

## Power Comparison of Autocorrelation Tests in Dynamic Models

Erum Toor<sup>®</sup> and Tanweer Ul Islam

National University of Sciences and Technology      National University of Sciences and Technology

Received: 24.07.2018      Accepted: 23.07.2019      Published: 30.09.2019

### ABSTRACT

The four most readily available tests of autocorrelation in dynamic models namely Durbin's  $M$  test, Durbin's  $H$  test, Breusch-Godfrey (BGF) test and Ljung and Box ( $Q$ ) test are compared in terms of their power for varying sample sizes, levels of autocorrelation and significance using Monte Carlo simulations in STATA. Power comparison reveals that the Durbin  $M$  test is the best option for testing the hypothesis of no autocorrelation in dynamic models for all sample sizes. Breusch-Godfrey's test has comparable and at times minutely better performance than Durbin's  $M$  test however in small sample sizes, Durbin's  $M$  test outperforms the Breusch-Godfrey test in terms of power. The Durbin  $H$  and the Ljung and Box  $Q$  tests consistently occupy the second last and last positions respectively in terms of power performance with maximum power gap of 63 and 60% respectively from the best test ( $M$  test).

**Key words:** *Durbin Test, Breusch-Godfrey Test, Ljung and Box Test*

JEL Classifications: C60, C63

## 1. INTRODUCTION

Dynamic models are frequently used in economics, social sciences, business and medicines for policy analysis (Scott and Heady, 1967; McCunn and Huffman, 2000; Spix and Wichmann, 1996; Solomou and Weale, 1993; Sivo and Willson, 1998; Ahlburg, 1985; Goetzmann and Jorion, 1993; Chirinko, 1980; Womer and Patterson, 1983; and Li, 1999). Their basic aim is to capture the remarkably subtle feedback effects, often missed by static models. The addition of lagged dependent variables in Ordinary Least Square (OLS) regression serves two purposes: capturing dynamic effect and ridding the model of autocorrelation, but the inference of the model in these cases is a long-standing problem (Rois et al., 2012).

Least Square estimation of parameters in case of both linear and dynamic models is highly inefficient in the presence of auto-correlated errors (Lee and Lund, 2004). The OLS estimates for the dynamic linear models are biased and generally inconsistent (Breusch, 1978). Moreover, other diagnostic statistics like  $R^2$ ,  $t$  and  $F$  tests are invalid (Greene, 2000). Consequently, the problem of testing for autocorrelation has received much needed attention from the researchers concerned with the analysis of time series data (Savin and White, 1978; Wallis, 1972; Park, 1975; Breusch, 1978; Durbin, 1970).

---

<sup>®</sup> Erum Toor, Graduate, National University of Sciences and Technology, Islamabad, (email: [erum\\_toor@hotmail.com](mailto:erum_toor@hotmail.com)), Tel: +01223 3456789, Fax: + 01225 398765.

Tanweer Ul Islam, Assistant Professor, Department of Economics, School of Social Sciences and Humanities, National University of Sciences and Technology, Islamabad 44000, Pakistan, (email: [tanweer@s3h.nust.edu.pk](mailto:tanweer@s3h.nust.edu.pk)), Tel: +92 51 90853567.

For linear models, Durbin-Watson (DW), Durbin  $d$ , Schmidt test and Geary's sign test are the statistics often used to detect the problem of autocorrelation in the residuals (Smith, 1976) and the literature provides the consensus regarding the supremacy of DW-test to other competitors in terms of power. However, the inability of these tests to include lagged dependent terms in model limits their application.

To overcome this limitation, several such tests have been devised (e.g. Durbin  $H$  test, Durbin  $M$  test, Breusch-Godfrey LM test and Box-Pierce-Ljung Statistic) which incorporate varying autoregressive error structures and functional forms. The availability of such a good number of autocorrelation tests motivated simulation studies to find the best test (see for example; Kenkel, 1974; McNown and Hunter, 1980; Inder, 1984; Dezhbakhsh, 1990; Rayner, 1993; Rois et al., 2012). Literature to a great extent reveals its strength when it comes to the individual tests, be it the underlying assumption, critical value generation, mathematical representation and other properties but when it comes to the comparison amongst different tests in the specified setting the literature is void of research. In particular, the literature is scanty for the last two decades in terms of power comparison of autocorrelation tests for dynamic model. Moreover, the most popular and readily available autocorrelation tests (Durbin  $H$  test, Durbin  $M$  test, Breusch-Godfrey test and Box-Pierce-Ljung  $Q$  Statistic) have not been compared all together in a single simulation study. This study compares the aforementioned autocorrelation tests in terms of their power under the given conditions i.e. specific sample sizes, autocorrelation coefficients and levels of significance

