Pathwise Grid Valuation of Fixed-Income Portfolios with Applications to Risk Management

Shiva Zamani
RiskLab, GSME, Sharif University of Technology

Ali Chaghazardi
RiskLab, GSME, Sharif University of Technology

Hamidreza Arian
RiskLab, GSME, Sharif University of Technology

July 19, 2021

Abstract

Numerical calculation of Value-at-Risk (VaR) for large-scale portfolios poses great challenges to financial institutions. The problem is even more daunting for large fixed-income portfolios as their underlying instruments have exposure to higher dimensions of risk factors. This article provides an efficient algorithm for calculating VaR using a historical grid-based approach with volatility updating and shows its efficiency in computational cost and accuracy. Our VaR computation algorithm is flexible and simple, while one can easily extend it to cover other nonlinear portfolios such as derivative portfolios on equities and FX securities.

Keywords: Financial Risk Management, Value-at-Risk (VaR), Profit and Loss (PnL) Calculation, Scenario Analysis, Fixed-Income Portfolio
JEL classification: C61, G11.

1 Introduction

This paper proposes a novel algorithm to estimate Value-at-Risk (VaR) and Expected Shortfall (ES) with low computation costs and high accuracy for fixed-income portfolios. The key idea is to use a set of realized base scenarios in a multidimensional space of risk factors to estimate profit and loss (PnL) on the portfolio level. This technique, developed in a fixed income framework, is easily extendable to derivatives portfolios with other underlying securities such as equities and currencies where several risk factors affect the portfolio value. Many of the proposed methods for computing VaR require numerous calculations of the portfolio value to achieve the expected PnL distribution. These calculations are time-consuming, especially for bonds with an embedded option or assets for which no explicit formula exists to calculate shocked values. Also, due to the nonlinear dependence of most financial derivatives on their underlying assets, there is a great deal of complexity in derivatives portfolio valuation. That is why various researchers have proposed methods for faster and calculation of risk measures such as VaR.

By examining the economic conditions and characteristics of measures such as VaR and ES, the Basel Committee obliges financial institutions to use them in the regulatory capital calculation process. An example is FRTB, in which regulators proposed the use of ES with a 97.5% confidence level and

Electronic copy available at: https://ssrn.com/abstract=3889826

varying time horizons after twenty years of using VaR with a 10-day time horizon and 99% confidence to determine market risk capital (we will discuss this further in Section 2). Banks must calculate the regulatory capital based on both standardized and internal model approaches specified by FRTB. Standardized approaches involve delta, gamma, and vega estimations, and internal model approaches are based on stressed ES, implying a heavy computational load for fixed-income portfolios. So any method targeting time reduction in VaR and ES calculations while maintaining the accuracy is well received.

Chen et al. [2019] points out that nonlinearity reduces the effectiveness of approaches such as delta-normal and argues that the lack of a closed-form pricing solution for derivatives securities complicates the problem of measuring risk. To solve this problem, they propose a generic simulation-based algorithm to calculate the VaR for nonlinear portfolios. Francq and Zakoian [2020] propose a new method called Virtual Historical Simulation to estimate the ES of large-scale portfolios. This method avoids facing challenges such as estimating a dynamic model for the vector of risk factors.

Jamshidian and Zhu [1996] proposes an alternative two-step approach. In the first step, a limited number of risk factors explain the yield curve variation. They draw informative factors and ignore the noninformative ones. Then they use principal component analysis (PCA) to identify the orthogonal factors that explain the variation in the yield curve and take into account just the three most important ones. The second step assumes a discrete binomial distribution for each main risk factor and forms a three-dimensional grid. Considering each node of the grid as a scenario and calculating the portfolio's value, the model obtains a discrete distribution for the portfolio value changes and calculates the VaR. But, according to Abken [2000], the Jamshidian and Zhu [1996] method may result in a biased estimation of the true risk measure. Poor performance of this method becomes perceptible when the portfolio shows more severe nonlinear behavior. Further, Gibson and Pritsker [2000] (GP) examines the two simplification steps of Jamshidian and Zhu [1996] and provides appropriate corrections by identifying some pitfalls in each of these steps.

One of the drawbacks of Jamshidian and Zhu [1996] approach is the insufficient explanatory power of their selected risk factors for the changes in portfolio value. Addressing the possible changes in the value of the portfolio and identifying relevant factors are necessary when building a risk measurement machinery. The first PCA factor typically captures an approximate parallel shift in the yield curve. The potential problem arises when considering a portfolio with both buy and sell positions in securities, like short positions in short-term securities and long positions in long-term securities. The changes related to the first risk factor will cause one part of the portfolio to increase and the other part to decrease in value, offsetting each other to some extent. As a result, when the investor has hedged the portfolio against parallel shifts in the yield curve, even though the first factor explains a large portion of the yield curve variations, it may not significantly affect the value of the investor portfolio. Hence, choosing a risk factor in isolation without studying its portfolio exposure is not optimal.

Gibson and Pritsker [2000] propose the Partial Least Squares (PLS) method in which the risk factors are chosen based on their explaining power for the portfolio value changes. While PCA chooses and ranks factors by considering their ability to explain the variance of the data, PLS ranks them by considering their ability to explain the covariance between the data (yield curve variations) and some other variables (changes in portfolio value). By this ranking method and using the most critical factors, this approach can perform dimension reduction.

The other drawback of the Jamshidian and Zhu [1996] method is that a limited number of nodes represent the entire space of risk factors. To address this shortcoming, GP proposes an approach called "Grid Monte Carlo (GMC)". Without significantly increasing the computational load, GMC uses

Electronic copy available at: https://ssrn.com/abstract=3889826

more market information, leading to higher accuracy of VaR measurement. This approach assumes a continuous distribution for each of the identified risk factors, forms a grid similar to Jamshidian and Zhu [1996], and calculates the portfolio value for the scenarios of the grid. In this way, by sampling the assumed distributions and interpolating the portfolio value for these samples, GP obtains a distribution for the portfolio PnL and estimates the VaR. Applying this process and partially taking advantage of the Monte Carlo (MC) simulation reduces the computational load while increasing the accuracy of PnL estimation.

Before Gibson and Pritsker [2000], other researchers have tried to explain and solve the problem mentioned above using similar methods. For example, Abken [2000] shows that the VaR estimated by Jamshidian and Zhu [1996] can fluctuate around the actual VaR erratically for grids which are more practical due to their point densities. Frye [1998] also uses PCA and an approach similar to GMC to estimate the VaR. Chishti [1999] applies this latter method and focuses on classifying the type of errors and calculating their values. Their paper considers each scenario as a point in the space of risk factors, whose dimensions are related to interest rates changes at different maturities. Then, they consider images of these points in a three-dimensional space. It is clear that during this process, the model loses some information of the original scenarios, leading to a projection error (Chishti [1999]). Chishti [1999] introduces another type of error called interpolation error, which is the result of using interpolation procedures to estimate the portfolio values.

So far, we have reviewed the proposed methods in the literature. Here we explain the advantages of our approach. To reduce the computational load of the VaR estimation process for fixed-income portfolios, most of the previous methods use dimension reduction processes leading to projection error. Therefore, it is worth investigating whether it is possible to reduce the computational load without going through the dimension reduction step. Examining the scenarios in the original space of risk factors interprets the portfolio changes more clearly. Hence, instead of working in a space whose dimensions do not necessarily have an economic interpretation, we directly deal with interest rates with different maturities. This way, we remain in the original space but propose a simple method leading to more accurate results compared to previous methods proposed in the literature.

There is also another notable aspect of our method. Like the Grid Monte Carlo (GMC) method, built upon MC, we will introduce a Grid Historical Simulation based on HS. This way, we can even expand the scope of our view and consider the possibility of combining the proposed ideas by Jamshidian and Zhu [1996], Frye [1998], Gibson and Pritsker [2000], etc. with other VaR calculation methods such as FHS, BRW (Boudoukh et al. [1998]), etc. The results of this study show that using such ideas improves our method.

We will evaluate the performance of our proposed method by simulating a portfolio of callable and puttable bonds with different characteristics. For this purpose, we use some existing models in the literature that have been introduced for evaluating VaR calculation methods, including Ranking Model (Şener et al. [2012]), Diebold-Mariano test (Diebold and Mariano [2002]), unconditional coverage (Kupiec [1995]), conditional coverage, and independence tests (Christoffersen [1998]). We also use Murphy diagrams (Murphy [1977]) to rank the methods with respect to consistent scoring functions (Ehm et al. [2016]).

The remainder of this paper is as follows. In Section 2, we introduce our method to estimate the VaR for fixed-income portfolios and improves its performance by updating input scenarios. In Section 3 we evaluate our method using various tests and models and compares it with the GP method. Section 4 concludes.

Electronic copy available at: https://ssrn.com/abstract=3889826

2 Methodology

This section provides a fast and robust approach for the PnL calculation of fixed-income derivatives portfolios under a scenario-based approach. The objective is to choose the grid points from historical scenarios to describe the portfolio value changes adequately.

2.1 Risk Measurement

Considering the expected distribution of portfolio value changes with time horizon T, the absolute value of the $(1-\alpha)$ percentile of this distribution is $\text{VaR}(\alpha, T)$ (Equation (1)). Due to its advantages, this measure was chosen by the Basel Committee on Banking Supervision (BCBS) in Basel II (Basel [2004]) as a standard tool for measuring risk which is expressed mathematically as

$$\text{VaR}_{\alpha,T} = -q_{1-\alpha}(X_T), \tag{1}$$

where $q_\beta(X_T) = \max\{x \mid F_{X_T}(x) \geqslant \beta\}$ is the β-quantile of X_T, and X_T is a random variable representing the portfolio value changes with time horizon T.

Despite its advantages, VaR has some drawbacks such as not being sub-additive, and hence, not being a coherent measure of risk (Artzner et al. [1999], Acerbi and Tasche [2002], and Szegö [2002]). Furthermore, VaR only shows the threshold of severe losses based upon a certainty level and does not consider the number of losses that have exceeded this threshold. This feature led to serious underestimation of risk during the crisis of 2007-2009 since the regulatory capital was determined based on VaR by Basel II rules. As a response, the BCBS introduced the Fundamental Review of the Trading Book (FRTB), a set of revisions to the market risk framework. Among the changes proposed in these revisions, the migration from VaR to ES is an alternative measure to calculate the regulatory capital requirement. In this way, ES becomes the most popular risk measure in financial regulation (Wang and Zitikis [2020]). The ES is equal to the average of severe losses that have exceeded the VaR. Mathematically for confidence level alpha and for the time horizon T,

$$\text{ES}_{\alpha,T} = -\mathbb{E}\left[X_T \mid X_T \leqslant q_{1-\alpha}(X_T)\right]. \tag{2}$$

This measure is sub-additive and gives us the average loss during critical times (see Artzner [1997] and Pflug [2000]). After the release of FRTB, ES comes into the focus of risk managers, and several researchers investigate it further. Among them are Akat and Memis [2018], who compare the performances of VaR and ES in calculating the appropriate capital. It is worth noting that, despite its weaknesses, VaR is still a standard tool for measuring financial risk in the industry (Bustreo et al. [2016]). Since the methods proposed in this study make it possible to estimate the empirical distribution of portfolio value changes, they can calculate both VaR and ES.

2.2 Path-wise Grid PnL Calculation

In order to calculate VaR for fixed-income portfolios, one should examine the yield curve changes. Each scenario for the yield curve belongs to the space of interest rate changes at different maturities. The dimension of this space equals the number of tenor points involved in representing yield curve changes. Most methods use some dimension reduction techniques to overcome the heavy computational load caused by the high dimensionality of the interest rate space (generally between 10 and 15 dimensions).

Electronic copy available at: https://ssrn.com/abstract=3889826

Then, the model selects a set of points that form an ordered grid as the representative of the entire space and calculates the portfolio value changes for the selected grid.

There is a significant difference between our Pathwise Grid VaR (PwG) method and the current literature. In our method, instead of revaluing the portfolio for the selected grid, which consists of hypothetical scenarios, we choose yield curve actual shocks as our base scenarios and calculate the PnL for these scenarios. Contrary to previous approaches, there is no need to relocate scenarios from one space to another. Thus, we avoid the projection error of Chishti [1999]. However, the PwG leads to another type of error with a completely different nature, which we will explain later.

