Statistical Inference in Autoregressive Models with Non-negative Residuals

S. Zamani Mehreyan† and A. Sayyareh†‡*,

† Razi University
‡ K. N. Toosi University of Technology

Abstract. Normal residual is one of the usual assumptions of autoregressive models but in practice sometimes we are faced with non-negative residuals case. In this paper we consider some autoregressive models with non-negative residuals as competing models and we have derived the maximum likelihood estimators of parameters based on the modified approach and EM algorithm for the competing models. Also, based on the simulation study, we have compared the ability of some model selection criteria to select the optimal autoregressive model. Then we consider a set of real data, level of lake Huron 1875-1930, as a data set generated from a first order autoregressive model with non-negative residuals and based on the model selection criteria we select the optimal model between the competing models.

Keywords. Autoregressive model; Kullback-Leibler information; model selection criterion; modified maximum likelihood.


1 Introduction

Let \( n \times 1 \) random vector \( X_t = (Y_t, Z_t), \ t = 1, \ldots, n \) are i.i.d with common unknown true distribution \( H(\cdot) \) on a complete probability space \( (X, \sigma_X) \), where \( X \) is the Euclidean space \( \mathbb{R}^m \) and \( \sigma_X \) is the Borel \( \sigma \)-field on \( X \). Let \( (Y, \sigma_Y) \) and \( (Z, \sigma_Z) \) be the measurable spaces associated with \( Y_t \) and \( Z_t \). We shall be
interested in the true conditional distribution $H_{Y|Z}(\cdot|V)$ of $Y_t$ given $Z_t$. Let $H_Z$ be the true marginal distribution of $Z_t$, and $\nu_Y$ be a $\sigma$-finite measure on $(Y, \sigma_Y)$. For $H_Z$-almost all $z$, $H_{Y|Z}(\cdot|z)$ has a Radon-Nikodym density $h(\cdot|z)$ relative to $\nu_Y$, which is strictly positive for $\nu_Y$-almost all $y$.

We now consider two competing parametric families of conditional distribution defined on $\sigma_Y \times \sigma_Z$ for $Y_t$ given $Z_t$:

$$\mathcal{G} = \left\{ g^\beta(y|z), \ \beta \in \mathcal{B} \subseteq \mathbb{R}^p \right\} \quad \text{and} \quad \mathcal{F} = \left\{ f^\gamma(y|z), \ \gamma \in \Gamma \subseteq \mathbb{R}^q \right\}. $$

A known measure of divergence is Kullback-Leibler (1951), $KL$, measure which is defined in term of conditional densities as:

$$KL\{h_{Y|Z}, g^\beta_{Y|Z}\} = E_h \left\{ \log \frac{h(Y|Z)}{g^\beta(Y|Z)} \right\},$$

where $E_h$ denotes the expectation with respect to the true joint distribution of $(Y,Z)$. The so-called reduced model approach, Commenges et al. (2008), is more satisfactory to define this measure. Consider a sample of i.i.d couples of variables $(Y_i, Z_i), \ i = 1, \ldots, n$ having joint pdf $h(y, z) = h(y|z)h(z)$. Consider the model $\mathcal{G}$ such that $g^\beta(y, z) = g^\beta(y|z)h(z)$; the model is called “reduced” because $h_Z(\cdot)$ is assumed known. So the Kullback-Leibler divergence is:

$$KL\{h, g^\beta\} = E_h \left\{ \log h(Y|Z) \right\} - E_h \left\{ \log g^\beta_{Y|X}(Y|Z) \right\},$$

that is the term in $h_Z(\cdot)$ disappears (so that we do not need to know it in fact) and we get the same definition as in Vuong (1989) using only the conventional Kullback-Leibler divergence. In the literature of the model selection theory we have the following definition:

**Definition 1.** The model $\mathcal{G}$ is well-specified if there is $\beta_0 \in \mathcal{B}$ and $g \in \mathcal{G}$ such that, $h(\cdot|\cdot) = g^{\beta_0}(\cdot|\cdot)$; otherwise it is misspecified.

**Definition 2.** (i) $\mathcal{F}$ and $\mathcal{G}$ are nonoverlapping if $\mathcal{F} \cap \mathcal{G} = \emptyset$; (ii) $\mathcal{F}$ is nested in $\mathcal{G}$ if $\mathcal{F} \subset \mathcal{G}$; (iii) Two models $\mathcal{G}$ and $\mathcal{F}$ are non-nested if and only if $\mathcal{G} \cap \mathcal{F} = \emptyset$.

If $\mathcal{G}$, is conditional model, its distance from the true conditional density $h(y|z)$, as measured by the minimum Kullback-Leibler risk criterion, equal $KL\{h(\cdot|\cdot), g^{\beta_*}(\cdot|\cdot)\}$, where $\beta_*$ is the pseudo-true value of $\beta$, see e.g White
Thus, an equivalent selection criterion can be based on the quantity $E_h \{ \log g^{\beta^*}(Y|Z) \}$, the best model being the one for which this quantity is the largest. $KL$ is a non-negative quantity. By definition, the more $g^\beta(\cdot|\cdot)$ agrees with $h(\cdot|\cdot)$ the smaller $KL\{h(\cdot|\cdot), g^\beta(\cdot|\cdot)\}$ is. Then the closest member in $G$ to the $h(\cdot|\cdot)$ is $g^{\beta^*}(\cdot|\cdot)$ where $\beta^* \in B$ is the minimizer of $KL\{h(\cdot|\cdot), g^\beta(\cdot|\cdot)\}$. For Kullback-Leibler divergence, $g^{\beta^*}(\cdot|\cdot)$ is the best approximation to $h(\cdot|\cdot)$ under model $G$. It is important to notice that when the model is well-specified we have $\beta_0 = \beta^*$. The Quasi Maximum Likelihood Estimator (QMLE), $\hat{\beta}_n$, is a consistent estimator of $\beta^*$, see White (1982). If the model is misspecified, $KL(h,g^\beta) > 0$. Hence $KL$ divergence takes its value in $[0, \infty]$. The $KL$ divergence is not a metric, but it is additive over marginal of product measures. $KL(h,g^\beta) = 0$ implies that $h = g^\beta$.

