Gazi University Journal of Science
GU J Sci
25(2):409-416 (2012)

ORIGINAL ARTICLE

Residual Types in Time Series and Their Applications

Mehmet Güray ÜNSAL[1,♠], Reşat KASAP[2]

[1]*Gazi University, Science Faculty, Statistics Department, 06100,Ankara, TURKEY*

[2]*Gazi University, Science Faculty, Statistics Department, 06100,Ankara, TURKEY*

Received: 27.11.2011 Revised: 26.12.2011 Accepted: 08.02.2012

ABSTRACT

Residual types in time series has not been investigated throughly in literature. This study aims to provide practical applications of residual types. In this study, firstly, basic information about different types of residuals was given and some features of the residuals were investigated with numerical applications. Then a simulation study was conducted to show differences in decisions when different residual types were considered in diagnostic checking.

Keywords: Time Series, Residual Types, Diagnostic Checking

1. INTRODUCTION

Time series is an applied field of statistics and could be employed in every branch of science. Box and Jenkins (1976) worked on building and forecasting time series models and found out the method which is called Box-Jenkins Modelling Process in time series analysis [Akdi, 2003]. One of the most important points in the process of analyzing a time series is diagnostic checking. There are two ways to determine which Box-Jenkins model is suitable for time series data. The first method is examining the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots of the time series. The second method is testing the (H0) null hypothesis that the correlations of residuals for lags from 1 to m are equal to zero simultaneously [Arranz, 2005]. Thus, residuals play a significant role in diagnostic checking in time series models.

Examination of the general structure of the residuals is very essential in practical applications, however there are not many studies in this subject. Ansley and Newbold (1979) used normalized residual vector instead of unconditional residual vector in their simulation study and showed that extends the range of cases for which statistics frequently used in model diagnostic checking can be usefully interpreted through the usual asymptotic significance levels. Moreover, Brockwell and Davis (2002) demonstrated that both innovations and normalized residuals can be calculated in analysis. Mauricio (2008) gave some basic distributional and theorotical properties about the residuals in time series analysis which can be classified in 4 different classes as, "conditional residuals", "unconditional residuals", "innovations" and "normalized residuals" . In addition, Mauricio indicated that when roots of a model contains moving average that are near the unit circle and the number of observations is small, the unconditional residuals and normalized residuals give different decisions in diagnostic checking. Unsal and Kasap (2010) investigated some special matrices and their elementwise cases, which are used to calculate residuals types.

♠Corresponding author, e-mail: mgunsal@gazi.edu.tr

Since very few studies have been conducted about the residuals, this subject has not been investigated sufficiently in the literature. Thus this article aims to fill this gap by providing practical applications of residual types in literature. This paper is structured as follows: Section 2 gives the structures of the residual types adopted by Mauricio (2008). Then in Section 3, numerical applications are given to indicate some features about the residual types and a simulation study is conducted for comparing the conditional and normalized residuals in diagnostic checking. Lastly, we concluded the study in Section 4.

2. DEFINING AND CALCULATING TYPES OF RESIDUALS FOR ARMA(p,q)

Let $\{W_t\}$ be a stationary time series process, following the model and let $w=[w_1,w_2,...,w_n]'$ be generated from $\{W_t\}$. The theoretical representation of ARMA (Autoregressive Moving Average) model is given below

$$\phi(B)\tilde{W}_t=\theta(B)A_t. \quad (1)$$

Here, $\phi(B)=1-\sum_{i=1}^{p}\phi_i B^i$ and $\theta(B)=1-\sum_{i=1}^{p}\theta_i B^i$ are polinomials with degrees of *p* and *q,* also *B* is a lag operator, $\tilde{W}_t=W_t-E[W_t]$ and $\{A_t\}$ is a white noise process with $\sigma^2>0$. Regarding model (1) for $t=1,2,...,n$, $\tilde{W}=[\tilde{W}_1,\tilde{W}_2,...,\tilde{W}_n]'$, $A=[A_1,A_2,...,A_n]'$ and $U_*=[\tilde{W}_{1-p},...,\tilde{W}_0,A_{1-q},...,A_0]'$, observed time series $w=[w_1,w_2,...,w_n]'$ can be seen as a particular realization of a random vector $W=[W_1,W_2,...,W_n]'$ following the model;

