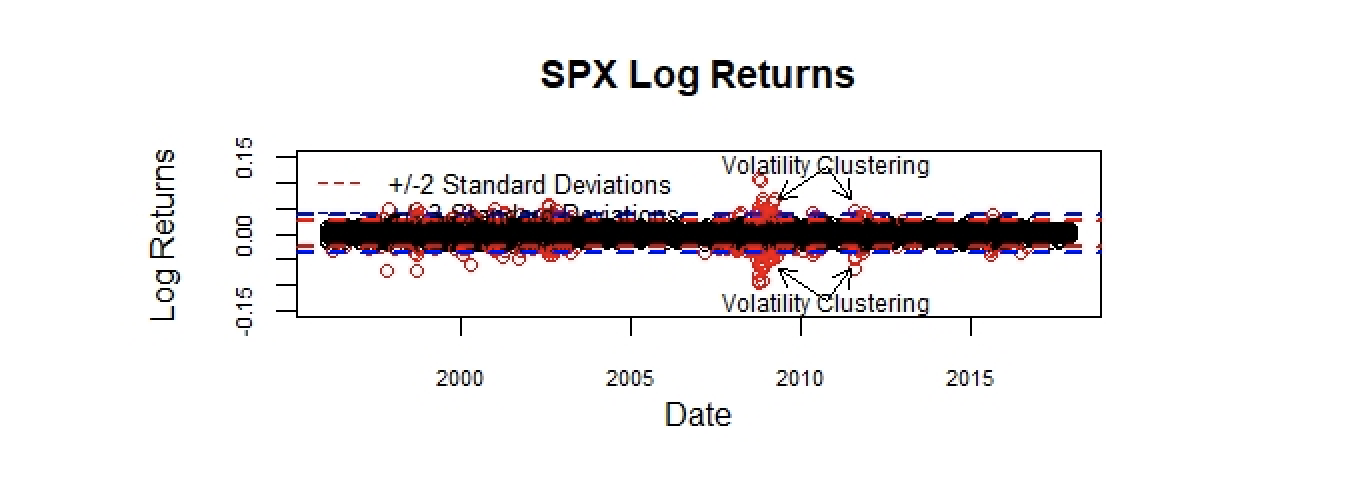
**Question 1: Describe the dataset and provide a table of summary statistics for the log-returns. Do the log-returns data appear normally distributed? If not, are the log-returns positive or negatively skewed and is there excess kurtosis? With the code, you will compute the Jarque-Bera test statistic. What is the Jarque-Bera test statistic and its null hypothesis? Do we fail to reject or fail to accept the null hypothesis?**

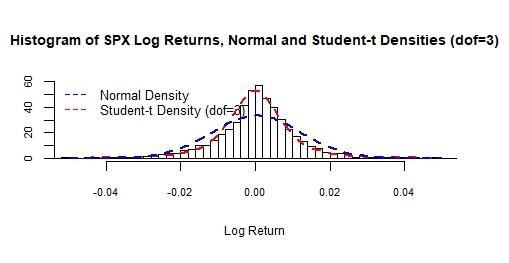
SPX return has an average return of a 0.02% and the SPX real vol mean is 0.16. The skewness of the distribution is -0.248 which is negative skewed. The kurtosis of the distribution is 8 which is excess kurtosis. The distribution is not normally distributed.

The null hypothesis of the Jarque-Bera test is a joint hypothesis of the skewness being zero and the excess kurtosis being zero. The p-value is 0 so we reject the null hypothesis.



**Question2: You will plot a histogram of returns along with a Normal and Student-t density superimposed. Does the histogram of log-returns appear better approximated by the Normal or Student-t density? Please provide the plot.**

As the following histogram shows, the distribution of SPX log returns appears to fit Student-t density better. This may be because our histogram is derived from the sample data that we collect, not the population. Academically speaking, the Normal distribution usually is used to depict the distribution of whole population, if we want to depict a distribution of a sample from whole population, the Student-t distribution may be better.