Now, we are going to describe our simple method. In the first step, we assign a value to each scenario, showing its effect on the changes in the portfolio value. To do this, let us assume the yield curve is represented by an n-dimensional vector $\mathbf{r} = (r_1, \cdots, r_n)$ of listed interest rates for standard maturities and the value of the portfolio, P, is defined as follows

$$P = f(r_1, r_2, \cdots, r_n). \tag{3}$$

Now, we approximate the change in the portfolio value by the following equation based on Taylor expansion,

$$\Delta P = \frac{\partial f(r_1, \cdots, r_n)}{\partial r_1}(\Delta r_1) + \frac{\partial f(r_1, \cdots, r_n)}{\partial r_2}(\Delta r_2) + \cdots + \frac{\partial f(r_1, \cdots, r_n)}{\partial r_n}(\Delta r_n). \tag{4}$$

Note that each scenario is a set of interest rate shocks at different maturities, represented by Δr_1 to Δr_n in Equation (4). In this way, we can form the equation for each scenario, and if the coefficients of Δr_1 to Δr_n are available, calculate the portfolio shock approximately. But Equation (4) has another interpretation as well. As can be seen, ΔP represents a weighted average of interest rate shocks at different maturities as is an *impact characteristic* of the scenarios, representing both the interest rates changes and the response to these changes. It is important to note that this approximation is used only to measure the impact of different scenarios and not to calculate VaR directly. Using this approximation leads to an error, which will be calculated and compared with the errors of other methods.

To find the coefficients of Δr_1 to Δr_n, there are two methods used in the GP method. In the first one, after selecting a few samples of the existing scenarios, we calculate the changes in the portfolio value. Then, we regress ΔP against Δr_1 to Δr_n and find the coefficients β_i, $i = 0, 1, \cdots, n$.

$$\Delta P = \beta_0 + \beta_1(\Delta r_1) + \beta_2(\Delta r_2) + \cdots + \beta_n(\Delta r_n), \tag{5}$$

the coefficient β_0 can be eliminated from this equation, because if the interest rate shocks at all maturities are zero, the portfolio shock is zero as well. This way the regression coefficients can be considered as the unknown coefficients of Δr_1 to Δr_n in Equation (4),

$$\frac{\partial f(r_1, \cdots, r_n)}{\partial r_j} = \beta_j \qquad j \in \{1, 2, \cdots, n\}. \tag{6}$$

Finally, the impact characteristics of each scenario is calculated as follows

$$\begin{cases} \mathcal{I} : \mathbb{R}^n \to \mathbb{R} \\ \mathcal{I}(\boldsymbol{\mathcal{S}}_i) = \langle \boldsymbol{\beta}, \boldsymbol{\mathcal{S}}_i \rangle, \end{cases} \tag{7}$$

Electronic copy available at: https://ssrn.com/abstract=3889826

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_n)$ and $\boldsymbol{\mathcal{S}} = (\Delta r_1, \cdots, \Delta r_n)$. In this approach one requires to recalculate the portfolio value for each scenario, so in order to reduce the computational load, in the second approach, we use the delta-gamma approximation to find the portfolio shock.

Choosing the base scenarios is the next step of the PwG method. Suppose the set of all scenarios is $\{\boldsymbol{\mathcal{S}}_i\}_{i=1}^{N}$. For selecting the set of base scenarios, $\{\boldsymbol{\mathcal{S}}_i^B\}_{i=1}^{n} \subset \{\boldsymbol{\mathcal{S}}_i\}_{i=1}^{N}$, we propose two approaches based upon their explanatory power for estimating PnL. Risk managers may use alternative selection criteria, which gives them more flexibility. Suppose that the yield curve shocks are available for the last 100 days, and we plan to extract 11 base scenarios. As a demonstration, Figure 1 shows the yield curve shocks over the 100 days on the left side and the one-dimensional scatter plot of the impact characteristics of each scenario on the right side.

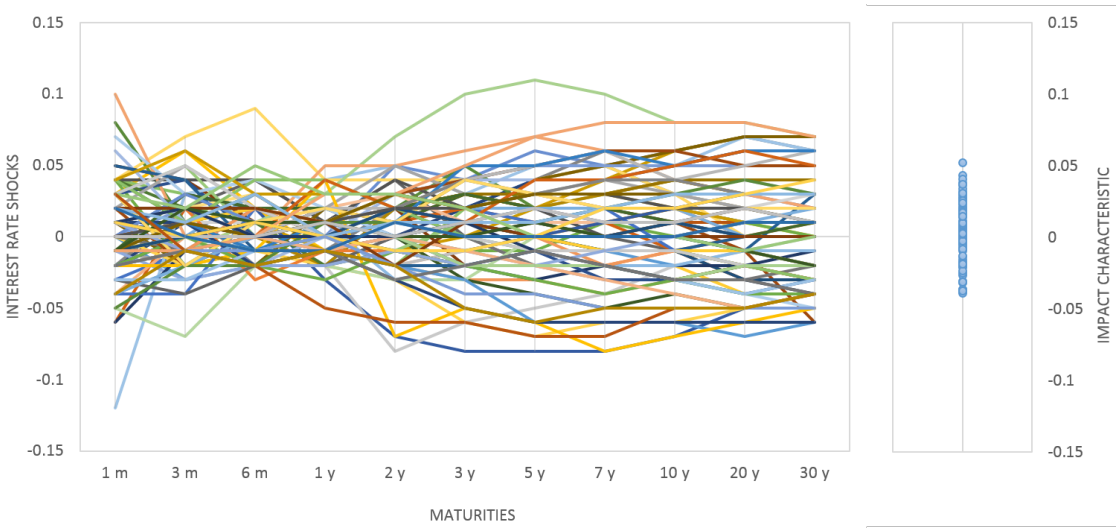
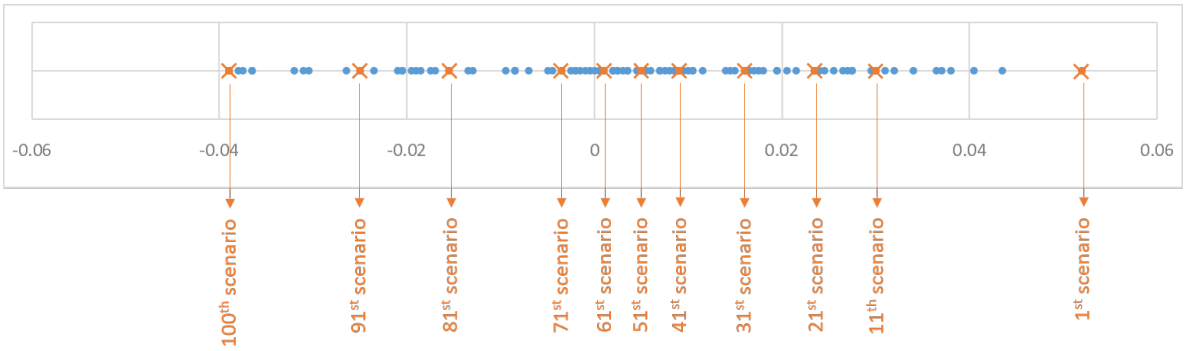


Figure 1: Yield curve shocks over the last 100 days and the impact characteristic of each scenario

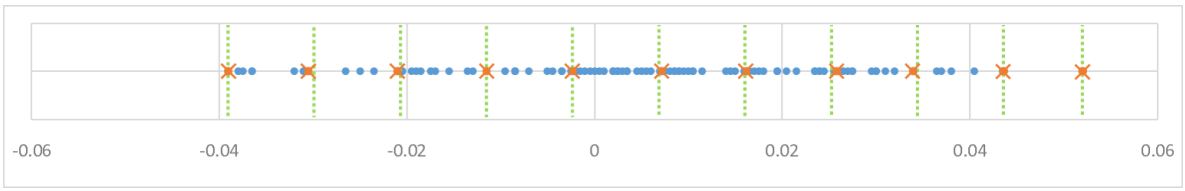
In the first approach, we sort the scenarios based on their impact characteristics, form several clusters with an equal number of scenarios, and select the first scenario of each cluster as a base scenario. In this example, we form 10 clusters associated with ten scenarios. Since more information about extreme scenarios leads to a more accurate approximation of the VaR, the last cluster's last scenario is also considered a base scenario. In this way, Figure 2(a) displays 11 base scenarios as cross signs. In the second approach, we plot the lines at equal distances from each other over the entire range of impact characteristics (green lines in Figure 2(b)). Note that we draw the first and last lines at the maximum and minimum of the impact characteristics, and the number of lines is equal to the number of base scenarios. We consider the closest scenarios to these lines as base scenarios (cross signs in Figure 2(b)). It seems that in this latter approach, due to the higher density of the base scenarios at the tails of the impact characteristics distribution, more information is captured from the extreme scenarios. On the other hand, the first approach provides higher accuracy in estimating the portfolio value for more frequent scenarios. We note that unlike many previous methods where changing the number of grid points is not easy due to their rigid structure, in our proposed method, adding one or

Electronic copy available at: https://ssrn.com/abstract=3889826

more scenarios to the base scenarios can be easily done.



(a) Choosing base scenarios according to the density of real scenarios



(b) Choosing base scenarios with equal intervals

Figure 2: Impact characteristics of different scenarios (cross signs are base scenarios)

Now, we calculate the changes in the portfolio value for base scenarios, and with the help of the GMC idea, estimate the portfolio shocks for non-base scenarios through interpolation or extrapolation. It is now possible to achieve distribution for the portfolio changes and calculate VaR and ES.

As mentioned earlier, the GP model employs PLS, while in our proposed PwG method, we use linear regression on actual market shock scenarios. Furthermore, in the GP model, the scenarios are projected onto an unknown space whose dimension does not necessarily have an economic interpretation, whereas, in the PwG method, we do not use the projection process. Thus, the implementation and interpretation of our method reflect its simplicity. Figure 3 and Figure 4 show the steps required to implement the GP and PwG methods, respectively.

2.3 Updating Input Scenarios

Nonparametric methods of VaR calculation do not consider any specific distribution for portfolio returns. One of these methods uses Monte Carlo simulation to obtain the required distribution, which requires an appropriate distribution for the variables under study. Choosing the right distribution is a major challenge for the financial variables with heavy tails and highly skewed empirical distributions. This challenge could be seen in the GP method when choosing distributions for each of the risk factors. Although there are studies such as Allen et al. [2020] and Gabrielsen et al. [2015] that have proposed some appropriate models for returns, HS remains to be a much simpler alternative in which there is no need to consider specific distributions for different variables. As we know, HS will lead to good

Electronic copy available at: https://ssrn.com/abstract=3889826

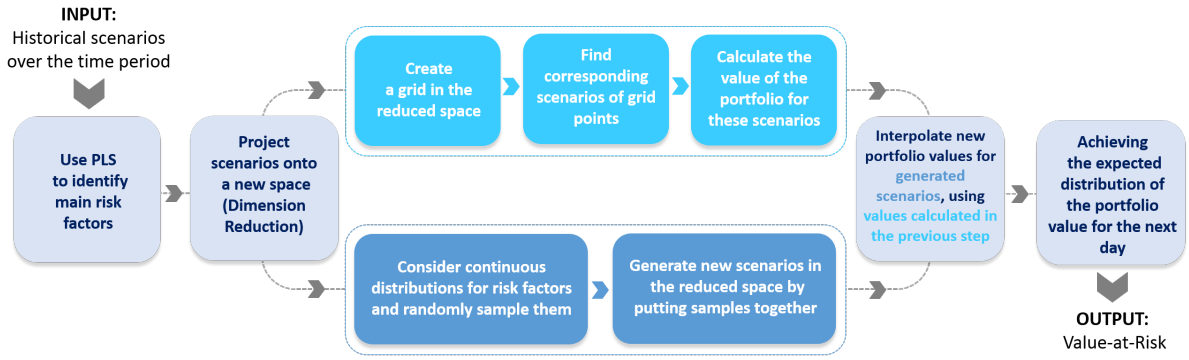


Figure 3: Steps required to implement the GP method

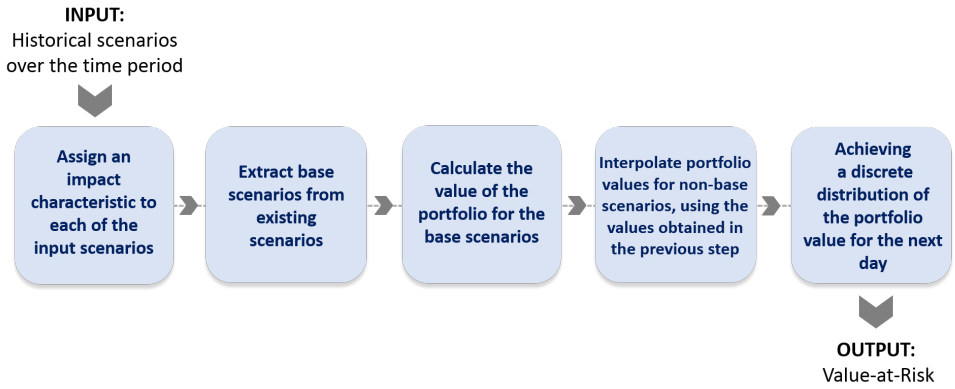


Figure 4: Steps required to implement the PwG method

results only if the historical observations are independently and identically distributed (i.i.d.). One of the factors that can violate this condition is a non constant volatility over time, resulting to an inconsistent estimate of VaR (Hendricks [1996] and McNeil and Frey [2000]). In the presence of the volatility clustering phenomenon, first noted by Mandelbrot [1997], the future is not the reflection of the past. Among the extensions proposed to overcome the non-i.i.d. feature of the data are the methods of Boudoukh et al. [1998] (BRW) and the Filtered Historical Simulation (FHS). Here, we briefly explain these methods and employ their ideas to improve the PwG method.

