Asset Management Area

Risk Management Team

Stefano Serio (Head of Asset Management)
+39 3385756999

Getano Barberi (Risk Analyst)

Leonardo Astegiano (Risk Analyst)

Summary:

I. Main Results

II. Introduction

III. Historical VAR

- Base Case
- Approaches to address varying distributions
- Rolling Filtered Historical

IV. Normality Hypothesis

- Rolling Parametric – EWMA – simplification
- Rolling VAR-COV
- Evidences against normality

V. Montecarlo Simulations

- Rolling Montecarlo – T-Student
- Bootstrapping

VI. Educational Series #1

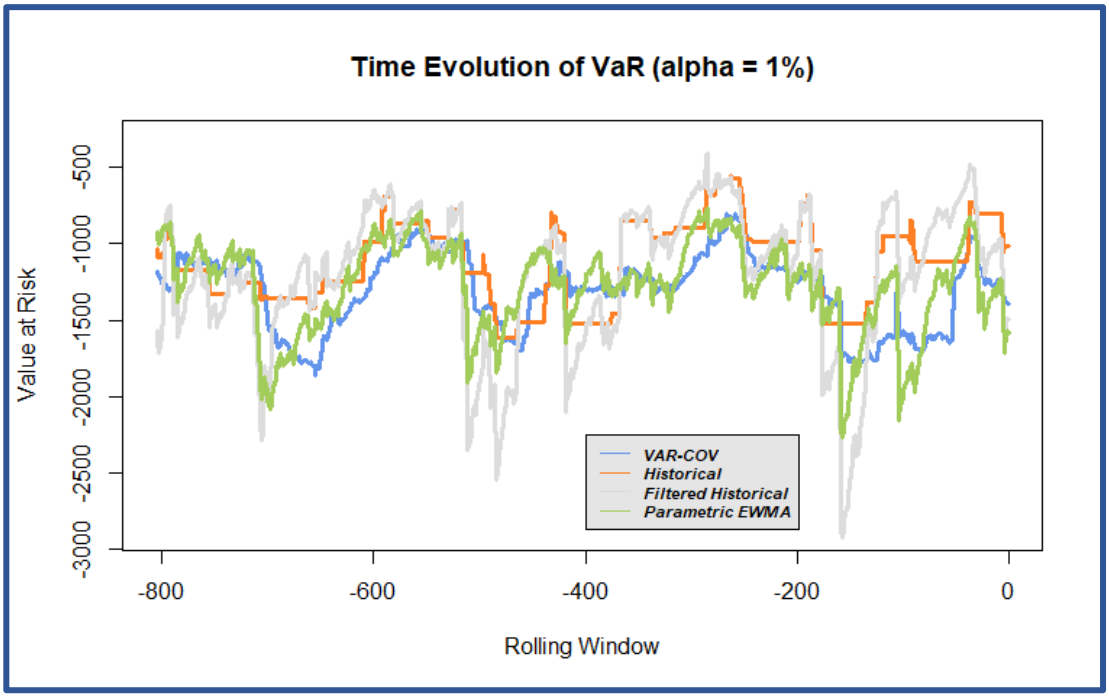
Warning

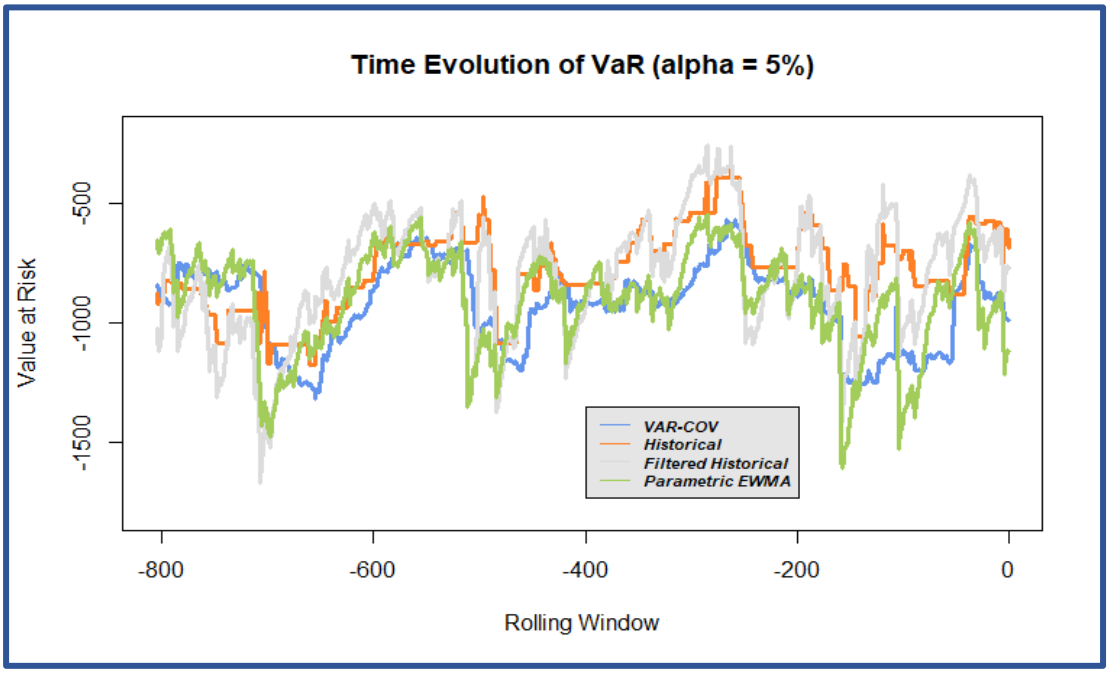
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

I. Main Results

Methods

	Historical	Filtered Historical	Parametric EWMA	Variance-Covariance	Montecarlo Simulations
Value at Risk (α = 5%)	-626 €	-676 €	-967 €	-901 €	-918 €
Value at Risk (α = 1%)	-895 €	-1,129 €	-1,367 €	-1,274 €	-1,520 €





Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

II. Introduction

The main objective of this report is to quantify the risk embedded in the Minerva IMS Active Portfolio composed by the Investment Division. The point of view chosen is that of a daily perspective on the potential extreme behaviour of a basket of assets, with the future goal of sending signals to the portfolio manager to marginally review its stocks' selection or the weights assigned to each asset. As Investment Risk division, our focus is the estimation of the Daily Value at Risk (henceforth, VaR): the maximum portfolio loss that occurs with α% of probability over a time horizon of 1 day.

