Durban Watson for Testing the Lack-of-Fit of Polynomial Regression Models without Replications

Ruba A. Alyafi, Maha A. Omair, Abdullah A. Al-Shiha
ralyafi@ksu.edu.sa, maomair@ksu.edu.sa, aalshiha@ksu.edu.sa
Department of Statistics and Operations Research, College of Sciences
King Saud University, Saudi Arabia

Abstract:
A parametric approach using Durban Watson test is developed to test the lack of fit of the following regression model when replication is not available.

\[ Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_i^j + \epsilon_i ; i = 1, 2, ..., n. \]

This paper is based on the idea of testing the lack of fit depending on the randomness of the sample \( v_1, v_2, ..., v_n \). We assumed that the errors \( \epsilon_1, \epsilon_2, ..., \epsilon_n \) are a random sample from a normal distribution with mean equal to zero and variance \( \sigma^2 \). Simulation results are presented to show the power and the size of the modified Durban Watson test (1951).

Key words and phrases: Parametric test, Durban Watson, lack of fit, model adequacy, regression model, power of the test, size of the test and replication.

1. Introduction
A very well-known Technique, the classical lack of fit F-test “Fisher (1922)”, is used to test the lack of fit for regression models when replication is available. On the other hand, when replication is not available most of the studies tend to partition the sample space of the continuous covariant. This study suggests a simple approach to test the adequacy of the regression model using Durban Watson test technique.

We will test the lack of fit of the following regression model (1985) when replication is not available.

\[ Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_i^j + \epsilon_i ; i = 1, 2, ..., n. \quad (1.1) \]

If the regression model is appropriate then the errors \( \epsilon_1, \epsilon_2, ..., \epsilon_n \) is a random sample from a normal distribution with mean equal to zero and variance \( \sigma^2 \). Approximately, it can be considered...
that the residuals $e_i's$ $(i = 1, 2, ..., n)$ is a random sample from a normal distribution with mean equal to zero and variance $\sigma^2$. The value $X_i$ is the value of the independent variable for the $i^{th}$ trial. Let $X(1) < X(2) ... < X(n)$ be the ordered values for the independent variable and $v_i = e(i)$ is the residuals associated with the ordered value $X(i)$.

In this test, we built a lack of fit test depending on the randomness of the sample $v_1, v_2, ..., v_n$.

To test the adequacy of the model based on the proposed lack of fit tests with the hypotheses $H_0$: the sample $v_1, v_2, ..., v_n$ is random vs. $H_1$: the sample $v_1, v_2, ..., v_n$ is not random, we run a simulation study and calculated the empirical size and power of the tests. In this simulation, we considered testing the lack of fit of the following three regression models, i.e. when $p = 1, 2$ and $3$ in (1.1):

Model (1): $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i; i = 1, 2, ..., n. \quad (1.2)$

Model (2): $y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \varepsilon_i; i = 1, 2, ..., n. \quad (1.3)$

Model (3): $y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i; i = 1, 2, ..., n. \quad (1.4)$

Note that Model (0) refers to the case $y_i = \beta_0; i = 1, 2, ..., n$.

The simulation is done using the software Mathematica 8 and the tests are repeated 10000 times. We generated the independent variable $X$ form random real numbers with minimum value 0 and a maximum value $c$ (where $c = 1, 5, 20$ and 50). The dependent variable $Y$ is calculated from the specified model (true model (1) or (2) or (3)) then we fit $Y$ to the true model and the false model (the false model is a model with a lower degree). In the simulation, we considered the following values:

$n$ is the sample size that takes the values 10, 30 and 100.

$\sigma^2$ is the variance that takes the values 0.1, 0.5, 1 and 5.

$\alpha$ is the significant level that takes the values 0.05 and 0.1.

We assumed $\beta_0 = 1$ and $\beta_1 = 1$ (when dealing with model (3)).

$\beta_j = 0.5, 1, 5$ for $j = 1, 2, ..., p$.

