ON TESTS FOR MODEL ADEQUACY CHECKING

Kareema Abdul-Kaddum AL-Kafaji
Department of Mathematics, College of Education

Abstract

In this paper we derive the proposed general formula for 1) The alternative statistic T for the MA(q) process that was introduced in [5]

$$T(f,f') = \log\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{f(w)}{f'(w)}dw - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\frac{f(w)}{f'(w)}dw \qquad \text{and in [6]}$$

$$T(f,f') = \log\frac{1}{2\pi}\int_{-\pi}^{\pi} f(w)dw - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(w)dw$$

2) The variance of the estimated statistic $\hat{T}$ of that statistic.

Introduction

Consider a real, zero mean stationary MA(q) process $\{Z_t\}_{t\varepsilon Z}$, generated by the following equation:

$$Z_t = (1-\theta_1 B - \theta_2 B^2 - .... - \theta_q B^q)a_t$$

where B is backward shift operator on t and $\{a_t\}$ is a sequence of independent and identically distributed random variables with finite variance σ_a^2.

The fitted model is:

$$Z_t = (1-\hat{\theta}_1 B - \hat{\theta}_2 B^2 - \hat{\theta}_3 B^3 - .... - \hat{\theta}_q B^q)a_t$$

where $\hat{\theta}_1,...,\hat{\theta}_q$ are estimated.

The goodness of this fitting to the set of data is generally tested by comparing the observed values $\{z_1,..., z_N\}$ with the corresponding predicted values of the fitted model. If the fitted model is good, then the residuals should have a behavior close to white nose.

In this way, we would like to find a statistical procedure to check if the model is valid or in other words to construct an adequacy test.

The approach is generally, founded on calculation of residuals

$$\hat{a}_t = (1-\hat{\theta}_1 B - \hat{\theta}_2 B^2 - \hat{\theta}_3 B^3 - .... - \hat{\theta}_q B^q)^{-1} Z_t \qquad (1)$$

with assumption $Z_{t-q},...,Z_0$ observable and $\hat{a}_{t-q},....,\hat{a}_0$ equal to zero and to check that the sequence $\hat{a}_1,....,\hat{a}_N$ behave like the sequence of realizations of independent random variables.

Box and Pierce (1970)[1] show that under the hypothesis of model specification, provided that m moderately large the statistic

$$\hat{Q} = N\sum_{k=1}^{m}\hat{r}_k^2 \qquad (2)$$

where $\hat{r}_k = \dfrac{\sum_{t=k+1}^{N} \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^{N} \hat{a}_t^2}$ is the residual autocorrelations, is asymptotically distributed as χ^2 with (*m-p-q*) degrees of freedom.
A simple modification studied later in detail by Ljung and Box(1978)[4]

$$\hat{Q}' = N(N+2)\sum_{k=1}^{m} \frac{\hat{r}_k^{\,2}}{N-k} \qquad (3)$$

appears to have a distribution very much closer to asymptotic χ^2.
Another different statistic was proposed by Mokkadem(1993)[5] based on the spectral density $f_Z(w)$ of the process $\{Z_t\}$he introduced the quantity

$$T(f,f') = \log \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{f(w)}{f'(w)}dw - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log \frac{f(w)}{f'(w)}dw \qquad (4)$$

and modified it in 1994[6].
we derive the general formulas of alternative statistic T introduced in [5],[6] and the variance of the estimated statistic $\hat{T}$ of that statistic.

Test based on the alternative statistic

Mokkadem[6] introduced and proposed a test of white noise based on the modified parameter
$T(f,f')$ simply denoted by T as

$$T(f,f') = \log \frac{1}{2\pi}\int_{-\pi}^{\pi} f(w)dw - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(w)dw \qquad (5)$$

Since the quantity $T(f,f')$ verity the following main property:
P1: $T(f,f')\geq 0$ and $T(f,f')=0$ if and only if f/f' =constant, that we
This statistic (5) estimated when a sample $\{Z_1,\ldots, Z_N\}$ is observed by

$$\hat{T}(f,f') = \log\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(w)dw\right| - \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|\hat{f}(w)\right|dw \qquad (6)$$

where $\hat{f}(w)$ is the truncated estimator of spectral density $f(w)$, given by

$$\hat{f}(w) = \frac{1}{2\pi}\sum_{k=-n}^{n} \hat{\gamma}_k e^{ikw} \qquad (7)$$

where

$$\hat{\gamma}_k = \frac{1}{N}\sum_{j=1}^{N-|k|} Z_j Z_{j+|k|} \qquad (8)$$

is the theoretical covariance of the process$\{Z_t\}$ which variance is

$\hat{\gamma}_0 = \int_{-\pi}^{\pi} \hat{f}(w)dw$ then

$$\hat{T} = \log(\frac{\hat{\gamma}_0}{2\pi}) - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left|\hat{f}(w)\right|dw \qquad (9)$$

which asymptotically $\hat{T} \cong \frac{1}{\gamma_0^2}\sum_{k+1}^{n}\hat{\gamma}_k^2$ as $\gamma_k = 0$ for $k \neq 0$ and $\hat{\gamma}_k = \hat{\gamma}_{-k}$.