$$D_\phi\hat{W}=D_\theta A+VU_* \quad (2)$$

Where D_ϕ and D_θ are nxn parameter matrices with ones as diagonal elements and $-\phi_j$ and $-\theta_j$ as elements that constitute the jth subdiagonal respectively, and V is a $nx(p+q)$ matrix with $V_{ij}=\phi_{p+i-j}(i=1,...,p;j=1,...,p)$ and $V_{ij}=-\theta_{q+i-j+p}(i=1,...,q;j=p+i,...,p+q)$ where the remaining elements are zero [Mauricio, 2008].

Let us assume that the theoretical autocovariance matrix is $\sum_w=\sigma^{-2}E[\tilde{W}\tilde{W}']$, and $\hat{\sum}_w$ is an estimation of $\sum_w$. The autocovariance matrix can be given as follows from equation (2) [Mauricio, 2008];

$$\sum_w=D_\phi^{-1}(D_\phi D_\theta'+V\Omega V')D_\phi^{-1}{}'=K^{-1}(I+Z\Omega Z')K^{-1}' \quad (3)$$

In equation (3) $K=D_\theta^{-1}D_\phi, Z=-D_\theta^{-1}V$ and $\Omega=\sigma^{-2}E[U_*U_*']$ are parameter matrices of dimensions $nxn, nx(p+q)$ and $(p+q)x(p+q)$ respectively, with being readily expressible in terms of $\phi_1,...,\phi_p,\theta_1,...,\theta_q$ for example, in Ljung and Box (1979). Besides, here $\sum_0=I+Z\Omega Z'=[I-Z(\Omega^{-1}+Z'Z)^{-1}Z']^{-1}$

Using the relation (3), $\tilde{w}'\hat{\Sigma}_w^{-1}\tilde{w}$ can be written as,

$$\tilde{w}'\hat{\Sigma}_w^{-1}\tilde{w}=\tilde{w}'\hat{K}'(I+Z\Omega Z')\hat{K}\tilde{w} \quad (4)$$

$\hat{K},\hat{Z}$ and $\hat{\Omega}$ symbolize the estimations of parameter matrices defined in equation (3). According to these theoretical information, residuals have been grouped by Mauricio (2008) in four different types as below:

2.1. Conditional Residuals

Conditional residuals are associated with relation (4) and defined as the elements of the *nx1* vector $\hat{a}_0=\hat{K}\tilde{w}$.

2.2. Unconditional Residuals

Unconditional residuals are associated with (4) and defined as the elements of the *nx1* vector $\hat{a}=(I+\hat{Z}\hat{\Omega}\hat{Z}')^{-1}\hat{K}\tilde{w}=\hat{\sum}_0^{-1}\hat{a}_0$.

2.3. Innovations

Innovation residuals are associated with (4) and defined as the elements of the *nx1* vector $\hat{e}=\hat{L}^{-1}\tilde{w}=(\hat{K}\hat{L})^{-1}\hat{a}_0$. Here, $\hat{L}$ is the estimation of the *nxn* unit lower- triangle matrix *L* in below as Cholesky decomposition,

$$\sum_w = LFL' = \begin{bmatrix} 1 & 0 & \dots & 0 \\ L_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \dots & 1 \end{bmatrix} \begin{bmatrix} F_1 & 0 & \dots & 0 \\ 0 & F_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & F_n \end{bmatrix} \begin{bmatrix} 1 & L_{21} & \dots & L_{n1} \\ 0 & 1 & \dots & L_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

$$F_t > 0 \ (t = 1, 2, ..., n)$$

From relation (4), the equations given below will be obtained [Mauricio, 2008]:

$$\tilde{w}' \hat{\sum}_w^{-1} \tilde{w} = \hat{a}_0' \hat{\sum}_0^{-1} \hat{a}_0 = \hat{a}' \hat{\sum}_0 \hat{a} = \hat{e}' \hat{F}^{-1} \hat{e}$$

where, $\hat{F}$ is the estimation of matrix *F*.

