and forecasting. We also used a Monte Carlo simulation to evaluate the finite sample performance of conditional likelihood estimators (CMLEs) and white noise test.

keywords: ARMA, Kumaraswamy distribution, conditional likelihood, double bounded data, seasonal time series, forecast.

1 Introduction

The Kumaraswamy distribution is one of the distributions that is especially useful in biomedical and epidemiological research (Nadarajah, 2008) for several natural phenomena and their outcomes have lower and upper limits or bounded outcomes. Let a random variable $\tilde{X}$ ˜follows a continuous distribution, i.e., Kumaraswamy distribution with supports in the interval (a, b), if its probability density function is expressed by (Mitnik, 2013).

*Corresponding author: mohammad@nith.ac.in

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

$$f(\tilde{x},\phi,\gamma)=\frac{\phi\gamma}{b-a}\Big(\frac{\tilde{x}-a}{b-a}\Big)^{\phi-1}\Big[1-(\frac{\tilde{x}-a}{b-a})^{\phi}\Big]^{\gamma-1} \tag{1}$$

for $a<\tilde{x}<b$. Where $\phi>0$ and $\gamma>0$ is the shape parameters. CDF (the cumulative distribution function) and quantile functions are, represented as

$$F(\tilde{x})=1-\Big[1-\Big(\frac{\tilde{x}-a}{b-a}\Big)^{\phi}\Big]^{\gamma}\ and\ F^{-1}(k)=a+(b-a)\Big[1-(1-\upsilon)^{\frac{1}{\gamma}}\Big]^{\frac{1}{\phi}},\ \ 0<k<1. \tag{2}$$

The mean and variance of $\tilde{X}$ are expressed by

$$\mathbb{E}(\tilde{X})=a+(b-a)\gamma B\Big(1+\frac{1}{\phi},\gamma\Big)\ and\ Var(\tilde{X})=(b-a)^2\Big\{\gamma B\Big(1+\frac{2}{\phi},\gamma\Big)-\Big[\gamma B(1+\frac{1}{\phi},\gamma\Big)\Big]^2\Big\} \tag{3}$$

here $B(.,.)$ denotes the beta function (Gupta and Nadarajah, 2004). The Kumaraswamy distribution is the most commonly used distribution and it is a more compatible substitute in place of the beta distribution (Gupta and Nadarajah, 2004; Jones, 2009; Mitnik, 2013).

Beta regression based model (Ferrari and Cribari-Neto, 2004) which shows similarity with the generalized linear model (GLM) (McCullagh and Nelder, 1989), has also been generalized, improved, and implemented in many works (Cribari-Neto and Zeileis, 2010). Even in the literature, time series models based on beta distribution have gotten a lot of attention. For example, it can be seen in (Palm and Bayer, 2018; Guolo and Varin, 2014; Rocha and Cribari-Neto, 2009; da Silva et al., 2011).

The beta distribution is very flexible and it has a broader range of applications to improves its analytical utility (Nadarajah, 2008; Jones, 2009; Lemonte et al., 2013). For the hydrological process, for example daily rainfall, daily stream-flow, beta distribution does not fit satisfactorily (Kumaraswamy, 1976, 1980; Lemonte et al., 2013) but Kumaraswamy distribution is considered as a good substitute for beta distribution in hydrology and related fields (Nadarajah, 2008; Lemonte et al., 2013).

Kumaraswamy distribution is also widely used to investigate the temporal and spatial pattern in demographic data observed over time which shows values laying in the double bounded interval (a, b) such as rates and proportions. Despite its wide variety of uses, the Kumaraswamy distribution is still unknown to statisticians. For the mean and variance, the absence of closed form expression, has obstructed its utilization in the modeling purposes. Mitnik and Baek (2013), resolve this problem by proposing a median-based re-parameterization to promote its use in regression-based models. An expression of the median for the Kumaraswamy distribution is expressed by:

$$md(\tilde{X})=\tilde{\upsilon}=a+(b-a)\Big(1-0.5^{\frac{1}{\gamma}}\Big)^{\frac{1}{\phi}} \tag{4}$$

where, the median of the rescaled variable $X=\frac{\tilde{X}-a}{b-a}\in(0,1)$ is $(1-0.5^{\frac{1}{\gamma}})^{\frac{1}{\phi}}=\upsilon$.

We adopted the same structure as the generalized autoregressive moving average model (GARMA) (Benjamin et al., 2003) to construct the proposed model. Still, in terms of its median, we can also allocate a parametrization of the Kumaraswamy distribution.

The objective of this paper is to introduce a class of KSARMA models to predict double-bounded relative humidity time-series data. Besides, obtain the estimates of its parameters, confidence intervals and hypothesis testing (including seasonality test). For the model selection, we proposed various residuals that are used for defining goodness-of-fit and apply white noise test, which is established on these residuals and demonstrate how we can generate out-of-sample forecasts. In addition, we presented results with the help of Monte Carlo simulation, which is used for estimation of the finite sample performance of conditional maximum likelihood. Eventually, an empirical application is introduced and discussed.

2 The Proposed model

My objective is to establish the proposed model for the random variables which is observed over time distributed by Kumaraswamy. To alleviate the existence of the variable's serial correlation in the Kumaraswamy distribution's conditional median, we shall present a seasonal autoregreesive moving average (SARMA) time series structure.
Let $\{\tilde{X}_t\}_{t\,\in\,\mathbb{Z}}$, be a stochastic process, where $\tilde{X}_t \in (a, b)$ and $a, b \in \mathbb{R}$ with $a < b$.. Assume that for a given previous information set $\mathcal{G}_{t-1}$, the conditional distribution of each $\tilde{X}_t$, follows the Kumaraswamy distribution i.e. $K(\tilde{\upsilon}_t, \phi, a, b)$. For given $\mathcal{G}_{t-1}$, the conditional density of $\tilde{X}_t$ is

$$f_{\upsilon_t}(\tilde{x_t} \mid \mathcal{G}_{t-1}) = (\frac{1}{b-a})\frac{\phi log(0.5)}{log(1-{\upsilon_t}^{\phi})}\ x_t^{\phi-1}(1-x_t^{\phi})^{\frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}-1},\quad 0 < \upsilon_t < 1,\ \phi > 0 \quad (5)$$

for $0 < x_t < 1$, where υ_t (mean) and ϕ (precision) are distribution parameters and $x_t = \frac{\tilde{x_t}-a}{b-a}$, $\upsilon_t = \frac{\tilde{\upsilon_t}-a}{b-a}$. This specific type of density is desirable because it permits modeling with no transformation, this is usually done in literature (Rocha and Cribari-Neto, 2009), however managing the easier distribution of X_t . CDF and quantile functions, are given by

$$F_{\upsilon_t}(\tilde{x_t} \mid \mathcal{G}_{t-1}) = 1 - (1-x_t^{\phi})^{\frac{log(0.5)}{log(1-\upsilon_t^{\phi})}} \quad (6)$$

$$F_{\upsilon_t}{}^{-1}(\tilde{x_t} \mid \mathcal{G}_{t-1}) = a + (b-a)\Big[1-(1-u)^{\frac{log(1-\upsilon_t^{\phi})}{log(0.5)}}\Big]^{\frac{1}{\phi}} \quad (7)$$

Conditional mean and conditional variance of $\tilde{X}_t$ are expressed as below

$$\mathbb{E}(\tilde{X}_t \mid \mathcal{G}_{t-1}) = a + (b-a)\frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}B\Big(1+\frac{1}{\phi}, \frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}\Big) \quad (8)$$

$$Var(\tilde{X}_t \mid \mathcal{G}_{t-1}) = \frac{1}{b-a}\Big\{\frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}B\Big(1+\frac{2}{\phi}, \frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}\Big) - \quad (9)$$

$$\Big[\frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}B\Big(1+\frac{1}{\phi}, \frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}\Big)\Big]^2\Big\}$$

Now we define the seasonal autoregressive moving average model i.e SARMA$(p,q) \times (P,Q)_S$.