Boudoukh et al. [1998] assume that the weight given to the scenario $n+1$ days ago is λ times the weight given to the scenario n days ago,

$$w_{n+1} = \lambda w_n, \quad \text{where} \quad 0 < \lambda < 1.$$

Figure 5 shows the steps required to use the BRW probabilities in the PwG method. One of the criticisms of this approach is that a short-run sequence of abnormally large returns that have occurred

Electronic copy available at: https://ssrn.com/abstract=3889826

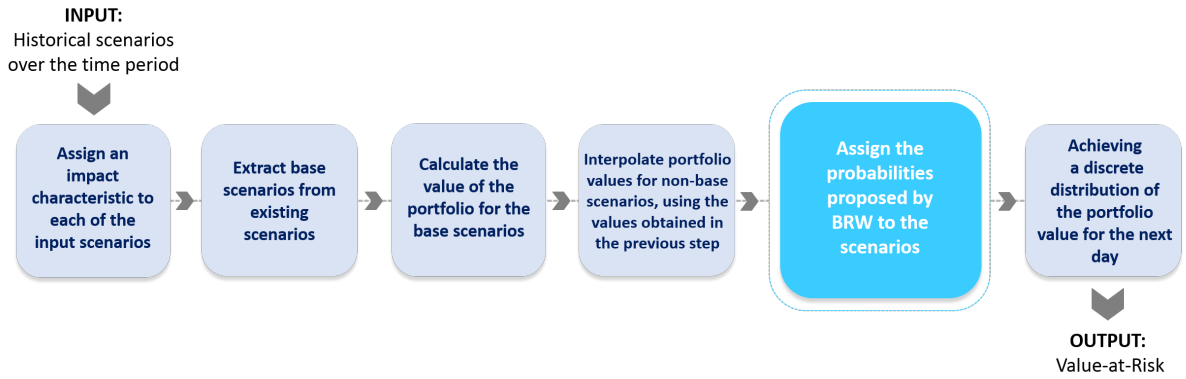


Figure 5: Steps required to use the BRW probabilities in the PwG method

in recent days can significantly skew the predicted distribution of the portfolio value to one side (Hull and White [1998]). We will examine the results of this method in Section 3.

Another way to treat the changes in the volatility of financial data is to update the volatility. Barone-Adesi et al. [1999] have combined the GARCH model for updating volatility and the HS method to calculate the VaR. To make returns *i.i.d.*, they use a moving average (MA) term in the conditional mean equation along with the GARCH processes to remove any serial correlation.

Another study that used filtered historical simulation to calculate the VaR for fixed-income portfolios is the paper of Kallur [2016]. This study updates the securities' volatilities separately and models the dependencies between them using a Student's t copula. This way, by the identified dependence structure and the expected scenarios for changes, a distribution for the expected changes in the portfolio value is simulated, and the VaR is easily estimated. Nevertheless, the PwG method uses changes in interest rates, not changes in the value of the instruments in the portfolio. This feature is an advantage of this method compared to Kallur [2016] because interest rate data is much richer to use.

Now back to our method, we update the variances of the observed interest rate shocks. If the interest rates at n maturities $T_1, T_2, \cdots, T_n$ for the last $m+1$ days are available, the interest rate shocks are defined by

$$\delta_i^k = r_i^k - r_{i-1}^k \qquad k \in \{1, 2, \cdots, n\},\ i \in \{1, 2, \cdots, m\}, \tag{8}$$

where r_i^k is the interest rate at the k^{th} maturity and i^{th} day. Therefore, for each maturity k we have a time series of interest rate shocks. It is now possible to consider the shocks obtained for each time series separately and update the variances for each of them. For this purpose, different models such as ARMA-GARCH, GARCH, EGARCH, EWMA, etc. can be used. For example, considering the ARMA-GARCH model, we have

$$\delta_i^k = \mu\delta_{i-1}^k + \theta\varepsilon_{i-1}^k + \varepsilon_i^k \qquad \varepsilon_i^k \sim N(0, h_i^k), \tag{9}$$

$$h_i^k = \omega + \alpha(\varepsilon_{i-1}^k)^2 + \beta h_{i-1}^k, \tag{10}$$

$$e_i^k = \frac{\varepsilon_i^k}{\sqrt{h_i^k}}, \tag{11}$$

Electronic copy available at: https://ssrn.com/abstract=3889826

$$\delta^k_{m+1,i} = e^k_i \sqrt{h^k_{m+1}}, \tag{12}$$

where ε^k_i is the random residual corresponding to the interest rate shock δ^k_i, and h^k_i is its variance. The parameters μ, θ, α, and β are the coefficients of AR(1), MA(1), ARCH and GARCH models, respectively, and ω is the constant of the GARCH(1,1). e^k_i is the standardized residual corresponding to ε^k_i, and $\delta^k_{m+1,i}$ is the i^{th} scenario of the set of interest rate shocks at the k^{th} maturity that may occur on the $m+1^{th}$ day, i.e., the next business day. This way, for a given i, we have formed a scenario for the yield curve change by considering the set $\{\delta^1_{m+1,i}, \delta^2_{m+1,i}, \cdots, \delta^n_{m+1,i}\}$, which corresponds to the i^{th} scenario of m different changes in the yield curve that have already occurred. By repeating the process for all $i \in \{1, 2, \cdots, m\}$ new scenarios are generated and will be used to calculate the VaR. Figure 6 shows the steps required to update scenarios and calculate the VaR using the GP and PwG methods.

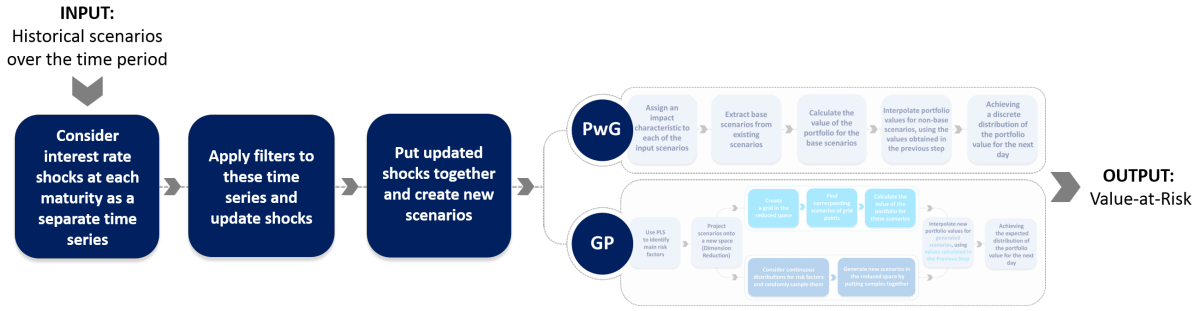


Figure 6: Steps required to update scenarios and calculate the VaR using the GP and PwG methods

We will show in Section 3 that this approach improves the accuracy of the results. It should be noted that because interest rate shocks at each maturity are considered as separate time series, changes in the dependence structure of interest rates at different maturities are not considered. Here, we are going to present a method to treat this problem. Let us consider the scenarios observed in the last m days. As mentioned earlier, each of these scenarios can be considered as a point in an n-dimensional space $\mathbb{R}^n$. Now, using the PCA method, these scenarios are mapped in a space with p orthogonal dimensions, that is

$$\begin{matrix} & \text{Original Space} & & & \text{Reduced Space} \\ \mathcal{M}: & \begin{pmatrix} \delta^1_1 & \cdots & \delta^n_1 \\ \vdots & \ddots & \vdots \\ \delta^1_m & \cdots & \delta^n_m \end{pmatrix}_{m\times n} & \xrightarrow{PCA} & \begin{pmatrix} \tilde{\delta}^1_1 & \cdots & \tilde{\delta}^p_1 \\ \vdots & \ddots & \vdots \\ \tilde{\delta}^1_m & \cdots & \tilde{\delta}^p_m \end{pmatrix}_{m\times p} \end{matrix} \tag{13}$$

In the original matrix, each row is assigned to a specific scenario that occurred in the last m days, the corresponding row in the second matrix shows the image of that scenario in the new reduced space. Each column of the original matrix is associated to one of the maturities for the interest rates, and the columns of the second matrix are associated to the orthogonal basis of the reduced space. As we know, each $\{\tilde{\delta}^k_i\}$, $k = 1, \cdots, p$ and $i = 1, \cdots, m$, is represented by a linear combination of the shocks

Electronic copy available at: https://ssrn.com/abstract=3889826

$\{\delta_i^1, \cdots, \delta_i^n\}$, so there exist $\beta_{i,j}^k$ such that

$$\tilde{\delta}_i^k = \sum_{j=1}^{n} \beta_j^k \delta_i^j. \tag{14}$$

Thus, with the same argument proposed earlier, the FHS method can be applied to images of the original scenarios in the reduced space. Because the identified risk factors are orthogonal, the correlations of these factors form the identity matrix. Therefore, without worrying about correlation changes, we can consider each column of the reduced space matrix as a separate time series and apply the variance updating process on it. Then, the updated shocks are placed next to each other, and a new scenario for the yield curve changes is generated. In this way, m new scenarios are generated in the reduced space for the next day. To identify their corresponding scenarios in the original space, we use an inverted PCA. If we consider $\{\delta^{*1}_{m+1,i}, \delta^{*2}_{m+1,i}, \cdots, \delta^{*p}_{m+1,i}\}$ as the i^{th} scenario of yield curve changes for the next day, $\{\delta^1_{m+1,i}, \delta^2_{m+1,i}, \cdots, \delta^n_{m+1,i}\}$ will be the corresponding original scenario, which is obtained by the following mapping

$$\begin{array}{ccccc} & \text{Reduced Space} & & \text{Original Space} & \\ \widetilde{\mathcal{M}}: & \begin{pmatrix} \delta^{*1}_{m+1,1} & \cdots & \delta^{*p}_{m+1,1} \\ \vdots & \ddots & \vdots \\ \delta^{*1}_{m+1,m} & \cdots & \delta^{*p}_{m+1,m} \end{pmatrix}_{m\times p} & \xrightarrow{PCA} & \begin{pmatrix} \delta^{1}_{m+1,1} & \cdots & \delta^{n}_{m+1,1} \\ \vdots & \ddots & \vdots \\ \delta^{1}_{m+1,m} & \cdots & \delta^{n}_{m+1,m} \end{pmatrix}_{m\times n} \end{array} \tag{15}$$

As the second advantage, this method reduces the computational cost of the volatility updating process from n to p. While changing the space of scenarios, some of the information will be lost, which may reduce the accuracy of the final estimate. This approach gives risk managers more flexibility to strike their desired balance between the imposed computational load and the accuracy of the final risk measure estimate. In Section 3, we will discuss the results obtained by this approach. Figure 7 shows the steps required to update scenarios by applying the filters to leading risk factors and calculating VaR using the GP and PwG methods.

