Statistics & Probability Letters 13 (1992) 391–396
North-Holland

7 April 1992

A note on the Ljung–Box–Pierce portmanteau statistic with missing data

David S. Stoffer

Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA

Clélia M.C. Toloi

Institute of Mathematics and Statistics, University of São Paulo, São Paulo, Brazil 20570

Received March 1991
Revised May 1991

Abstract: The overall test for lack of fit for time series models proposed by Box and Pierce (1970) and Ljung and Box (1978) is modified to include the case when observations are missing. The missing data mechanism considered here is general and nonparametric.

Keywords: ARMA models, missing observations, residual autocorrelation, test for lack of fit.

1. Introduction

A method for verifying the adequacy of time series regression models is the portmanteau test proposed by Box and Pierce (1970). The test was subsequently modified by Ljung and Box (1978) in response to Davies, Triggs and Newbold (1977) who argued that for moderate sample lengths, the true significance levels were likely to be much lower than predicted by asymptotic theory. Ljung and Box (1978) concluded that the modified test provided an improved approximation that would be adequate for most practical purposes. Since then, the modified portmanteau test has become a popular aid in model diagnostics and is used extensively in practice.

Let $\hat{e}(1),\ldots,\hat{e}(n)$ be the standardized residuals from fitting a time series regression model, and let

$$\hat{r}(k) = \sum_{t=k+1}^{n} \hat{e}(t)\hat{e}(t-k) \Big/ \sum_{t=1}^{n} \hat{e}^2(t), \quad k = 1, 2, \ldots,$$

be their autocorrelations. If the model is correct, the Ljung–Box–Pierce Q-statistic

$$Q(\hat{r}) = n(n+2) \sum_{k=1}^{m} (n-k)^{-1} \hat{r}^2(k) \qquad (1.1)$$

is asymptotically distributed as χ^2 with $m-p$ degrees of freedom (d.f.) where p denotes the number of parameters in the model (if the model is autoregressive-moving average, then p is the sum of the autoregressive order and the moving average order); see Box and Pierce (1970) and Ljung and Box (1978) for details. This approximation yields a general test for lack of fit.

In practice, one often encounters time series data where observations are missed. Various methods for fitting time series models to data with missing observations exist (Jones, 1980; Dunsmuir and Robinson, 1981a; Shumway and Stoffer, 1982; Harvey and Pierce, 1984). These methods can be used in conjunction with overfitting and with information based goodness-of-fit criteria such as AIC or BIC that address the question of the model order in fitting autoregressive-mov-

ing average models to time series data. The purpose of this note is to recommend a missing data modification of the Ljung–Box–Pierce portmanteau test for checking whether the residuals from a model fit are white.

2. Missing data mechanism

Let $\{y(t)\}$ denote the data, and let $\{a(t)\}$ be a stochastic sequence such that

$$a(t) = \begin{cases} 0 & \text{if } y(t) \text{ is missed at time } t, \\ 1 & \text{if } y(t) \text{ is observed at time } t. \end{cases}$$

Let $y(t \mid t-1)$ be the prediction of $y(t)$ based on the observations available at time $t-1$, and $e(t)$ be the standardized innovation,

$$e(t) = \sigma^{-1}(t)\{y(t) - y(t \mid t-1)\},$$

where

$$\sigma^2(t) = \operatorname{Var}\{y(t) - y(t \mid t-1)\}.$$

We assume that $e(t)$ has finite fourth moment. In this way, we define the set of standardized innovations in the missing data problem as

$$z(t) = a(t)e(t).$$

This technique of accounting for missing data was first introduced in Parzen (1962); his primary objective was to perform spectral analysis with missing or irregular observations. Since then, a number of authors have used this technique in the spectral domain (see Dunsmuir and Robinson, 1981a, for a review) and a few have used this methodology successfully in the time domain (Dunsmuir and Robinson, 1981a; Stoffer, 1986).

