Robustness of the Ljung-Box Test and its Rank Equivalent

Patrick Burns
patrick@burns-stat.com
http://www.burns-stat.com

This Draft: 6 October 2002

Abstract:

The Ljung-Box test is known to be robust. This paper reports on simulations that show just how robust it is in finite samples. Even so, we demonstrate some practical applications where the robustness of the test fails dramatically. The Ljung-Box test on the ranks of the data provides a suitably robust alternative when the distribution is extremely long-tailed. In particular, the rank Ljung-Box test is highly recommended over the Ljung-Box test for evaluating the adequacy of GARCH models. Simulations also explore properties of the test when applied to binary data. There is some evidence that the test starts to deteriorate as the number of lags exceeds 5% of the length of the series whether or not the data are long-tailed.

1. Introduction

The portmanteau test of Ljung and Box (1978) is commonly used to test the quality of fit of a time series model. If significant autocorrelation is not found in the residuals from the model, then the model is declared to pass the test. The Ljung-Box test is known to be robust to outliers, nonetheless, several robust alternatives have been proposed—examples are Li (1988) and Chan (1994). Pena and Rodriquez (2002) introduce a test that is generally more powerful in the Gaussian setting.

The present article investigates how robust the Ljung-Box test and its rank equivalent are. Applications in finance motivate the study. Simulations are used to explore both the null distribution and the power of the Ljung-Box test and the rank test in finite samples. This is done where the data come from the Gaussian distribution, several long-tailed distributions and some binary distributions. Runde (1997) looks at the asymptotic distribution under some infinite variance distributions.

The remainder of the article is arranged as follows. Some applications are highlighted in section 2 which informed the choices for the simulations. The Ljung-Box test is defined and its robustness discussed in section 3. Section 4 provides details of the simulations. Simulation results under the null hypothesis are presented for continuous distributions in section 5, and for binary distributions in section 6. The power of the test under continuous distributions is explored in section 7. Section 8 concludes.

2. Some Financial Applications

Three applications in which the Ljung-Box test may play a part are discussed. Although most distributions in finance are non-Gaussian, the distributions resulting from these applications are particularly far from the Gaussian. Extremely long-tailed distributions are encountered in the test of squared residuals from GARCH models. Binary series are used to test the quality of Value at Risk estimates, and the consistency of the quality of predictions.

Choices for the simulations were influenced by these applications.

GARCH Modeling

Engle (1982) and Bollerslev (1986) introduced GARCH models—these account for the volatility clustering that is often seen in the return series of market-priced assets. An example is the popular GARCH(1,1) model:

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}$$

where h_t is the variance at time t conditional on past information, and ε_t is the residual at time t. The three parameters of the model are α, β and ω—there would also generally be a parameter for the mean of the series.

Given an estimate of the parameters for a model, it is desirable to determine if the model adequately explains the variance process. A common approach is to divide each residual by the estimated standard deviation for that time point, and square these standardized residuals. Finally perform a Ljung-Box test on the squared standardized residuals (minus their mean). If the statistic is large, then there is evidence that the model is inadequate. Wong and Li (1995) studied the rank Ljung-Box test in this setting.

Consider the example of the S&P 500 for dates from 2 January 1985 through 31 December 2001. The return series for this data has 4292 observations. We'll start with a six-parameter model. The Ljung-Box test statistic with 15 lags for the model is 30.57, giving a p-value of 1%. This is as we expect since the model is known not be very good—it is a GARCH(0,4) model (that is, an ARCH(4) model) assuming a Gaussian distribution for the residuals. This model has four lags of the squared residual and no lags of the conditional variance.

Now we move to a model that is known to be (relatively) good for this dataset. It is a component model of Engle and Lee (1999) with leverage assuming the t-distribution for the residuals. The leverage is in the style of Glosten, Jagannathan and Runkle (1993). The model follows the two equations:

$$q_t = \omega + \rho q_{t-1} + \phi(\varepsilon_{t-1}^2 - q_{t-1})$$
$$h_t = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(h_{t-1} - q_{t-1}) + \lambda(\varepsilon_{t-1}^-)^2$$

where the parameters to be estimated are ω, ρ, ϕ, α, β and λ. Again, h_t is the conditional variance. The notation e^- means $-e$ when e is negative and zero otherwise. The fitted model has two additional parameters—one for the mean of the series, the other for the degrees of freedom of the t-distribution—making a total of 8 parameters.

The Ljung-Box test statistic for this new model is now 4.71, which has a p-value of 0.9943. We have done very well indeed by this measure. In fact the statistic is so small that we might consider rejecting in the other direction. That is, the statistic seems to be trying to tell us that we have overfit the data, but with 8 parameters on four thousand observations, this hardly seems likely.

The story changes when we consider the Ljung-Box test on the ranks of the squared standardized residuals. For the first model—the ARCH(4) —the rank Ljung-Box has a statistic of 97.6 which has a p-value of zero to double precision. The rank Ljung-Box has a statistic of 37.2 (p-value 0.0012) for the components model. Rather than saying the components model is perfect, the rank test implies that there is still some unexplained autocorrelation in the variance.

When fitting GARCH models, the number of daily observations should be at least a thousand, and a few thousand observations are often used. Maximum likelihood estimates of the degrees of freedom often fall in the range of 5 to 8 when assuming a t distribution for the residuals of daily returns. It has become typical to use 15 lags in the Ljung-Box test for GARCH models.

Value at Risk Estimation

Value at Risk (VaR) has become a standard tool in the field of risk management. Perhaps its greatest strength is that it is intuitive and easy to explain. It is the amount of money that we expect to lose more than with some given probability—usually 5% or 1%. However, its ease of explanation belies the difficulty of estimating it. The statistical task is to estimate a given quantile of a distribution that constantly changes. See Jorion (2000) for background and references.

