

On the Distribution of Residual Autocorrelations in Box–Jenkins Models

By A. I. McLEOD†

Department of Statistics, University of Waterloo, Ontario, Canada

[Received August 1977. Revised January 1978]

SUMMARY

The large sample distribution of the residual autocorrelations in the ARMA model is derived. The main advantage of this derivation over that of Box and Pierce (1970) is that it extends directly to more general situations. Generalizations of the derived distribution are presented for the residual autocorrelations in the multiplicative seasonal ARMA model and for the autocorrelations of a subseries of the residuals.

Keywords: ARMA MODEL; BOX–JENKINS MODELLING; MODEL CRITICISM; INTERVENTION ANALYSIS; MULTIPLICATIVE SEASONAL ARMA MODEL; RESIDUAL AUTOCORRELATIONS

1. INTRODUCTION

BOX AND JENKINS (1970) have described methods of fitting parametric time series models by model selection and estimation followed by model criticism through significance tests and diagnostic checks on the adequacy of the fitted model. The residual autocorrelations are useful at the model criticism stage since possible departures from the key assumption that the white noise disturbances in the specified model are uncorrelated may be detected. In Section 2 the large sample distribution of the residual autocorrelations in the autoregressive-moving average (ARMA) model is derived. This derivation is broadly similar to that of Box and Pierce (1970) and Durbin (1970), but nevertheless its advantage over these two approaches is that it extends directly to other important situations of interest such as SARMA and intervention analysis models. Also, McLeod (1977b) uses a method similar to that of Section 2 to obtain the large sample distribution of the residual cross-correlations in univariate ARMA models. The distribution of the residual autocorrelations in the multiplicative seasonal ARMA (SARMA) model is given in Section 3. An application to the intervention analysis model is described in Section 4.

2. RESIDUAL AUTOCORRELATIONS IN THE ARMA MODEL

2.1. Introduction

Suppose that the time series w_t , $t = 1, \dots, n$, is generated by the ARMA time series model

$$\phi(B)w_t = \theta(B)a_t, \quad (1)$$

where

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \quad \theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q,$$

and B is the backshift operator on t . The white noise series, a_t , is assumed to be independent and identically distributed with mean 0 and variance 1 and the ARMA model is assumed to be stationary, invertible and not redundant. Let $\beta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ be the $p+q$ dimensional vector of true model parameters and let $\hat{\beta} = (\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q)$ be any value in the admissible parameter space.

For any such $\hat{\beta}$, the estimated white noise series, \hat{a}_t , can be approximately calculated by

$$\hat{a}_t = \theta_1 \hat{a}_{t-1} + \dots + \theta_q \hat{a}_{t-q} + w_t - \phi_1 w_{t-1} - \dots - \phi_p w_{t-p}, \quad (p+1 \leq t \leq n), \quad (2)$$

† Present address: Statistics and Actuarial Science Group, University of Western Ontario, Canada.

where $\hat{a}_t = 0$ ($t \leq p$). Alternatively, a better approximation is obtained by the back-forecasting algorithm of Box and Jenkins (1970). If the true value, β , were known, the white noise series, a_t , could be calculated apart from a transient error (which is $O(\varepsilon^t)$ as $t \rightarrow \infty$ for some $0 < \varepsilon < 1$) from equation (2). The white noise autocorrelations,

$$r(l) = \frac{\sum_{t=1}^{n-l} a_t a_{t+l}}{\sum_{t=1}^n a_t^2} \quad (l = 1, 2, \dots), \tag{3}$$

are (Hannan, 1970, p. 229) asymptotically jointly normal with mean 0 and covariances

$$\text{cov}(r(l), r(k)) = \delta_{l,k}/n, \tag{4}$$

where $\delta_{l,k} = 1$ or 0 according as $l = k$ or $l \neq k$. Moreover it is easily seen that the effect of the aforesaid transient error on $r(l)$ is $O(1/n)$ as $n \rightarrow \infty$.