2.4. Normalized Residuals

If we define lower-triangle matrix *P* as, $\sum_0 = I + Z\Omega Z' = PP'$, the definition of vector $\hat{v}$ (normalized residuals) is $\hat{v} = \hat{P}^{-1}\hat{a}_0 = \hat{P}'\hat{a} = \hat{F}^{-1/2}\hat{e}$, where $\hat{P}$ is the estimator of matrix *P* [Mauricio, 2008; Box et al., 1994; Wei,1990; Kasap, 1998].

$$V_{P^{-1}} = \begin{bmatrix} 0.779334306883787 \\ 0.871036251437872 \\ 0.914606656149130 \\ 0.939710732340943 \\ 0.955757237069755 \\ 0.966696844415371 \\ 0.974484506469331 \\ 0.980200094152302 \\ 0.984489332730189 \\ 0.987762164172522 \end{bmatrix}_{10x1}$$

The vector $V_{P^{-1}}$ and vector V_F are diagonal elements vectors of P^{-1} and *F* matrices respectiely. As seen from the vectors, diagonal elements of P^{-1} and F converge to one from below and above, respectively, as *i* increases. The diagonal elements of P^{-1} and *F* vectors are found as below when number of obsevation *(n)* is 25:

$$V_F = \begin{bmatrix} 1.64646465000000 \\ 1.31803681087231 \\ 1.19544963743166 \\ 1.13243067826744 \\ 1.09472442901383 \\ 1.07008776406892 \\ 1.05305274090786 \\ 1.04080775678749 \\ 1.03175829809329 \\ 1.02493241052978 \end{bmatrix}_{10x1}$$

3. NUMERICAL APPLICATIONS AND SIMULATION STUDY

In this section, firstly, diagonal elements which belong to P^{-1} and *F* matrices are presented for ARMA model from the vectors which are generated by diagonal elements of these matrices. In Mauricio (2008), for an invertible model, the *(i,i)th* element of $\sum_0 = I + Z\Omega Z' = PP'$ converges to one from above as *i* increases, it turns out that the *(i,i)th* element of P^{-1} (which is strictly positive of all *i*) converges to one from below as *i* increases. As a contribution to this finding of Mauricio, similarly for *F* matrix, it could be said that *(i,i)th* element of *F* converges to one from above as *i* increases [Unsal and Kasap, 2010].

To see these features, P^{-1} and *F* matrices are composed under the assumption that the appropriate model is *ARMA(1,1)* with ϕ=0.1 and θ=0.9 parameter values. In this situation, when the number of observation *(n)* is 10, the diagonal elements vectors are as below:

$$V_{P^{-1}} = \begin{bmatrix} 0.779334306883787 \\ 0.871036251437872 \\ 0.914606656149130 \\ 0.939710732340943 \\ 0.955757237069755 \\ 0.966696844415371 \\ 0.974484506469331 \\ 0.980200094152302 \\ 0.984489332730189 \\ 0.987762164172522 \\ 0.990291250152694 \\ 0.992264773143076 \\ 0.993816528425645 \\ 0.995043961782710 \\ 0.996019450712885 \\ 0.996797621003420 \\ 0.997420243544843 \\ 0.997919602738620 \\ 0.998320869785524 \\ 0.998643810287995 \\ 0.998904035573201 \\ 0.999113934336597 \\ 0.999283375845123 \\ 0.999420246982722 \\ 0.999530866517954 \end{bmatrix}_{25x1} \qquad V_F = \begin{bmatrix} 1.646464650000000 \\ 1.318036810872310 \\ 1.195449637431660 \\ 1.132430678267440 \\ 1.094724429013830 \\ 1.070087764068920 \\ 1.053052740907860 \\ 1.040807756787490 \\ 1.031758298093290 \\ 1.024932410529780 \\ 1.019703984693670 \\ 1.015651824295520 \\ 1.012482602183250 \\ 1.009986253340680 \\ 1.008008886437010 \\ 1.006435655579300 \\ 1.005179547237160 \\ 1.004173814791230 \\ 1.003366737840700 \\ 1.002717907170050 \\ 1.002195537540520 \\ 1.001774489449620 \\ 1.001434790433700 \\ 1.001160515155260 \\ 1.000938927636010 \end{bmatrix}_{25x1}$$

