RISK MANAGEMENT MAGAZINE

Vol. 16, Issue 1
January – April 2021

EXCERPT

https://www.aifirm.it/newsletter/progetto-editoriale/

Design of an algorithm for an adaptive Value at Risk measurement through the implementation of robust methods in relation to asset cross-correlation

Marco Bagnato, Anna Bottasso and Pier Giuseppe Giribone

Design of an algorithm for an adaptive Value at Risk measurement through the implementation of robust methods in relation to asset cross-correlation

Marco Bagnato (SoftJam), Anna Bottasso (University of Genoa) and Pier Giuseppe Giribone (Banca CARIGE – University of Genoa)

Article submitted to double-blind peer review, received on 5th March 2021 and accepted on 8th April 2021

Abstract

This study proposes an algorithmic approach for selecting among different Value at Risk (VaR) estimation methods. The proposed metaheuristic, denominated as "Commitment Machine" (CM), has a strong focus on assets cross-correlation and allows to measure adaptively the VaR, dynamically evaluating which is the most performing method through the minimization of a loss function. The CM algorithm compares five different VaR estimation techniques: the traditional historical simulation method, the filtered historical simulation (FHS) method, the Monte Carlo method with correlated assets, the Monte Carlo method with correlated assets which uses a GARCH model to simulate asset volatility and a Bayesian Vector autoregressive model. The heterogeneity of the compared methodologies and the proposed dynamic selection criteria allow us to be confident in the goodness of the estimated risk measure. The CM approach is able to consider the correlations between portfolio assets and the non-stationarity of the analysed time-series in the different models. The paper describes the techniques adopted by the CM, the logic behind model selection and it provides a market application case of the proposed metaheuristic, by simulating an equally weighted multi-asset portfolio.

Key Words:

Value at Risk (VaR), Historical VaR, Filtered Historical Simulation (FHS), Monte Carlo VaR, GARCH volatility model, Bayesian Vector Autoregressive (BVAR), Commitment Machine (CM)

DOI 10.47473/2020rmm0083

1) Introduction

The adoption of VaR measures in financial markets dates back to the 90s; since then, the available models for VaR estimation have increased both in quantity and in complexity. However, there is no preferred VaR method that performs "better" in absolute terms; each methodology considers a particular aspect of the return distribution and its performance depends on the context and on the specifications used to implement each model.

An important risk factor for any investment portfolio is the cross correlation of assets, which tends to increase during periods of financial crisis [1]. Another important risk factor is the time-varying behavior of conditional volatility (see section 3). Therefore, the objective of this work is to create a Risk Management system based on models that take into account the cross-correlation between assets and the trend of variance over time, and to propose a possible solution to the problem of selecting among several VaR models.

In particular, the proposed solution consists in a flexible and adaptive selection criterion that allows to choose, day by day, the estimation of the VaR that reasonably represents financial market conditions. We adopted five VaR models, whose structural principles are described in section 3: traditional historical VaR, filtered historical simulations (FHS) VaR, Monte Carlo VaR with cross correlated assets, Monte Carlo GARCH VaR with cross correlated assets, Bayesian VaR. These different models are evaluated considering different aspects of the time series; three out of these five methods examine the correlation among portfolio assets, while the other two, based on the "historical" approach (historical VaR and filtered historical VaR), are less sophisticated but quite flexible and easy to implement.

Starting from the cited methods, a validation phase is carried out, based on the frequency of VaR threshold violation, with an approach based on backtesting and statistical tests to verify the hypothesis of unconditional coverage (binomial test and traffic light approach). Subsequently, a CM is proposed.

The CM evaluates the five approaches day by day by calculating every day, for each of them, the value assumed by a given loss function in the previous days. The model with the best performance is then chosen to estimate the VaR of the following day. The use of loss functions for evaluating a VaR model has many examples in related literature. A particular attention has been directed towards the use of a loss function that weights the negative differences between VaR and returns according to a quadratic function [17]. Section 4.1 presents a theoretical discussion of the properties of the loss function adopted in this study. Two different variants have been tested based on the number of observations used to calibrate the CM. For each of these two variants, the overall performance and selection capacity of the CM have been assessed, by analyzing the losses with respect to the VaR threshold and the frequency of VaR violations. The results clearly indicate that the CM is able to make an efficient selection among the various methods, by choosing VaR thresholds that are less likely to be violated and register the smallest losses in terms of negative differences between VaR and actual returns.

The proposed risk management approach is extremely customizable. Indeed, thanks to the flexibility of the code written in MATLAB, the CM approach can be used to select among a great variety of VaR methods.

2) Dataset

For the purposes of the analyses introduced in section 1, we built an equally weighted portfolio by using 4 historical time series retrieved from the info-provider Bloomberg®. These series track four different indices representing three of the main asset classes available to investors (equities, gold, bonds). The components of these indices are representative of the investment choices of most financial intermediaries and allow us to represent a balanced portfolio:

- European Stock Index (SXXP Index): the Stoxx 600 index tracks the trend of large, mid and small cap stocks in 17 different European countries. With its 600 components, it allows to simulate a highly diversified equity portfolio across the UK, Switzerland and the Eurozone.
- US Stock Index (RAY Index): The Ray Index includes 3000 listed companies which represent (in terms of market cap) 98% of the universe of US listed shares, allowing US stock markets to be incorporated into the portfolio.
- World Bond Index (Legatruh Index): the Bloomberg Barclays Global Aggregate Index collects investment grade debt listed on 24 markets, in both developed and emerging economies. The inclusion of this index allows to increase diversification by adding a second asset class distinct from the stock market and diversifying the geographic risk.
- Gold (XAU USD currency): this series tracks the historical exchange rate between gold and the US dollar. Gold has traditionally been considered a safe-haven asset and its inclusion can offer significant diversification potential.

Since the European stock index is denominated in euro, we have retrieved the historical Euro/Dollar exchange rates for the analysed period and used it to convert all the data into dollars. In order to achieve a reasonable sample size, we decided to analyze the data of the daily closing prices for the period from 1st June 2000 to 30st September 2020. This time span contains a total of 5305 market observations for each asset considered. The analysis requires the choice of a time window that allows to dynamically evaluate the evolution of the risk measures and of all other relevant variables (particularly the evolution of cross correlations in order to have a correct measurement of portfolio risk). This observation window ('rolling windows') must be large enough to be statistically significant, but at the same time it should not be too wide in order to concretely capture the effects of relatively short-term shocks (for example the collapse and subsequent recovery of the markets due to the Covid-19 pandemic in the spring of 2020).

In order to balance the above-mentioned trade-off, in accordance with the practice used for this type of analysis, we decided to use a rolling window of 260 observations, which is equal to one year and it is considered large enough for an overall analysis of market risk. Scientific literature [8] shows a remarkable consistence of VaR estimations calculated on data from a 250-day time window. Such consistence is measured by root mean squared relative bias, that is a measure of the relative distance between different risk measures calculated for a given interval. (See tab A2 of [8] for a comparison with other observation windows).

Unless otherwise specified, in the context of this study, all the measurements obtained will always refer to the 260-day rolling window. This means that, for example, the VaR calculated for the 270th daily observation will be calculated on the data ranging from the 10th to the 269th day. At each iteration, the calculation moves forward by a step equal to one day, so the VaR of the 271st day will be calculated on the data ranging from the 11th day to the 270th day, and so on; the same logic is maintained when we consider all the inputs that contribute to VaR estimation (mean, variance, correlation, etc.) that are always calculated on the same rolling window.

