Comparative Study of Normality Tests As Applied to Some Set of Time Series Data

Carolina B. Baguio
Arlene V. Cahoy

Abstract

Time Series data are usually autocorrelated or dependent. Applications of the different normality tests for independent and identically distributed data if employed posed many drawbacks. A comparative study of these various tests applied to time series data showed that the Geary's test performed better than its counterparts followed by the Wilk-Shapiro test.

Keywords: autocorrelated, stationary and non-stationary data, normality, Geary's Test, Wilk-Shapiro

1 Introduction

The first step in the analysis of data before any estimation or inferences are made is to test for the normality of the data which are at least interval val-

✍ **CAROLINA B. BAGUIO** is a Professor of Mathematics in the Department of Mathematics, College of Science and Mathematics, MSU-IIT, Iligan City, Philippines. She holds a Ph.D. in Applied Mathematics from the Mindanao Polytechnic State College, Cagayan de Oro City. **ARLENE V. CAHOY**, a faculty member of the University of the Philippines, Cebu City, has a master's degree in Applied Statistics from MSU-IIT, Iligan City.

ued if one intends to employ the parametric method of estimation where the assumption of normality distribution is required (Siegel and Castellan Jr., 1988). Hence, the most important distribution in the field of statistics is the normal distribution (Walpole and Myers, 1994). Moreover, it is also required that these data must be independent and identically distributed (iid). Balagbis, S. S.(1994), in her undergraduate thesis demonstrated procedures and came up with comparative results of the different tests for normality specifically for independent and identically distributed normal data only. She did not investigate the effect of using these methods if the data violate the assumption of independence.

It is inevitable that data collected has substantial dependence as in the case of the time series or simulated data using the Monte Carlo simulation. Moore (1982) pointed out that tests of fit that assumes iid observations may be affected by dependence among observations as in the case of the Chi-Square test. It was also shown in this reference that observations come from a finite general class of Gaussian Processes, positive correlation among the observations confounded with lack of normality.

Gasser (1975) in his study of the goodness-of-fit for time series data, found out that for strongly correlated observations, the Chi-Square test for normality fails. However, he found out that for Autoregressive Interactive Moving Average (ARIMA) model the alternative tests such as Kolmogorov-Smirnov and skewness were shown to succeed in testing for normality. It was pointed out also that autocorrelated violations of stationarity would invalidate the test for normality. Hence, it was recommended to transform non-stationary series into stationary by some appropriate transformation.

This paper aims to investigate the performance of the different tests for normality appropriate for independent and identically distributed data when applied to stationary and non-stationary time series data. The basis for the selection of the tests is on hypothesis testing as to whether to accept or reject the null hypothesis that the data comes from a normal distribution.

2 Definitions and Concepts

Stationary series is a time series which has a constant mean and variance over time. Visually, the series plot should remain within the level band of constant width. A stationary series is meaningfully described by the autocorrelation function (ACF), which says that the correlation between any two points in a series is a function of how far apart the two points are, and is not a function of the actual times at which two points occur. Moreover, the ACF plot of a stationary series dies down fairly rapidly (exponentially) so that beyond some lag all autocorrelations are negligible (SPSS Glossary). On the other hand, the non-stationary series do not have a constant mean over time.

The different tests for normality compared in the study are the graphical and numerical tests enumerated as follow:

A. Graphical Tests

1. Rankit Plots

   This is a graphical check on the normality of the data which is applicable for small size, that is $n \leq 5$ (Wardraw, 1985). If the observations are a random sample from an underlying normal distribution, then the points should approximate a straight line.

2. Probability Plots (PP)
Plots a variable's cumulative proportions against the cumulative proportions of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line (SPSS help).

3. Q-Q plots
Plots the quantiles of a variable's distribution against the quantiles of any of a number of test distributions. Probability plots are generally used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points cluster around a straight line (SPSS help).

B. Numerical Tests

1. Chi-Square Test
This is a goodness-of-fit test which can be applied for testing normality of a given set of data by breaking up the range of variable into intervals, and then comparing the observed frequencies in cells with those of the best fitting distribution. It is required that at least 20% of the cells must have frequencies less than 5. Comparing the computed Chi-Square value with the tabulated with degrees of freedom equal to $(k-1)(n-1)$, where n and k are categories in the contingency table. The null hypothesis that the

data come from the normal distribution is rejected if the computed value is greater than the tabulated value. If the probability of significance is computed then the null hypothesis is rejected if it is less than or equal to .05.

