© Statistical Research and Training Center

مجله‌ی پژوهش‌های آماری ایران
پژوهشکده‌ی آمار
دوره‌ی ۱۷، شماره‌ی ۲، پاییز و زمستان ۱۳۹۹، صص ۲۷۷–۳۱۳

J. Statist. Res. Iran **17** (2020): 277–313
DOR: 20.1001.1.25385771.2021.17.2.1.8

Statistical Inference in the First-Order Autoregressive-Moving Average Models with Skewed Non-Negative Innovations

Maryam Kasraie[†] and Abdolreza Sayyareh[‡,*]

[†] Islamic Azad University
[‡] K.N. Toosi University of Technology

Received: 2021/13/07 Approved: 2022/10/11

Abstract. The most critical assumption in the autoregressive-moving average models (ARMA) is that innovations involved in the modeling process are normally distributed. Nonetheless, in some trials, we encounter instances where the normality assumption of the innovations is violated. In this article we were interested in model selection among the ARMA(1,1) models with innovations from the Weibull and exponential families. This may be beneficial for both forecasting and modeling perspectives. For this purpose, the innovation distribution parameters and the model coefficients were estimated by modified maximum likelihood (MML) method. Afterward, the asymptotic distribution of the acquired estimators in stationary mode was evaluated. Based on a simulation study, the ability of different model selection methods has been compared. Using the simulation, it was demonstrated that Vuong test correctly selected the optimal ARMA(1,1) model with non-negative innovations. Sometimes two competing models were picked as equivalent by the Vuong test. However, it did not indicate that they were close to the true model or away from it. Therefore, we used the Cox test to find out if the equivalent models are suitable or unsuitable. The obtained results of the tracking intervals has confirmed the results of the Vuong test. A real dataset was analyzed and modeled using the specific model selection approaches, and

[*] Corresponding author

the proper ARMA(1,1) model was selected from the competing ARMA(1,1) models for data fitting.

Keywords. ARMA(1,1), Cox test, MMLE, tracking interval, Vuong test.

1 Introduction

A common and widespread problem in statistics is determining a suitable probability distribution to describe or characterize an experimental dataset. Selecting the best probability distribution based on priori considerations is the main goal of the model selection. Hypothesis testing and model selection criteria are procedures for choosing an appropriate distribution based on observations. Kullback–Leibler divergence criterion (KL)(Kullback & Leibler, 1951) is perhaps the most prevalent information criterion used for assessing model discrepancy. In statistics, it is presented as the expected logarithm of the likelihood ratio. Assume that $\{Z_s\}_{s=1}^m$ are independent and identically distributed random variables (i.i.d) with common true probability density function (pdf) $q(.)$, and $\mathcal{F} = \big(f^{\omega}(.)\big)_{\omega\in\mathcal{W}}$ is a proposed model for approximation $q(.)$, then the KL divergence criterion between f^{ω} and q for random variable Z is defined as:

$$KL\{q, f^{\omega}\} = E_q\Big\{ \log\big(\frac{q(Z)}{f^{\omega}(Z)}\big)\Big\} = E_q\big\{ \log q(Z)\big\} - E_q\big\{ \log f^{\omega}(Z)\big\}.$$

E_q demonstrates the expectation with regard to the true distribution of Z. The closest member of $\mathcal{F}$ to $q(.)$ is $f^{\omega^*}(.)$ where $\omega^* \in \mathcal{W}$ is the minimizer of the KL criterion or maximizer of the $E_q\big\{ \log f^{\omega}(Z)\big\}$ equivalently. $E_q\big\{ \log f^{\omega^*}(Z)\big\}$ is the influential part of the KL criterion and has an estimator as follows:

$$\frac{1}{m}\sum_{s=1}^{m} \log f^{\hat{\omega}_m}(z_s),$$

where $\hat{\omega}_m$ is a consistent quasi-maximum likelihood estimator of the ω^* which is the pseudo-true value of ω (White, 1982). If $q(.) \in \mathcal{F}$, then there is a $\omega^0 \in \mathcal{W}$ for which $q(.) = f^{\omega^0}(.)$. In this case, the model $\mathcal{F}$ is well-specified

© 2020, SRTC Iran

and $\hat{\omega}_m \xrightarrow{p} \omega^0 (= \omega^*)$. Otherwise, if $q(.) \notin \mathcal{F}$ the model is classified as misspecified and $\hat{\omega}_m \xrightarrow{p} \omega^*$. It needs to be considered that $\frac{1}{m}\sum_{s=1}^{m} \log f^{\hat{\omega}_m}(z_s)$ as an estimator of expected log-likelihood function is biased. Akaike et al (1973) estimated the amount of bias and introduced Akaike information criterion (AIC) to approximate the KL divergence criterion. Competing models may be nested or non-nested, and in other cases they may overlap or not. The method of testing non-nested hypotheses was first examined by Cox (1961) and Atkinson (1970). Vuong (1989) introduced the non-nested model selection test for complete data. Cox and Vuong tests are a generalization of the likelihood ratio tests (LRT) for various contexts. Commenges et al (2008) quantified the tracking interval for estimating a difference of risks between two models.

When modeling of time series data is conducted, usually the normal distribution of the innovations generated by the process is supposed and order selections are noticed. In the case of the time series model linearity, and if the normal distribution is considered, the data distribution will be normal as well. However, in modeling with ARMA models, sometimes non-normality of residuals are noticed. This concept can be considered about stream-flow and hydrological datasets. They are prone to heavy tails and sometimes skewness (Braga & Calmon, 2017; Duca et al., 2019; Sarlak & Sorman, 2007). Also, the residual distribution of the stock price data modeling emerges as leptokurtic (Spierdijk, 2016; Trindade et al., 2010; Zhang & Li, 2019). Therefore, the use of the classical ARMA models requires some type of original series transformation into normal mode before applying the model (G. E. Box & Cox, 1964). Sometimes monotonous transformations do not fix asymmetry (Weiss, 1975). Hence, models with ARMA-type dependence structures, applicable to non-normal situations, are considered.

Grunwald et al (2000) and Bayer et al (2018) have discussed the time series ARMA modeling with innovations of the exponential family. Sim (1987) suggested a mixed Gamma model of the ARMA(1,1) process which can be used to simulate stationary river flow sequences with high skewness. The pure autoregressive process has been a solid base to model non-normal time series in most studies. However, a high order model can be resulted from modeling based on pure moving average (MA) or autoregressive (AR) processes. In general, the large number of the parameters both increase the rate of computations and reduce the efficiency of the estimation (Wei, 2006). Accordingly, to construct a non-normal model, it is beneficial to consider MA and AR sentences in one pattern. In this respect two main coins should

be pondered; 1) in modeling a non–normal time series by the MA and AR compilation processes, what the apposite innovation distribution is; 2) and how we should estimate the distribution parameters and model coefficients in non-normal mode.

we are interested in selecting the distribution of the innovations in the case that the modeling is performed with the ARMA(1,1) process. For this goal, it is quite essential to find solutions for the estimation problems under non-normality of the innovation distribution. In the ARMA(1,1) model, $x_s = \Phi x_{s-1} + e_s - \Theta e_{s-1}$, the coefficients Θ and Φ are unknown. For non-normal innovations, maximum likelihood estimators (MLE) cannot be acquired in some situations because of the complexity of the likelihood equations for which there is no explicit solution. MML estimation method can solve this problem (Tiku, 1967). This method is based on the complex phrase linearization in the log-likelihood function by Taylor series expansion. Zamani Mehreyan and Sayyareh (2017) expanded this method to estimate the AR(1) model parameters when the innovations are non-normal. Kasraie and Sayyareh (2020) implemented the modified method to approximate the ARMA(1,1) model parameters in non-normal situations. They introduced specific relationships for estimating the parameters of the ARMA(1,1) model when the innovations are originated from Weibull and exponential families.

Non-nested Cox and Vuong tests and the tracking interval are approaches that we use for choosing the optimal model among estimated rival models. In the Cox test, the null hypothesis is supposed to comprise the correct innovation model. The background concept for this test is to compare the difference between the log-likelihood ratio, under the null hypothesis, and its expected value. Vuong (1989) introduced a statistic to test the hypothesis that the two non-nested models are equally close to the true data generating model, as opposed to alternative of one model being closer. Accepting the null hypothesis means that two rival models are equivalent, but it does not indicate their vicinity to the true model. Therefore, Cox test is used to find out if the equivalent models are suitable or unsuitable (Sayyareh et al., 2010). Since precise AIC value does not have direct interpretation, Commenges et al (2008) proposed the tracking interval that is a standardized difference of estimated AICs in two different models to approximate the difference of expected KL risks. The existence of zero in the tracking interval infers that two rival models are equivalent, otherwise one of the models is selected as the optimal one.

The objectives of this research are, 1) to introduce the MMLEs to esti-

© 2020, SRTC Iran

mate the ARMA(1,1) model parameters briefly and to calculate the asymptotic distribution of the MMLEs when the innovations are originated from Weibull and exponential families, and 2) to conduct model selection among the estimated ARMA(1,1) processes with non-negative innovations and to compare the capabilities of different model selection methods by simulation.

2 Parameter Estimation

The preliminary stage of applying the model selection methods, including criteria and hypothesis testings, is to estimate the model parameters under investigation. However, in estimating the parameters of the ARMA(1,1) models by the ML method, even in the case of data distribution normality, the likelihood functions are complex and difficult to be solved. To facilitate the solution of the likelihood function, Kasraie and Sayyareh (2020) implemented a modified method briefly described here. The purpose of this section is to calculate the asymptotic distribution of the MMLEs used to estimate the ARMA(1,1) model parameters under different competitive models of the innovations.

Consider the ARMA(1,1) model in stationary mode depicted with Equation (1).

$$x_s = \Phi x_{s-1} + e_s - \Theta e_{s-1}, \quad s = 2, 3, ..., m \tag{1}$$

In this equation Θ and Φ, are moving average and autoregressive coefficients of the model such that $\mid \Theta \mid < 1$, $\mid \Phi \mid < 1$, and $\Phi \neq \Theta$. The model innovations, e_s's, are i.i.d of the Weibull or exponential family. If we consider ω as the vector of the innovation distribution parameters, the goal is to approximate the vector $\delta = (\omega, \Phi, \Theta)$.

Proposition 1. Assume that the innovation distribution is from the exponential family as follows:

$$f^{\eta}(e_s) = \exp\{\sum_{i=1}^{k} \eta_i C_i(e_s) - \mathcal{A}(\eta)\} h(e_s) = \exp\{\eta^T C(e_s) - \mathcal{A}(\eta)\} h(e_s), \quad s = 2, 3, ..., m.$$

$\eta = (\eta_1, \eta_2, ..., \eta_k)$ is natural parameter vector and $C(e_s) = (C_1(e_s), C_2(e_s), ..., C_k(e_s))$ points to the sufficient statistic. $\mathcal{A}(\eta)$ is considered to be cumulant function and $h(e_s)$ is a function of e_s. Under these assumptions, the non-

recursive MMLEs are defined as:

$$\hat{\Phi} = \frac{\sum_{s=2}^{m} \mathcal{M}_1(\bar{e}_c) + \mathcal{M}_1^{'}(\bar{e}_c)[x_s^* + \hat{\Theta} e_{s-1}^* - \bar{e}_c]}{\sum_{s=2}^{m} \mathcal{M}_1^{'}(\bar{e}_c) x_{s-1}^*}$$

$$\hat{\Theta} = - \frac{\sum_{s=2}^{m} \mathcal{M}_2(\bar{e}_c) + \mathcal{M}_2^{'}(\bar{e}_c)[x_s^* - \hat{\Phi} x_{s-1}^* - \bar{e}_c]}{\sum_{s=2}^{m} \mathcal{M}_2^{'}(\bar{e}_c) e_{s-1}^*}$$

$$\frac{\partial}{\partial \eta_i} \mathcal{A}(\eta) = \frac{\sum_{s=2}^{m} C_i(e_s)}{m-1}, \quad i = 1, 2, ..., k,$$

where

$$\mathcal{M}_1(e_s^*) = \eta^T \frac{\partial}{\partial \Phi} C(e_s^* + \mu_e) + \frac{\partial}{\partial \Phi} \log\ h(e_s^* + \mu_e),$$

$$\mathcal{M}_2(e_s^*) = \eta^T \frac{\partial}{\partial \Theta} C(e_s^* + \mu_e) + \frac{\partial}{\partial \Theta} \log\ h(e_s^* + \mu_e)$$

and

$$e_s^* = x_s^* - \Phi x_{s-1}^* + \Theta e_{s-1}^*, \quad s = 2, 3, ..., m.$$

e_s^* and x_s^* are the centralized forms of e_s and x_s around the mean, respectively, and $\bar{e}_c$ is the mean of e_s^*. The recursive MMLEs of Φ and Θ can also be obtained from Equations (2) and (3).

$$\mathcal{M}_1(e_s^*) \simeq \mathcal{M}_1(\bar{e}_c) + \mathcal{M}_1^{'}(\bar{e}_c)(e_s^* - \bar{e}_c) \tag{2}$$

$$\mathcal{M}_2(e_s^*) \simeq \mathcal{M}_2(\bar{e}_c) + \mathcal{M}_2^{'}(\bar{e}_c)(e_s^* - \bar{e}_c) \tag{3}$$

In which,

$$e_s^* = \sum_{i=0}^{s-2} \Theta^i [x_{s-i}^* - \Phi x_{s-i-1}^*], \quad s = 2, 3, ..., m. \tag{4}$$

Proof. Proof is given in Kasraie and Sayyareh (2020). □

The MMLEs are similarly determined in the Weibull family.