$$\Phi(B^S)\varphi(B)g(x_t) = \alpha + \Theta(B^S)\theta(B)\epsilon_t$$

$$g(x_t) = \alpha + \sum_{i=1}^{p}\varphi_i g(x_{t-i}) + \sum_{j=1}^{q}\theta_j\epsilon_{t-j} + \sum_{I=1}^{P}\Phi_I g(x_{t-IS}) + \sum_{J=1}^{Q}\Theta_J\epsilon_{t-JS} - \sum_{i=1}^{p}\sum_{I=1}^{P}\varphi_i\Phi_I g\Big(x_{t-(i+IS)}\Big) + \epsilon_t + \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\epsilon_{t-(j+JS)}$$

$$\eta_t = g(\upsilon_t) = \alpha + \sum_{i=1}^{p}\varphi_i g(x_{t-i}) + \sum_{j=1}^{q}\theta_j\epsilon_{t-j} + \sum_{I=1}^{P}\Phi_I g(x_{t-IS}) + \sum_{J=1}^{Q}\Theta_J\epsilon_{t-JS} - \sum_{i=1}^{p}\sum_{I=1}^{P}\varphi_i\Phi_I g\Big(x_{t-(i+IS)}\Big) + \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\epsilon_{t-(j+JS)} \tag{10}$$

where, $g\colon \mathbb{R} \to (0,1)$ be link function which is twice differentiable and strictly monotonic then $g^{-1}\colon \mathbb{R} \to (0,1)$ exists and is also twice differentiable, $\alpha \in \mathbb{R}$, $\epsilon_t = g(x_t) - g(\upsilon_t)$, $\varphi(B)$, $\theta(B)$ is the autoregressive and moving average polynomial given by $\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - ... - \varphi_p B^p)$, $\theta(B) = (1 + \theta_1 B + \theta_2 B^2 + ... + \theta_q B^q)$ and $\Phi(B)$, $\Theta(B)$ is the seasonal autoregressive and seasonal moving average polynomial is given by $\Phi(B) = (1 - \Phi_1 B^S - \Phi_2 B^{2S} - ... - \Phi_P B^{PS})$, $\Theta(B) = (1 + \Theta_1 B^S + \Theta_2 B^{2S} + ... + \Theta_Q B^{QS})$, S is representing the seasonality frequency (For monthly data, we will take S=12 and for quarterly data S=4).

We noticed that all the functions could be added to the model, such as logit, probit, loglog, since $\upsilon_t \in (0,1)$.

The proposed Kumaraswamy seasonal autoregressive moving average (KSARMA) model is given by equations (5) and (10).

2.1 Some particular cases

KSARMA $(p,q) \times (P,Q)_S$ includes a number of important models. The following is a list of some of them.

2.1.1 KSARMA $(1,1) \times (1,1)_{12}$

The KSARMA $(1,1) \times (1,1)_{12}$ model can be expressed as

$$\Phi(B^{12})\varphi(B)g(x_t) = \alpha + \Theta(B^{12})\theta(B)\epsilon_t$$

$$(1 - \Phi_1 B^{12})(1 - \varphi_1 B)g(x_t) = \alpha + (1 + \Theta_1 B^{12})(1 + \theta_1 B)\epsilon_t$$

$$g(x_t) - \varphi_1 g(x_{t-1}) - \Phi_1 g(x_{t-12}) + \varphi_1\Phi_1 g(x_{t-13}) = \alpha + \epsilon_t + \theta_1\epsilon_{t-1} + \Theta_1\epsilon_{t-12} + \theta_1\Theta_1\epsilon_{t-13}$$

$$g(\upsilon_t) = \alpha + \varphi_1 g(x_{t-1}) + \Phi_1 g(x_{t-12}) - \varphi_1\Phi_1 g(x_{t-13}) + \theta_1\epsilon_{t-1} + \Theta_1\epsilon_{t-12} + \theta_1\Theta_1\epsilon_{t-13}.$$

2.1.2 KSARMA $(2,1)\times(2,2)_{12}$

The KSARMA $(2,1)\times(2,2)_{12}$ model can be expressed as

$\Phi(B^{12})\varphi(B)g(x_t) = \alpha + \Theta(B^{12})\theta(B)\epsilon_t$

$(1-\Phi_1 B^{12} - \Phi_2 B^{24})(1-\varphi_1 B - \varphi_2 B^2)g(x_t) = \alpha + (1+\Theta_1 B^{12} + \Theta_2 B^{24})(1+\theta_1 B)\epsilon_t$

$g(x_t) - \varphi_1 g(x_{t-1}) - \varphi_2 g(x_{t-2}) - \Phi_1 g(x_{t-12}) - \Phi_2 g(x_{t-24}) + \varphi_1\Phi_1 g(x_{t-13}) + \varphi_2\Phi_1 g(x_{t-14})$

$+\varphi_1\Phi_2 g(x_{t-25}) + \varphi_2\Phi_2 g(x_{t-26}) = \alpha + \epsilon_t + \theta_1\epsilon_{t-1} + \Theta_1\epsilon_{t-12} + \Theta_2\epsilon_{t-24} + \theta_1\Theta_1\epsilon_{t-13} + \theta_1\Theta_2\epsilon_{t-25}$

$g(\upsilon_t) = \alpha + \varphi_1 g(x_{t-1}) + \varphi_2 g(x_{t-2}) + \Phi_1 g(x_{t-12}) + \Phi_2 g(x_{t-24}) - \varphi_1\Phi_1 g(x_{t-13}) - \varphi_2\Phi_1 g(x_{t-14})$

$-\varphi_1\Phi_2 g(x_{t-25}) - \varphi_2\Phi_2 g(x_{t-26}) + \theta_1\epsilon_{t-1} + \Theta_1\epsilon_{t-12} + \Theta_2\epsilon_{t-24} + \theta_1\Theta_1\epsilon_{t-13} + \theta_1\Theta_2\epsilon_{t-25}.$

3 Parameter Estimation

In the proposed KSARMA model, we can derive the conditional maximum likelihood estimator for the parameters. Let $\Omega = (\alpha, \varphi^T, \theta^T, \Phi^T, \Theta^T, \phi)^T$ be the $(p+q+P+Q+2)$ dimensional vector. By maximizing of the logarithm of the conditional likelihood function, we can acquire the conditional maximum likelihood estimators (CMLE) of Ω. The m initial observations for the log-likelihood function is conditional for Ω which can be expressed as

$$l = l(\Omega; \tilde{x}_t) = \sum_{t=m+1}^{n} log(f(\tilde{x}_t \mid \mathcal{G}_{t-1})) = \sum_{t=m+1}^{n} l_t(\upsilon_t, \phi) \tag{11}$$

where,

$$l_t(\upsilon_t,\phi) = log(\phi) - log(b-a) + log\Big(log(0.5)\Big) - log\Big(log(1-\upsilon_t^{\phi})\Big) + (\phi-1)log(x_t).$$
$$+\tfrac{log(0.5)}{log(1-\upsilon_t^{\phi})} log(1-x_t^{\phi}) - log(1-{x_t}^{\phi})$$

and m = max (PS + p, QS + q)

3.1 Conditional score vector

For the first derivative of log-likelihood $l(\Omega)$, we can calculate the Conditional score vector. Let $\Omega = (\alpha, \varphi^T, \Phi^T, \theta^T, \Theta^T)$. Differentiating equation (11), with regard to the *ith* element of the vector Ω, $\Omega_i \neq \phi$,with i = 1, ..., $(p+q+P+Q+1)$, we get

$$\frac{\partial l}{\partial \Omega_i} = \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \Omega_i} \tag{12}$$

observe that,

$$\frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t} = \frac{\phi {\upsilon_t}^{\phi-1}}{(1-{\upsilon_t}^{\phi})log(1-\upsilon_t)^{\phi}}\Big(\gamma_t log(1-{x_t}^{\phi}) + 1\Big) = \phi c_t \tag{13}$$

where,

$$c_t = \frac{{\upsilon_t}^{\phi-1}}{(1-{\upsilon_t}^{\phi})log(1-\upsilon_t)^{\phi}}\Big(\gamma_t log(1-{x_t}^{\phi})+1\Big), \gamma_t = \frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})} \tag{14}$$

since, $\eta_t = g(\upsilon_t)$, it follows that $\frac{d\upsilon_t}{d\eta_t} = \frac{1}{g'(\upsilon_t)}$

Using equation (13) and (14) in (12), we get

$$\frac{\partial l}{\partial \Omega_i} = \sum_{t=m+1}^{n} \frac{\phi {\upsilon_t}^{\phi-1}}{(1-{\upsilon_t}^{\phi})log(1-\upsilon_t)^{\phi}}\Big(\gamma_t log(1-{x_t}^{\phi})+1\Big)\frac{\partial \eta_t}{\partial \Omega_i} = \sum_{t=m+1}^{n} \phi \frac{c_t}{g'(\upsilon_t)}\frac{\partial \eta_t}{\partial \Omega_i}$$

For the computation of the score vector, we obtain the $\frac{\partial \eta_t}{\partial \Omega_i}$ for every coordinate $\Omega_i \neq \phi$ of Ω.