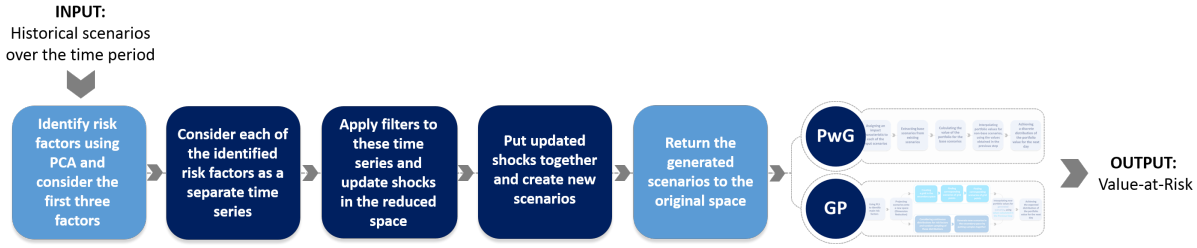


Figure 7: Steps required to update scenarios by applying filters to main risk factors and calculate the VaR using the GP and PwG methods

focusing on the reduced space, the scenario $\{\delta^{*1}_{m+1,i}, \delta^{*2}_{m+1,i}, \cdots, \delta^{*p}_{m+1,i}\}$ is derived from the volatility updating of the scenario $\{\delta^{*1}_i, \delta^{*2}_i, \cdots, \delta^{*p}_i\}$. Considering that the shocks of different columns are uncorrelated, we introduce two methods to generate new scenarios for the yield curve changes. In

Electronic copy available at: https://ssrn.com/abstract=3889826

the first method, we randomly select a shock from each column, and regardless of its time position, place it next to the other selected shocks. Thus, having information of the last m days and considering p main factors, we can generate m^p new scenarios.

In the second method, considering a specific distribution for each column and sampling from these distributions, new scenarios are generated for the yield curve changes. One of the challenges of using this method is choosing a suitable distribution for the risk factors. This distribution is commonly assumed to be the normal distribution. Due to the fat tails of empirical distributions, researchers also suggest the Student's t-distribution. Fiori and Iannotti [2007] pay attention to the skewness of empirical distributions and propose to generate a large number of new scenarios for the yield curve. Fiori and Iannotti [2007] derive the kernel densities and obtain the corresponding nonparametric probability distribution functions. They create a new scenario for the yield curve changes by sampling from the distribution of each factor separately. The generated scenarios are the images of the original scenarios. Applying the PC decomposition of the interest rate term structure and reconstructing the dependence structure, one reaches the original scenarios. In this way, many new scenarios are generated and used to estimate the VaR.

Finally, we emphasize that the two new proposed methods are not a subset of the general HS method. Since both methods produce new scenarios that have not happened before, they can be considered similar to the MC simulation. Given these explanations, it seems that the ideas in these methods can also be applied to the GP approach and make it more accurate.

3 Empirical Analysis

In this section, we present the results. We select one of the best methods for re-evaluating fixed income portfolios proposed by Gibson and Pritsker [2000] as our benchmark for comparison. We use a 500 days window of data to estimate the VaRs and compare them with the realized returns to judge the performance of different methods. The data consist of interest rates reported by the US Treasury Department for 11 different maturities from 2011 to the 24th of June 2020. Examining the data in other periods, we observe no substantial difference in the behavior of the portfolio; therefore, one can consider any other period as well. Furthermore, we simulate a fixed portfolio consisting of callable and putable bonds with different characteristics. We designed our portfolio to account for various interest rate changes at different maturities. In all comparisons, the computational load of the different methods is almost the same.

3.1 Tests and Models Used to Evaluate Different Methods

The first test is the unconditional coverage test, which evaluates the performance of a VaR calculation method based on the number of its violations compared to the VaR confidence level. The second test, the independence test (Christoffersen [1998]), is designed to answer the question of whether the violations are bunching or spreading independently over time. Finally, the conditional coverage test incorporates the previous ones. These tests examine the validity of each method separately and do not compare them with each other. Note that rejecting the null hypothesis indicates the failure of the method (Appendix B).

Another test proposed by Diebold and Mariano [2002] compares the accuracy of two different time series for forecasting a particular variable of interest. This test obtains the difference between the

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 1: Loss functions used in this paper

Regulator's Loss Functions		Firm's Loss Functions	
Lopez's quadratic	$\begin{cases} 1+(\mathrm{VaR}_t - r_t)^2 & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Mean Squared Error**	$(\mathrm{VaR}_t - r_t)^2$
Lineal	$\begin{cases} (\mathrm{VaR}_t - r_t) & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Sarma**	$\begin{cases} (\mathrm{VaR}_t - r_t)^2 & r_t < \mathrm{VaR}_t \\ -\alpha \times \mathrm{VaR}_t & r_t \geq \mathrm{VaR}_t \end{cases}$
Quadratic	$\begin{cases} (\mathrm{VaR}_t - r_t)^2 & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Caporin1**	$\lvert 1 - \lvert \frac{r_t}{\mathrm{VaR}_t} \rvert \rvert$
Caporin1	$\begin{cases} \left\lvert 1 - \lvert \frac{r_t}{\mathrm{VaR}_t} \rvert \right\rvert & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Caporin2**	$\frac{(\lvert r_t \rvert - \lvert \mathrm{VaR}_t \rvert)^2}{\lvert \mathrm{VaR}_t \rvert}$
Caporin2	$\begin{cases} \frac{(\lvert r_t \rvert - \lvert \mathrm{VaR}_t \rvert)^2}{\lvert \mathrm{VaR}_t \rvert} & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Caporin3**	$\lvert \mathrm{VaR}_t - r_t \rvert$
Caporin3	$\begin{cases} \lvert \mathrm{VaR}_t - r_t \rvert & r_t < \mathrm{VaR}_t \\ 0 & r_t \geq \mathrm{VaR}_t \end{cases}$	**Abad Benito Lopez**	$\begin{cases} (\mathrm{VaR}_t - r_t)^2 & r_t < \mathrm{VaR}_t \\ \alpha \times (r_t - \mathrm{VaR}_t) & r_t \geq \mathrm{VaR}_t \end{cases}$
Linear	$[\boldsymbol{\theta} - \mathbf{1}(r_t < \mathrm{VaR}_t)](r_t - \mathrm{VaR}_t)$		

Note: This table introduces several loss functions from the perspective of regulators and firms. The parameter θ in the linear regulatory loss function denotes the VaR quantile. In the loss functions of Sarma and Abad, Benito, Lopez, the cost of capital α is used to address the real value of capital that has been allocated without financial justification.

forecasted values and the actual values for each time series and applies a loss function to the errors. Then a statistical test is used to check whether the performance of one model is significantly better than the other model (Appendix B outlines the steps required to perform this test).

The selected superior model depends on the choice of the loss function. An underestimated VaR exposes the firm to significant risk; this is considered a serious problem by the regulators. On the other hand, an overestimated VaR forces the firm to store more regulatory capital than necessary and reduces its earning power. Thus, to evaluate the performance of different methods, one needs to determine the desired viewpoint. In the Diebold-Mariano test, the loss function determines this viewpoint, and changing the loss function can change the choice of the superior method. In this study, 13 different loss functions proposed by Lopez [1999], Caporin [2008], Sarma et al. [2003], Romero et al. [2014], etc. are used to consider both the regulators and the firms' concerns (Table 1).

The Ranking Model proposed by (Şener et al. [2012]) is another model which we will employ to

Electronic copy available at: https://ssrn.com/abstract=3889826

rank the VaR calculation methods. This model considers the concerns of both regulators and financial institutions, along with the number, size, and dependence of violations. In addition, it calculates a penalty for each day when the forecasted VaR is not exactly equal to the realized return, and according to the total penalty over the whole period, the superior method is identified. One of the interesting features of this model is that, unlike previous ones, it does not use statistical tests, and each VaR calculation method is scored based on its performance. In this way, the Ranking Model compares several different methods simultaneously. Appendix B describes the steps required to implement this model.

In addition to the above tests, using the Murphy diagrams (Murphy [1977]), we investigate our methods concerning the *consistency* of scoring functions. For predicting a real-valued outcome, forecasters try to present their best estimate. Without defining an explicit directive, each forecaster may report a different function of their estimated distribution as the final forecast (see Murphy and Daan [1985] and Engelberg et al. [2009]). Gneiting [2011] provides guidance for effective point forecasting. In this paper, we will skip the details of this guidance and focus only on the consistency of a scoring function. Here we explain the scoring functions and the statistical functions.

A representation for the performance criteria of point forecasting procedures is as follows

$$\bar{S} = \frac{1}{n} \sum_{i=1}^{n} S\left(x_i, y_i\right), \tag{16}$$

where y_i and x_i are the realized and forecasted values, respectively, and n is the number of observations. The scoring function $S\left(x_i, y_i\right)$ is defined as a penalty that is assigned to the forecaster due to the prediction of x_i instead of y_i. Finally, by taking average of these penalties for different forecasters, their performances are compared. The lower the value of the scoring function, the better the forecaster.

Another subject we need to review is statistical function. A statistical functional is a set-valued mapping from a class of probability distributions to a Euclidean space. Following Gneiting [2011], we assume that the functional T maps each distribution $F \in \mathcal{F}$ to a subset $T(F)$ of the domain $D \subseteq \mathbb{R}$ (Equation (17)). The subset $T(F)$ is often single-valued, such as a specific quantile or expectile.

$$T : \mathcal{F} \to \mathcal{P}(D), \quad F \mapsto T(F) \subseteq D. \tag{17}$$

One says that for the functional T, the scoring function S is consistent relative to the class $\mathcal{F}$ if

$$\mathbb{E}_F[\ \mathrm{S}(t, Y)] \leq \mathbb{E}_F[\ \mathrm{S}(x, Y)], \tag{18}$$

for all probability distributions $F \in \mathcal{F}$, all $t \in T(F)$, and all $x \in D$.

Gneiting [2011] claims that meaningful assessment of point estimators requires using consistent scoring functions. This highlights the importance of understanding the close relationship between appropriate functional selection and proper evaluation procedure. Many studies such as Savage [1971], Thomson [1978], Gneiting [2011], Roccioletti [2015], Ehm et al. [2016], etc. have examined the general form of consistent scoring functions for different functionals. Since VaR is a quantile of the PnL distribution, we focus on consistent scoring functions for quantiles as functional. The scoring function S_α^Q is consistent for the quantile functional at a level $\alpha \in (0, 1)$ relative to the class $\mathcal{F}_0$ if and only if it is of the form

$$\mathrm{S}_\alpha^{\mathrm{Q}}(x, y) = (\mathbf{1}(y < x) - \alpha)(g(x) - g(y)), \tag{19}$$

Electronic copy available at: https://ssrn.com/abstract=3889826

where g is a non-decreasing function. According to Ehm et al. [2016], each scoring function which satisfies Equation (19) admits a representation of the form

$$\mathrm{S}^{\mathrm{Q}}_{\alpha}(x,y) = \int_{-\infty}^{+\infty} \mathrm{S}^{\mathrm{Q}}_{\alpha,\theta}(x,y)\mathrm{d}H(\theta) \quad (x,y \in \mathbb{R}), \tag{20}$$

where

$$\mathrm{S}^{\mathrm{Q}}_{\alpha,\theta}(x,y) = (\mathbf{1}(y<x)-\alpha)(\mathbf{1}(\theta<x)-\mathbf{1}(\theta<y)), \tag{21}$$

and H is nonnegative, unique and satisfies $\mathrm{d}H(\theta) = \mathrm{d}g(\theta)$ for $\theta \in \mathbb{R}$, g is the function in Equation (19). Furthermore, we have $H(x)-H(y) = \mathrm{S}(x,y)/(1-\alpha)$ for $x > y$. After applying several simplifications to evaluate two different empirical forecasters, Ehm et al. [2016] argue that the first forecaster performed better than the second one if

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{S}^{\mathrm{Q}}_{\alpha,\theta}\left(x_{1i},y_i\right) \leq \frac{1}{n}\sum_{i=1}^{n} \mathrm{S}^{\mathrm{Q}}_{\alpha,\theta}\left(x_{2i},y_i\right), \tag{22}$$

for $\theta \in \{x_{11}, x_{12}, y_1, \ldots, x_{n1}, x_{n2}, y_n\}$, where y_i is the realized value of the i^{th} observation, and x_{ji} is the j^{th} forecaster's point prediction of y_i. Finally, to compare the two forecasters, Ehm et al. [2016] use the Murphy diagram, in which the penalties calculated in Equation (22) are plotted against θ. The smaller the area under the curve, the better the performance of the method. We will use the Murphy diagram to compare different VaR estimation methods in this paper.