The estimators of the ARMA(1,1) model parameters are presented here under one innovation distribution of the exponential family based on the

© 2020, SRTC Iran

acquired theoretical results. Log-normal is a innovation distribution selected from the exponential family.

Example 1. Suppose that the innovations in model (1) have log-normal distribution, $\mathcal{LN}(\mathcal{M},\mathcal{S})$,

$$f_1^{\omega}(e_s)=\frac{1}{e_s\sqrt{2\pi\mathcal{S}^2}}\exp\{\frac{-1}{2S^2}(\log e_s-\mathcal{M})^2\},\quad e_s,\mathcal{S}>0,\quad \mathcal{M}\in R,\quad s=2,3,...,m.$$

This density function in the canonical form is as follows:

$$f_1^{\eta}(e_s)=h_1(e_s)\exp\{\eta_{11}\big(\log e_s\big)^2+\eta_{12}\log e_s-\mathcal{A}_1(\eta)\}.$$

In this format, $\eta_{11}=-1/2\mathcal{S}^2$, $\eta_{12}=\mathcal{M}/\mathcal{S}^2$, $h_1(e_s)=1/(e_s\sqrt{2\pi})$ and

$$\mathcal{A}_1(\eta)=\frac{\mathcal{M}^2}{2\mathcal{S}^2}+\frac{1}{2}\log\mathcal{S}^2=-\frac{\eta_{12}^2}{4\eta_{11}}+\frac{1}{2}\log(-\frac{1}{2\eta_{11}}).$$

For estimating the aforementioned model parameters, the MMLEs are obtained from proposition 2.1. Consequently, the non-recursive estimators are,

$$\hat{\Phi}_{m,1}=\frac{\Big((\bar{e}_c+\mu_e)(\hat{\mathcal{S}}_m^2-\hat{\mathcal{M}}_m+\log(\bar{e}_c+\mu_e)\Big)\sum_{s=2}^m x_{s-1}^*}{\Big(-\hat{\mathcal{S}}_m^2+\hat{\mathcal{M}}_m-\log(\bar{e}_c+\mu_e)+1\Big)\sum_{s=2}^m {x_{s-1}^*}^2}+\frac{\sum_{s=2}^m x_{s-1}^*(x_s^*+\hat{\Theta}_{m,1}e_{s-1}^*-\bar{e_c})}{\sum_{s=2}^m {x_{s-1}^*}^2},$$

$$\hat{\Theta}_{m,1}=-\Big[\frac{\Big((\bar{e}_c+\mu_e)(-\hat{\mathcal{S}}_m^2+\hat{\mathcal{M}}_m-\log(\bar{e}_c+\mu_e)\Big)\sum_{s=2}^m e_{s-1}^*}{(\hat{\mathcal{S}}_m^2-\hat{\mathcal{M}}_m+\log(\bar{e}_c+\mu_e)-1)\sum_{s=2}^m {e_{s-1}^*}^2}+\frac{\sum_{s=2}^m e_{s-1}^*(x_s^*-\hat{\Phi}_{m,1}x_{s-1}^*-\bar{e_c})}{\sum_{s=2}^m {e_{s-1}^*}^2}\Big],$$

$$\hat{\mathcal{S}}_m^2=\frac{\sum_{s=2}^m\big(\log(x_s-\hat{\Phi}_{m,1}x_{s-1}+\hat{\Theta}_{m,1}e_{s-1})-\hat{\mathcal{M}}_m\big)^2}{m-1}$$

and

$$\hat{M}_m=\frac{\sum_{s=2}^m\log(x_s-\hat{\Phi}_{m,1}x_{s-1}+\hat{\Theta}_{m,1}e_{s-1})}{m-1}$$

where $\mu_e=\exp\{\mathcal{M}+\frac{1}{2}\mathcal{S}^2\}$. In real circumstances, the non-recursive forms of the estimators are not efficient due to the existence of innovations in their structure. In these cases it is suggested to use recursive estimators which can only be approximated using time series values. If Equations (2) and (3) are resolved according to e_s^* recursive definition brought up in Equation (4),

then the recursive MMLEs are,

$$\hat{\Phi}_{m,1} = \frac{\Big((\bar{e}_c+\mu_e)(\hat{\mathcal{S}}_m^2 - \hat{\mathcal{M}}_m + \log(\bar{e}_c+\mu_e)\Big)\sum_{s=2}^{m}\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}x_{s-i-1}^{*}}{\Big(-\hat{\mathcal{S}}_m^2 + \hat{\mathcal{M}}_m - \log(\bar{e}_c+\mu_e)+1\Big)\sum_{s=2}^{m}(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}x_{s-i-1}^{*})^2} +$$

$$\frac{\sum_{s=2}^{m}\Big[(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}x_{s-i-1}^{*})(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}x_{s-i}^{*}-\bar{e}_c)\Big]}{\sum_{s=2}^{m}(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}x_{s-i-1}^{*})^2},$$

$$\sum_{s=2}^{m}\Bigg[\Bigg((\bar{e}_c+\mu_e)(-1-\frac{1}{\hat{\mathcal{S}}_m^2}\log(\bar{e}_c+\mu_e)+\frac{\hat{\mathcal{M}}_m}{\hat{\mathcal{S}}_m^2}\Bigg)\sum_{i=0}^{s-2}i\Theta^{i-1}(x_{s-i}^{*}-\hat{\Phi}_{m,1}x_{s-i-1}^{*})-$$

$$(-1-\frac{1}{\hat{\mathcal{S}}_m^2}\log(\bar{e}_c+\mu_e)+\frac{\hat{\mathcal{M}}_m}{\hat{\mathcal{S}}_m^2}+\frac{1}{\hat{\mathcal{S}}_m^2})\Bigg(\sum_{i=0}^{s-2}i\Theta^{i-1}(x_{s-i}^{*}-\hat{\Phi}_{m,1}x_{s-i-1}^{*})\Bigg)$$

$$\times\Bigg(\sum_{i=0}^{s-2}\Theta^{i}(x_{s-i}^{*}-\hat{\Phi}_{m,1}x_{s-i-1}^{*})-\bar{e}_c\Bigg)\Bigg]=0,$$

$$\hat{\mathcal{S}}_m^2 = \frac{\sum_{s=2}^{m}\big[\log(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}(x_{s-i}-\hat{\Phi}_{m,1}x_{s-i-1})+\hat{\Theta}_{m,1}^{s-1}\mu_e)-\hat{\mathcal{M}}_m\big]^2}{m-1}$$

and

$$\hat{\mathcal{M}}_m = \frac{\sum_{s=2}^{m}\log\big(\sum_{i=0}^{s-2}\hat{\Theta}_{m,1}^{i}(x_{s-i}-\hat{\Phi}_{m,1}x_{s-i-1})+\hat{\Theta}_{m,1}^{s-1}\mu_e\big)}{m-1}.$$

In the following, the asymptotic distribution of the MMLEs is discussed. Before stating the main theorem, a theorem without proof is given.

***Theorem* 1** (Hoffding and Robbins)**.** *suppose that* $\{X_s\}_{s=1}^{m}$ *is a stationary and n-dependent sequence where* $E(X)=0$ *and* $Cov(X_s,X_{s+h})=E(X_sX_{s+h})=:\Gamma(h)$. *Also suppose* $\nu_n := \Gamma(0)+2\sum_{j=1}^{n}\Gamma(j)$. *Then*

$$\frac{1}{\sqrt{m}}\sum_{s=1}^{m}X_s \xrightarrow{d} N(0,\nu_n),$$

and

$$mVar(\bar{X}_m)\xrightarrow{p}\nu_n,$$

where $\bar{X}_m = \dfrac{\sum_{s=1}^{m}X_s}{m}$

Proof. Proof is given in Resnick (2013). □

***Theorem* 2** (Asymptotic distribution)**.** *Suppose* $\{X_s\}_{s=1}^{m}$ *is a stationary*

© 2020, SRTC Iran

time series that has been modeled using the ARMA(1,1) process, and the $\boldsymbol{e}_s$*'s are i.i.d random variables generated by q(.). If* δ^* *is the pseudo-true value of* $\delta = (\omega, \Phi, \Theta)$ *under the proposed innovation model f(.), the asymptotic distribution of* $\hat{\delta}_m$ *is as follows:*

$$(\hat{\delta}_m - \delta^*) \xrightarrow{d} N(o, C),$$

where

$$C = I_f^{-1}(\delta^*) J_f(\delta^*) I_f^{-1}(\delta^*), J_f(\delta) = E_q\left[\left(\frac{\partial \ell_m^f(\delta)}{\partial \delta}\right)\left(\frac{\partial \ell_m^f(\delta)}{\partial \delta^T}\right)\right], I_f(\delta) = -E_q\left[\frac{\partial^2 \ell_m^f(\delta)}{\partial \delta \partial \delta^T}\right].$$

Furthermore, the asymptotic distributions of the model coefficient estimators, assuming that the other model parameters are known, are

$$\sqrt{m-1}(\hat{\Phi}_m - \Phi) \xrightarrow{d} N(0, \sigma^2_{e_s^*} \Gamma^{-1}(0))$$

and

$$\sqrt{m-1}(\hat{\Theta}_m - \Theta) \xrightarrow{d} N(0, 1)$$

where $\sigma^2_{e_s^*} = Var(\boldsymbol{e}_s^*)$ *and* $\Gamma(h) = Cov(X_s^*, X_{s+h}^*)$.