Mokkadem (1994) [6] proposed a different approach based on the spectral density $f_a(w)$ of the process $\{a_t\}$ which are i.i.d. if and only if $f_a(w)$=constant.

The test for randomess is obtained as a test of

H_0: $f_a(w)$=constant against

H_1: $f_a(w) \neq$constant.

We will reject H_0 if $\hat{T} > t$, where t is positive real number under the test

H_0: T=0

H_1: T>0

In [2] the test is:

H_0: The sequence $\{Z_t\}$ is a white noise

H_1: The alternative hypothesis under which the sequence $\{Z_t\}$ is an ARMA process

Power of the test on the statistic $\hat{T}$

The power of the test

H_o: The process $\{Z_t\}$is white noise.

H_1: The process $\{Z_t\}$ is MA(q),

On $\hat{T}$ at the significance level α is given by the following probability

$\hat{P} = P_{H_1}(\hat{T} > t_\alpha)$

where t_α is such that

$$t_\alpha = \frac{\sqrt{2\pi}F^{-1}(1-\alpha)}{N} \qquad (10)$$

Since $\frac{N\hat{T} - n}{\sqrt{2n}}$ is asymptotically standard normal distribution under H_o by Theorem 2 [see Drouiche, K. and Mohdeb,Z.(1995)[2].

Under H_1 for N large the distribution of $\hat{T}$ can be approximation i.e $N(T, \Gamma^2)$ where

$$\Gamma^2 = \frac{1}{N\pi}\int_{-\pi}^{\pi}(1 - \frac{f(w)}{k})^2 dw \qquad (11)$$

and

$$k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(w)dw$$

Now we must calculate T and Γ^2

1) Calculating the statistic T of MA(q) Model

Let $\{Z_t\}$be a MA(q) process which generated by the following equation:

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

$$= (\ 1 - \theta_1 B - \theta_2 B^2 \ - \dots - \theta_q B^q)a_t$$

$$Z_t = \theta(B)a_t \qquad (12)$$

it's spectral density is

$$f(w)=\frac{\sigma_a^2}{2\pi}\left|1-\theta_1 e^{-iw}-\theta_2 e^{-2iw}-\ldots-\theta_q e^{-qiw}\right|$$
$$=\frac{\sigma_a^2}{2\pi}\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2 \qquad (13)$$

where σ_a^2 is the variance of the process $\{a_t\}$, since

$$T=\log\frac{\gamma_o}{2\pi}-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f(w)dw \qquad (14)$$

where

$\gamma_o=\sigma_a^2(1+\theta_1^2+\theta_2^2+\ldots+\theta_q^2)$ is the variance of MA(q) under the alternative hypothesis, then

$$T \quad = Log\frac{\sigma_a^2(1+\sum_{k=1}^{q}\theta_k^2)}{2\pi}-\frac{1}{2\pi}\int_{-\pi}^{\pi}Log\frac{\sigma_a^2}{2\pi}\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2 dw$$

$$T=Log(1+\sum_{k=1}^{q}\theta_k^2)-\frac{1}{2\pi}\int_{-\pi}^{\pi}Log\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2 dw \qquad (15)$$

If $\left|\sum_{k=1}^{q}\theta_k e^{-ikw}\right|<1$, then by using a Taylor series expansion of $Ln(1-x)$ for $|x|<1$

And the identity $\int_{-\pi}^{\pi}e^{\pm ikw}dw=0,\quad k\neq 0$ where $Log|1-x|=Ln|1-x|$, then

$$Log\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2=Log\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|\left|1-\sum_{k=1}^{q}\theta_k e^{kiw}\right|$$
$$=Log\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|+Log\left|1-\sum_{k=1}^{q}\theta_k e^{kiw}\right|$$
$$=\sum_{j=1}^{\infty}\frac{(\sum_{k=1}^{q}\theta_k e^{-ikw})^j}{j}+\sum_{L=1}^{\infty}\frac{(\sum_{k=1}^{q}\theta_k e^{ikw})^L}{L}$$

Then the statistic

$$T=Log(1+\sum_{k=1}^{q}\theta_k^2) \qquad (16)$$

Since $\int_{-\pi}^{\pi}Log\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2 dw=0$ by the fact that

$\int_{-\pi}^{\pi}e^{\pm kiw}dw=\frac{1}{\pm kiw}[\sin(kw)\pm i\cos(w)]_{-\pi}^{\pi}=0$, where

1) cos(π)=-1

2) sin(-π)=-sin(π)=0

The equation(16) is the general formula for the statistic that introduced by Mokkadem [5],[6] for MA(q) model, where