The Akaike Information Criterion, AIC, (Akaike, 1973) initially was proposed as an estimate of minus twice the expected log-likelihood. We notice that the important part of the $KL$ divergence is $E_h \{ \log g^{\beta^*}(Y|Z) \}$ which has an consistent estimator as

$$\frac{1}{n} \sum_{i=1}^{n} \log g^{\hat{\beta}_n}(Y_i|Z).$$

It can be considered as an estimator of the divergence between the true density and the competing model. Now the stress is on $\hat{\beta}_n$ because $\frac{1}{n} \sum_{i=1}^{n} \log g^{\hat{\beta}_n}(Y_i|Z)$ provides an overestimate and then the maximized likelihood function has a positive bias as an estimator of the expected log-likelihood. Since $\hat{\beta}_n$ corresponds to the empirical distribution, say, $F_n$ which introduces the estimator. In fact both of them depend on the same sample. The AIC is defined as

$$AIC = -2 \log \text{ likelihood} + 2 \text{ Number of estimated parameters}.$$

It indicates that the bias of the log-likelihood approximately becomes the number of free parameters contained in the model. The bias is derived under the assumption that the true distribution is contained in the specified parametric model. Hurvich and Tsai (1993) proposed a corrected Akaike Information Criterion, $AIC_c$, for small sample, which can be expressed as

$$AIC_c = -2 \log \text{ likelihood} + 2k \frac{n}{n-k-1}$$

where $k$ is number of estimated parameters. The Bayesian Information Criterion (BIC) or Schwarz’s Information Criterion (SIC) proposed by Schwarz.
(1978) is an evaluation criterion for models defined in terms of their posterior probability. The SIC is actually defined as

\[
SIC = -2 \log \text{likelihood} + k \log(n)
\]

where \( k \) is number of estimated parameters. De Gooijer et al. (1985) have considered automatic model selection criteria such as AIC and Bayesian information Criterion, SIC. Claeskens et al. (2007) proposed an adapted version of the Focused Information Criterion that defined by Claeskens and Hjort (2003).

In modelling the time series it is usually assumed that residual terms follow normal distribution and noticed to order selection. We consider a model of time series models such as autoregressive model

\[
y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \epsilon_t
\]

where \( \phi = (\phi_1, \ldots, \phi_p) \) is autoregressive coefficients and \( \epsilon_t \)'s are i.i.d random variables with normal distribution, \( N(0, \sigma^2) \). The conditional log-likelihood function is

\[
l(\phi, \sigma^2) = -\frac{n}{2} \log(2\pi \sigma^2) - \frac{S(\phi)}{2\sigma^2},
\]

where \( S(\phi) = \sum_{t=p+1}^{n} \epsilon_t^2 \). Determination of the model order is an important step in autoregressive, AR, modelling. So we select optimal order based on information criteria or hypotheses tests. In this case all competing models are nested. During recent years, a number of non-normal models with AR-type correlation structure have been proposed. In fact model selection for residuals of autoregressive model is important as determination of the model order. Here we consider the non-nested competing models.

Under autoregressive models with non-normal residuals, the maximum likelihood estimator, MLE, is not appropriate since explicit solutions from the likelihood equations cannot be obtained. We can use some other method such as modified maximum likelihood, MML, method and EM algorithm.

The modified maximum likelihood method has been developed by Tiku (1967) and applied to some non-normal time series models. This method is based on linearization of intractable terms of the log-likelihood function using first-order Taylor series expansion. Bayrak and Akkaya (2010) studied the multiple autoregressive model and estimated the parameters of this model by the modified maximum likelihood.

In this paper the residual model selection is interested. We consider the true
model as \( x_i = \phi x_{i-1} + \varepsilon_i \), where \( \phi \) is the autoregressive coefficient and \( \varepsilon_i \)'s are i.i.d non-negative random variable. Also we consider autoregressive model with Gamma, Weibull, Log-normal and skew normal as the four competing models. The results in this paper are organized as follows. We estimate the parameters of competing models in the Section 2. Using simulation study, in Section 3, we provided that the information criteria such as AIC, AIC\(_c\) and SIC are suitable criteria for autoregressive model selection with non-negative residuals. In Section 4, to confirm the theoretical results the real data is studied.

2 Model and Parameter Estimation

Bayrka and Akkaya (2010) have studied multiple autoregressive model with non-normal residuals. They consider three different types of non-normal distribution (i) long-tailed symmetric, (ii) skew distributions represented by the Generalized Logistic and (iii) short-tailed symmetric and only derived the modified maximum likelihood estimators for these models. In this paper, we obtain the MMLE of parameters of three competing models and EM estimator of parameters of autoregressive model with skew normal residuals. We select optimal model based on information criteria. Here we compute the MML and EM estimator of parameters.

2.1 Autoregressive Model with Gamma Residual

Consider the first order Gamma autoregressive model

\[ x_i = \phi_1 x_{i-1} + \varepsilon_i \]

where \( \phi_1 \) is the autoregressive coefficient and the \( \varepsilon_i \)'s are i.i.d error terms with Gamma distribution, \( G(\alpha, \beta) \). We get the log-likelihood of \( \varepsilon_1, \ldots, \varepsilon_n \), as

\[ l(\phi_1, \alpha, \beta) = -n \log(\beta) - n \log(\Gamma(\alpha)) + (\alpha - 1) \sum_{i=2}^n \log(z_i) - \sum_{i=2}^n z_i \]

where \( z_i = (x_i - \phi_1 x_{i-1})/\beta \). The differentiating log-likelihood function with respect to \( \phi_1 \) is functions in terms of \( z_i^{-1} \) and it has no explicit solutions. To obtain the explicit solution, we order \( \varepsilon_i \) (for a given \( \phi \)) in order of increasing magnitude. So we obtain modified maximum likelihood estimators by solving
the estimating equations

