© Statistical Research and Training Center
مجله‌ی پژوهش‌های آماری ایران
J. Statist. Res. Iran **12** (2015): 83–104
دوره‌ی ۱۲، شماره‌ی ۱، بهار و تابستان ۱۳۹۴، صص ۸۳-۱۰۴
پژوهشکده‌ی آمار

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.

MSC 2010: 62F10.

1 Introduction

Let $n\times 1$ random vector $\mathbf{X}_t = (\mathbf{Y}_t, \mathbf{Z}_t)$, $t = 1, \ldots, n$ are i.i.d with common unknown true distribution $H(\cdot)$ on a complete probability space $(\mathbf{X}, \sigma_X)$, where $\mathbf{X}$ is the Euclidean space $\Re^m$ and σ_X is the Borel σ -field on $\mathbf{X}$. Let $(\mathbf{Y}, \sigma_Y)$ and $(\mathbf{Z}, \sigma_Z)$ be the measurable spaces associated with $\mathbf{Y}_t$ and $\mathbf{Z}_t$. We shall be

* Corresponding author

interested in the true conditional distribution $H_{Y|Z}(\cdot|V)$ of $\mathbf{Y}_t$ given $\mathbf{Z}_t$. Let H_Z be the true marginal distribution of $\mathbf{Z}_t$, and ν_Y be a σ -finite measure on $(\mathbf{Y}, \sigma_Y)$. For H_Z-almost all z, $H_{Y|Z}(\cdot|z)$ has a Radon-Nikodym density $h(\cdot|z)$ relative to ν_Y, which is strictly positive for ν_Y-almost all y.
We now consider two competing parametric families of conditional distribution defined on $\sigma_Y \times \mathbf{Z}$ for $\mathbf{Y}_t$ given $\mathbf{Z}_t$:

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

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

$$\mathcal{KL}\{h_{Y|Z}, g^{\beta}_{Y|Z}\} = \mathcal{E}_h \left\{ \log \frac{h(Y|Z)}{g^{\beta}(Y|Z)} \right\},$$

where $\mathcal{E}_h$ 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:

$$\mathcal{KL}\{h, g^{\beta}\} = \mathcal{E}_h \left\{ \log h_{Y|Z}(Y|Z) \right\} - \mathcal{E}_h \left\{ \log g^{\beta}_{Y|X}(Y|Z) \right\},$$

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 $\mathcal{KL}\{h(\cdot|\cdot), g^{\beta_*}(\cdot|\cdot)\}$, where β_* is the pseudo-true value of β, see e.g White

(1982). Thus, an equivalent selection criterion can be based on the quantity $\mathcal{E}_h\left\{\log g^{\beta_*}(Y|Z)\right\}$, the best model being the one for which this quantity is the largest. $\mathcal{KL}$ is a non-negative quantity. By definition, the more $g^{\beta}(\cdot|\cdot)$ agrees with $h(\cdot|\cdot)$ the smaller $\mathcal{KL}\{h(\cdot|\cdot), g^{\beta}(\cdot|\cdot))\}$ is. Then the closest member in $\mathcal{G}$ to the $h(\cdot|\cdot)$ is $g^{\beta_*}(\cdot|\cdot)$ where $\beta_* \in \mathcal{B}$ is the minimizer of $\mathcal{KL}\{h(\cdot|\cdot), g^{\beta}(\cdot|\cdot))\}$. For Kullback-Leibler divergence, $g^{\beta_*}(\cdot|\cdot)$ is the best approximation to $h(\cdot|\cdot)$ under model $\mathcal{G}$. 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 β_*, see White (1982).
If the model is misspecified, $\mathcal{KL}(h, g^{\beta}) > 0$. Hence $\mathcal{KL}$ divergence takes its value in $[0, \infty]$. The $\mathcal{KL}$ divergence is not a metric, but it is additive over marginal of product measures. $\mathcal{KL}(h, g^{\beta}) = 0$ 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 $\mathcal{KL}$ divergence is $\mathcal{E}_h\{\log g^{\beta}(Y|Z)\}$ 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

