Journal of the Korean Data & Information Science Society
2016, **27**(2), 559–564

http://dx.doi.org/10.7465/jkdi.2016.27.2.559
한국데이터정보과학회지

Remarks on correlated error tests†

Tae Yoon Kim[1] · Jeongcheol Ha[2]

[12]Department of Statistics, Keimyung University

Received 19 February 2016, revised 21 March 2016, accepted 22 March 2016

Abstract

The Durbin-Watson (DW) test in regression model and the Ljung-Box (LB) test in ARMA (autoregressive moving average) model are typical examples of correlated error tests. The DW test is used for detecting autocorrelation of errors using the residuals from a regression analysis. The LB test is used for specifying the correct ARMA model using the first some sample autocorrelations based on the residuals of a fitted ARMA model. In this article, simulations with four data generating processes have been carried out to evaluate their performances as correlated error tests. Our simulations show that the DW test is severely dependent on the assumed AR(1) model but isn't sensitive enough to reject the misspecified model and that the LB test reports lackluster performance in general.

Keywords: Correlated error, Durbin-Watson test, Ljung-Box test, model specification.

1. Introduction

In a regression model, a basic assumption is that error terms in the regression model are independent. When this assumption-among others-is satisfied, the regression function recovery procedure such as least square method is valid and a great deal of its use has been made. In other words, it might be said that the regression function recovery procedures are eventually meant to recover the independence of residuals. Along this line of argument, whether the residuals after a given model fitting achieves independence has been regarded as a critical step for proper model specification. There have been two popular methods checking residual independence in the literatures; the Durbin-Watson (DW) test in linear regression model and the Ljung-Box (LB) test in ARMA (autoregressive moving average) model (Ljung and Box, 1978). The DW test is established in the residuals from a regression analysis. The small sample distribution of this test was derived by on Neumann (1941) under the null. Durbin and Watson (1950, 1951) applied this test to the residuals from least square regressions, and developed a bound test for the null hypothesis that the errors are serially uncorrelated against the alternative that they follow a first order autoregressive process. The LB test is commonly used in ARMA modeling. It is applied to the residuals

† This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A3A01020028)

[1] Professor, Department of Statistics, Keimyung University, Daegu 42601, Korea.
[2] Corresponding author: Associate professor, Department of Statistics, Keimyung University, Daegu 42601, Korea. E-mail: jeicy@kmu.ac.kr

of a fitted ARMA model, not the original series, and in such applications the hypothesis actually being tested is that the residuals from the ARMA model have no autocorrelation. Since autocorrelation makes the underlying model difficult to discern, various efforts have been conducted (Hwang, 2014; Ha, 2015).

The DW test calculates its asymptotic distribution under $AR(1)$ error assumption while the LB test calculates the null distributions under the iid error. For the LB test, it has been frequently noticed that asymptotic distribution has less variance than expected. See, for instance, Pena and Rodriguez (2002).

In this article, we are mainly concerned about investigating the strength or weakness of these two tests via simulations. Section 2 discusses the DW test and the LB test separately. Section 3 provides simulation works supporting our findings.

2. Correlated error tests

2.1. Durbin-Watson test

To derive a suitable test criterion, it is important to consider the set of alternative hypotheses against which it is desired to discriminate. The concerning alternative for DW test is the stationary Markov process

$$\epsilon_i = \rho\epsilon_{i-1} + u_i \ (i = \cdots, -1, 0, 1, \cdots) \tag{2.1}$$

where $|\rho| < 1$ and u_i is normal with mean zero and variance σ_u^2 and is independent of $u_{i-1}, u_{i-2}, \cdots$. The null hypothesis is then the hypothesis that $\rho = 0$ in (2.1). It is known that for certain regression problems with error distributions close to that given by (2.1) tests can be obtained which are uniformly most powerful against one-sided alternatives and which give regions for two-sided alternatives. The DW is derived based on such theoretical background (Durbin and Watson, 1950; 1951).