Proof. We start the proof by obtaining the asymptotic distribution of $(\hat{\delta}_m - \delta^*)$ under the assumption that the regularity conditions are established. Using the Taylor expansion of $(\partial/\partial\delta)\sum_{s=2}^m \log f^{\hat{\delta}_m}(e_s) = (\partial/\partial\delta)\ell_m^f(\hat{\delta}_m)$ around δ^*, we have

$$\frac{\partial}{\partial \delta}\ell_m^f(\hat{\delta}_m) = \frac{\partial}{\partial \delta}\ell_m^f(\delta) \mid_{\delta^*} + \frac{\partial^2}{\partial \delta \partial \delta^T}\ell_m^f(\delta)(\hat{\delta}_m - \delta) \mid_{\delta^*} + o_p(1).$$

Since $\frac{\partial}{\partial \delta}\ell_m^f(\hat{\delta}_m) = 0$, the equation can be rewritten as below:

$$(\hat{\delta}_m - \delta^*) = -\frac{\frac{\partial}{\partial \delta}\ell_m^f(\delta) \mid_{\delta^*}}{\frac{\partial^2}{\partial \delta \partial \delta^T}\ell_m^f(\delta) \mid_{\delta^*}} + o_p(1).$$

In this equation $-\frac{\partial^2}{\partial \delta \partial \delta^T}\ell_m^f(\delta) \mid_{\delta^*} \xrightarrow{p} I_f(\delta^*)$. Also, using the multivariate

central limit theorem (CLT) $\frac{\partial}{\partial\delta}\ell_m^f(\delta)\mid_{\delta^*}\xrightarrow{d} N(0, J_f(\delta^*))$. Thus, we conclude that

$$(\hat{\delta}_m - \delta^*) \xrightarrow{d} N_p(0, I_f^{-1}(\delta^*)J_f(\delta^*)I_f^{-1}(\delta^*)).$$

In other words,

$$(\hat{\delta}_m - \delta^*) \xrightarrow{d} N_p(0, C).$$

Then, we get asymptotic distribution of the MMLEs used to estimate the ARMA(1,1) model coefficients. All Φ estimators obtained under different distributions can be summarized as follows:

$$\hat{\Phi}_{m,i} = \frac{A_i \sum_{s=2}^m x_{s-1}^*}{\sum_{s=2}^m {x_{s-1}^*}^2} + \frac{\sum_{s=2}^m x_{s-1}^*(x_s^* + \Theta e_{s-1}^* - \bar{e}_c)}{\sum_{s=2}^m {x_{s-1}^*}^2}, \qquad i = 1, 2, 3, ... \quad (5)$$

The first term of Equation (5) converges to zero in probability. This conclusion is due to the fact that the ARMA(1,1) process, with regard to the stationary condition and considering $\sum_{h=0}^{\infty}|\Gamma(h)| < \infty$, is ergodic for the mean. A process is called ergodic for the mean if the time series average tends to the population mean by increasing the sample size. Therefore, as $n \to \infty$, $\sum_{s=2}^m (x_{s-1} - \mu_x) = o_p(1)$, $\sum_{s=2}^m (x_{s-1} - \mu_x)^2 = O_p(1)$ and also by weak law of large numbers $\bar{e}_c \xrightarrow{p} 0$. Consequently, the equivalent relationship for (5) is

$$\hat{\Phi}_{m,i} \simeq \frac{\sum_{s=2}^m x_{s-1}^*(x_s^* + \Theta e_{s-1}^*)}{\sum_{s=2}^m {x_{s-1}^*}^2}, \qquad i = 1, 2, 3, ... \quad (6)$$

By rewriting the equation $e_s^* = x_s^* - \Phi x_{s-1}^* + \Theta e_{s-1}^*$ as $x_s^* + \Theta e_{s-1}^* = \Phi x_{s-1}^* + e_s^*$ and replacing it in Equation (6), we have

$$\hat{\Phi}_{m,i} \simeq \frac{\sum_{s=2}^m x_{s-1}^*(\Phi x_{s-1}^* + e_s^*)}{\sum_{s=2}^m {x_{s-1}^*}^2} = \Phi + \frac{\sum_{s=2}^m x_{s-1}^* e_s^*}{\sum_{s=2}^m {x_{s-1}^*}^2}.$$

Then

$$\sqrt{m-1}(\hat{\Phi}_{m,i} - \Phi) = \frac{(m-1)^{-\frac{1}{2}} \sum_{s=2}^m x_{s-1}^* e_s^*}{(m-1)^{-1} \sum_{s=2}^m {x_{s-1}^*}^2} = \frac{(m-1)^{-\frac{1}{2}} \sum_{s=2}^m u_s}{(m-1)^{-1} \sum_{s=2}^m {x_{s-1}^*}^2},$$

where $u_s = x_{s-1}^* e_s^*$. Since $\boldsymbol{e}_s^*$ and X_{s-1}^* are independent, $E(U_s) = E(X_{s-1}^*)E(\boldsymbol{e}_s^*)$

© 2020, SRTC Iran

$= 0$. Also,

$$E(U_s^2) = E(X_{s-1}^* \boldsymbol{e}_s^* X_{s-1}^* \boldsymbol{e}_s^*) = E({X_{s-1}^*}^2 \boldsymbol{e}_s^{*2}) = E({X_{s-1}^*}^2) E(\boldsymbol{e}_s^{*2}) = \Gamma(0)\sigma_{e_s^*}^2,$$

and for $h > 0$

$$E(U_{s+h}U_s) = E(X_{s+h-1}^* \boldsymbol{e}_{s+h}^* X_{s-1}^* \boldsymbol{e}_s^*) = E(X_{s+h-1}^* X_{s-1}^* \boldsymbol{e}_s^*) E(\boldsymbol{e}_{s+h}^*) = 0.$$

A similar result is obtained for $h < 0$. Now, theorem 2.1 can be used to say

$$(m-1)^{-\frac{1}{2}} \sum_{s=2}^{m} u_s \xrightarrow{d} N(0, \Gamma(0)\sigma_{e_s^*}^2).$$

On the other hand, in linear models

$$(m-1)^{-1} \sum_{s=2}^{m} {x_{s-1}^*}^2 \xrightarrow{p} \Gamma(0)$$

(Shumway & Stoffer, 2017). Thus,

$$\sqrt{m-1}(\hat{\Phi}_{m,i} - \Phi) \xrightarrow{d} N(0, \sigma_{e_s^*}^2 \Gamma(0)^{-1}\Gamma(0)\Gamma(0)^{-1}) = N(0, \sigma_{e_s^*}^2 \Gamma(0)^{-1}).$$

All the estimators obtained for Θ can be summerized according to the following formula:

$$\hat{\Theta}_{m,i} = \frac{B_i \sum_{s=2}^{m} e_{s-1}^*}{\sum_{s=2}^{m} {e_{s-1}^*}^2} - \frac{\sum_{s=2}^{m} e_{s-1}^* (x_s^* - \Phi x_{s-1}^* - \bar{e_c})}{\sum_{s=2}^{m} {e_{s-1}^*}^2}; \qquad i = 1, 2, 3, ... \quad (7)$$

The first term of the Equation (7) is convergent to zero in probability, because $\sum_{s=2}^{m} e_{s-1}^* = o_p(1)$ and $\sum_{s=2}^{m} {e_{s-1}^*}^2 = O_p(1)$. Furthermore, from the displacement of the relation $x_s^* = \Phi x_{s-1}^* + e_s^* - \Theta e_{s-1}^*$, relation $x_s^* - \Phi x_{s-1}^* = e_s^* - \Theta e_{s-1}^*$ is derived. Then

$$\hat{\Theta}_{m,i} \simeq \frac{\sum_{s=2}^{m} e_{s-1}^* (x_s^* - \Phi x_{s-1}^*)}{\sum_{s=2}^{m} {e_{s-1}^*}^2} = \frac{\sum_{s=2}^{m} e_{s-1}^* (e_s^* - \Theta e_{s-1}^*)}{\sum_{s=2}^{m} {e_{s-1}^*}^2}$$

and

$$\sqrt{m-1}(\hat{\Theta}_{m,i} - \Theta) = \frac{(m-1)^{-\frac{1}{2}} \sum_{s=2}^{m} e_{s-1}^* e_s^*}{(m-1)^{-1} \sum_{s=2}^{m} {e_{s-1}^*}^2} = \frac{(m-1)^{-\frac{1}{2}} \sum_{s=2}^{m} v_s}{(m-1)^{-1} \sum_{s=2}^{m} {e_{s-1}^*}^2},$$

where $v_s = e^*_{s-1}e^*_s$. $\boldsymbol{e}^*_s$'s are i.i.d random variables, so we have got

$$E(V_s) = E(\boldsymbol{e}^*_{s-1}\boldsymbol{e}^*_s) = E(\boldsymbol{e}^*_{s-1})E(\boldsymbol{e}^*_s) = 0$$

and

$$E(V_s^2) = E({\boldsymbol{e}^*_{s-1}}^2{\boldsymbol{e}^*_s}^2) = E({\boldsymbol{e}^*_{s-1}}^2)E({\boldsymbol{e}^*_s}^2) = (\sigma^2_{\boldsymbol{e}^*_s})^2.$$

Also for $h > 0$

$$E(V_{s+h}V_s) = E(\boldsymbol{e}^*_{s+h-1}\boldsymbol{e}^*_{s+h}\boldsymbol{e}^*_{s-1}\boldsymbol{e}^*_s) = 0.$$

If there is $h = 1$, then

$$E(V_{s+1}V_s) = E(\boldsymbol{e}^*_s\boldsymbol{e}^*_{s+1}\boldsymbol{e}^*_{s-1}\boldsymbol{e}^*_s) = E(\boldsymbol{e}^*_{s-1}{\boldsymbol{e}^*_s}^2\boldsymbol{e}^*_{s+1}) = E(\boldsymbol{e}^*_{s-1})E({\boldsymbol{e}^*_s}^2)E(\boldsymbol{e}^*_{s+1}) = 0.$$

These results are also established for $h < 0$. According to the CLT

$$(m-1)^{-\frac{1}{2}}\sum_{s=2}^{m} v_s \xrightarrow{d} N(0, (\sigma^2_{e^*_s})^2),$$

and by weak law of large numbers $(m-1)^{-1}\sum_{s=2}^{m} {e^*_{s-1}}^2 \xrightarrow{p} \sigma^2_{e^*_s}$. Therefore,

$$\sqrt{m-1}(\hat{\Theta}_{m,i} - \Theta) \xrightarrow{d} N(0, \sigma^{-2}_{e^*_s}\sigma^4_{e^*_s}\sigma^{-2}_{e^*_s}) = N(0,1),$$

and the proof is complete. □

3 Model Selection Tests and Tracking Interval

The competing models from which a model is being selected are the ARMA(1,1) models with some non-negative innovations of the Weibull and exponential families. Consider the following models:

$$\mathcal{F} = \left\{f^{\delta}(X_s|x_1, e_1) : \delta \in \Delta \subseteq \mathbb{R}^p\right\} = \left(f^{\delta}(X_s|x_1, e_1)\right)_{\delta\in\Delta} \tag{8}$$

and

$$\mathcal{G} = \left\{g^{\vartheta}(X_s|x_1, e_1) : \vartheta \in V \subseteq \mathbb{R}^q\right\} = \left(g^{\vartheta}(X_s|x_1, e_1)\right)_{\vartheta\in V} \tag{9}$$

as two suggested competing parametric families of densities for the $q(.)$ unknown true innovation density. $\mathcal{F}$ and $\mathcal{G}$, are nested if one model can be

© 2020, SRTC Iran

acquired by applying constraints from the other model. In other words, $\mathcal{F}$ is nested in $\mathcal{G}$ if $\mathcal{F} \subset \mathcal{G}$. Furthermore, the two models are non-nested iff $\mathcal{F} \cap \mathcal{G} = \emptyset$. In the next case, they may be partial nested (overlapping) models iff $\mathcal{F} \cap \mathcal{G} \neq \emptyset$, $\mathcal{F} \nsubseteq \mathcal{G}$ and $\mathcal{G} \nsubseteq \mathcal{F}$. In the classical time series ARMA modeling and when the innovation distribution is normal, all competing models are nested and the only consideration is to select the ideal order of the ARMA model. Whereas, non-nested innovation models lead to non-nested time series models. In this section some of model selection methods such as Cox and Vuong tests and tracking interval are improved to choose the most proper distribution of the non-negative innovations in the ARMA(1,1) model.