Define

$$C_a(k) = (n-k)^{-1} \sum_{t=k+1}^{n} a(t)a(t-k), \qquad k = 0, 1, \ldots,$$

and assume throughout that as $n \to \infty$, $C_a(k) \to \theta_a(k) \neq 0$ almost surely, for each finite $k = 0, 1, \ldots$. For the $\{z(t)\}$ process, we define

$$C_z(k) = n^{-1} \sum_{t=k+1}^{n} z(t)z(t-k), \quad k = 0, 1, \ldots.$$

Parzen (1962) and Dunsmuir and Robinson (1981a,b) suggested that the autocorrelations $\rho_e(k) = \operatorname{corr}\{e(t), e(t-k)\}$ be estimated by

$$r_e(k) = C_e(k)/C_e(0)$$

where

$$C_e(k) = C_z(k)/C_a(k)$$

provided that $C_a(k) \neq 0$. Note that if there are no missing observations then $C_a(k) = 1$ for all k, and $r_e(k)$ corresponds to the usual definition of sample autocorrelation.

If the model is correct, then under the conditions stated above, $n^{1/2}r_e(k)$, for $k \geqslant 1$, are asymptotically independent normal random variables with mean 0 and variance $\theta_a^{-1}(k)$ (Dunsmuir and Robinson, 1981b, Corollary 2). From this result and following Box and Pierce (1970) we could establish a portmanteau χ^2-statistic in the missing data case based on the the fact that the distribution of

$$Q^*(r_e) = n \sum_{k=1}^{m} C_a(k) r_e^2(k) \tag{2.1}$$

is approximately χ_m^2. However, heeding the warnings of Davies, Triggs and Newbold (1977) and Ljung and Box (1978), we investigate finite sample properties of such a quantity.

3. Mean and variance of $r_e(k)$

To investigate the finite sample properties of $r_e(k)$, we consider approximating sequences of expectations via Taylor expansions (Fuller, 1976, Section 5.4). If $\{\boldsymbol{x}_n\}$ is a sequence of q-dimensional random vectors and g is a function mapping $\mathbb{R}^q$ into $\mathbb{R}$, then under appropriate conditions (that can be easily verified for this problem — see Fuller, 1976, Theorem 5.4.3)

$$E\{g(\boldsymbol{x}_n)\} = g(\boldsymbol{\mu}) + E\{D^2 g(\boldsymbol{\mu}) \cdot \tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^2\} + O(\boldsymbol{\xi}_n^3) \tag{3.1}$$

where $\boldsymbol{\mu} = E(\boldsymbol{x}_n)$, $\boldsymbol{\xi}_n^3 = E \mid \boldsymbol{x}_n - \boldsymbol{\mu} \mid^3$, and

$$D^2 g(\boldsymbol{\mu}) \cdot (\boldsymbol{x} - \boldsymbol{\mu})^2 = \sum_{i_1=1}^{q} \sum_{i_2=1}^{q} g^{(i_1,i_2)}(\boldsymbol{\mu})(x_{i_1} - \mu_{i_1})(x_{i_2} - \mu_{i_2})$$

where

$$g^{(i_1,i_2)}(\boldsymbol{\mu}) = \partial^2 g(\boldsymbol{x})/\partial x_{i_1}\,\partial x_{i_2}\,|_{\boldsymbol{x}=\boldsymbol{\mu}}.$$

We further assume that $\theta_a(k) = E\{a(t)a(t-k)\}$ (that is, $C_a(k)$ is unbiased for $\theta_a(k)$), and that the sequence $\{a(t)\}$ and $\{e(t)\}$ are independent. First we evaluate $\mathrm{Cov}\{r_e(h), r_e(k)\}$ via (3.1). Let

$$\boldsymbol{x}_n = \big(C_z(h), C_z(k), C_z(0), C_a(h), C_a(k), C_a(0)\big)',$$

let $g(\boldsymbol{x}) = (x_1x_2x_6^2)/(x_3^2x_4x_5)$, and note that $g(\boldsymbol{\mu}) = 0$. To evaluate the second term in the expansion (3.1), the only terms that are non-zero have

$$g^{(1,2)}(\boldsymbol{\mu}) = g^{(2,1)}(\boldsymbol{\mu}) = \theta_a^2(0)/\{\theta_a(h)\theta_a(k)\gamma_z^2(0)\}$$

where $\gamma_z(0) = E\{C_z(0)\} = \theta_a(0)$. Hence,

$$\mathrm{Cov}\{r_e(h), r_e(k)\} = \{\theta_a(h)\theta_a(k)\}^{-1}\,\mathrm{Cov}\{C_z(h), C_z(k)\} + \mathrm{O}(\xi_n^3).$$