For each run, the empirical size of the test $\alpha$ (the test’s probability to reject $H_0$ when $H_0$ is true ) is calculated as follows:
The empirical size of the test

\[
\text{number of rejecting the true model when the y values are generated from that model} / \text{number of runs}
\]

The empirical power of the test \(1 - \beta\) (the test’s probability to reject \(H_0\) when \(H_0\) is false) is calculated as follows:

\[
\text{the empirical power of the test} = \frac{\text{number of rejecting the false model when the y values are generated from the true model}}{\text{number of runs}}
\]

For both, the true and false models, the residuals and the standardized residuals associated with the ordered \(X\) values are calculated. The Durbin Watson test statistic is calculated for both the true and the false model. Using the two-sided rejection rule the empirical size and power of the test is obtained. A counter is initiated, whenever the test is inconclusive the counter adds up and the sample is omitted.

Note: The Durbin-Watson test is not applicable when using Model (0) as the false model and Model (1) as the true model, i.e. when \(p = 1\), since Durbin-Watson tables are given only when a constant term in the regression model exist and at least one independent variable.

2. The Test

In this test, we propose testing the lack of fit of the regression model (1.1) i.e. the randomness of the residuals \(v_1, v_2, \ldots, v_n\) associated with the ordered \(x\) values by testing the serial correlation of the ordered residuals using the Durbin-Watson (1951) test. This test tests the null hypothesis that the residuals are not auto-correlated \(H_0: \text{all } \rho_s = 0\) against the alternative \(H_1: \rho_s = \rho^s\) which is equivalent to \(H_0 = \text{the proposed model is appropriate against } H_1 = \text{the proposed model is not appropriate.}\) The alternative takes place from the assumption that the errors in the regression model (1.1) are generated by first-order autoregressive processes that are observed at equally spaced time periods. such that:

\[
\epsilon_t = \rho \epsilon_{t-1} + z_t \quad (2.2)
\]

Where \(\epsilon_t\) is the error component at time \(t\) and \(z_t\) are i.i.d. \(N(0, \sigma^2)\). \(\epsilon_t\) and \(z_t\) are independent.

The Durbin Watson test statistic \(d\) is used on model (1.1). Using the two-sided rejection rule, at level \(\alpha\), we reject \(H_0\) if \(d < d_L \) or \(4 - d < d_U\), not reject \(H_0\) if \(d > d_U\) and \(4 - d > d_U\), otherwise the test is said to be inconclusive.
The figure and the table below display some cases of the empirical size and the power of the test to illustrate our method.

Figure 1 shows the power and size of the test for testing the lack of fit for model (1.1) when \( p = 2 \), for \( \alpha = .05 \), \( \beta_1 = .5 \) \( \beta_2 = 5 \) and different values of \( n \), \( c \) and \( \sigma^2 \).

Table 1 shows the power and size of the test for both residuals and standardized residuals when testing the lack of fit for model (1.1) when \( p = 2 \), for \( \sigma^2 = .1 \), \( \beta_1 = .5 \) and different values of \( n \), \( c \), \( \beta_2 \) and \( \alpha \).
Figure(1) The size and the power of the modified Durban Watson test ($\alpha = 0.05, \beta_1 = 0.5 \beta_2 = 5$)
Table 1 The size and the power of the modified Durban Watson test (\(\sigma^2 = .5\) \(\beta_1 = .5\))