Figure 2.1 Historical trend of portfolio components

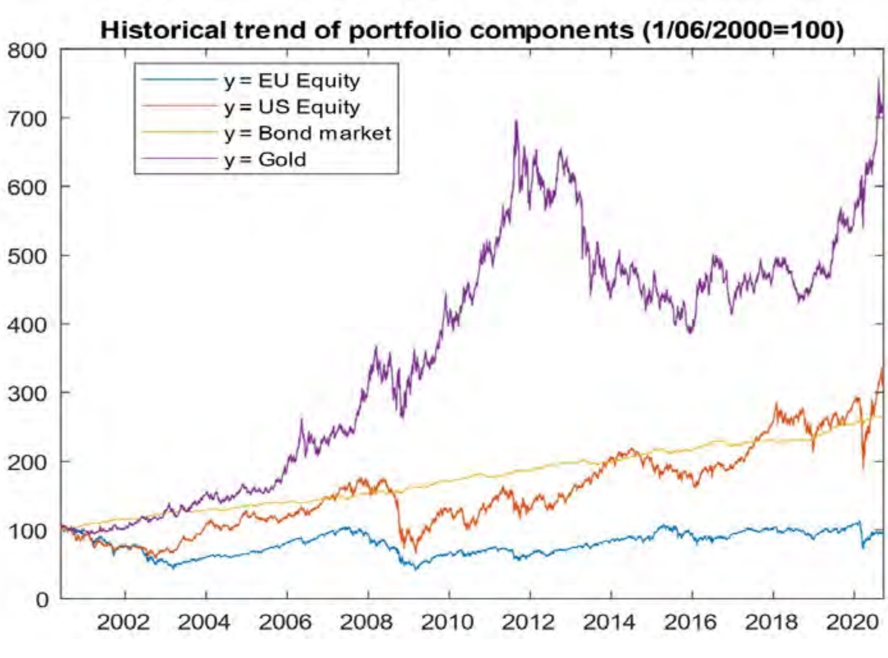


Figure 2.1 shows the trend of the 4 indices over time, after normalizing their price to 100 on the first observation day (1st June 2000).

In order to conduct a correct risk analysis, it is necessary to consider the characteristics of daily returns distribution both for the four indices considered and for the portfolio. Table 2.1 shows the main descriptive statistics (minimum, maximum, median and the first four moments) for each historical time series. The various data are calculated on the entire history available without the use of rolling windows.

Table 2.1 Analysis of portfolio and components time series

	Distributional and descriptive statistics of daily returns						
	Min	Max	Median	Average	St. Dev.	Kurtosis	Skewness
US equity	-11%	10%	0,026%	0,006%	1,21%	10,49	-0,22
EU equity	-12%	13%	0,061%	0,032%	1,44%	12,06	-0,12
Bond market	-2%	2%	0,024%	0,020%	0,16%	11,60	-0,41
Gold	-9%	11%	0,045%	0,042%	1,07%	9,19	-0,19
Portfolio	-7%	6%	0,04%	0,025%	0,65%	12,77	-0,27

All series show a high level of kurtosis and this seems to suggest a non-normal distribution for our data. In order to check for data distribution, we performed a Kolmogorov Smirnov test at 5% significance level that suggested us to reject the normality hypothesis for all the four time series (p value ~ 0 for all the series).

This allows us to reject the hypothesis of normality of our dataset and prevents the adoption of the variance-covariance method in the VaR calculation. Consequently, this approximated method of estimation has not been adopted in this study. The second consideration is about the most volatile among the indices i.e., the two equity indices. The matrix of daily correlations across the indices, reported in Table 2.2, shows that the greatest correlation is observed between the US stock index and the European stock index.

Table 2.2 Cross correlation matrix of portfolio components

Cross correlation matrix of portfolio components				
	EU Equity	US Equity	Bond market	Gold
EU Equity	1,00	0,51	-0,29	-0,01
US Equity	0,51	1,00	-0,18	0,002
Bond market	-0,29	-0,18	1,00	0,18
Gold	-0,01	0,002	0,18	1,00

The correlation between portfolio assets can be considered as a possible risk factor. The greater riskiness of the two equity indices has been confirmed by the Euler decomposition of portfolio risk, that suggests that more than 75% of the portfolio volatility is generated by the two most volatile equity indices. More specifically, the Euler decomposition attributes portfolio risk among the portfolio components as follows: 32,3% to the European stock index, 43% to the American stock index, 1,2% to the bond market index, 23,5% to the gold.

This calculation highlights how the correlation structure is by itself a risk factor, as it has the potential to amplify the losses due to the most volatile indices in the portfolio.

Another important feature of the cross-correlations between assets is the opportunity to use them to build less procyclical VaR models. Over the sample period, the two major negative events (the 2008 crisis and the Covid shock) came after a long period of positive equity market returns; in both cases, the value of the portfolio reached an all-time high just before the crisis. This is an obvious issue for risk management: the most common models of VaR are strongly backward looking, with obvious negative effects when suddenly indices shift from growth to collapse. However, cross-correlations between assets can help to solve this problem: in both cases mentioned above, the correlation between the two equity indices started to rise before the onset of the crisis. In the case of the 2008 crisis, in the previous two years both the variance of each of the two indices and their covariance increased, while before the 2020 crisis the two variances were stable.

3) VaR methodologies

After choosing the dataset we have built a risk management system that can serve as a basis for the analyses. In particular, we focused on the calculation of VaR and Expected Shortfall (ES). Although the analysis of this work focuses on the VaR threshold, the CM can also estimate an ES value from the corresponding VaR threshold for every time step. This makes the approach more versatile and offers - in perspective - greater possibilities for evaluating its performance and a better risk analysis with a coherent and subadditive risk measure (the ES). In fact, the CM selects a method for simulating price distribution that can be used both for VaR and for ES calculation.

As already mentioned, the value of the VaR threshold has been calculated on a daily basis, thus returning a total of 5043 VaR measurements for each considered method. All the methodologies have been estimated on a 95% confidence interval to calculate the losses occurring in the market day following the calculation date.

For the purpose of calculating these two quantities, 5 different methods have been implemented in order to take into account both sudden changes in variance (or, in other words, to distinguish between conditional and unconditional variance) and the effects of cross-correlations. These 5 methods can be classified into 3 families of models:

- Base Methodology: Historical VaR. This is a very simple method that is affected by rather simplifying assumptions on the future distribution of returns. However, given its computational simplicity and its sensitivity to negative tails of market returns, we decided to include it in the system and employ it together with other more sophisticated methods. A more advanced version has also been adopted (Historical VaR with filtered simulations, or FHS) which takes into account the short-term conditional volatility for each calculation date.
- Stochastic Differential Equation (SDE) Integration: Monte Carlo methods. The Monte Carlo method is a numerical simulation technique based on the integration of stochastic differential equations, which in this case is applied to the performance of the returns of the selected assets. This is an extremely flexible simulation method; in this work it has been implemented by simulating the effects of the correlation between assets on the dynamics of process innovations in two different versions (with and without GARCH volatility to consider the heteroskedasticity of the return time series).
- Econometric Bayesian methods: Bayesian Vector Autoregressive. This methodology is more complex than the Monte Carlo method, both theoretically and in terms of implementation. In particular, the Bayesian method allows us to simulate the joint trend of correlations and volatility in an extremely flexible way, including them in the stochastic component of the model.

We now proceed to expose the theory, the operational principles and the implementation of the methodologies discussed.

3.1) Basic Methodology: historical VaR

The first family of models included in the risk management system is the historical method. This technique is the simplest of the five models adopted in this paper: it models the VaR threshold for a certain day as the quantile of the returns of the n immediately preceding market days (in this case n = 260). This is a backward-looking method which assumes that the distributional characteristics of past returns are a good proxy for analysing future returns. It relies on the assumption that the past situation of the market reflects the future one. Such an assumption is impossible in the case of portfolios of new financial products for which we have no previous past prices realization [3]; in our case, the main issues are related to the non-stationarity of the historical time series considered. However, considering its popularity and its simplicity, we decided to combine this approach with more complex techniques whose calibration could however create potential model risks.

In order to further diversify the approach adopted, a historical Var method based on "filtered historical simulations" (FHS) has also been employed. The FHS method “filters” the various returns, rescaling them using short-term volatility [13]. In this paper, the short-term volatility is calculated taking into account 100 market days, while the long-term volatility is based on a sample of 260 market days, in line with the same time windows used in the other methods.

In this way, the distribution of returns also incorporates information on the most recent volatility. In more formal terms, by assuming to have a distribution of unfiltered returns x(t), it is possible to derive the distribution of filtered returns X(t) as:

$$X(t) = x(t) * \frac{\sigma(t)_{SHORT\ TERM}}{\sigma(t)_{LONG\ TERM}} \quad (1)$$

Thereby, if the ratio of conditional short-term volatility to long term volatility increases, the corresponding return increases in absolute value, and if it is negative it assumes more weight in the calculation of the VAR; the opposite holds if volatility decreases. This allows to consider the effects of changes in volatility on VaR: in periods of financial crisis when the market returns decline and volatility increases (and therefore $\sigma_{SHORT\ TERM} > \sigma_{LONG\ TERM}$), negative returns are rescaled and increase in absolute value, making VaR thresholds theoretically more robust.

