After conducting the backtesting, the conclusion is that different approaches produce different results, as well as that the results are dependent on number of days used for estimation. The most convenient way to present the results will be to produce them for different horizons and then to discuss the implications following from the results.

For the testing and estimation purposes, data for S&P 500 was taken for a period from 9th February 1990 to 7th March 2017, 10 days VaR and ES values were estimated using 2 years of observations for returns. The paper is organized in the following way. First part discusses methods and analyses the data, then VaR and ES estimations are compared using the whole sample and using only a part of the sample from 2009, last section concludes.

# Comparison of approaches and data properties

# Historical simulation (HS) in an essence is taking an alpha quantile of historical returns as an estimation for a VaR. As it is noted Danielsson (2011), HS “relies on the assumption that history repeats itself” (p.95). Here 3 different approaches to historical simulation are discussed.

# The first approach involves estimation of 1-day VaR using historical simulation (HS) and then scaling it to 10 days. For a precise estimation of VaR using simple HS there should be no structural breaks in volatility. It might be useful to at least visually check if there were structural breaks in volatility over this 20 years horizon.

# 

Figure : S&P 500 daily volatility for each overlapping 10-day period

# As it is seen from the Figure 1 there were periods of low volatility as well as periods of high volatility, this might lead to inaccuracies with a simple HS approach, because periods with sudden volatility increases can lead to an overestimation or underestimation of VaR. Another issue with this approach is scaling by square root of time horizon, because such scaling is appropriate only when returns are normal and i.i.d, however as Danielsson (2011) argues, this way is almost correct for 1% VaR (p.179), so it should not be an issue here. As it discussed in Danielsson (2011), HS approach tends to perform better than parametric alternatives in the absence of structural breaks, because it does not contain an estimation error.

# The second approach is to use non-overlapping 10 day returns and HS to estimate 10-day VaR, same issue as before applies here, there are structural breaks in volatility which make VaR estimation inaccurate. Other than that, using only non-overlapping 10-day periods requires large amount of data for a meaningful estimation, i.e. to find 1% VaR at least 1000 days are required for estimation, however if only 1000 days are used ES will be equal to VaR. According to Danielsson (2011), the minimum sample size for HS should be 3/p (p.98), which in this case is 300 non-overlapping 10-day periods or 3000 daily returns, that is 12 years. Using such an old data to estimate VaR with HS is hardly justifiable, especially in the presence of structural breaks.

# The third approach is filtered historical simulation (FHS) introduced in Barone-Adesi *et al.* (1998, 1999). It uses asymmetric GARCH to model the volatility of the returns. The observed returns are then divided by the GARCH estimated volatility to obtain standardised residuals, which are i.i.d. GARCH forecast of the next day volatility is multiplied by a random residual from the sample estimated earlier, this procedure simulates the return for the next day. This return is then used to update GARCH prediction for the second day, which is then multiplied by a new random residual from the sample. This procedure is repeated until the end of VaR horizon. Then estimated returns are summed to obtain the return for the whole VaR horizon. The whole operation is repeated for thousands of simulations to obtain 10-day returns distribution, alpha quantile of this distribution gives a VaR estimate. This approach to VaR estimation can capture the regime-switching behaviour of volatility as well as returns and volatility clustering that are presented in the sample used here (Figure 2).

# H:\New folder\portfolio risk project\autocorr3.jpg

Figure : Autocorrelation in returns and squared returns of S&P 500

# From the discussion of this section, following predictions can be made. Out of the three approaches FHS should give the most accurate value for VaR and ES estimations when the whole data sample is used, however simple HS should outperform FHS when only part of the sample without structural breaks in volatility used. The second approach is expected to produce poor results in comparison with the first and the third ones. To properly test the second approach and compare it with others, large sample should be used for VaR estimation.

# Estimation and Backtesting

To estimate and backtest VaR, the whole data sample is divided into estimation and testing windows, 9 days are taken from the end of the sample to calculate 10-day returns. For the first test, estimation window is set to 500 days and testing window is 6321 days. To estimate VaR for the first 10 days of a testing window with HS, returns from estimation window were sorted in ascending order, and the 5th return corresponding to alpha quantile is taken as a VaR estimation, this VaR is then scaled to 10 days by multiplying it by square root of 10. For the second approach 500 days were split into fifty 10-day periods, which are sorted in ascending order, because 50 is not enough observations to find a number corresponding to an alpha quantile, the smallest return is taken as both VaR and ES.