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

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

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

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

$$y_t = \phi_1 y_{i-1} + \cdots + \phi_p y_{i-p} + \epsilon_i$$

where $\phi = (\phi_1, \ldots, \phi_p)$ is autoregressive coefficients and ϵ_i'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 $S(\phi) = \sum_{i=p+1}^{n} \epsilon_i^2$. 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 ϕ is the autoregressive coefficient and ε_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 ϕ_1 is the autoregressive coefficient and the ε_i's are i.i.d error terms with Gamma distribution, G(α,β). We get the log-likelihood of $\epsilon_1, \ldots, \epsilon_n$, 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 ϕ_1 is functions in terms of z_i^{-1} and it has no explicit solutions. To obtain the explicit solution, we order ε_i (for a given ϕ) in order of increasing magnitude. So we obtain modified maximum likelihood estimators by solving the estimating equations

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

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

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

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

$$z_{(i)}^{-1} \simeq \alpha_i - \beta_i z_{(i)} \qquad (1)$$

where $\alpha_i = \frac{2}{t_{(i)}}$ and $\beta_i = \frac{1}{t_{(i)}^2}$. Incorporated Eq.(1) in estimating equations. We can obtain MMLE of α, β and ϕ_1 as

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

where Γ_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

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

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 ϵ_i's have the Weibull distribution, W(γ,τ)

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

The log-likelihood of $\epsilon_1, \ldots, \epsilon_n$ 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,

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

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

$$\frac{\partial}{\partial\phi_2}l(\phi_2, \gamma, \tau) = -\frac{(\gamma-1)}{\tau}\sum_{i=2}^{n} x_{[i]-1} z_{[i]}^{-1} + \frac{\gamma}{\tau}\sum_{i=2}^{n} x_{[i]-1} z_{[i]}^{\gamma-1} = 0,$$

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

$$z_{(i)}^{-1} \simeq v_{i0} - B_{i0} z_i,$$

$$z_{(i)}^{\gamma-1} \simeq v_{i0}^* + B_{i0}^* z_i$$

where $v_{i0} = \frac{2}{t_{(i)}}$, $B_{i0} = \frac{1}{t_{(i)}^2}$, $t_{(i)} = (-\log(1 - \frac{i}{n+1}))^{\frac{1}{\gamma}}$, $v_{i0}^* = (2-\gamma)t_{(i)}^{\gamma-1}$ 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 γ, τ and ϕ_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 $\Delta = B^2 - 4AC$, $A = -n$, $B = -\sum_{i=2}^{n}(x_{[i]} - \hat{\phi}_2 x_{[i]-1})v_i$ and $C = \sum_{i=2}^{n}(x_{[i]} - \hat{\phi}_2 x_{[i]-1})^2\beta_i$,

$$\hat{\phi}_2 = \frac{\sum_{i=2}^{n}\beta_i x_{[i]-1}x_{[i]} - \hat{\tau}\sum_{i=2}^{n} v_i x_{[i]-1}}{\sum_{i=2}^{n}\beta_i x_{[i]-1}^2}.$$

2.3 Autoregressive Model with Log-normal Residual

Here we consider the first order autoregressive model as $x_i = \phi_3 x_{i-1} + \varepsilon_i$ with Log-normal, LN(μ,σ) residuals. We get the log-likelihood of $\epsilon_1, \ldots, \epsilon_n$, as

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

We derive modified maximum likelihood estimators by solving the estimating equations

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

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

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

where $z_{(i)} = x_{[i]} - \phi_3 x_{[i]-1}$. Similarly we linearize the intractable terms, $z_{(i)}^{-1}$ and $\log(z_{(i)})$ in likelihood equations using the two terms of the Taylor series expansion, we have

$$z_{(i)}^{-1} \simeq \alpha_i - B_i z_{(i)}$$

and

$$\log(z_{(i)}) = c_i + \frac{\alpha_i}{2}z_{(i)}$$

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

$$\int_0^{t_{(i)}} \frac{1}{z\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(\log(z) - \mu)^2)dz = \frac{i}{n+1},$$

so we can obtain MMLE of μ, σ^2 and ϕ_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