\[
\frac{\partial}{\partial \alpha} l(\phi_1, \alpha, \beta) = -n \frac{\partial}{\partial \alpha} \log(\Gamma(\alpha)) + \sum_{i=2}^{n} \log(z_{(i)}) = 0
\]

\[
\frac{\partial}{\partial \beta} l(\phi_1, \alpha, \beta) = -\beta n \frac{1}{\beta} - (\alpha - 1) \beta + \sum_{i=2}^{n} x_{[i]} - \phi_1 x_{[i]-1} = 0
\]

\[
\frac{\partial}{\partial \phi_1} l(\phi_1, \alpha, \beta) = -\frac{1}{\beta} \sum_{i=2}^{n} x_{[i]-1} - \beta \sum_{i=2}^{n} \frac{x_{[i]-1}}{\beta} = 0
\]

where the \(z_{(i)}\) are ordered \(z_i\)-values and \((x_{[i]}, x_{[i]-1})\) is that pair of \((x_i, x_{i-1})\) observations which corresponds to the ordered \(\varepsilon_{(i)}\). Define \(t_{(i)} = E\{z_{(i)}\}\) which will be obtain from

\[
\frac{1}{\Gamma(k)} \int_{0}^{t_{(i)}} \exp(-z)z^{k-1}dz = \frac{i}{n+1}
\]

to more illustrations, see Akkaya and Tiku (2007). We use two terms of the Taylor series expansion,

\[
z_{(i)}^{-1} = \frac{2}{t_{(i)}} - \frac{z_{(i)}}{t_{(i)}^2} + o(|z_{(i)} - t_{(i)})|
\]

so

\[
z_{(i)}^{-1} \simeq \alpha_i - \beta_i z_{(i)}
\]

(1)

where \(\alpha_i = \frac{2}{t_{(i)}}\) and \(\beta_i = \frac{1}{t_{(i)}}\). Incorporated Eq.(1) in estimating equations.

We can obtain MMLE of \(\alpha, \beta\) and \(\phi_1\) as

\[
\hat{\Gamma}_D = \frac{1}{n} \sum_{i=2}^{n} \log(z_{(i)})
\]

where \(\Gamma_D\) is \(\frac{\partial}{\partial \alpha} \log(\Gamma(\alpha))\),

\[
\hat{\beta} = \frac{1}{n\alpha} \sum_{i=2}^{n} (x_{[i]} - \phi_1 x_{[i]-1})
\]

and

\[
\hat{\phi}_1 = \frac{\beta}{\alpha-1} \sum_{i=2}^{n} x_{[i]-1} + \sum_{i=2}^{n} \beta_i x_{[i]-1} - \beta \sum_{i=2}^{n} \alpha_i x_{[i]-1}. \frac{\sum_{i=2}^{n} \beta_i x_{[i]-1}}{}.
\]

© 2015, SRTC Iran
2.2 Autoregressive Model with Weibull Residual

Consider the first order Weibull autoregressive model $x_i = \phi_2 x_{i-1} + \varepsilon_i$ where the i.i.d residuals $\varepsilon_i$’s have the Weibull distribution, $W(\gamma, \tau)$

$$f(\varepsilon_i) = \frac{\gamma - 1}{\tau^\gamma} \varepsilon_i^{(\gamma - 1)} \exp\left(-\frac{\varepsilon_i}{\tau}\right).$$

The log-likelihood of $\varepsilon_1, \ldots, \varepsilon_n$ is

$$l(\phi_2, \gamma, \tau) = n \log(\gamma) - n \log(\tau) + (\gamma - 1) \sum_{i=2}^{n} \log(z_i) - \sum_{i=2}^{n} z_i^{\gamma}$$

where $z_i = (x_i - \phi_2 x_{i-1})/\tau$. The following estimating equations,

$$\frac{\partial}{\partial \gamma} l(\phi_2, \gamma, \tau) = \frac{n}{\gamma} + \sum_{i=2}^{n} \log(z_i) - \sum_{i=2}^{n} z_i^{\gamma} \log(z_i) = 0$$

$$\frac{\partial}{\partial \phi_2} l(\phi_2, \gamma, \tau) = -\frac{(\gamma - 1)}{\tau} \sum_{i=2}^{n} x_i z_i - \frac{\gamma}{\tau} \sum_{i=2}^{n} x_i z_i^{\gamma - 1} = 0$$

have no explicit solution. The modified likelihood equations are obtained by linearizing the intractable terms, $z_i^{\gamma - 1}$ and $z_i^{\gamma - 1}$ in likelihood equations using the first two terms of the Taylor series expansion,

$$z_i^{\gamma - 1} \simeq v_{i0} - B_{i0} z_i,$$

$$z_i^{\gamma - 1} \simeq v_{i0}^* - B_{i0}^* z_i$$

where $v_{i0} = \frac{2}{t_i}, B_{i0} = \frac{1}{t_i}, t_i = (-\log(1 - \frac{i}{n+1}))^{\frac{1}{\gamma}}, v_{i0}^* = (2 - \gamma) t_i^{\gamma - 1}$ and $B_{i0}^* = (\gamma - 1)t_i^{\gamma - 2}$. Define

$$v_i = (\gamma - 1)v_{i0} - \gamma v_{i0}^*,$$

$$\beta_i = (\gamma - 1)\beta_{i0} + \gamma \beta_{i0}^*$$

so we can obtain MMLE of $\gamma$, $\tau$ and $\phi_2$ as

$$\frac{\partial}{\partial \gamma} l(\phi_2, \gamma, \tau) = \frac{n}{\gamma} + \sum_{i=2}^{n} \log(z_i) - \sum_{i=2}^{n} z_i \log(z_i)(v_{i0}^* + \beta_{i0}^* z_i) = 0$$

\[ \hat{\tau} = \frac{(-B \pm \sqrt{\Delta})}{2A} \]

where \( \Delta = B^2 - 4AC \), \( A = -n \), \( B = -\sum_{i=2}^{n}(x[i] - \hat{\phi}_2x[i-1])v_i \) and \( C = \sum_{i=2}^{n}(x[i] - \hat{\phi}_2x[i-1])^2 \beta_i \), \( \hat{\phi}_2 = \frac{\sum_{i=2}^{n}\beta_i x[i-1] - \hat{\tau} \sum_{i=2}^{n}v_ix[i-1]}{\sum_{i=2}^{n}\beta_i x[i-1]^2} \).