To test for positive autocorrelation at significance level α, that is $H_0 : \rho = 0$ versus $H_1 : \rho > 0$, the test statistic which is calculated by the residuals $\{e_i\}$

$$d_n = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

is compared to lower and upper critical values ($d_{L,\alpha}^{(n)}$ and $d_{U,\alpha}^{(n)}$):

- If $d_n < d_{L,\alpha}^{(n)}$, there is statistical evidence that the error terms are positively autocorrelated.
- If $d_n > d_{U,\alpha}^{(n)}$, there is no statistical evidence that the error terms are positively autocorrelated.
- If $d_{L,\alpha}^{(n)} < d_n < d_{U,\alpha}^{(n)}$, the test is inconclusive.

Positive serial correlation is serial correlation in which a positive error for one observation increases the chances of a positive error for another observation.

To test for negative autocorrelation at significance α, that is $H_0 : \rho = 0$ versus $H_1 : \rho < 0$, the test statistic $(4 - d_n)$ is compared to lower and upper critical values ($d_{L,\alpha}^{(n)}$ and $d_{U,\alpha}^{(n)}$):

- If $(4 - d_n) < d_{L,\alpha}^{(n)}$, there is statistical evidence that the error terms are negatively autocorrelated.
- If $(4 - d_n) > d_{U,\alpha}^{(n)}$, there is no statistical evidence that the error terms are negatively autocorrelated.
- If $d_{L,\alpha}^{(n)} < (4 - d_n) < d_{U,\alpha}^{(n)}$, the test is inconclusive.

Negative serial correlation implies that a positive error for one observation increases the chance of a negative error for another observation and a negative error for one observation increases the chances of a positive error for another.

The critical values, $d_{L,\alpha}^{(n)}$ and $d_{U,\alpha}^{(n)}$ vary by the level of significance (α), number of observations, and number of predictors in the regression equation. Their derivation under the null is complex and statisticians typically obtain them from the appendices of statistical texts. Actually these two significance points or bounds are obtained by quite accurate approximation to the distribution of d_n. The bounds are known to be 'best' in two senses: first they can be attained, and secondly when they are attained the test criterion adopted is uniformly most powerful against suitable alternative hypotheses.

2.2. Ljung-Box test

Given a discrete time series $Z_t, Z_{t-1}, Z_{t-2}, \cdots$ and using B for the backward shift operator such that $BZ_t = Z_{t-1}$, a general autoregressive integrated moving average (ARIMA) model of order (p, r, q) may be written

$$\phi(B)\nabla^r Z_t = \theta(B)\epsilon_t \tag{2.2}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$, ϵ_t is a sequence of independent errors with common variance σ_u^2, to be referred to as 'white noise', and where the roots of $\phi(B) = 0$ and $\theta(B) = 0$ lie outside the unit circle. In other words, if $X_t = \nabla^r Z_t = (1 - B)^r Z_t$, is the rth difference of the series Z_t, then X_t is a stationary, invertible, mixed autoregressive moving average process given by

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{2.3}$$

and permitting $r > 0$ allows the original series to be (homogeneously) nonstationary. In some instances the model (2.2) will be appropriate after a suitable transformation is made on Z; in others Z may represent the noise structure after allowing for some systematic model (Box and Pierce, 1970).

After a model of this form has been fitted to a series $X_1, \cdots, X_n$, it is useful to study the adequacy of the fit by examining the residuals $e_1, \cdots, e_n$, and, in particular, their autocorrelations

$$\hat{r}_k = \sum_{t=k+1}^{n} e_t e_{t-k} \Big/ \sum_{t=1}^{n} e_t^2 \quad (k = 1, 2, \cdots).$$

Note that $\hat{r}_k$ is obtained as a least square estimator of r_k from

$$e_i = r_k e_{i-k} + u_i \ (i = \cdots, -1, 0, 1, \cdots, k = 1, 2, \cdots) \tag{2.4}$$

An informal graphical analysis of these quantities combined with overfitting is usually known to be most effective in detecting possible deficiencies in the model. Furthermore, it is known that when the $p+q$ parameters of an appropriate model are estimated, then under the null hypothesis for the autocorrelations of $\{\epsilon_t\}$, that is, $\rho_1 = \rho_2 = \cdots = \rho_m = 0$,

$$Q_n(\hat{r}) = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^2 \;\; \Rightarrow \;\; \chi^2_{m-p-q} \tag{2.5}$$

yielding an approximate test for lack of fit. Note that asymptotic distribution derivation of $Q_n(\hat{r})$ does not depend on the estimation method employed for ARIMA.