## **2. LITERATURE REVIEW**

There exists immense literature concerned with the implications of serially correlated disturbances both in linear as well as in dynamic regression models and alternative estimation techniques. However, if we do not know whether the model we are dealing with has autoregressive disturbance term, we may not be able to decide which estimation method to employ. For researchers, choosing the best test for detecting the autocorrelation among the available options is a dilemma. This problem has led to the interests in relative power and other properties of autocorrelation tests (see for example; Durbin and Watson, 1951; Jenkins, 1954; Koerts and Abrahamse, 1968; Hannan, 1968; Geary, 1970; Schmidt and Guilkey, 1975; and Blattberg, 1973). Econometricians have not only proposed various tests to check for autocorrelation under varying assumptions but also performed tests to check relative performance of some of the tests to see which test performs better in terms of size and power but there is a very limited literature available in this regard.

In the case of linear models, a power comparison of Durbin-Watson, Schmidt test and Durbin Alternative Exact test reveals that the specification of error autoregressive structure affects the performance of each test. The DW and Durbin Alternative exact test are quite comparable in their overall performance, as indicated by their estimated powers (Smith, 1976). These results were quite similar to the ones obtained by Blattberg (1973) who suggests using a wider array of test statistics based on their autoregressive structure.

Gastwirth and Selwyn (1980) compare the Geary and Durbin Watson (DW) tests for serial correlation in terms of robustness through Monte Carlo simulations. The results reveal that the DW procedure is generally more powerful and robust than Geary's test under number of specifications.

Maddala and Rao (1973) compare the tests put forth by Durbin with LR test using data generated by two Monte Carlo experiments for the sample size of 32 and 76 running 500 simulations using the following model:

$$y_t = \alpha y_{t-1} + x_t' \beta + u_t$$

Results reveal that both of the Durbin's test perform slightly better than the LR test, of the two Durbin's test, the  $H$  statistic is easier to apply but is rejected in many of the cases where as  $M$  test although a lengthy procedure fares much better and can be applicable in all circumstances.

Comparison of alternative Durbin tests namely  $D$ ,  $H$  and  $M$  tests, in finite samples, receives considerable attention as well (Rayner, 1993). Inder (1984) noted that the much of the apparent differences in the empirical estimates of power were largely due to the different approximations of the critical values used by the econometricians.

Strelec and Adamec (2013) performed Monte Carlo exercise to compare the power of several tests of autocorrelation including DW test, Durbin  $H$ , Langrage Multiplier test and Ljung Box Pierce test in first order serial correlation models for sample sizes ranging from 20 to 100 and positive autocorrelation coefficients and conclude that Durbin Watson test with two sided alternative appears to be the best test closely followed by Ljung Box test.

Dezhabksh (1990) while commenting on the inappropriate use of autocorrelation tests in dynamic models compared the performance of various tests including DW, Durbin's  $H$  test and  $M$  test and revealed that Durbin  $M$  test, a test that is most infrequently used in literature is the best test having greater power and a size very close to its nominal size.

Rayner (1993) uses bootstrap method to compare the performance of  $D$ ,  $H$ ,  $M$  and a newly introduced  $S$  test by Inder (1990) using the dynamic model given by

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + x_t' \beta + u_t$$

where  $u_t$  follows an AR(1) error structure, to evaluate the various approximations to critical points to accept or reject the hypothesis of no autocorrelation at 5% level of significance, Rayner (1993) uses the simulated data along with the quarterly GNP of USA for sample sizes of 32 and 60 for 5000 replications. Results reveal that  $S$  test appears to perform better in models with two lagged variables however with three lagged variables  $M$  test performs better.

Inder (1990) tests for AR(1) errors in models with lagged dependent variables

$$y_t = \alpha y_{t-1} + x_t' \beta + u_t$$

and purposes a new test. He compares his test with DW,  $H$  and  $T$  test and concludes that the DW test is reliable and powerful test in small samples, however, the new test turns out to be the most powerful test under selected circumstances of a relatively small sample sizes with AR(1) error structure at 95% confidence interval.