### 2.3 Autoregressive Model with Log-normal Residual

Here we consider the first order autoregressive model as \( x_i = \phi_3x_{i-1} + \epsilon_i \) with Log-normal, \( \text{LN}(\mu, \sigma) \) residuals. We get the log-likelihood of \( \epsilon_1, \ldots, \epsilon_n \), as

\[
l(\phi_3, \mu, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \sum_{i=1}^{n} \log(x_i - \phi_3x_{i-1}) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (\log(x_i - \phi_3x_{i-1}) - \mu)^2.
\]

We derive modified maximum likelihood estimators by solving the estimating equations

\[
\frac{\partial}{\partial \mu} l(\phi_3, \mu, \sigma) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (\log(z(i)) - \mu) = 0
\]

\[
\frac{\partial}{\partial \sigma^2} l(\phi_3, \mu, \sigma) = -\frac{n}{2\sigma^4} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (\log(z(i)) - \mu)^2 = 0
\]

\[
\frac{\partial}{\partial \phi_3} l(\phi_3, \mu, \sigma) = \sum_{i=1}^{n} x[i-1]z^{-1}(i) + \sum_{i=1}^{n} x[i-1]z^{-1}(i)(\log(z(i)) - \mu) = 0
\]

where \( z(i) = x[i] - \phi_3x[i-1] \). Similarly we linearize the intractable terms, \( z^{-1}(i) \) and \( \log(z(i)) \) in likelihood equations using the two terms of the Taylor series expansion, we have

\[
z^{-1}(i) \simeq \alpha_i - B_iz(i)
\]

and

\[
\log(z(i)) = c_i + \frac{\alpha_i}{2}z(i)
\]

where \( \alpha_i = \frac{2}{t(i)}, B_i = \frac{1}{t(i)}, c_i = \log(t(i)) - 1 \) and

\[
\int_{0}^{t(i)} \frac{1}{z\sigma\sqrt{2\pi}} \exp\left( -\frac{1}{2\sigma^2}(\log(z) - \mu)^2 \right) dz = \frac{i}{n+1}.
\]
so we can obtain MMLE of $\mu$, $\sigma^2$ and $\phi_3$ as

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \log(z_{(i)})$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (\log(z_{(i)}) - \hat{\mu})^2$$

$$\hat{\phi}_3 = \frac{(-B + \sqrt{\Delta})}{2A}$$

where

$$\Delta = B^2 - 4AC$$

$$A = \sum_{i=1}^{n} \frac{\alpha_i B_i}{2\sigma^2} x_{[i]}^{3/2}$$

$$B = -\sum_{i=2}^{n} x_{[i]}^{3/2} \frac{\alpha_i^2 - 2B_i (c_i - \hat{\mu} - 2\hat{\sigma}^2)}{2\sigma^2} - \frac{\alpha_i B_i}{\sigma^2} x_{[i]}^{3/2}$$

and

$$C = \sum_{i=2}^{n} x_{[i]}^{3/2} \left( \frac{\alpha_i (c_i - \mu + 2\hat{\sigma}^2) - B_i \alpha_i x_{[i]}^2}{2\sigma^2} \right)$$

$$+ \sum_{i=2}^{n} x_{[i]}^{3/2} \left( \frac{\alpha_i^2 - 2B_i (c_i - \mu - 2\hat{\sigma}^2)}{2\sigma^2} \right).$$

For the fact that, the obtained modified maximum likelihood estimator of mentioned models have not closed form, in order to show the values of MMLE are close to the true vector parameters we have done simulation study. We have considered different values for $n$. For each $n$, we estimate the unknown parameters. The results are presented in Table 1.

It is possible that data is generated from family near normal with more or less skewed so we consider a class of skew-normal models that include the normal distribution as a particular member. See Table 7 for kurtosis and skewness values. In the next subsection we find the MLE of the parameters based on the EM algorithm.
Table 1. The value of modified maximum likelihood estimators