Elements of $V_{P^{-1}}$ and V_F vectors converge to one from below and above again, respectively, as *i* increases. The convergence of the elements is better when *n=25* than when *n=10.* This result supports another finding of Mauricio. Mauricio (2008) indicated that when roots of a model contains moving average that are near the unit circle (or close to non-invertibilty) and the number of observation is small, the residuals causes different decision making in diagnostic checking. This situation is a result of cases of P^{-1} and *F* matrices. Let us suppose that we are interested in calculating conditional and normalized residuals. The relationship between them is $v = P^{-1}a_0$. So, the difference between a_0 and v is caused from P^{-1}. Besides, because of matrix *Z*, model's moving average parameters are influenced by the value of P^{-1} as presented in Section 2. When the moving average parameter is close to non-invertibility and the number of observations is small, the calculated residual values could be different from each other. The calculated values of conditional and normalized residual values under ARMA(1,1) model with parameters ϕ *=0.1* and θ *=0.9* are given below for *n=10* and *25*.

Table 1. Conditional and Normalized Residuals for n=10

n	Conditional	Normalized
1	1.97379	1.53824
2	1.57116	0.76100
3	0.69266	-0.11415
4	0.93222	0.20191
5	1.17986	0.48899
6	1.01038	0.35683
7	1.28699	0.67054
8	-1.85790	-2.38003
9	1.96528	1.51393
10	-0.12716	-0.54772

Table 2. Conditional and Normalized Residuals for n=25

n	Conditional	Normalized	n	Conditional	Normalized	n	Conditional	Normalized
1	1.87087	145804	9	-0.67701	-0,58488	17	0.37777	0.39825
2	-1.23734	-1.65363	10	0.42791	0,51264	18	0.40583	0.42246
3	-1.47123	-1.51271	11	0.95894	1,01955	19	1.51810	1.52971
4	0.35036	0.40340	12	0.67653	0,71659	20	1.32287	1.32920
5	1.50438	1.46252	13	0.76587	0,79201	21	-2.14862	-2.14220
6	-0.46384	-0.54110	14	-0.96780	-0,94398	22	-0.64083	-0.63238
7	-1.33207	-1.35004	15	-0.63317	-0,60510	23	0.40906	0.41687
8	-1.98629	-1.93243	16	0.52340	0,54906	24	-0.47641	-0.46937
						25	1.16136	1.16739

As seen from Table 1 and Table 2, conditional and normalized residuals are quite different. This difference between residuals might cause a significant difference between calculated test statistics (Portmanteu test statistics). In other words, residual types could lead to different decisions in diagnostic checking. Another important finding is that the differences between residual types vanish when the number of observations increases. A similar deduction could be made for calculated test statistics values.