Kenkel (1974) compares the DW test, Durbin  $H$  test and Durbin alternative test and suggests that a new test be devised as the researchers are faced with a dilemma to make a choice between a test that has no theoretical support in case of dynamic models (DW test) a test that is inapplicable in many environments (Durbin  $H$ ) and one that has a low power.

McNown and Hunter (1980) compare the power performance of  $M$ ,  $D$  and  $H$  test using Monte Carlo simulations for sample sizes of 32 and 76 and devised a new test from the model

$$y_t = (b_1 + \rho)y_{t-1} - b_1 y_{t-2} + b_2 x_t - b_2 \rho x_{t-1} + e_t$$

that performs better under certain circumstances, the proposed test is comparable in power properties to  $H$  test, as long as severe autocorrelation was not present. They use the critical value approach for rejecting the null hypothesis at 5% level of significance.

Rois et al. (2012) evaluate the performance of Breusch Godfrey test in dynamic models and propose a distance based one sided LM test. By using four different data sets, twenty thousand Monte Carlo simulations are run with AR(2) error structure with varying first and second order of autocorrelation from 0.1 to 0.9 and 0.2 to 0.8 respectively revealing that Distance based LM (DLM) test performs better in majority of experiments.

For linear models, there is almost consensus in the literature regarding the supremacy of DW test however, the literature is almost scanty for the last two decades in terms of comparison of autocorrelation tests for dynamic models. In particular, the following selected autocorrelation tests have not been compared together in a single study.

## 2.1. Selected Autocorrelation Tests

Following is the brief description of each test of autocorrelation under consideration for comparison of power in context of dynamic models.

### 2.1.1. Durbin H Test

Durbin derived a large sample test to detect the presence of autocorrelation; it is a test against the first order autocorrelation and need the specification of the complete set of regressors in the model (Durbin, 1970). This test is restricted to one lag of dependent variable.

Consider the model:

$$y_t = \alpha y_{t-1} + \beta x_t$$

with this error structure

$$u_t = \rho_1 u_{t-1} + \varepsilon_t$$

Durbin proposes the following normally distributed H-statistic to test the hypothesis of no autocorrelation ( $\rho_1 = 0$ )

$$h = \hat{p} \sqrt{\frac{n}{1-n\hat{V}(\hat{\alpha})}}$$

where  $n$  is the sample size,  $\hat{V}(\hat{\alpha})$  is the estimated variance of the coefficient of  $y_{t-1}$ ,  $n\hat{V}(\hat{\alpha}) < 1$  and  $\hat{p}$  is the ratio of errors computed as

$$\hat{p} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2}$$

The  $h$ -statistics follows the standard normal distribution.

### 2.1.2. Breusch Godfrey Test

Breusch (1978) and Godfrey (1978a, 1978b) develop the most widely used LM test to check for autocorrelation. This likelihood based LM type test is one of the most appropriate tests to check for the presence of autocorrelation in dynamic models. This test allows for regressors that are not strictly exogenous, higher order of autoregressive schemes, simple or higher order of moving averages of white noise error term.

The test follows the usual mechanics of any LM test, so for the model given as

$$y_t = \alpha y_{t-1} + \beta x_t + u_t$$

with a suspected autoregressive scheme of order 1

$$u_t = \rho_1 u_{t-1} + \varepsilon_t$$

The model could be rewritten as

$$y_t = \alpha y_{t-1} + \beta x_t + \rho_1 u_{t-1} + \varepsilon_t$$

This formation of the model is termed as the ‘unrestricted model’, because of the fact that  $\rho$  can take any value without restriction, resultantly the error term can be dependent (autocorrelated) or independent with  $\rho$  being 0 in the latter case.

The restricted alternative being

$$y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t$$

Breusch and Godfrey explained that the residual error be extended to AR( $p$ ) structure

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t$$

Simple OLS regression of the dynamic model having specified error structure can give us a set of sample residuals, which can be used in an auxiliary regression

$$\widehat{u}_t = \alpha y_{t-1} + \beta x_t + \rho_1 \widehat{u}_{t-1} + \rho_2 \widehat{u}_{t-2} + \dots + \rho_p \widehat{u}_{t-p} + \varepsilon_t$$

If the usual  $R^2$  is calculated for the model, it can be used for the test statistic  $TR^2 \sim \chi_p^2$ .