We downloaded the historical daily prices (from 01/07/2015 to 26/10/2018) of the 40 stocks that compose Minerva IMS Active Portfolio and, for any of them, we computed daily log returns over time. Hence, our panel dataset is composed by 857 observations (*i.e* log returns) for every stock. Minerva IMS Active Portfolio is a so-called "Zero-Investment Portfolio" with 20 long positions and 20 short positions, whose inception date is 26/10/2018. Please read the report of Minerva Asset Management division for more details about the Active Portfolio composition. Since it is a zero-exposure portfolio, it is interesting to point out that, if all the asset had the same distribution of returns, we would have a zero VaR exposure too.

Unlike the previous report, we will describe the time evolution of Portfolio VaR in order to enhance the precision of our final estimate. To do so, Value at Risk analyses will be performed on homogeneous rolling windows of 52 consecutive log returns for every stock. A generic rolling window is a subset of the full dataset that contains 52 consecutive observations for every stock. Since the full dataset has 857 log returns for every stock and a single rolling window contains 52 consecutive observations, VaR computation will be performed on 806 rolling windows. By construction, rolling windows can be ordered in time from the most recent to the least recent. Computing daily Value at Risk for each rolling window will provide us with 806 different estimates. Hence, we will apply the exponential smoothing average on the 806 different estimates in order to find a number that best approximate the maximum portfolio loss that occurs with α% of probability over a time horizon of 1 day.

In our framework, the formula for the exponential smoothing average is:

$$\frac{\sum_{i=-805}^{0} \lambda^{-i} \times PortfolioVar_i(\alpha\%)}{\sum_{i=-805}^{0} \lambda^{-i}}$$

where $\lambda = 0.94$ is the smoothing coefficient, and $PortfolioVar_i(\alpha\%)$ is the estimated VaR of the i^{th} rolling window.

Warning

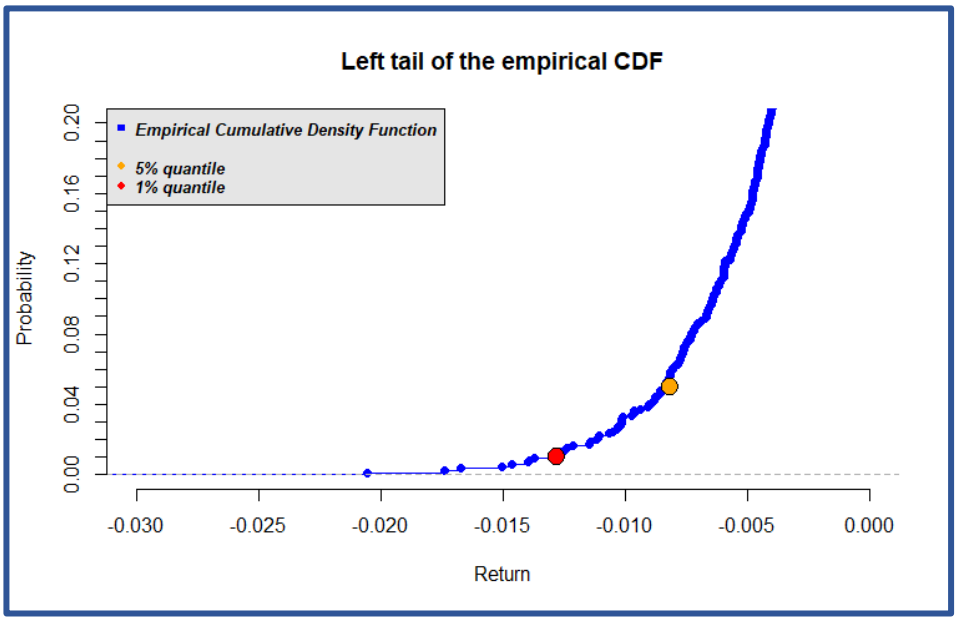
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

III. Historical VAR

- *Base Case*

This approach is valuable since it does not require any assumptions on the distribution of returns. We simply suppose that returns are identically and independently distributed. This estimate is robust to different hypothesis but, on the other side, sampling variability is high.

We computed the empirical distribution of returns using 3 years of historical data, then we estimated the cumulative distribution function using the empirical frequency.



This method implies that each observation has the same weight. As a consequence, past and new observations improve the estimate in the same way.

- *Approaches to address varying distributions*

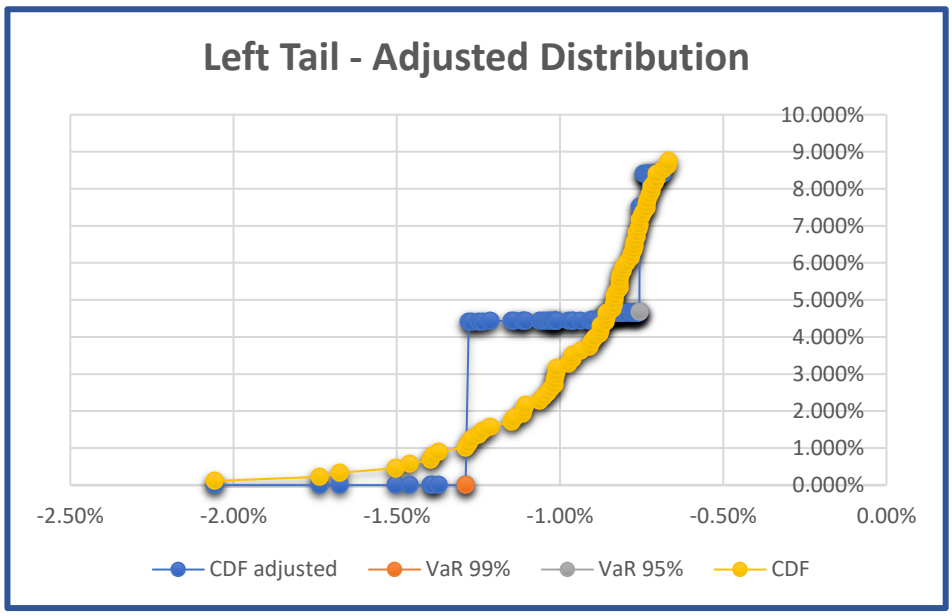
An alternative view is that variance of returns is time dependent. As a consequence, past returns reflect market conditions that may be different from current ones. We used two different approaches to have an overall understanding of this behaviour, and both are based on a smoothing estimate:

Warning

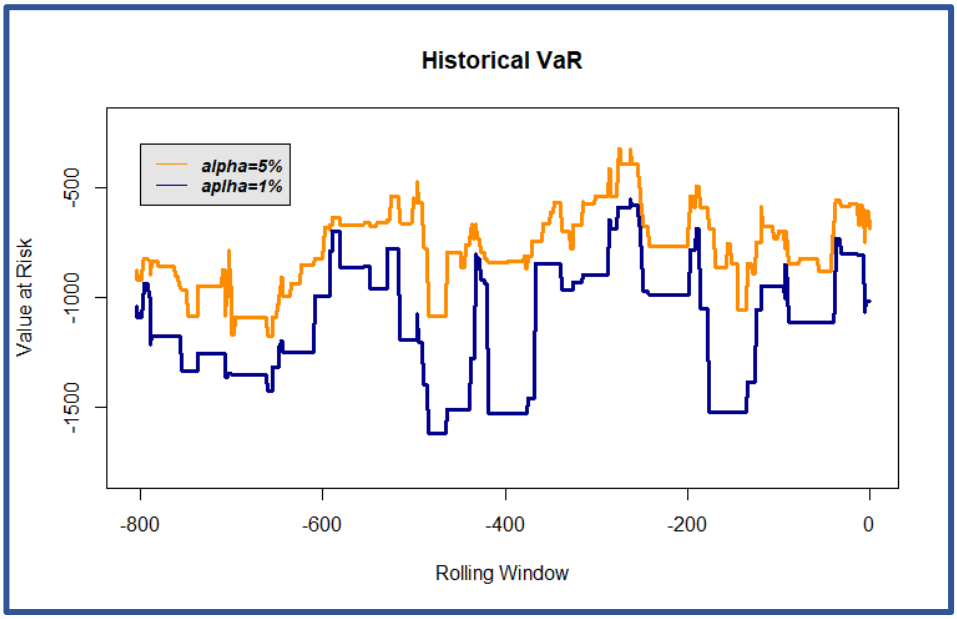
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

1) The first approach requires to assign an exponentially decreasing weight to each return observation in the dataset and to compute the empirical cumulative distribution function using these weights. The estimated Portfolio VaR (α%) corresponds to the α quantile of the empirical cumulative density

   function of returns (henceforth CDF). This means that the probability to observe a more negative portfolio return that the one identified by Portfolio VaR (α%) is α%.



2) The second approach is slightly more complex. The assumptions are the same, but instead of using directly the full sample of returns, we divide it in subgroups. For each subgroup, we extract the α-quantile obtaining a vector of 806 VaR estimates. This procedure allows us to understand the time evolution of daily VaR in the period under review. The line graph below shows the time evolution of daily Historical VaR between 01/07/2015 and 26/10/2018.



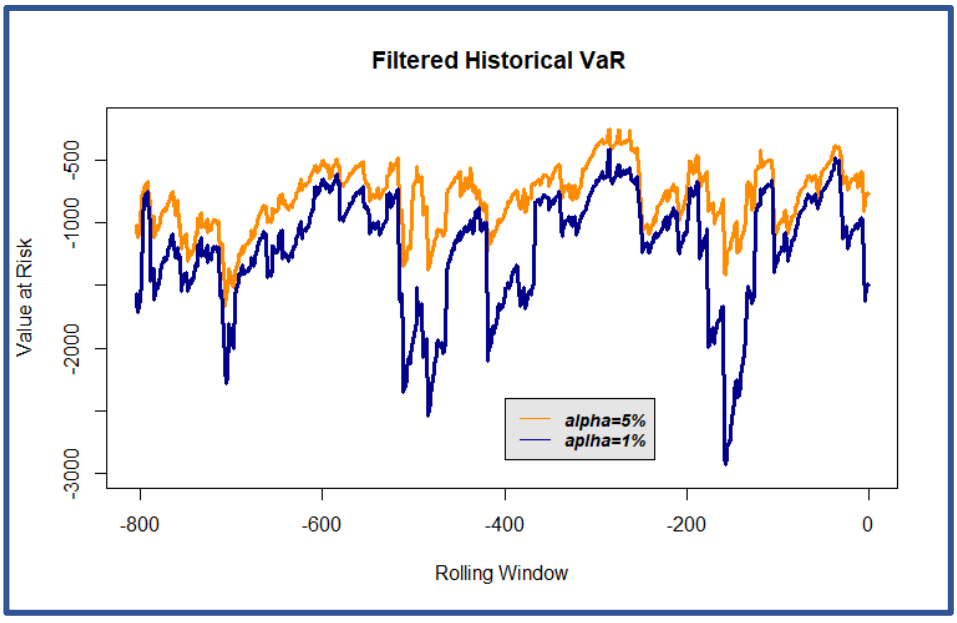
Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

One of the main drawbacks of this approach, however, is the Echo effect. Looking at the previous plot, in fact, we notice a strange behavior of the Roll VaR, it happens to be generally stable for may rolls and then suddenly drops or explode stabilizing to another level. This is due to the fact that, since we are using stable weights for past observations, when they get older, they still receive the same weight in the computation of Var. Thus, when extreme observation leaves the sample, the Var experience a strong variation.

- *Filtered Historical*

This approach is based on historical returns, but they are scaled using some *filter* or *scaling factor* in order to consider possible changes in the current level of risk. The filter used is the ratio between consecutive variance estimate, so that if current conditions are less volatile, they dump down the risk estimate. We used the *RiskMetrics* method to estimate variance, under the assumption of zero mean. We standardized the returns using the variance estimate of the previous period and we computed the α-quantile on a rolling base. Once we obtained the vector of observation, we transformed back them using the current variance. The line graph below shows the time evolution of daily Filtered Historical VaR between 01/07/2015 and 26/10/2018.



Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

IV. Normality Hypothesis

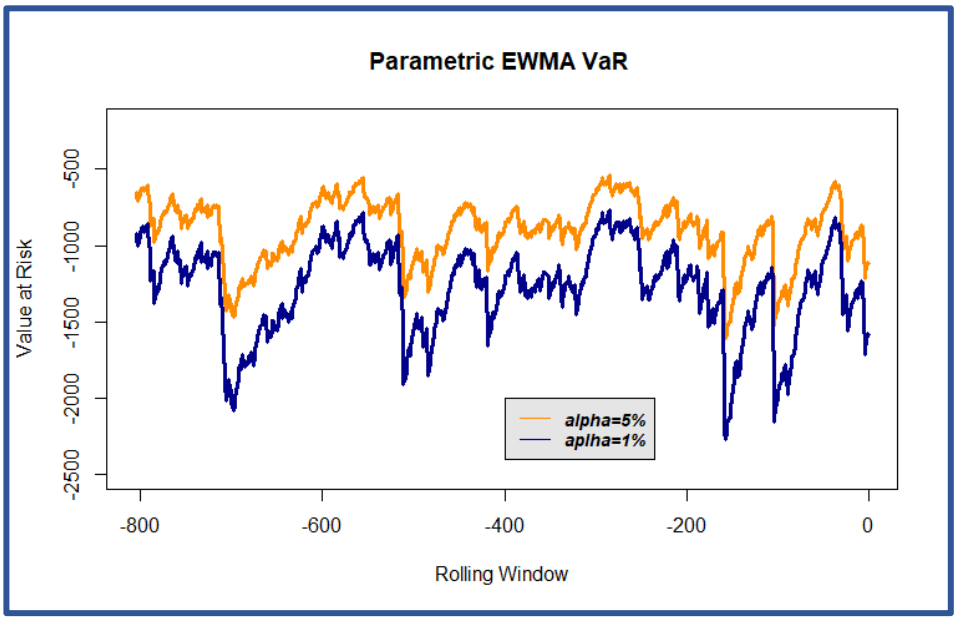
- *Rolling Parametric – EWMA – simplification*

This approach requires returns to be normally distributed. We estimated the variance on a rolling basis using *RiskMetrics* smoothing estimate, and we adopted a conservative approach assuming the expected return to be constant in time and equal to zero. Hence, a consistent estimate for the α-quantile is:

$$r_\alpha = z_\alpha \hat{\sigma}$$

where z_α is the α-quantile of a standard normal distribution $Z \sim N(0;1)$.

The line graph below shows the time evolution of daily Parametric EWMA VaR between 01/07/2015 and 26/10/2018.



This approach is a simplified one, because we model directly the returns of the portfolio, instead of decomposing them into sub-assets returns. We will relax this assumption in the following section. However, compared to the base historical approach, we are overcoming the Echo effect. In fact, observations receive ever reducing weights when they get older and older. Thus, leaving the sample, they will only marginally affect the Var.

Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

- *Rolling VAR-COV*

Under the assumption that log returns of each stock are normally distributed, we implemented rolling variance-covariance approach to estimate Portfolio VaR.

For a generic rolling window of 52 observations, the procedure can be summarized as follow:

1) We estimated historical standard deviations and correlation matrix of the stocks in our portfolio.

2) According to *RiskMetrics* procedure, we set the expected return of every asset equal to 0. Hence, exploiting Normality assumption, we computed the alpha quantile r_α of the return distribution of every stocks as

$$r_{i,\alpha} = \sigma_i \times z_\alpha \quad \forall \mathrm{i} = 1 \dots 40$$

where σ_i is the estimated standard deviation of the i^{th} stock and z_α is the alpha quantile of a standard normal distribution $Z \sim N(0; 1)$. Note that using rolling windows technique, we did not constrain the standard deviations of each stock to be constant throughout the analysis rather we allow it to change. As an example, there is no reason to believe the standard deviation of the stock Lockheed Martin in rolling window 0 is the same as the standard deviation of the same stock in rolling window 806 because of the three years gap between the two rolling windows.

3) We obtained an estimate of Value at Risk for each single position in our portfolio multiplying the alpha quantile of a stock by the monetary amount invested in it. Since our portfolio include an equal number of long and short positions, and the amount invested in each stock is constant and equal to |5000$|, the alpha quantile of a long position is multiplied by 5000$ while the alpha quantile of a short position is multiplied by -5000$. Assuming that we have a long position in the first 20 stock in our portfolio and a short position in the last 20:

$$VaR(i) = r_{i,\alpha} \times (+5000) \; \forall i = 1 \dots 20$$
$$VaR(j) = r_{j,\alpha} \times (-5000) \; \forall j = 21 \dots 40$$

Where $VaR(i)$ is the Value at Risk of a generic long position and $VaR(j)$ is the Value at Risk of a generic short position.

4) We aggregated each single position VaR into a row vector $\boldsymbol{w} =: [VaR(1), VaR(2), \dots, VaR(40)]$ and, according to Variance-Covariance approach, computed Portfolio Value at Risk as:

$$Portfolio\ VaR = \sqrt{w} \times CORR \times \sqrt{w'}$$

Where $CORR$ is the estimated correlation matrix of the 40 assets in the portfolio.

5) Every step of the above procedure has been iterated through recursive techniques (such as loops) for every rolling window of our analysis, and for $\alpha = 5\%$ and $\alpha = 1\%$.

Warning

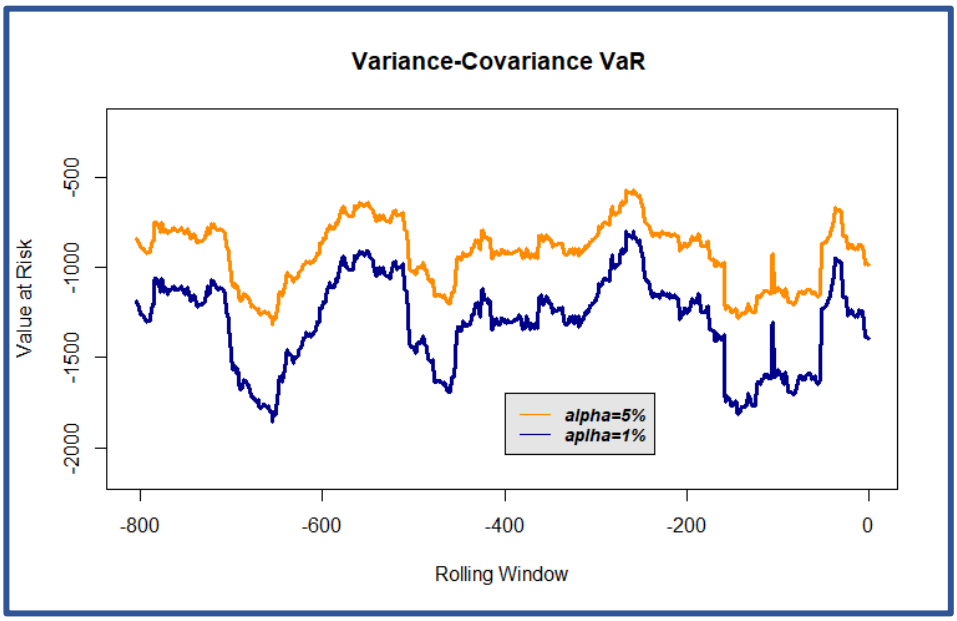
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

