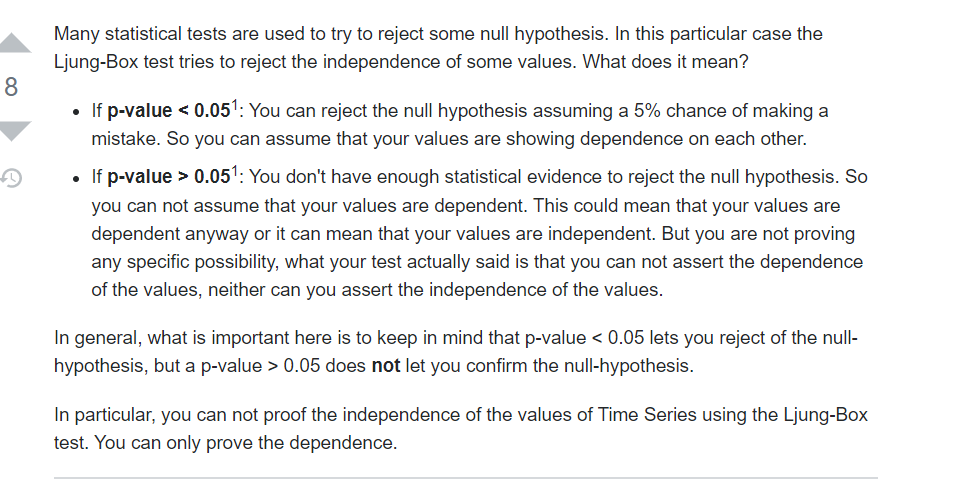
"scale\_down”

Residual diagnostics: Ljung Box test for white noise behaviour in residuals. Since the residuals have p-values>0.05 and we fail to reject the null hypothesis, there is no evidence of autocorrelation in the residuals.Hence, we may conclude that the residuals behave as white noise.

Test for ARCH beaviour in residuals: Looking at the standaridized squared residuals and ARCH LM Tests, the p-values>0.05 and we fail to reject the null hypothesis hence there is no evidence of serial correlation in squared residuals. This confirms that the residuals behave as a white noise process.



Arma

ARMA models are used to model the conditional  
expectation of a process given the past, but in an ARMA model the conditional variance given the past is constant. What does this mean for, say,  
modeling stock returns? Suppose we have noticed that recent daily returns  
have been unusually volatile. We might expect that tomorrow’s return is also  
more variable than usual. However, an ARMA model cannot capture this  
type of behavior because its conditional variance is constant. So we need better time series models if we want to model the nonconstant volatility. In this  
chapter we look at GARCH time series models that are becoming widely used  
in econometrics and finance because they have randomly varying volatility. (pag 405 em statdata)

process, such as a GARCH  
process, in which the conditional mean is constant but the conditional variance  
is nonconstant is an example of an uncorrelated but dependent process. The  
dependence of the conditional variance on the past causes the process to be  
dependent. The independence of the conditional mean on the past is the reason  
that the process is uncorrelated – 408

As we have seen, an AR(1) process has a nonconstant conditional mean but a  
constant conditional variance, while an ARCH(1) process is just the opposite.

If both the conditional mean and variance of the data depend on the past, then  
we can combine the two models. In fact, we can combine any ARMA model  
with any of the GARCH models in Sect. 14.6. In this section we combine an  
AR(1) model with an ARCH(1) model

Because past values of the *σt* process are fed back into the present value (with  
nonnegative coefficients *βj*), the conditional standard deviation can exhibit  
more persistent periods of high or low volatility than seen in an ARCH process. – pag 111

In fact, GARCH processes exhibit heavy  
tails even if *{t}* is Gaussian. Therefore, when we use GARCH models, we can  
model both the conditional heteroskedasticity and the heavy-tailed distributions of financial market data. Nonetheless, many financial time series have  
tails that are heavier than implied by a GARCH process with Gaussian *{t}*.  
To handle such data, one can assume that, instead of being Gaussian white  
noise, *{t}* is an i.i.d. white noise process with a heavy-tailed distribution. Pag.112

A GARCH model assumes the standardized errors (shocks, innovations) are i.i.d. with zero mean and unit variance. After having fit a GARCH model, it makes sense to test whether this is the case. Some common checks are to examine presence of autocorrelation and/or autoregressive conditional heteroskedasticity in the standardized errors; under the i.i.d. assumption, there should be none. If any is found, the model assumptions are violated, so the face value of the modeling results cannot be trusted.