Let $\hat{\beta}$ denote an asymptotically efficient estimate of β such as the least squares estimate of Box and Jenkins (1970) and let \hat{a}_t denote the value of a_t when $\beta = \hat{\beta}$. Then, if the model is correct, the residual autocorrelations,

$$\hat{r}(l) = \frac{\sum_{t=1}^{n-l} \hat{a}_t \hat{a}_{t+l}}{\sum_{t=1}^n \hat{a}_t^2} \quad (l = 1, 2, \dots), \tag{5}$$

will have mean 0. On the other hand, for an incorrect model some of the residual autocorrelations will have nonzero means. This suggests that a test of whether some of the residual autocorrelations are significantly different from 0 will be useful in model criticism.

In large samples, it is well known (Hannan, 1970, pp. 345–348) that $\hat{\beta}$ is normally distributed with mean β and covariance matrix \mathbf{I}^{-1}/n where,

$$\mathbf{I} = \begin{pmatrix} \gamma_{vv}(i-j) & \gamma_{vu}(i-j) \\ \gamma_{uv}(i-j) & \gamma_{uu}(i-j) \end{pmatrix} \begin{matrix} p \\ q \end{matrix} \tag{6}$$

where the (i, j) element in each partitioned matrix is indicated and $\gamma_{vv}, \gamma_{uu}, \gamma_{vu}, \gamma_{uv}$ are the theoretical auto- and cross-covariances defined by

$$\phi(B)v_t = -a_t, \tag{7}$$

$$\theta(B)u_t = a_t, \tag{8}$$

$$\gamma_{vv}(k) = E(v_t v_{t+k}), \tag{9}$$

$$\gamma_{uu}(k) = E(u_t u_{t+k}), \tag{10}$$

$$\gamma_{vu}(k) = E(v_t u_{t+k}), \tag{11}$$

$$\gamma_{uv}(k) = \gamma_{vu}(-k). \tag{12}$$

2.2. Distribution of the Residual Autocorrelations

For any fixed $m \geq 1$, let

$$\mathbf{r} = (r(1), \dots, r(m)) \tag{13}$$

and

$$\hat{\mathbf{f}} = (\hat{r}(1), \dots, \hat{r}(m)). \tag{14}$$

Theorem 1. The large sample distribution of $\hat{\mathbf{f}}$ is normal with mean vector $\mathbf{0}$ and covariance matrix

$$\text{var}(\hat{\mathbf{f}}) = (\mathbf{I} - \mathbf{X}\mathbf{I}^{-1}\mathbf{X}^T)/n, \tag{15}$$

where $\mathbf{1}$ is the $m \times m$ identity matrix, \mathbf{I} is defined by equation (6) and

$$\mathbf{X} = \begin{pmatrix} -\phi'_{i-j} & \theta'_{i-j} \\ p & q \end{pmatrix} m, \tag{16}$$

where the coefficients ϕ' and θ' are defined by

$$1/\phi(B) = \sum_{i=0}^{\infty} \phi'_i B^i, \tag{17}$$

$$1/\theta(B) = \sum_{i=0}^{\infty} \theta'_i B^i \tag{18}$$

and $\phi'_l = \theta'_l = 0$ for $l < 0$. The coefficients ϕ'_l and θ'_l are readily computed using the recursive procedure of Box and Jenkins (1970, pp. 132-134).

The proof of Theorem 1 is based on the following three lemmas.

Lemma 1. For any fixed m ,

$$r(l) = c(l) + O_p(1/n) \quad (1 \leq l \leq m), \tag{19}$$

where

$$c(l) = \sum_{i=1}^{n-l} a_i a_{i+l} / n. \tag{20}$$

Proof. Since the means of $c(l)$ and $c(0)$ are 0 and 1 respectively and the asymptotic variances and covariance of $c(l)$ and $c(0)$ are $O(1/n)$, equation (19) follows from the Taylor series expansion of $r(l)$ as a function of $(c(l), c(0))$ about $(0, 1)$.