2. Wilk-Shapiro Test

   A goodness-of-fit test for normality that may be used for n less than or equal to 50. This is one of the most powerful test for normality. It arises out of the considerations of the probability plot (Walpole, 1986) The value must be reasonably near 1.0 for the data to be normal.

3. Kolmogorov-Smirnov One Sample test (K-S test)

   A goodness-of-fit test which is alternative to the Chi-Square. It can be used even if the sample size is small. It is more powerful than the Chi-Square test (Walpole, 1986).

4. Skewness

   Measures the degree of asymmetry of a given set of data. Values of skewness reasonably near zero indicates normality (Mood, et. al.,1974). This can be computed as the ratio of the third moment and the square root of the second moment.

5. Geary's Test

   This is the most powerful test for normality as compared to the Chi-Square test. This test is based on a very simple statistic which is a ratio of two estimators of the population standard deviation. Values of U differing considerably from 1.0 represent the signal that the hypothesis of normality should be rejected (Walpole,1982)

Since this method is not available in the SPSS and SX, a program in Turbo Pascal was made for the computation of the test statistic.

i. For small sample

Calculate

$$U = \frac{\sqrt{\pi/2}\sum_{i=1}^{n}|x_i - \overline{x}|/n}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2/n}}$$

The algorithm on how to compute the value of U is done step by step such as enumerated below.

a. Input the set of data x_i, $i = 1, 2, \ldots, n$.

b. Compute the sum and the mean of these data.

c. Get the absolute of the difference of each x_i and the mean.

d. Get the sum of the difference and divide by n then multiply by the square root of $\pi/2$.

e. Get the sum of the square of the difference of each x_i and the mean and divide this by n. Then take the square root of this quantity.

f. Divide the results in (d) and (e).

ii. Interpretation

Values of U differing considerably from 1.0 represents a signal of possible rejection of the hypothesis that the data are normally distributed.

For large sample $n \geq 35$:

a. Calculate the standardized U.

$$Z = \frac{U - 1}{0.2661\sqrt{n}}$$

b. Interpretation

Values of Z outside the region

$-Z_{\alpha/2} < Z < Z_{\alpha/2}$ signals non-normality of the data.

3 Methods

In this comparative study, eight normality tests were used, three graphical methods and five numerical methods as outlined in the previous section. The data used for the investigation were taken from the book entitled Time Series Analysis by Wei (1990) with varying sample sizes for both stationary and non-stationary data. Three different statistical software namely Statistical Package for Social Sciences (SPSS) version 10, Statistix version 4.1 (SX4.1), and Time Series Program (TSP) were utilized for the computation of the different statistics. Since the Geary's Test is not found in any of these programs, a program in Turbo Pascal was made to facilitate the computation of the statistics U for the test. The Turbo Pascal Program for the Geary's test is found in Cahoy (1998).

4 Summary of Findings

The findings of the comparative study of normality tests as applied to some time series data are summarized in Table 1 below.

1. In some cases, the normality tests for unordered and independent observations such as P-P plot, Q-Q plot, Rankit plot, Wilk-Shapiro Test, Kolmogorov-Smirnov one sample test, Chi-Square Test, and Skewness are applicable to time series data. Almost all of these tests are applicable to stationary time series data with no positive autocorrelations. Some tests are also applicable to non-stationary series. Table 1 shows that most of these tests failed to show normality for non-stationary time series data.

2. Geary's Test is both efficient in non-stationary and stationary time series data as displayed in Table 1. Kolmogorov-Smirnov One Sample test is also efficient in both non-stationary and stationary data when the observation is not more than 200. P-P plot, Q-Q plot, Rankit plot, Skewness and Chi-Square test are only applicable to stationary time series data.