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

$$A = \sum_{i=1}^{n}\frac{\alpha_i B_i}{2\hat{\sigma^2}} x^3_{[i]-1}$$

$$B = -\sum_{i=2}^{n} x^2_{[i]-1}\left(\frac{\alpha_i^2 - 2B_i(c_i - \hat{\mu} - 2\hat{\sigma^2})}{2\hat{\sigma^2}} - \frac{\alpha_i B_i}{\hat{\sigma^2}} x_{[i]}\right)$$

and

$$C = \sum_{i=2}^{n} x_{[i]-1}\left(\frac{\alpha_i(c_i - \mu + 2\hat{\sigma^2}) - B_i\alpha_i x^2_{[i]}}{2\hat{\sigma^2}}\right) + \sum_{i=2}^{n} x_{[i]-1}x_{[i]}\left(\frac{\alpha_i^2 - 2B_i(c_i - \mu - 2\hat{\sigma^2})}{2\hat{\sigma^2}}\right).$$

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 estimaors

True model	n	$\hat{\phi}_1$	$\hat{\alpha}$	$\hat{\beta}$
	25	0.7702	1.3962	2.3184
	50	0.7461	1.7571	1.8154
	100	0.7424	1.7579	1.7205
	200	0.7232	1.8493	1.8540
	300	0.7212	1.9220	1.8561
$x_i = 0.7x_{i-1} + \epsilon_i$	400	0.7202	1.9232	1.8661
$\epsilon_i \sim G(2,2)$	500	0.7191	1.9566	1.8732
	600	0.7179	1.9624	1.8971
	700	0.7155	1.9583	1.8942
	800	0.7151	1.9737	1.8971
	900	0.7131	1.9760	1.9133
	1000	0.7120	1.9665	1.9171
	n	$\hat{\phi}_2$	$\hat{\gamma}$	$\hat{\tau}$
	50	0.7333	2.7373	4.4884
	100	0.7220	2.7771	4.6382
	200	0.7135	2.8691	4.7745
	300	0.7102	2.8935	4.8382
$x_i = 0.7x_{i-1} + \epsilon_i$	400	0.7084	2.9082	4.8602
$\epsilon_i \sim W(3,5)$	500	0.7075	2.9191	4.8833
	600	0.7061	2.9332	4.9012
	700	0.7043	2.9474	4.9171
	800	0.7041	2.9539	4.9271
	900	0.7031	2.9634	4.9365
	1000	0.7030	2.9652	4.9413
	n	$\hat{\phi}_3$	$\hat{\mu}$	$\hat{\sigma}$
	50	0.7324	1.6500	1.3691
	100	0.7244	1.7312	1.2983
	200	0.7183	1.8036	1.2435
	300	0.7162	1.8332	1.2151
$x_i = 0.7x_{i-1} + \epsilon_i$	400	0.7140	1.8432	1.2041
$\epsilon_i \sim LN(2,1)$	500	0.7132	1.8603	1.1822
	600	0.7124	1.8674	1.1791
	700	0.7123	1.8675	1.1732
	800	0.7121	1.8715	1.1671
	900	0.7120	1.8716	1.1620
	1000	0.7120	1.8742	1.1582

2.4 Autoregressive Model with Skew Normal Residual

The first autoregressive model with skew-normal residuals is expressed as $x_i = \phi_4 x_{i-1} + \epsilon_i$, $i = 2, \ldots, n$ where residuals have a skew-normal distribution with the location parameter, μ, scale parameter, σ^2, and λ as the skewness parameter. Its density function is

$$f(\epsilon_i) = \frac{2}{\sigma_s}\phi(\frac{\epsilon_i - \mu_s}{\sigma_s})\Phi(\lambda\frac{\epsilon_i - \mu_s}{\sigma_s})$$

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

$$l(\phi_4, \mu_s, \sigma_s^2, \lambda) = n\log(2) - n\log(2\pi\sigma_s^2) - \frac{1}{2\sigma_s^2}\sum_{i=2}^{n}(x_i - \phi_4 x_{i-1} - \mu_s)^2 + \sum_{i=2}^{n}\log\Phi(\lambda\frac{x_i - \phi_4 x_{i-1} - \mu_s}{\sigma_s})$$