To obtain the derivative of the linear predictor η_t with regard to α, let the error term $\epsilon_t = g(x_t) - g(\upsilon_t)$, so that

$$\frac{\partial \eta_t}{\partial \alpha} = 1 + \sum_{j=1}^{q}\theta_j \frac{\partial \epsilon_{t-j}}{\partial \alpha} + \sum_{J=1}^{Q}\Theta_J \frac{\partial \epsilon_{t-JS}}{\partial \alpha} + \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J \frac{\partial \epsilon_{t-(j+JS)}}{\partial \alpha}$$

$$\frac{\partial \eta_t}{\partial \alpha} = 1 - \sum_{j=1}^{q}\theta_j \frac{\partial \eta_{t-j}}{\partial \alpha} - \sum_{J=1}^{Q}\Theta_J \frac{\partial \eta_{t-JS}}{\partial \alpha} - \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J \frac{\partial \eta_{t-(j+JS)}}{\partial \alpha} \qquad \text{(since, } \epsilon_t = g(x_t) - \eta_t)$$

Regarding the remaining parameters, the derivatives of linear predictor are given by

$$\frac{\partial \eta_t}{\partial \varphi_i} = g(x_{t-i})\Phi(B^S) - \sum_{j=1}^{q}\theta_j \frac{\partial \eta_{t-j}}{\partial \varphi_i} - \sum_{J=1}^{Q}\Theta_J \frac{\partial \eta_{t-JS}}{\partial \varphi_i} - \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J \frac{\partial \eta_{t-(j+JS)}}{\partial \varphi_i}, \quad i \in \{1, ..., p\},$$

$$\frac{\partial \eta_t}{\partial \Phi_I} = g(x_{t-IS})\varphi(B) - \sum_{j=1}^{q}\theta_j \frac{\partial \eta_{t-j}}{\partial \Phi_I} - \sum_{J=1}^{Q}\Theta_J \frac{\partial \eta_{t-JS}}{\partial \Phi_I} - \sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J \frac{\partial \eta_{t-(j+JS)}}{\partial \Phi_I}, \quad I \in \{1, ..., P\},$$

$$\frac{\partial \eta_t}{\partial \theta_j} = \epsilon_{t-j}\Theta(B^S) - \sum_{i=1}^{q}\theta_i \frac{\partial \eta_{t-i}}{\partial \theta_j} - \sum_{J=1}^{Q}\Theta_J \frac{\partial \eta_{t-JS}}{\partial \theta_j} - \sum_{i=1}^{q}\sum_{J=1}^{Q}\theta_i\Theta_J \frac{\partial \eta_{t-(i+JS)}}{\partial \theta_j}, \quad j \in \{1, ..., q\},$$

$$\frac{\partial \eta_t}{\partial \Theta_J} = \epsilon_{t-JS}\theta(B) - \sum_{j=1}^{q}\theta_j \frac{\partial \eta_{t-j}}{\partial \Theta_J} - \sum_{i=1}^{Q}\Theta_i \frac{\partial \eta_{t-iS}}{\partial \Theta_J} - \sum_{j=1}^{q}\sum_{i=1}^{Q}\theta_j\Theta_i \frac{\partial \eta_{t-(j+iS)}}{\partial \Theta_J}, \quad J \in \{1, ..., Q\}$$

Differentiating (11) with respect to precision parameter ϕ, we get

$$\frac{\partial l_t}{\partial \phi} = \frac{1}{\phi} + log(y_t) + c_t\upsilon_t log(\upsilon_t) - (\gamma_t - 1)\frac{{x_t}^{\phi}log(x_t)}{(1-{x_t}^{\phi})}$$

Now, differentiating (11) with respect to parameter α, we get

$$\frac{\partial l}{\partial \alpha} = \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \alpha}$$

$$\frac{\partial l}{\partial \alpha} = \sum_{t=m+1}^{n} \Big[\phi c_t \frac{1}{g'(\upsilon_t)}\Big\{1-\sum_{j=1}^{q}\theta_j\frac{\partial \eta_{t-j}}{\partial \alpha}-\sum_{J=1}^{Q}\Theta_J\frac{\partial \eta_{t-JS}}{\partial \alpha}-\sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial \eta_{t-(j+JS)}}{\partial \alpha}\Big\}\Big] \tag{15}$$

$$\frac{\partial l}{\partial \varphi_i} = \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \varphi_i}$$

$$\frac{\partial l}{\partial \varphi_i} = \sum_{t=m+1}^{n} \Big[\phi c_t \frac{1}{g'(\upsilon_t)}\Big\{g(x_{t-i})\Phi(B^S)-\sum_{j=1}^{q}\theta_j\frac{\partial \eta_{t-j}}{\partial \varphi_i}-\sum_{J=1}^{Q}\Theta_J\frac{\partial \eta_{t-JS}}{\partial \varphi_i}- \tag{16}$$

$$\sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial \eta_{t-(j+JS)}}{\partial \varphi_i}\Big\}\Big]$$

$$\frac{\partial l}{\partial \Phi_I} = \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \Phi_I}$$

$$\frac{\partial l}{\partial \Phi_I} = \sum_{t=m+1}^{n} \Big[\phi c_t \frac{1}{g'(\upsilon_t)}\Big\{g(x_{t-IS})\varphi(B)-\sum_{j=1}^{q}\theta_j\frac{\partial \eta_{t-j}}{\partial \Phi_I}-\sum_{J=1}^{Q}\Theta_J\frac{\partial \eta_{t-JS}}{\partial \Phi_I}-\sum_{j=1}^{q}\sum_{J=1}^{Q}\theta_j\Theta_J\frac{\partial \eta_{t-(j+JS)}}{\partial \Phi_I}\Big\}\Big]\frac{\partial l}{\partial \theta_j}$$

$$= \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \theta_j}(17)$$

$$\frac{\partial l}{\partial \theta_j} = \sum_{t=m+1}^{n} \Big[\phi c_t \frac{1}{g'(\upsilon_t)}\Big\{\epsilon_{t-j}\Theta(B^S)-\sum_{i=1}^{q}\theta_i\frac{\partial \eta_{t-i}}{\partial \theta_j}-\sum_{J=1}^{Q}\Theta_J\frac{\partial \eta_{t-JS}}{\partial \theta_j}-\sum_{i=1}^{q}\sum_{J=1}^{Q}\theta_i\Theta_J\frac{\partial \eta_{t-(i+JS)}}{\partial \theta_j}\Big\}\Big] \tag{18}$$

$$\frac{\partial l}{\partial \Theta_J} = \sum_{t=m+1}^{n} \frac{\partial l_t(\upsilon_t,\phi)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \Theta_J}$$

$$\frac{\partial l}{\partial \Theta_J} = \sum_{t=m+1}^{n} \Big[\phi c_t \frac{1}{g'(\upsilon_t)}\Big\{\epsilon_{t-JS}\theta(B)-\sum_{j=1}^{q}\theta_j\frac{\partial \eta_{t-j}}{\partial \Theta_J}-\sum_{i=1}^{Q}\Theta_i\frac{\partial \eta_{t-iS}}{\partial \Theta_J}-\sum_{j=1}^{q}\sum_{i=1}^{Q}\theta_j\Theta_i\frac{\partial \eta_{t-(j+iS)}}{\partial \Theta_J}\Big\}\Big] \tag{19}$$

Therefore the element of score vector in matrix form can be written as
$U(\Omega) = \Big(U_\alpha(\Omega), U_\varphi(\Omega)^T, U_\theta(\Omega)^T, U_\Phi(\Omega)^T, U_\Theta(\Omega)^T, U_\phi(\Omega)\Big)^T$,
where, $U_\alpha(\Omega) = v^T TC$, $U_\varphi(\Omega) = A^T TC$, $U_\Phi(\Omega) = B^T TC$, $U_\theta(\Omega) = P^T TC$, $U_\Theta(\Omega) =$

R^TTC,

$U_\phi(\gamma) = \frac{n-m}{\phi} + \sum_{t=m+1}^{n} \Big\{log(x_t) + c_t\upsilon_t log(\upsilon_t) - \frac{(\gamma_t - 1){x_t}^{\phi} log(x_t)}{(1-{x_t}^{\phi})}\Big\}$

where, $T = diag\{\frac{1}{g'(\upsilon_{m+1})}, ..., \frac{1}{g'(\upsilon_n)}\}$, $v = (\frac{\partial\eta_{m+1}}{\partial\alpha}, ..., \frac{\partial\eta_n}{\partial\alpha})$, and A is an matrix of order $(n-m)\times p$ where (i,j) component is provided by $A_{i,j} = \frac{\partial\eta_{i+m}}{\partial\varphi_j}$, B is an matrix of order $(n-m)\times P$ where (i,j) component is provided by $B_{i,j} = \frac{\partial\eta_{i+m}}{\partial\Phi_j}$, P is an $(n-m)\times q$ order matrix where (i,j) component is provided by $P_{i,j} = \frac{\partial\eta_{i+m}}{\partial\theta_j}$, and R is an matrix of order $(n-m)\times Q$ where (i,j) component is provided by $R_{i,j} = \frac{\partial\eta_{i+m}}{\partial\Theta_j}$, $C = (\phi c_{m+1}, ..., \phi c_n)^T$

CMLE of Ω is found by solving the given system of equations $U(\Omega) = \mathbf{0}$, wherever $\mathbf{0}$ in $\mathbb{R}^{p+q+P+Q+2}$ is the null vector.