The hit series of past VaR's is created in order to investigate the quality of a VaR estimator. This series is zero when the actual loss does not exceed the VaR, and one when it does. The mean of this series should be equal to the probability level. Also the series should not have autocorrelation. If it does, then a better VaR estimator exists since the autocorrelation implies that the probability of a hit is not constant and hence not always equal to the stated probability level. An alternative approach to testing VaR—the dynamic quantile test—is proposed in Engle and Manganelli (1999).

The Ljung-Box test and the rank Ljung-Box test are the same for hits as there are only two distinct values. Burns (2002) investigates a number of VaR estimators—tests of the 10-day VaR estimates where there are 1550 observations have a suspiciously high number of p-values very close to one for the better estimates. This is more pronounced for the 1% estimates than for the 5% estimates. In the 1% case it isn't hard to believe that the distribution might be off as we are expecting only about 15 non-zero values.

Correct Predictions

It is common to try to predict if the price of an asset will go up or down in a given time period in the future. A backtest of the predictions will give a binary series that is one if the prediction turned out to be correct and zero otherwise. Hence a binary series results that is approximately 50% ones. In this application autocorrelation is not necessarily a bad thing, but it is useful to know if it exists. Should autocorrelation be found, that knowledge may provide clues to improve the prediction.

3. An Examination of the Test

The lag k autocorrelation statistic of a time series x_t (with mean zero) of length n is:

$$r_k = \frac{\sum_{t=k+1}^{n} x_t x_{t-k}}{\sum_{t=1}^{n} x_t^2}$$

The M lag Ljung-Box statistic is defined as:

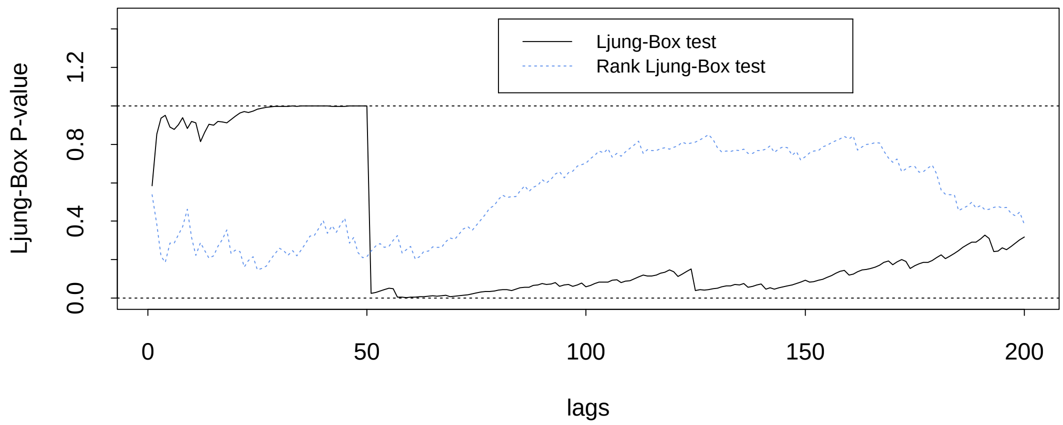
$$Q_M = n(n+2)\sum_{k=1}^{M} \frac{r_k^2}{n-k}$$

The equivalent rank test merely substitutes the mean-centered rank of x_t within the time series for the actual value of x_t. Thus it is simple to produce the rank test with existing software.

Let's examine the robustness of the Ljung-Box test from an abstract point of view. Suppose we have a white noise series of standard Gaussians, and two outliers are introduced that are on the order of 100, say. How does this affect the statistics r_k?

The denominator of r_k is greatly increased because of the two outliers—this tends to decrease the Ljung-Box statistic. For values of k not equal to the distance between the outliers, there will be no large terms in the numerator—only the products of the outliers with typical values. The value of r_k will be very large when k is equal to the distance between the outliers. Thus in this simplified world as the number of lags increases, the p-value will drift towards 1, suddenly drop towards 0, then drift larger again. Exhibit 1 shows similar behavior for the Ljung-Box test when the data are the squares of values from Student's t with 4 degrees of freedom. Since the series is a random sample, the p-value should ideally be neither small nor large—as we see with the rank Ljung-Box test.

Exhibit 1. Ljung-Box and rank Ljung-Box p-values for a dataset of random squared t with 4 degrees of freedom.



4. Simulation Details

A number of distributions are simulated. Each distribution, whether under the null hypothesis or given a particular parametric model in the alternative, is estimated with 10,000 replicates. In the AR(1) models for a given combination of distribution and number of observations, the same innovations are used for different AR parameters. Series of lengths 100, 1000 and 10,000 are studied. The lags of the Ljung-Box test and the rank test are 5, 15 and 50.

The continuous distributions that are studied are the Gaussian, Student's t with 10 degrees of freedom, the t with 4 degrees of freedom, and the t with 1 degree of freedom (i.e., the Cauchy). The square of each of the three t distributions is also studied. For the power simulations on the squared distributions, the AR(1) process is generated as usual and then each element of the series is squared. The probabilities of a one in the binary series are 50%, 5% and 1%.

The simulations were carried out in S-PLUS version 3.4 for Sun Solaris.