The first order derivatives are listed below,

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

$$\frac{\partial}{\partial\sigma_s^2}l(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\frac{n}{2\sigma_s^2} + \sum_{i=2}^{n}\frac{(x_i - \phi_4 x_{i-1} - \mu_s)^2}{2\sigma_s^4} - \frac{\lambda(x_i - \phi_4 x_{i-1} - \mu_s)}{2\sigma_s^3}W(k_i)$$

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

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

where $k_i = \lambda\frac{x_i - \phi_4 x_{i-1} - \mu_s}{\sigma_s}$ 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

$$\epsilon_i | Z_i = z_i \sim N(\mu_s + \frac{\lambda\sigma_s}{\sqrt{1+\lambda^2}}z_i, \frac{\sigma_s^2}{\sqrt{1+\lambda^2}}),$$

$$Z_i \sim HN(0,1),$$

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 treat 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

$$l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\frac{n}{2}\log\sigma_s^2 + \frac{n}{2}\log(1+\lambda^2) - \frac{1+\lambda^2}{2\sigma_s^2}\sum_{i=1}^{n}\left(x_i - \phi_4 x_{i-1} - \mu_s - \frac{\lambda\sigma_s}{\sqrt{1+\lambda^2}}z_i\right)^2$$

- 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 $\boldsymbol{\upsilon}^{(k)} = (\hat{\phi}^{(k)}, \hat{\mu_s}^{(k)}, \hat{\sigma_s^2}^{(k)} \hat{\lambda}^{(k)})$:

  $$\hat{z}_i = E(Z_i|X, \hat{\upsilon}^{(k)}) = T_1 + \frac{\phi(\frac{T_1}{T_2})}{\Phi(\frac{T_1}{T_2})}T_2$$

  $$\hat{z^2}_i = E(Z_i^2|X, \hat{\upsilon}^{(k)}) = T_1^2 + T_2^2 + \frac{\phi(\frac{T_1}{T_2})}{\Phi(\frac{T_1}{T_2})}T_1T_2,$$

  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.

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

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

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

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

$$\frac{\partial^2}{\partial\phi_4 \partial\phi_4} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \frac{1+\lambda^2}{\sigma_s^2} x_{i-1}^2$$

$$\frac{\partial^2}{\partial\phi_4 \partial\mu_s} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\sum_{i=2}^{n} \frac{1+\lambda^2}{\sigma_s^2} x_{i-1}$$

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

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

$$\frac{\partial^2}{\partial\mu_s \partial\mu_s} l_c(\phi_4, \mu_s, \sigma_s^2, \lambda) = -\frac{n(1+\lambda^2)}{\sigma_s^2}$$

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

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

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

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

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

where $a_i = x_i - \phi_4 x_{i-1} - \mu_s$. Thus, the $(k+1)$th estimate of parameter υ can be obtained by $\hat{\upsilon}^{(k+1)} = \hat{\upsilon}^{(k)} + J(\hat{\upsilon}^{(k)})^{-1}U(\hat{\upsilon}^{(k)})$ where $U(\upsilon) = \frac{\partial}{\partial \upsilon} l_c(\upsilon)$ and $J(\upsilon) = -\frac{\partial^2}{\partial \upsilon \partial \upsilon} l_c(\upsilon)$.

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 ε_i's are i.i.d W(3,5). Consider first order Gamma autoregressive model, GAR(1), first order Weibull autoregressive model, WAR(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), WAR(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), WAR(1) and LNAR(1) are given in Table 2. It shows that for each n, WAR(1) model is optimum model because $IC(f_2^{\psi}) < IC(f_1^{\eta}) < IC(f_3^{\theta})$. The IC shows all of the information criteria AIC, SIC and AIC_c, where $h(x)$ is true density, W(3,5), and $f_1^{\eta}(x)$, $f_2^{\psi}(x)$ and $f_3^{\theta}(x)$ 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

$$\mathcal{KL}(g, f_2^{\psi}) < \mathcal{KL}(g, f_1^{\eta}) < \mathcal{KL}(g, f_3^{\theta}).$$