Therefore, the outputs of this procedure are two column vectors which contain Portfolio VaR for every rolling window considered:

$$\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha} = \mathbf{5}\%) =: [PortfolioVar_0(5\%), \dots, PortfolioVar_{-805}(5\%)]'$$
$$\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha} = \mathbf{1}\%) =: [PortfolioVar_0(1\%), \dots, PortfolioVar_{-805}(1\%)]'$$

where the first element of each vector is the most recent estimated of Portfolio Value at Risk because it has been computed using the observations of the first rolling window (rolling window 0), and the last element of this vector is the oldest estimated Portfolio Value at Risk because it has been computed using the observations of the last rolling window (rolling window -805).

Plotting these vectors will give a graphical idea of the time evolution of the estimated Portfolio VaR.



For every rolling window, *i.e* for every set of 52 consecutive observations, we have an estimate of Portfolio VaR. However, it is necessary find the number that better approximate the overall daily risk exposure of our portfolio in the next month. One possibility would be picking the first element of the vector $\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha}\%)$, which is the VaR estimated using the most recent rolling window. However, since the above graph shows that VaR has been volatile throughout the rolling windows considered, we believe that an exponential smoothing average of the elements of the vector $\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha}\%)$ is the best approximation of the overall daily risk exposure of our portfolio in the next month. With a discount factor $\lambda = 0.94$, the final estimates of Portfolio VaR, for $\alpha = 5\%$ and $\alpha = 1\%$, are:

Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

$$Final\ VaR(\alpha = 5\%) = \frac{\sum_{i=-805}^{0} \lambda^{-i} \times PortfolioVar_i(5\%)}{\sum_{i=-805}^{0} \lambda^{-i}} = 901€$$

$$Final\ VaR(\alpha = 1\%) = \frac{\sum_{i=-805}^{0} \lambda^{-i} \times PortfolioVar_i(1\%)}{\sum_{i=-805}^{0} \lambda^{-i}} = 1274€$$

- *Evidences Against Normality*

It is necessary to perform normality tests on the distributions of log returns of the stocks in our portfolio because both Parametric EWMA and Variance-Covariance VaR approaches requires multivariate normality of log returns distribution. This means that the joint probability density function of the log returns of the 40 stocks in our portfolio must follow a multivariate normal distribution.
A set of random variables (log returns in our case) follows a multivariate normal distribution if and only if every random variable is marginally gaussian and any linear combination of them is normally distributed as well.
After having performed Mardia's Multivariate Normality test on the set of log returns, we cannot assess multivariate normality to our model because we reject the null hypothesis. Moreover, Shapiro-Wilk normality test confirms that the random variable log return of each stock is not normally distributed because we reject H0 in all the 40 cases. The output of these tests is reported below:

$multivariateNormality

	Test	Statistic	p value	Normality
1	Mardia Skewness	69239.3404207161	0	NO
2	Mardia Kurtosis	275.543194060744	0	NO
3	MVN	<NA>	<NA>	NO

$univariateNormality

	Test	Variable	Statistic	p value	Normality
1	Shapiro-Wilk	Column1	0.9485	<0.001	NO
2	Shapiro-Wilk	Column2	0.9267	<0.001	NO
3	Shapiro-Wilk	Column3	0.9745	<0.001	NO
4	Shapiro-Wilk	Column4	0.9208	<0.001	NO
5	Shapiro-Wilk	Column5	0.9492	<0.001	NO
6	Shapiro-Wilk	Column6	0.8491	<0.001	NO
7	Shapiro-Wilk	Column7	0.8426	<0.001	NO
8	Shapiro-Wilk	Column8	0.9759	<0.001	NO
9	Shapiro-Wilk	Column9	0.9820	<0.001	NO

Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

10	Shapiro-Wilk	Column10	0.9797	<0.001	NO
11	Shapiro-Wilk	Column11	0.9634	<0.001	NO
12	Shapiro-Wilk	Column12	0.7582	<0.001	NO
13	Shapiro-Wilk	Column13	0.9465	<0.001	NO
14	Shapiro-Wilk	Column14	0.8312	<0.001	NO
15	Shapiro-Wilk	Column15	0.9706	<0.001	NO
16	Shapiro-Wilk	Column16	0.9790	<0.001	NO
17	Shapiro-Wilk	Column17	0.9779	<0.001	NO
18	Shapiro-Wilk	Column18	0.9416	<0.001	NO
19	Shapiro-Wilk	Column19	0.9593	<0.001	NO
20	Shapiro-Wilk	Column20	0.9541	<0.001	NO
21	Shapiro-Wilk	Column21	0.9624	<0.001	NO
22	Shapiro-Wilk	Column22	0.9649	<0.001	NO
23	Shapiro-Wilk	Column23	0.9806	<0.001	NO
24	Shapiro-Wilk	Column24	0.9257	<0.001	NO
25	Shapiro-Wilk	Column25	0.8054	<0.001	NO
26	Shapiro-Wilk	Column26	0.9179	<0.001	NO
27	Shapiro-Wilk	Column27	0.9183	<0.001	NO
28	Shapiro-Wilk	Column28	0.8703	<0.001	NO
29	Shapiro-Wilk	Column29	0.8805	<0.001	NO
30	Shapiro-Wilk	Column30	0.9519	<0.001	NO
31	Shapiro-Wilk	Column31	0.9594	<0.001	NO
32	Shapiro-Wilk	Column32	0.9048	<0.001	NO
33	Shapiro-Wilk	Column33	0.9670	<0.001	NO
34	Shapiro-Wilk	Column34	0.8955	<0.001	NO
35	Shapiro-Wilk	Column35	0.8652	<0.001	NO
36	Shapiro-Wilk	Column36	0.5216	<0.001	NO
37	Shapiro-Wilk	Column37	0.9560	<0.001	NO
38	Shapiro-Wilk	Column38	0.8897	<0.001	NO
39	Shapiro-Wilk	Column39	0.9666	<0.001	NO
40	Shapiro-Wilk	Column40	0.9394	<0.001	NO