Ljung-Box (LB) test on standardized residuals tests for autocorrelation in standardized errors, while LB test on standardized ***squared*** residuals and ARCH-LM test test for autoregressive conditional heteroskedasticity. Autocorrelation and autoregressive conditional heteroskedasticity are not the same. You can have one, the other or both in a time series. Hence, you should not be surprised if some tests find presence of one but not the other.

A problem with applying any of these tests to standardized (squared) residuals from a GARCH model is that the test statistics have nonstandard distributions under the null. (They have their standard null distributions when applied to raw data, but not when applied to residuals of a GARCH model.)\* As far as I know, this is not accounted for in the rugarch package. Hence, you should take the test results with a grain of salt.

\*There are papers and (I think) textbooks showing that ARCH-LM test should be substituted by Li-Mak test to have the correct distribution under the null if the mean of the process is modelled as a constant (not as ARMA as in your case). Similar corrections are needed for the LB tests. When the mean is not modelled as a constant, I am not sure whether there exists any test at all with a known null distribution. See my answer in the thread ["Remaining heteroskedasticity even after GARCH](https://stats.stackexchange.com/questions/271875/remaining-heteroskedasticity-even-after-garch-estimation/271952#271952)

Use the Ljung-Box q statistic to test whether a series of observations over time are random and independent. If observations are not independent, one observation can be correlated with a different observation k time units later, a relationship called autocorrelation. Autocorrelation can decrease the accuracy of a time-based predictive model, such as time series plot, and lead to misinterpretation of the data.

For example, an electronics company tracks monthly sales of batteries for five years. They want to use the data to develop a time series model to help forecast future sales. However, monthly sales might be affected by seasonal trends. For example, each year an increase in sales occurs when people buy batteries for Christmas toys. Thus a monthly sales observation in one year could be correlated with a monthly sales observations 12 months later (a lag of 12).

Before choosing their time series model, they can assess autocorrelation for the monthly differences in sales. The Ljung-Box Q (LBQ) statistic tests the null hypothesis that autocorrelations up to lag k equal zero (that is, the data values are random and independent up to a certain number of lags--in this case 12). If the LBQ is greater than a specified critical value, autocorrelations for one or more lags might be significantly different from zero, indicating the values are not random and independent over time.

LBQ is also used to assess assumptions after fitting a time series model, such as ARIMA, to ensure that the residuals are independent.

The Ljung-Box is a Portmanteau test and is a modified version of the Box-Pierce chi-square statistic.

The Goodness-of-Fit tests2 compare the empirical distribution of the standardized residuals with the theoretical ones from the specified  
density, which is Gaussian by default. The small *p*-values strongly reject the  
null hypothesis that the white noise standardized innovation process *{t}* is  
Gaussian p415

Interpret the key results for Chi-Square Goodness-of-Fit Test

[Learn more about Minitab](https://www.minitab.com/)

Complete the following steps to interpret a chi-square goodness-of-fit test. Key output includes the p-value and a bar chart of expected and observed values.

**In This Topic**

* [Step 1: Determine whether the observed values are statistically different from the expected values](https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/basic-statistics/tables/chi-square-goodness-of-fit-test/interpret-the-results/key-results/#step-1-determine-whether-the-observed-values-are-statistically-different-from-the-expected-values)
* [Step 2: Examine the difference between observed and expected values for each category](https://support.minitab.com/en-us/minitab-express/1/help-and-how-to/basic-statistics/tables/chi-square-goodness-of-fit-test/interpret-the-results/key-results/#step-2-examine-the-difference-between-observed-and-expected-values-for-each-category)

Step 1: Determine whether the observed values are statistically different from the expected values

Use the p-value to determine whether to reject or fail to reject the null hypothesis, which states that the population proportions in each category are consistent with the specified values in each category.

To determine whether the observed values from the sample and expected values from the specified distribution are statistically different, compare the p-value to the significance level. Usually, a significance level (denoted as α or alpha) of 0.05 works well. A significance level of 0.05 indicates a 5% risk of incorrectly rejecting the null hypothesis.

**P-value ≤ α: The observed data are statistically different from the expected values (Reject H0)**

If the p-value is less than or equal to the significance level, you reject the null hypothesis and conclude that the data does not follow a distribution with certain proportions. Use your specialized knowledge to determine whether the difference is practically significant.