3.2 Conditional Fisher's information matrix

By evaluating the expected value for all the second-order derivative, we obtained conditional Fisher information matrix. Let $\Omega = (\alpha, \varphi^T, \theta^T, \Phi^T, \Theta^T)^T$. It can be shown that, for $i, j \in \{1, ..., p+q+P+Q+1\}$

$$\begin{aligned}\frac{\partial^2 l(\Omega)}{\partial\Omega_i\partial\Omega_j} &= \sum_{t=m+1}^{n} \frac{\partial}{\partial\upsilon_t}\Big(\frac{\partial l_t(\upsilon_t,\phi)}{\partial\upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j}\Big)\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_i}\\ &= \sum_{t=m+1}^{n}\Big[\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2}\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j} + \frac{\partial l_t(\upsilon_t,\phi)}{\partial\upsilon_t}\frac{\partial}{\partial\upsilon_t}\Big(\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j}\Big)\Big]\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_i}\end{aligned}$$

Under the standard regularity condition $\mathbb{E}\Big(\frac{\partial l_t(\upsilon_t,\phi)}{\partial\upsilon_t}|\mathcal{G}_{t-1}\Big) = 0$ and thus,

$$\mathbb{E}\Big(\frac{\partial^2 l(\Omega)}{\partial\Omega_i\partial\Omega_j}|\mathcal{G}_{t-1}\Big) = \sum_{t=m+1}^{n}\mathbb{E}\Big(\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2}|\mathcal{G}_{t-1}\Big)\Big(\frac{d\upsilon_t}{d\eta_t}\Big)^2\frac{\partial\eta_t}{\partial\Omega_i}\frac{\partial\eta_t}{\partial\Omega_j} \tag{20}$$

Differentiating (13) with respect to υ_t, we get

$$\begin{aligned}\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2} &= \phi A_t + \Big(\frac{log(0.5)}{log(1-{\upsilon_t}^{\phi})}log(1-x_t^{\phi}) + 1\Big) + \frac{\phi^2 log(0.5)\lambda_2}{log(1-{\upsilon_t}^{\phi})}log(1-x_t^{\phi})\\ &= \phi A_t + \phi\gamma_t\Big[A_t + \phi\lambda_2\Big]log(1-x_t^{\phi})\end{aligned}$$

where, $A_t = \frac{\partial}{\partial\upsilon_t}\Big(\frac{\upsilon_t^{\phi-1}}{(1-\upsilon_t^{\phi})log(1-\upsilon_t^{\phi})}\Big) = \phi\lambda_2\Big[1 + log(1-\upsilon_t^{\phi})\Big] + (\phi-1)\lambda_1$ and $\lambda_k = \frac{\upsilon_t^{k\phi-2}}{(1-\upsilon_t^{\phi})^k log(1-\upsilon_t^{\phi})^k}$, we have $\mathbb{E}\Big[log(1-x_t^{\phi} \mid \mathcal{G}_{t-1})\Big] = \frac{-1}{\gamma_t}$ (Bayer et al., 2017), so that

$\mathbb{E}\Big(\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2} \mid \mathcal{G}_{t-1}\Big) = -\phi^2\lambda_2 = w_t.$

From equation (20), we have

$$\mathbb{E}\Big(\frac{\partial^2 l(\Omega)}{\partial\Omega_i\partial\Omega_j}|\mathcal{G}_{t-1}\Big) = \sum_{t=m+1}^{n}\frac{w_t}{g'(\upsilon_t)^2}\frac{\partial\eta_t}{\partial\Omega_i}\frac{\partial\eta_t}{\partial\Omega_j} \tag{21}$$

Now, differentiating l_t two times with respect to ϕ, we get

$$\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi^2} = \frac{-1}{\phi^2} + \frac{\partial c_t}{\partial\phi}\upsilon_t log(\upsilon_t) - \frac{\partial}{\partial\phi}\Big((\gamma_t - 1)\frac{x_t^{\phi} log(x_t)}{1 - x_t^{\phi}}\Big) \tag{22}$$

Differentiating (14) with respect to ϕ, we get

$$\frac{\partial c_t}{\partial\phi} = \gamma_t\upsilon_t\lambda_2 log(\upsilon_t) log(1-x_t^{\phi}) - \gamma_t\upsilon_t\lambda_1\Big[\frac{x_t^{\phi} log(x_t)}{1 - x_t^{\phi}}\Big] + c_t log(\upsilon_t)\Big(\frac{\upsilon_t}{(1-\upsilon_t^{\phi}) log(1-\upsilon_t^{\phi})} + \frac{1}{1-\upsilon_t^{\phi}}\Big) \tag{23}$$

and

$$\frac{\partial}{\partial\phi}\Big[\frac{x_t{}^{\phi} log(x_t)}{(1-x_t{}^{\phi})}(\gamma_t - 1)\Big] = (\gamma_t - 1)\frac{x_t^{\phi} log(x_t)^2}{(1-x_t^{\phi})^2} + \frac{\gamma_t\upsilon_t^{\phi} log(\upsilon_t)}{(1-\upsilon_t^{\phi}) log(1-\upsilon_t^{\phi})}\Big[\frac{x_t^{\phi} log(x_t)}{1-x_t^{\phi}}\Big] \tag{24}$$

Using (23) and (24) in (22), we obtain

$$\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi^2} = \frac{-1}{\phi^2} + \gamma_t\upsilon_t^2\lambda_2 log(\upsilon_t)^2 log(1-x_t^{\phi}) - 2\gamma_t\upsilon_t^2 log(\upsilon_t)\lambda_1\Big[\frac{x_t^{\phi} log(x_t)}{1-x_t^{\phi}}\Big]$$
$$-(\gamma_t - 1)\frac{x_t^{\phi} log(x_t)^2}{(1-x_t^{\phi})^2} + c_t\upsilon_t log(\upsilon_t)^2\Big(\lambda_1\upsilon_t^2 + \frac{1}{1-\upsilon_t^{\phi}}\Big) \tag{25}$$

On taking the conditional expectation of (25) and substituting the result from lemma 2 (Bayer et al., 2017), we have

$$\mathbb{E}\Big(\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi^2} \mid \mathcal{G}_{t-1}\Big) = \frac{m-n}{\phi^2} - \sum_{t=m+1}^{n}\Bigg\{\upsilon_t^2\lambda_2 log(\upsilon_t)^2 + 2\delta_t\upsilon_t^2 log(\upsilon_t)\lambda_1\Big(\frac{1-\psi(\gamma_t+1)-\kappa}{(\gamma_t-1)\phi}\Big)$$
$$+\frac{\psi(\gamma_t)\Big[\psi(\gamma_t)+2(\kappa-1)\Big]-\psi'(\gamma_t)+\kappa_0}{(\gamma_t-2)\phi^3}\Bigg\}$$

where ψ denotes the digamma function ψ' is represented as the trigamma function, $\kappa = 0.5772156649...$ is the Euler-Mascheroni constant (Gradshteyn and Ryzhik, 2007) and $\kappa_0 = \frac{\pi^2}{6} + \kappa^2 - 2\kappa$.

Since, $\frac{\partial l_t}{\partial\phi} = \frac{1}{\phi} + log(x_t) + c_t\upsilon_t log(\upsilon_t) - (\Omega_t - 1)\frac{x_t{}^{\phi} log(x_t)}{(1-x_t{}^{\phi})}$
therefore, differentiating above equation with respect to $\Omega_j \neq \varphi$, we have
$\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi\partial\Omega_j} = c_t\frac{\partial\upsilon_t log(\upsilon_t)}{\partial\Omega_j} + \upsilon_t log(\upsilon_t)\frac{\partial c_t}{\partial\Omega_j} - \frac{x_t^{\phi} log(x_t)}{1-x_t^{\phi}}\frac{\partial\gamma_t}{\partial\Omega_j}$
From equation (13), we have $c_t = \frac{1}{\phi}\frac{\partial l_t(\upsilon_t,\phi)}{\partial\upsilon_t}$ so that

$$\frac{\partial c_t}{\partial\Omega_j} = \frac{\partial c_t}{\partial\upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j} = \frac{1}{\phi g'(\upsilon_t)}\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2}\frac{\partial\eta_t}{\partial\Omega_j},$$

$$\frac{\partial\upsilon_t log(\upsilon_t)}{\partial\Omega_j} = \frac{\partial\upsilon_t log(\upsilon_t)}{\partial\upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j} = \Big(\frac{log(\upsilon_t)+1}{g'(\upsilon_t)}\Big)\frac{\partial\eta_t}{\partial\Omega_j}$$

From equation (13), we have $c_t = \frac{1}{\phi}\frac{\partial l_t(\upsilon_t,\phi)}{\partial\upsilon_t}$ so that

$$\frac{\partial c_t}{\partial\Omega_j} = \frac{\partial c_t}{\partial\upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial\eta_t}{\partial\Omega_j} = \frac{1}{\phi g'(\upsilon_t)}\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2}\frac{\partial\eta_t}{\partial\Omega_j},$$