Since the $e(t)$'s are independent, we have that $\mathrm{Cov}\{C_z(h), C_z(k)\} = 0$ for $h \neq k$. When $h = k \geqslant 1$, note that

$$E\left\{\sum_{t=k+1}^{n} a(t)a(t-k)e(t)e(t-k)\right\}^2 = (n-k)\theta_a(k)$$

so that

$$\mathrm{Var}\{C_z(k)\} = (n-k)\theta_a(k)/n^2.$$

Next, expanding $\{r_e(h)r_e(k)\}$ through third order terms and using Fuller (1976, Theorem 5.4.1), we may establish that the expected value of the third order moment is $\mathrm{O}(n^{-2})$. Combining these results we have

$$\mathrm{Var}\{r_e(k)\} = (n-k)\theta_a^{-1}(k)n^{-2} + \mathrm{O}(n^{-2}), \tag{3.2}$$

while $\mathrm{Cov}\{r_e(k), r_e(h)\} = \mathrm{O}(n^{-2})$ for $h \neq k \geqslant 1$. In a similar manner, we may use (3.1) to establish the fact that $E\{r_e(k)\} = \mathrm{O}(n^{-2})$ for all finite $k \geqslant 1$.

4. The modified portmanteau statistic

In view of (3.2) and following Box and Pierce (1970) and Ljung and Box (1978), we define the quantity

$$Q(r_e) = n^2 \sum_{k=1}^{m} (n-k)^{-1} C_a(k) r_e^2(k),$$

which, based on our assumptions, has an asymptotic χ_m^2 distribution.

Now let $\hat{e}(t)$ be the innovation based on the parameter estimates obtained from the model fit and define the set of estimated residuals in the missing data problem as

$$\hat{z}(t) = a(t)\hat{e}(t),$$

in a manner analogous to Section 2. Continuing the analogy, put

$$\hat{C}_z(k) = n^{-1} \sum_{t=k+1}^{n} \hat{z}(t)\hat{z}(t-k)$$

and

$$\hat{r}_e(k) = \hat{C}_e(k)/\hat{C}_e(0),$$

where $\hat{C}_e(k) = \hat{C}_z(k)/C_a(k)$ provided that $C_a(k) \neq 0$. Hence, define the statistic

$$Q(\hat{r}_e) = n^2 \sum_{k=1}^{m} (n-k)^{-1} C_a(k) \hat{r}_e^2(k). \tag{4.1}$$

As in the case of Ljung–Box–Pierce statistic, the asymptotic null distribution of (4.1) is χ^2 with $m-p$ degrees of freedom. This can be shown by following the Taylor expansion argument of Box and Pierce (1970, Section 2), and we will sketch the details shortly. But first, note that if there are no missing observations (that is, $a(t) = 1$ for all t), then $C_a(k) = 1$, $\hat{r}_e^2(k) = \hat{r}^2(k)$, but the multiplier of the weighted sum of autocorrelations is n^2 rather than the familiar $n(n+2)$ that appears in the original Q-statistic, (1.1). The reason for this difference is that (1.1) is based on normal theory whereas (4.1) is not. Nevertheless, the asymptotic properties of the statistics $Q(\hat{r})$ and $Q(\hat{r}_e)$ are the same in this case.