<table>
<thead>
<tr>
<th>True model</th>
<th>n</th>
<th>$\hat{\phi}_1$</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>25</td>
<td>0.7702</td>
<td>1.3962</td>
<td>2.3184</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.7461</td>
<td>1.7571</td>
<td>1.8154</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.7424</td>
<td>1.7579</td>
<td>1.7205</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7232</td>
<td>1.8493</td>
<td>1.8540</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7212</td>
<td>1.9220</td>
<td>1.8561</td>
</tr>
<tr>
<td>$\epsilon_i \sim G(2, 2)$</td>
<td>400</td>
<td>0.7202</td>
<td>1.9232</td>
<td>1.8661</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>0.7191</td>
<td>1.9566</td>
<td>1.8732</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.7179</td>
<td>1.9624</td>
<td>1.8971</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.7155</td>
<td>1.9683</td>
<td>1.8942</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.7151</td>
<td>1.9737</td>
<td>1.8971</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.7131</td>
<td>1.9760</td>
<td>1.9133</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.7120</td>
<td>1.9665</td>
<td>1.9171</td>
</tr>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>400</td>
<td>0.7084</td>
<td>2.9082</td>
<td>4.8602</td>
</tr>
<tr>
<td>$\epsilon_i \sim W(3, 5)$</td>
<td>500</td>
<td>0.7075</td>
<td>2.9191</td>
<td>4.8833</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.7061</td>
<td>2.9332</td>
<td>4.9012</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.7043</td>
<td>2.9474</td>
<td>4.9171</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.7041</td>
<td>2.9539</td>
<td>4.9271</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.7031</td>
<td>2.9634</td>
<td>4.9365</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.7030</td>
<td>2.9652</td>
<td>4.9413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True model</th>
<th>n</th>
<th>$\hat{\phi}_2$</th>
<th>$\hat{\gamma}$</th>
<th>$\hat{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>50</td>
<td>0.7333</td>
<td>2.7373</td>
<td>4.8884</td>
</tr>
<tr>
<td>$\epsilon_i \sim W(3, 5)$</td>
<td>100</td>
<td>0.7220</td>
<td>2.7771</td>
<td>4.6382</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7135</td>
<td>2.8691</td>
<td>4.7745</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7102</td>
<td>2.8935</td>
<td>4.8382</td>
</tr>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>400</td>
<td>0.7084</td>
<td>2.9082</td>
<td>4.8602</td>
</tr>
<tr>
<td>$\epsilon_i \sim N(2, 1)$</td>
<td>500</td>
<td>0.7075</td>
<td>2.9191</td>
<td>4.8833</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.7061</td>
<td>2.9332</td>
<td>4.9012</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.7043</td>
<td>2.9474</td>
<td>4.9171</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.7041</td>
<td>2.9539</td>
<td>4.9271</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.7031</td>
<td>2.9634</td>
<td>4.9365</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.7030</td>
<td>2.9652</td>
<td>4.9413</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>True model</th>
<th>n</th>
<th>$\hat{\phi}_3$</th>
<th>$\hat{\mu}$</th>
<th>$\hat{\sigma}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>50</td>
<td>0.7324</td>
<td>1.6500</td>
<td>1.3691</td>
</tr>
<tr>
<td>$\epsilon_i \sim N(2, 1)$</td>
<td>100</td>
<td>0.7244</td>
<td>1.7312</td>
<td>1.2983</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.7183</td>
<td>1.8036</td>
<td>1.2435</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.7162</td>
<td>1.8332</td>
<td>1.2151</td>
</tr>
<tr>
<td>$x_i = 0.7x_{i-1} + \epsilon_i$</td>
<td>400</td>
<td>0.7140</td>
<td>1.8432</td>
<td>1.2041</td>
</tr>
<tr>
<td>$\epsilon_i \sim LN(2, 1)$</td>
<td>500</td>
<td>0.7132</td>
<td>1.8603</td>
<td>1.1822</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.7124</td>
<td>1.8674</td>
<td>1.1791</td>
</tr>
<tr>
<td></td>
<td>700</td>
<td>0.7123</td>
<td>1.8675</td>
<td>1.1732</td>
</tr>
<tr>
<td></td>
<td>800</td>
<td>0.7121</td>
<td>1.8715</td>
<td>1.1671</td>
</tr>
<tr>
<td></td>
<td>900</td>
<td>0.7120</td>
<td>1.8716</td>
<td>1.1620</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.7120</td>
<td>1.8742</td>
<td>1.1582</td>
</tr>
</tbody>
</table>
2.4 Autoregressive Model with Skew Normal Residual

The first autoregressive model with skew-normal residuals is expressed as

\[ x_i = \phi_4 x_{i-1} + \epsilon_i, \quad i = 2, \ldots, n \]

where residuals have a skew-normal distribution with the location parameter, \( \mu \), scale parameter, \( \sigma^2 \), and \( \lambda \) as the skewness parameter. Its density function is

\[
f(\epsilon_i) = \frac{2}{\sigma_s} \phi\left(\frac{\epsilon_i - \mu_s}{\sigma_s}\right) \Phi\left(\lambda \frac{\epsilon_i - \mu_s}{\sigma_s}\right)
\]

where \( \phi(\cdot) \) and \( \Phi(\cdot) \) are the standard normal density and normal distribution function, respectively. The log-likelihood function can be written as

\[
l(\phi_4, \mu_s, \sigma^2_s, \lambda) = n \log(2) - n \log(2\pi\sigma^2_s) - \frac{1}{2\sigma^2_s} \sum_{i=2}^{n} (x_i - \phi_4 x_{i-1} - \mu_s)^2 + \sum_{i=2}^{n} \log \Phi\left(\lambda \frac{x_i - \phi_4 x_{i-1} - \mu_s}{\sigma_s}\right)
\]

The first order derivatives are listed below,

\[
\frac{\partial}{\partial \mu_s} l(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \frac{(x_i - \phi_4 x_{i-1} - \mu_s)}{\sigma^2_s} - \frac{\lambda}{\sigma_s} W(k_i)
\]

\[
\frac{\partial}{\partial \sigma^2_s} l(\phi_4, \mu_s, \sigma^2_s, \lambda) = -\frac{n}{2\sigma^2_s} - \frac{n}{2\sigma^4_s} \sum_{i=2}^{n} (x_i - \phi_4 x_{i-1} - \mu_s)^2 - \frac{\lambda}{2\sigma^3_s} W(k_i)
\]

\[
\frac{\partial}{\partial \lambda} l(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \frac{(x_i - \phi_4 x_{i-1} - \mu_s)}{\sigma_s} W(k_i)
\]

\[
\frac{\partial}{\partial \phi_4} l(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \left[ \frac{(x_i - \phi_4 x_{i-1} - \mu_s)}{\sigma^2_s} - \frac{\lambda}{\sigma_s} W(k_i) \right] x_{i-1},
\]

where \( k_i = \lambda \frac{x_i - \phi_4 x_{i-1} - \mu_s}{\sigma} \) and \( W(k_i) = \frac{\phi(k_i)}{\Phi(k_i)} \). We compute the maximum likelihood estimates based on the EM algorithm. The model can be expressed as

\[
\epsilon_i | Z_i = z_i \sim N(\mu_s + \frac{\lambda \sigma_s}{\sqrt{1 + \lambda^2}} z_i, \frac{\sigma^2_s}{\sqrt{1 + \lambda^2}})
\]
\[ Z_i \sim HN(0, 1), \]

where \( HN(0, 1) \) denotes the standardized univariate half-normal distribution. See Cancho et al. (2008). Note that \( Z_i, i = 1, \ldots, n \) and \( X_i, i = 1, \ldots, n \) can be treated as missing and observed data, respectively and \( Y_c = (X_i, Z_i) \) denotes the complete data. The complete data log likelihood, ignoring additive constant terms, is

\[
l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = -\frac{n}{2} \log \sigma^2_s + \frac{n}{2} \log(1 + \lambda^2) - \frac{1 + \lambda^2}{2\sigma^2_s} \sum_{i=1}^{n} \left( x_i - \phi_4 x_{i-1} - \mu_s - \frac{\lambda \sigma_s}{\sqrt{1 + \lambda^2}} \hat{z}_i \right)^2
\]