$\frac{\partial \upsilon_t log(\upsilon_t)}{\partial \Omega_j} = \frac{\partial \upsilon_t log(\upsilon_t)}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \Omega_j} = \left(\frac{log(\upsilon_t)+1}{g'(\upsilon_t)}\right)\frac{\partial \eta_t}{\partial \Omega_j}$

and $\frac{\partial \gamma_t}{\partial \Omega_j} = \frac{\partial \gamma_t}{\partial \upsilon_t}\frac{d\upsilon_t}{d\eta_t}\frac{\partial \eta_t}{\partial \Omega_j} = \frac{1}{g'(\upsilon_t)}\frac{\partial \eta_t}{\partial \Omega_j}\left[\frac{log(0.5)\phi\upsilon_t^{\phi-1}}{log(1-\upsilon_t^{\phi})^2(1-\upsilon_t^{\phi})}\right] = \left(\frac{\phi\upsilon_t\Omega_t\lambda_1}{g'(\upsilon_t)}\right)\frac{\partial \eta_t}{\partial \Omega_j}$

Hence,

$\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi\partial\Omega_j} = \left\{\frac{1}{g'(\upsilon_t)}\left[c_t\Big(log(\upsilon_t)+1\Big) + \frac{\upsilon_t log(\upsilon_t)}{\phi}\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\upsilon_t^2} - \phi\upsilon_t\gamma_t\lambda_1\frac{x_t{}^{\phi}log(x_t)}{(1-x_t{}^{\phi})}\right]\right\}\frac{\partial\eta_t}{\partial\Omega_j}$

On taking conditional expectation of above equation,we get

$\mathbb{E}\Big(\frac{\partial^2 l_t(\upsilon_t,\phi)}{\partial\phi\partial\Omega_j} \mid \mathcal{G}_{t-1}\Big) = \sum_{t=m+1}^{n}\frac{d_t}{g'(\upsilon_t)}\frac{\partial\eta_t}{\partial\Omega_j}$

where, $d_t = -\phi\upsilon_t log(\upsilon_t)\lambda_2 - \phi\upsilon_t\gamma_t\lambda_1\Big(\frac{1-\Psi(\gamma_t+1)-\kappa}{(\gamma_t-1)\phi}\Big)$

Let $D = diag\{d_{m+1},...,d_n\}$, $W = diag\{w_{m+1},...,w_n\}$, $N = diag\Big\{\frac{\partial^2 l_t(\upsilon_{m+1},\phi)}{\partial\phi^2},...,\frac{\partial^2 l_n(\upsilon_n,\phi)}{\partial\phi^2}\Big\}$.
The conditional Fisher information matrix for Ω is

$$H = H(\Omega) = \begin{bmatrix} H_{(\alpha,\alpha)} & H_{(\alpha,\varphi)} & H_{(\alpha,\Phi)} & H_{(\alpha,\theta)} & H_{(\alpha,\Theta)} & H_{(\alpha,\phi)} \\ H_{(\varphi,\alpha)} & H_{(\varphi,\varphi)} & H_{(\varphi,\Phi)} & H_{(\varphi,\theta)} & H_{(\varphi,\Theta)} & H_{(\varphi,\phi)} \\ H_{(\Phi,\alpha)} & H_{(\Phi,\varphi)} & H_{(\Phi,\Phi)} & H_{(\Phi,\theta)} & H_{(\Phi,\Theta)} & H_{(\Phi,\phi)} \\ H_{(\theta,\alpha)} & H_{(\theta,\varphi)} & H_{(\theta,\Phi)} & H_{(\theta,\theta)} & H_{(\theta,\Theta)} & H_{(\theta,\phi)} \\ H_{(\Theta,\alpha)} & H_{(\Theta,\varphi)} & H_{(\Theta,\Phi)} & H_{(\Theta,\theta)} & H_{(\Theta,\Theta)} & H_{(\Theta,\phi)} \\ H_{(\phi,\alpha)} & H_{(\phi,\varphi)} & H_{(\phi,\Phi)} & H_{(\phi,\theta)} & H_{(\phi,\Theta)} & H_{(\phi,\phi)} \end{bmatrix}$$

where $K_{(\alpha,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l(\Omega)}{\partial\alpha^2} \mid \mathcal{G}_{t-1}\Big) = -v^TWT^2v$, $H_{(\alpha,\varphi)} = H^T_{(\varphi,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\alpha\partial\varphi} \mid \mathcal{G}_{t-1}\Big) = -v^TWT^2A$, $H_{(\alpha,\Phi)} = H^T_{(\Phi,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\alpha\partial\Phi} \mid \mathcal{G}_{t-1}\Big) = -v^TWT^2B$, $H_{(\alpha,\theta)} = H^T_{(\theta,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\alpha\partial\theta} \mid \mathcal{G}_{t-1}\Big) = -v^TWT^2P$, $H_{(\alpha,\Theta)} = H^T_{(\Theta,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\alpha\partial\Theta} \mid \mathcal{G}_{t-1}\Big) = -v^TWT^2R$, $H_{(\alpha,\phi)} = H^T_{(\phi,\alpha)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\alpha\partial\varphi} \mid \mathcal{G}_{t-1}\Big) = -v^TDT\mathbf{1}$, $H_{(\varphi,\varphi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\phi^2} \mid \mathcal{G}_{t-1}\Big) = -A^TWT^2A$, $H_{(\varphi,\Phi)} = H^T_{(\Phi,\varphi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\phi\partial\Phi} \mid \mathcal{G}_{t-1}\Big) = -A^TWT^2B$, $H_{(\varphi,\theta)} = H^T_{(\theta,\varphi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\phi\partial\theta} \mid \mathcal{G}_{t-1}\Big) = -A^TWT^2P$, $H_{(\varphi,\Theta)} = H^T_{(\Theta,\varphi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\phi\partial\Theta} \mid \mathcal{G}_{t-1}\Big) = -A^TWT^2R$, $H_{(\varphi,\phi)} = H^T_{(\phi,\varphi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\varphi\partial\phi} \mid \mathcal{G}_{t-1}\Big) = -A^TDT\mathbf{1}$, $H_{(\Phi,\Phi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Phi^2} \mid \mathcal{G}_{t-1}\Big) = -B^TWT^2B$, $H_{(\Phi,\theta)} = H^T_{(\theta,\Phi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Phi\partial\theta} \mid \mathcal{G}_{t-1}\Big) = -B^TWT^2P$, $H_{(\Phi,\Theta)} = H^T_{(\Theta,\Phi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Phi\partial\Theta} \mid \mathcal{G}_{t-1}\Big) = -B^TWT^2R$, $H_{(\Phi,\phi)} = H^T_{(\phi,\Phi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Phi\partial\varphi} \mid \mathcal{G}_{t-1}\Big) = -B^TDT\mathbf{1}$, $H_{(\theta,\theta)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\theta^2} \mid \mathcal{G}_{t-1}\Big) = -P^TWT^2P$, $H_{(\theta,\Theta)} = H^T_{(\Theta,\theta)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\theta\partial\Theta} \mid \mathcal{G}_{t-1}\Big) = -P^TWT^2R$, $H_{(\theta,\phi)} = H^T_{(\phi,\theta)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\theta\partial\phi} \mid \mathcal{G}_{t-1}\Big) = -P^TDT\mathbf{1}$, $H_{(\Theta,\Theta)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Theta^2} \mid \mathcal{G}_{t-1}\Big) = -R^TWT^2R$, $H_{(\Theta,\phi)} = H^T_{(\phi,\Theta)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\Theta\partial\phi} \mid \mathcal{G}_{t-1}\Big) = -R^TDT\mathbf{1}$, $H_{(\phi,\phi)} = \mathbb{E}\Big(\frac{\partial^2 l_t}{\partial\phi^2} \mid \mathcal{G}_{t-1}\Big) = -tr(N)$, tr(.) is the

trace function and **1** represents a vector of ones of order $(n-m) \times 1$.
The conditional maximum likelihood estimates are consistent under some mild regularity conditions and are asymptotically normally distributed (Andersen, 1970). Thus, in large sample sizes,

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\varphi} \\ \hat{\Phi} \\ \hat{\theta} \\ \hat{\Theta} \\ \hat{\phi} \end{pmatrix} \sim N_k \left(\begin{pmatrix} \hat{\alpha} \\ \hat{\varphi} \\ \hat{\Phi} \\ \hat{\theta} \\ \hat{\Theta} \\ \hat{\phi} \end{pmatrix}, \ H^{-1} \right) \tag{26}$$

approximately, where CMLE of α, φ, Φ, θ, Θ and ϕ are $\hat{\alpha}$, $\hat{\varphi}$, $\hat{\Phi}$, $\hat{\theta}$, $\hat{\Theta}$, $\hat{\phi}$. The asymptotic covariance matrix of Ω is H^{-1}.