3.2) Monte Carlo method with and without GARCH volatility

The second class of models that has been implemented is the Monte Carlo method, of which two possible variants are proposed. The Monte Carlo method in this context is interpreted as a numerical method that allows to simulate the possible trajectories of one or more assets that follow a Brownian geometric motion. A Brownian geometric motion is meant as a stochastic process defined by the SDE:

$$dS_t = \mu S_t dt + \sigma S_t \mathrm{d}W_t \quad (2)$$

Where μ is the mean of asset returns, σ>0 is standard deviation, S(t) is the price of the asset at time t and W(t) is a Wiener process, that is a stochastic process defined by independent increments over time with mean equal to 0 and variance equal to the time interval considered: W(T)-W(0) is normally distributed with mean 0 and variance T.

It can be shown that equation (2) can be rewritten in its final form as:

$$S_T = S_0 \exp\left[\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\varepsilon_T\sqrt{T}\right] \quad (3)$$

Where ε_T is distributed as a standardized normal with zero mean and unit variance. By means of equation (3), using the mean and variance of returns, the price at the beginning of the period and the time interval of the simulation as input parameters, it is possible to simulate multiple paths of asset price evolution.

Each individual path produces a simulated value S_T which represents a possible final price value. The final price S_T of the single simulation is used to estimate the various possible returns; these returns are combined into a single sample from which the 5% quantile (VaR) and the Expected Shortfall are calculated.

However, the setting of this work also requires the evaluation of the effects of correlations between assets. Consequently, two different variants of the Monte Carlo method have been implemented which generalize the "basic" method described, including the correlations in the calculation of the error term. In order to incorporate the correlation matrix for the four assets in the Monte Carlo simulation, the Cholesky decomposition can be used.

Assuming you have a set of unrelated random numbers $\vec{\varepsilon} = \varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots \varepsilon_T$, Cholesky decomposition can be used to transform them into a set of correlated variables $\vec{a} = a_1, a_2, a_3, \ldots a_T$. If $\vec{a}$ and $\vec{\varepsilon}$ are column vectors with N rows, and **R** is the correlation matrix, it is possible to apply the following transformation:

$$\vec{a} = \boldsymbol{M}\vec{\varepsilon} \quad (4)$$

Where $\boldsymbol{M}$ is a matrix that must satisfy the condition $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{R}$. The matrix $\boldsymbol{M}$ can be obtained by applying the Cholesky decomposition to $\boldsymbol{R}$. Subsequently, the correlated shocks ($\vec{a}$) are substituted to the errors (ε) in equation (3). From this point, the various possible paths of the assets are simulated, thus obtaining a set of possible values of the returns from which to calculate VaR and ES with the quantile method.

With regards to the implementation of the model in the MATLAB environment, the Cholesky function has been used to transform the correlation matrix $\boldsymbol{R}$ into an upper triangular matrix $\boldsymbol{M}$ that would guarantee the respect of the condition $\boldsymbol{M}\boldsymbol{M}^T = \boldsymbol{R}$. Subsequently, the Hadamard product has been used to multiply, for each simulation, the innovations and the $\boldsymbol{M}$ matrix.

The number of simulations adopted for each VaR and ES value has been set equal to 50000. Once the value of the single innovation has been obtained, the simulated value of the returns of each asset in T + 1 has been calculated for each simulation, applying equation (3).

In each simulation, the mean of the prospective returns of the individual assets has been calculated in order to jointly obtain the return values of the assets and the portfolio. From these replicated simulations, VaR and ES have been calculated.

In order to implement the stochastic differential equations for the Monte Carlo engine, we did not use 'built in' functions already coded in the MATLAB toolboxes. Such a choice guarantees maximum flexibility in the design phase and allows to customize, for example, the dynamics that describes the stochastic process.

For this reason, the code has been validated through the pricing of three different options: a call option written on one asset, a European arithmetic average spread option written on two assets and an exotic option written on three assets (option on Maximum of two spread options) [7]. In all cases the results have been compared with a different numerical method, obtaining aligned values up to the basis points.

However, the presented Monte Carlo method could have some flaws, as it does not consider the non-stationarity of the time series. In particular, sudden increases in variance are observed in periods of crisis, a clear sign of the possible presence of heteroskedasticity.

In order to test the presence of this phenomenon, an Engle test with a confidence interval of 95% has been conducted. For each analyzed return series, the test led to the rejection of the null hypothesis of heteroskedasticity. Table 3.1 reports the P value of the two tests performed in this section.

Another very common feature in historical financial time-series is the autocorrelation of returns. With regards to this aspect, the autocorrelation of the residuals has been tested using a Ljung-Box test at 95% confidence interval applied to the first lag of the data.

The results of the test lead to reject the null hypothesis of independence of the residuals for the US equity index and the European equity index, the two more volatile components of our portfolio as outlined in section 2 (Tab 2.1). P values of this test are shown in Table 3.1.

Table 3.1 P-values for the statistical tests

P value of Engle Test and Ljung Box test for the analyzed time series					
		US equity Index	European Equity index	Bond index	Gold index
P-values of statistical test	Engle Test	0.001	0.002	0.008	0.007
	L-B, 1st lag	0.026	0.015	0.2	0.05

Starting from these results we have decided to estimate a GARCH model (Generalized AutoRegressive Conditional Heteroskedasticity) [6]. In a GARCH process, the conditional variance depends on the long-term unconditional volatility, p most recent values of the variance and the square of the last q past returns, according to equation (5):

$$\sigma_t^2 = V_L\gamma + \sum_{i=1}^{p} \alpha_i * u_{t-i}^2 + \sum_{i=1}^{q} \beta_i * \sigma_{t-i}^2 \quad (5)$$

Where V_L is the unconditional (or long term) volatility, u_{t-1}^2 is the squared log-return observed in t-1, and σ_{t-1}^2 is the conditional volatility observed in t-1. γ_i, α_i and β_i are the three weights whose sum is equal to 1. In the case of a GARCH (1,1) and assuming ω=$V_L\gamma$, the equation (5) can be rewritten as (6):

$$\sigma_t^2 = \omega + \alpha_i * u_{t-1}^2 + \beta_i * \sigma_{t-1}^2 \quad (6)$$

By applying a Maximum Likelihood (ML) approach, it is possible to estimate the three parameters ω, α and β, obtaining then γ, where γ =1-α-β. Writing the estimated variance in T as $v_t = \sigma_t^2$ and assuming that the probability distribution of u conditional to the variance u_t^2 is normal, the ML equation that has to be maximized becomes:

$$L = \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi v_i}} \exp\left(\frac{u_1^2}{-2v_i}\right) \quad (7)$$

By applying the natural logarithm, eq. (7) can be rewritten as:

$$L = \sum_{i=1}^{m} \left[-\ln(v_i - \frac{u_1^2}{-v_i})\right] = \sum_{i=1}^{m} L_i \quad (8)$$

The next step has a computational nature: by using a traditional numerical optimization approach such as the Nelder-Mead simplex, it is possible to obtain the value of the weights that maximize L. Once these weights have been estimated, Equation (6) has been substituted in Equation (3); in other terms, the GARCH volatility structure has been plugged in the Monte Carlo simulation that describes the dynamics of correlated assets with the Cholesky decomposition.

3.3) Econometric methodology: Bayesian VaR.

In order to further diversify the approach used to describe the dynamics of correlated assets, we decided to use an additional approach related to Econometric methods: a BVAR - Bayesian Vector Autoregressive model. This model allows to include in the stochastic component the uncertainty that results not only from the shocks, but also from the variation of the correlation between the indices over time. Taking as a reference an econometric model written in the form:

$$y = X\beta + \varepsilon \qquad \varepsilon \sim N(0,\Sigma) \quad (9)$$

The main parameters are the vector of the coefficients β and the variance-covariance matrix of the errors Σ, which in this case is distributed according to a multivariate Normal. The principle of Bayesian analysis consists in putting together the information that is available in advance on the distribution of these parameters (the so-called prior distribution) with the information that we can obtain from the data (i.e., the likelihood function). In this way it is possible to obtain a new probability function that takes into account both factors, the so-called posterior distribution. The essential step for putting together the prior distribution and the likelihood function is the Bayesian rule. For a vector of parameters θ and a dataset y, given the density function f (y | θ), the Bayesian rule can be expressed as:

$$\pi(\theta|y) = \frac{f(y|\theta)}{f(y)}\pi(\theta) \quad (10)$$

The formula states that, given y, the probability that the "true value" of the parameter vector is θ is equal to the likelihood function of the data multiplied by the a priori distribution of the vector of parameters π (θ) and divided by the density of the data f (y). The vector of the parameters mentioned above is made up of two different elements: the vector of the coefficients β and the variance-covariance matrix of the errors Σ. For each of these elements, it is necessary to specify a prior probability distribution that allows - together with the likelihood function - to implement the “Bayesian rule”. One of the most widely used prior distribution is the "Minnesota prior".