5. Null Distribution with Continuous Distributions

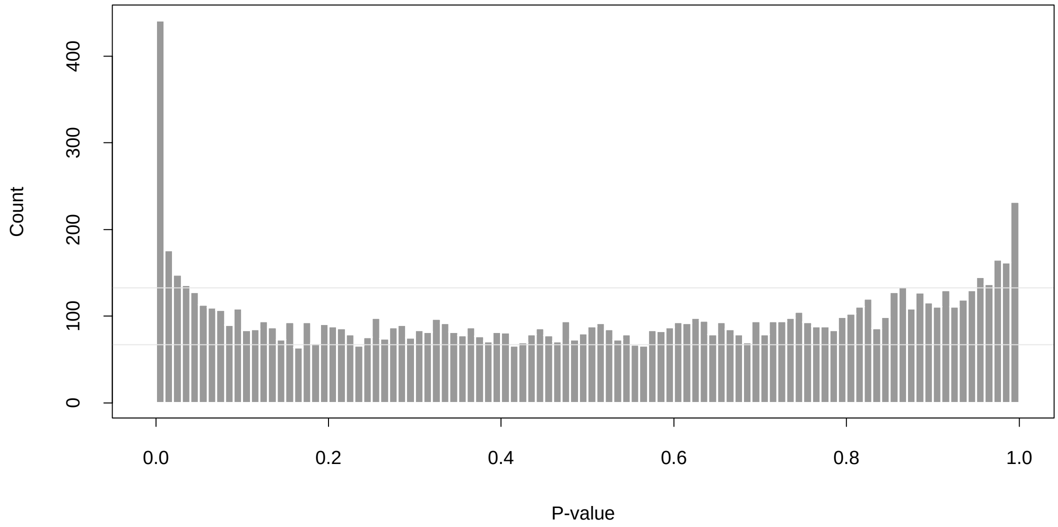
If the null hypothesis holds, then the p-value of a test should have a uniform(0,1) distribution. Exhibit 2 gives the results of Kolmogorov-Smirnov tests of the distribution of the p-value of the Ljung-Box and rank Ljung-Box under the Gaussian distribution when the null hypothesis of no autocorrelation applies. Bold entries in this and later exhibits indicate that the Kolmogorov-Smirnov test is not significant at the 5% level—that is, when there is little evidence to suppose that the test for autocorrelation is performing poorly.

Exhibit 2. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with Gaussian data.

# observations, # lags	Ljung-Box test	Rank Ljung-Box test
100 obs., 5 lags	**8.9e-3 (0.41)**	**0.011 (0.18)**
100 obs., 15 lags	0.021 (2.9e-4)	0.030 (3.4e-7)
100 obs., 50 lags	0.058 (0)	0.060 (0)
1000 obs., 5 lags	**5.9e-3 (0.88)**	**6.5e-3 (0.79)**
1000 obs., 15 lags	**0.010 (0.27)**	**7.0e-3 (0.71)**
1000 obs., 50 lags	**0.011 (0.17)**	**0.012 (0.12)**
10,000 obs., 5 lags	**8.6e-3 (0.45)**	**5.3e-3 (0.94)**
10,000 obs., 15 lags	**8.4e-3 (0.49)**	**7.0e-3 (0.71)**
10,000 obs., 50 lags	**5.6e-3 (0.92)**	**6.6e-3 (0.77)**

Both tests do well in this setting, though the rank test performs a little worse when the series is only 100 long. It is known that the number of lags for the Ljung-Box test should be small relative to the number of observations—Exhibit 2 provides some guidance in this regard. Exhibit 3 shows the distribution of p-values under the null hypothesis in the extreme case of 50 lags with only 100 observations.

Exhibit 3. Distribution of the 50-lag Ljung-Box p-value under the Gaussian distribution with 100 observations.



Exhibits 4 through 9 indicate the quality of the tests under the null hypothesis with non-Gaussian distributions. As the tails get longer, the Ljung-Box test deteriorates but it deteriorates slower for series with a large number of observations. The rank test remains stable relative to the tail length.

Exhibit 4. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with Student t 10 degrees of freedom data.

# observations, # lags	**Ljung-Box test**	**Rank Ljung-Box test**
100 obs., 5 lags	**8.6e-3 (0.44)**	0.016 (0.012)
100 obs., 15 lags	0.028 (6.0e-7)	0.025 (7.4e-6)
100 obs., 50 lags	0.079 (0)	0.061 (0)
1000 obs., 5 lags	**5.7e-3 (0.90)**	**9.2e-3 (0.36)**
1000 obs., 15 lags	**9.4e-3 (0.34)**	**7.8e-3 (0.58)**
1000 obs., 50 lags	0.015 (0.028)	0.021 (2.2e-4)
10,000 obs., 5 lags	**7.9e-3 (0.56)**	**4.8e-3 (0.98)**
10,000 obs., 15 lags	**8.4e-3 (0.48)**	**6.9e-3 (0.73)**
10,000 obs., 50 lags	**6.4e-3 (0.81)**	**9.6e-3 (0.31)**

Exhibit 5. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with Student t 4 degrees of freedom data.

# observations, # lags	**Ljung-Box test**	**Rank Ljung-Box test**
100 obs., 5 lags	0.034 (2.9e-7)	**0.010 (0.22)**
100 obs., 15 lags	0.059 (0)	0.028 (6.2e-7)
100 obs., 50 lags	0.12 (0)	0.066 (0)
1000 obs., 5 lags	**0.012 (0.13)**	**7.0e-3 (0.71)**
1000 obs., 15 lags	0.015 (0.017)	**7.6e-3 (0.61)**
1000 obs., 50 lags	0.028 (7.8e-7)	**0.011 (0.15)**
10,000 obs., 5 lags	**6.7e-3 (0.76)**	**8.8e-3 (0.41)**
10,000 obs., 15 lags	**9.1e-3 (0.38)**	**6.7e-3 (0.76)**
10,000 obs., 50 lags	**9.0e-3 (0.39)**	**8.7e-3 (0.44)**

Exhibit 6. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with Student t 1 degree of freedom (Cauchy) data.

# observations, # lags	**Ljung-Box test**	**Rank Ljung-Box test**
100 obs., 5 lags	0.49 (0)	0.016 (0.011)
100 obs., 15 lags	0.53 (0)	0.025 (1.2e-5)
100 obs., 50 lags	0.63 (0)	0.066 (0)
1000 obs., 5 lags	0.69 (0)	**0.011 (0.20)**
1000 obs., 15 lags	0.72 (0)	**6.1e-3 (0.86)**
1000 obs., 50 lags	0.71 (0)	0.016 (0.013)
10,000 obs., 5 lags	0.83 (0)	**8.6e-3 (0.45)**
10,000 obs., 15 lags	0.86 (0)	**0.012 (0.10)**
10,000 obs., 50 lags	0.85 (0)	**6.0e-3 (0.87)**

Exhibit 7. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with the square of Student t 10 degrees of freedom data.