Lemma 2. Let $\hat{\beta}$ be the least squares estimate of β obtained by minimizing

$$S(\hat{\beta}) = \sum_{i=1}^n a_i^2. \tag{21}$$

Then

$$\hat{\beta} - \beta = \mathbf{I}^{-1} s_c + O_p(1/n), \tag{22}$$

where the i th element of the vector s_c is

$$s_{c,i} = \begin{cases} -\sum a_i v_{i-i} / n & (1 \leq i \leq p), \\ -\sum a_i u_{i-p-i} / n & (p+1 \leq i \leq p+q). \end{cases} \tag{23}$$

Proof. Let $\partial S / \partial \beta$ and $\partial S / \partial \hat{\beta}$ denote the vector of partial derivatives of $S(\hat{\beta})$ evaluated at β and $\hat{\beta}$ respectively. Since $\hat{\beta} - \beta$ is $O_p(1/\sqrt{n})$ and the partial derivatives of $S(\beta)$ of all orders are $O_p(1/\sqrt{n})$, it follows from the Taylor series expansion of $\partial S / \partial \hat{\beta}$ about β and evaluated at $\hat{\beta}$ that

$$\mathbf{0} = \frac{\partial S}{\partial \beta} + \frac{\partial^2 S}{\partial \beta \partial \beta^T} (\hat{\beta} - \beta) + O_p(1). \tag{24}$$

It may be shown that

$$\frac{1}{2} \frac{\partial^2 S}{\partial \beta \partial \beta^T} = n\mathbf{I} + O_p(\sqrt{n}). \tag{25}$$

Evaluating $\partial S / \partial \beta$ using,

$$v_{i-i} = \partial a_i / \partial \phi_i \tag{26}$$

and

$$u_{i-i} = \partial a_i / \partial \theta_i, \tag{27}$$

it is seen that (22) holds.

Lemma 3. The joint asymptotic distribution of $\sqrt{n}(\hat{\beta} - \beta, \mathbf{r})$ is normal with mean $\mathbf{0}$ and covariance matrix

$$\begin{pmatrix} \mathbf{I}^{-1} & -\mathbf{I}^{-1}\mathbf{X}^T \\ -\mathbf{X}\mathbf{I}^{-1} & \mathbf{1} \end{pmatrix} \begin{matrix} p+q \\ m \end{matrix} \tag{28}$$

where \mathbf{I} , \mathbf{X} and $\mathbf{1}$ are as in Theorem 1.

Proof. From Lemmas 1 and 2 any linear combination of the elements of $(\hat{\beta} - \beta, \mathbf{r})$ is, apart from terms $O_p(1/n)$, an average of a series of martingale differences and hence, by the martingale central limit (Billingsley, 1961), \sqrt{n} times this linear combination is asymptotically normal. This proves that $\sqrt{n}(\hat{\beta} - \beta, \mathbf{r})$ is asymptotically jointly normal.

The asymptotic covariance of $\sqrt{(n)}s_{c,i}$ ($1 \leq i \leq p$) and $\sqrt{(n)}c(j)$ ($1 \leq j \leq m$) is

$$\lim_{n \rightarrow \infty} nE(s_{c,i}c(j)) = -\lim_{n \rightarrow \infty} E(\sum_t a_t v_{t-i} a_s a_{s+j})/n = \gamma_{av}(j-i) = \phi'_{j-i}. \tag{29}$$

The last equality follows from a well-known fourth moment result (Hannan, 1970, p. 23). Similarly it may be shown that

$$\lim_{n \rightarrow \infty} nE(s_{c,i}c(j)) = -\theta'_{j-i} \quad (p+1 \leq i \leq p+q; 1 \leq j \leq m). \tag{30}$$

Thus the asymptotic covariance matrix of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{(n)}\mathbf{r}$ is $-\mathbf{I}^{-1}\mathbf{X}^T$. This proves Lemma 3.