- **Expectation step:** in this step we calculate the expected value of the log likelihood function, with respect to the conditional distribution of \( Z \) given \( X \) under the current estimate of the parameters \( \hat{\nu}^{(k)} = (\hat{\phi}_4^{(k)}, \hat{\mu}_s^{(k)}, \hat{\sigma}^2_s^{(k)}, \hat{\lambda}^{(k)}) \) :

\[
\hat{z}_i = E(Z_i|X, \hat{\nu}^{(k)}) = T_1 + \frac{\phi_4(T_1)}{\Phi(T_2)} T_2
\]

\[
\hat{z}_i^2 = E(Z_i^2|X, \hat{\nu}^{(k)}) = T_1^2 + T_2^2 + \frac{\phi_4(T_1)}{\Phi(T_2)} T_1 T_2
\]

where \( T_1 + \frac{\lambda}{\sigma_s \sqrt{1 + \lambda^2}} \epsilon_i \) and \( T_2^2 = \frac{1}{1 + \lambda^2} \).

- **Maximization step:** The first and second derivatives are presented below.

\[
\frac{\partial}{\partial \phi_4} l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s} \left( \frac{a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i \right) x_{i-1}
\]

\[
\frac{\partial}{\partial \mu_s} l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s} \left( \frac{a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i \right)
\]

\[
\frac{\partial}{\partial \sigma^2_s} l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = -\frac{n}{2\sigma^2_s} + \sum_{i=2}^{n} \frac{1 + \lambda^2}{2\sigma^3_s} \left( \frac{a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i \right) a_i
\]
\[
\frac{\partial}{\partial \lambda} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = \sum_{i=2}^{n} \left[ \frac{\lambda}{1 + \lambda^2} + \frac{1}{\sqrt{1 + \lambda^2}} \left( \frac{a_i}{\sigma_s} \hat{z}_i - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i^2 \right) \right.
\]
\[
- \frac{\lambda a_i^2}{\sigma_s^2} - \frac{\lambda^3}{1 + \lambda^2} \hat{z}_i + \frac{2a_i \lambda^2}{\sigma_s \sqrt{1 + \lambda^2}} \hat{z}_i \right]
\]
\[
\frac{\partial^2}{\partial \phi_4 \partial \phi_4} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s^2} x_{i-1}^2
\]
\[
\frac{\partial^2}{\partial \phi_4 \partial \mu_s} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s} x_{i-1}
\]
\[
\frac{\partial^2}{\partial \phi_4 \partial \sigma_s^2} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = \sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s^2} \left( \frac{a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda}} \hat{z}_i \right) x_{i-1}
\]
\[
\frac{\partial^2}{\partial \phi_4 \partial \lambda} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = \sum_{i=2}^{n} \frac{2 \lambda}{\sigma_s} \frac{1 + \lambda^2}{2 \sigma_s^2} \left( \frac{a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda}} \hat{z}_i \right) x_{i-1}
\]
\[
- \sum_{i=2}^{n} \frac{x_{i-1}}{\sigma_s \sqrt{1 + \lambda^2}} \hat{z}_i
\]
\[
\frac{\partial^2}{\partial \mu_s \partial \sigma_s^2} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \frac{1 + \lambda^2}{\sigma_s^2} \left( \frac{2a_i}{\sigma_s} - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i \right)
\]
\[
\frac{\partial^2}{\partial \mu_s \partial \lambda} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \left[ \frac{1}{\sigma_s \sqrt{1 + \lambda^2}} \hat{z}_i - \frac{2\lambda}{\sigma_s^2} a_i + \frac{2 \lambda^2}{\sigma_s \sqrt{1 + \lambda^2}} \hat{z}_i \right]
\]
\[
\frac{\partial^2}{\partial \sigma_s^2 \partial \sigma_s^2} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = \frac{n}{2 \sigma_s^4} \sum_{i=2}^{n} \frac{1 + \lambda^2}{4 \sigma_s^5} \frac{4a_i}{\sigma_s} \left( \frac{3\lambda}{\sigma_s \sqrt{1 + \lambda^2}} \hat{z}_i \right) a_i
\]

\[
\frac{\partial^2}{\partial \sigma^2 \partial \lambda} l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = \sum_{i=2}^{n} \frac{\lambda}{\sigma_s^2} \left( 2 \sigma_i^2 \sqrt{1 + \lambda^2} \hat{z}_i \right) + \sum_{i=2}^{n} \frac{a_i}{\sigma_s^2} \left( \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i - \frac{a_i}{\sqrt{1 + \lambda^2}} \hat{z}_i \right)
\]

\[
\frac{\partial^2}{\partial \lambda \partial \lambda} l_c(\phi_4, \mu_s, \sigma^2_s, \lambda) = \frac{n}{(1 + \lambda^2)} + \sum_{i=2}^{n} \frac{\lambda}{\sqrt{(1 + \lambda^2)^3}} \left( \frac{a_i}{\sigma_s^2} \hat{z}_i - \frac{\lambda}{\sqrt{1 + \lambda^2}} \hat{z}_i^2 \right)
\]

where \( a_i = x_i - \phi_4 x_{i-1} - \mu_s \). Thus, the \((k+1)\)th estimate of parameter \( \upsilon \) can be obtained by

\[
\hat{\upsilon}(k+1) = \hat{\upsilon}(k) + J(\hat{\upsilon}(k))^{-1} U(\hat{\upsilon}(k))
\]

where \( U(\upsilon) = -\frac{\partial^2}{\partial \upsilon \partial \upsilon} l_c(\upsilon) \) and \( J(\upsilon) = -\frac{\partial}{\partial \upsilon} l_c(\upsilon) \).