### 2.1.3. Durbin M Test

Durbin  $M$  test commonly referred to as Durbin’s alternative test, it employs a procedure quite similar to that of Breusch-Godfrey LM test. It is a more general test and can be performed to detect presence of autocorrelation even if  $n\widehat{V}(\widehat{\alpha}) > 1$  (Durbin, 1970). The distribution of this test does not depend on regression parameters and error variance. Asymptotically, the test statistic follows the standard normal distribution.

### 2.1.4. The Q Test

The  $Q$  test proposed by Box and Pierce (1970) and later modified by Ljung and Box (1978) is a way to test for autocorrelation at multiple lags.

Consider a discrete time series  $x_t$  generated by a stationary ARMA model:

$$\varphi(B)x_t = \theta(B)e_t$$

where,  $\varphi(B)$  and  $\theta(B)$  are the finite lag polynomial with order  $p$  and  $q$  respectively, and  $e_t \sim iidN(0, \sigma^2)$ . Box and Pierce, 1970 show that for large  $n$  the following statistics

$$Q = n(n+2) \sum_{h=1}^m (n-h)^{-1} r_h^2$$

where

$$r_h = \frac{\sum_{t=h+1}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2}$$

is distributed as  $\chi_m^2$  and the null hypothesis of zero autocorrelation up to  $m$  order is rejected for larger values.

### 3. SIMULATIONS

Let  $X=(X_1, X_2, \dots, X_n)$  is a vector of observations with density function  $f(x, \varphi)$  and  $\varphi$  belongs to parameter space  $\emptyset$  with the null and alternative hypothesis as given below:

$$\begin{aligned} H_0: \varphi \in \emptyset_0 \subseteq \emptyset \\ H_a: \varphi \in \emptyset_a \subseteq \emptyset \end{aligned}$$

A hypothesis test is any function  $\psi(x)$  which takes values  $\{0,1\}$  and  $\Psi$  is a set of all such hypothesis tests. The size of a test is defined as the maximum probability of rejecting the null hypothesis when the null hypothesis is true.

$$\alpha = \text{size of } \psi(x) = \text{Sup}_{\varphi \in \emptyset_0} p(\psi(x) = 0/\varphi)$$

If  $\Psi_\alpha$  is a set of all test with size  $\alpha$ , the power of test is defined as the probability of rejecting the null hypothesis when the null hypothesis is false.

$$\text{Power of } \psi(x) = E^\varphi[\psi(x)] = 1 - p(\psi(x) = 0/\varphi \in \emptyset_a)$$

The research aims at testing the power performances of the selected autocorrelation tests in dynamic models by considering the following data generating process (DGP):

$$y_t = \alpha y_{t-1} + \beta x_t + u_t \quad (3.1)$$

where  $t=1, 2, \dots, n$ ;  $y_t$  is the  $t^{\text{th}}$  observation of the dependent variable; and  $u_t$  is a stochastic disturbance which follows the AR (1) process.  $u_t = \rho_1 u_{t-1} + \varepsilon_t$  where  $|\rho_1| < 1$  and  $\varepsilon_t \sim IN(0, \sigma^2)$ . The value of  $\rho_1$  ranges from 0.1 to 0.9, with a 0.1 interval each based on assumed degree of correlation. The error structure is limited to order one as Durbin's  $H$  test is not applicable for testing second or higher order of autocorrelation in dynamic models. Following Davidson and MacKinnon (1999), the null hypothesis of no autocorrelation is rejected based on  $p$ -value criterion.

Following Dezhbakhsh (1990) and Rayner (1993), the coefficient of lagged dependent variable is set at 0.5 in equation (3.1). Only positive values of autocorrelation coefficient are considered because this is the phenomenon of common occurrence (Maddala and Rao, 1973). Power and size calculations are based on 10,000 Monte Carlo simulations for selected sample sizes (25, 50, 75, 100, 250, 500 and 1000) and level of significance (0.01, 0.05, 0.1, 0.2 and 0.4) in STATA. Simulated critical values are used to explore the size properties with 100,000 replications.