3.2 Comparison of the PwG and GP

In this section we present the results of the PwG and compare them with the GP ones. Figure 8 shows the estimated 99%-VaRs using two mentioned methods along with the daily realized returns over a period of 1800 days.

Examining the performance of these methods at different confidence levels may provide more information. Therefore, Figure 9(a) and Figure 9(b) are displayed with 95% and 90% confidence level, respectively.

The fluctuations of the VaR predicted by the GP are much greater than that of the PwG. However, as the confidence level for calculating VaR decreases, the observed difference also declines. If these fluctuations were due to the greater accuracy of the GP method, they could show the strength of this method. However, by experimenting, we will show that these fluctuations result from a random sampling process. Interestingly, there is a similar step in the PwG method, but it does not lead to large fluctuations. For further investigation, we run the GP and PwG methods several times over ten days and represent the diagram in Figure 10. The PwG method produces almost similar results so that the paths related to this method are nearly superimposed. On the other hand, the GP results vary for different iterations, but the average values of different iterations fluctuate very little. Thus, one can argue that these observations occurred as a result of a random sampling process. The GP method performs random sampling in two different steps of the GP method. One of them is when we create new scenarios from the main risk factor distributions. Since this step generates many scenarios, it cannot be the cause of the fluctuations. The other step is selecting several scenarios to implement the PLS and identifying the reduced space. To show the effect of this step on the observed fluctuations, we repeat the above experiment in three cases using 70, 200, and all the 500 scenarios. As shown in Figure

Electronic copy available at: https://ssrn.com/abstract=3889826

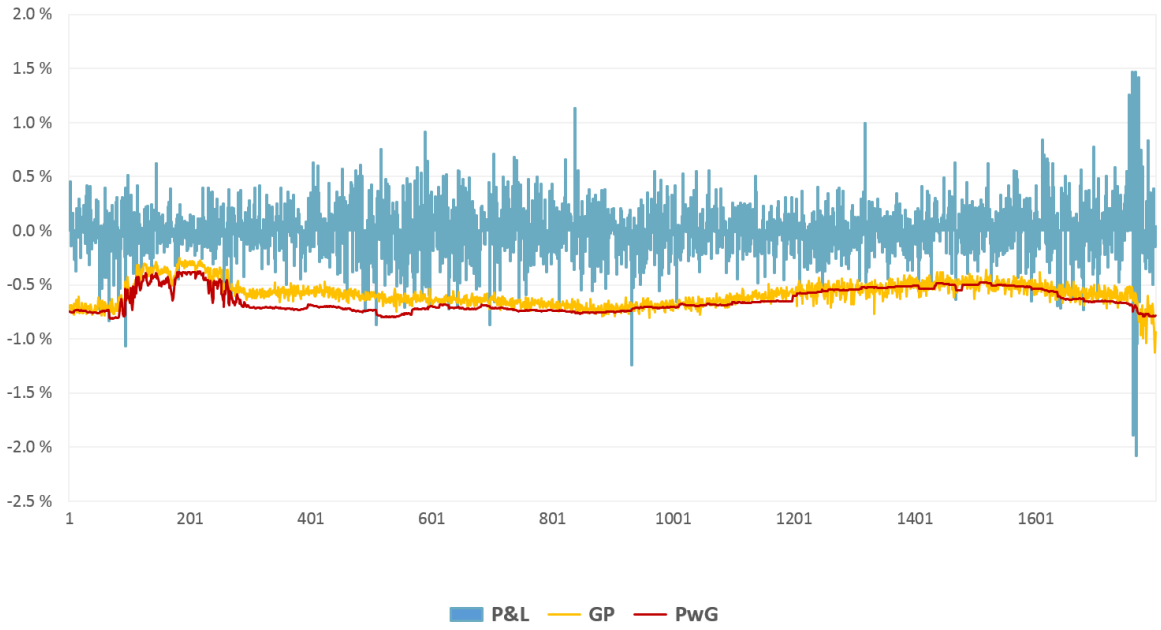


Figure 8: Realized returns and estimated 99% VaRs by the PwG and GP methods, Y-axis: Return and VaR, X-axis: Observation

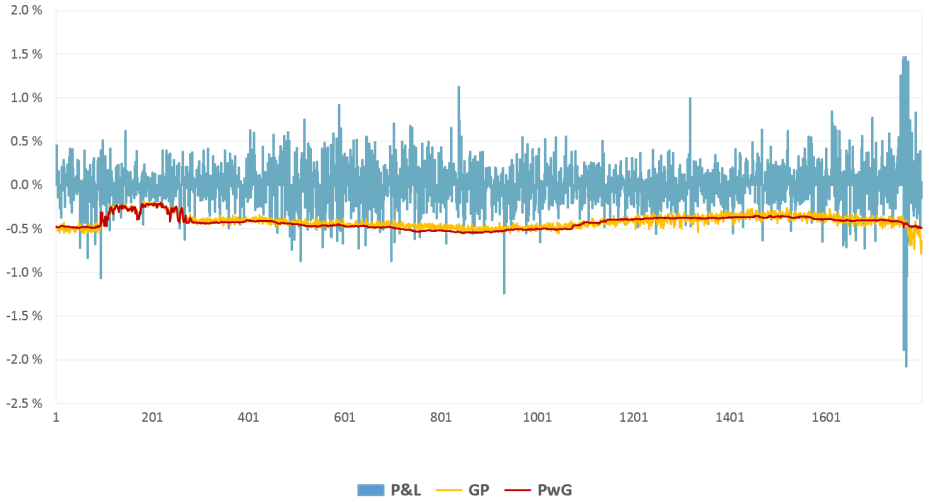
11, when we use a higher number of scenarios, the VaR fluctuations decrease, showing that this step causes the fluctuations. PwG method performs random sampling to identify impact characteristics, but it does not cause significant fluctuations. To summarize, the absence of large fluctuations in the final results is one of the advantages of the PwG method over the GP method.

Now, we are going to test the performance of the GP and PwG methods. In the first part, we examine the Murphy diagram by the use of consistent scoring functions. (Figure 12). As can be seen, the area under the PwG diagram is lower, which indicates the superiority of this method.

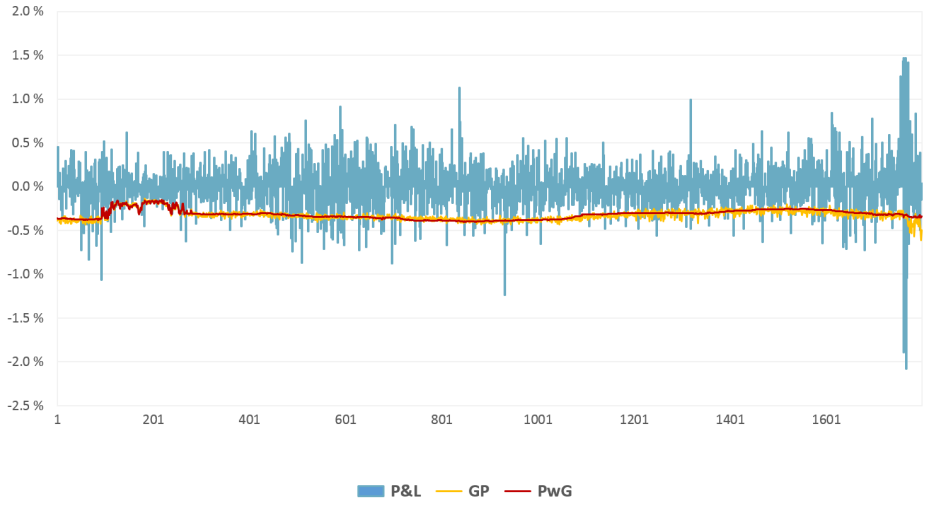
We now use the unconditional coverage, independence, and conditional coverage tests to evaluate the methods (Table 2). At 95% and 90% confidence levels, there is not much difference between the two methods, but at 99% confidence level, the results show the superiority of the PwG method. The null hypothesis for the GP method is rejected in the conditional coverage test, whereas this test does not reject the null hypothesis for the PwG method.

In this part, we want to directly compare the performance of the GP and PwG methods using the Diebold-Mariano test. The test results are reported in Table 3. If statistics in this test are significantly greater than zero, the first method is relatively superior. We see no significant difference in these methods at the 95% and 90% confidence levels. However, the results obtained at a confidence level of 99% show that we will achieve different results by considering different loss functions. As can be seen, the PwG method is more desirable in the regulator's eyes, and in contrast, the performance

Electronic copy available at: https://ssrn.com/abstract=3889826



(a) Confidence level: 95%



(b) Confidence level: 90%

Figure 9: Realized returns and estimated VaRs by the PwG and GP methods, Y-axis: Return and VaR, X-axis: Observation

of the GP method is more appropriate from the perspective of institutions.

We use the Ranking Model to compare the performance of the PwG and GP methods. As can be

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 2: The unconditional coverage, independence, and conditional coverage tests results for the GP and PwG methods

Confidence Level	Method	Conditional Coverage Test		Independence Test		Unconditional Coverage Test		Violation Rate
		Stat.	P-Value	Stat.	P-Value	Stat.	P-Value	
99%	**GP**	9.88	0.0071	0.23	0.6256	9.65	0.0019	0.0180
	PwG	1.95	0.37	1.23	0.2657	0.71	0.3982	0.0120
95%	**GP**	2.35	0.3084	0.96	0.3264	1.38	0.2385	0.0561
	PwG	0.58	0.7463	0.15	0.6904	0.4267	0.5136	0.0534
90%	**GP**	0.45	0.7952	0.3671	0.5446	0.09	0.7644	0.0978
	PwG	0.05	0.9755	0.04	0.84	0.009	0.9249	0.1006

Note: In this table, the performance of the GP and PwG methods in VaR estimation with 99%, 95%, and 90% confidence levels is evaluated using unconditional coverage, independence, and conditional coverage tests.

Table 3: The Diebold-Mariano test results for comparison of the GP and PwG methods

Confidence Level:	**99%**		**95%**		**90%**	
Method 1:	**PwG**		**PwG**		**PwG**	
Method 2:	**GP**		**GP**		**GP**	
Loss Functions	**Stat.**	**P-Value**	**Stat.**	**P-Value**	**Stat.**	**P-Value**
R-Lopez's quadratic	3.326	0.000	0.832	0.203	-1.944	0.026
R-Lineal	4.419	0.000	3.419	0.000	-2.291	0.011
R-Quadratic	2.244	0.012	2.734	0.003	-1.103	0.135
R-Caporin 1	4.080	0.000	3.756	0.000	-1.788	0.037
R-Caporin 2	1.933	0.027	2.251	0.012	-0.134	0.447
R-Caporin 3	4.419	0.000	3.419	0.000	-2.291	0.011
R-Linear	-5.457	0.000	1.618	0.053	-1.203	0.115
F-Mean Squared Error	-28.685	0.000	-6.227	0.000	5.991	0.000
F-Sarma	-6.555	0.000	-0.227	0.410	1.704	0.044
F-Caporin 1	-16.919	0.000	-1.009	0.157	1.251	0.106
F-Caporin 2	-25.445	0.000	-2.813	0.002	1.768	0.039
F-Caporin 3	-35.086	0.000	-7.976	0.000	6.533	0.000
F-Abad-Benito-Lopez	-6.041	0.000	0.003	0.499	1.156	0.124

Note: In this table, using Diebold-Mariano test, the performances of the GP and PwG methods in estimating VaR at 99%, 95%, and 90% confidence levels are compared. For this purpose, different loss functions are used. The prefixes R and F represent the viewpoints of regulators and firms, respectively. In this table, the positive (negative) statistics show the better performance of the first (second) method.