3.3 Confidence intervals and hypothesis testing inference

Suppose the r^{th} component of Ω is denoted by Ω_r. We have, From (26)

$$\frac{\hat{\Omega}-\Omega_r}{\sqrt{H(\hat{\Omega})^{rr}}} \sim N_k(0,1)$$

where the r^{th} diagonal part of the conditional information matrix inverse is $H(\hat{\Omega})^{rr}$. Suppose z_δ represents the standard normal quantile of δ. A $100(1-\alpha)\%,\ 0<\alpha<1/2$, confidence interval for Ω, r = 1,...,(p+q+P+Q+2), is

$$\left[\hat{\Omega}_r - z_{1-\alpha/2}\sqrt{H(\hat{\Omega})^{rr}}; \hat{\Omega}_r + z_{1-\alpha/2}\sqrt{H(\hat{\Omega})^{rr}}\right].$$

For detailed information about asymptotic confidence intervals, see in (Pawitan, 2001; Davison and Hinkley, 1997). Going to test for null hypothesis $F_0 : \Omega_r = {\Omega_r}^0$ Vs alternative hypothesis $F_1 : \Omega_r \neq {\Omega_r}^0$ are being established on signed square root test of Wald's statistic, provided by (Pawitan, 2001).

$$z = \frac{\hat{\Omega}_r - {\Omega_r}^0}{\hat{se}(\hat{\Omega}_r)},$$

where, $\hat{se}(\hat{\Omega}_r) = \sqrt{H(\hat{\Omega})^{rr}}$. The limiting distribution of Z is standard normal under F_0. Hypothesis testing inference can be performed using the likelihood ratio (Neyman and Pearson, 1928), Rao's score (Rao, 1948), Wald (Wald, 1943) and gradient (Terrell, 2002) tests. These test statistics are well distributed as chi-square (χ^2) under the condition of F_0 and in large samples.
The test for seasonal movement i.e. the Wald test can be displayed as following.

$$F_0 : (\Phi_1, ..., \Phi_P, \Theta_1, ..., \Theta_Q)^T = \mathbf{0}_{P+Q} \quad \text{(non-seasonal)},$$

$F_1 : (\Phi_1, ..., \Phi_P, \Theta_1, ..., \Theta_Q)^T \neq \mathbf{0}_{P+Q}$ (seasonal),
here $\mathbf{0}_{P+Q}$ is the (p+Q)-vector of zeros. There is no seasonal variance below F_0. Rejecting F_0 (null hypothesis) means that seasonality must be taken into account. The Wald test statistic is

$$X = (\hat{\Phi}_1, ..., \hat{\Phi}_P, \hat{\Theta}_1, ..., \hat{\Theta}_Q)\Big(H(\hat{\Omega})^{\Phi\Theta}\Big)^{-1}(\hat{\Phi}_1, ..., \hat{\Phi}_P, \hat{\Theta}_1, ..., \hat{\Theta}_Q)^T,$$

where, $H(\hat{\Omega})^{\Phi\Theta}$ is the block of inverse Fisher information matrix $(P + Q) \times (P + Q)$ in respect to the seasonal parameters calculated at $\hat{\Omega}$. X is Chi-squared asymptotically distributed with P+Q no. of degrees of freedom (df) under the condition of standard regularity and the condition of null hypothesis. If we are going to calculate X, then the seasonal model just needs to be calculated.

4 Model selection, diagnostic analysis and forecasting

Some model selection criteria for identification of model and some diagnostic tools for adapted KSARMA models are discussed in this section. A fitted model can be used for diagnostic checks to assess if the data dynamics are completely captured. For out-of-sample prediction that passes all diagnostic tests, an appropriate model may also be used.

4.1 Model selection criteria

Criteria for the selection of the models are rules used for choosing a statistical model. Among the several competing models, the selection of the best-fitted model can be decided on the basis of Akaike's information criteria (AIC) (Akaike, 1974).
AIC is defined as

$$\text{AIC} = -2\hat{l}_* + 2(p + q + P + Q + 2) \tag{27}$$

where, $\hat{l}_* = \frac{\hat{l} \times n}{(n-m)}$ and the number of model parameters is $(p + q + P + Q + 2)$.
Information criteria such as Akaike's (AIC), Schwartz's (SIC) (Schwarz, 1978), and Hannan and Quinn's (HQ) (Hannan and Quinn, 1979) are also obtained from the maximized conditional log-likelihood function in a regular manner.

4.2 Deviance

Deviance E is specified two times the difference between saturated model $\tilde{l}$ (for which $\hat{v}_t = x_t$) and fitted model $\hat{l}$ in the conditional log-likelihood which is expressed as

$$E = 2(\tilde{l} - \hat{l}), \quad \text{where } \tilde{l} = \sum_{t=1}^{n} l_t(x_t, \phi) \quad \text{and } \hat{l} = \sum_{t=1}^{n} l_t(\hat{v}_t, \phi)$$

E is approximately distributed as chi-square $\left(\chi^2_{n-m-(p+q+P+Q+2)}\right)$, if the fitted model is defined correctly.

4.3 Residuals

Analysis of the residual is an essential tool for ensuring that for the data the predicted model is adequate (Kedem and Fokianos, 2005). Different kinds of residuals are available for various class of models (Mauricio, 2008). Here, we can consider quantile residual (Dunn and Smyth, 1996) for proposed KSARMA$(p,q)\times(P,Q)_{12}$ model which is delivered by

$$w_t^{(q)} = \Theta^{-1}(F_{\upsilon_t}(\tilde{x_t} \mid \mathcal{G}_{t-1}))$$

where Θ represents the normal quantile function. Under the adequate model specification, these kinds of residuals is distributed around standard normal.
The residuals should act as white noise whenever the model is defined correctly, that is, they are predicted to be non-stationary in series and implement a protocol of mean zero and its variance is constant (Kedem and Fokianos, 2005). A Ljung-Box test (Ljung and Box, 1978) which is based on the residuals is a suitable alternative for testing the sufficiency of the fitted model which can be seen in (Greene, 2003) for detailed information.

4.4 Forecasting

For conditional median υ_t, we can obtain estimator of υ_t, $\hat{\upsilon}_t$ for $t \in \{m, ..., n\}$ (in sample) by applying conditional maximum likelihood estimator in equation (4) and $\hat{\epsilon}_t = g(x_t) - g(\hat{\upsilon}_t)$. we can obtained forecast h-step ahead as

$$\hat{\upsilon}_{n+h} = g^{-1}\Bigg(\hat{\alpha}+\sum_{i=1}^{p}\hat{\varphi}_i\Big[g(x_{n+h-i})\Big]+\sum_{j=1}^{q}\hat{\theta}_j\Big[\epsilon_{n+h-j}\Big]+\sum_{I=1}^{P}\hat{\Phi}_I\Big[g(x_{n+h-IS})\Big]+\sum_{J=1}^{Q}\hat{\Theta}_J\Big[\epsilon_{n+h-JS}\Big]$$
$$-\sum_{i=1}^{p}\sum_{I=1}^{P}\hat{\varphi}_i\hat{\Phi}_I\Big[g(x_{n+h-(i+IS)})\Big] + \sum_{j=1}^{q}\sum_{J=1}^{Q}\hat{\theta}_j\hat{\Theta}_J\Big[\epsilon_{n+h-(j+JS)}\Big]\Bigg)$$

where

$$[g(x_t)] = \begin{cases} g(\hat{\upsilon}_t) & \text{if } t > n \\ g(x_t) & \text{if } t \leq n. \end{cases}$$

and

$$[\epsilon_t] = \begin{cases} 0 & \text{if } t > n \\ g(x_t) - g(\hat{\upsilon}_t) & \text{if } t \leq n. \end{cases}$$