Proof of Theorem 1. For any $\hat{\beta}$, let

$$\dot{r}(l) = \sum_{i=1}^{n-1} \dot{a}_i \dot{a}_{i+l} / \sum_{i=1}^n \dot{a}_i^2 \quad (1 \leq l \leq m), \tag{31}$$

where \dot{a}_i is defined in (2). It is easily shown that

$$\begin{aligned} \frac{\partial \dot{r}(i)}{\partial \phi_j} &= \sum a_i v_{i+i-j} + O_p(1/\sqrt{n}) \\ &= -\phi'_{i-j} + O_p(1/\sqrt{n}) \end{aligned} \tag{32}$$

and

$$\frac{\partial \dot{r}(i)}{\partial \theta_j} = \theta'_{i-j} + O_p(1/n). \tag{33}$$

By a Taylor series expansion of $(\dot{r}(1), \dots, \dot{r}(m))$ about $\hat{\beta} = \beta$ and evaluated at $\hat{\beta} = \hat{\beta}$ it follows from (32) and (33) that

$$\mathbf{f} = \mathbf{r} + \mathbf{X}(\hat{\beta} - \beta) + O_p(1/n). \tag{34}$$

Theorem 1 now follows from Lemma 3.

It can be algebraically demonstrated (McLeod, 1977a) that for m large, Theorem 1 is equivalent to the previous result of Box and Pierce (1970, pp. 1522-1525).

ARMA models with some of the model coefficients $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ constrained to 0 occasionally arise with seasonal data and with other applications. In this case the large sample covariance matrix of the residual autocorrelations may be shown to be

$$\text{var}(\mathbf{f}) = (\mathbf{1} - \mathbf{X}_0 \mathbf{I}_0^{-1} \mathbf{X}_0^T)/n, \tag{35}$$

where \mathbf{I}_0 is obtained from the matrix \mathbf{I} of equation (6) by deleting rows and columns corresponding to the constrained coefficients and similarly \mathbf{X}_0 is obtained from equation (16) by deleting columns corresponding to the constrained coefficients.

2.3. Application to Model Criticisms

The estimated large sample covariance matrix of the estimate $\hat{\beta}$, say V , is often generated by the nonlinear least squares algorithm. Alternatively V can be estimated by

$$V = \hat{I}^{-1}/n, \tag{36}$$

where \hat{I} denotes the value of I in equation (6) when $\beta = \hat{\beta}$. Thus the estimated large sample covariance matrix of f can be conveniently calculated from

$$\{\text{var}(f)\}_{\text{est}} = 1/n - \hat{X}V\hat{X}^T, \tag{37}$$

where \hat{X} denotes the value of X when $\beta = \hat{\beta}$.

Since often large values of $r(l)$ for some particular lag l will suggest model inadequency, a useful model criticism procedure is to plot $r(l)$ ($1 \leq l \leq m$) and their estimated 95 per cent confidence intervals. The estimated 95 per cent confidence interval for $r(l)$ is consequently $\pm 1.96 \times \sqrt{\{\text{var}(f(l))\}_{\text{est}}}$. In many situations (see, for example, Box and MacGregor, 1971; Hipel *et al.*, 1977; McLeod *et al.*, 1977; McLeod, 1977a) this approach yields more insight than the use of only the portmanteau test of Box and Jenkins (1970, p. 290).

If m is large enough so that $\phi'_i \doteq \theta'_i \doteq 0$ for $i > m$, $X^T X \doteq I$ and so $\text{var}(f)$ is idempotent and of rank $m - p - q$. Thus

$$Q_m = n \sum_{l=1}^m \hat{r}^2(l) \tag{38}$$

is $\chi^2(m - p - q)$. Box and Jenkins (1970) suggest using Q_m for a portmanteau test of model adequency. However, Ljung and Box (1976) and Davies *et al.* (1977) have found that this test is quite conservative in small samples (so that the chance of rejecting the null hypothesis of model adequacy is overestimated).

3. RESIDUAL AUTOCORRELATIONS IN THE SARMA MODEL

3.1. Introduction

The multiplicative seasonal ARMA or SARMA model of order $(p, q)(p_s, q_s)_s$ for the time series w_t , $t = 1, \dots, n$, is defined by

$$\Phi(B^s)\phi(B)w_t = \Theta(B^s)\theta(B)a_t, \tag{39}$$

where

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_{p_s} B^{sp_s},$$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_{q_s} B^{sq_s},$$

s is the length of the seasonal period and $\phi(B)$, $\theta(B)$ and a_t are defined as in Section 2.1. It is assumed that the model defined by (39) is stationary, invertible and not redundant. Box and Jenkins (1970) introduced the SARMA model for describing seasonal time series for which the seasonal component is stochastically rather than deterministically specified. The SARMA model has been found to provide a suitable model for many seasonal economic time series (Box *et al.*, 1976, Cleveland and Tiao, 1976).