# observations, # lags	**Ljung-Box test**	**Rank Ljung-Box test**
100 obs., 5 lags	0.13 (0)	0.017 (7.7e-3)
100 obs., 15 lags	0.19 (0)	0.027 (1.1e-6)
100 obs., 50 lags	0.28 (0)	0.065 (0)
1000 obs., 5 lags	0.062 (0)	**0.011 (0.20)**
1000 obs., 15 lags	0.083 (0)	**7.0e-3 (0.71)**
1000 obs., 50 lags	0.11 (0)	0.018 (2.6e-3)
10,000 obs., 5 lags	0.021 (4.4e-4)	**9.9e-3 (0.28)**
10,000 obs., 15 lags	0.033 (2.9e-7)	**0.011 (0.22)**
10,000 obs., 50 lags	0.023 (4.0e-5)	**7.2e-3 (0.67)**

Exhibit 8. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with the square of Student t 4 degrees of freedom data.

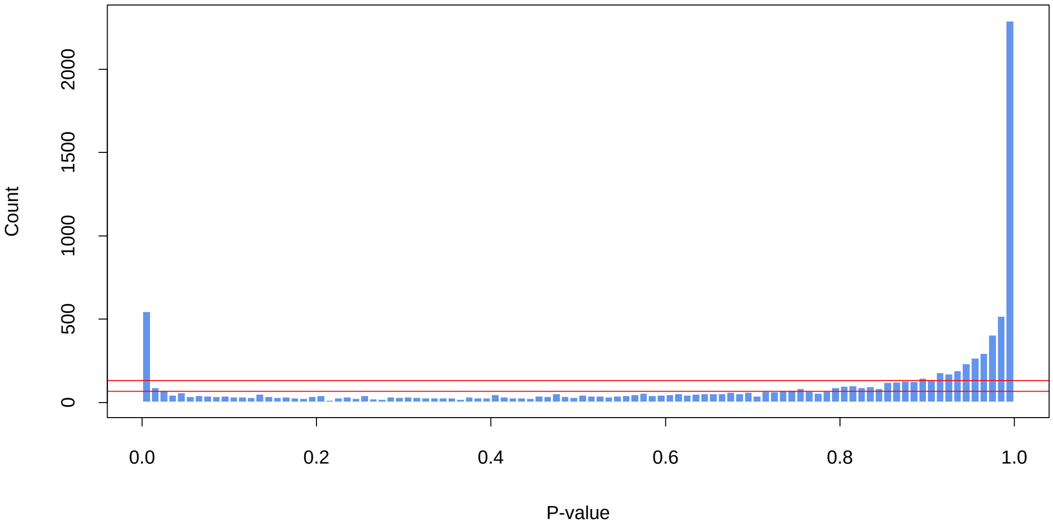
# observations, # lags	Ljung-Box test	Rank Ljung-Box test
100 obs., 5 lags	0.28 (0)	**0.012 (0.090)**
100 obs., 15 lags	0.34 (0)	0.027 (9.8e-7)
100 obs., 50 lags	0.45 (0)	0.063 (2.9e-7)
1000 obs., 5 lags	0.31 (0)	**9.2e-3 (0.37)**
1000 obs., 15 lags	0.36 (0)	**9.1e-3 (0.39)**
1000 obs., 50 lags	0.40 (0)	**0.011 (0.18)**
10,000 obs., 5 lags	0.31 (0)	**0.012 (0.10)**
10,000 obs., 15 lags	0.39 (0)	**7.2e-3 (0.68)**
10,000 obs., 50 lags	0.43 (0)	**6.5e-3 (0.79)**

Exhibit 9. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with the square of Cauchy data.

# observations, # lags	Ljung-Box test	Rank Ljung-Box test
100 obs., 5 lags	0.76 (0)	**0.013 (0.068)**
100 obs., 15 lags	0.78 (0)	0.022 (7.0e-5)
100 obs., 50 lags	0.83 (0)	0.065 (0)
1000 obs., 5 lags	0.92 (0)	**7.0e-3 (0.71)**
1000 obs., 15 lags	0.91 (0)	**0.013 (0.079)**
1000 obs., 50 lags	0.88 (0)	0.014 (0.036)
10,000 obs., 5 lags	0.98 (0)	**0.012 (0.11)**
10,000 obs., 15 lags	0.97 (0)	**8.4e-3 (0.48)**
10,000 obs., 50 lags	0.96 (0)	**8.1e-3 (0.52)**

Exhibit 10 shows the distribution of the p-value for the 15-lag Ljung-Box test when the data are the square of a t with 4 degrees of freedom and 10,000 observations. The distribution when there are 1000 observations is virtually identical. The null distribution of the p-value of the Ljung-Box tests is significantly far from the uniform for all of the squared distributions examined.

Exhibit 10. Distribution of the Ljung-Box 15 lag p-value for the square of a t with 4 degrees of freedom with 10,000 observations.



6. Null Distribution with Binary Data

Exhibits 11 through 13 evaluate the p-value distribution under the null hypothesis of the Ljung-Box test when evaluated on binary data. The distributions appear to be okay when the probability of a one is 50%—however, the farther from 50% the probability is, the worse the null distribution. The length of the series also has a large effect.

Exhibit 11. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with binary data—probability 50% of a one.