The Minnesota prior assumes that the variance-covariance matrix Σ is already known. Therefore, only the vector of the coefficients β remains to be estimated: for this purpose, it is necessary to identify the likelihood function of β, f (y| β), and a prior distribution π(β). The starting point is the likelihood function: equation (9) implies that y is distributed as a normal multivariate distribution with mean Xβ and variance-covariance matrix Σ. Various techniques can be employed in order to estimate the matrix Σ. With enough computational power it is possible to relax the hypothesis of the diagonality of the Σ matrix described by [5], which is then derived from the variance-covariance matrix of a similar VAR model estimated via Ordinary Least Squares – OLS. Consequently, the maximum likelihood function can be written as:

$$f(y|\beta,\Sigma) = (2\pi)^{-nT/2}|\Sigma|^{-\frac{1}{2}} \exp[-1/2(y-\bar{X}\beta)'\bar{\Sigma}^{-1}(y-\bar{X}\beta)] \quad (11)$$

The notation can be simplified by collecting the terms that do not depend on β:

$$f(y|\beta,\Sigma) = \alpha \exp[-1/2(y-\bar{X}\beta)'\bar{\Sigma}^{-1}(y-\bar{X}\beta)] \quad (12)$$

Looking at the distribution of β, it is supposed to follow a normal multivariate distribution with mean β_0 and variance Ω_0. In the original formulation [10] the expected value for each parameter (which contributes to the specification of the β_0 vector) is equal to 1 for the first lag and equal to 0 for the following lags since most of the time series are characterized by the presence of a unit root. The variance-covariance matrix Ω_0 is a diagonal matrix whose terms are defined by a set of parameters usually derived from the econometric theory. For a discussion of the estimation of these parameters, see the reference literature [5]

The chosen approach uses a slightly more complex variant of prior distribution with respect to the Minnesota prior: the normal-inverse-Wishart prior [9]. The main difference is related to the variance-covariance matrix of the errors Σ, which is also an unknown and is no longer known in advance (in line with the approach adopted in the Monte Carlo GARCH which also stochastically considers non-homoskedastic components). It is assumed that Σ follows an inverse Wishart distribution which has as input parameters the matrix Ω and the number of degrees of freedom v. In mathematical notation: $\Sigma \sim W^{-1}$ (Ω, ν), where the matrix Ω is equal to the amount (v – number of parameters – 1) multiplied by the diagonal matrix that contains in the main diagonal the variance of the errors of each single variable calculated with an equivalent number of AR models.

Furthermore, in this setting, the variance-covariance matrix of β is not symmetric: the hypothesis of independence between the variables is relaxed; the new variance-covariance matrix of β is obtained by multiplying Σ by φ_0, a matrix with each dimension equal to the number of variables. The new distribution of the coefficients is therefore:

$$\pi(\beta) \sim N(\beta_0, \Sigma \otimes \varphi_0) \quad (13)$$

Where the elements of φ_0 depend on a series of hyperparameters set a priori according to the model. In this case, a notable aspect of this approach is the relationship between the prior distribution of the parameters β and the matrix Σ.

A rather common problem in the calculation of the correlation coefficient between assets is the effect of variance: in conditions of relatively low volatility on the market, an apparent increase in correlation between the assets can be noticed, as the declining volatility decreases the denominator of the correlation coefficient. Conversely, if volatility increases, the denominator increases and it is possible to observe an apparent decline in correlation between assets. The term "apparent" is used here because, in fact, there is not a real variation in the risk linked to the correlation between assets: simply, in proportion, the co-movements between the assets are larger or smaller with respect to the variability of each asset.

However, depending on what the stochastic component of the model is, the effects on VaR estimation can be quite significant. In the Monte Carlo VaR, discussed in 3.2, this risk is reduced by the fact that the correlated shocks of an asset are multiplied by the corresponding standard deviation of each asset, thus offsetting any "apparent" effects due to changes in volatility. However, if one wants to analyse a BVAR model where the only varying parameter is the cross correlation between two or more assets, then the difference between the "real" and the "apparent" variations would become much more marked. This can be a very important theoretical problem while trying to generalize the models.

By way of example, looking at the S&P500 index data from 1992 to 2010, it is possible to observe a series of recurring patterns in the correlation matrix between different sectors and, starting from these patterns, to identify 8 "states" of the market characterized by particular configurations of this correlation matrix [12]. More importantly, the authors of this study show that the market only moves along continuous states. This is an interesting starting point for any subsequent calibration of the a posteriori distribution of the coefficients of interest. In the case of the Bayesian VaR, which adopts the normal-inverse-Wishart prior, this problem is implicitly solved by creating a proportional relationship between the variance - covariance matrix of the coefficients and the Σ matrix.

More generally, the calculation of a VaR with the Bayesian approach starts from the a posteriori distribution: once a distribution for the parameters has been defined, n possible values of each parameter are calculated, thus obtaining n possible simulations of the returns of each asset. Starting from these values, with the techniques already described for the other methods based on quantiles, VaR and Expected Shortfall are computed.

In order to implement the BVAR model, the first step has been a model selection analysis in order to identify the most suitable number of lags for the model. First, a maximum number of 10 lags has been set for the model. Subsequently, a specific VAR model has been estimated in a structured form, for each of these lags.

Once a VAR model has been estimated for each lag, the value of two information criteria has been calculated for each model: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).
The best model has been identified for each of the two information criteria: if both suggested a model with a certain number of lags, that model has been adopted; in case of disagreements, the model with the best BIC has been chosen: it is the criterion that attributes the greatest penalty (log (n)) for the number of parameters, thus making the modelling more prudent. Note that this caution does not only derive from a modelling issue: a big problem encountered in building the Bayesian model is the computational time.

Models that consider more than one lag make it problematic to conduct a sufficiently large number of simulations. In about 99% of the more than 5000 iterations replicated for the entire dataset, this strategy chooses a VAR (1) as the best model for all four assets in the portfolio, which has been consequently adopted for the entire dataset. This criterion indicates rather low orders of representation, orienting us towards parsimonious models in terms of the number of parameters. Obtaining a simulation of the returns for each asset is a computationally complex challenge for the implementation of the Bayesian method. Estimating a Bayesian VAR for each iteration and conducting 10,000 simulations may require many hours, if the code is not parallelized, depending on the processors used.

4) The Commitment Machine

4.1) Design phase

As already highlighted, the need to have a precise VaR and ES estimation for T + 1 in T makes it necessary to selectively use the models described in section 3. To this end, it has been decided to opt for an algorithmic solution, that we refer to as Commitment Machine (CM), which, given a set of VaR methods considered statistically reliable, allows to select on each day T a method that has been more performing in the previous period and to use it in order to calculate VaR and ES thresholds for the T+1 observation.

The time interval in which the CM evaluates the performance of the models is defined as the 'observation window': so, for example, if we use data ranging from T-49 to T, the observation window is equal to 50.

Given these specifications, the CM algorithm has been defined starting from 3 elements:

1. A set of methods for calculating the risk measures used (VaR and ES), whose adequacy has been tested by the CM, together with the portfolio return data to be used for the backtesting and for the calculation of the loss function.
2. A loss function to be minimized that allows the meta-heuristic to select the different calculation methods of VaR and ES.
3. An observation window of n observations ranging from T- (n-1) to T which is used for estimating the loss function.

The first step to proceed with the construction of the CM is the selection of the components, i.e., the choice of which methods of VaR and ES calculation are considered valid for the purpose of inclusion in the CM. To this end, it has been considered appropriate to evaluate the single models on the basis of their overall forecast performance. These models have been calibrated to provide a minimum return threshold corresponding to a 95% confidence interval. This implies that we must expect to have lower returns than the corresponding VaR thresholds in 5% of the observations.