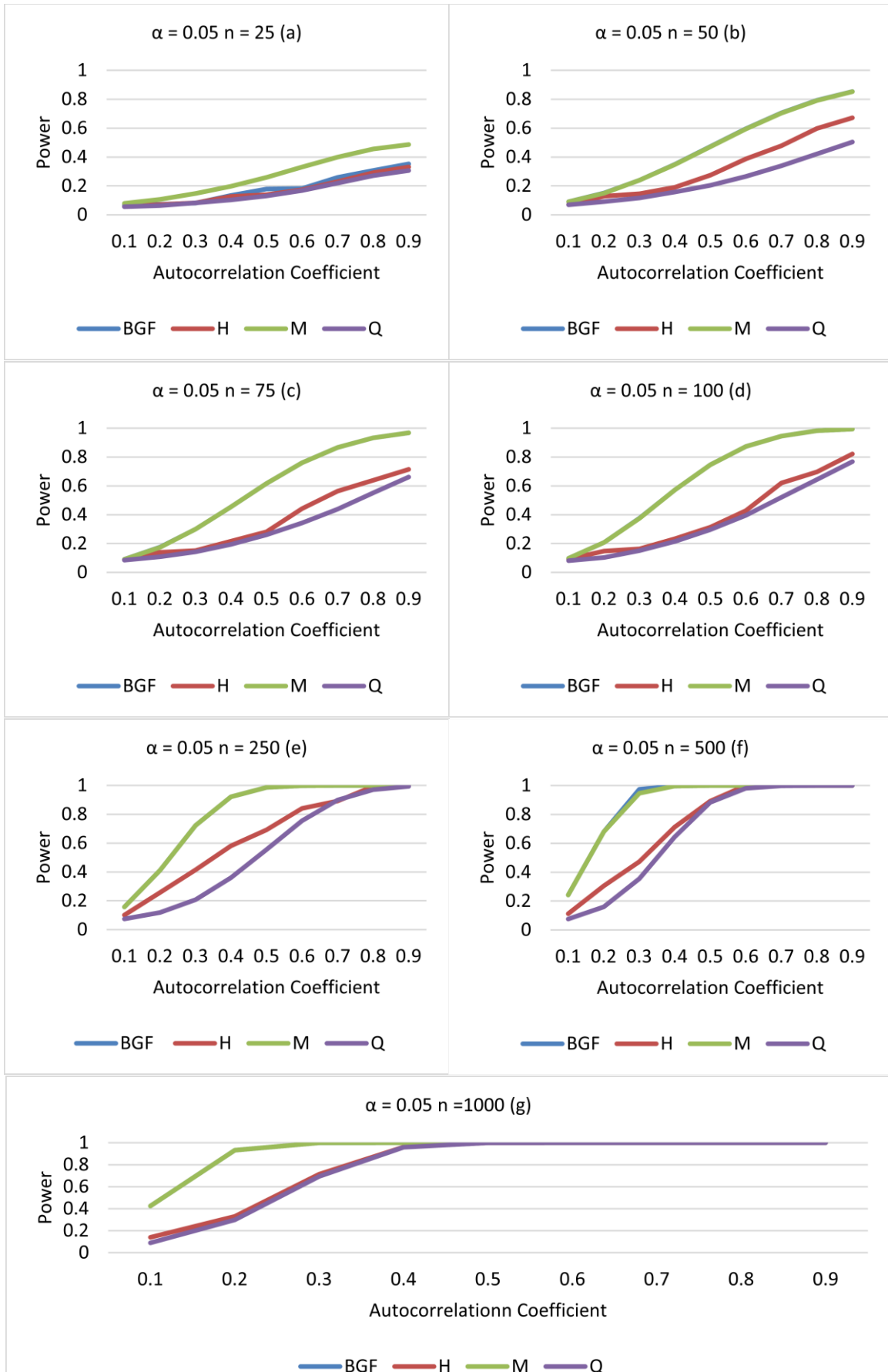
### 4. RESULTS DISCUSSION

When evaluating the power of selected autocorrelation tests, the Durbin's  $M$  test has shown better performance even for small sample size and low autocorrelation values. On balance, the results corroborate with the findings in literature (for example; Breusch, 1978; McNown and Hunter, 1980; Inder, 1990; Davidson and MacKinnon, 2004; Godfrey, 2007).