To see the differences between calculated residuals, again assume that appropriate model is ARMA(1,1) with parameters ϕ*=0.1* and θ*=0.9*, and conditional ($\hat{a}_0$), unconditional ($\hat{a}$), innovations ($\hat{e}$) and normalized ($\hat{v}$) residuals are calculated from the generated time series. The mean absolate differences are given in Table 3 for different numbers of observation *(n)* as below:

Table 3. Mean Absolate Differences For Residuals

Number of Observation	$\sum_{i=1}^{n}\lvert\hat{a}_{0i}-\hat{a}_i\rvert/n$	$\sum_{i=1}^{n}\lvert\hat{a}_{0i}-\hat{e}_i\rvert/n$	$\sum_{i=1}^{n}\lvert\hat{a}_{0i}-\hat{v}_i\rvert/n$	$\sum_{i=1}^{n}\lvert\hat{a}_i-\hat{e}_i\rvert/n$	$\sum_{i=1}^{n}\lvert\hat{a}_i-\hat{v}_i\rvert/n$	$\sum_{i=1}^{n}\lvert\hat{e}_i-\hat{v}_i\rvert/n$
n=10	0.16006	0.23615	0.25728	0.12524	0.11438	0.05892
n=25	0.09124	0.12356	0.13177	0.05363	0.04900	0.02568
n=50	0.04890	0.06501	0.06912	0.02690	0.02458	0.01286
n=100	0.02457	0.03263	0.03469	0.01345	0.01229	0.00643
n=250	0.00983	0.01305	0.01387	0.00538	0.00492	0.00257
n=500	0.00491	0.00653	0.00694	0.00269	0.00246	0.00129

In Table 3 *(i=1,2,...,n)*, the mean absolate differences between calculated residuals tend to decrease as the number of observations increases. Table 3 shows that the differences between calculated test statistics which are used for diagnostic checking decrease as the number of observations increase.

As a last investigation of this study, the conditional and normalized residuals are compared by a simulation study for the *ARMA(1,1)* model under certain parameter values for different number of observations and lags and their conditions in diagnostic checking are examined using the Ljung-Box test statistic [Ljung and Box, 1978]. The test statistics, which are used for diagnostic checking, are called Portmanteau test statistics. There are a lot of Portmanteau test statistics in the literature and detailed information about structures and distributions of other Portmanteau test statistics could be found in many studies [Ljung and Box, 1978; Davies, Triggs and Newbold 1977; Battaglia ,1990; Arranz 2005].

Ljung-Box test statistic, one of the most popular Portmanteu test statistics tests the hypothesis "Ho: Model is appropriate" and the test statistics could be written as below:

$$Q_{LB} = n(n+2)\sum_{k=1}^{m} r_k^2/(n-k)$$

k is the number of lag, n is the number of observations and r_k is the sample autocorrelation values obtained as follows:

$$r_k = \sum_{t=k+1}^{n} u_t u_{t-k} / \sum_{t=1}^{n} u_t^2, \qquad k = 1,2,...,m$$

Here, u_t represents the residuals. In the simulation study, we used two pairs of model parameter values as $\phi=0.5, \theta=0.6$ and $\phi=0.2, \theta=0.9$ respectively. To see the rejection ratio of the null hypothesis "Ho: Model is appropriate", trials are repeated 100 times and total number of rejections are divided by 100. Furthermore, to observe the effect of the sample size and number of lags, different numbers of observations and lags are taken.

Table 4. Rejection Ratios of Null Hypothesis for $\phi=0.5$ and $\theta=0.6$
(n: Number of Observation, m: Number of Lag)

Residuals / n and m	Conditional Residuals	Normalized Residuals
n=25, m=10	0.03	0.03
n=25, m=5	0.01	0.01
n=25, m=3	0.05	0.05
n=25, m=2	0.02	0.02
n=50, m=10	0.12	0.12
n=50, m=5	0.09	0.09
n=50, m=3	0.05	0.05
n=50, m=2	0.08	0.08
n=100, m=10	0.09	0.09
n=100, m=5	0.05	0.05
n=100, m=3	0.03	0.03
n=100, m=2	0.05	0.05

Table 5. Rejection Ratios of Null Hypothesis for $\phi=0.2$ and $\theta=0.9$
(n: Number of Observation, m: Number of Lag)