3. Simulation

In this section we perform simulations in order to see the behavior of the test statistics as a measure for model specification. That is, null hypothesis is the imposed model explains the DGP correctly rather than the original hypothesis of DW and LB. Our simulation results will provide useful insights into error correlation test problems. Basically we are concerned in applying and comparing two error correlation tests; DW test d_n and LB test Q_n. We set m in Q_n as a minimum of 10 and half of the sample size, which is large enough to cover the order of autocorrelation concerned.

We consider four cases and each cases are summarized in Table 3.1. (i) linear regression model is correctly imposed but data generating process (DGP) is contaminated with correlated errors (Table 3.2). (ii) linear regression model is wrongly imposed when DGP is generated by a quadratic function plus error (Table 3.3). (iii) AR(1) model is correctly imposed on the underlying DGP (Table 3.4). (iv) AR(1) model is wrongly imposed on the underlying DGP (Table 3.5).

In view of model specification we consider that (i) and (iii) are the cases when null hypothesis holds and that (ii) and (iv) are the cases when alternative hypothesis holds. From the relation with DGP and imposed model, Tables 3.2 and 3.3 report the cases where DW is expected to function and Tables 3.4 and 3.5 report the cases where LB is expected to function. We use R 3.1.2 and experimental simulations are repeated 1000 times at sample size $n = 10, 30, 50$ and 100 in each case.

Table 3.1 DGP and imposed model for each case.

	DGP	Imposed model
(i)	$Y_t = 1 + 2x_t + \epsilon_t$ with $\epsilon_t = \phi\epsilon_{t-1} + u_t$	$Y_t = a + bx_t + \epsilon_t$
(ii)	$Y_t = 1 - x_t + x_t^2 + u_t$	$Y_t = a + bx_t + \epsilon_t$
(iii)	$X_t = \phi X_{t-1} + u_t$	$X_t = a + bX_{t-1} + \epsilon_t$
(iv)	$X_t = \phi X_{t-1} + u_t + \theta u_{t-1}$	$X_t = a + bX_{t-1} + \epsilon_t$

Table 3.2 Simulated power and size comparison of d_n and Q_n for testing correlated error at $\alpha = 0.05$ when linear regression model is imposed for data generating process $Y_t = 1 + 2x_t + \epsilon_t$ where $\epsilon_t = \phi\epsilon_{t-1} + u_t$ and iid $u_t \sim N(0, \sigma^2)$.

		d_n			Q_n		
σ	n	$\phi = -0.5$	$\phi = 0$	$\phi = 0.5$	$\phi = -0.5$	$\phi = 0$	$\phi = 0.5$
0.1	10	0.309	0.058	0.259	0.383	0.16	0.218
	30	0.821	0.048	0.753	0.633	0.125	0.453
	50	0.948	0.054	0.948	0.817	0.125	0.676
	100	1	0.054	0.999	0.984	0.099	0.958
1	10	0.288	0.048	0.251	0.375	0.176	0.223
	30	0.812	0.062	0.794	0.604	0.122	0.499
	50	0.957	0.04	0.94	0.811	0.103	0.693
	100	0.999	0.055	1	0.977	0.097	0.961

At Table 3.2, DGP is $Y_t = 1 + 2x_t + \epsilon_t$ where $x_t = t/n$, $\epsilon_t = \phi\epsilon_{t-1} + u_t$ and $u_t \sim N(0, \sigma^2)$ is iid. We generate Y_t for $\phi = -0.5, 0, 0.5$ and $\sigma = 0.1, 1$. Note here that $\phi = 0$ corresponds to iid error. A classical linear regression analysis is done for practicing DW d_n and LB Q_n in checking correlated error with significance level $\alpha = 0.05$. Table 3.2 shows that DW is very sensitive to correlated error ($\phi = -0.5, 0.5$) and rejects the correctly imposed model whenever it is contaminated with the correlated error, which is normal in the ordinary hypothesis of DW. In other words, DW accepts the correctly imposed model only when the true DGP is with iid error ($\phi = 0$). LB reports similar results but its level of performance is not so precise as DW.