Electronic copy available at: https://ssrn.com/abstract=3889826

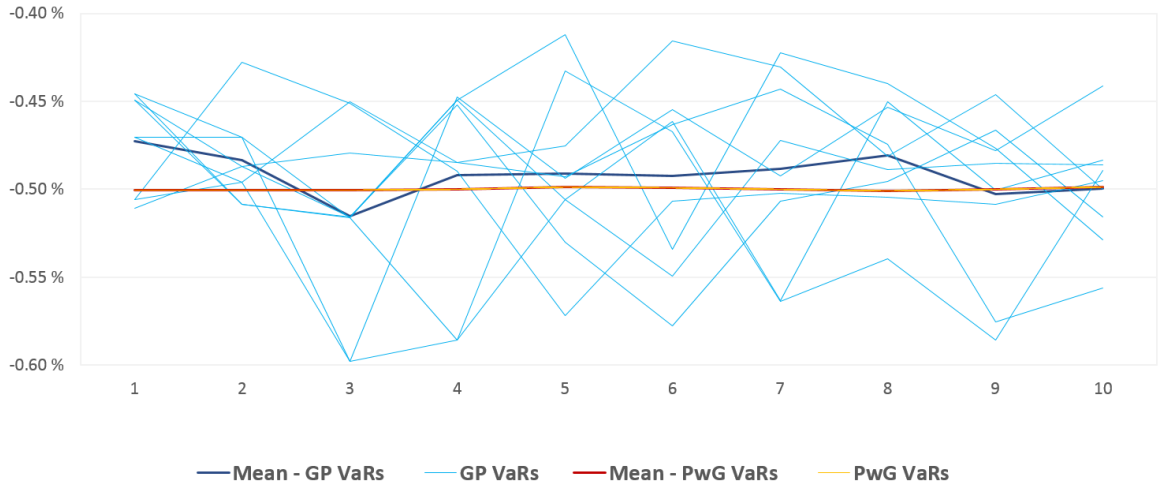


Figure 10: Fluctuation of VaRs estimated by the GP and PwG methods, Y-axis: VaR, X-axis: Observation

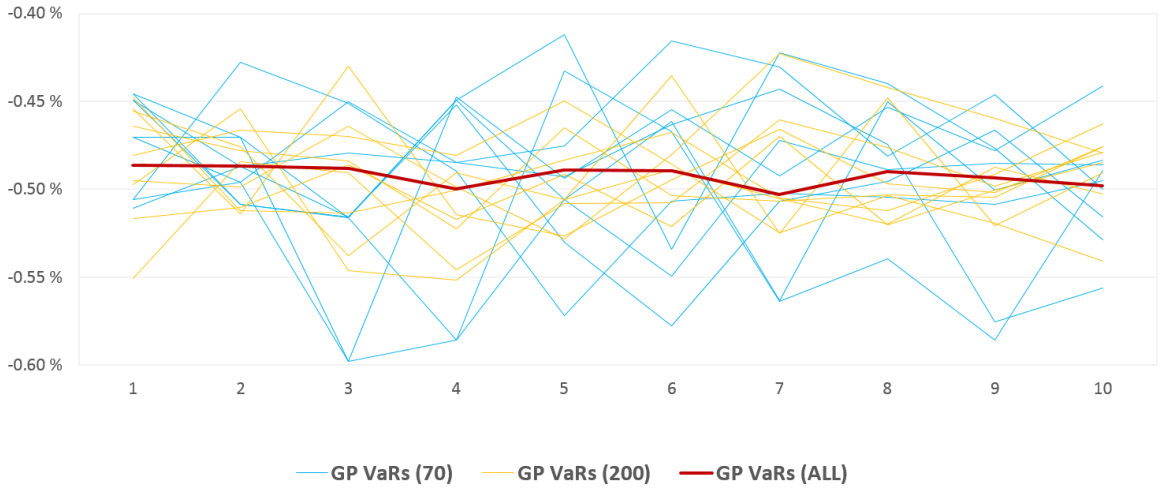


Figure 11: Fluctuation of VaRs estimated by the GP method using 70, 200 and all scenarios to identify the reduced space, Y-axis: VaR, X-axis: Observation

seen in Table 4, this model does not report a significant difference in the performance of these methods. Of course, the result is not unexpected because we saw that each method works better in one of the perspectives in the previous test. Given that the Ranking Model takes both perspectives into account, the strengths and weaknesses of each method may offset each other when considered together.

Electronic copy available at: https://ssrn.com/abstract=3889826

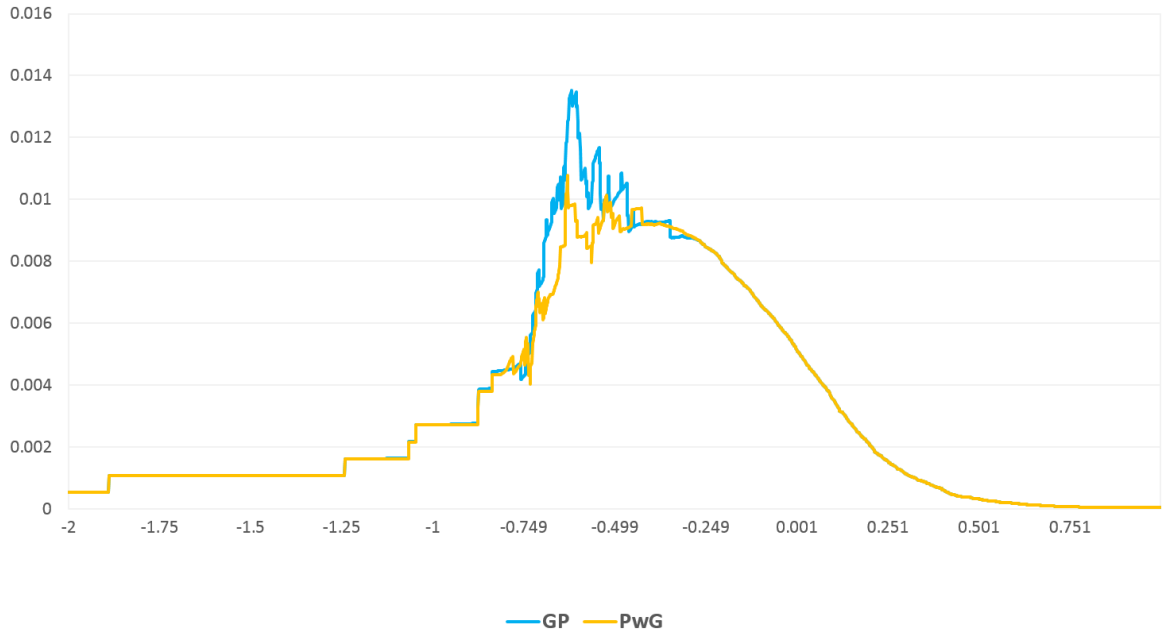


Figure 12: Murphy diagrams for the PwG and GP methods, Y-axis: Expected Score, X-axis: Parameter θ

3.3 Comparison of Unfiltered and Filtered Modes of the PwG and GP Methods

We review the dynamic variance methods in this subsection, like GARCH, ARMA-GARCH, and EWMA filters. Comparing these methods before and after the filter application, we will examine the effectiveness of this technique. Figures 13(a) and 13(b), show the daily realized returns over a period of 1800 days and the estimated 99%-VaRs using various methods. At a glance, we find that unlike the unfiltered GP and PwG methods, VaRs estimated by filters respond properly to variance changes. For example, at the end of the period, large returns are observed as a cluster of variances. As can be seen, applying filters reduces the difference between the realized losses and the estimated VaRs. We can apply filters to interest rates with different maturities and investigate their effect on portfolio returns.

Table 5 show the results of unconditional coverage, independence, and conditional coverage tests for unfiltered and filtered GP and PwG methods. Given the statistics and rates of violations, it seems that using filters has improved the performance of both methods. The results show that for the PwG method, the improvement brought upon by the GARCH filter is more significant than the EWMA one. In contrast, for the GP method, the EWMA filter performs better. A possible reason for this observation could be the way these methods use scenarios. The PwG method directly uses scenarios after the updating process, but the GP method projects the input scenarios into another space and

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 4: The Ranking Model results for the GP and PwG methods

Confidence Level	Method	Penalty	Percentage	Rank
99%	**GP**	2.28×10^{-4}	51.2	2
	PwG	2.17×10^{-4}	48.8	1
95%	**GP**	2.43×10^{-4}	50.81	2
	PwG	2.35×10^{-4}	49.19	1
90%	**GP**	3.63×10^{-4}	48.5	1
	PwG	3.86×10^{-4}	51.5	2

Note: In this table, using the Ranking Model, the performances of the GP and PwG methods in estimating VaR at 99%, 95%, and 90% confidence levels are compared. The smaller the penalty, the better the performance of the method.

generates new ones. Thus, the filter that performs better for the original scenarios may not be effective for the generated ones.

Looking at Figures 13(a) and 13(b), we observe that there is a difference between the performance of unfiltered and filtered methods at the end of the period, in which the volatility of the portfolio value is high. In Figure 14 an EWMA process is represented for the volatility of the portfolio. The difference between filtered and unfiltered methods is obvious in the mentioned sub-period of time. Here we will design an experiment to understand how filters perform in volatile and normal market environments. First, we consider 30% of the period with high volatility as the first category and the rest of the period as the second one. Then, we perform the unconditional coverage test separately for each of these categories. The test results are shown in Table 6. The independence and conditional coverage tests are not useful because of the missing sequence of days. An interesting point to note is that the unfiltered methods in the category that include days with low volatility are not significantly different from the filtered methods. In the other category, however, the performance of the methods improves considerably after the filter is applied. We observe no improvements if we focus on 30% of the days with low volatility and rerun the test (Table 7). In conclusion, using filters will be more effective in situations where we are facing high volatilities.

Table 8 reports the results of the Ranking Model. This model confirms the improvement of the unfiltered methods by adding filters and introduces the GARCH filter as the best candidate. The application of filters increases the accuracy of the GP method more than the PwG. As shown in the conditional coverage test results, the PwG method's performance was acceptable before the application of the filter, but for the GP, it was not the case. This difference may be a reason why we observe the greater impact of filtering on the GP method. Earlier, we tested the hypothesis that filters would be more effective over periods of high volatility. Here, we intend to test this hypothesis with a new experiment. For this purpose, we first consider the last 300 days, which show very high volatility according to Figure 14, then the Ranking Model is used to compare the performance of filtered and unfiltered methods in this period (Table 9). Comparing Tables 8 and 9, it can be argued that filters are much more effective on days with higher volatility. Thus, the results of this experiment, like the previous one, confirm the validity of the hypothesis.

Assigning the probabilities proposed by BRW to the input scenarios is another approach we have used to consider the effect of volatility changes. Since this approach is known as an extension of the

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 5: The unconditional coverage, independence, and conditional coverage tests results for unfiltered and filtered modes of the GP and PwG

Method	Conditional Coverage Test		Independence Test		Unconditional Coverage Test		Violation Rate
	Stat.	P-Value	Stat.	P-Value	Stat.	P-Value	
PwG	1.95	0.37	1.23	0.2657	0.71	0.3982	0.0120
PwG-GARCH	0.42	0.8080	0.39	0.5277	0027	0.8683	0.0104
PwG-ARMA GARCH	0.59	0.7426	0.44	0.5060	0.15	0.6922	0.0109
PwG-EWMA	2.64	0.26	0.158	0.69	2.48	0.1148	0.0066
GP	9.88	0.0071	0.23	0.6256	9.65	0.0019	0.0180
GP-GARCH	5.35	0.0686	0.87	0.3508	4.47	0.0343	0.0153
GP-ARMA GARCH	6.31	0.0426	0.93	0.3337	5.37	0.0204	0.0158
GP-EWMA	0.87	0.6457	0.48	0.4849	0.38	0.53	0.0115

Note: In this table, the performance of the GP and PwG methods, before and after applying filters, in VaR estimation with 99% confidence level is evaluated using unconditional coverage, independence, and conditional coverage tests.