When normality of a_t is assumed, the large sample information matrix per observation on the coefficients $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \Phi_1, \dots, \Phi_{p_s}, \Theta_1, \dots, \Theta_{q_s}$ is

$$I = \left(\begin{array}{c|c} I_1 & I_2 \\ \hline I_2^T & I_3 \end{array} \right), \tag{40}$$

where I_1 and I_3 are the information matrices, defined in equation (6), for the ARMA (p, q) and

ARMA (p_s, q_s) models with model parameters $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ and $\Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q$, respectively and

$$\mathbf{I}_2 = \begin{pmatrix} \gamma_{vV}(i-j) & \gamma_{vU}(i-j) \\ \gamma_{uV}(i-j) & \gamma_{uU}(i-j) \end{pmatrix} \begin{matrix} p \\ q \end{matrix} \quad (41)$$

where the (i, j) th element in each partitioned matrix is indicated and $\gamma_{vV}, \gamma_{vU}, \gamma_{uV}, \gamma_{uU}$ are the cross-covariance functions of the processes v_t, u_t, V_t, U_t defined by equations (7), (8),

$$\Phi(B^s)V_t = -a_t \quad (42)$$

and

$$\Theta(B^s)U_t = a_t \quad (43)$$

In Section 3.2 the results of Section 2.2 are generalized to yield the asymptotic distribution of the residual autocorrelations in SARMA models and in Section 3.3 the application of this new result is discussed.

3.2. Distribution of the Residual Autocorrelations

For any fixed $m \geq 1$, let $\mathbf{f} = (\mathbf{f}(1), \dots, \mathbf{f}(m))$ be the m -dimensional vector of residual autocorrelations in the SARMA model.

Theorem 2. The large sample distribution of \mathbf{f} is normal with mean vector $\mathbf{0}$ and covariance matrix

$$\text{var}(\mathbf{f}) = (\mathbf{1} - \mathbf{X}\mathbf{I}^{-1}\mathbf{X}^T)/n, \quad (44)$$

where \mathbf{I} is given by (39) and

$$\mathbf{X} = \begin{pmatrix} -\phi'_{i-j} & \theta'_{i-j} \\ \phi'_{i-j} & -\theta'_{i-j} \end{pmatrix} \begin{matrix} p \\ q \end{matrix} \begin{matrix} -\Phi'_{i-j} & \Theta'_{i-j} \\ \Phi'_{i-j} & -\Theta'_{i-j} \end{matrix} m, \quad (45)$$

where ϕ' and θ' are defined in (17) and (18) and Φ' and Θ' are defined by

$$1/\Phi(B^s) = \sum_{i=0}^{\infty} \Phi'_i B^i \quad (46)$$

and

$$1/\Theta(B^s) = \sum_{i=0}^{\infty} \Theta'_i B^i. \quad (47)$$

The detailed proof of this theorem is omitted since the required modifications of Lemmas 2 and 3 and the proof of Theorem 1 are straightforward. However, it should be noted that the method of Box and Pierce (1970) cannot be extended to the case of SARMA models when $s > 1$ because of the multiplicative constraints on the parameters.

3.3. Application to Model Criticism

The residual autocorrelations and their estimated confidence intervals can be plotted as described in Section 2.3.