### 3 Simulation Study

Based on the simulation study, we have shown that the information criteria such as AIC, AIC\(_c\) and SIC are appropriate criteria for optimum model selection for autoregressive models with non-negative residuals based on the modified maximum likelihood estimators. Assume that the set of data \( \{x_1, \ldots, x_n\} \) is generated by Weibull autoregressive model. In the other hand \( x_i = 0.7x_{i-1} + \varepsilon_i \), where \( \varepsilon_i \)'s are i.i.d \( W(3,5) \). Consider first order Gamma autoregressive model, GAR(1), first order Weibull autoregressive model, W AR(1), and first order Log-normal autoregressive model, LNAR(1), as three competing models. By using obtained MMLE in the provide section and an available data, we estimate parameter of GAR(1), W AR(1) and LNAR(1) and compute the value of AIC, AIC\(_c\) and SIC for the three competing models for different \( n \). The value of these information criteria for GAR(1), W AR(1) and LNAR(1) are given in Table 2. It shows that for each \( n \), W AR(1) model is optimum model because \( IC(f^W_1) < IC(f^G_1) < IC(f^L_1) \). The IC shows all of the information criteria AIC, SIC and AIC\(_c\) and SIC for the three competing models for different \( n \). The value of these information criteria for GAR(1), W AR(1) and LNAR(1) are given in Table 2. It shows that for each \( n \), W AR(1) model is optimum model because \( IC(f^W_1) < IC(f^G_1) < IC(f^L_1) \). The IC shows all of the information criteria AIC, SIC and AIC\(_c\), where \( h(x) \) is true density, \( W(3,5) \), and \( f^W_1(x) \), \( f^G_1(x) \) and \( f^L_1(x) \) are Gamma, Weibull and Log-normal autoregressive models, respectively, \( \eta = (\alpha, \beta, \phi_1) \), \( \psi = (\gamma, \tau, \phi_2) \) and \( \theta = (\mu, \sigma, \phi_3) \). For more illustration see Figure 1. Since AIC is an estimator of Kullback-Leibler criterion we can conclude that

\[
\mathcal{K}L(g, f^W_1) < \mathcal{K}L(g, f^G_1) < \mathcal{K}L(g, f^L_1).
\]
Table 2. The value of information criteria for competing models

<table>
<thead>
<tr>
<th>n</th>
<th>$f_1^n$</th>
<th>$f_2^n$</th>
<th>$f_3^n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>245.2012</td>
<td>186.6434</td>
<td>276.3434</td>
</tr>
<tr>
<td>100</td>
<td>478.3437</td>
<td>375.1828</td>
<td>548.6551</td>
</tr>
<tr>
<td>AIC</td>
<td>250</td>
<td>1106.8554</td>
<td>927.1864</td>
</tr>
<tr>
<td>500</td>
<td>2145.6165</td>
<td>1926.8272</td>
<td>2475.0494</td>
</tr>
<tr>
<td>50</td>
<td>250.9370</td>
<td>209.0045</td>
<td>282.0794</td>
</tr>
<tr>
<td>100</td>
<td>486.1592</td>
<td>384.7092</td>
<td>556.4705</td>
</tr>
<tr>
<td>SIC</td>
<td>250</td>
<td>1117.4193</td>
<td>945.6185</td>
</tr>
<tr>
<td>500</td>
<td>2158.2601</td>
<td>1946.8844</td>
<td>2487.6937</td>
</tr>
<tr>
<td>50</td>
<td>245.7227</td>
<td>187.1652</td>
<td>276.8651</td>
</tr>
<tr>
<td>100</td>
<td>286.9998</td>
<td>268.8021</td>
<td>442.4647</td>
</tr>
<tr>
<td>AICc</td>
<td>250</td>
<td>1106.9528</td>
<td>927.2839</td>
</tr>
<tr>
<td>500</td>
<td>2145.6655</td>
<td>1926.8759</td>
<td>2475.0988</td>
</tr>
</tbody>
</table>

The Kolmogorov-Smirnov test confirms these results. The values of the Kolmogorov-Smirnov test for different $n$ are given in Table 3. It shows that all of the P-values of Kolmogorov-Smirnov test of estimated Weibull autoregressive are greater than 0.05.

Figure 1. Weibull, Gamma and Log-normal autoregressive model curves

The Kolmogorov-Smirnov test confirms these results. The values of the Kolmogorov-Smirnov test for different $n$ are given in Table 3. It shows that all of the P-values of Kolmogorov-Smirnov test of estimated Weibull autoregressive are greater than 0.05.

Table 3. The P-values of the Kolmogorov-Smirnov test

<table>
<thead>
<tr>
<th>n</th>
<th>$f_1^\eta$</th>
<th>$f_2^\psi$</th>
<th>$f_3^\theta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.1396</td>
<td>0.3321</td>
<td>0.0830</td>
</tr>
<tr>
<td>100</td>
<td>0.0686</td>
<td>0.3358</td>
<td>0.0003</td>
</tr>
<tr>
<td>250</td>
<td>0.1188</td>
<td>0.7304</td>
<td>0.0000</td>
</tr>
<tr>
<td>500</td>
<td>0.1617</td>
<td>0.9184</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Now consider first order autoregressive models with $W(3,5)$ residual as true model and with $G(5,1.2)$, $W(3,0.2)$ and $LN(1,0.2)$ residuals as three competing models. Based on the Figure 2 we can say, first order autoregressive models with $G(5,1.2)$ is optimal model. It is result that we can achieve by information criteria well. because

$$IC(f_1^\eta) < IC(f_3^\theta) < IC(f_2^\psi).$$

![Figure 2](image-url)  
**Figure 2.** Weibull, Gamma and Log-normal autoregressive model curves

The values of information criteria for three competing models are presented in Table 4.
Table 4. The value of information criteria for competing models