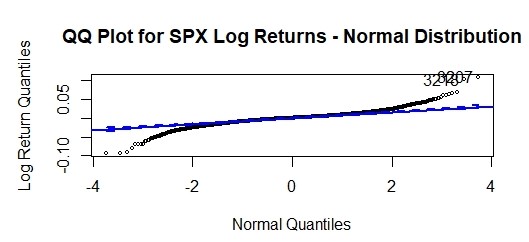
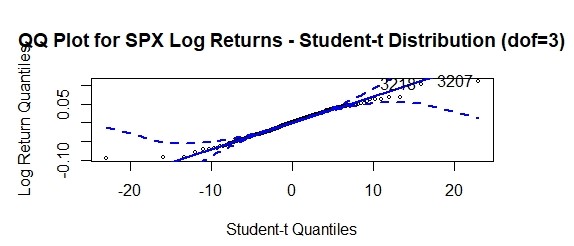


**Question 3. You will provide Quantile-Quantile (QQ) plots for each of the Normal and Student-t. What is a QQ plot? Which appears to fit better the Normal or Student-t?**

QQ plot is a visual tool for assessing non-normality. The actual name is quantile-quantile (QQ) plot, and the idea is to plot the quantiles of the calculated returns against the quantiles of the normal distribution. If the returns are truly normal, then the graph should look like a straight line on a 45-degree angle.

As the following two graphs show, the Student-t distribution appears to fit better than the Normal distribution. We can find that line of normal distribution cannot capture a lot of data points at both sides of tails as the second graph shows. However, in the first graph, we find that there are only several data points deviating from the line. Obviously, the Student-t distribution appears to fit better.

The reason why the Student-t distribution appears to fit better may be that the dataset is just a sample, not a population. And the Normal distribution is usually used to describe the distribution of a population. So we think if we used the data from whole population to get the results, the Normal distribution would appear to fit better than the Student-t distribution.

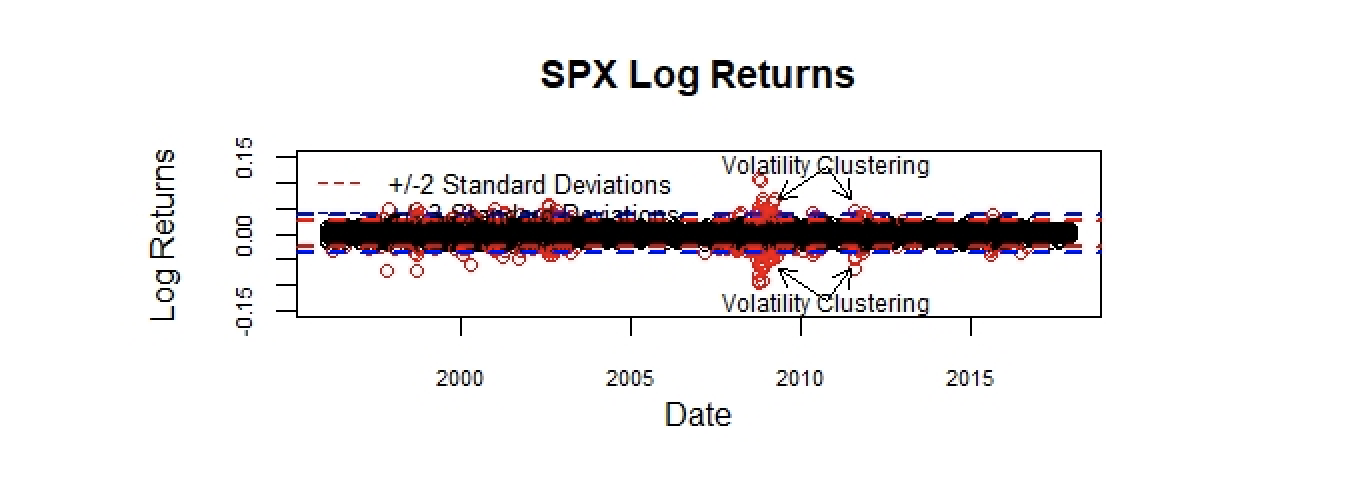


**Question 4: You will provide Autocorrelation Function plots for the log-returns and squared log-returns. What is the ACF? What do you find? Do the log-returns appear to display significant autocorrelation? Do the squared log-returns appear to display significant autocorrelation? If so, what does this mean?**