Table 2. The value of information criteria for competing models

	n	f_1^{η}	f_2^{ψ}	f_3^{θ}
	50	245.2012	186.6434	276.3434
	100	478.3437	375.1828	548.6551
AIC	250	1106.8554	927.1864	1312.6081
	500	2145.6165	1926.8272	2475.0494
	50	250.9370	209.0045	282.0794
	100	486.1592	384.7092	556.4705
SIC	250	1117.4193	945.6185	1323.1730
	500	2158.2601	1946.8844	2487.6937
	50	245.7227	187.1652	276.8651
	100	286.9998	268.8021	442.4647
AIC_c	250	1106.9528	927.2839	1312.7064
	500	2145.6655	1926.8759	2475.0988

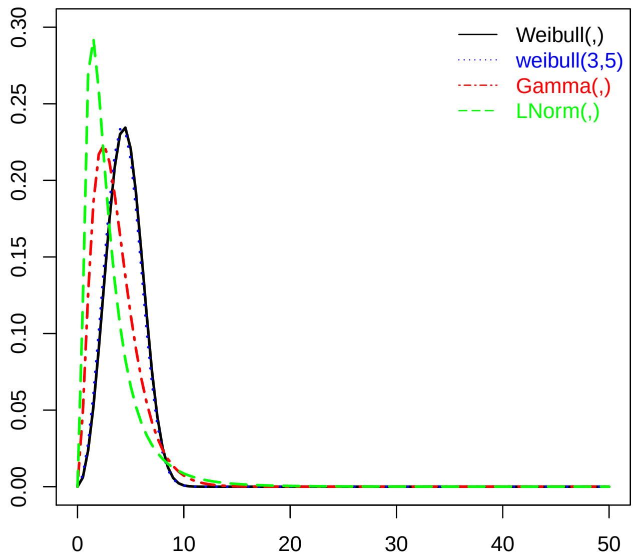

Figure 1. Weibull, Gamma and Log–normal autoregressive model curves

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

n	f_1^η	f_2^ψ	f_3^θ
50	0.1396	0.3321	0.0830
100	0.0686	0.3358	0.0003
250	0.1188	0.7304	0.0000
500	0.1617	0.9184	0.0000

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

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

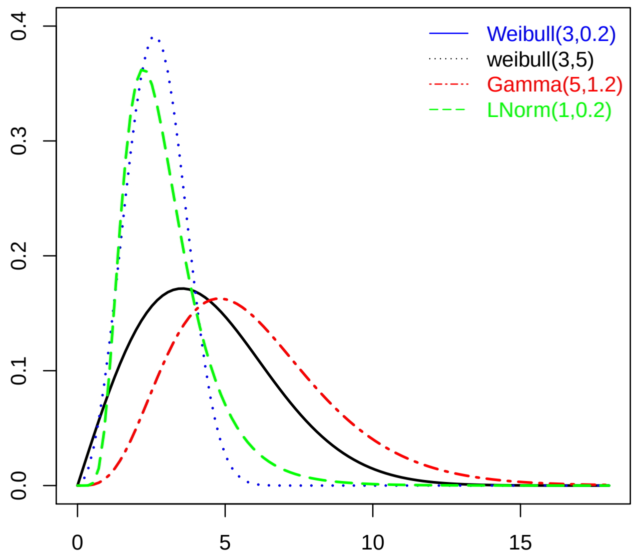

Figure 2. Weibull, Gamma and Log–normal autoregressive model curves

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

Table 4. The value of information criteria for competing models

	n	W(3,0.2)	G(5,1.2)	LN(1,0.2)
AIC	50	2434835	256.2216	765.2733
	100	5248667	510.1005	1564.9226
	250	11575308	1350.6731	4252.7522
	500	21725749	2602.6340	7820.6584
SIC	50	2436308	261.9577	771.0094
	100	5250142	517.9161	1572.7382
	250	11576786	1361.2376	4263.3172
	500	21727228	2615.2785	7833.3023
AIC_c	50	2434836	256.7434	765.7951
	100	5248667	510.3505	1565.1722
	250	11575308	1350.7725	4252.8595
	500	21725749	2602.6826	7820.7075

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.