# observations, # lags	**Ljung-Box test**	
100 obs., 5 lags	**0.013 (0.063)**	
100 obs., 15 lags	0.027 (1.8e-6)	
100 obs., 50 lags	0.072 (0)	
1000 obs., 5 lags	**0.011 (0.21)**	
1000 obs., 15 lags	**7.6e-3 (0.60)**	
1000 obs., 50 lags	**0.014 (0.051)**	
10,000 obs., 5 lags	**0.012 (0.12)**	
10,000 obs., 15 lags	**0.011 (0.21)**	
10,000 obs., 50 lags	**5.4e-3 (0.93)**	

Exhibit 12. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with binary data—probability 5% of a one.

# observations, # lags	**Ljung-Box test**	
100 obs., 5 lags	0.28 (0)	
100 obs., 15 lags	0.25 (0)	
100 obs., 50 lags	0.33 (0)	
1000 obs., 5 lags	0.043 (0)	
1000 obs., 15 lags	0.040 (0)	
1000 obs., 50 lags	0.048 (0)	
10,000 obs., 5 lags	**0.013 (0.076)**	
10,000 obs., 15 lags	**9.4e-3 (0.34)**	
10,000 obs., 50 lags	**9.5e-3 (0.32)**	

Exhibit 13. Kolmogorov-Smirnov test statistic (p-value) of the p-value distribution under the null hypothesis with binary data—probability 1% of a one.

# observations, # lags	**Ljung-Box test**	
100 obs., 5 lags	0.91 (0)	
100 obs., 15 lags	0.83 (0)	
100 obs., 50 lags	0.77 (0)	
1000 obs., 5 lags	0.57 (0)	
1000 obs., 15 lags	0.31 (0)	
1000 obs., 50 lags	0.33 (0)	
10,000 obs., 5 lags	0.099 (0)	
10,000 obs., 15 lags	0.074 (0)	
10,000 obs., 50 lags	0.079 (0)	

7. Power under Continuous Distributions

The power of the Ljung-Box and rank Ljung-Box tests was found for an AR(1) model. The AR parameter ranges from 0 to 0.5 in increments of 0.05. The tests displayed have size 5%—tests with size 1% exhibit similar behavior. The results are also consistent across the number of lags. The first distribution to examine, of course, is the Gaussian—see Exhibits 14 through 16. There is very little cost for using the rank test rather than the Ljung-Box test when the data truly are Gaussian, at least under an AR(1) alternative.

Exhibit 14. Power of the Ljung-Box and rank Ljung-Box tests with Gaussian data with 1000 observations.

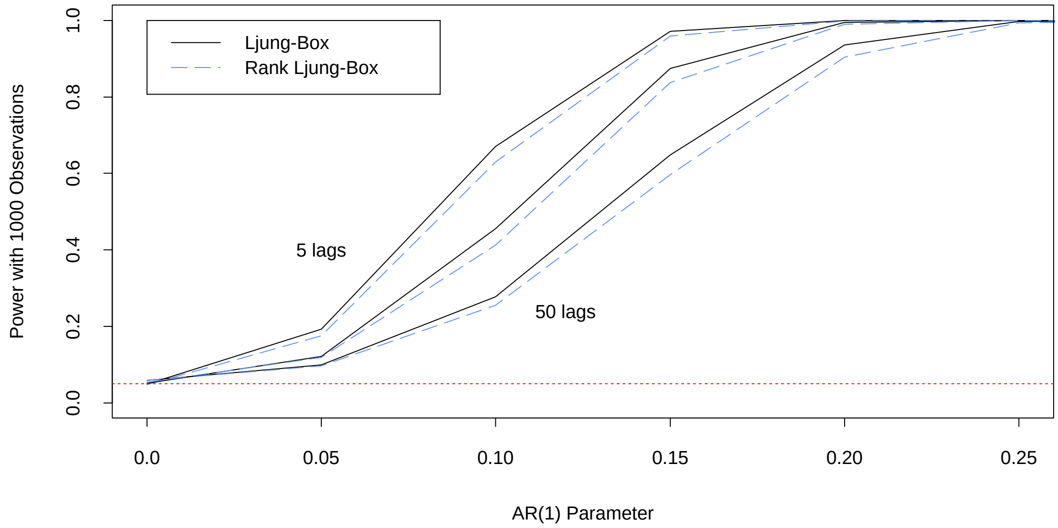


Exhibit 15. Power of the Ljung-Box and rank Ljung-Box tests with Gaussian data and 50 lags.