5 Numerical evaluation

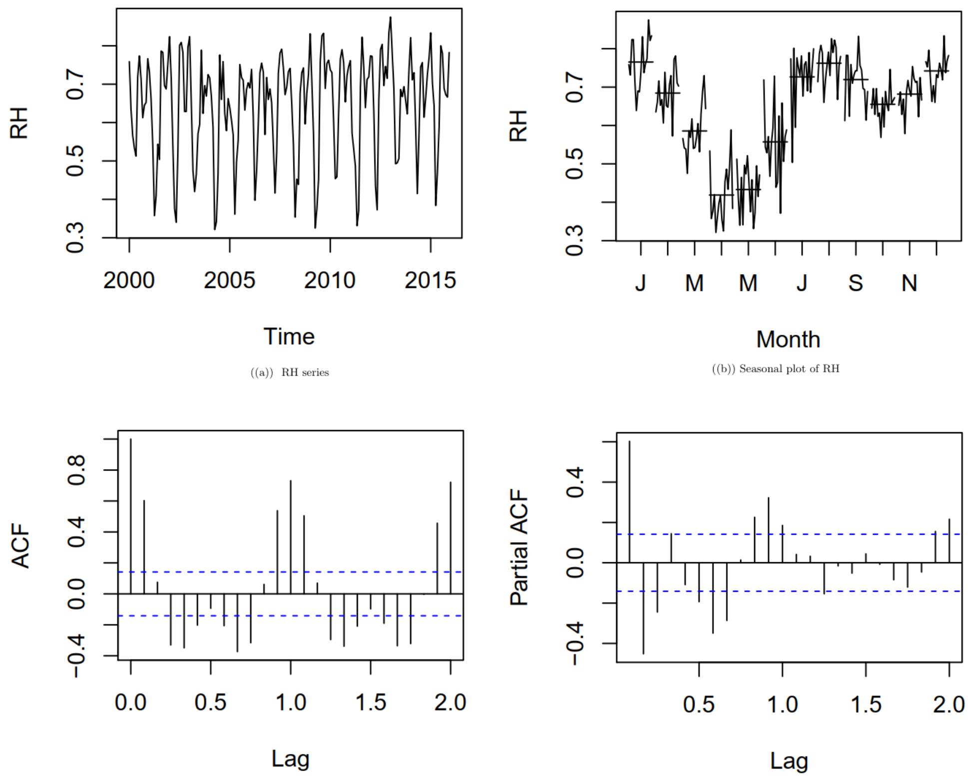
We present a Monte Carlo simulation analysis in this section to evaluate the CMLE's finite sample efficiency for KSARMA models. We simulate 5000, Monte Carlo replications of KSARMA $(1,1)\times(1,1)_{12}$ model with parameters $\alpha = 0.7000$, $\varphi_1 = 0.5000$, Φ_1 = -0.7800, $\theta_1 = 0.6000$, $\Theta_1 = 0.8000$ and sample sizes n = 60, 120, 240, 480. From the KSARMA model, produce a vector representation of n iterations of the variable x_t with the link function logit. A numerical optimization method, i.e., BFGS with

the first derivative was used for all conditional log-likelihood maximization. We obtained the initial value of the coefficients of the autoregressive by regressing $g(x_t)$ on $g(x_{t-1}), ..., g(x_{t-p}), g(x_{t-(p+1)}), ..., g(x_{t-(p+P)})$ and at the beginning of the conditional log-likelihood maximizations, we equate zero of all the parameters of moving average. The simulation has been done by R- software (Team, 2017). The results of the simulation are represented in the form of Table 1.
In Table 1, statistical performance is represented as mean, bias, relative bias (RB), standard deviation (SD) and root mean square error (RMSE). Overall performance of CMLE was observed to satisfactory except as our expectation for every smaller size that is n = 60. For the sample size n = 60, the estimation was improved significantly with the increase in sample size. The smallest relative bias, i.e., φ is the overall estimator of the parameters. The estimation of the performance was found to be superior in the part of autoregressive estimator than the moving average estimator. Simulation results analysis represents that inferences about moving average parameters with respect to the other parameters are found to be poorer. Table 1 mention, Parametric estimation of RMSE is found to be smaller in all the situations.

6 Empirical application

For the proposed KSARMA model, we shall use the monthly average relative humidity data of Delhi, India, which is taken from January, 2000 to December, 2016. This data were acquired from India Meteorological Department (IMD), Pune. We have reserved last 12 observation from the data for the performance of forecasting. Relative humidity (RH) is an important factor for the human health, hydrological studies, pharmaceuticals industry, agriculture and irrigation scheduling. Humidity levels that are too high or too low may have severe health consequences. Low RH has been linked to lots of health issues, including cold flu, nasal bleeding, vomiting, asthma attacks, allergies etc. In low RH conditions, the human body is more susceptible to covid-19 infection (Mangla et al., 2021). In addition to causing respiratory problems, high RH causes an increase in precipitation, which can be dangerous if it occurs in excess. The time series plot of RH (Figure 1(a)) shows in Figure 1, seasonal plot of RH (Figure 1(b)), autocorrelation function (ACF) (Figure 1(c)) and partial autocorrelation function (PACF) plot (Figure 1(d)).
We have selected the KSARMA $(2, 1) \times (2, 2)_{12}$ model for relative humidity data. The fitted KSARMA $(2, 1) \times (2, 2)_{12}$ models as well as some diagnostic are shown in Table 2. Figure 2 represents some diagnostic plots which contains five plots, that is, (a) Observed data versus fitted values, (b) index plot of Quantile residual, (c) residual ACF, (d) residual PACF and (e) residual QQ (quantile-quantile) plot. From figure 2(b), we can see that there is no distinct pattern over time and residual behaves like white noise. Figure 2(c) and figure 2(d) display a plot of ACF and PACF that can help to visual verification of the residual white noise hypothesis, the Ljung-Box test was also used for its verification given in Table 2 . From Figure 2(e), we can see that, the presence of a nearly straight line in the plot indicates that the residuals are approximately normally

distributed. Hence, we can confidently use the fitted model for forecasting according to all plots and tests.. Figure 3(a) shows fitted KSARMA $(2,1) \times (2,2)_{12}$ model.
Using the KSARMA $(2,1) \times (2,2)_{12}$ and SARIMA $(2,0,1) \times (2,0,2)_{12}$ models, we predicted the monthly relative humidity data for next 12 observations, which are shown in Figure 3(b). The forecasting performance comparison from different models are shown in Table 3. Thus, we observed that the proposed model perform better than SARIMA in both measure.



((a)) RH series

((b)) Seasonal plot of RH

((c)) plot of ACF

((d)) plot of PACF

Figure 1: Time series plot of monthly relative humidity data of Delhi, India. (a) RH series, (b) Seasonal plot of RH, (c) plot of ACF and (d) plot of PACF

Table 2: **Fitted KSARMA** $(2,1) \times (2,2)_{12}$ **model**

Parameters	Estimate	Standard error	z stat.	p-value
α	0.0440	0.0073	6.0387	<0.0001
φ_1	0.2636	0.0553	4.7669	<0.0001
φ_2	-0.0675	0.0339	1.9919	0.0464
Φ_1	0.3217	0.0774	4.1582	<0.0001
Φ_2	0.6145	0.0383	16.0259	<0.0001
θ_1	-0.0268	0.0594	0.4519	0.6513
Θ_1	-0.2986	0.0857	3.4835	0.0005
Θ_2	-0.3227	0.0885	3.6478	0.0003
ϕ	9.9881	0.6107	16.3542	< 0.0001

Log-likelihood = 244.2485

Deviance = 155.5892

AIC = -470.4971

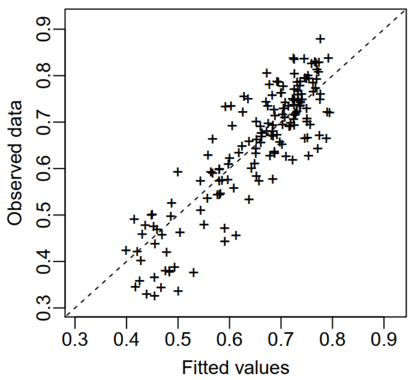
BIC = -441.1796

Seasonality test: W = 296.349 (p-value < .0001)

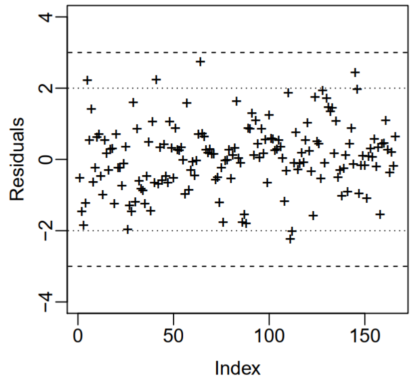
Ljung-Box test: Q = 19.283 (p-value = 0.5035)

Table 1: **Results of simulation on point estimation based on KSARMA**$(1,1) \times (1,1)_{12}$

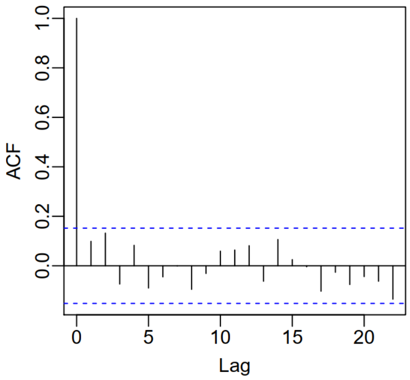
	α	φ	Φ	θ	Θ	ϕ
Parameters	0.7000	0.5000	-0.7800	0.6000	0.8000	12.0000
n = 60						
Mean	0.3543	0.4991	0.0061	0.4312	-0.0267	10.2730
Bias	-0.3457	-0.009	0.7861	-0.1688	-0.8267	-1.7270
RB(%)	-49.3798	-0.1745	-100.7857	-28.1339	-103.3435	-14.3916
SD	0.1512	0.1817	0.1961	0.2260	0.2180	1.0062
RMSE	0.3773	0.1816	0.8102	0.2821	0.8550	1.9987
n=120						
Mean	0.3783	0.4732	0.0164	0.4027	-0.0474	10.0988
Bias	-0.3217	-0.0267	0.7964	-0.1973	-0.8474	-1.9012
RB(%)	-45.9594	-5.5081	-102.1023	-32.8851	-105.9306	-15.8434
SD	0.1329	0.1573	0.1671	0.2001	0.1622	0.7683
RMSE	0.3481	0.1597	0.8137	0.2810	0.8628	2.0506
n=240						
Mean	0.4104	0.4368	0.0104	0.3734	-0.0448	10.0130
Bias	-0.2896	-0.0632	0.7904	-0.2266	-0.8448	-1.9870
RB(%)	-41.3695	-12.6483	-101.3384	-37.7595	-105.6052	-16.5584
SD	0.1199	0.1347	0.1454	0.1783	0.1457	0.7218
RMSE	0.3134	0.1488	0.8037	0.2883	0.8573	2.1140
n= 480						
Mean	0.4264	0.4177	0.0200	0.3422	-0.0576	9.9256
Bias	-0.2736	-0.0823	0.8000	-0.2578	-0.8576	-2.0744
RB(%)	-39.0888	-16.4622	-102.5624	-42.9714	-107.1980	-17.2868
SD	0.1142	0.1157	0.1471	0.1598	0.1434	0.6677
RMSE	0.2956	0.1420	0.8134	0.3033	0.8695	2.1792