The Vuong test is a model selection test based on the difference between the KL criteria of the two competing models and inspired by the LRT. In this test, the two models may be nested, non-nested or partially nested relative to each other, and one or both of the distribution families may include the correct model. Also, it is possible that neither of the two distribution families contains the correct model. Considering $\mathcal{F}$ and $\mathcal{G}$ in Equations (8) and (9) as two competing models, Vuong test hypotheses are

$$H_0 : E_q\left\{ \log \frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)} \right\} = 0,$$

$$H_g : E_q\left\{ \log \frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)} \right\} < 0$$

or

$$H_f : E_q\left\{ \log \frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)} \right\} > 0.$$

H_0 hypothesis means that the two competing models, $f^{\delta}(X_s|x_1, e_1)$ and $g^{\vartheta}(X_s|x_1, e_1)$, are equivalent in proximity to the true model. Despite H_0, H_g implies that model $g^{\vartheta}(X_s|x_1, e_1)$ is better than model $f^{\delta}(X_s|x_1, e_1)$ and H_f implies that model $f^{\delta}(X_s|x_1, e_1)$ is better than model $g^{\vartheta}(X_s|x_1, e_1)$ for data fitting. The Vuong test statistic, when competing models are nested or non-nested, is made on the basis of the likelihood ratio (LR) statistic defined as:

$$\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m) = \sum_{s=2}^{m} \log \frac{f^{\hat{\delta}_m}(x_s|x_{s-1}, ..., x_1, e_1)}{g^{\hat{\vartheta}_m}(x_s|x_{s-1}, ..., x_1, e_1)} = \sum_{s=2}^{m} \log \frac{f^{\hat{\delta}_m}(e_s)}{g^{\hat{\vartheta}_m}(e_s)}.$$

In this section, the asymptotic distribution of the LR statistic is obtained un-

der the condition that the innovations competing models of the ARMA(1,1) process are non-nested.

***Theorem* 3** (Asymptotic distribution of th LR statistic)**.** *Under the assumptions of the Vuong (1989) and stationarity of the process, if $f^{\delta^*}(e_s) \neq g^{\vartheta^*}(e_s)$ then*

$$(m-1)^{-1/2}\mathrm{LR}_n(\hat{\delta}_m, \hat{\vartheta}_m) - (m-1)^{1/2} E_q\left[\log \frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)}\right] \xrightarrow{d} N(0, W_*^2),$$

where $W_*^2 = Var_q\left[\log \frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)}\right]$. E_q and Var_q demonstrate the expectation and variance regarding the true model.

Proof. Proof is provided in Appendix A. □

Using theorem (3.1), in the case of the non-nested innovations competing models, the standard form of the Vuong test statistic under H_0 is

$$V_m = \frac{(m-1)^{-1/2}\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m)}{\hat{W}_m},$$

where

$$\hat{W}_m = \frac{1}{m-1}\sum_{s=2}^{m}\left[\log \frac{f^{\hat{\delta}_m}(e_s)}{g^{\hat{\vartheta}_m}(e_s)}\right]^2 - \left[\frac{1}{m-1}\sum_{s=2}^{m}\log \frac{f^{\hat{\delta}_m}(e_s)}{g^{\hat{\vartheta}_m}(e_s)}\right]^2.$$

V_m converges asymptotically to the standard normal distribution, $N(0,1)$. Given the significance level of α, if $V_m > Z_\alpha$ then H_0 will be rejected against H_f and $f^\delta(e_s)$ will be selected as an optimal model. $V_m < -Z_\alpha$ results in rejection of H_0 against H_g. Also, $|V_m| < Z_\alpha$ indicates the equivalence of the two models. In other words, the two competing models are identical in terms of proximity to the true model.

The Cox test is a generalization of the likelihood ratio test for separate hypotheses, or in other words for testing non-nested models. The statistic of this test, which is the standardized form of the log-likelihood ratio statistic under hypothesis $H_0^f : q = f^\delta$ vs. $H_1^g : q = g^\vartheta$, is as follows:

$$T_f = \frac{\{\ell_m^f(\hat{\delta}_m) - \ell_m^g(\hat{\vartheta}_m)\} - E_f\{\ell_m^f(\hat{\delta}_m) - \ell_m^g(\hat{\vartheta}_m)\}}{\sqrt{V_{fg}}}. \tag{10}$$

© 2020, SRTC Iran

V_{fg} is the variance of $T_{fg} = \{\ell_m^f(\hat{\delta}_m) - \ell_m^g(\hat{\vartheta}_m)\} - E_f\{\ell_m^f(\hat{\delta}_m) - \ell_m^g(\hat{\vartheta}_m)\}$ and is equal with

$$V_{fg} = m\big\{V_\delta(F-G) - \frac{C_\delta^2(F-G, F_\delta)}{V_\delta(F_\delta)}\big\},$$

where $F = \log f^\delta(e_s), G = \log g^{\vartheta_\delta}(e_s), F_\delta = \dfrac{\partial \log f^\delta(e_s)}{\partial \delta}$. Also, V_δ and C_δ represent the variance and covariance under $f^\delta(e_s)$. Cox (1962) proved that this statistic has a standard normal asymptotic distribution. According to Cox's idea, the role of null and alternative hypotheses should be changed because there is not any reason to place either of the competing models in the null hypothesis. Therefore, after conducting the test $H_0^f : q = f^\delta$ vs. $H_1^g : q = g^\vartheta$, the test $H_0^g : q = g^\vartheta$ vs. $H_1^f : q = f^\delta$ should also be considered with the test statistic T_g.

$$T_g = \frac{\{\ell_m^g(\hat{\vartheta}_m) - \ell_m^f(\hat{\delta}_m)\} - E_g\{\ell_m^g(\hat{\vartheta}_m) - \ell_m^f(\hat{\delta}_m)\}}{\sqrt{V_{gf}}},$$

where V_{gf} is the variance of $T_{gf} = \{\ell_m^g(\hat{\vartheta}_m) - \ell_m^f(\hat{\delta}_m)\} - E_g\{\ell_m^g(\hat{\vartheta}_m) - \ell_m^f(\hat{\delta}_m)\}$. According to these two statistics, T_f and T_g, the Cox test has four rejection-acceptance regions as follows:

(i) If $|T_f| > C_\alpha$ and $|T_g| > C_\alpha$, both H_0^f and H_1^g are rejected,
(ii) If $|T_f| < C_\alpha$ and $|T_g| < C_\alpha$, neither of H_0^f and H_1^g are rejected,
(iii) If $|T_f| > C_\alpha$ and $|T_g| < C_\alpha$, H_0^f is rejected but there is not any reason to reject H_1^g,
(iv) If $|T_f| < C_\alpha$ and $|T_g| > C_\alpha$, H_1^g is rejected but there is not any reason to reject H_0^f.

In these rejection-acceptance regions, C_α refers to the critical value of the standard normal distribution at the significant level of α. The quantity ϑ_δ and similar quantities that are based on δ and ϑ, play an essential role in the Cox test. If H_0^f is correct, the accepted model parameter is δ. Therefore, ϑ_δ is the quantity that, under correctness of H_0^f, the $\hat{\vartheta}_m$ estimator converges to it in probability. However, estimating the parameters of one competitive model under the correctness of the other model complicates the calculation of the Cox test statistic and makes it difficult to compare the different probability functions using this test.

Hashem Pesaran and Pesaran (1993) proposed a stochastic simulation approach, also known as the Monte Carlo method, for computing the Cox test statistic. To describe this method in our work, we define the standardized Cox statistic for testing H_0^f against H_1^g as

$$N_f = \frac{\sqrt{(m-1)}N_{fg}}{\sqrt{V_f}},$$

where

$$N_{fg} = \bar{d} - E_f(\bar{d}), \qquad \bar{d} = \frac{1}{m-1}\sum_{s=2}^{m} d_{s-1},$$

$$\begin{aligned} d_{s-1} &= \log f^{\hat{\delta}_m}(x_s|x_{s-1},..,x_1,e_1) - \log g^{\hat{\vartheta}_m}(x_s|x_{s-1},..,x_1,e_1) \\ &= \log f^{\hat{\delta}_m}(e_s) - \log g^{\hat{\vartheta}_m}(e_s), \quad s = 2,3,...,m. \end{aligned}$$

V_f is the variance of $\sqrt{(m-1)}N_{fg}$ and $\hat{\delta}_m$ and $\hat{\vartheta}_m$ are the MLEs of δ and ϑ under H_0^f and H_1^g respectively. The essential issue in calculating the N_f statistic is computing the $\hat{\vartheta}_m$ under H_0^f. A simple method of computing $\hat{\vartheta}_m$ under H_0^f would be to compute the MLE of ϑ using the e_s observations that are artificially simulated assuming $f^{\hat{\delta}_m}(e_s)$ is the data generation process. If we consider $\boldsymbol{e}_i = (e_{1i}, e_{2i}, ..., e_{mi})$ as a vector of innovations generated artificially under H_0^f with $\delta = \hat{\delta}_m$, then $\hat{\vartheta}_i$ is the MLE of ϑ under H_1^g which is calculated using $\boldsymbol{e}_i$. Therefore, the estimator of ϑ under H_0^f is

$$\hat{\vartheta}_m(K) = \frac{1}{K}\sum_{i=1}^{K}\hat{\vartheta}_i,$$

where K is the number of replications. Also, $\hat{\vartheta}_i$ is a function of $\boldsymbol{e}_i$ and thus $\hat{\delta}_m$, and $\hat{\vartheta}_m(K) \xrightarrow{p} \vartheta$ under H_0^f. Now, N_{fg} can be consistently estimated by

$$\hat{N}_{fg} = \frac{1}{(m-1)}\big(\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(e_s) - \sum_{s=2}^{m}\log g^{\hat{\vartheta}_m}(e_s)\big) - C_K(\hat{\delta}_m, \hat{\vartheta}_m(K)),$$

© 2020, SRTC Iran

where

$$C_K(\hat{\delta}_m, \hat{\vartheta}_m(K)) = \frac{1}{(m-1)K} \sum_{i=1}^{k} \big(\sum_{s=2}^{m} \log f^{\hat{\delta}_m}(e_{si}) - \sum_{s=2}^{m} \log g^{\hat{\vartheta}_m(K)}(e_{ti}) \big).$$

V_f can also be estimated by

$$\hat{V}_f = (m-1)^{-1} d^T \big\{ I_{m-1} - R(\hat{\delta_m}) \big[R^T(\hat{\delta_m}) R(\hat{\delta_m}) \big]^{-1} R^T(\hat{\delta_m}) \big\} d,$$

where $d^T = (d_1, d_2, ..., d_{m-1})$ and $R(\delta)$ is the $(m-1) \times (p+1)$ matrix as

$$R(\delta) = \begin{bmatrix} 1 & \partial \log f^{\delta}(e_2)/\partial\delta_1 & . & . & . & \partial \log f^{\delta}(e_2)/\partial\delta_p \\ 1 & \partial \log f^{\delta}(e_3)/\partial\delta_1 & . & . & . & \partial \log f^{\delta}(e_3)/\partial\delta_p \\ . & . & & & & . \\ . & . & & & & . \\ . & . & & & & . \\ 1 & \partial \log f^{\delta}(e_m)/\partial\delta_1 & . & . & . & \partial \log f^{\delta}(e_m)/\partial\delta_p \end{bmatrix}.$$

Under H_0^f, the N_f statistic,

$$N_f = \frac{\sqrt{(m-1)}\hat{N}_{fg}}{\sqrt{\hat{V}_f}},$$

is asymptotically distributed as standard normal.

As another method for testing mis-specified non-nested competitive models, we investigate a tracking interval. Since the AIC value does not have a specific interpretation, and nonetheless, can be utilized in comparison with another value for model selection, Commenges et al (2008) proposed the standardized form of the difference between the two AICs as an estimate of the difference between the expected KL risks.