To establish the null distribution of (4.1) we

restrict attention to the case of autoregressive models. Expanding $\hat{r}_e(k)$ we have

$$\hat{r}_e(k) = r_e(k) + \sum_{j=1}^{p} \left(\phi_j - \hat{\phi}_j\right)\hat{\delta}_{jk} + O_p(n^{-1})$$

where $\hat{\delta}_{jk}$ is the negative of the partial derivative of $r_e(k)$ with respect to $\hat{\phi}_j$. Using the asymptotic equivalence of $r_e(k)$ and $r(k)$ (Dunsmuir and Robinson, 1981b, Theorem 1) and the consistency of $\hat{\phi}_j$ for ϕ_j we obtain the same approximation for $\hat{\delta}_{jk}$ as described in Box and Pierce (1970), namely, $\hat{\delta}_{jk} \approx \psi_{k-j}$, where the ψ_{k-j} are the usual ψ-weights obtained by writing the autoregressive process in terms of an infinite order moving average.

Define the $m \times m$ matrix $\boldsymbol{C}_a = \text{diag}(C_a(1), \ldots, C_a(m))$, and $m \times 1$ vectors $\hat{\boldsymbol{r}}_e = (\hat{r}_e(1), \ldots, \hat{r}_e(m))$, and $\boldsymbol{r}_e = (r_e(1), \ldots, r_e(m))$, so that approximately, for large n,

$$\boldsymbol{C}_a^{1/2}\hat{\boldsymbol{r}}_e = \boldsymbol{C}_a^{1/2}\boldsymbol{r}_e + \boldsymbol{C}_a^{1/2}\boldsymbol{X}\left(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}\right) \tag{4.2}$$

where $\boldsymbol{\phi}$ is the $p \times 1$ vector of parameters, and $\boldsymbol{X}$ is an $m \times p$ matrix whose jkth element is ψ_{k-j} where $\psi_{k-j} = 0$ if $k - j < 0$.

If the model is correct, the $\hat{z}(t)$'s are asymptotically uncorrelated so that $\hat{r}_e(k) \to 0$ almost surely as $n \to \infty$ for $k > 0$. Thus, since $C_a(k) \to \theta_a(k)$ almost surely as $n \to \infty$ we have that, approximately, for large n,

$$\boldsymbol{X}'\boldsymbol{C}_a\hat{\boldsymbol{r}}_e = \boldsymbol{0}. \tag{4.3}$$

Define the $m \times p$ matrix $\boldsymbol{A} = \boldsymbol{C}_a^{1/2}\boldsymbol{X}$. Multiplying (4.2) by $\boldsymbol{D} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$, in view of (4.3), establishes that, approximately, for large n, $\boldsymbol{C}_a^{1/2}\hat{\boldsymbol{r}}_e = [\boldsymbol{I} - \boldsymbol{D}]\boldsymbol{C}_a^{1/2}\boldsymbol{r}_e$. Since $\boldsymbol{C}_a^{1/2}\boldsymbol{r}_e \sim \text{AN}(\boldsymbol{0}, n^{-1}\boldsymbol{I})$, where AN denotes asymptotic normality, it follows that $\boldsymbol{C}_a^{1/2}\hat{\boldsymbol{r}}_e \sim \text{AN}(\boldsymbol{0}, n^{-1}[\boldsymbol{I} - \boldsymbol{D}])$, where $\boldsymbol{I} - \boldsymbol{D}$ is idempotent of rank $m - p$. It then follows that the approximate distribution of $Q(\hat{r}_e)$ is χ^2 with $m - p$ degrees of freedom.

5. Empirical studies

To study the null behavior of $Q(\hat{r}_e)$ for moderate sample sizes, Monte Carlo studies were performed on AR(1) models based on 1000 replications with $n = 100$, and $m = 21$ (i.e., 20 d.f.). For each case, the $a(t)$ were independent Bernoulli random variables with $\Pr\{a(t) = 0\} = 0.25$, that is, on the average, 25% of the data were missing. Parameter estimation was accomplished using the missing data techniques of Shumway of Stoffer (1982). Figures 1, 2, and 3 compare the empirical distribution function of $Q(\hat{r}_e)$ with the χ^2_{20} distribution when the autoregressive parameter is $\phi = 0.3$, 0.5 and 0.8, respectively. For contrast, the