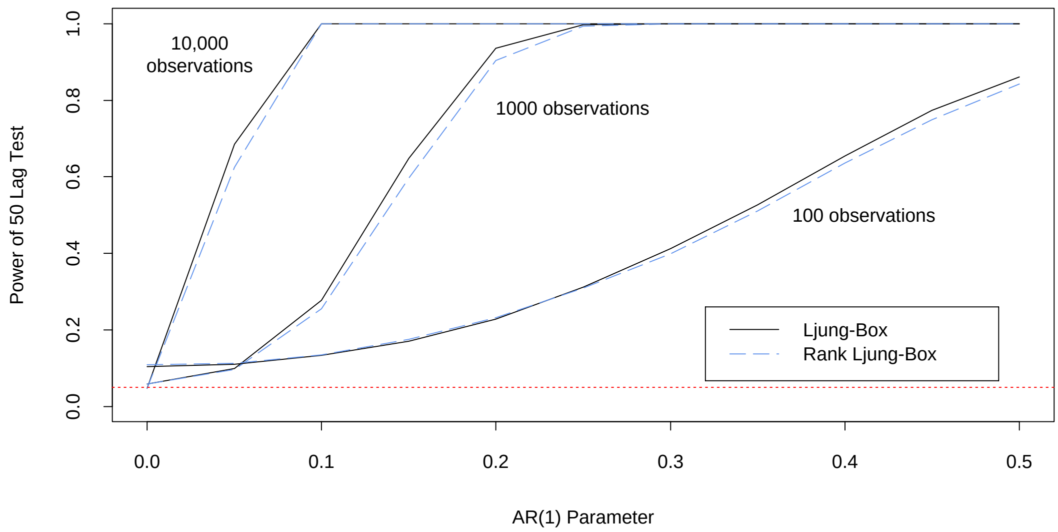
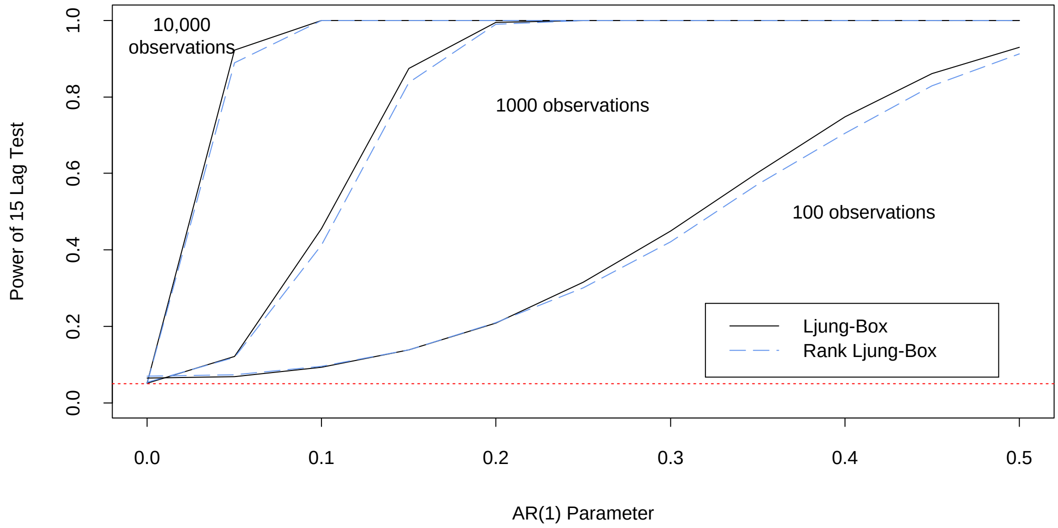


Exhibit 16. Power of the Ljung-Box and rank Ljung-Box tests with Gaussian data and 15 lags.



Exhibits 16 through 19 show the power for 15 lag tests for the symmetric distributions that were studied. The power of the Ljung-Box test remains remarkably stable as the tails lengthen. The power of the rank test gradually increases, though at the Cauchy it is much more powerful. Under the Cauchy the rank test with 1000 observations is as powerful as the Ljung-Box test with 10,000 observations.

Exhibit 17. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from a t distribution with 10 degrees of freedom.

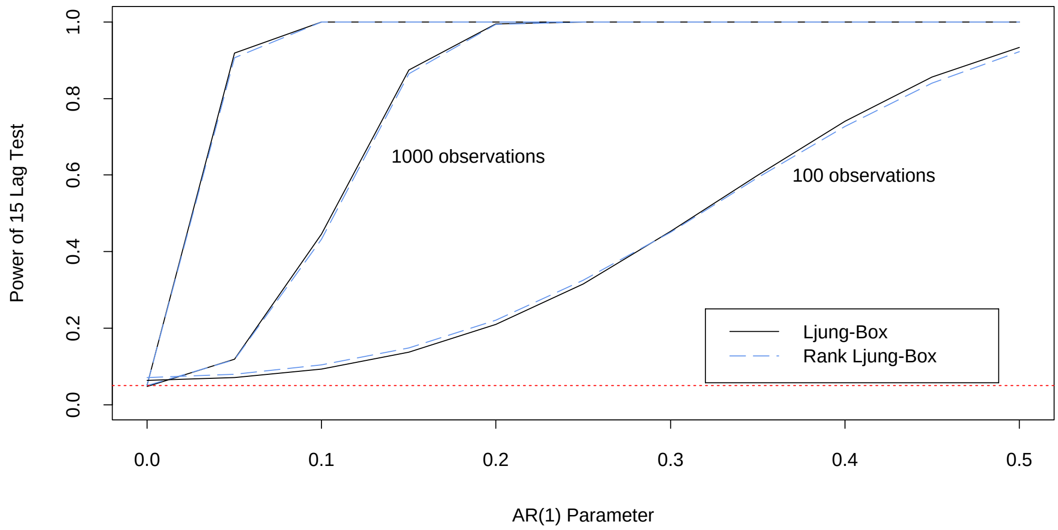


Exhibit 18. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from a t distribution with 4 degrees of freedom.