The KL divergence criteria for competing models q and f, and q and g are

$$KL\{q, f^{\delta}\} = E_q\big\{ \log q(X_s|x_1, e_1) \big\} - E_q\big\{ \log f^{\delta}(X_s|x_1, e_1) \big\},$$

and

$$KL\{q, g^{\vartheta}\} = E_q\big\{ \log q(X_s|x_1, e_1) \big\} - E_q\big\{ \log g^{\vartheta}(X_s|x_1, e_1) \big\}.$$

In each of the presented relations, the first sentence of the KL criterion depends only on the true distribution and is therefore a constant value. The difference between the two KL criteria is defined as follows:

$$\begin{aligned}
&\nabla KL(f^{\delta}(X_s|x_1,e_1), g^{\vartheta}(X_s|x_1,e_1))\\
&= KL(q(X_s|x_1,e_1), f^{\delta}(X_s|x_1,e_1)) - KL(q(X_s|x_1,e_1), g^{\vartheta}(X_s|x_1,e_1))\\
&= E_q\big\{\log g^{\vartheta}(X_s|x_1,e_1)\big\} - E_q\big\{\log f^{\delta}(X_s|x_1,e_1)\big\}.
\end{aligned}$$

Therefore, the estimator of the difference between the two KL divergence criteria can be defined as:

$$\begin{aligned}
&D(f^{\hat{\delta}_m}(X_s|x_1,e_1), g^{\hat{\vartheta}_m}(X_s|x_1,e_1)) = \frac{1}{2(m-1)}\big[AIC(f_m^{\hat{\delta}}(X_s|x_1,e_1)) - AIC(g_m^{\hat{\vartheta}}(X_s|x_1,e_1))\big]\\
&= -\frac{1}{m-1}\big[\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(x_s|x_{s-1},...,x_1,e_1) - \sum_{s=2}^{m}\log g^{\hat{\vartheta}_m}(x_s|x_{s-1},...,x_1,e_1) - (p-q)\big],\\
&= -\frac{1}{m-1}\big[\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(e_s) - \sum_{s=2}^{m}\log g^{\hat{\vartheta}_m}(e_s) - (p-q)\big]
\end{aligned}$$

In which,

$$\begin{aligned}
AIC(f_m^{\hat{\delta}}(X_s|x_1,e_1)) &= -2\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(x_s|x_{s-1},...,x_1,e_1) + 2p\\
&= -2\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(e_s) + 2p\\
&= -2\sum_{s=2}^{m}\log f^{\hat{\delta}_m}(\sum_{i=0}^{s-2}\Theta^i(x_{s-i} - \Phi x_{s-i-1}) + \Theta^{s-1}e_1) + 2p.
\end{aligned}$$

In a similar way,

$$AIC(g_n^{\hat{\vartheta}}(X_s|x_1,e_1)) = -2\sum_{s=2}^{m}\log g^{\hat{\vartheta}_m}(\sum_{i=0}^{s-2}\Theta^i(x_{s-i} - \Phi x_{s-i-1}) + \Theta^{s-1}e_1) + 2q.$$

Therefore, the tracking interval is defined as follows:

$$D(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) - Z_{\frac{\alpha}{2}}(m-1)^{\frac{-1}{2}}\hat{W}_m < \nabla KL(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) < D(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) + Z_{\frac{\alpha}{2}}(m-1)^{\frac{-1}{2}}\hat{W}_m,$$

© 2020, SRTC Iran

in which,

$$\hat{W}_m^2 = \frac{1}{m-1}\sum_{s=2}^{m}\left[\log\frac{f^{\hat{\boldsymbol{\delta}}_n}(x_s|x_{s-1},...,x_1,e_1)}{g^{\hat{\boldsymbol{\vartheta}}_m}(x_s|x_{s-1},...,x_1,e_1)}\right]^2 - \left[\frac{1}{m-1}\sum_{s=2}^{m}\log\frac{f^{\hat{\boldsymbol{\delta}}_m}(x_s|x_{s-1},...,x_1,e_1)}{g^{\hat{\boldsymbol{\vartheta}}_m}(x_s|x_{s-1},...,x_1,e_1)}\right]^2$$
$$= \frac{1}{m-1}\sum_{s=2}^{m}\left[\log\frac{f^{\hat{\boldsymbol{\delta}}_m}(e_s)}{g^{\hat{\boldsymbol{\vartheta}}_m}(e_s)}\right]^2 - \left[\frac{1}{m-1}\sum_{s=2}^{m}\log\frac{f^{\hat{\boldsymbol{\delta}}_m}(e_s)}{g^{\hat{\boldsymbol{\vartheta}}_m}(e_s)}\right]^2,$$

and

$$P_q\Big(A_m < \nabla KL(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) < B_m\Big) \longrightarrow 1-\alpha,$$

that

$$A_m = D(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) - Z_{\frac{\alpha}{2}}(m-1)^{\frac{-1}{2}}\hat{W}_m,$$

and

$$B_m = D(f^{\hat{\delta}_m}, g^{\hat{\vartheta}_m}) + Z_{\frac{\alpha}{2}}(m-1)^{\frac{-1}{2}}\hat{W}_m,$$
$$1-\Phi(Z_{\frac{\alpha}{2}}) = \frac{\alpha}{2}.$$

Φ is the standard normal distribution function. If the estimated tracking interval includes both positive and negative values, two competing models will be selected as equivalents. Otherwise, the tracking interval with positive values confirms the selection of the model g and the negative tracking interval confirms the selection of the model f.

4 Numerical Experiments

To analyze the obtained results from theoretical discussions, in this section, a simulation has been used. Then, the optimal ARMA(1,1) model has been selected from among the competitive models using the model selection tests. This study shows that the Cox and Vuong tests and the tracking interval are appropriate methods for selecting the optimal ARMA(1,1) model with non-negative innovations when the model parameters are estimated by the MMLEs.

At first, it is assumed that the data are generated from the ARMA(1,1) model with Gamma innovation distribution, $\mathcal{G}(1.5, 2)$. The ARMA(1,1) models with Gamma, Weibull, Rayleigh, inverse Gaussian, and log-normal innovation distributions are considered as competing models. Using the MMLEs and the dataset, the values of the Vuong test statistic for competing models are calculated, and the results are summarized in Table 1.

From the results presented in Table 1 it is concluded that if the ARMA(1,1) models with Gamma and Weibull innovations are considered as two compet-

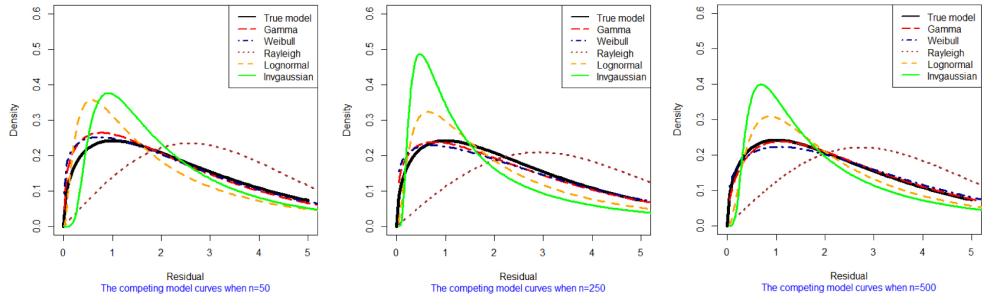

Figure 1. The true model diagram and the estimated competing models with Gamma, Weibull, Rayleigh, log-normal, inverse Gaussian innovations

ing models, in most cases, the two models will be selected as equivalent models. The reason is that in these cases the values of the Vuong test statistics are less than the critical value of the standard normal distribution, 1.96. In other cases, where one of the competing models is the ARMA(1,1) model with Gamma innovations, the model with Gamma innovations is selected as the optimal model. In rows 5, 6, and 7 of Table 1, the ARMA(1,1) model with Weibull innovations has been compared to the ARMA(1,1) models with log-normal, Rayleigh, and inverse Gaussian innovations, and in all cases, the model with Weibull innovations has been selected as the optimal model. In the last row of table 1, the ARMA(1,1) model with inverse Gaussian and Rayleigh innovations as competing models have been tested, and in some cases, two models have been selected as equivalent models. Figure 1 shows a graph of the true model and competing models in different sample sizes. This chart confirms the results obtained from the Vuong test. Additionally, to demonstrate the variability of the Vuong test statistics based on different combinations of the model coefficients, a 3D scatter plot of the Vuong test statistics for several different selections of competing models was drawn (Figure 2).

Following that, using simulation, the performance of the Vuong and Cox tests in selecting a model in non-nested cases has been studied. For this purpose, the ARMA(1,1) model, $x_s = 0.2x_{s-1} + e_s - 0.4e_{s-1}$, with Gamma innovations, $\mathcal{G}(1.5, 2)$, is considered as the true model and the ARMA(1,1) models with Gamma, Weibull, Rayleigh, log-normal and inverse Gaussian innovations are considered as competing models. Then, in each double selection of the competing models, the Cox and Vuong test statistics are calculated

© 2020, SRTC Iran

Table 1. The values of Vuong test statistic for competing models

$x_s = \Phi x_{s-1} + e_s - \Theta e_{s-1}, \quad e_s = \mathcal{G}(1.5, 2)$						
$f(e_s)$ & $g(e_s)$	m	$\Phi = -0.6$	$\Phi = -0.4$	$\Phi = 0.2$	$\Phi = 0.4$	$\Phi = 0.6$
		$\Theta = -0.5$	$\Theta = -0.3$	$\Theta = -0.1$	$\Theta = 0.3$	$\Theta = 0.5$
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	0.98	0.38	1.10	1.43	0.29
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	1.97	1.98	1.34	1.10	2.72
	500	1.88	2.80	2.61	2.38	1.74
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	2.82	2.13	2.66	3.24	2.18
$g(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	250	2.10	2.05	2.49	2.41	3.87
	500	3.47	4.55	3.28	3.37	5.02
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	2.80	2.66	3.42	3.70	3.07
$g(e_s) = \mathcal{R}(\zeta)$	250	2.61	2.49	2.89	4.18	4.24
	500	5.44	5.37	5.69	4.85	5.37
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	2.98	2.98	3.46	5.42	4.67
$g(e_s) = \mathcal{INV}(\tau, \vartheta)$	250	4.79	3.80	6.06	4.29	4.80
	500	7.86	5.96	5.58	6.15	5.46
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	−2.96	−1.97	−2.04	−2.04	−1.98
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	−2.46	−2.40	−2.41	−2.26	−3.63
	500	−3.51	−3.81	−2.55	−2.64	−4.06
$f(e_s) = \mathcal{R}(\zeta)$	50	−2.97	−2.13	−2.95	−2.50	−2.68
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	−2.75	−2.75	−4.60	−2.36	−3.85
	500	−5.71	−5.17	−5.76	−4.90	−4.93
$f(e_s) = \mathcal{INV}(\tau, \vartheta)$	50	−3.18	−2.96	−3.50	−2.76	−4.16
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	−4.56	−3.85	−5.63	−3.52	−3.89
	500	−7.51	−6	−5.94	−5.61	−5.57
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	3.58	2.33	2.15	2.38	3.75
$g(e_s) = \mathcal{R}(\zeta)$	250	2.72	2.57	5.73	3.17	2.69
	500	2.82	3.28	2.43	3.37	3.21
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	3.82	2.48	3.02	6.53	7.85
$g(e_s) = \mathcal{INV}(\tau, \vartheta)$	250	3.65	3.90	5.78	3.93	3.31
	500	2.89	5.85	2.49	8.08	5.06
$f(e_s) = \mathcal{INV}(\tau, \vartheta)$	50	−1.98	−1.78	−1.97	−1.61	−3.56
$g(e_s) = \mathcal{R}(\zeta)$	250	−1.97	−1.99	−2.37	−1.98	−2.43
	500	−1.45	−2.01	−2.56	−2.67	−1.91

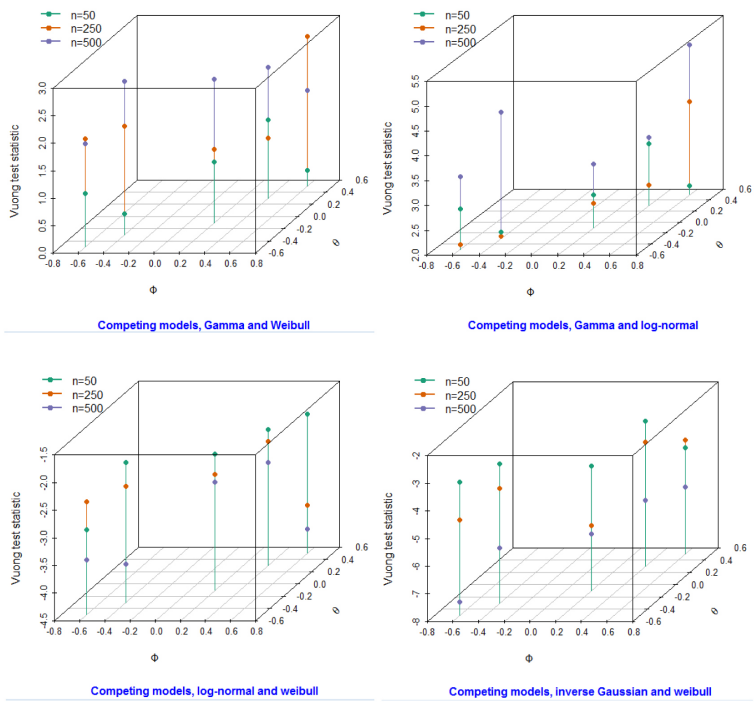

Figure 2. 3D scatter plot of the Vuong test statistics for four pairs of competing models.