Residuals / n and m	Conditional Residuals	Normalized Residuals
n=25, m=10	0.15	0.06
n=25, m=5	0.12	0.06
n=25, m=3	0.11	0.07
n=25, m=2	0.14	0.05
n=50, m=10	0.14	0.09
n=50, m=5	0.09	0.05
n=50, m=3	0.13	0.05
n=50, m=2	0.07	0.03
n=100, m=10	0.12	0.10
n=100, m=5	0.08	0.07
n=100, m=3	0.08	0.07
n=100, m=2	0.09	0.06

In Table 4 and 5, the rejection ratios of the null hypothesis "Ho: Model is appropriate" could be seen for different number of observations and lags. According to simulation study, when the parameter values are $\phi=0.5$ *and* $\theta=0.6$, the residual types have the same rejection ratios. But when the model parameters are close to non-invertibility as $\phi=0.2$ and $\theta=0.9$, the residual types have different rejection ratios. This situation could be generalized for different number of observations and different number of lags as seen in Table 4 and 5.

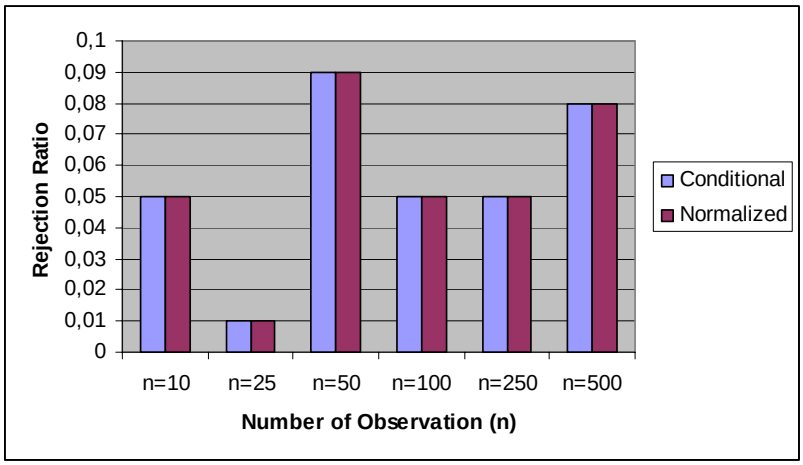

Figure 1. Rejection Ratios for Conditional and Normalized residuals
(ϕ=0.5, θ=0.6, Lag=5)

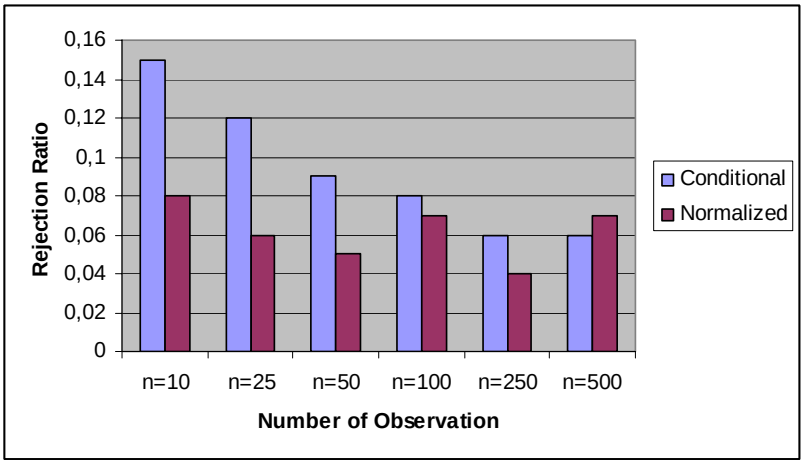

Figure 2. Rejection Ratios for Conditional and Normalized residuals
(ϕ=0.2, θ=0.9, Lag=5)