1) Calculating the variance of $\hat{T}$

The variance of $\hat{T}$ given in [3] by the following formula

$$\Gamma^2 = \frac{1}{N\pi}\int_{-\pi}^{\pi}(1-\frac{f_Z(w)}{k})^2\,dw \qquad (17)$$

where

$$k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_Z(w)dw$$

$$\begin{aligned} f_Z(w) = \frac{\sigma_a^2}{2\pi}[&\sum_{k=0}^{q}\theta_k^2 - 2\sum_{k=1}^{q}\theta_k\cos(kw) + 2\sum_{k=1}^{q-1}\theta_k\theta_{k+1} + \sum_{k=1}^{q-2}\theta_k\theta_{k+2}\cos(2w) \\ &+\sum_{k=1}^{q-3}\theta_k\theta_{k+3}\cos(3w) + \ldots + \sum_{k=1}^{q-(q-1)}\theta_k\theta_{k+(q-1)}\cos((q-1)w)] \end{aligned} \qquad (18)[3]$$

in [3] where θ_0=1 and $\sum_{k=1}^{q-(q-1)}\theta_k\theta_{k+(q-1)}\cos((q-1)w) = \theta_1\theta_q\cos((q-1)w)$

$$\begin{aligned} k = \frac{\sigma_a^2}{(2\pi)^2}[&2\pi\sum_{k=0}^{q}\theta_k^2 + \sum_{k=1}^{q}\frac{2}{k}\theta_k\sin(kw) + 2\sum_{k=1}^{q-1}\theta_k\theta_{k+1}\sin(w) + \frac{1}{2}\sum_{k=1}^{q-2}\theta_k\theta_{k+2}\sin(2w) \\ &+\ldots+\frac{\theta_1\theta_q}{(q-1)}\sin((q-1)w)]_{-\pi}^{\pi} = \frac{\sigma_a^2}{2\pi}\sum_{k=0}^{q}\theta_k^2 \end{aligned}$$

$$\Gamma^2 = \frac{1}{N\pi}\int_{-\pi}^{\pi}(1-\frac{\sigma_a^2\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2}{2\pi(\frac{\sigma_a^2}{2\pi}\sum_{k=0}^{q}\theta_k^2)})^2\,dw = \frac{1}{N\pi}\int_{-\pi}^{\pi}(1-\frac{\left|1-\sum_{k=1}^{q}\theta_k e^{-ikw}\right|^2}{\sum_{k=0}^{q}\theta_k^2})^2\,dw$$

$$\Gamma^2 = \frac{2\pi}{N\pi} - \frac{2}{N\pi\sum_{k=0}^{q}\theta_k^2}\int_{-\pi}^{\pi}\left|1-\sum_{-\pi}^{\pi}\theta_k e^{-ikw}\right|^2 dw + \frac{1}{N\pi}\int_{-\pi}^{\pi}\frac{1}{(\sum_{k=0}^{q}\theta_k^2)^2}\left|1-\sum\theta_k e^{-ikw}\right|^4 dw$$

$$\Gamma^2 = \frac{1}{N}[\frac{4\sum_{k=1}^{q}\theta_k^2}{[1+\sum_{k=1}^{q}\theta_k^2]^2}]^2 \qquad (19)$$

that is the variance of the statistic $\hat{T}$ and the power of the test as function of $\theta_1, \ldots, \theta_k$ is given by

$$P_{\hat{T}}(\theta) = 1 - F(\frac{t_\alpha - T}{\Gamma}) \qquad (20)$$

where t_α , T and Γ are given respectively by (10),(16) and (19).
F is the distribution function of the random variable N(0,1).

Conclusion

This paper has been devoted to model adequacy checking. The two proposed formulas are:

1) For alternative statistic T introduced in [5],[6] used only when $\{Z_t\}$ is a MA(q) process under H_1 with the condition $\left|\sum_{k=1}^{q}\theta_k e^{-ikw}\right| < 1$

2) For the variance of the estimated $\hat{T}$.

These formulas are very easy and simple in using.

References

[1] Box, G.E.P.& Pierce, D.A.(1970),Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models. J.Am.Stat.Ass. 65, p.1059-26.

[2] Drouiche, K.& Mohdeb, Z.(1995),"On Test for Model Adequacy Checking", Prepublic Actions, University' de Paris-Sud Mathematiques 425, 91405 Orsay France.

[3] Kafji, K.A.K.(2000),"The proposed general formula for the population spectrum of MA(q) Model" Unpublished research, College of Education, Univ. of Babylon.

[4] Ljung, G.M.& Box, G.E.P.(1978),On Measure of Lack of Fit in Time series Models, Biometrika 65, p.297-303.

[5] Mokkadem, A.(1993), On some tests for time series. Submitted.

[6] ____________,(1994), Sur Quelques Nouveaux Tests pour les Series Chronologiques, C.R. Acad. Sci. Paris, t.318, se'rie I, pp.755-58.