The ACF measures how returns on one day are correlated with returns on previous days. If such correlations are statistically significant, we have strong evidence for predictability. From the chart below, we can find out that log returns do not appear to display significant autocorrelation , but squared log returns appear to display significant correlation. That explains why we prefer squared data rather than the original. We have more confidence to predict squared log returns than the log returns themselves because they have strong autocorrelation so that more predictable from historic data.

**Question 5: What is meant by “volatility clustering”? Describe and provide a plot.**

Volatility clustering refers that "large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes." It also means volatility changes over time. Furthermore, given the apparent cycles, volatility is partially predictable.



**Question 6: How do we measure realized volatility? Is volatility directly observable or a property of returns? Is there more than one way to calculate actual realized volatility?**

Realized volatility measures what actually happened in the past and is based on taking intraday data, sampled at regular intervals (e.g., every 10 minutes), and using the data to obtain the covariance matrix. The main advantage is that it is purely data driven and there is no reliance on parametric models. The downside is that intraday data need to be available; such data are often difficult to obtain, hard to use, not very clean and frequently very expensive.

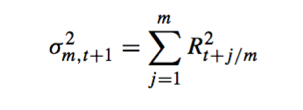
In addition, it is necessary to deal with diurnal patterns in volume and volatility when using realized volatility (i.e., address systematic changes in observed trading volume and volatility throughout the day). Moreover, the particular trading platform in use is likely to impose its own patterns on the data. All these issues complicate the implementation of realized volatility models.

Volatility is a property of returns. Consider the case where we have observations every 5 minutes on the price of a liquid asset, for example, the dollar/yen exchange rate. Let m be the number of observations per day. If we have 24-hour trading and 5-minute observations, then

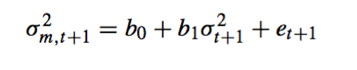
m = 24 ∗ 60/5 = 288. Let the j th observation on day t + 1 be denoted St+j/m. Then the closing price on day t + 1 is St+m/m = St+1, and the jth return is

Rt+j/m = ln(St+j/m) − ln(St+(j−1)/m)

Having m observations available within a day, we can calculate an estimate of the daily variance from the intraday squared returns simply as



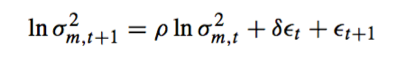
This variance measure could of course also be used instead of the squared return for evaluating the forecasts from variance models. Thus, we could run the regression,



But again, we want to go further and use the new variance measure directly for variance forecasting. The so-called realized variance measure noted earlier is, of course, only an estimate of the true variance. Under fairly general conditions it can be shown that as the number of intraday observations, m, gets infinitely large, the realized variance measure will converge to the true variance for day t + 1. Furthermore, for liquid securities, the distribution of the logarithm of σ 2 across days appears m,t +1 to be very close to the normal distribution. Thus, a very practical and sensible forecasting model of variance based on the realized variance measure would be, for example,



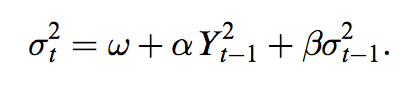
or perhaps



which are, respectively, an AR(1) and an ARMA(1,1) model in the log-realized volatilities. The AR(1) can be estimated using simple linear regression, while the ARMA(1,1) can be estimated using MLE.The ARMA(1,1) can be viewed as an AR(1) allowing for measurement error in the realized volatilities. When calculating the actual realized data, we can also use two days data or three days data or change other conditions and do the calculation.