© 2020, SRTC Iran

Table 2. Relative frequency of Cox and Vuong test results

$x_s = 0.2x_{s-1} + e_s - 0.4e_{s-1}, \quad e_s = \mathcal{G}(1.5, 2)$								
		Vuong test			Cox test			
$f(e_s)$ & $g(e_s)$	m	f is better	g is better	f and g are equivalent	H_0^f : Accepted H_1^g : Rejected	H_0^f : Rejected H_1^g : Accepted	Both accepted	Both rejected
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	0.170	0.025	0.850	0.216	0.056	0.717	0.011
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	0.238	0.003	0.759	0.286	0.019	0.688	0.007
	500	0.134	0.004	0.862	0.293	0.017	0.676	0.014
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	0.873	0.010	0.117	0.773	0.033	0.008	0.186
$g(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	250	0.949	0.00	0.051	0.889	0.004	0.001	0.106
	500	1.000	0.00	0.00	0.951	0.00	0.00	0.049
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	0.897	0.003	0.100	0.838	0.00	0.00	0.162
$g(e_s) = \mathcal{R}(\zeta)$	250	0.959	0.003	0.038	0.895	0.00	0.00	0.105
	500	0.988	0.00	0.012	0.920	0.00	0.00	0.080
$f(e_s) = \mathcal{G}(\alpha, \beta)$	50	0.915	0.00	0.085	0.917	0.00	0.00	0.083
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau, \vartheta)$	250	0.959	0.00	0.041	0.928	0.00	0.00	0.0715
	500	0.984	0.00	0.016	0.954	0.00	0.00	0.046
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	0.001	0.828	0.171	0.064	0.803	0.003	0.130
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	0.00	0.986	0.014	0.006	0.845	0.00	0.145
	500	0.00	1	0.00	0.00	0.712	0.00	0.288
$f(e_s) = \mathcal{R}(\zeta)$	50	0.022	0.745	0.233	0.011	0.677	0.033	0.279
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	0.001	0.937	0.062	0.005	0.707	0.00	0.288
	500	0.00	0.987	0.013	0.00	0.757	0.00	0.243
$f(e_s) = \mathcal{IN}\mathcal{V}(\tau, \vartheta)$	50	0.038	0.781	0.181	0.00	0.732	0.008	0.260
$g(e_s) = \mathcal{W}(\nu, \zeta)$	250	0.001	0.852	0.147	0.00	0.709	0.00	0.291
	500	0.00	1	0.00	0.00	0.700	0.00	0.300
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	0.804	0.125	0.071	0.169	0.017	0.00	0.810
$g(e_s) = \mathcal{R}(\zeta)$	250	0.850	0.114	0.036	0.052	0.006	0.00	0.942
	500	0.940	0.00	0.060	0.001	0.007	0.00	0.992
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	50	0.961	0.006	0.032	0.067	0.00	0.134	0.763
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau, \vartheta)$	250	0.987	0.00	0.013	0.063	0.00	0.033	0.849
	500	1	0.00	0.00	0.010	0.00	0.007	0.893
$f(e_s) = \mathcal{IN}\mathcal{V}(\tau, \vartheta)$	50	0.005	0.581	0.414	0.00	0.099	0.059	0.873
$g(e_s) = \mathcal{R}(\zeta)$	250	0.003	0.835	0.163	0.00	0.073	0.007	0.920
	500	0.00	0.989	0.011	0.00	0.013	0.00	0.987

and decisions have been made on acceptance or rejection of the hypothesis under consideration. This simulation has been repeated 10^4 times and the relative frequency results have been obtained and summarized in Table 2.

For example, to test the ARMA(1,1) models with Gamma innovations versus the model with inverse Gaussian innovations in the fourth row, it is observed that if m= 250, Then the relative frequency is 0.959, and in most cases based on the Vuong test, the ARMA(1,1) model with Gamma innovations is selected as the optimal model. Similarly, the relative frequency of selection of the ARMA(1,1) model with Gamma innovations versus the model with inverse Gaussian innovations based on the Cox test is 0.928. Therefore, the Cox test selects the model with Gamma innovations as the optimal one. The results presented in the first row of Table 2, show that if the ARMA(1,1)

models with Gamma and Weibull innovations are considered as competing models, based on the Vuong test, the two models will be selected as equivalent. Also, the Cox test with higher relative frequency values in the sixth column of Table 2, selects both models as the optimal models. Therefore, in the first row of Table 2, based on the Cox and Vuong tests, two models with Gamma and Weibull innovations are selected as two equivalent and suitable models for fitness on the data. In all rows that models with Gamma and Weibull innovations have been tested as competing models, the Cox and Vuong tests with high relative frequencies have selected Gamma and Weibull models as the optimal models. In the last row of Table 2, the ARMA(1,1) models with the Rayleigh and inverse Gaussian model innovations have been tested as competing models. Based on the Vuong test, a model with Rayleigh innovations has been selected as an optimal model. Whereas, based on Cox's test, both models are unsuitable for fitness on the data.

As noted in the previous simulation, in some cases, the Vuong test selects two competing models as equivalent, but based on this test, it is not possible to determine whether the two equivalent models are suitable or not. The Cox test helps to solve this problem. It is shown below when the Vuong test selects two competing models as equivalents, the Cox test rejects or accepts two equivalent models. We consider the ARMA(1,1) model

$$x_s = 0.3x_{s-1} + e_s + 0.5e_{s-1}$$

whose innovations have Weibull distribution of $\mathcal{W}(2, 2)$, as true model and models

$$x_s = 0.307x_{s-1} + e_s + 0.507e_{s-1}, \quad e_s \sim \mathcal{W}(2.32, 2.17),$$

and

$$x_s = 0.295x_{s-1} + e_s + 0.493e_{s-1}, \quad e_s \sim \mathcal{G}(4.11, 0.467),$$

As two competing models. Figure 3 shows that the two models are equivalent. The relative frequencies of the Cox and Vuong test results are presented in Table 3. The results of Table 3 show that for different sample sizes, based on the Vuong test, the two models are selected as equivalent models and the Cox test confirms that two equivalent models are suitable models for fitness on the data.

If the competing models, are the ARMA(1,1) models

© 2020, SRTC Iran

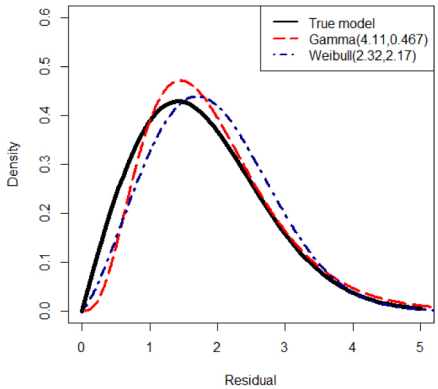

Figure 3. Graph of suitable equivalent models with Gamma and Weibull innovations

Table 3. Relative frequency of the Cox and Vuong test results for suitable equivalent models

$x_s = 0.3x_{s-1} + e_s + 0.5e_{s-1}, \quad e_s = \mathcal{W}(2,2)$								
		Vuong test			Cox test			
$f(e_s)$ & $g(e_s)$	m	f is better	g is better	f & g are equivalent	H_0^f :Accepted H_1^g :Rejected	H_0^f : Rejected H_1^g : Accepted	Both accepted	Both rejected
	50	0.003	0.103	0.89	0.003	0.018	0.979	0.00
	150	0.003	0.108	0.88	0.001	0.029	0.97	0.00
$f(e_s) = \mathcal{G}(4.1, 0.46)$	250	0.002	0.144	0.85	0.001	0.064	0.93	0.00
$g(e_s) = \mathcal{W}(2.32, 2.17)$	500	0.001	0.20	0.79	0.005	0.123	0.87	0.00
	1000	0.002	0.28	0.71	0.003	0.200	0.79	0.002

Table 4. Relative frequency of the Cox and Vuong test results for unsuitable equivalent models

$x_s = 0.3x_{s-1} + e_s + 0.5e_{s-1}, \quad e_s = \mathcal{W}(2,2)$								
		Vuong test			Cox test			
$f(e_s)$ & $g(e_s)$	m	f is better	g is better	f & g are equivalent	H_0^f :Accepted H_1^g :Rejected	H_0^f : Rejected H_1^g : Accepted	Both accepted	Both rejected
$f(e_s) = \mathcal{G}(7.09, 0.18)$ $g(e_s) = \mathcal{W}(3.09, 1.49)$	50	0.00	0.00	1	0.00	0.01	0.00	0.99
	150	0.00	0.002	0.99	0.00	0.00	0.00	1
	250	0.00	0.020	0.98	0.00	0.00	0.00	1
	500	0.00	0.03	0.97	0.00	0.00	0.00	1
	1000	0.00	0.082	0.918	0.00	0.00	0.00	1

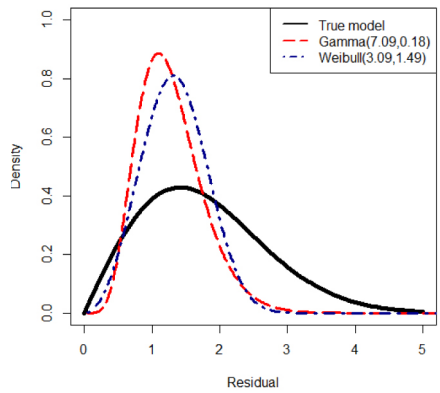

Figure 4. Graph of unsuitable equivalent models with Gamma and Weibull innovations

$$x_s = 0.19x_{s-1} + e_s - 0.37e_{s-1} \quad e_s \sim \mathcal{W}(3.09, 1.49),$$

and

$$x_s = 0.201x_{s-1} + e_s - 0.44e_{s-1} \quad e_s \sim \mathcal{G}(7.09, 0.18),$$

then, the Vuong test selects the two models as equivalent models and the Cox test selects the two models as unsuitable equivalent models. In other words, the Cox test rejects the two models. The obtained results are reported in Table 4. Figure 4 confirms the obtained results.

Finally, the Cox and Vuong model selection tests along with the tracking interval, which is another criterion for selecting a model, are analyzed. Using the simulation shown, the tracking interval confirms the Vuong test results. To this end, the model

© 2020, SRTC Iran

$$x_s = 0.3x_{s-1} + e_s + 0.5e_{s-1},$$

With Weibull innovations, $\mathcal{W}(2,2)$ as the true model and the ARMA(1,1) models with Weibull, Gamma, Rayleigh, inverse Gaussian and Log normal innovations are considered as competing models. Using MMLEs and dataset, the values of the Cox and Vuong test statistics and the tracking intervals are calculated and presented in Table 5. The first row of Table 5 is devoted to the ARMA(1,1) models with Gamma and Weibull innovations. In the sample size of 50 and based on all three model selection criteria, two competing models of Gamma and Weibull are recognized as equivalent and suitable. The Vuong test statistics and each of the Cox test statistics, in this case, have values less than 1.96 and the tracking interval contains zero. However, in other sample sizes based on all three criteria, the Weibull model is selected as the optimal model. The tracking interval with positive values confirms the selection of the Weibull model. In the other rows of the table and in the sample size of 50, three models of Weibull, Gamma and Rayleigh are selected as equivalent and suitable models compared to other competitive models. The tracking interval with positive values confirms the choice of the Weibull model. However, by increasing the sample size, based on the Cox test, model with Gamma innovations is not selected as the optimal model and also based on the Vuong test, Gamma model is not considered as equivalent to Rayleigh and Weibull models. In competition with other models, Gamma model based on the Vuong test and the tracking interval is still a better model. In the last row, the log-normal model is a better model compared to the inverse Gaussian model, based on the Vuong test and the tracking interval, but based on the Cox test, it is an unsuitable model for fitting on the data. Figure 5 confirms the obtained results.