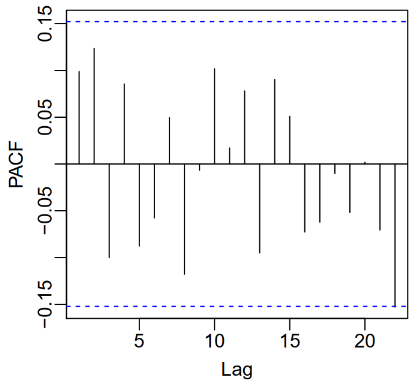
((a)) Observed versus fitted



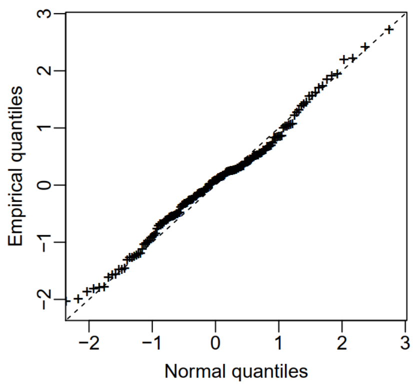
((b)) Quantile residual



((c)) Residual ACF

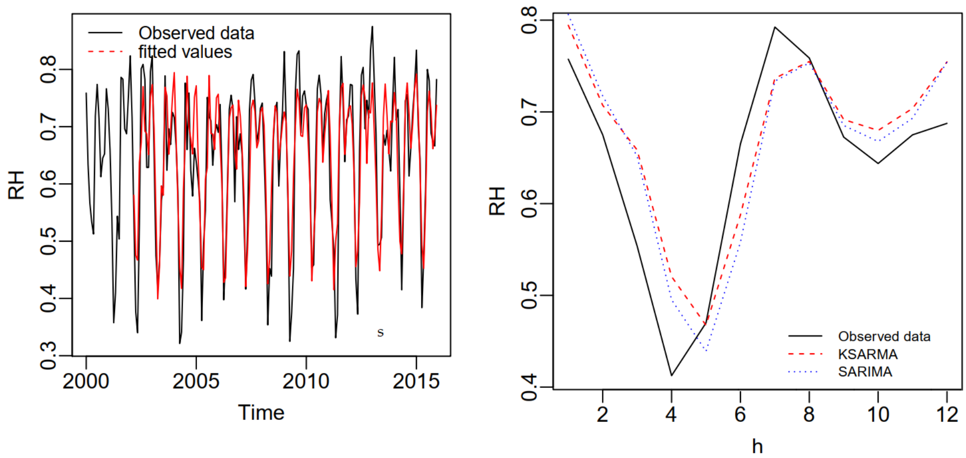


((d)) Residual PACF



((e)) QQ-plot

Figure 2: Plots of diagnostic based on the quantile residuals. (a) Observed versus fitted, (b) Quantile residual, (c) Residual ACF, (d) Residual PACF and (e) QQ-plot



((a)) fitted KSARMA (2,1) x (2,2) 12

((b)) Observed and forecasted values from January to December 2016

Figure 3: Observed and forecast values of relative humidity in Delhi, India. (a) fitted KSARMA $(2,1)\times(2,2)_{12}$ and (b) Observed and forecasted values from January to December 2016

Table 3: **Forecasting performance comparison from different Models**

Models	RMSE	MAPE
KSARMA	0.05854	0.07390
SARIMA	0.05921	0.07967

7 Conclusions

This study introduced a new class of dynamic seasonal time series KSARMA model to predict double-bounded relative humidity time-series data. This model is a generalized version of KARMA model for seasonal data and it can be used to model and forecasting seasonal time series data with value laying in the double bounded interval. In the proposed model, we used the conditional maximum likelihood approach for the parameters estimation of the model. Besides, closed type expression were obtained for the conditional score vector and fisher information matrix. We have also discussed the confidence interval, hypothesis testing and selection of model. A quantile residual approach to evaluating goodness-of-fit was discussed in this paper, as well as a white noise test which can be implemented to the residual obtained from its fitted model. For the evaluation of the finite sample results of the CMLE and test of white noise, we have used a Monte Carlo simulation.

Acknowledgement

The financial grant in the form of fellowship to the first author by the CSIR, India is thankfully acknowledged.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2):283–301.

Bayer, F. M., Bayer, D. M., and Pumi, G. (2017). Kumaraswamy autoregressive moving average models for double bounded environmental data. *Journal of Hydrology*, 555:385–396.

Benjamin, M. A., Rigby, R. A., and Stasinopoulos, D. M. (2003). Generalized autoregressive moving average models. *Journal of the American Statistical association*, 98(461):214–223.

Cribari-Neto, F. and Zeileis, A. (2010). Beta regression in r. *Journal of statistical software*, 34:1–24.

da Silva, C. Q., Migon, H. S., and Correia, L. T. (2011). Dynamic bayesian beta models. *Computational Statistics & Data Analysis*, 55(6):2074–2089.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*. Number 1. Cambridge university press.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5(3):236–244.

Ferrari, S. and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7):799–815.

Gradshteyn, I. S. and Ryzhik, I. (2007). Table of integrals, series, and products. seventh. *Elsevier/Academic Press, Amsterdam*, 48:1171.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Guolo, A. and Varin, C. (2014). Beta regression for time series analysis of bounded data, with application to canada google® flu trends. *The Annals of Applied Statistics*, 8(1):74–88.

Gupta, A. K. and Nadarajah, S. (2004). *Handbook of beta distribution and its applications*. CRC press.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):190–195.

Jones, M. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical methodology*, 6(1):70–81.

Kedem, B. and Fokianos, K. (2005). *Regression models for time series analysis.* John Wiley & Sons.

Kumaraswamy, P. (1976). Sinepower probability density function. *Journal of Hydrology*, 31(1-2):181–184.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of hydrology*, 46(1-2):79–88.

Lemonte, A. J., Barreto-Souza, W., and Cordeiro, G. M. (2013). The exponentiated kumaraswamy distribution and its log-transform. *Brazilian Journal of Probability and Statistics*, 27(1):31–53.

Ljung, G. M. and Box, G. E. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

Mangla, S., Pathak, A. K., Arshad, M., Ghosh, D., Sahoo, P. K., Garg, V. K., and Haque, U. (2021). Impact of environmental indicators on the covid-19 pandemic in delhi, india. *Pathogens*, 10(8):1003.

Mauricio, J. A. (2008). Computing and using residuals in time series models. *Computational statistics & data analysis*, 52(3):1746–1763.

McCullagh, P. and Nelder, J. (1989). Generalized linear models [ivq.

Mitnik, P. A. (2013). New properties of the kumaraswamy distribution. *Communications in Statistics-Theory and Methods*, 42(5):741–755.

Mitnik, P. A. and Baek, S. (2013). The kumaraswamy distribution: median-dispersion reparameterizations for regression modeling and simulation-based estimation. *Statistical Papers*, 54(1):177–192.

Nadarajah, S. (2008). On the distribution of kumaraswamy. *Journal of Hydrology*, 348(3):568–569.

Neyman, J. and Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference, part i. *Biometrika A*, 20:1–2.

Palm, B. G. and Bayer, F. M. (2018). Bootstrap-based inferential improvements in beta autoregressive moving average model. *Communications in Statistics-Simulation and Computation*, 47(4):977–996.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood.* Oxford University Press.

Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge University Press.

Rocha, A. V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models. *Test*, 18(3):529–545.

Schwarz, G. (1978). Estimating the dimension of a model ann stat 6: 461–464. *Find this article online.*

Team, R. C. (2017). R: a language and environment for statistical computing. r found. *Stat. Comput. Vienna, Austria.*

Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34(34):206–215.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical society*, 54(3):426–482.