Warning

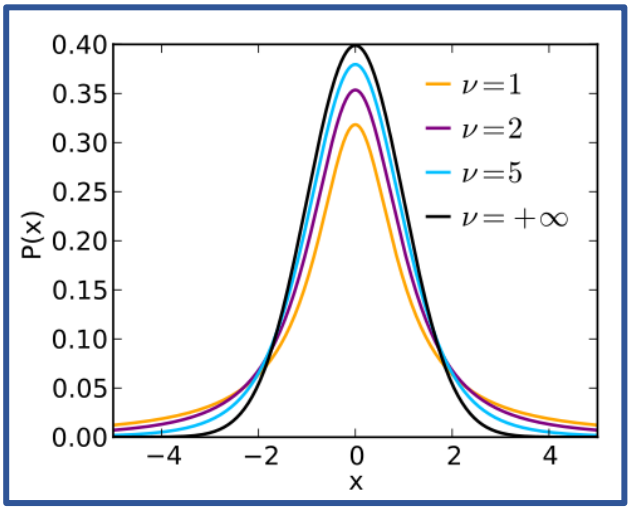
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

V. Montecarlo Simulations

- *Rolling Montecarlo – T-Student*

Because of the above-mentioned normality tests, we abandon the Gaussian hypothesis and we switch to the T-Student Case. In this way, we can consider also higher unusual moments of the distribution and account for heavy tails. As a further proof of non-normality of returns, trying to fit the data with a multivariate t-student, we obtain df = 6, recalling that for a Gaussian distribution df should tend to infinity.

To this end, we now move to the simulation approach, since the parametric Var-Cov methodology only suits the normality assumption.

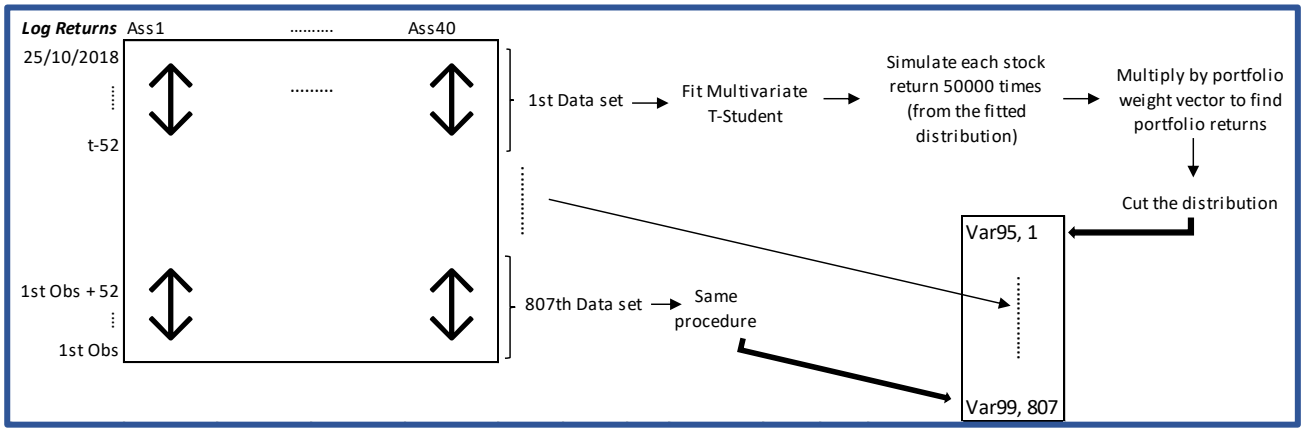


We are asking R to:

- fit rolling data with 806 Multivariate T-Student distributions, containing 40 stocks and 52 data points each
- simulate values for each stock and for each roll (for each Multivariate distribution) 50000 times
- assemble simulating values into portfolio values using portfolio weights and/or €amounts
- cut the distribution at the desired percentile.

The procedure to obtain these results is much more straight-forward from a coding perspective compared to the Var-Cov approach, since we need one single loop. On the other hand, however, it is more time consuming because of the complexity of the problem, and it takes approximately 15 minutes for R to reach the solution.

How the loop works:



Warning

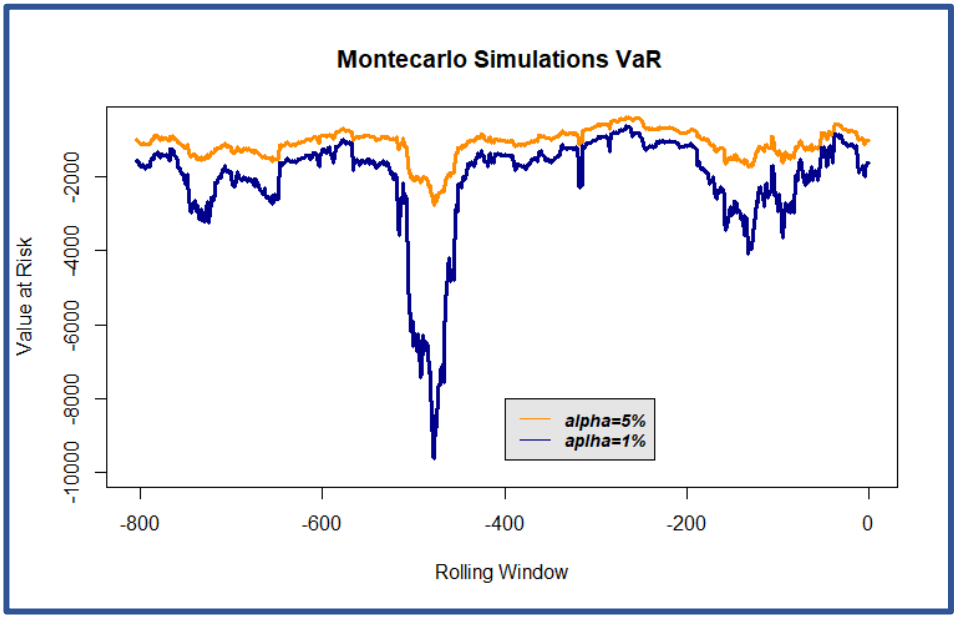
This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

Therefore, the outputs of this loop are two column vectors which contain Portfolio VaR for every rolling window considered:

$$\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha = 5\%}) =: [PortfolioVar_0(5\%), \dots, PortfolioVar_{-805}(5\%)]'$$
$$\boldsymbol{PortfolioVaR}(\boldsymbol{\alpha = 1\%}) =: [PortfolioVar_0(1\%), \dots, PortfolioVar_{-805}(1\%)]'$$

where the first element of each vector is the most recent estimated of Portfolio Value at Risk because it has been computed using the observations of the first rolling window (rolling window 0), and the last element of this vector is the oldest estimated of Portfolio Value at Risk because it has been computed using the observations of the last rolling window(rolling window -805).