Consequently, the frequency of VaR overruns has been selected to be the performance measure. This frequency is expected to be slightly higher than 5%: it is likely that in the event of a market collapse, the implemented VaR methods will need more than one day to recalibrate the estimated VaR and ES values. However, this percentage must not be much higher than 5%: otherwise, the system of risk management methods could not be considered robust and reliable. Furthermore, a frequency of violation greater than the theoretical value (in this case, 5%) by a wide margin could lead to an inspective intervention in order to understand the low effectiveness of the risk measure adopted.

Table 4.1 displays the percentage of days in which portfolio returns are lower than the daily VaR threshold calculated by the model, for each method, during the whole time span of our dataset

Table 4.1 VaR violation frequency for the five daily VaR estimation models

VaR Threshold violation frequency				
Historic	FHS	Monte Carlo	Monte Carlo GARCH	BVAR
5.44%	5.46%	5.37%	5.31%	5.21%

The results show that all five methods have a violation frequency greater than 5%, but apparently not much higher. Consequently, we decided to select all of the five models as potential candidates for the CM VaR selection. The correctness of this selection method has also been verified in a more rigorous way using two different statistical tests:

1) Binomial Test: The binomial test compares the observed number of violations, x, with the expected number. The observed number of violations is assumed to follow a binomial distribution. By using the properties of the binomial distribution, it is possible to construct a test statistic p expressed as:

$$Z_{bin} = \frac{x-Np}{\sqrt{Np(1-p)}} \quad (14)$$

N is the number of observations, p = (1 - VaR level) is the probability of observing a violation if the model is correct, and x is the number of violations observed. This value is compared with the expected value $x - Np$, that is supposed to follow a normal distribution with zero mean. The null hypothesis states that it is possible to observe $x - Np$ violations, that is, Z_{bin} lies in the confidence interval of the theoretical normal distribution. In the case of the five VaR models mentioned above, the binomial test leads to accept the null hypothesis in all cases.

2) "Traffic Light test": the TL test, proposed by the Basel committee [2], measures the probability of observing a number of violations equal to or less than x (i.e., the cumulative probability from 0 to x). The probability of observing a certain number

of violations is supposed to follow a binomial distribution similar to the one seen for the binomial test. The test is called "traffic light test" because, depending on the realized performance (the number of overruns), the test classifies a VaR estimation model in one of the following three zones, denoted by a different colour:

- Red zone: the probability of observing a number of violations equal to or smaller than the number actually observed is 99.99%. Results are in the tail of the distribution: it is very unlikely that a VaR model with x violations on N observations is correct. The model is therefore discarded.
- Yellow zone: the probability is between 95% and 99%. The number of violations is very high but not excessively so the model is acceptable, but on the condition that the calculated VaR thresholds are strengthened.
- Green area: The probability is 95% or less. The model is accepted.

All selected VaR models fall within the green zone.

The second step needed in order to design the CM is the choice of the loss function to be minimized: it is necessary to specify a performance objective that can be used to evaluate different VaR estimation methods, taking into account the goal of obtaining a method that produces a percentage of overruns as close to 5% as possible (i.e., a theoretically solid model, able to reliably predict the actual value of the estimated quantile).

A first possible performance measure could be, again, the violation frequency. However, this is a measure with low sensitivity for limited time intervals: in certain periods, it is likely that the percentage of violations of the five methods is the same depending on the market phase considered. A second problem with the violation frequency is that it is a partial measure: the magnitude of violations is not factored in. There might be many small violations or few very large ones, where the return even falls below the Expected Shortfall thresholds.

The alternative approach is to focus on losses: in such case, the CM evaluates the VaR models based on the cumulated losses in the observation window. The use of a loss function to compare the performance of different VaR estimation methods has already been discussed in literature, starting with [11] that suggest three different possible loss functions (LF) that reflect specific concerns of risk management and tend to grow in the case of VaR estimation model failure. From this work various analyses of possible LF have been derived and the most common is the "regulatory loss function" proposed by [17], which incorporates one of the LFs proposed by [11].

The regulatory loss function has a functional form similar to the one described in equation (16) that is adopted in this work, but it weights the losses according to a quadratic function; it is supported by another LF which also takes into account the opportunity costs related to the capital absorbed by VaR in periods in which the VaR threshold is not exceeded. Compared to these LFs proposed by [17] and [11], the two LFs tested in this work (equations (15) and (16)) are easier and allow an immediate comparison between the different VaR techniques.

In order to adopt an approach that is as flexible as possible, we decided to test 2 different loss functions, one characterized by the objective of minimizing violations and one with the objective of minimizing the overall losses.

Defining R_T ad V_T respectively as the returns at date T and the corresponding VaR threshold, the loss function for a certain VaR model takes the functional form described in equation (15) if we want to minimize violations, while equation (16) is more suitable in case we want to focus on losses minimization.

$$f_1(x) = \begin{cases} 1, & R_T < V_T \\ 0, & R_T \geq V_T \end{cases} \qquad (15)$$

$$f_2(x) = \begin{cases} \mathrm{Vt} - \mathrm{Rt}, & R_T < V_T \\ 0, & R_T \geq V_T \end{cases} \qquad (16)$$

Minimizing the objective function, therefore, is equivalent to iteratively choosing the method of VaR estimation which leads to lower losses or overruns. It is assumed that a "weak" model in the previous period maintains this weakness also in T + 1: since the five models used are structurally very different, it is expected that they can capture the market trend with a different degree of precision, depending on the market conditions in the period considered.

The LF is not the only possible way to take into consideration temporal variations in the ability of single models to predict the actual value of the VaR: in literature, for example, methods of estimating the VaR are based on a weighted average of the VaR thresholds estimated by several models [4].

This approach has been discarded because it is more difficult to calibrate and returns less informative results than simply selecting the best model (meaning that it is more difficult to interpret, also in terms of performances).

Although the performance measure takes into account the VaR threshold, the CM is designed to calculate both VaR and ES: when a certain method is selected as "the best performing", not only the VaR value in T+1, but also the corresponding expected shortfall value on the same date is calculated by using that method.

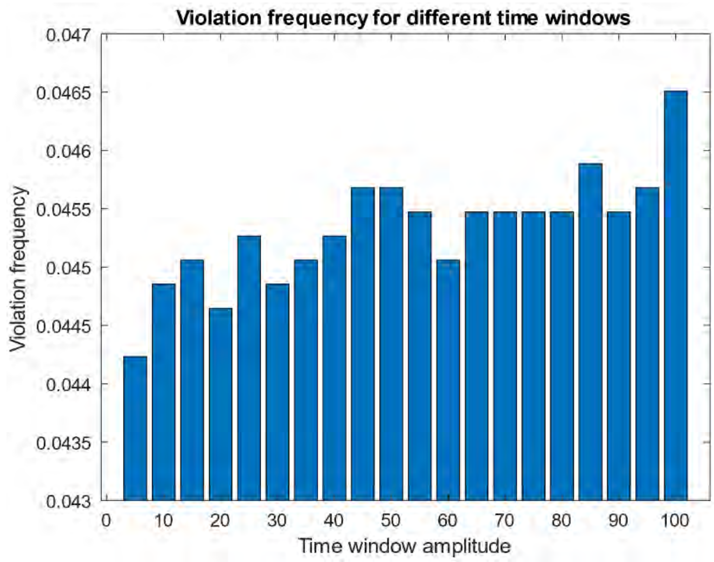
4.2) Calibration phase

The last step for the CM design is the determination of the time interval used for its calibration. This process highlights an important trade-off. While, as the observation window expands, there is a more consistent estimate of the relative effectiveness of the various models, on the other hand a further argument in favour of a short observation window is the rapidity of reaction; among the main issues for risk management are sudden moments of market collapse. Such events occur very quickly, making it advisable, under certain circumstances, to have a CM that adapts as quickly as possible.

In order to balance this trade-off, we decided to test the algorithm on 20 possible observation windows of various length between 5 and 100 days (5, 10, 15 .... 100 days). To evaluate the performance of the CM ex-post, we decided to calculate, for each simulation, both the frequency of VaR violations (given the regulatory and management importance of this indicator) and the overall average of the losses occurred (i.e., the sum of the VaR overruns divided by the number of observations) for the VaR thresholds calculated by each loss function. The results of this calibration show that the loss function based on losses below the VaR threshold performs better than the one based on violations. More specifically, by taking the mean of the results of all of the possible simulations (20, one for each simulated observation window) we find that the CM adopting a LF based on the overruns has a frequency of violations of 4.82%, while the CM based on losses has a frequency equal to 4.53%. For all observation windows, as can be easily imagined, the LF that tries to minimize losses below the VaR threshold has minor losses. The superiority of a loss-based function with respect to both performance measures considered (losses and overruns) is supported by the data. Therefore, we decided to adopt a LF that aims to minimize losses.