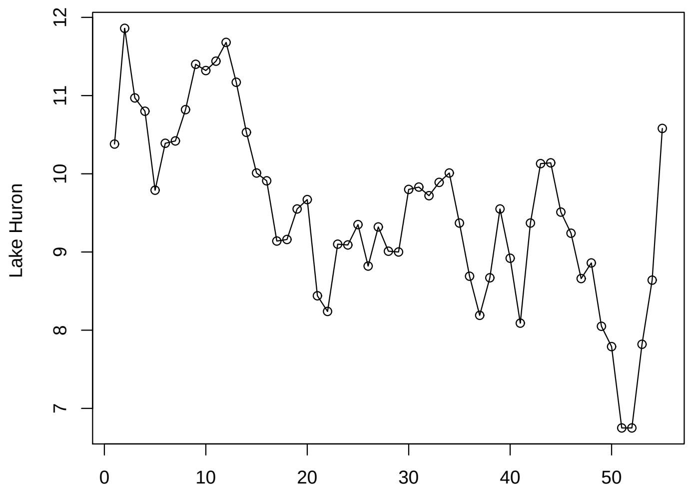

Figure 3. Level of Lake Huron 1875 − 1930 curve

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.

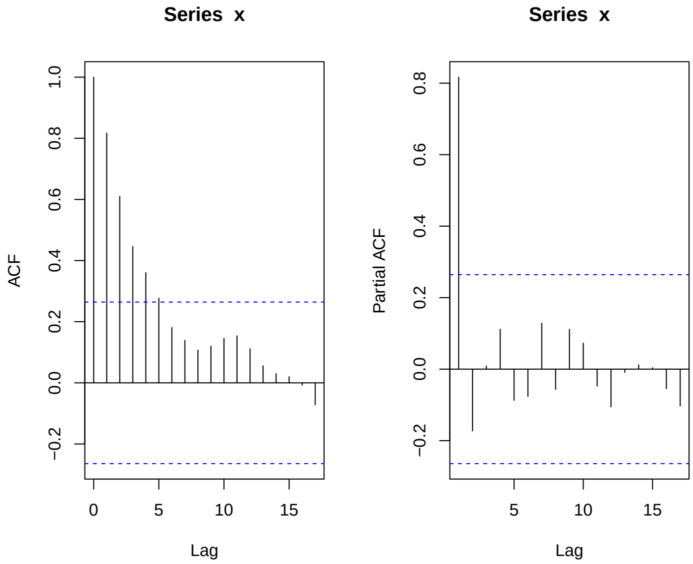

Figure 4. The sample autocorrelation function and partial autocorrelation function of Lake Huron 1875 − 1930

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

	first parameter	second parameter	third parameter	ϕ
$G(\alpha,\beta)$	1.2516	1.6272	-	0.8466
$W(\upsilon,\tau)$	2.4898	1.7283	-	0.8468
$LN(\mu,\sigma)$	0.4574	0.2603	-	0.8253
$SN(\mu_s,\sigma_s,\lambda)$	1.4996	0.7507	0.9862	0.8098

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

$$IC(f_2^{\psi}) < IC(f^{\upsilon}) < IC(f_3^{\theta}) < IC(f_1^{\eta})$$

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

	AIC	SIC	AIC_c	P-value of $K-S$
$G(\alpha,\beta)$	152.1973	158.2193	152.6679	0.0007
$W(\upsilon,\tau)$	107.9387	116.9607	108.4093	0.9091
$LN(\mu,\sigma^2)$	117.0986	126.1206	117.5692	0.4111
$SN(\mu_s,\sigma_s,\lambda)$	116.5389	124.5683	117.3389	0.4383

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.