Table 6: The unconditional coverage test results for unfiltered and filtered modes of the GP and PwG with high (30% of all scenarios) and low volatilities

Method	Unconditional Coverage Test			
	Low Volatility (70% of all scenarios)		High Volatility (30% of all scenarios)	
	Stat.	P-Value	Stat.	P-Value
PwG	0.26	0.6044	4.34	0.0371
PwG-GARCH	0.05	0.8203	0.39	0.5315
GP	0.36	0.5474	18.06	0.0000
GP-GARCH	2.64	0.1042	1.91	0.16

Note: We calculate the volatility of the portfolio value, considering the 30% of the total days with the highest volatility as one group and the rest as the other group. The performance of the filtered and unfiltered modes of the GP and PwG methods in 99%-VaR estimation is evaluated using unconditional coverage test.

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 7: The unconditional coverage test results for unfiltered and filtered modes of the GP and PwG with high and low volatilities (30% of all scenarios)

Method	Unconditional Coverage Test			
	Low Volatility (30% of all scenarios)		High Volatility (70% of all scenarios)	
	Stat.	P-Value	Stat.	P-Value
PwG	0.44	0.5045	0.89	0.1687
PwG-GARCH	0.44	0.5045	0.36	0.5474
GP	0.048	0.8260	12.06	0.0005
GP-GARCH	0.048	0.8260	5.49	0.0190

Note: We calculate the volatility of the portfolio value, considering the 30% of the total days with the lowest volatility as one group and the rest as the other group. The performance of the filtered and unfiltered modes of the GP and PwG methods in 99%-VaR estimation is evaluated using unconditional coverage test.

Table 8: The Ranking Model results for unfiltered and filtered modes of the GP and PwG methods

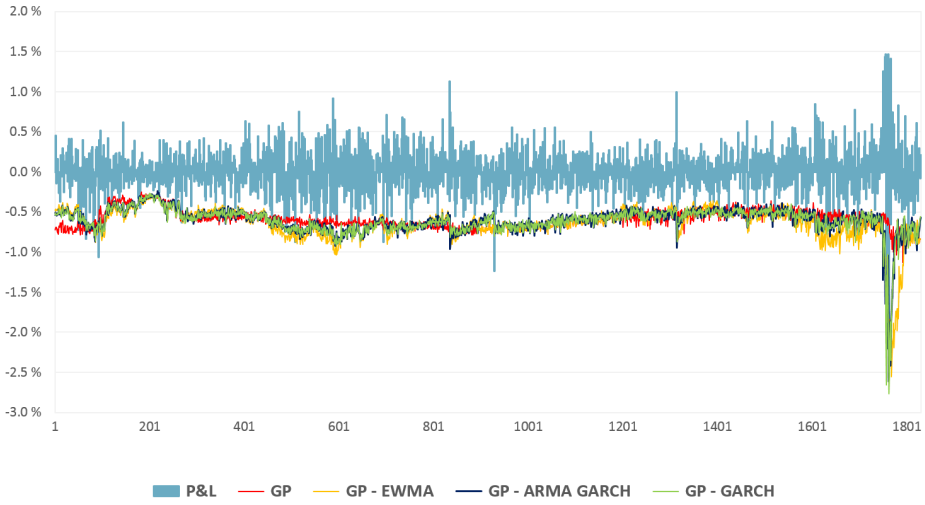
Method	Penalty	Percentage	Rank
PwG	5.32×10^{-5}	27.26	4
PwG-GARCH	4.48×10^{-5}	22.95	1
PwG-ARMA GARCH	4.76×10^{-5}	24.37	2
PwG-EWMA	4.96×10^{-5}	25.42	3
GP	6.86×10^{-5}	33.32	4
GP-GARCH	4.38×10^{-5}	21.28	1
GP-ARMA GARCH	4.68×10^{-5}	22.74	3
GP-EWMA	4.66×10^{-5}	22.67	2

Note: In this table, using the Ranking Model, the unfiltered and filtered modes of the GP and PwG methods in estimating 99%-VaR are compared. The smaller the penalty, the better the performance of the method.

Electronic copy available at: https://ssrn.com/abstract=3889826



(a) PwG method



(b) GP method

Figure 13: Realized returns and estimated VaRs(99%) by unfiltered and filtered modes of the GP and PwG methods, Y-axis: Return and VaR, X-axis: Observation

HS approach, only the PwG method can implement it. Figures 15(a) and 15(b) show the results for

Electronic copy available at: https://ssrn.com/abstract=3889826



Figure 14: Daily volatility of the portfolio value, Y-axis: Volatility, X-axis: Observation

99% and 95% VaRs. The BRW technique shows a piece-wise constant behavior over time, whereas other methods do not show this feature.

Table 10 shows the results of evaluating the methods using unconditional coverage, independence, and conditional coverage tests. As can be seen, applying this extension improves results for 95% confidence level, while using it at 99% confidence level leads to inferior performance. As mentioned in Section 2, a significant return that has occurred in recent days can skew the distribution of portfolio value changes in this technique. It is clear that as the level of confidence increases, this effect becomes more visible, and the accuracy of the final estimate decreases. This effect justifies the poor performance of the BRW technique at 99% confidence levels.

We use the Ranking Model to compare the methods (Table 11). Based on the results, the PwG-BRW method works better than the simple PwG method but is not more accurate than the PwG-GARCH method.

We have already found that applying filters on interest rates with different maturities improves the PwG and GP methods. Previously, we proposed a solution to prevent the destructive impact of correlation changes over time, in which instead of applying filters to interest rates, we applied them to the main risk factors. Furthermore, this method reduces the computational load. To examine this idea, after finding the three main risk factors, using PCA, we will apply the GARCH filter on them (the PwG-PC and GP-PC methods) and compare the outcomes with the previously obtained results. Figures 16(a) and 16(b) show the VaRs estimated using the above methods.

By choosing the three main risk factors and ignoring others, one would lose some of the information in the scenarios. Therefore, the primary purpose of the comparisons is to answer whether the PwG-PC and GP-PC methods can have a performance close to the filtered mode of PwG and GP approaches. Table 12 shows the unconditional coverage, independence, and conditional coverage tests to measure the performance of these methods. As can be seen, the results are auspicious and show success in all tests.

Table 13 reports the results of the Ranking Model. Due to the approximate similarity of the

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 9: The Ranking Model results for unfiltered and filtered modes of the GP and PwG methods for the last 300 days of the time period

Method	Penalty	Percentage	Rank
PwG	8.28×10^{-5}	36.14	4
PwG-GARCH	3.90×10^{-5}	17.03	1
PwG-ARMA GARCH	5.49×10^{-5}	23.94	2
PwG-EWMA	5.24×10^{-5}	22.89	3
GP	1.73×10^{-4}	54.31	4
GP-GARCH	3.85×10^{-5}	12.08	1
GP-ARMA GARCH	5.46×10^{-5}	17.13	3
GP-EWMA	5.25×10^{-5}	16.48	2

Note: In this table, using the ranking model, the unfiltered and filtered modes for the GP and PwG methods in estimating 99%-VaR are compared over a 300-day period that shows the highest volatility. The smaller the penalty, the better the performance of the method.

Table 10: The unconditional coverage, independence, and conditional coverage tests results for the PwG method with and without GARCH filter and BRW probabilities

Confidence Level	Method	Conditional Coverage Test		Independence Test		Unconditional Coverage Test		Violation Rate
		Stat.	P-Value	Stat.	P-Value	Stat.	P-Value	
95%	**PwG**	0.74	0.6886	0.59	0.4401	0.15	0.6984	0.0475
	PwG-BRW	0.39	0.8211	0.10	0.7458	0.28	0.5907	0.0534
99%	**PwG**	2.82	0.2433	1.59	0.2068	1.23	0.2669	0.0133
	PwG-BRW	6.35	0.0418	0.74	0.3867	5.6	0.0170	0.0175

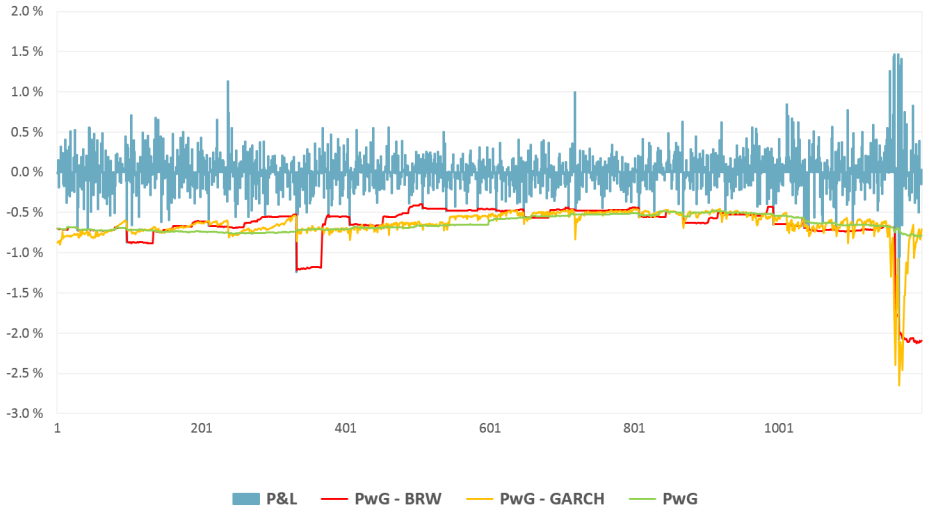
Note: In this table, the performance of the PwG method, with and without the use of BRW probabilities, in VaR estimation with 95% and 99% confidence levels is evaluated using unconditional coverage, independence, and conditional coverage tests.

Table 11: The Ranking Model results for the PwG method with and without GARCH filter and BRW probabilities

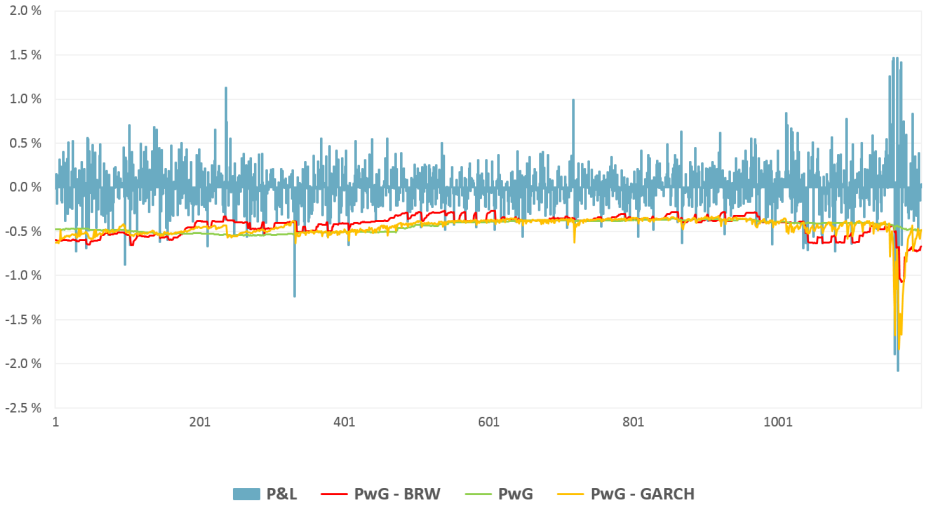
Confidence Level	Method	Penalty	Percentage	Rank
95%	**PwG**	2.35×10^{-4}	37.84	3
	PwG-GARCH	1.75×10^{-4}	28.08	1
	PwG-BRW	2.12×10^{-4}	34.08	2
99%	**PwG**	5.66×10^{-5}	37.83	3
	PwG-GARCH	4.20×10^{-5}	28.10	1
	PwG-BRW	5.10×10^{-5}	34.07	2

Note: In this table, using the Ranking Model, the performances of the PwG method before and after application of GARCH filter and allocation of BRW probabilities are compared in estimating the 99%-VaR. The smaller the penalty, the better the performance of the method.