The results in Table 4 and 5 could be summarized with Figure 1 and Figure 2 for Lag=5. Similarly Table 4 and 5, in Figure 1, when the model parameter values are taken to be ϕ=0.5, θ=0.6, there is no difference between the rejection ratios of the two residual types for any number of observations. In contrast, in Figure 2 (for ϕ=0.2, θ=0.9), when the model parameters are taken such that the model is close to non-invertibility, in other words, when roots of a model contain moving average near the unit circle, the residuals have different rejection ratios. However this situation vanishes as the number of observations increases. These results could be seen better in the figures. Figure 1-2 are drawn for Lag=5, but the similar situations could be seen in Table 4-5 for the other different number of lags. Figure 1-2 are just given for the provision of convenience as the visual.

Different rejection ratios are caused by different calculated Ljung-Box test statistic values. Genarally speaking calculated test statistics are different for conditional residuals than for normalized residuals. Obviously, the different rejection ratios occur due to differences between the residuals. As mentioned in this study, MA parameters and number of observations effect the residual types' values. So, in parallel with Mauricio (2008), we concluded that when the model is close to non-invertibility and the number of observations is small, different types of residuals could cause different decision making in diagnostic checking.

4. CONCLUSION

In this study, we investigated the residual types in time series. Basic computational information about the different types of residuals is given. Besides, some calculation results are given to indicate some features about the residuals in literature. A simulation study is conducted for comparing the conditional and normalized residuals in diagnostic checking. This study can be summarized briefly as follows:

If we consider the matrix *F*, which is obtained from the decomposition of $\sum_w = LFL'$, its diagonal elements converge to one from above. Moreover, for the matrix *P*, which is obtained from the decomposition of $\sum_0 = I + Z\Omega Z' = PP'$, the diagonal elements of P^{-1} converge to one from below. It is also seen that the last diagonal elements F and P^{-1} matrices show better covergence rate to one as the number of observations increases. Another investigation for practical purposes is that, the mean absolate difference between calculated residual values tend to decrease as the number of observations increases as seen in Table 3. It is also concluded that when the model is close to non-invertibility and the number of observations is small, different types of residuals could cause different decision making in diagnostic checking.

REFERENCES

Ansley, C.F., Newbold, P., On the finite sample distribution of residual autocorrelations in autoregressive moving average models. Biometrika, 66:547-553 (1979).

Akdi, Y., , Time Series Analysis (Unit Roots and Cointegration), *Bıçaklar Bookstore*, Ankara, 1-18 (2003).

Arranz, M.A., Portmanteau Test Statistics in Time Series: Time-Oriented Language, 25-32 (2005) .

Battaglia, F., "Approximate power of portmanteau tests for time series", *Statistics and Probability Letters*, 9: 337-341 (1990).

Box, G.E.P., Jenkins, G.M and Reinsel, G.C., Time Series analysis: Forecasting and a Control , Prentice Hall: New Jersey, 7-88, 224-307 (1994).

Brockwell, P.J., Davis, R.A., Introduction to Time Series and Forecasting. Sec ond ed. Springer, NewYork (2002).

Davies, N., Triggs, C. M. And :Newbold, P., "Significance levels of the Box-Pierce "Portmanteau statistic in finite samples", *Biometrika*, 64: 517-522 (1977).

Kasap, R., An analysis of the Istanbul Stock Exchange (ISE) national-100 index: a statistical approach. ISE Review, 6:27-33 (1998).

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

Mauricio, J.A., Computing and Using Residuals in Time Series Models, Computational Statistics & Data Analysis, 52, 3:1746 – 1763 (2008).

Unsal, M.G., Kasap, R., Analysis Of The Matrices Used For Computing Of Residuals Types For ARMA(1,1) Model, 19. Statistics Research Symposium, TÜİK, Ankara (2010).

Wei, W.W.S., Time Series Analysis: Univariate and Multivariate Methods, Addison-Wesley Publishing Company, Inc.: UK, 32-64, 135-156 (1990).