Plotting these vectors will give a graphical idea of the time evolution of the estimated Portfolio VaR.



With a discount factor $\lambda = 0.94$, the final estimates of Portfolio VaR, for $\alpha = 5\%$ α and $\alpha = 1\%$, are:

$$Final\ VaR(\alpha = 5\%) = \frac{\sum_{i=-805}^{0} \lambda^{-i} \times PortfolioVar_i(5\%)}{\sum_{i=-805}^{0} \lambda^{-i}} = 918€$$
$$Final\ VaR(\alpha = 1\%) = \frac{\sum_{i=-805}^{0} \lambda^{-i} \times PortfolioVar_i(1\%)}{\sum_{i=-805}^{0} \lambda^{-i}} = 1520€$$

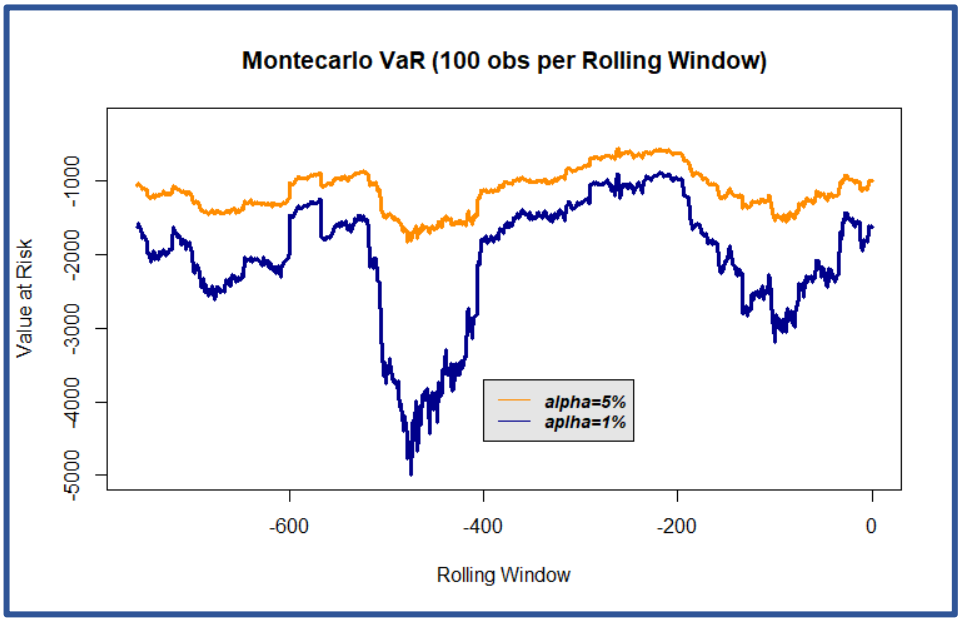
Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

However, the graph above shows a peak around the 500th rolling window VaR captured by the Montecarlo Simulation for the 99% Confidence Level. This extreme behavior, that is not reflected in the other methodologies results, may be due to 3 specific issues/reasons:

1) Using the T-Student distribution, we are correctly accounting for Heavy Tails, assigning higher probabilities to more negative returns, compared to the normality case.
2) Since we are using 52 observations per roll, the fitting function may have encountered issues in assigning reliable parameters to each single roll.
3) Since we are using 52 observations per roll and since, differently from the historical case, Montecarlo simulations use data as parameters to estimate a distribution, it might happen that, because of very positive correlations and very negative returns recorded during a roll, the fitting function is forced to estimate a truly extreme distribution for a specific set of data.

To investigate this issue and obtain a more reliable estimate of this kind of Var, we repeated the procedure using rolls composed by 100 observation, rather than 52. In the following plot we notice that the peak has been shrunk presenting fewer extreme values.



Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

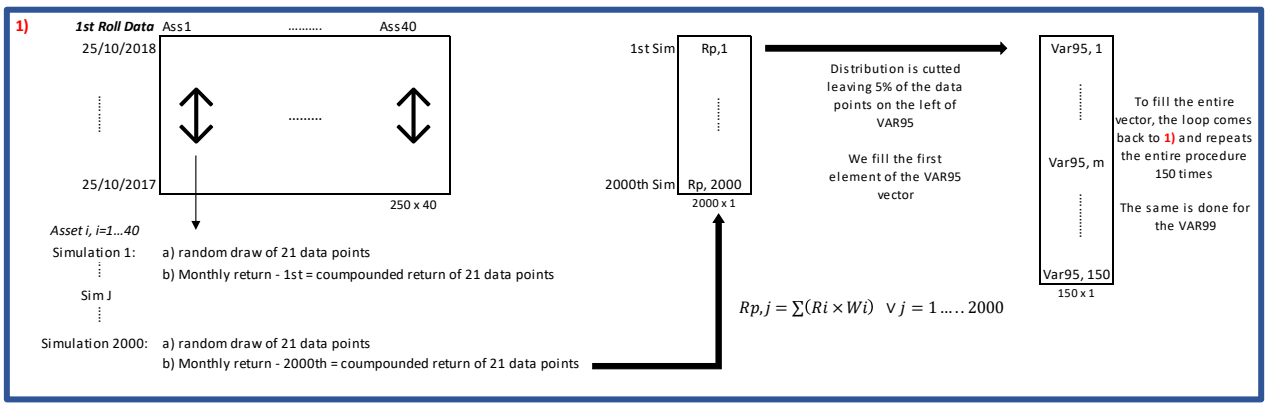
- *Bootstrapping*

To give an idea about what a Monthly VaR would look like for this active portfolio, we implemented a Bootstrapping procedure. Bootstrapping stands in the middle between Montecarlo Simulations and Historical Simulation, it combines a random element (the return is drawn randomly) and an historical element (return is drawn from an historical time series).