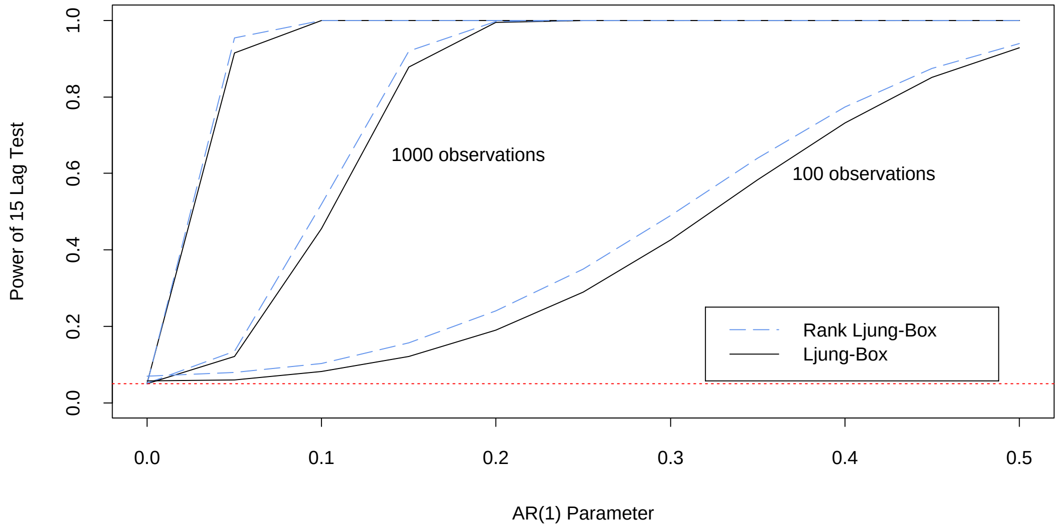
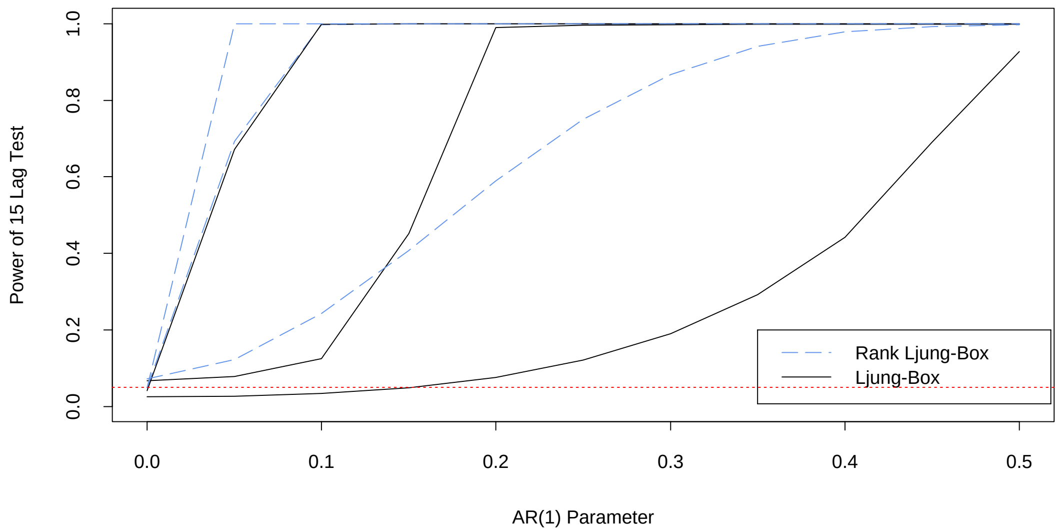


Exhibit 19. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from a Cauchy distribution. The series lengths are 10,000, 1000 and 100.



Exhibits 20 through 22 indicate the power for the squared t distributions. For these cases the Ljung-Box test has lost power relative to the symmetric distributions. The longer the tail in the squared distribution, the less power the Ljung-Box test has. The rank test, in contrast, gains power as the tails lengthen. Curiously, the Ljung-Box test has more power for an AR(1) alternative than the rank test under the square of the t with 10 degrees of freedom. However the null distribution for the Ljung-Box is decidedly wrong in this case, so the rank test will still be preferred.

Exhibit 20. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from the square of a t distribution with 10 degrees of freedom.

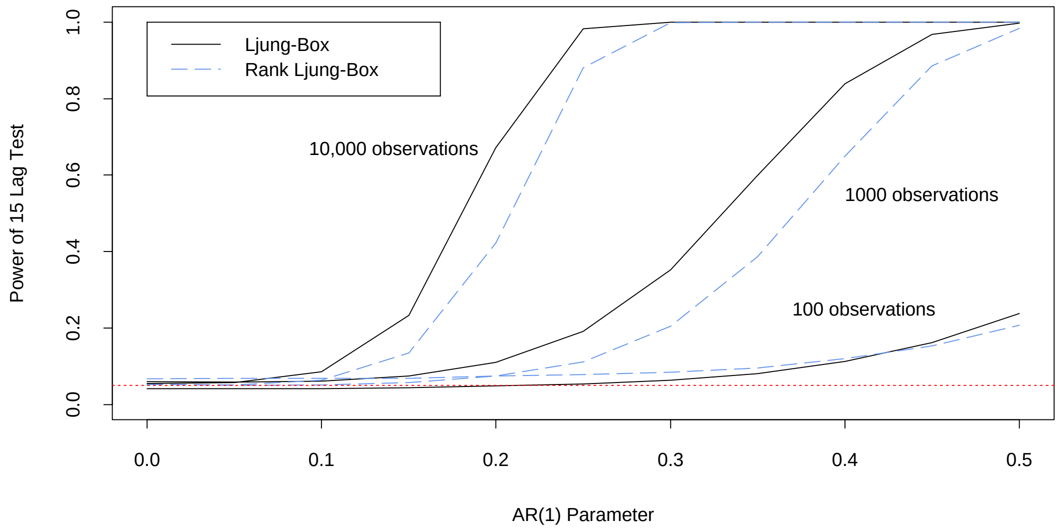


Exhibit 21. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from the square of a t distribution with 4 degrees of freedom.