Table 2 gives the summary the U values of the Geary's test computed using the Turbo Pascal Program. Table 2 shows the values of U which are close to 1.0 since the null hypothesis is accepted since the value of the probability of significance is greater than 0.05. T-test result confirmed that the values of U are not significantly different from zero. Hence the stationary and non-stationary time series data ranging from 55 to 300 indicate normality. Moreover, the values of U for both stationary and non-stationary time series data are not significantly different

	Stationary					Non-Stationary				
Normality Test	$N=$ 55	$N=$ 80	$N=$ 100	$N=$ 200	$N=$ 300	$N=$ 55	$N=$ 80	$N=$ 100	$N=$ 200	$N=$ 300
K-S Test	**	**	**	**	**	**	**	**		
Wilk-Shapiro	**	**	**	**	**	**	**	**	**	**
Chi-Square	**	**		**						
Geary's test	**	**	**	**	**	**	**	**	**	**
Skew-ness	**	**	**	**	**			**		**
Rankit Plot	**	**	**	**	**		**	**		
P-P Plot	**	**	**	**	**		**	**		
Q-Q Plor	**	**	**	**	**		**	**		

** indicates normality

Table 1: Normality Test Results Across Varying Sample Sizes (N)

as shown by the t-test result with probability of significance equal to 0.3676 which is greater than 0.05.

Sample Size	**Stationary**	**Non-Stationary**	**T-Test Result**
55	1.08	1.03	$H_{01}: U = 0$
80	1.01	0.99	$Prob. = 0.1918 > .05$
100	0.94	1.08	
200	1.04	0.99	$H_{02}: U1 = U2$
300	0.97	1.12	$Prob. = 0.3676 > .05$

Table 2: Summary of the U Values of the Geary's Test

3. Skewness performs well to detect normality for stationary data but for only relatively large sample sizes for nonstationary data as shown

in Table 1. The results shown in Cahoy (1998) indicated that the skewness is more reliable to use for stationary time series data. It must be mentioned that the values of the skewness must be close to zero for normal data.

4. In Table 1 the Wilk-Shapiro indicated normality of the time series data from small to large sizes since the values are close to 1 except for sample size 55 for non-stationary data which is relatively smaller than the rest of the values as displayed in Table 3. Although the Wilk-Shapiro Test gives values which are reasonably near 1.0 for both stationary and non-stationary cases, t-test result indicated that there is no sufficient evidence to conclude that the difference of the values of the Wilk-Shapiro for both stationary and non-stationary series is really zero (0) as shown in Table 3 since the probability of significance is less than 0.05.

Sample Size	**Stationary** WS1	**Non-Stationary** WS2	T-**Test Result**
55	0.9715	0.8806	$H_0 : WS1 = WS2$
80	0.9909	0.9746	$Prob. = 0.0206 < .05$
100	0.9973	0.9663	
200	0.9847	0.9091	
300	0.9934	0.9276	

Table 3: Summary of the Wilk-Shapiro Values

5 Conclusions

1. Geary's was shown to be a good test for normality if the time series data are either stationary or non-stationary for small and relatively large sample sizes.

2. The Wilk-Shapiro Test ranked second as an alternate test for normality for time series data which are either stationary or non-stationary.

3. The characteristic of the autocorrelations of the data has no effect to the Geary's and Wilk-Shapiro Tests.

4. For varying sample sizes the values of the Geary's test statistic is not significantly different for both the stationary and non-stationary time series data.

6 Future Direction

In light of the findings of this study, it is suggested that a theoretical study on the Geary's and Wilk-Shapiro Tests statistics be made which should include the derivation of the formula, the investigation on the robustness property of the statistic and the attempt to derive the asymptotic distribution of the statistic.

References

[1] Balagbis, S. S. *Survey on Some Normality Tests*, Unpublished Seminar Paper, MSU-IIT , 1994.

[2] Cahoy, Arlene V. *Normality Tests As Applied to Some Set of Time Series Data*, Unpublished Undergraduate Thesis submitted to the College of Science and Mathematics. MSU-IIT, 1998.

[3] Gasser, Theo *Biometrika*, United Kingdom, 1975.

[4] Mood,et. Al. *Introduction to the Theory of Statistics*, Tokyo: McGraw Hill, Inc., 1974.

[5] Moore, David S. *The Annals of Statistics*, 1998.

[6] Siegel, S. and Castellan J. *Nonparametric Statistics for Behavioral Sciences*, Singapore: McGraw Hill Inc., 1988.

[7] Walpole, Ronald E. *Introduction to Statistics*. New York. McMillan Publishing Company, Inc., 1986.

[8] Walpole, R. and Meyers, R. *Probability and Statistics for Engineers and Scientists*, McMillan Publishing Company, New York, 1994.

[9] Wardraw, A.C. *Practical Statistics for Experimental Biologists*. New York: John Wiley & Sons, 1985.

[10] Wei, William W.S. *Time Series Analysis*. California. Addison-Wiley Publishing Co., Inc., 1990.