In the SARMA model, the residual autocorrelations at lags $s, 2s, 3s, \dots$ are of special interest since these residual autocorrelations may indicate inadequacy of the seasonal component. It may be shown that if $p \ll s$ and $q \ll s$, and if the roots of the equation $\phi(z)\theta(z) = 0$ are not close to the unit circle, then the residual autocorrelations $(\hat{r}(s), \hat{r}(2s), \dots, \hat{r}(ms))$ have approximately the same covariance matrix as the first m residual autocorrelations in the nonseasonal model

$$\Phi(B)w_t = \Theta(B)a_t \quad (48)$$

Thus in the SARMA model of order $(0, 1)(0, 1)_{12}$,

$$\text{var}(\hat{\rho}(12)) \doteq \Theta_1^2/n \quad (49)$$

and

$$\text{var}(\hat{\rho}(24)) \doteq (1 - \Theta_1^2 + \Theta_1^4)/n, \quad (50)$$

provided that θ_1 is not close to ± 1 . Note that the variance of $\hat{\rho}(12)$ can be much less than $1/n$ (cf. Box and Pierce, 1970, p. 1516).

4. CONCLUDING REMARKS

The method given in Section 2 can be used to obtain the distribution of the autocorrelations of the residuals after the intervention in the intervention analysis model of Box and Tiao (1975). If the number of observations before and after the intervention, say, T and $n-T$ respectively, is large then the covariance matrix of the first m autocorrelations of the residuals after the intervention is

$$\mathbf{1}/(n-T) - \mathbf{X}\mathbf{I}^{-1}\mathbf{X}^T/n, \quad (51)$$

where \mathbf{X} and \mathbf{I} are calculated exactly as in Sections 2.2 and 3.2 from the ARMA and SARMA parameters which describe the autocorrelated disturbances in the intervention model. This result can be used to check the assumption that the intervention did not cause a change in the ARMA and SARMA component.

Computer programs for fitting Box-Jenkins models and calculating the residual autocorrelations and their estimated standard deviations are given in the forthcoming book by Hipel and McLeod (1979).

REFERENCES

- BILLINGSLEY, P. (1961). The Lindeberg-Levy theorem for martingales. *Proc. Amer. Math. Soc.*, **12**, 788-792.
- BOX, G. E. P., HILLMER, S. C. and TIAO, G. C. (1976). Analysis and modelling of seasonal time series. NBER-CENSUS Conference on Seasonal Time Series, Washington, D.C.
- BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control*. San Francisco: Holden-Day.
- BOX, G. E. P. and MACGREGOR, J. F. (1972). Letter to the Editor. *Technometrics*, **14**, 985-986.
- BOX, G. E. P. and PIERCE, D. A. (1970). Distribution of the residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Ass.*, **65**, 1509-1526.
- BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *J. Amer. Statist. Ass.*, **70**, 70-79.
- CLEVELAND, W. P. and TIAO, G. C. (1976). Decomposition of seasonal time series: a model for the Census X-11 Program. *J. Amer. Statist. Ass.*, **71**, 581-587.
- DAVIES, N., TRIGGS, C. M. and NEWBOLD, P. (1977). Significance levels of the Box-Pierce portmanteau statistic in finite samples. *Biometrika*, **64**, 517-522.
- DURBIN, J. (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*, **38**, 410-421.
- HANNAN, E. J. (1970). *Multiple Time Series*. New York: Wiley.
- HYPEL, K. W. and McLEOD, A. I. (1979). *Applied Box-Jenkins Modelling for Water Resources Engineers*. Amsterdam: Elsevier. Forthcoming.
- HYPEL, K. W., McLEOD, A. I. and McBEAN, E. A. (1977). Comment on "Hydrologic estimation and economic regret" by R. U. Jettmar and G. K. Young. *Water Resources Research*, **13**, 687-688.
- JIANG, G. M. and BOX, G. E. P. (1976). A modification of the overall χ^2 test for lack of fit in time series models. Technical Report 477, Department of Statistics, University of Wisconsin.
- McLEOD, A. I. (1977a). Topics in time series and econometrics. Ph.D. Thesis, University of Waterloo.
- (1977b). Distribution of the residual cross-correlations in univariate ARMA time series models. *J. Amer. Statist. Ass.* (in press).
- McLEOD, A. I., HYPEL, K. W. and LENNOX, W. C. (1977). Advances in Box-Jenkins modelling. Part two: Applications. *Water Resources Research*, **13**, 577-586.