#### 4.1. Small Sample

On balance, for small sample sizes, at 0.05 level of significance, Durbin's  $M$  test outperforms other autocorrelation statistics in terms of power comparison (Figure 4.1.a,b) with the BGF test following closely except for sample size of size 25 (Figure 4.1.a). For this particular incidence, the gap between the powers of  $M$  and BGF tests increases with the increase in autocorrelation and reaches the maximum level of 24% when the autocorrelation is 0.8. It is evident from Figure 4.1(a,b) that the gap between the power of the best performing test ( $M$  test) and the worst

Figure 4.1 Power Comparison of *M*, BGF, *H* and *Q* tests ( $\alpha=0.05$ )



performing test ( $Q$  test) maximizes when the autocorrelation ranges between 0.7 to 0.8 in small samples with a maximum gap of 37% ( $\rho = 0.8$  and  $n = 50$ ). The performance of  $H$  test although improves with extent of autocorrelation but the maximum gap of power with the best test is 22.6% ( $\rho=0.8$  and  $n=50$ ). For small samples, the tests can be ranked as  $M > \text{BGF} > H > Q$ .

#### 4.2. Medium Sample

With the increase in sample size, the BGF test catches up with the performance of the  $M$  test (Figure 4.1.c,d) which corroborates with the findings in Rayner (1993), Godfrey (2007) and Maddala and Rao (1973). On balance, it is evident from Figure 4.1.c,d, the overall performance of all tests improve with the increase in sample size and the level of autocorrelation present, in the case of  $H$  test the percentage of times where the test is inapplicable slightly reduces as well. It is interesting to note that the gap between the best performing tests ( $M$  and BGF) and the  $Q$  and  $H$  tests maximizes when the level of autocorrelation ranges from 0.6 to 0.8 with maximum gap of 47.7% and 44.5% ( $\rho = 0.6$  and  $n=100$ ) respectively.

McNown and Hunter (1980) compare the performance of Durbin  $M$  and  $H$  tests using Monte Carlo simulations for the sample size of 32 and 76 and prove the superiority of the  $M$  test to  $H$  test. Štřelec and Adamec (2013) compare the power properties of several tests in dynamic models and the results obtained are in line with our findings- the BGF test has the best performance for sample of sizes more than sixty. They attribute the superior performance to its test statistic's capability of detecting correlation between current and lagged errors (Greene, 2000), this line of reasoning can be applied to Durbin  $M$  test as well. In general, for medium sample sizes, the ranking of the four selected tests of autocorrelation is  $M = \text{BGF} > H > Q$ .

#### 4.3. Large Sample

Increase in sample size beyond 100 observations is a scenario seldom tested in the literature; however, the literature agrees upon the notion that the performance of all test improves with the increase in sample size.

It is evident from Figure 4.1(e,f,g) that all four tests reject the null of no autocorrelation with almost 100 % accuracy for larger autocorrelation values. There is hardly any difference between the power performance of the  $M$  and BGF tests for large samples. At lower levels of autocorrelation, Durbin's  $M$  and the BGF test retains their top position with the  $Q$  and  $H$  test lagging behind with maximum power gap of 63 and 60 percent respectively. On balance, the Durbin's  $M$  and the BGF test statistics remain the trusted choices because of their encompassing nature to cater to all levels of autocorrelation with greater accuracy.

### 5. ROBUSTNESS

Additional properties of the autocorrelation tests were observed by varying the previously fixed parameters (coefficient of lagged dependent variable, structure of independent variable to gauge the effect of the changes on the power performance. The power of  $M$  test is unhampered by the changes in the coefficient of lagged dependent variable, it is insensitive to the variation of lagged dependent variable, these findings are in agreement with the work of Maddala and Rao (1973) and Kenkel (1974) making it a more general and encompassing test. The change in the structure of independent variable does not impact the power drastically as well.



In the case of  $Q$  test the power improves drastically if the coefficient of lagged dependent variable is above 0.5, lower coefficients result in lower powers where as a slight increase in the coefficient results in a drastic increase in power. This is similar to the effect observed in Durbin  $H$  test where the power of the test improves with the increase in the coefficient of lagged dependent variable and low power is the problem of the  $H$  test whenever low values of lagged dependent variable are encountered confirming the results of McNown and Hunter (1980), Kenkel (1974), Maddala and Rao (1973) however, the AR structure of the dependent variable has no impact on the power properties.