<table>
<thead>
<tr>
<th>n</th>
<th>W(3,0.2)</th>
<th>G(5,1.2)</th>
<th>LN(1,0.2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>2434835</td>
<td>256.2216</td>
<td>765.2733</td>
</tr>
<tr>
<td>100</td>
<td>5248667</td>
<td>510.1005</td>
<td>1564.9226</td>
</tr>
<tr>
<td>AIC</td>
<td>250</td>
<td>11575308</td>
<td>4252.7522</td>
</tr>
<tr>
<td>500</td>
<td>21725749</td>
<td>2602.6340</td>
<td>7820.6584</td>
</tr>
<tr>
<td>100</td>
<td>5250142</td>
<td>517.9161</td>
<td>1572.7382</td>
</tr>
<tr>
<td>SIC</td>
<td>250</td>
<td>11576786</td>
<td>4263.3172</td>
</tr>
<tr>
<td>500</td>
<td>21727228</td>
<td>2615.2785</td>
<td>7833.3023</td>
</tr>
<tr>
<td>50</td>
<td>2436308</td>
<td>261.9577</td>
<td>771.0094</td>
</tr>
<tr>
<td>100</td>
<td>5250142</td>
<td>517.9161</td>
<td>1572.7382</td>
</tr>
<tr>
<td>AICc</td>
<td>250</td>
<td>11575308</td>
<td>4252.8595</td>
</tr>
<tr>
<td>500</td>
<td>21725749</td>
<td>2602.6826</td>
<td>7820.7075</td>
</tr>
</tbody>
</table>

4 Real Data Analysis

The Lake series shows that the level in feet of Lake Huron in the years 1875-1930. This data can be found in Itsm data libraries. A graph of the level in feet of Lake Huron is displayed in Figure 3.

Figure 3. Level of Lake Huron 1875 – 1930 curve
The sample autocorrelation function, ACF, suggests that an autoregressive model might provide a reasonable model for given data. The sample partial autocorrelation function, PACF, of the data is slightly outside the bounds \( \pm 1.96/\sqrt{55} \) at lag 1. So, we can suggest first order autoregressive model for the data. The ACF and PACF are shown in Figure 4.

Based on the Yule-Walker method for parameter estimation, the autoregressive coefficient is 0.81. We can compute the residuals \( \epsilon_i = x_i - 0.81x_{i-1} \). Since all of residuals are non-negative, we suggest non-negative competing models. Consider first order autoregressive model with Gamma, Weibull, Log-normal and Skew-Normal residuals as four competing models. The Table 5 shows the estimated values of parameters of first order autoregressive model with Gamma, Weibull and Log-normal residuals based on modified maximum likelihood method and first order autoregressive model with Skew-Normal residuals based on EM algorithm.
Table 5. Estimation of parameters

<table>
<thead>
<tr>
<th></th>
<th>first parameter</th>
<th>second parameter</th>
<th>third parameter</th>
<th>ϕ</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G(\alpha, \beta)$</td>
<td>1.2516</td>
<td>1.6272</td>
<td>-</td>
<td>0.8466</td>
</tr>
<tr>
<td>$W(\upsilon, \tau)$</td>
<td>2.4898</td>
<td>1.7283</td>
<td>-</td>
<td>0.8468</td>
</tr>
<tr>
<td>$LN(\mu, \sigma)$</td>
<td>0.4574</td>
<td>0.2603</td>
<td>-</td>
<td>0.8253</td>
</tr>
<tr>
<td>$SN(\mu_\epsilon, \sigma_\epsilon, \lambda)$</td>
<td>1.4996</td>
<td>0.7507</td>
<td>0.9862</td>
<td>0.8098</td>
</tr>
</tbody>
</table>

The AIC, SIC, AIC$_c$ and P-value of Kolmogorov-Smirnov test are given in Table 6. Because

$$IC(f^\psi_1) < IC(f^\upsilon_2) < IC(f^\theta_3) < IC(f^\eta_1)$$

the first order autoregressive with Weibull distribution as a suitable model for residuals, is the best model among the other competing models. The Kolmogorov-Smirnov test confirms this result.

Table 6. The value of information criteria and P-value of Kolmogorov-Smirnov test

<table>
<thead>
<tr>
<th></th>
<th>AIC</th>
<th>SIC</th>
<th>AIC$_c$</th>
<th>P-value of $K - S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$G(\alpha, \beta)$</td>
<td>152.1973</td>
<td>158.2193</td>
<td>152.6679</td>
<td>0.0007</td>
</tr>
<tr>
<td>$W(\upsilon, \tau)$</td>
<td>107.9387</td>
<td>116.9607</td>
<td>108.4093</td>
<td>0.9091</td>
</tr>
<tr>
<td>$LN(\mu, \sigma^2)$</td>
<td>117.0986</td>
<td>126.1206</td>
<td>117.5692</td>
<td>0.4111</td>
</tr>
<tr>
<td>$SN(\mu_\epsilon, \sigma_\epsilon, \lambda)$</td>
<td>116.5389</td>
<td>124.5683</td>
<td>117.3389</td>
<td>0.4383</td>
</tr>
</tbody>
</table>

Figure 5 shows the histogram of residuals and four competing models. It suggests Weibull model, $W(2.4898,1.7283)$, as optimal model.

The summary of information about Lake Huron data, residuals of Lake Huron data and four competing models are given in Table 7. As we see, the mean and variance of autoregressive with Weibull residuals are near to the mean and variance of observation respectively.
Figure 5. Histogram of residuals and Gamma, Weibull and Log-normal model curve

Table 7. Summary of information about the Lake Huron data

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1st Qu</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Qu</th>
<th>Max</th>
<th>Var</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAR(1)</td>
<td>9.5832</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAR(1)</td>
<td>12.7305</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNAR(1)</td>
<td>2.5411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNAR(1)</td>
<td>10.0992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References


S. Zamani Mehreyan  
Department of Statistics,  
Razi University,  
Kermanshah, Iran.  
email: s.zamani121@yahoo.com

A. Sayyareh  
Faculty of Mathematics,  
K. N. Toosi University of Technology,  
Tehran, Iran.  
email: asayyareh@kntu.ac.ir