Figure 4.1 shows the various violation frequencies achieved by the CM for observation windows ranging from 5 to 100 days. On the horizontal axis we show the different amplitude values of the observation windows for which the percentage of violations has been calculated (vertical axis).

Figure 4.1 Violation frequency of the CM



The first important consideration that can be deduced from the figure is that the logic behind the CM appears to be quite reasonable; indeed, for any interval tested, the frequency of violations is less than 5%, and it is between 4.42%, measured with a 5-day interval, and 4.65%, measured with a 100-day interval. The CM is effective in reducing overruns, although this effectiveness decreases as the observation window is extended. The high sensitivity of the LF leads to very prudent results if applied over very short time intervals; however, the proposed goal is to have a robust method for estimating a 95% confidence interval and a violation frequency far below 5% means having an excessively cautious model compared to the initial specifications. Considering Figure 4.1, a violation frequency near to 4.5-4.6% would represent a reasonable result, given that it is better to have an excess of prudence than an excess of risk: in this sense, having a percentage of overruns of 4.6%, for example, is much better than having a percentage of 5.4%, although in both cases the absolute distance from the target is the same. To this end, two different time windows have been chosen to calibrate the CM: one at 50 days and one at 85 days. The reason for this double choice is that the violation frequency is relatively stable between 45 and 95 days and choosing values within this range seems to ensure that we adopt a reliable model at the same time.

Table 4.2 shows the main characteristics of the two different CMs, differentiated according to the number of observations considered each time (50 for the first variant, 85 for the second).

Table 4.2: Summary of the two selected CM variants

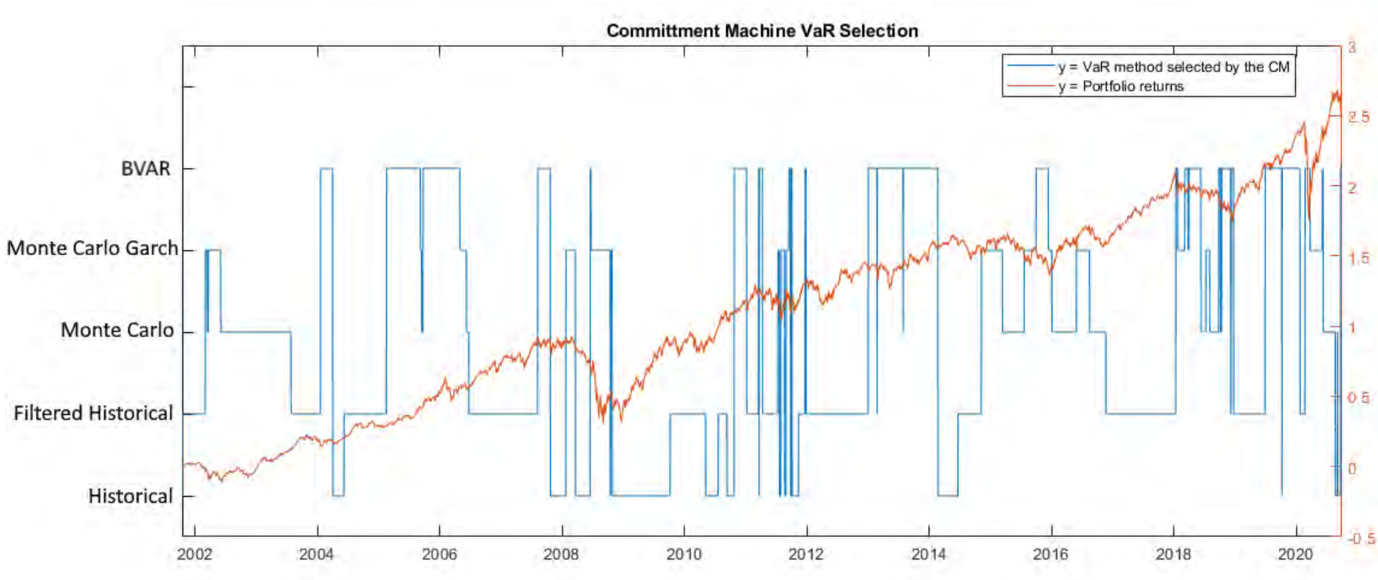
	Violation Frequency	**Frequency of VaR selection by CM**				
		Historical	Filtered Historical	Monte Carlo	Monte Carlo Garch	Bayesian
CM (50)	4,92%	13,4%	37,3%	14,3%	11,7%	23,3%
CM (85)	4,82%	13,0%	37,4%	12,5%	14,1%	23,0%

The two CM variants CM(50) and CM(85) are very similar in terms of how frequently the various VaR methods are chosen. The only difference is that the version based on the shorter observation window chooses more frequently the Monte Carlo method without GARCH volatility and less frequently the Monte Carlo method which includes the GARCH. It is important to notice that using two different CMs makes it possible to verify that the heuristic underlying the CM is efficient. In section 5, the results (in terms of VaR selection) of both of these CMs were evaluated. From the point of view of the robustness of the methodology, using multiple variants can confirm that the design flexibility of the CM does not translate into an excessive variability of the results. While in fact the adoption of a different observation window is optimal for the needs of different subjects and increases the flexibility of the algorithm during implementation, it is clear that the CM (with the same selected methods) should provide a consistent approach. Indeed, choosing a window observation of 50 or 85 days should not lead to too different selections, otherwise the algorithm of the CM would be excessively unstable and the possibility to choose between multiple observation windows would add a further layer of complexity to the model risk, instead of making the implementation more flexible.

5) Results

This section focuses on the analysis of the results of the two different CM variants discussed above. In particular, the analysis is divided into two steps: a disaggregated analysis of the performances of each CM on the days in which one of the five methods has been selected (historical VaR, filtered historical VaR, Monte Carlo, Monte Carlo GARCH, Bayesian) and a general analysis of the performances of the two different CMs over the entire dataset. In order to distinguish between the different CMs, the notation proposed in section 4 is maintained: CM(50) is the CM that uses an observation window - in accordance with the adopted heuristics - of 50 trading days, while CM(85) is the CM that uses a window of 85 days.
Starting from the analysis of the methods selected by the various CMs, we first analyze the CM(50), i.e., the CM calibrated over a time interval of 50 days. This variant of CM is particularly interesting because it shows how the optimization algorithm performs on relatively short time intervals, where the information set available is more volatile. It should be noticed that in 83.4% of the periods considered, the CM(50) and the CM(85) choose the same VaR method. While this data highlights the consistency of the algorithm adopted, on the other hand it allows us to take the VaR selection of one of the two as representative of both CMs, thus making this analysis much easier.
Figure 5.1 shows, for the CM(50), the type of VaR model selected by the algorithm (labelled on the y axis) and the performance of the portfolio (measured by cumulative returns) from the beginning of the dataset to the end.
By observing Figure 5.1, two significant aspects of the CM can be noticed:
1) Stability: for many long periods, the CM always chooses the same method for calculating the VaR. This is a very important factor, as it highlights how CM's choices are consistent over time. There is no random selection, but a logical choice. This stability is less evident when the Bayesian method is selected since Bayesian estimators tend to produce more volatile values [10].
2) In the majority of the phases characterized by positive portfolio returns, the CM chooses the historical methods, particularly in 2007, in the period 2009-end 2011, in 2014 and in 2017. The fact that in these periods mean returns are positive makes it more difficult for the Monte Carlo VaR to generate a prudential forecast, compared to the historical VaR. Conversely, the VaR estimated with the historical method remains very conservative, even in a context characterized by positive returns [15]. In times of market downturn, however, the CM switches to the Monte Carlo VaR, with or without GARCH (particularly in the subprime mortgage crisis and in the fall of 2016) and, during the past two years of declining trend, it also adopts the Bayesian VaR, whose estimates tend to be more conservative than the other methods employed [14]. In the case of the 2008 crisis, for example, the historical VaR is initially chosen and after some observations, as the market trend reverses, the CM switches to the Monte Carlo VaR, which is more prudent also due to the trend of cross-correlations.