Third approach involves fitting an asymmetric GARCH model to 500 days of returns, instead of asymmetric GARCH used in Barone-Adesi *et al.* (1998), model by Glosten *et al.* (1993) is used here (GJR-GARCH). As it is discussed in Alexander (2008), GJR-GARCH is an “alternative formulation” of asymmetric GARCH and both models can be used (p.150). GJR-GARCH has a following form:

,

where stands for volatility, is a constant corresponding to a long-run volatility level, I (.) is an indicator function, ε is a residual and , is a random variable, stand for ARCH, leverage and GARCH components coefficients, respectively. Estimated model is used to find conditional variances for an each day of the estimation period. For calculation of conditional variances mean of squared returns used as an input for ARCH, leverage and GARCH components, as it is advised in Hamilton (1994, p.667). Historical returns are then divided by estimated conditional variances to obtain standardized residuals which are used for simulation as it was discussed above.

To backtest VaR, statistics for violation ratio (VR) and Bernoulli Coverage[[1]](#footnote-1) test are used, for ES backtesting, normalized shortfall (NS) is used. VR is simply number of exceedances divided by the expected number of exceedances. NS is a mean of ratios of observed returns on VaR violation dates to ES (Danielsson 2011, p.160). Results are summarized in Table 1.

Table : Results of estimation and backtesting for 500 days estimation window and 6321 days testing window

|  |  |  |  |
| --- | --- | --- | --- |
| Approach | VaR Violation Ratio | Bernoulli Coverage Test (p-value) | Normalized Shortfall |
| Historical Simulation | 1.0615 | 0.2362 (0.6270) | 1.0692 |
| Historical Simulation with non-overlapping 10 day periods | 2.5665 | -- | 1.4659 |
| Filtered Historical Simulation | 1.0298 | 0.056 (0.8129) | 0.9523 |

As it was predicted FHS performed better than both of HS methods, especially HS with non-overlapping 10-day periods that had number of violations 2.5 times higher than expected. From the results of Bernoulli coverage test it follows that null hypothesis of violation ratio equal to one cannot be rejected for both first and third approaches, the test statistics for the second approach could not be calculated.

To properly estimate VaR and ES values with the second approach, another test is conducted. Estimation window is set to 3000 days, and testing window to 3821. New results are shown in Table 2.

Table

|  |  |  |  |
| --- | --- | --- | --- |
| Approach | VR | Bernoulli Coverage Test (p-value) | Normalized Shortfall |
| Historical Simulation | 1.0493 | 0.0922 (0.7614) | 1.0931 |
| Historical Simulation with non-overlapping 10 day periods | 1.0493 | 0.0922 (0.7614) | 1.2175 |
| Filtered Historical Simulation | 0.5247 | 10.5266 (0.0012) | 1.036 |

Both HS and HS with non-overlapping periods has performed quite well now, VR for both is the same and close to one, even though exceedances occurred on different days, also Bernoulli Coverage test statistics show no rejection of H0. However NS for the second approach is high enough to say that it underestimates the ES. FHS performed quite poorly now with violation ratio equal to 0.52, which indicates that VaR was overestimated. A possible explanation for such results of FHS is a parametric error that builds up over a long horizon, so a GARCH component coefficient of GJR-GARCH model might be too high because of a relatively calm volatility periods, so an increase in volatility after a structural break persists longer and consequently leads to an overestimation of the VaR (Figure 3, Figure 4).