**Question 7: Write down the formulaic expression for the standard GARCH(1,1) model. Explain the model and why it may prove useful in modeling the** **conditional variance of returns. What are the drawbacks to the standard GARCH(1,1) based on normal innovations? Discuss alternatives to the standard GARCH(1,1) to include other model formulations and distributions for the innovations.**

The simple GARCH model discussed is often referred to as the GARCH(1,1) model. GARCH(1,1) means the conditional variance of today’s return is equal to a constant, plus yesterday’s return squared and yesterday’s volatility squared. It relies on only one lag of returns squared and one lag of variance itself. The formula is:



Conditional volatility, σt, is typically, but not always, obtained from application of a statistical procedure to a sample of previous return observations, making up the estimation window. By including lagged volatility during ARCH model creation, we have the potential to incorporate the impact of historical returns. The drawback is that the assumption of the maximum likelihood is normal distribution but in reality it is not normally distributed. Therefore, there should be have bias.

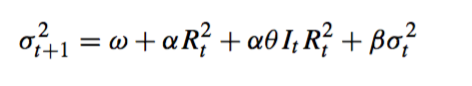
Normally, a negative return increases variance by more than a positive return of the same magnitude. This was referred to as the leverage effect, as a negative return on a stock implies a drop in the equity value, which implies that the company becomes more highly levered and thus riskier (assuming the level of debt stays constant). We can modify the GARCH models so that the weight given to the return depends on whether the return is positive or negative in the following simple manner:



which is sometimes referred to as the NGARCH (nonlinear GARCH) model.

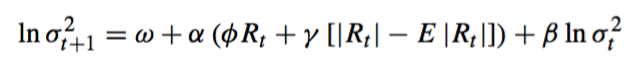
Notice that it is strictly speaking a positive piece of news, zt > 0, rather than raw return Rt , which has less of an impact on variance than a negative piece of news, if θ > 0. The persistence of variance in this model is α(1 + θ2) + β, and the long-run variance is σ2 = ω/(1 − α(1 + θ2) − β).

Another way of capturing the leverage effect is to define an indicator variable, It, to take on the value 1 if day t′s return is negative and zero otherwise. The variance dynamics can now be specified as



Thus, a θ larger than zero will again capture the leverage effect. This is sometimes referred to as the GJR-GARCH model.

A different that also captures the leverage is the exponential GARCH model or EGARCH,

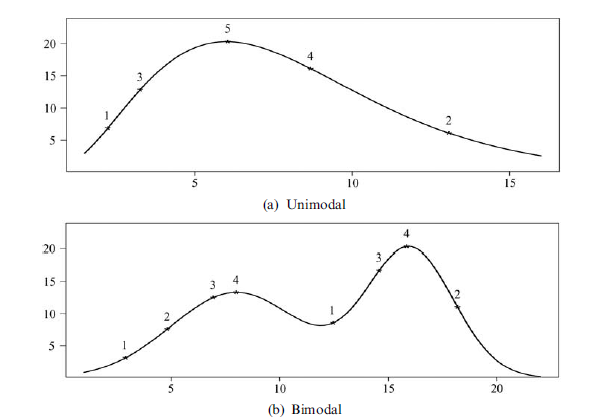


which displays the usual leverage effect if αφ < 0. The EGARCH model has the advantage that the logarithmic specification ensures that variance is always positive, but it has the disadvantage that the future expected variance beyond one period cannot be calculated analytically.

**Question 8**

**What is Maximum Likelihood Estimation (MLE)? What is the function that is being maximized for our GARCH (1, 1) with normal innovations?**