Figure 5.1 CM model selection



A significant element for evaluating the performances of the two different variants of CM is their ability to select, for each day, a model whose ex-post performance is better than the performances of the other models. Before making this comparison, it is advisable to have a general overview of the performance of the single VaR methods over the days they have been selected, by using two metrics: the violation frequency and the average loss.
The average loss is simply the average, for each method, of the values assumed by the loss function introduced in section 4 (equation (16)) on the days in which that method is selected; the average loss of the historical VaR, for example, represents the total sum of the negative differences between returns and Historical VaR in the days in which the CM selects the historical VaR, divided by the total number of days in which the historical VaR is selected. Tables 5.1 and 5.2 show the values of these two indicators for each VaR method selected by each of the two CMs.

Table 5.1 Violation frequency of different VaR methods in the day in which they are selected by the CMs

	Average violation % when selected				
	Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	2,91%	3,02%	2,89%	9,75%	6,48%
CM(85)	2,53%	3,61%	2,82%	8,36%	6,01%

Table 5.2 Average VaR loss for different methods in the day in which they are selected by the CMs

	Average VaR loss when selected				
	Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	0,014%	0,015%	0,013%	0,074%	0,039%
CM(85)	0,010%	0,016%	0,011%	0,077%	0,039%

Tables 5.1 and 5.2 show that, on average, both the CMs choose the Bayesian method and the Monte Carlo GARCH method in the days in which VaR violations are more frequent (Tab 5.1) and there are greater VaR losses (Tab 5.2). This is consistent with the fact that these two methods are selected during periods of market crisis, as shown in Figure 5.1.
To better contextualize these results, it is important to evaluate the performances of other VaR methods observed in the same days. To this end, we proceed to further break down the analysis and we consider, for each CM, the days in which each VaR method is selected. The dataset has been divided into 5 sets, each containing the days in which one of the 5 methods is selected. Then, the average violation and VaR losses have been calculated for each method in each of the five sets, and these data have been compared with the performances of the other methods on the days of the same set.
Table 5.3 shows the violation percentage frequency. Here is an explanation of how to read Table 5.3: the first of the five rows of the CM(50) shows, for each VaR estimation method, the percentages of overrun of the VaR threshold calculated on the days in which the method selected is the unfiltered "base" historic VaR.
It is expected that, on the days in which the CM chooses the historical method, this percentage will be lower for the historical VaR, i.e., the CM is able to choose, in T+1, the method for which it has the smallest chance of seeing one exceeding the VaR threshold.
In order to simplify the analysis, the performances of each method on the days in which it is selected have been highlighted in bold. For each row, these performances are expected to be the best. The analysis of the results shows a significant selection ability of the CM; with the exception of the historical unfiltered VaR, in all considered cases, the CM is able to manage the choice of the method that guarantees the lowest violation frequency. The historical VaR is an exception, probably because in 2008 it has been chosen by the CM in the very first days of the outbreak of the financial crisis (see Figure 5.1).

Table 5.3 Analysis of violation frequency of VaR methods

		Average violation % when selected				
		Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	Historic	**2,91%**	2,75%	3,21%	3,52%	2,91%
	F. Historic	3,85%	**3,02%**	4,40%	4,34%	3,90%
	Montecarlo	4,04%	3,75%	**2,886%**	3,319%	4,473%
	Montecarlo Garch	10,99%	12,77%	10,284%	**9,752%**	10,461%
	BVAR	7,55%	8,35%	7,282%	6,927%	**6,483%**
CM(85)	Historic	**2,53%**	2,37%	3,165%	3,481%	2,532%
	F. Historic	4,44%	**3,61%**	4,874%	4,819%	4,436%
	Montecarlo	3,65%	3,48%	**2,819%**	3,317%	4,312%
	Montecarlo Garch	9,384%	11,144%	8,798%	**8,358%**	9,238%
	BVAR	7,271%	7,810%	6,732%	6,373%	**6,014%**

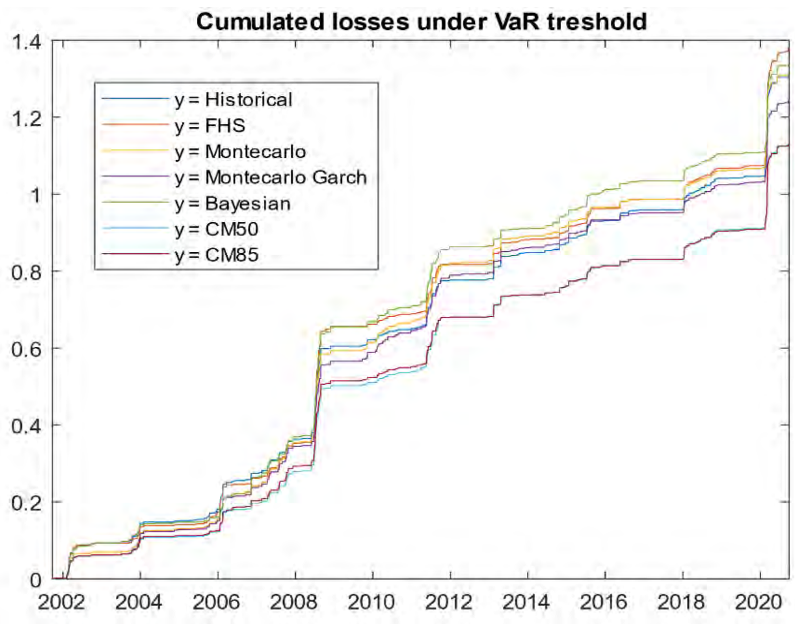
The Monte Carlo GARCH method is selected on the days in which, on average, the frequency of violations of the various methods is very high. As already indicated in the design phase, its ability to take into account both the correlation between returns and the non-stationarity issues makes it quite useful in times of crisis, when VaR violations are generally more frequent. The Bayesian VaR, when selected by the CM, is also the best and this represents a noteworthy result, considering the high sensibility of the Bayesian model to parameter estimations. However, it should be considered that the heuristic underlying the CM is different since the aim is to minimize not the frequency of violations, but the overall losses.
In this sense, a more convincing performance measure is the average of the losses realized in the days in which our two CMs select the various VaR methods. Resuming the objective function discussed in section 4, this measure is equal to the average value of such function on the analysed days.
In this way, it is possible to derive an ex-post value that indicates how the selected method really minimizes the value of the objective function compared to the other methods. Higher values indicate higher losses in the days considered and vice versa. Table 5.4 shows the values of the average losses, organized in a similar way to that seen in Table 5.3.

Table 5.4 Analysis of the average VaR loss methods

		Average VaR loss when selected				
		Historic	F. Historic	Montecarlo	Montecarlo Garch	BVAR
CM(50)	Historic	**0,014%**	0,014%	0,017%	0,017%	0,016%
	F. Historic	0,015%	**0,013%**	0,018%	0,018%	0,017%
	Montecarlo	0,013%	0,012%	**0,011%**	0,011%	0,015%
	Montecarlo Garch	0,074%	0,090%	0,066%	**0,061%**	0,075%
	BVAR	0,039%	0,040%	0,038%	0,035%	**0,034%**
CM(85)	Historic	**0,0099%**	0,0098%	0,013%	0,013%	0,012%
	F. Historic	0,018%	**0,016%**	0,021%	0,020%	0,020%
	Montecarlo	0,012%	0,011%	**0,0094%**	0,0095%	0,014%
	Montecarlo Garch	0,063%	0,077%	0,056%	**0,052%**	0,064%
	BVAR	0,038%	0,039%	0,036%	0,034%	**0,032%**

Results suggest that the CM effectively manages to accurately select the various methods. In each row, the minimum loss is associated with the method selected by the CM; for example, in the last row it can be seen that, on the days in which the CM(85) selects the Bayesian method, this method is the one with the lowest average losses (0.032% against 0.034-0.039% of the other methods). From Table 5.4 we further notice the difficulty of the two CMs in making the best use of the unfiltered historical VaR.
Starting from this data, it is appropriate to take a step back and aggregate the analysis once again in order to evaluate the overall performance of the five VaR methods (Historical, Filtered Historical, Monte Carlo, Monte Carlo GARCH, Bayesian) and of the two CMs over the entire time series.
Figure 5.2 shows, for the time interval considered, the cumulative losses with respect to the VaR threshold of the five methods considered by the CM and of the two variants of CM. The cumulative loss on a certain date t is equal to the sum of the historical values from 0 to t of the loss function introduced in section 4, equation (16) and it represents the total amount of losses below the VAR threshold up to that day.