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>(\beta_2)</th>
<th>(n)</th>
<th>(\alpha)</th>
<th>(1 - \beta)</th>
<th>(\alpha)</th>
<th>(1 - \beta)</th>
<th>(\alpha)</th>
<th>(1 - \beta)</th>
<th>(\alpha)</th>
<th>(1 - \beta)</th>
<th>(\alpha)</th>
<th>(1 - \beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1</td>
<td>.5</td>
<td>10</td>
<td>0.0977</td>
<td>0.0620</td>
<td>0.0925</td>
<td>0.5666</td>
<td>0.1039</td>
<td>0.9997</td>
<td>0.0992</td>
<td>0.9997</td>
<td>0.0793</td>
<td>0.0582</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.0698</td>
<td>0.0760</td>
<td>0.0702</td>
<td>0.9992</td>
<td>0.0699</td>
<td>1</td>
<td>0.0701</td>
<td>0.0679</td>
<td>0.0750</td>
<td>0.0682</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.0775</td>
<td>0.0865</td>
<td>0.0740</td>
<td>1</td>
<td>0.0800</td>
<td>1</td>
<td>0.0811</td>
<td>0.0773</td>
<td>0.0864</td>
<td>0.0738</td>
</tr>
<tr>
<td>.1</td>
<td>1</td>
<td>10</td>
<td>0.0986</td>
<td>0.0596</td>
<td>0.1040</td>
<td>0.9448</td>
<td>0.0990</td>
<td>0.9994</td>
<td>0.1082</td>
<td>0.9997</td>
<td>0.0806</td>
<td>0.0547</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.0712</td>
<td>0.0764</td>
<td>0.0673</td>
<td>1</td>
<td>0.0650</td>
<td>1</td>
<td>0.0728</td>
<td>1</td>
<td>0.0689</td>
<td>0.0759</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.0776</td>
<td>0.0929</td>
<td>0.0764</td>
<td>1</td>
<td>0.0736</td>
<td>1</td>
<td>0.0785</td>
<td>1</td>
<td>0.0768</td>
<td>0.0926</td>
</tr>
<tr>
<td>.1</td>
<td>5</td>
<td>10</td>
<td>0.1026</td>
<td>0.0858</td>
<td>0.1006</td>
<td>1</td>
<td>0.1026</td>
<td>0.9994</td>
<td>0.1020</td>
<td>0.9997</td>
<td>0.0798</td>
<td>0.0849</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30</td>
<td>0.0703</td>
<td>0.5339</td>
<td>0.0688</td>
<td>1</td>
<td>0.0678</td>
<td>1</td>
<td>0.0661</td>
<td>1</td>
<td>0.0686</td>
<td>0.5459</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.0750</td>
<td>0.9650</td>
<td>0.0757</td>
<td>1</td>
<td>0.0717</td>
<td>1</td>
<td>0.0767</td>
<td>1</td>
<td>0.0747</td>
<td>0.9667</td>
</tr>
</tbody>
</table>
3. Conclusions:
Our following conclusions applies for $p = 2$ and 3.
1- Comparing the performance of the test using standardized residuals and regular residuals, we get almost the same results.
2- When $n=10$ sometimes the test is said to be inconclusive and this is due to the large range between the $d_l$ and $d_u$ value relatively with the Durbin Watson test statistic, as $n$ increases the inconclusive cases decreases.
Example:
When $n = 10$, $\beta_1 = .5$, $\beta_2 = .5$, $\sigma^2 = .1$, $c = 5$.

For $\alpha = .05$:
Number of inconclusive for true model i.e. model 2 (1.3) is 4850.
Number of inconclusive for false model i.e. model 1 (1.2) is 2928.
For $\alpha = .1$:
Number of inconclusive for true model i.e. model 2 (1.3) is 6142.
Number of inconclusive for false model i.e. model 1 (1.2) is 1265.

When $n = 100$, $\beta_1 = .5$, $\beta_2 = .5$, $\sigma^2 = .1$, $c = 5$.

For $\alpha = .05$:
Number of inconclusive for true model i.e. model 2 (1.3) is 528.
Number of inconclusive for false model i.e. model 1 (1.2) is 0.
For $\alpha = .1$:
Number of inconclusive for true model i.e. model 2 (1.3) is 900.
Number of inconclusive for false model i.e. model 1 (1.2) is 0.

3- The empirical size of the test almost decreases as $n$ increases.
4- The empirical size of the test increases as $\alpha$ increases.
5- The empirical size of the test increases as $c$ increases.
6- The empirical power of the test increases when $n$ increases.
7- The empirical power of the test increases when $c$ increases.
8- The empirical power of the test slightly increases when $\beta_1$ increases.
9- The empirical power of the test increases when $\beta_2$ increases.
11- As n increases the empirical power of the test almost always increases especially for larger c values.

12- As $\sigma^2$ increases the empirical power of the test decreases.

References