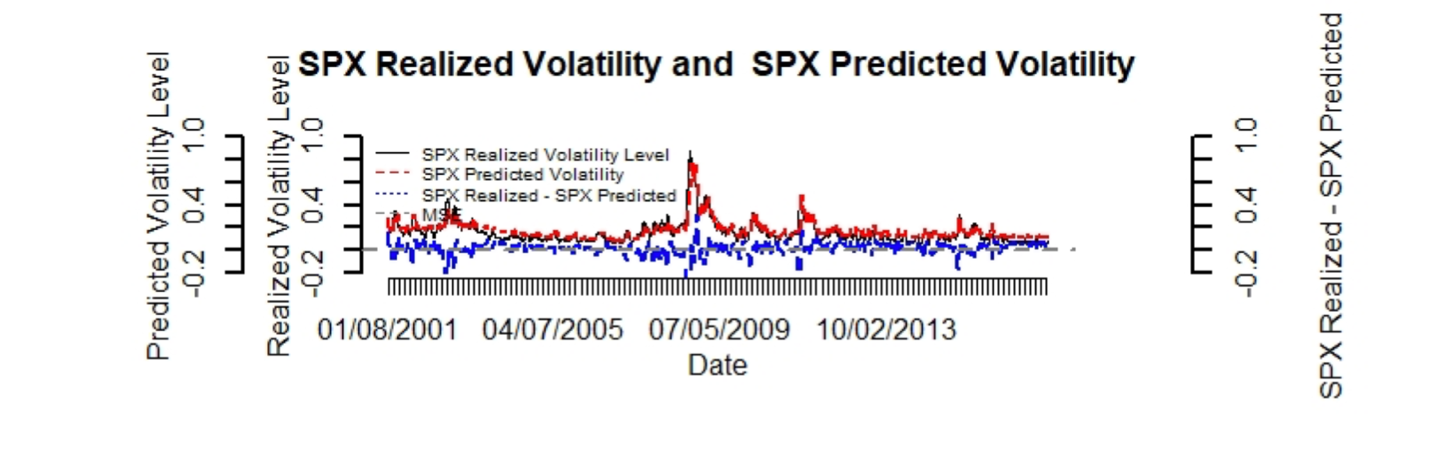
Maximum likelihood estimation is a method that determines values for the parameters of a model. The parameter values are found such that they maximize the likelihood that the process described by the model produced the data that were actually observed.



Some likelihood functions have multiple local minimum ---as in figure 8--- or long flat area. In such cases, finding the maximum is difficult to find the highest. While it may be possible to evaluate the likelihood function at the peak, we are searching for a solution and need to evaluate the likelihood for a large number of other combinations of parameters first. These problems are rare for smaller models such as GARCH(1,1), but become increasingly likely as models become richer.4 One way to guard against this is to try multiple starting values for the optimization.

**Question 9: After running the code provide a plot of the predicted versus actual realized one-month annualized volatility values? Does the GARCH(1,1) model appear to do a good job in modeling the dynamics of the conditional variance? Does it tend to over or underestimate the actual realized volatility? What happens during the 2008 financial crisis?**

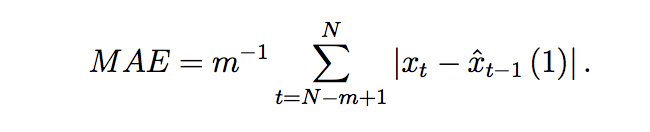
The GARCH (1,1) model does not do a good job in modeling the dynamics of the conditional variance. On average, the forecast overestimates the volatility during ordinary days and underestimate during the crisis. On the year of 2008, the spread of the reality and predicted volatility deepened, so the forecast on the volatility does not work and underestimate a lot during the crisis.

****

**Question 10**

**We calculate the Mean Squared Error (MSE) to evaluate the predictive power of the model. What are other loss functions that one may use?**

Mean absolute prediction errors (MAE), in the same notation as above, the mean absolute error is defined as



As m → ∞, The MAE should converge to E|xt − E(x|It−1)| or a slightly larger value, assuming this value exists. For a Gaussian world, this moment is proportional to the standard deviation, with a fixed proportionality factor. Also for other distributions, the absolute moment will measure the dispersion of the forecast errors.

The MAE is based on the loss function g (x) = |x|, which is more sensitive to small deviations from 0 and much less sensitive to large deviations than the usual squared loss. Therefore, the MAE can be viewed as a ‘robust’ measure of predictive accuracy. The MAE tends to prefer forecasting procedures that produce occasional large forecast failures, while they are reasonably good on average. By contrast, the MSE tends to prefer forecasting procedures that avoid large forecast failures, even though they produce a less satisfactory fit otherwise.

Because the estimation procedures are usually based on least-squares criteria, an emphasis on the MAE may involve a slight logical inconsistency. The best class of models is then selected according to a criterion that is different from the one that selects among the different members of an individual model class.