Where as in the case of Breusch-Godfrey test, power shows no sensitivity to the coefficient of lagged dependent variable. The power of the test was checked varying the AR structure of independent variable  $X$ , if  $X$  is independent of its lag term, the power shows a slight decrease, but not something drastic. Similarly, if the independent variable completely follows its past (coefficients of 0.9 to 1) there is only a slight improvement in the power, hence it can be concluded that varying the AR structure of dependent variable has a very negligible impact on the power properties confirming the pioneer workings of Breusch (1978).

## 6. CONCLUSION

This study explores the power and size properties of selected autocorrelation tests, Durbin  $M$  and  $H$  test, Breusch-Godfrey test (BGF) and Ljung and Box test ( $Q$ ), in context of dynamic models by calling the Monte Carlo simulation procedures for different sample sizes, correlation values and levels of significance. Followings are the key findings of the study:

1. In general, power of each test improves with the increase in sample size and autocorrelation value.
2. Asymptotically, all tests are equivalent in power performance for high correlation values however, the  $M$  test is better than others with small samples and low autocorrelations.
3. The BGF test closely follows the  $M$  test both in small and large samples sizes with  $M$ -test having an edge in small sample sizes. For all levels of autocorrelations, there is hardly any difference between the powers of the two test for medium and large sample size.
4. The Durbin  $H$  and  $Q$  tests consistently occupy the second last and last positions respectively in terms of power performance with maximum power gap of 63 and 60% respectively from the best test ( $M$  test). This worst performance of the  $Q$  and  $H$  statistics is for the lower levels of autocorrelations.
5. Varying the coefficient of lagged dependent variable has almost negligible impact on power of the  $M$  and BGF test.
6. There is a significant gain in power of the  $Q$  test if the coefficient of lagged dependent variable is more than 0.5.

In light of the findings in this study, we recommend the Durbin  $M$  test for all samples size and all autocorrelation values. For sample size greater than 25, the BGF test is equivalent to the  $M$  test in terms of power performance. The Durbin  $H$  test and the  $Q$  tests are only useful if the sample sizes are large and autocorrelations are high as well.

## REFERENCES

- Ahlburg, D.A. (1985). The Effect of Strikes on Suicide: Time Series Evidence from the United States. *Sociological Focus*, 18 (1), 29-36.
- Blattberg, R.C. (1973). Evaluation of the power of the Durbin-Watson statistic for non-first order serial correlation alternatives. *The Review of Economics and Statistics*, 55 (4), 508-515.
- Box, G.E. and D.A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65 (332), 1509-1526.
- Breusch, T.S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17 (31), 334-355.
- Chirinko, R.S. (1980). The real wage rate over the business cycle. *The Review of Economics and Statistics*, 62 (3), 459-461.
- Davidson, R. and J.G. MacKinnon (1999). The size distortion of bootstrap tests. *Econometric Theory*, 15 (3), 361-376.
- Davidson, R. and J.G. MacKinnon (2004). *Econometric Theory and Methods (Vol. 5)*. New York: Oxford University Press.
- Dezhbakhsh, H. (1990). The inappropriate use of serial correlation tests in dynamic linear models. *The Review of Economics and Statistics*, 72 (1) 126-132.
- Durbin, J. (1970). Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica* (pre-1986), 38 (3), 410.
- Durbin, J. and G.S. Watson (1951). Testing for serial correlation in least squares regression. II. *Biometrika*, 38 (1/2), 159-177.
- Gastwirth, J.L. and M.R. Selwyn (1980). The robustness properties of two tests for serial correlation. *Journal of the American Statistical Association*, 75 (369), 138-141.
- Geary, R.C. (1970). Relative efficiency of count of sign changes for assessing residual autoregression in least squares regression. *Biometrika*, 57 (1), 123-127.
- Godfrey, L.G. (1978a). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, 46 (6), 1293-1301.
- Godfrey, L. G. (1978b). Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society*, 46 (6), 1303-1310.