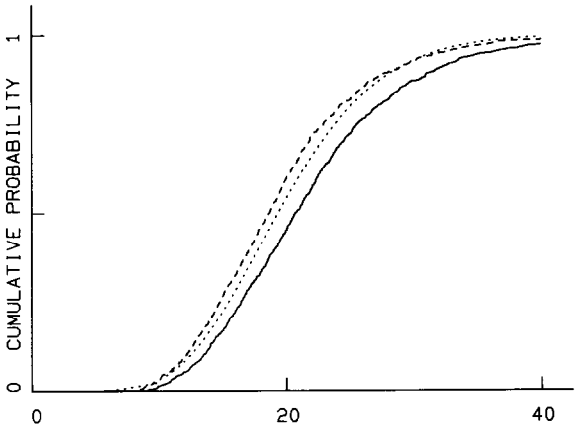

Fig. 1. Monte Carlo distribution of $Q(\hat{r}_e)$ for an AR(1) model, $\phi = 0.3$, based on 1000 replications, $n = 100$, $m - p = 20$, and 25% missing data (solid line); distribution function of χ^2_{20} (dotted line); corresponding empirical distribution of $Q^*(\hat{r}_e)$ (dashed line).

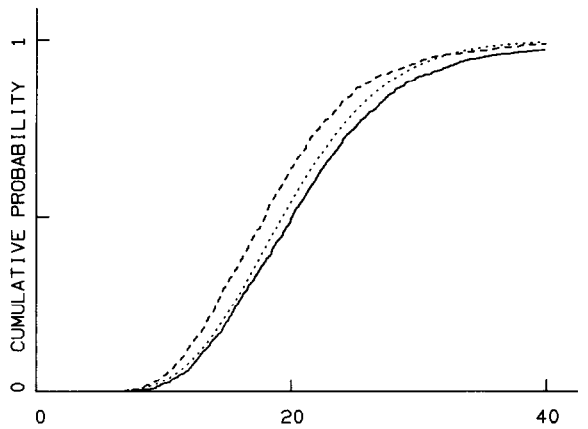

Fig. 2. Monte Carlo distribution of $Q(\hat{r}_e)$ for an AR(1) model, $\phi = 0.5$, based on 1000 replications, $n = 100$, $m - p = 20$, and 25% missing data (solid line); distribution function of χ^2_{20} (dotted line); corresponding empirical distribution of $Q^*(\hat{r}_e)$ (dashed line).

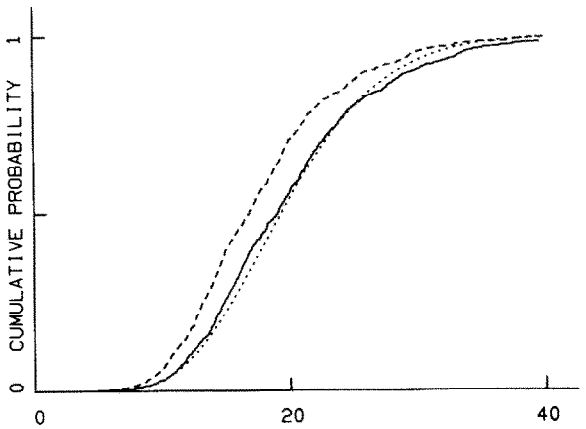

Fig. 3. Monte Carlo distribution of $Q(\hat{r}_e)$ for an AR(1) model, $\phi = 0.8$, based on 1000 replications, $n = 100$, $m - p = 20$, and 25% missing data (solid line); distribution function of χ^2_{20} (dotted line); corresponding empirical distribution of $Q^*(\hat{r}_e)$ (dashed line).

empirical distribution function of $Q^*(\hat{r}_e)$ given in (2.1) is also shown in each figure. The Monte Carlo results for $Q(\hat{r}_e)$ are similar to those reported by Ljung and Box (1978) for the non-missing data case and we draw the same conclusion; that is, the approximation should be quite adequate for most practical purposes. We also note that in these simulations, the true upper-tail probabilities (significance values) of $Q(\hat{r}_e)$ are slightly larger than predicted by asymptotic theory. This is in contrast to the simulation results of Ljung and Box (1978) and Davies, Triggs and Newbold (1977) for the non-missing data case, where (1.1) was likely to be smaller than predicted by asymptotic theory. Furthermore, we note that the values of $Q^*(\hat{r}_e)$ are typically smaller than predicted by asymptotics.