Table 3.3 Simulated power comparison of d_n and Q_n for testing correlated error at $\alpha = 0.05$ when linear regression model is imposed and data generating process is $Y_t = 1 - x_t + x_t^2 + u_t$ with *iid* $u_t \sim N(0, \sigma^2)$.

	d_n			Q_n		
n	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$	$\sigma = 0.1$	$\sigma = 1$	$\sigma = 2$
10	0.361	0.047	0.048	0.182	0.171	0.188
30	0.674	0.06	0.048	0.255	0.125	0.118
50	0.834	0.05	0.037	0.66	0.096	0.117
100	0.975	0.075	0.047	0.995	0.105	0.097

At Table 3.3, DGP is $Y_t = 1 - x_t + x_t^2 + u_t$ where $x_t = t/n$ and $u_t \sim N(0, \sigma^2)$ is iid. We generate Y_t for $\sigma = 0.1, 1, 2$. Note here that correlated error is not considered. Again, a classical linear regression analysis is done for practicing DW d_n and LB Q_n with significance level $\alpha = 0.05$. Table 3.3 shows that DW is not sensitive enough to reject the wrongly imposed model particularly when error variance σ^2 is relatively large ($\sigma^2 = 1, 4$). LB reports similar results but its performance is relatively less misleading (accurate) than DW when σ^2 is large (small). This shows that checking error correlatedness would not be a good measure for model specification.

Table 3.4 Simulated power comparison of d_n and Q_n for testing correlated error at $\alpha = 0.05$ when AR(1) model is imposed and data generating process is $X_t = \phi X_{t-1} + u_t$ with *iid* $u_t \sim N(0, \sigma^2)$.

		d_n				Q_n			
σ	n	$\phi = -0.5$	$\phi = -0.1$	$\phi = 0.1$	$\phi = 0.5$	$\phi = -0.5$	$\phi = -0.1$	$\phi = 0.1$	$\phi = 0.5$
0.1	10	0.01	0.012	0.002	0.02	0.057	0.038	0.054	0.055
	30	0.006	0	0	0.008	0.052	0.044	0.044	0.059
	50	0.005	0	0	0.001	0.062	0.045	0.054	0.052
	100	0.002	0	0	0.002	0.053	0.051	0.049	0.053
1	10	0.008	0.002	0.006	0.015	0.04	0.059	0.041	0.044
	30	0.004	0	0.001	0.008	0.049	0.074	0.057	0.064
	50	0.001	0	0	0.001	0.058	0.047	0.05	0.044
	100	0	0	0	0.001	0.041	0.059	0.058	0.06

At Table 3.4, DGP is $X_t = \phi X_{t-1} + u_t$ with *iid* $u_t \sim N(0, \sigma^2)$. We generate X_t for $\phi = -0.5, -0.1, 0.1, 0.5$ and $\sigma = 0.1, 1$. Note here that $\phi = 0$ (iid error, equivalently) is not considered. A classical least square parameter estimation for the imposed AR(1) model is done for practicing DW d_n and LB Q_n in checking correlated error with significance level $\alpha = 0.05$. Table 3.4 shows that LB performs reasonably well in accepting the correctly imposed model across various values of ϕ and σ^2. DW almost always accepts the correctly imposed model regardless of value of ϕ and σ^2.