- Godfrey, L.G. (2007). Alternative approaches to implementing Lagrange multiplier tests for serial correlation in dynamic regression models. *Computational Statistics and Data Analysis*, 51 (7), 3282-3295.
- Goetzmann, W.N. and P. Jorion (1993). Testing the Predictive Power of Dividend Yields. *The Journal of Finance*, 48 (2), 663-679.
- Greene, W. H. (2000). *Econometric Analysis*. New Jersey: Prentice Hall.
- Hannan, E.J. and R.D. Terrell (1968). Testing for serial correlation after least squares regression. *Econometrica: Journal of the Econometric Society*, 36 (1), 133-150.
- Inder, B.A. (1984). Finite-sample power of tests for autocorrelation in models containing lagged dependent variables. *Economics Letters*, 14 (2-3), 179-185.
- Inder, B.A. (1990). A new test for autocorrelation in the disturbances of the dynamic linear regression model. *International Economic Review*, 31 (2), 341-354.
- Jenkins, G.M. (1954). Tests of hypotheses in the linear autoregressive model: 1. Null hypothesis distributions in the Yule scheme. *Biometrika*, 41(3/4), 405-419.
- Kenkel, J.L. (1974). Some small sample properties of Durbin's tests for serial correlation in regression models containing lagged dependent variables. *Econometrica* (pre-1986), 42 (4), 763.
- Koerts, J. and A.P.J. Abrahamse (1968). On the power of the BLUS procedure. *Journal of the American Statistical Association*, 63 (324), 1227-1236.
- Lee, J. and R. Lund (2004). Revisiting simple linear regression with autocorrelated errors. *Biometrika*, 91 (1), 240-245.
- Li, K. (1999). Testing symmetry and proportionality in PPP: a panel-data approach. *Journal of Business and Economic Statistics*, 17 (4), 409-418.
- Ljung, G.M. and G.E. Box (1978). On a measure of lack of fit in time series models. *Biometrika*, 65 (2), 297-303.
- Maddala, G.S. and A.S. Rao (1973). Tests for serial correlation in regression models with lagged dependent variables and serially correlated errors. *Econometrica* (pre-1986), 41 (4), 761.
- McCunn, A. and W.E. Huffman (2000). Convergence in US productivity growth for agriculture: implications of interstate research spillovers for funding agricultural research. *American Journal of Agricultural Economics*, 82 (2), 370-388.
- McNown, R.F. and K.R. Hunter (1980). A test for autocorrelation in models with lagged dependant variables. *The Review of Economics and Statistics*, 62 (2), 313-317.
- Park, S.-B. (1975). On the Small-Sample Power of Durbin's h Test. *Journal of the American Statistical Association*, 70 (349), 60-63.

- Rayner, R.K. (1993). Testing for Serial Correlation in Regression Models with Lagged Dependant Variables. *The Review of Economics and Statistics*, 75 (4), 716-721.
- Rois, R., T. Basak, M.M. Rahman and A.K. Majumder (2012). Modified Breusch-Godfrey Test for Restricted Higher Order Autocorrelation in Dynamic Linear Model—A Distance Based Approach. *International Journal of Business and Management*, 7 (17), 88.
- Savin, N.E. and K.J. White (1978). Testing for Autocorrelation with Missing Observations. *Econometrica*, 46 (1), 59-67.
- Schmidt, P. and D.K. Guilkey (1975). Some further evidence on the power of the Durbin watson and Geary tests. *The Review of Economics and Statistics*, 57 (3), 379-381.
- Scott Jr, J.T. and E.O. Heady (1967). Regional demand for farm buildings in the United States. *Journal of Farm Economics*, 49 (1\_Part\_I), 184-198.
- Sivo, S.A. and V.L. Willson (1998). Is parsimony always desirable? Identifying the correct model for a longitudinal panel data set. *The Journal of Experimental Education*, 66 (3), 249-255.
- Smith, V.K. (1976). The estimated power of several tests for autocorrelation with non-first-order alternatives. *Journal of the American Statistical Association*, 71 (356), 879-883.
- Solomou, S. and M. Weale (1993). Balanced estimates of national accounts when measurement errors are autocorrelated: the UK, 1920–38. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156 (1), 89-105.
- Spix, C. and H.E. Wichmann (1996). Daily mortality and air pollutants: findings from Köln, Germany. *Journal of Epidemiology and Community Health*, 50 (Suppl 1), 52-58.
- Střelec, L. and V. Adamec (2013). Exploration into power of homogeneity and serial correlation tests. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, 61 (4), 1129-1136.
- Wallis, K. (1972). Testing for Fourth Order Autocorrelation in Quarterly Regression Equations. *Econometrica*, 40 (4), 617-36.
- Womer, N.K. and J.W. Patterson (1983). Estimation and testing of learning curves. *Journal of Business and Economic Statistics*, 1 (4), 265-272.