Table 1
Power of $Q(\hat{r}_e)$ assuming an AR(1) model when the true model is AR(2) with real and with complex roots, based on 1000 replications, $m = 21$, and 25% missing data

n	level	Real roots	Complex roots
100	0.05	58.4%	62.4%
	0.01	41.1%	45.6%
200	0.05	88.8%	93.0%
	0.01	79.8%	82.3%

Finally, to investigate the power of $Q(\hat{r}_e)$ to detect departures from the null, we simulated two AR(2) models, one with real roots ($\phi_1 = 0.4$, $\phi_2 = 0.55$), and one with complex roots ($\phi_1 = 1$, $\phi_2 = -0.5$). We then fit AR(1) models to the generated data and counted the number of times that $Q(\hat{r}_e)$ with $m = 21$ rejected the AR(1) model (based on the χ^2_{20} distribution) for 1000 replications; in each case we used independent Bernoulli $a(t)$'s with $\Pr\{a(t) = 0\} = 0.25$. Table 1 shows the results of the simulations for $n = 100$ and 200, with significance levels of 0.01 and 0.05. Again we are lead to the same conclusion: for moderate sample sizes, the statistic $Q(\hat{r}_e)$ is adequate for most practical purposes.

Acknowledgements

The work of the first author was supported in part by National Science Foundation Grant DMS-9000522, the Centers for Disease Control through a cooperative agreement with the Association for Schools of Public Health, and Projecto BID-USP for travel support to Brazil. The work of the second author was supported in part by FAPESP-Brazil through travel support to the USA. The authors would also like to thank Sergio M. Koyama for help with the simulations. Finally, we acknowledge the comments of a referee that significantly improved the presentation of this article.

References

Box, G.E.P. and D.A. Pierce (1970), Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, *J. Amer. Statist. Assoc.* **65**, 1509–1526.

Davies, N., C.M. Triggs and P. Newbold (1977), Significance levels of the Box–Pierce portmanteau statistic in finite samples, *Biometrika* **64**, 517–522.

Dunsmuir, W. and P.M. Robinson (1981a), Estimation of time series models in the presence of missing data, *J. Amer. Statist. Assoc.* **76**, 560–568.

Dunsmuir, W. and P.M. Robinson (1981b), Asymptotic theory for time series containing missing and amplitude modulated observations, *Sankhyā Ser. A* **43**, 260–281.

Fuller, W.A. (1976), *Introduction to Statistical Time Series* (Wiley, New York).

Harvey, A.C. and R.G. Pierce (1984), Estimating missing observations in economic time series, *J. Amer. Statist. Assoc.* **79**, 125–131.

Jones, R.H. (1962), Spectral analysis with regularly missed observations, *Ann. Math. Statist.* **33**, 455–461.

Jones, R.H. (1980), Maximum likelihood fitting of ARMA models to time series with missing data, *Technometrics* **22**, 389–396.

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

Parzen, E. (1962), Spectral analysis of asymptotically stationary time series, *Bull. Internat. Statist. Inst.* **39**, 87–103.

Parzen, E. (1963), On spectral analysis with missing observations and amplitude modulation, *Sankhyā Ser. A* **25**, 383–392.

Shumway, R.H. and D.S. Stoffer (1982), An approach to time series smoothing and forecasting using the EM algorithm, *J. Time Series Anal.* **3**, 253–264.

Stoffer, D.S. (1986), Estimation and identification of space-time ARMAX models in the presence of missing data, *J. Amer. Statist. Assoc.* **81**, 762–772.