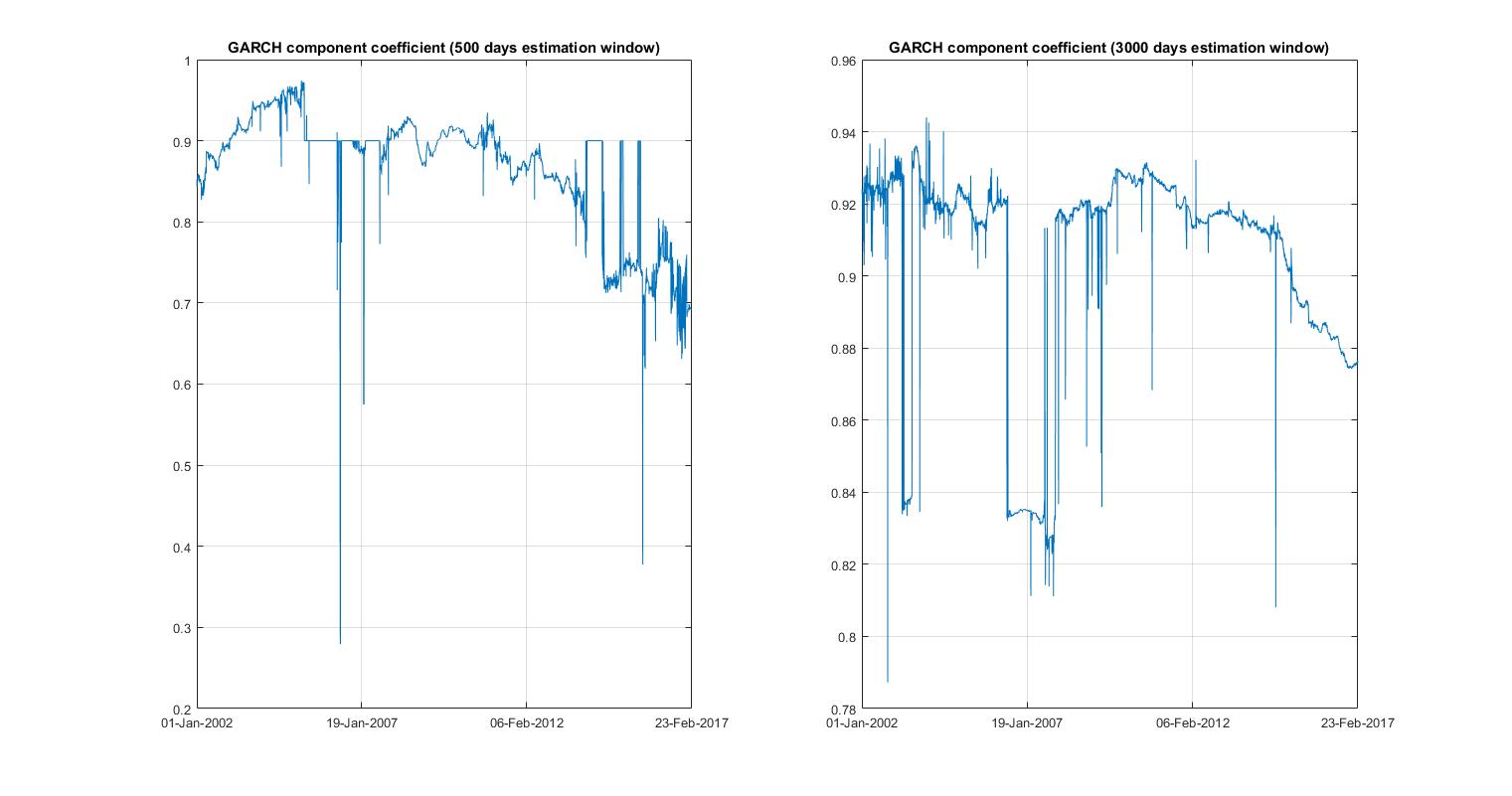


Figure : Comparison of GARCH component coefficients with different estimation windows

However, comparing the results of two tests in this way might be wrong because testing windows covered different time periods, reducing the whole sample for the first test so as to make testing periods the same gives following results (Table 3).

Table 3

|  |  |  |  |
| --- | --- | --- | --- |
| Approach | VR | Bernoulli Coverage Test (p-value) | Normalized Shortfall |
| Historical Simulation | 1.0231 | 0.0204 (0.8865) | 1.1838 |
| Historical Simulation with non-overlapping 10 day periods | 2.3610 | 51.5892 (0.0000) | 1.4897 |
| Filtered Historical Simulation | 0.8657 | 0.7277 (0.3936) | 0.9599 |

It might be useful to consider how two tests used earlier would work when data is reduced to a period only after the Financial Crisis of 2007-2008.

* Describe the characteristics of each approach (using the relevant literature to identify the key issues)
* Discuss the practicalities of implementation of each approach
* Discuss and compare the results. This necessitates backtesting the models over a one year period.
* Evidence of additional reading/research
* 1day VaR scaled to 10day using √time
* Non-overlapping 10day periods
* Overlapping 10day periods and Filtered Historical Simulation

1. Bernoulli coverage test is discussed in details in Danielsson (2011, pp. 154-155) [↑](#footnote-ref-1)