5 Real Data Analysis

To examine the competency of the tracking interval and model selection tests in selecting the ARMA(1,1) model in non-negative mode, a numerical instance has been discussed in this section. The dataset of this example is the annual unemployment rate among the US citizens in the 15-64 age group (https://fred.stlouisfed.org/series). Kasraie and sayyareh (2020) worked on modeling this dataset by the ARMA(1,1) model and showed the positive skewness of the acquired residuals. Following that, they considered the ARMA(1,1) model with Gamma, log-normal, inverse Gaussian, Weibull

Table 5. Tracking interval values, Vuong and Cox test statistics

$x_s = 0.3x_{s-1} + e_s + 0.5e_{s-1}, \quad e_s = \mathcal{W}(2,2)$					
$f(e_s)$ & $g(e_s)$	m	Vuong test	T_f Cox test	T_g Cox test	Tracking interval
$f(e_s) = \mathcal{G}(\alpha,\beta)$	50	-0.406	-1.203	-0.484	$(-0.051, 0.078)$
$g(e_s) = \mathcal{W}(\nu,\zeta)$	250	-2.320	-3.715	0.620	$(0.003, 0.040)$
	500	-2.206	-4.433	0.224	$(0.001, 0.032)$
$f(e_s) = \mathcal{G}(\alpha,\beta)$	50	5.584	1.483	-3.063	$(-0.204, -0.096)$
$g(e_s) = \mathcal{LN}(\mathcal{M},\mathcal{S})$	250	6.028	3.837	-10.449	$(-0.101, -0.051)$
	500	5.197	3.847	-10.806	$(-0.099, -0.044)$
$f(e_s) = \mathcal{G}(\alpha,\beta)$	50	0.795	-1.77	-1.072	$(-0.182, 0.079)$
$g(e_s) = \mathcal{R}(\zeta)$	250	-2.311	-4.505	0.311	$(0.003, 0.039)$
	500	-2.076	-5.21	-0.17	$(0.0008, 0.031)$
$f(e_s) = \mathcal{G}(\alpha,\beta)$	50	1.99	1.54	-5.502	$(-0.197, 0.00)$
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau,\vartheta)$	250	5.286	3.072	-9.202	$(-0.171, -0.078)$
	500	4.075	2.674	-8.573	$(-0.193, -0.067)$
$f(e_s) = \mathcal{W}(\nu,\zeta)$	50	4.590	-0.340	-3.172	$(-0.314, -0.123)$
$g(e_s) = \mathcal{LN}(\mathcal{M},\mathcal{S})$	250	4.550	-0.991	-9.634	$(-0.140, -0.055)$
	500	4.227	-0.0127	-11.050	$(-0.130, -0.047)$
$f(e_s) = \mathcal{W}(\nu,\zeta)$	50	1.426	-0.96	-1.364	$(-0.156, 0.026)$
$g(e_s) = \mathcal{R}(\zeta)$	250	0.794	-1.828	-0.313	$(-0.003, 0.001)$
	500	0.416	-0.117	-0.210	$(-0.004, 0.002)$
$f(e_s) = \mathcal{W}(\nu,\zeta)$	50	2.205	0.583	-4.676	$(-0.283, -0.013)$
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau,\vartheta)$	250	4.641	-0.899	-9.409	$(-0.209, -0.084)$
	500	3.864	-0.109	-10.109	$(-0.22, -0.072)$
$f(e_s) = \mathcal{LN}(\mathcal{M},\mathcal{S})$	50	-3.748	-3.416	-1.107	$(0.108, 0.358)$
$g(e_s) = \mathcal{R}(\zeta)$	250	-4.572	-15.46	0.772	$(0.055, 0.139)$
	500	-4.156	-15.86	-0.300	$(0.046, 0.129)$
$f(e_s) = \mathcal{LN}(\mathcal{M},\mathcal{S})$	50	0.133	3.865	-3.760	$(-0.247, 0.216)$
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau,\vartheta)$	250	4.136	4.656	-5.593	$(-0.071, -0.025)$
	500	3.033	3.503	-4.459	$(-0.096, -0.020)$
$f(e_s) = \mathcal{R}(\zeta)$	50	2.658	-1.182	-8.165	$(-0.146, -0.020)$
$g(e_s) = \mathcal{IN}\mathcal{V}(\tau,\vartheta)$	250	4.646	-0.360	-13.709	$(-0.207, -0.084)$
	500	3.826	-0.674	-113.724	$(-0.221, -0.071)$

© 2020, SRTC Iran

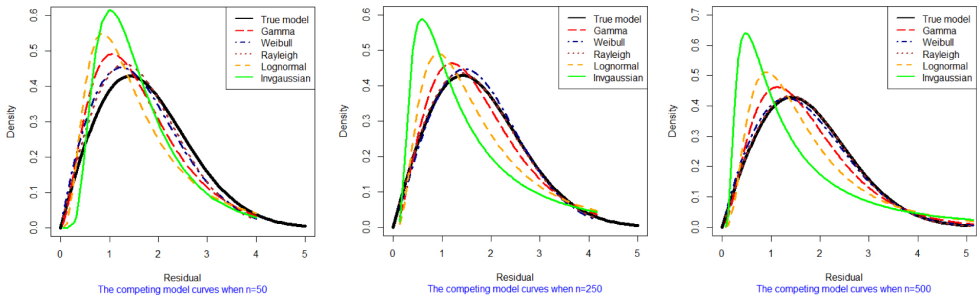

Figure 5. The true model diagram and estimated competing models with Gamma, Weibull, Rayleigh, log-normal and inverse Gaussian innovations

Table 6. The MMLEs of the ARMA(1,1) model parameters

distribution of the innovations	Φ	Θ	First parameter	Second parameter
$\mathcal{W}(\nu,\zeta)$	0.6791	0.4931	1.638	1.410
$\mathcal{R}(\sigma)$	0.6818	0.4891	-	1.063
$\mathcal{G}(\alpha,\beta)$	0.6722	0.4925	2.654	0.472
$\mathcal{LN}(\mathcal{M},\mathcal{S})$	0.6821	0.4824	0.027	0.664
$\mathcal{INVG}(\tau,\vartheta)$	0.6827	0.4923	1.255	2.126
standard deviation of the coefficient estimators	0.0043	0.0045	-	-

and Rayleigh innovation distributions as candidate models and estimated the parameters by MMLEs (Table 6).

In this article, we use Cox and Vuong tests and tracking interval to select one model from among the estimated ARMA(1,1) models to fit on the dataset.

For this purpose, Cox and Vuong test statistics and tracking interval values are calculated for each double selection of competing models (Table 7). According to the results presented in the second row of Table 7, it is concluded that if the ARMA(1,1) models with Gamma and log-normal innovations are considered as competing models, both models will be selected as suitable equivalent models. The reason is that in this case the values of the Vuong test statistic and each of Cox test statistics are less than the critical value of the standard normal distribution, 1.96, and the tracking interval contains zero. In the other rows, where one of the competing models is the ARMA(1,1) model with log-normal or Gamma innovations, the model with log-normal or Gamma innovations is selected as the optimal one. Therefore, models

$$x_s = 0.6821x_{s-1} + e_s - 0.4824e_{s-1}, \quad e_s \sim \mathcal{LN}(0.027, 0.664)$$

and

$$x_s = 0.6722x_{s-1} + e_s - 0.4925e_{s-1}, \quad e_s \sim \mathcal{G}(2.654, 0.472),$$

are optimal models to be fitted on unimployment rate data in the US.

Now, we are interested in modeling the dataset using the classical ARMA(1,1) model under the normality assumption of the innovations. It requires the transformation of original series to a normal mode before modeling. If the original series is defined as X, then an appropriate transformation for X will be X^{-1}. This is a kind of dataset on which the monotonous transformation to fix data asymmetry is effective and solves the non-normality problem. To test the normality hypothesis of the vector X^{-1}, the Kolmogorov-Smirnov test at the significant level of $\alpha = 0.05$ is used. The null hypothesis for this test is that the data are normally distributed. The evaluated p-value is 0.6221, which is greater than 0.05. Therefore, the normality hypothesis of the transformed data vector is accepted. The estimated ARMA(1,1) model to fit the dataset is acquired as

$$x_s = 0.6682x_{s-1} + e_s - 0.5259e_{s-1}, \quad e_s \sim \mathcal{N}(0.11, 0.0018). \tag{11}$$

A comparison of the estimated model in the normal mode with the selected models in the non-normal mode shows that the estimated coefficients in both modes are close to each other and the differences are only related to the innovation distributions. Although distributional assumptions about the innovations do not have a great influence on forecasts, the prediction intervals calculated under the assumption of the innovation distribution, will be affected. After estimating the parameters, we need to assess the adequacy of the estimated model by checking the validity of the model assumptions. A detailed analysis of the model residuals verify the accuracy of the model assumptions.

The main assumption is that the innovations are uncorrelated random variables. To investigate this assumption, a graph of the sample autocorrelation function (ACF) of the residuals is shown in Figure (6). By comparing the sample ACFs with the horizontal lines obtained from the standard error $\pm 2/\sqrt{n}$, the lake of autocorrelation in the residuals of this model is concluded.

The normality of the innovations can be evaluated using Q-Q plot of the

© 2020, SRTC Iran

Table 7. tracking interval values, Cox and Vuong test statistics

$x_s = \Phi x_{s-1} + e_s - \Theta e_{s-1},$				
$f(e_s)$ & $g(e_s)$	Vuong test	T_f Cox test	T_g Cox test	Tracking interval
$f(e_s) = \mathcal{G}(\alpha, \beta)$ $g(e_s) = \mathcal{W}(\nu, \zeta)$	2.205	0.583	-4.571	$(-0.271, -0.015)$
$f(e_s) = \mathcal{G}(\alpha, \beta)$ $g(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	0.506	1.206	0.628	$(-0.061, 0.087)$
$f(e_s) = \mathcal{G}(\alpha, \beta)$ $g(e_s) = \mathcal{R}(\zeta)$	3.74	1.04	-5.506	$(-0.193, -0.067)$
$f(e_s) = \mathcal{G}(\alpha, \beta)$ $g(e_s) = \mathcal{INV}(\tau, \vartheta)$	2.99	0.794	-4.676	$(-0.209, -0.089)$
$f(e_s) = \mathcal{W}(\nu, \zeta)$ $g(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$	-2.076	-3.77	-0.17	$(0.108, 0.365)$
$f(e_s) = \mathcal{W}(\nu, \zeta)$ $g(e_s) = \mathcal{R}(\zeta)$	0.794	-3.828	-9.409	$(-0.005, 0.002)$
$f(e_s) = \mathcal{W}(\nu, \zeta)$ $g(e_s) = \mathcal{INV}(\tau, \vartheta)$	1.42	4.67	-2.76	$(-0.192, 0.081)$
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$ $g(e_s) = \mathcal{R}(\zeta)$	2.658	-1.183	-8.165	$(-0.221, -0.072)$
$f(e_s) = \mathcal{LN}(\mathcal{M}, \mathcal{S})$ $g(e_s) = \mathcal{INV}(\tau, \vartheta)$	1.99	0.87	-3.62	$(-0.004, -0.001)$
$f(e_s) = \mathcal{R}(\zeta)$ $g(e_s) = \mathcal{INV}(\tau, \vartheta)$	-3.748	-10.36	-8.409	$(0.108, 0.308)$

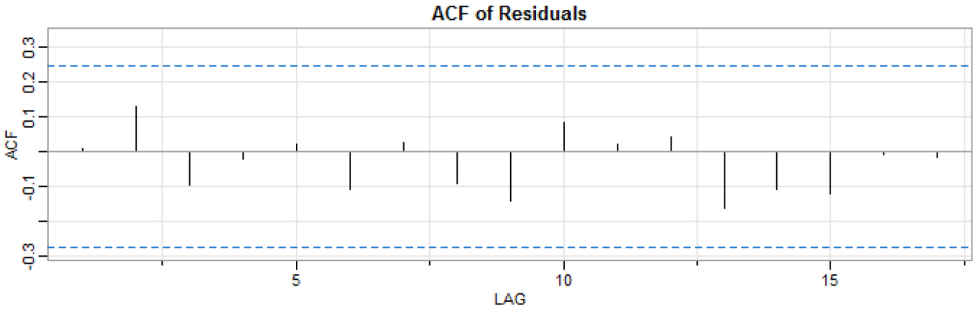

Figure 6. Sample ACFs of the residuals of model (11)

ARMA(1,1) model residuals (Figure 7). The Q-Q plot of the model residuals indicates that the points follow the straight line relatively well. According to this plot, the residual normality is not rejected. Since the p-value of the Kolmogorov-Smirnov test used for the residuals is 0.96, the normality assumption of the residuals also is not rejected.