**P-value > α: You cannot conclude that the observed data are statistically different from the expected values (Fail to reject H0)**

If the p-value is larger than the significance level, you fail to reject the null hypothesis because you do not have enough evidence to conclude that the data do not follow the distribution with specified proportions. However, you cannot conclude that the distributions are the same. A difference might exist, but your test might not have enough power to detect it.

|  |  |
| --- | --- |
| |  | | --- | | Chi-Square Test | |
| |  |  |  |  | | --- | --- | --- | --- | | N | DF | Chi-Sq | P-Value | | 225 | 3 | 0.65 | 0.8853 | |

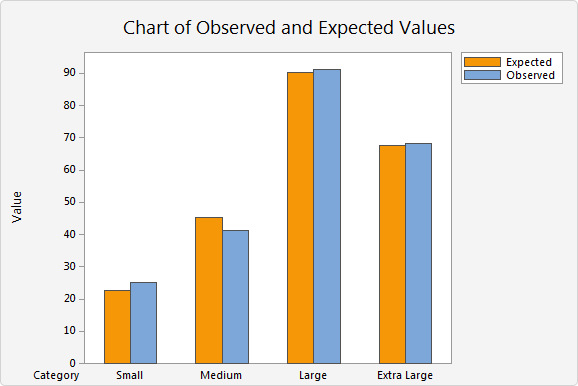
**Key Result: P-Value**

In these results, the p-value is 0.8853. Because the p-value is greater than the significance level of 0.05, you fail to reject the null hypothesis. Therefore, you cannot conclude that the observed proportions are significantly different from the specified proportions.

Step 2: Examine the difference between observed and expected values for each category

Use a bar chart that plots the observed and expected values for each category to determine whether there is a difference in a particular category.

If you determined that the difference between the observed and expected counts is statistically significant, then you can use this bar chart to determine which categories have the largest difference between observed and expected values.



This bar chart indicates that the observed values for each category are very similar to the expected values for each category. Thus, the bar chart visually confirms what the p-value indicates — that you cannot conclude that the observed proportions are different from the specified proportions.

**GARCH**

The GARCH(p,q) model has two characteristic parameters; p is the number of GARCH terms and q is the number of ARCH terms. GARCH(1,1) is defined by the following equation.

[GARCH equation](https://investexcel.net/wp-content/uploads/2011/10/GARCH-equation.png)

h is variance, ε is the residual squared, t denotes time. ω, α and β are empirical parameters determined by maximum likelihood estimation. The equation tells us that tomorrow’s variance is a function of

* today’s squared residual,
* today’s variance,
* the weighted average long-term variance

GARCH(1,1) captures only once square residual and one square variance.

This is not a magic wand, and financial analysts should be use the approach with a high degree of caution.  Given the appropriate circumstance, the predicted variance can greatly differ from the actual variance. Techniques such as the Ljung box text are used to determine if any autocorrelation remains in the residuals.

[Several researchers](http://ideas.repec.org/p/wpa/wuwpem/0411015.html) have highlighted deficiencies in GARCH(1,1) models, including its failure to predict the volatility in the S&P500 more accurately than other methods.

Lag is essentially delay. Just as correlation shows how much two timeseries are similar, autocorrelation describes how similar the time series is with itself.

Consider a discrete sequence of values, for lag 1, you compare your time series with a lagged time series, in other words you shift the time series by 1 before comparing it with itself. Proceed doing this for the entire length of time series by shifting it by 1 every time. You now have autocorrelation function.

From the values of autocorrelation function, you can see how much it correlates with itself. For any time series you will have perfect correlation at lag/delay = 0, since you're comparing same values with each other. As you shift your time series you begin to see the correlation values decreasing. Note that if timeseries comprises of completely random values, you will only have correlation at lag=0, and no correlation everywhere else. In most of the datasets/time series this is not the case, as values tend to decrease over time, thus having some correlation at low lag values.

Now, consider a long periodic time series, for example outdoor temperature over a few years, sampled hourly. Your time series will correlate with itself on daily basis (day/night temperature drop) as well as yearly (summer/winter temperatures). Lets say your first datapoint is at 1 pm in mid summer. Lag=1 represents one hour. The autocorrelation function at lag=1 will experience a slight decrease in correlation. At lag=12 you will have the lowest correlation of the day, after what it will begin to increase. Move forward 6 month to 1 pm. Your time series is still somewhat correlated. Move lag to 6 months and 1 am. This might be your lowest correlation point in the time series. At lag of 12 months your timeseries is again close to the peak value.

You might have noticed from the previous example that autocorrelation function reveals frequency components of a time series. Indeed, it is closely tied to frequency domain, and is just fourier transform from becoming a power spectra.

For a random time series, autocorrelation function will show you how quickly it becomes unsimilar with itself, while periodic time series will show at what delay/lag values time series is similar with itself.

Hope this isn't as confusing as it seems.