Table 3.5 Simulated power comparison of d_n and Q_n for testing correlated error at size $\alpha = 0.05$ when AR(1) model is imposed and data generating process is $X_t = \phi X_{t-1} + u_t + \theta u_{t-1}$ with $iid\ u_t \sim N(0, \sigma^2)$.

		$d_n(\phi,\theta)$				$Q_n(\phi,\theta)$			
σ	n	(-0.5,-0.3)	(-0.5,0.3)	(-0.1,-0.3)	(-0.1,0.3)	(-0.5,-0.3)	(-0.5,0.3)	(-0.1,-0.3)	(-0.1,0.3)
0.1	10	0.02	0.011	0.005	0.005	0.082	0.042	0.064	0.069
	30	0.063	0.001	0.001	0.008	0.131	0.059	0.074	0.082
	50	0.116	0.001	0.002	0.001	0.16	0.045	0.08	0.095
	100	0.238	0	0	0	0.221	0.063	0.068	0.083
1	10	0.061	0.029	0.005	0.011	0.089	0.086	0.047	0.072
	30	0.108	0.045	0.002	0	0.122	0.134	0.069	0.114
	50	0.168	0.113	0.001	0	0.152	0.164	0.054	0.106
	100	0.334	0.268	0.001	0	0.253	0.261	0.061	0.119
σ	n	(0.1,-0.3)	(0.1,0.3)	(0.5,-0.3)	(0.5,0.3)	(0.1,-0.3)	(0.1,0.3)	(0.5,-0.3)	(0.5,0.3)
0.1	10	0.001	0.019	0.001	0.066	0.053	0.056	0.039	0.106
	30	0	0.006	0	0.119	0.078	0.082	0.044	0.12
	50	0	0.003	0	0.174	0.073	0.096	0.053	0.156
	100	0	0.003	0	0.348	0.077	0.133	0.048	0.241
1	10	0.002	0.013	0.01	0	0.061	0.053	0.035	0.047
	30	0	0.004	0.002	0	0.079	0.071	0.052	0.049
	50	0	0.004	0	0	0.063	0.099	0.047	0.052
	100	0	0.005	0	0	0.077	0.096	0.06	0.044

At Table 3.5, DGP is $X_t = \phi X_{t-1} + u_t + \theta u_{t-1}$ with $iid\ u_t \sim N(0, \sigma^2)$. We generate X_t for $\phi = -0.5, -0.1, 0.1, 0.5$, $\theta = -0.3, 0.3$ and $\sigma = 0.1, 1$. A classical least square parameter estimation for the imposed AR(1) model is done for practicing DW d_n and LB Q_n in checking correlated error with significance level $\alpha = 0.05$. Table 3.5 shows that both LB and DW fail to perform properly in rejecting the null hypothesis.

In view of model specification our simulation results is summarized as follows. (a) DW is very sensitive to correlated error as long as the underlying correlated error is restricted to AR(1). (b) DW doesn't tend to be sensitive enough to reject the wrongly imposed linear regression model. (c) LB performs reasonably well when correct ARMA model is specified but does not have enough sensitivity when ARMA model is wrongly specified.

The DW test only works when errors are correlated via AR(1) model and the LB test needs to improve its level of performance. Both of them need some modifications and it will be done in separate papers.

References

Box, G. E. P and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models *Journal of the American Statistical Association*, **65**, 1509-1526.

Durbin, J. and Watson, G. S. (1950). Testing for serial correlation in least square regression: I. *Biometrika*, **37**, 409-428.

Durbin, J. and Watson, G. S. (1951). Testing for serial correlation in least square regression: II. *Biometrika*, **38**, 159-177.

Ha, J. and Jung, J. M. (2015). A study on the slope sign test for explosive autoregressive models. *Journal of the Korean Data & Information Science Society*, **26**, 791-799.

Hwang, H. (2014). Support vector quantile regression for autoregressive data. *Journal of the Korean Data & Information Science Society*, **25**, 1539-1547.

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

Pena, D. and Rodriguez, J. (2002). A Powerful Portmanteau Test of Lack of Fit for Time Series. *Journal of the American Statistical Association*, **97**, 601-610.