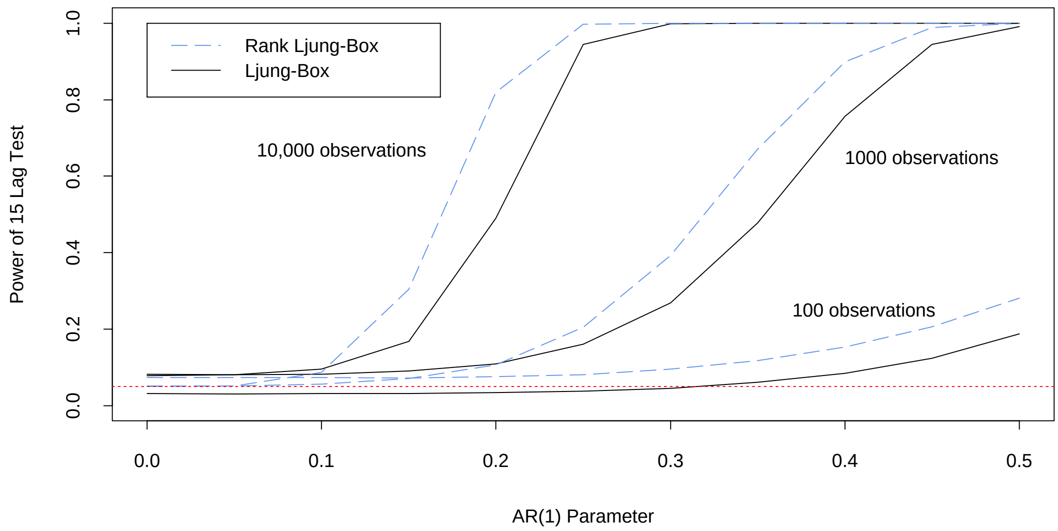
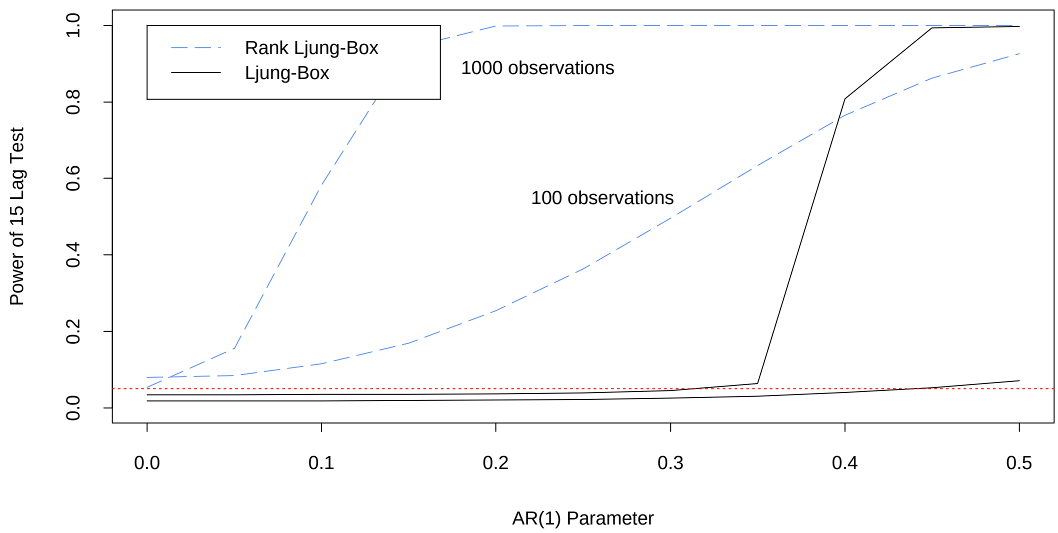


Exhibit 22. Power of 15 lag Ljung-Box and rank Ljung-Box tests for data from the square of a t distribution with 1 degree of freedom.



8. Conclusions

As long as the data have a distribution with tails no longer than a Student's t with 10 degrees of freedom, the Ljung-Box test is preferred to its rank equivalent. However there appears to be little loss of power with the rank test even for the Gaussian. As the tails become longer than a t with 10 degrees of freedom, the null distribution of the Ljung-Box suffers, and its power under an AR(1) alternative becomes inferior to that of the rank test.

Under the square of t-distributions—as when testing GARCH models—the rank Ljung-Box test should be used rather than the Ljung-Box test. As the tails get heavier, the Ljung-Box test loses power while the rank Ljung-Box gains power. The null distribution of the p-value for the Ljung-Box test is seriously suspect when testing autocorrelation of variance in cases of practical interest. The rank Ljung-Box test is well behaved for testing GARCH models.

It appears that the number of lags of Ljung-Box tests—whether standard or the rank equivalent—should be no more than about 5% of the length of the series, certainly less than 15%.

Applying Ljung-Box to binary data appears to be fine as long as the probability of each of the digits is approximately equal. The more unequal the probabilities become, the longer the series needs to be in order to have a reliable p-value. When the length of the series is 10,000, the null distribution is acceptable for a probability of 5%, but not 1%.

There are some limitations of the simulations that were performed that could benefit from further research. In particular, models for the alternative that go beyond the AR(1) would be interesting. A more careful study to determine the maximum number of lags that should be used relative to the length of the series would also be valuable.

References

Bollerslev, Tim (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **31** p307-327.

Burns, P. (2002). The quality of Value at Risk via univariate GARCH, Burns Statistics working paper. http://www.burns-stat.com

Chan, W. S. (1994). On portmanteau goodness-of-fit tests in robust time series modeling, *Computational Statistics* **9** 301-310.

Engle, Robert F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* **50**(4) p987-1007.

Engle, R. and G. J. Lee (1999). A permanent and transitory component model of stock return volatility, in ed. R. Engle and H. White, *Cointegration , Causality and Forecasting: A Festschrift in Honor of Clive W. J. Granger*. Oxford University Press. pp 475-497.

Engle, Robert F. and Simone Manganelli (1999). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, University of California, San Diego, Department of Economics Working Paper 99-20.

Glosten, L., R. Jagannathan, and D. Runkle (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48** 1779-1801.

Jorion, Philippe (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*. (2nd Edition). McGraw-Hill, New York.

Li, W. K. (1988). A goodness-of-fit test in robust time series modeling, *Biometrika* **75** 355-361.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models, *Biometrika* **65** p297-303.

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

Runde, R. (1997). The asymptotic null distribution of the Box-Pierce Q-statistic for random variables with infinite variance: An application to German stock returns, *Journal of Econometrics* **78** 205-216.

Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions, *Annals of Mathematical Statistics* 19 279-281.

Wong, H. and W. K. Li (1995). Portmanteau test for conditional heteroscedasticity, using ranks of squared residuals, *Journal of Applied Statistics* **22** 121-134.