6 Conclusion Remarks and Future Research Directions

The model selection among the ARMA(1,1) models with non-negative residuals from the Weibull and exponential families have been studied in this research. Then, the unknown model coefficients and residual distribution parameters were estimated using the MML method and the estimators asymptotic distribution was obtained. After that, based on the MMLEs, the Cox and Vuong model selection tests and the tracking interval to select the best model were investigated. Using simulations, it was depicted that the model selection tests select the best model to fit on the data. Sometimes, according to the Vuong test, two competing models are selected as equivalent; however, selecting the two models as equivalent models does not infer the proximity of the two models to the correct model. Therefore, the Cox test has been used as an inference after model selection. In this scenario, using the Cox test, two equivalent models are rejected or accepted. In any case, the tracking interval confirms the results of the Vuong test. A real dataset was evaluated and modeled using the implemented model selection techniques, and the best ARMA(1,1) model for data fitting was picked from the competing ARMA(1,1) models. For further research in the future, the calculation of the MMLEs in other models with non-negative residuals, like

© 2020, SRTC Iran

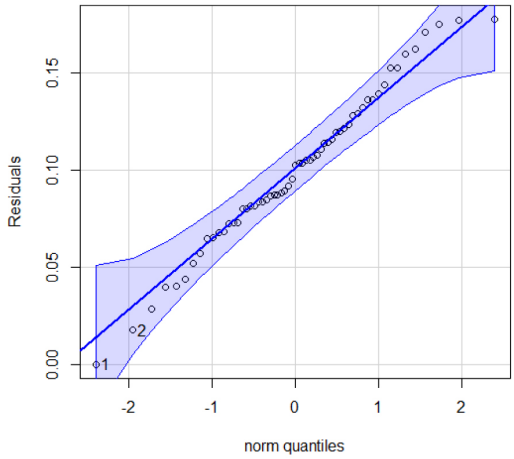

Figure 7. Q-Q plot of the ARMA(1,1) model residuals

the first-order MA model, can be investigated. Alternatively, selecting the model from the other orders of ARMA models with non-negative residuals is an important issue.

References

Akaike, H., Petrov, B.N., and Csaki, F. (1973). Second international symposium on information theory. Akadémiai Kiadó, Budapest.

Atkinson, A.C. (1970). A method for discriminating between models. *Journal of the Royal Statistical Society*, **32**, 323–344.

Bayer, F.M., Cintra, R. J., and Cribari-Neto, F. (2018). Beta seasonal autoregressive moving average models. ArXiv E-Prints, 1806, arXiv:1806.07921.

Box, G.E., and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, **26**, 211–243.

Braga, D., and Calmon, W. (2017). Periodic Gamma autoregressive model: an application to the Brazilian hydroelectric system. *RAIRO - Operations Research*, **51**, 469–483.

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*, **2**, 1123–1142.

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)*, **24**, 406–424.

Cox, D.R. (1961). Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, 23.

Duca, V.E.L. de A., Souza, R.C., Ferreira, P.G.C., and Oliveira, F.L.C. (2019). Simulation of time series using periodic Gamma autoregressive models. *International Transactions in Operational Research*, **26**, 1315–1338.

Grunwald, G.K., Hyndman, R.J., Tedesco, L., and Tweedie, R.L. (2000). Theory & methods: non-Gaussian conditional linear AR(1) models. *Australian & New Zealand Journal of Statistics*, **42**, 479–495.

Hashem Pesaran, M., and Pesaran, B. (1993). A simulation approach to the problem of computing Cox's statistic for testing nonnested models. *Journal of Econometrics*, **57**, 377–392.

Kasraie, M., and Sayyareh, A. (2020). Modified maximum likelihood estimation in first-order autoregressive moving Average models with some non-normal residuals. *JIRSS*, **19**, 33–66.

Kullback, S., and Leibler, R.A. (1951). On information and sufficiency, *Annals Maths. Statist.*, **22**, 79–86.

Resnick, S.I. (2013). *A probability path.* Springer Science & Business Media.

Sarlak, N., and Sorman, A. (2007). Gamma autoregressive models and application on Kizilirmak Basin.

Sayyareh, A., Obeidi, R., and Bar-Hen, A. (2010). Empiricial comparison between some model selection criteria. *Communications in Statistics—Simulation and Computation*, **40**, 72–86.

Shumway, R.H., and Stoffer, D.S. (2017). *Time series analysis and its applications: with R examples (4th ed.).* Springer International Publishing.

Sim, C.H. (1987). A mixed Gamma ARMA (1, 1) model for river flow time series. *Water Resources Research*, **23**, 32–36.

Spierdijk, L. (2016). Confidence intervals for ARMA–GARCH value-at-risk: The case of heavy tails and skewness. *Computational Statistics & Data Analysis*, **100**, 545–559.

Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Science*, **153**, 12–18.

Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, **54**, 155–165.

© 2020, SRTC Iran

Trindade, A.A., Zhu, Y., and Andrews, B. (2010). Time series models with asymmetric Laplace innovations. *Journal of Statistical Computation and Simulation*, **80**, 1317–1333.

Vuong, Q.H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society*, 307–333.

Wei, W.W. (2006). Time series analysis. In the Oxford Handbook of Quantitative Methods in Psychology: Vol. 2.

Weiss, G. (1975). Time-reversibility of linear stochastic processes.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.

Zamani Mehreyan, S., and Sayyareh, A. (2017). Separated hypotheses testing for autoregressive models with non-negative residuals. *Journal of Statistical Computation and Simulation*, **87**, 689–711.

Zhang, Z., and Li, W.K. (2019). An experiment on autoregressive and threshold autoregressive models with non-Gaussian error with application to realized volatility. *Economies*, **7**, 58.

Appendix A

Proof of Theorem 3.
Using the Taylor expansion of $\ell_m^f(\delta^*) = \sum_{s=2}^m \log f^{\delta^*}(x_s|x_{s-1},...,x_1,e_1) = \sum_{s=2}^m \log f^{\delta^*}(e_s)$ around $\hat{\delta}_m$, we have

$$\begin{aligned}\ell_m^f(\delta^*) &= \ell_m^f(\hat{\delta}_m) + \frac{\partial}{\partial\delta}\ell_m^f(\delta)\mid_{\hat{\delta}_m}(\delta^*-\hat{\delta}_m) + \frac{1}{2}(\delta^*-\hat{\delta}_m)\frac{\partial^2}{\partial\delta\partial\delta^T}\ell_m^f(\delta)\mid_{\hat{\delta}_m}(\delta^*-\hat{\delta}_m)^T + o_p(1)\\ &= \ell_m^f(\hat{\delta}_m) + \frac{1}{2}(\delta^*-\hat{\delta}_m)\frac{\partial^2}{\partial\delta\partial\delta^T}\ell_m^f(\delta)\mid_{\hat{\delta}_m}(\delta^*-\hat{\delta}_m)^T + o_p(1),\end{aligned}$$

where

$$\frac{\partial^2}{\partial\delta\partial\delta^T}\ell_m^f(\delta)\mid_{\hat{\delta}_m} = \begin{bmatrix} F^{\omega\omega} & F^{\omega\Phi} & F^{\omega\Theta}\\ F^{\Phi\omega} & F^{\Phi\Phi} & F^{\phi\Theta}\\ F^{\Theta\omega} & F^{\Theta\Phi} & F^{\Theta\Theta}\end{bmatrix}_{\delta=\hat{\delta}_m}.$$

In this equation $\frac{\partial^2}{\partial\delta\partial\delta^T}\ell_m^f(\delta)\mid_{\hat{\delta}_m} \xrightarrow{p} E_q\left[\frac{\partial^2\ell_m^f(\delta)}{\partial\delta\partial\delta^T}\mid_{\delta^*}\right] = A_f(\delta^*)$. Therefore,

$$\ell_m^f(\delta^*) = \ell_m^f(\hat{\delta}_m) + \frac{1}{2}(\delta^*-\hat{\delta}_m)A_f(\delta^*)(\delta^*-\hat{\delta}_m)^T + o_p(1).$$

Similary for g, we obtain

$$\ell_m^g(\vartheta^*) = \ell_m^g(\hat{\vartheta}_m) + \frac{1}{2}(\vartheta^* - \hat{\vartheta}_m)A_g(\vartheta^*)(\vartheta^* - \hat{\vartheta}_m)^T + \ o_p(1).$$

Thus, the difference between $\ell_m^f(\delta^*)$ and $\ell_m^g(\vartheta^*)$ is equal to

$$\begin{aligned}\ell_m^f(\delta^*) - \ell_m^g(\vartheta^*) = \ell_m^f(\hat{\delta}_m) - \ell_m^g(\hat{\vartheta}_m) + \frac{1}{2}(\delta^* - \hat{\delta}_m)A_f(\delta^*)(\delta^* - \hat{\delta}_m)^T \\ - \frac{1}{2}(\vartheta^* - \hat{\vartheta}_m)A_g(\vartheta^*)(\vartheta^* - \hat{\vartheta}_m)^T + o_p(1),\end{aligned}$$

or equivalently

$$\begin{aligned}\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m) = \mathrm{LR}_m(\delta^*, \vartheta^*) - \frac{1}{2}(\delta^* - \hat{\delta}_m)A_f(\delta^*)(\delta^* - \hat{\delta}_m)^T \\ + \frac{1}{2}(\vartheta^* - \hat{\vartheta}_m)A_g(\vartheta^*)(\vartheta^* - \hat{\vartheta}_m)^T + o_p(1).\end{aligned}$$

Since $(\delta^* - \hat{\delta}_m)$ and $(\vartheta^* - \hat{\vartheta}_m)$ are bounded in probability, $O_p(1)$, then

$$(m-1)^{-1/2}\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m) = (m-1)^{-1/2}\mathrm{LR}_m(\delta^*, \vartheta^*) + o_p(1).$$

Thus, it can be written that

$$\begin{aligned}&(m-1)^{-1/2}\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m) - (m-1)^{1/2}E_q\left[\log\frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)}\right] \\ &= (m-1)^{-1/2}\mathrm{LR}_m(\delta^*, \vartheta^*) - (m-1)^{1/2}E_q\left[\log\frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)}\right] + o_p(1).\end{aligned}$$

Now, from the multivariate CLT theorem, we have got

$$(m-1)^{-1/2}\mathrm{LR}_m(\hat{\delta}_m, \hat{\vartheta}_m) - (m-1)^{1/2}E_q\left[\log\frac{f^{\delta^*}(X_s|x_1, e_1)}{g^{\vartheta^*}(X_s|x_1, e_1)}\right] \xrightarrow{d} N(0, W_*^2).$$

□

© 2020, SRTC Iran

Maryam Kasraie
Department of Statistics,
Science and Research Branch,
Islamic Azad University,
Tehran, Iran.
email: *maryamkasraie@yahoo.com*

Abdolreza Sayyareh
Department of Computer Science and Statistics,
Faculty of Mathematics,
K. N. Toosi University of Technology,
Tehren, Iran.
email: *asayyareh@khntu.ac.ir*