Figure 5.2 Cumulative losses of VaR methods and CMs



The data show two periods in which the portfolio registers high losses, specifically the subprime mortgage crisis and the Covid shock in the spring of 2020. Between these two periods there is a further phase of losses – of smaller magnitude than the other two - corresponding to the crisis of the sovereign debts in the Eurozone. During these two negative market events, the cumulative losses increased considerably both for the five VaR methods used and for the CM built on these methods. The subprime mortgage crisis is a key moment for this analysis. Indeed, the two CMs suffer lower losses in such period, while the two methods with the worst performances are the Monte Carlo VaR (without GARCH) and the Bayesian Var.

This is a further endorsement of the validity of the heuristic approach adopted in the design phase, which over the course of the length of our time series allows the algorithm to avoid choosing, among the five methods, the ones that have the worst performance.

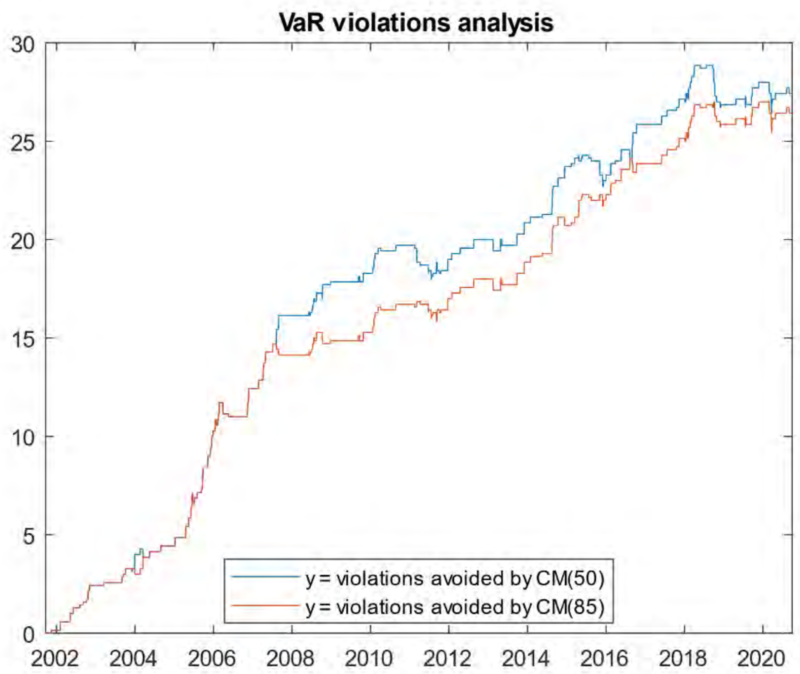
A further element of interest is the temporal trend of the cumulated losses: considering Figure 5.2, almost a third of the losses are realized in the two rare market events (the subprime mortgage crisis and the 2020 crisis). It is important to underline how, during the 2020 crisis, the VaR method (out of the three initially chosen) that accumulates the smallest losses is the Bayesian VaR. This data confirms the validity of a Bayesian approach in a scenario in which the market situation changes (more or less literally) overnight.

Another evidence of the goodness of our results can be seen by taking into account the amount of VaR violations that are avoided by the CMs. This measure can be estimated by the difference between the average number of VaR violations computed by the five methods and the VaR violation committed by the CMs. For example, a difference of 3 means that, on average, there are 3 days in which the VaR threshold calculated by the CM is not violated, while the other methods have VaR threshold violations. Figure 5.3 shows this difference: CM algorithm is successful at avoiding VaR violations and the smooth growth of the measure over time suggests a stability of the positive performances of the CM optimization algorithm.

It is worth to notice that the logic of CM approach allows to calculate the Expected shortfall (ES) for a given portfolio: once a VaR method is selected by the CM, the mean of returns under VaR threshold can be easily calculated.

This is a valuable feature considering that ES is a coherent risk measure and thus better suited as a risk measure for portfolio optimization [16].

Figure 5.3: Analysis of daily VaR threshold violations



6) Conclusions

In this paper we have designed an algorithm which performs an automatic choice among different VaR Methods, based on the minimization of a loss function that takes into account the negative returns below the VaR threshold.

The implemented Commitment Machine (CM) handles five different estimation techniques: traditional historical VaR, filtered historical simulations (FHS) VaR, Monte Carlo VaR with cross correlated assets, Monte Carlo GARCH VaR with cross correlated assets, and Bayesian VaR. These different models are able to take into account different econometric aspects of the time series, particularly non-stationarity and cross correlation.

The analysis of the CM performances, tested on the realized returns of an equally weighted portfolio made up of four market indices from 2001 to 2020, shows that the use of the algorithm provides interesting advantages compared to the implementation of a single VaR method.

Thanks to its adaptive logic selection, the CM records lower VaR violations and losses than the five methods individually implemented.

The flexibility of the code written in MATLAB environment guarantees the possibility of generalizing the analysis by including other VaR estimation methods, even different from the ones used in this work. It also offers further possibilities for alternative implementations, for example by modifying the loss function in order to consider the needs of the various entities involved in the risk assessment.

In conclusion, our applied analysis provides significant evidence in favor of the goodness of the design of the proposed CM, thus making it a useful tool for managing portfolio risk.

Marco Bagnato, Anna Bottasso and Pier Giuseppe Giribone

Bibliography

[1] Ang A., Chen J. (2002) "Asymmetric Correlations of Equity Portfolios". Journal of Financial Economics, vol. 63, issue 3, pp. 443-494

[2] Basilee Committee on Banking Supervision (1996) "Supervisory Framework for the use of backtesting in conjunction with the internal models approach to market risk capital requirements"

[3] Bottasso A., Giribone P. G., Martorana M. (2019) "The emerging market of guarantees of origin: analysis and design of a quantitative measurements system for monitoring financial risks", Risk Management Magazine Vol. 14 N. 2

[4] Casarin R., Chang C. L., Martin J. J., McAleer M., Amaral T. P. (2012) "Risk management of risk under the Basel Accord: A Bayesian approach to forecasting Value-at-Risk of VIX futures", Mathematics and Computers in Simulation, Vol. 94

[5] Dieppe A., Legrand B., Roye R. (2018) "The Bayesian Estimation, Analysis and Regression (BEAR) toolbox", Technical guide

[6] Engle R. F. (1982) "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", Econometrica 50, pp. 987-1007.

[7] Haug E. G. (2007) "The complete guide to Option Pricing Formulas", McGraw Hill, cap 8.1.4

[8] Hendricks D. (1996) "Evaluation of Value-at-Risk Models Using Historical Data", Economic Policy Review, Vol. 2, No. 1

[9] Karlsson S. (2012) "Forecasting with Bayesian Vector Autoregressions", Handbook of Economic Forecasting
Volume 2, Part B, 2013, pp. 791-897

[10] Litterman R. B. (1986) "Forecasting with Bayesian Vector Autoregressions – Five years of experience", International Journal of Forecasting, Vol. 4, No. 1, pp. 25-38

[11] Lopez J. A. (1998) "Methods for evaluating Value-at-Risk estimates", Federal Reserve Bank of New York, research paper No. 9082

[12] Münnix M. C., Shimada T., Schäfer R., Leyvraz F., Seligman T. H., Guhr T., Stanley H. E. (2012) "Identifying States of a Financial Market", Scientific Reports, Vol. 2, No. 644

[13] Perez P., Murphy D. (2015) "Filtered historical simulation Value-at-Risk models and their competitors", Bank of England Working Paper No. 525

[14] Pollard M. (2007) "Bayesian VaR and Capital Charge Puzzle" - papers.ssrn.com/sol3/papers.cfm?abstract_id=1116542

[15] Pritsker M. (2006) "The hidden danger of historical simulation", Journal of Banking & Finance, Vol. 30, Issue 2, pp. 561-582

[16] Quaranta A. G. , Zaffaroni A. (2008) "Robust optimization of conditional value at risk and portfolio selection", Journal of Banking & Finance, Vol. 32, Issue 10, pp. 2046–2056

[17] Sarma M., Thomas S., Shah A. (2002) Selection of Value-at-Risk models, journal of financial forecasting, Volume 22, Issue 4, pp. 277-358