How it works:

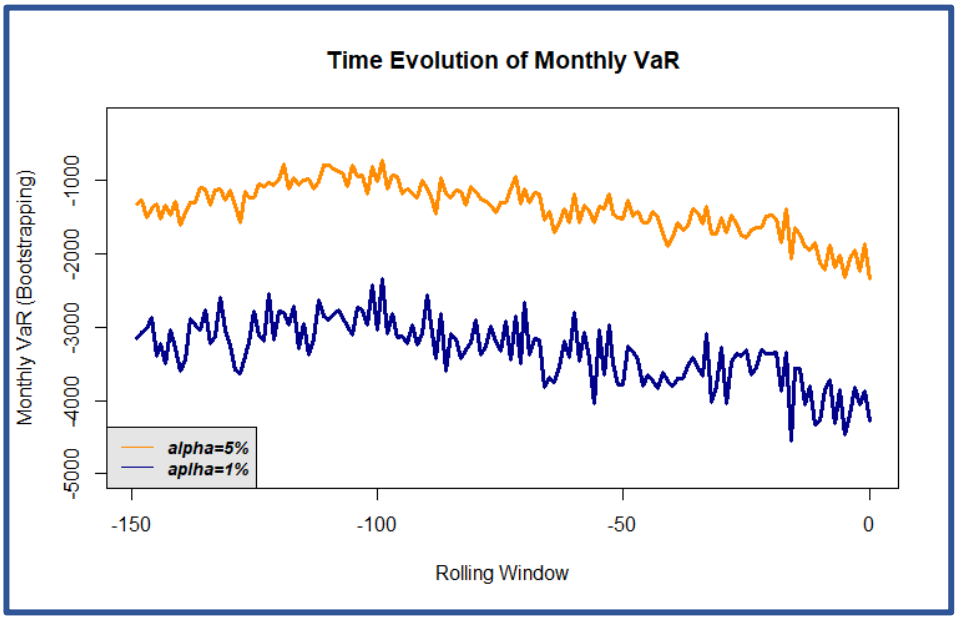
It is a sampling procedure with replacement from original data. In order to implement it we coded four loops able to provide a rolling VaR measure as for the other methodologies. For each roll (150 rolls), of 250 data points each (1 year) and for each asset, we extracted 21 random data points from the returns dataset and compounded those 21 daily logs returns to compute a monthly return. The random draw is performed 2000 times so that for each roll we have 2000 monthly returns available for each asset, that are then multiplied by either the portfolio weights (+5%,-5%) and/or the euro amounts invested in each asset (5000,-5000). In this way, we are overcoming the main limitation of historical VaR, in fact we are simulating a time series that is consistent with the original one but at the same time we do not have constraints in terms of number of observations (length of the sample). Without Bootstrapping, indeed, we should have used 1 year of monthly returns, i.e. 12 data points for each time series of returns, a too short sample for VaR calculation.

(total # sim = 40 x 2000 x 150 = 12,000,000)



Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.



Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

VI. Educational Series#1

- *Theoretical foundations of Shapiro-Wilk Test for Normality*

Given a sample $(X_1, \dots, X_n)$ of observed realizations of the random variable X, Shapiro-Wilk test (Shapiro and Wilk, 1965) challenges the null hypothesis that $X \sim i.i.d.Normal(\mu; \sigma^2)$.
Consider the order statistics $[X_{(1)}, X_{(2)}, \dots, X_{(i)}, \dots X_{(n)}]$ which takes the vector $(X_1, \dots, X_n)$ and rearranges all its element with the increasing criterion $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(i)} \leq \dots \leq X_{(n)}$.
Then consider the standard normal order statistics $[Z_{(1)}, Z_{(2)}, \dots, Z_{(i)}, \dots Z_{(n)}]$ which takes the vector $(Z_1, \dots, Z_n)$, such that $Z_1, \dots, Z_n \sim i.i.d.Normal(0; 1)$, and rearranges all its element with the increasing criterion $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(i)} \leq \dots \leq Z_{(n)}$.
Let m be the vector of the expected values of the standard normal order statistics. In other words, $m := [m_1, \dots, m_n] = [\mathbb{E}(Z_{(1)}), \dots, \mathbb{E}(Z_{(n)})]$.
Let V be the $n\ x\ n$ covariance matrix of the standard normal order statistics.
The Shapiro-Wilk test statistics is defined as:

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{[\sum_{i=1}^{n}(X_i - \bar{X}]^2]}.$$

Where:

- $X_{(i)}$ represents the i^{th} lower number of the vector $(X_1, \dots, X_n)$.
- X_i is a generic observed realization of the variable X.
- $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$. $\bar{X}$ is the sample mean.
- a_i is a constant associated with $X_{(i)}$ such that $a_i \in [a_1, \dots, a_n] := \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{1/2}}$. The vector $[a_1, \dots, a_n]$ basically summarizes the information about the standard normal order statistics $[Z_{(1)}, Z_{(2)}, \dots, Z_{(i)}, \dots Z_{(n)}]$ of the vector of variables $Z_1, \dots, Z_n$ which are, by assumption, independent and normally distributed with mean 0 and variance 1.

It can be proved that the support of W ranges from 0 to 1. Under the null hypothesis, the Shapiro-Wilk test statistics follows a Beta distribution of parameters $\gamma = \frac{1}{2}, \delta = \frac{1}{2}$:

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{[\sum_{i=1}^{n}(X_i - \bar{X}]^2]} \sim \beta_{\frac{1}{2},\frac{1}{2}}.$$

The intuition behind Shapiro-Wilk test for normality is basically trying to compare a sample belonging to an unknown distribution with a sample belonging to a standard normal distribution. The W statistics incorporates the information of the standard normal distribution in the constant a_i and can be seen as a proxy of the correlation between the two samples. As many correlation indexes, $support(W) = [0; 1]$.

Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

Therefore, for values of W close to 1 we cannot reject H0 because there is a significant correlation between the two samples, which means that the sample whose distribution was unknown comes actually from a normally distributed population.
The numerator of W can be considered as a sort of covariance between the two samples to be compared. Indeed, the higher covariance, the most likely the two samples come from the same distribution, the most likely we cannot reject H0.
The main critique addressed to Shapiro-Wilk test for normality is that its accuracy decreases as the sample size increases. Indeed, as many other tests for normality, Shapiro-Wilk test systematically rejects H0 when the sample size n is sufficiently large, even if the data follows an approximately normal distribution.
Therefore, for large values of n, the conclusion of Shapiro-Wilk test on a set of data shall be compared with the respective Q-Q plot in order to assess the validity of the test.

Warning

This is an academic paper related to an academic project. This paper is not an investment suggestion and it does not in any way represent an invitation to purchase the securities we virtually invest in.

MINERVA
INVESTMENT MANAGEMENT SOCIETY