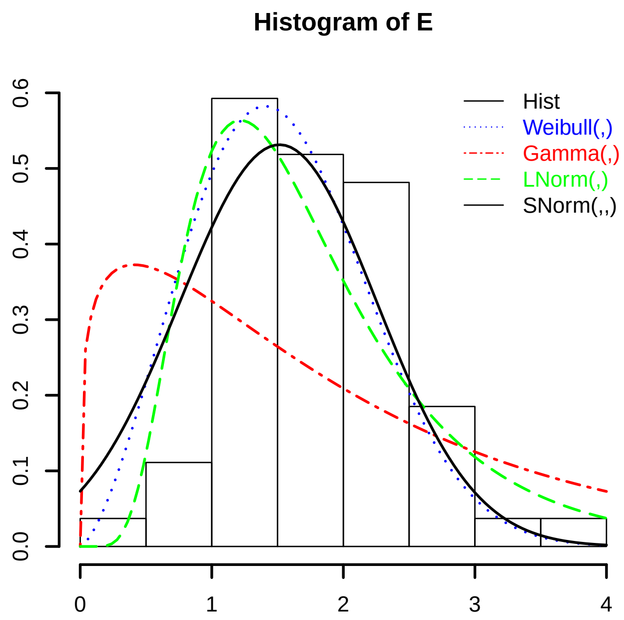

Figure 5. Histogram of residuals and Gamma، Weibull and Log-normal model curve

Table 7. Summary of information about the Lake Huron data

	Min	1st Qu	Median	Mean	3rd Qu	Max	Var	skewness	kurtosis
Observation	6.750	8.840	9.510	9.524	10.260	11.860	1.3077	-0.117	2.8110
WAR(1)				9.5832			1.4729		
GAR(1)				12.7305			11.2588		
LNAR(1)				2.5411			0.7947		
SNAR(1)				10.0992			1.5628		

References

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle, in the Second International Symposium on Information Theory, eds. by B.N. Petrov and F. Csake, Akademiai Kiado, Hungary, 267-281.

Akkaya, A. and Tiku, M. (2007). Estimating Parameters Autoregressive Models in Non-normal Situation: Asymmetric Innovations, **30**, 517-536.

Bayrak, O.T. and Akkaya, A. (2010). Estimating Parameters of a Multiple Autoregressive Model by the Modified Maximum Likelihood Method, *Journal of Computational and Applied Mathematics*, **233**, 1763-1772.

Cancho, V.G., Lachos, V.H. and ortega, EM. (2008). A Nonlinear Regression Model with Skew-Normal Errors. *Statistical Papers*, doi:10.1007/s00362-008-0139-y.

Claeskens, G., Croux, C. and Kerckhoven, J.V. (2007). Prediction Focussed Model Selection for Autoregressive Models, *Australian New Zealand Journal of Statistics*, **49**, 359-379.

Claeskens, G. and Hjort, N.L. (2003). The Focused Information Criterion [with discussion], *Journal of the American Statistical Association*, **98**, 900-916.

Commenges, D., Sayyareh, A., Letenneur, L., Guedj, J., and Bar-Hen, A. (2008). Estimating a Difference of Kullback-Leibler Risks Using a Normalized Difference of AIC, *The Annals of Applied Statistics*, **3**, 1123-1142.

De Gooijer, J.G., Abrraham, B., Gould, A. and Rabinson, L. (1985). Method for Determining the Order of an Autoregressive-Moving Average Process: A Survey, *International Statistical Review*, **53**, 301-329.

Hurvich, C.M. and Tsai, C.L. (1993). A Corrected Akaike Information Criterion for Vector Autoregressive Model Selection, *Journal of Time Series Analysis*, **14**, 271-279.

Kullback, S. and Leibler, R.A. (1951). Information and Sufficiency Modelling. *Annals of Mathematical Statistics*, **22**, 79-86.

Schwarz, G. (1978). Estimating the Dimension of a Model. *Annals of Statistics*, **6**, 461-464.

Tiku, M.L. (1967). Estimating the Mean and Standard Deviation from a Censored Normal Sample, *Biometrica*, **54**, 155-165.

Vuong, Q.H. (1989). Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses, *Econometrica*, **57**, 307-333.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, **50**, 1-25.

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*