Electronic copy available at: https://ssrn.com/abstract=3889826



(a) Confidence level: 99%

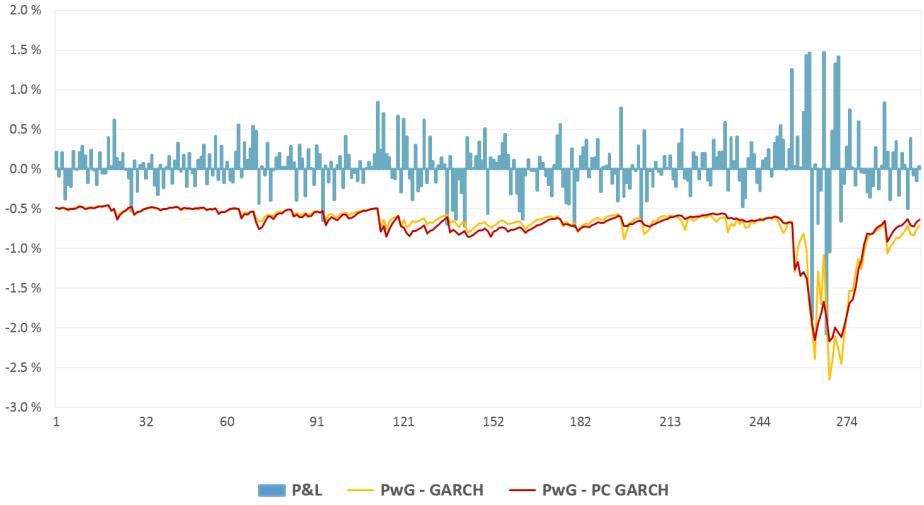


(b) Confidence level: 95%

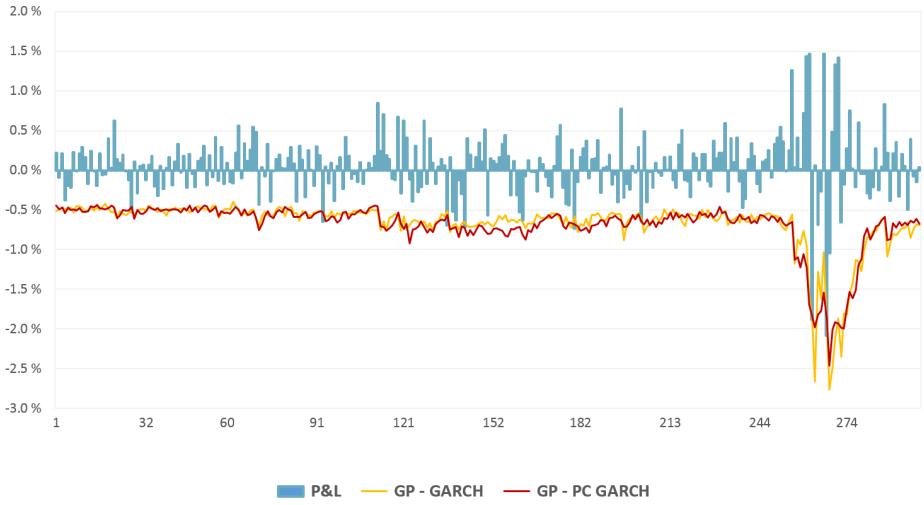
Figure 15: Realized returns and estimated VaRs by the PwG method before and after application of GARCH filter and allocation of BRW probabilities, Y-axis: Return and VaR, X-axis: Observation

results, this model cannot determine the superior method with certainty. In conclusion, the results of

Electronic copy available at: https://ssrn.com/abstract=3889826



(a) PwG and PwG-PC methods



(b) GP and GP-PC methods

Figure 16: Realized returns and estimated VaRs(99%) by the PwG-PC and GP-PC methods, Y-axis: Return and VaR, X-axis: Observation

all tests and models encourage us to use the PwG-PC and GP-PC methods because they reduce the

Electronic copy available at: https://ssrn.com/abstract=3889826

Table 12: The unconditional coverage, independence, and conditional coverage tests results for the GP, GP-PC, PwG, and PwG-PC methods

Method	Conditional Coverage Test		Independence Test		Unconditional Coverage Test		Violation Rate
	Stat.	P-Value	Stat.	P-Value	Stat.	P-Value	
PwG-GARCH	2.63	0.2677	0.24	0.6195	2.38	0.1222	0.0201
PwG-PC-GARCH	1.31	0.5169	0.1707	0.6795	1.1490	0.2838	0.0167
GP-GARCH	4.30	0.1160	0.33	0.5617	3.97	0.04	0.0234
GP-PC-GARCH	2.63	0.2677	0.24	0.6195	2.38	0.1222	0.0201

Note: In this table, the 99%-VaR estimation performance of the GP and PwG methods is evaluated, applying GARCH filter once in the original space (PwG-GARCH) and once in the reduced space (PwG-PC-GARCH).

Table 13: The Ranking Model results for the GP, GP-PC, PwG, and PwG-PC methods

Method	Penalty	Percentage	Rank
PwG-GARCH	3.66×10^{-5}	51.18	2
PwG-PC-GARCH	3.49×10^{-5}	48.82	1
GP-GARCH	3.67×10^{-5}	49.82	1
GP-PC-GARCH	3.70×10^{-5}	50.18	2

Note: In this table, the 99%-VaR estimation performances of the GP and PwG methods are compared, applying GARCH filter once in the original space (PwG-GARCH) and once in the reduced space (PwG-PC-GARCH).

computational load without resulting in a significant error in the final estimation. In addition, changes in the correlation of interest rates at different maturities have not reduced the accuracy of the final estimate because, otherwise, there would be a significant difference between the performances of the two methods.

Electronic copy available at: https://ssrn.com/abstract=3889826

4 Conclusion

Calculating the value of non-linear instruments in fixed-income portfolios is time-consuming and imposes a significant computational load due to the complexity of the design and the lack of explicit formulae. The first part of this paper proposes a simple method for calculating PnL to reduce this computational load. This method can also facilitate employing the FRTB proposed approaches for calculating the regulatory capital. In the second part, we try to improve the performance of our method by using filters such as GARCH and EWMA. Finally, we measure our performance in each of the mentioned parts using the unconditional coverage, independence, conditional coverage, Diebold-Mariano tests, and the Ranking Model. The method of Gibson and Pritsker [2000], one of the best methods proposed in this field, has been chosen as the benchmark.

In our proposed method, called PwG, we first assign an impact characteristic to each scenario. We use interest rate shocks associated with a scenario and the change in the portfolio value to define this indicator. We select some scenarios as the base scenarios, and in the next step, calculate the portfolio value for them. Then, according to the impact characteristics and changes in the portfolio value for the base scenarios, the portfolio PnLs are interpolated for the other scenarios resulting in a VaR estimate. The results of our numerical tests show the better performance of the PwG method at high confidence levels (e.g., 99%), except for the firm's point of view in the Diebold-Mariano test. Unlike the PwG method, VaRs estimated using the GP method show significant fluctuations over time. These fluctuations are not due to the high accuracy of this method and are due to noise resulting from performing a random sampling process. Therefore, one can consider the absence of fluctuations in the PwG VaRs to advantage over the GP VaRs. In addition, the Murphy diagram shows the superiority of the PwG by using consistent scoring functions to evaluate the two methods. Given these results, the regulators can propose the PwG method as an appropriate procedure to integrate VaR estimation and capital allocation process for fixed-income portfolios. The simplicity of this method reduces the possibility of deviating from the specified procedure and guarantees a more desirable integration. The appropriate performance of this method from the regulator's viewpoint and the absence of large fluctuations in the overall trend of estimated VaRs confirm our claim.

The PwG method is a historical simulation method. One of the ideas proposed to increase the accuracy of such methods is to consider the volatility changes over time by filtered historical simulation. The PwG method uses filtration by updating interest rate shocks at different maturities and then considers them as possible scenarios for the coming day. The results of various tests confirm increased accuracy in VaR estimation. Furthermore, several other experiments show that this improvement is more pronounced in cases where we are experiencing higher volatility. When applying the unfiltered PwG and GP methods, risk managers might not observe appropriate reactions to the significant realized losses, leading to fundamental problems for institutions. All the tests report a significant improvement for the filtered methods, and the estimated VaRs show appropriate responses to the large realized losses. Furthermore, we apply these filters to main risk factors instead of interest rates. There are two main objectives of using this approach. First, preventing the destructive impact of the changes in the correlation of the interest rates term structure; Second, reducing the number of filters and consequently the computational load.

In another way to consider volatility changes, we assign the probabilities proposed by the Boudoukh et al. [1998] to historical scenarios. As we move away from the present time, these probabilities decrease exponentially. According to the results, the performance of this method is poor at 99% confidence level but acceptable at 95% confidence level.

Electronic copy available at: https://ssrn.com/abstract=3889826

References

[1] Peter A Abken. An empirical evaluation of value at risk by scenario simulation. *The Journal of Derivatives*, 7(4):12–29, 2000.

[2] Carlo Acerbi and Dirk Tasche. Expected shortfall: a natural coherent alternative to value at risk. *Economic notes*, 31(2):379–388, 2002.

[3] Muzaffer Akat and Cahit Memis. Will switching from the var to the expected shortfall provide the efficiency in the capital adequacy? evidence from the fx positions. *Eurasian Journal of Business and Management*, 6(2):1–13, 2018.

[4] David Allen, Colin Lizieri, and Stephen Satchell. A comparison of non-gaussian var estimation and portfolio construction techniques. *Journal of Empirical Finance*, 58:356–368, 2020.

[5] Philippe Artzner. Thinking coherently. *Risk*, pages 68–71, 1997.

[6] Philippe Artzner, Freddy Delbaen, Jean-Marc Eber, and David Heath. Coherent measures of risk. *Mathematical finance*, 9(3):203–228, 1999.

[7] Giovanni Barone-Adesi, Kostas Giannopoulos, and Les Vosper. Var without correlations for portfolios of derivative securities. *Journal of Futures Markets*, 19(5):583–602, 1999.

[8] II Basel. International convergence of capital measurement and capital standards: a revised framework basle committee on banking supervision. *URL: http://www. bis. org/publ/bcbs*, 107, 2004.

[9] Jacob Boudoukh, Matthew Richardson, and Robert Whitelaw. The best of both worlds. *Risk*, 11 (5):64–67, 1998.

[10] R Bustreo et al. Value-at-risk for fixed-income portfolios: A kalman filtering approach. *IMA Journal of Management Mathematics*, 27(4):557–573, 2016.

[11] Massimiliano Caporin. Evaluating value-at-risk measures in the presence of long memory conditional volatility. *The Journal of Risk*, 10(3):79, 2008.

[12] Junyao Chen, Tony Sit, and Hoi Ying Wong. Simulation-based value-at-risk for nonlinear portfolios. *Quantitative Finance*, 19(10):1639–1658, 2019.

[13] Adnan Chishti. Simulation of fixed-income portfolios using grids. *Algo Research Quarterly*, 2(2): 41–50, 1999.

[14] Peter F Christoffersen. Evaluating interval forecasts. *International economic review*, pages 841–862, 1998.

[15] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

[16] Werner Ehm, Tilmann Gneiting, Alexander Jordan, and Fabian Krüger. Of quantiles and expectiles: consistent scoring functions, choquet representations and forecast rankings. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, pages 505–562, 2016.

Electronic copy available at: https://ssrn.com/abstract=3889826

[17] Joseph Engelberg, Charles F Manski, and Jared Williams. Comparing the point predictions and subjective probability distributions of professional forecasters. *Journal of Business & Economic Statistics*, 27(1):30–41, 2009.

[18] Roberta Fiori and Simonetta Iannotti. Scenario-based principal component value-at-risk when the underlying risk factors are skewed and heavy-tailed: an application to italian banks' interest rate risk exposure. *The Journal of Risk*, 9(3):63, 2007.

[19] Christian Francq and Jean-Michel Zakoian. Virtual historical simulation for estimating the conditional var of large portfolios. *Journal of Econometrics*, 217(2):356–380, 2020.

[20] Jon Frye. Monte carlo by day. *Risk*, 11(11):66–71, 1998.

[21] Alexandros Gabrielsen, Axel Kirchner, Zhuoshi Liu, and Paolo Zagaglia. Forecasting value-at-risk with time-varying variance, skewness and kurtosis in an exponential weighted moving average framework. *Annals of Financial Economics*, 10(01):1550005, 2015.

[22] Michael S Gibson and Matt Pritsker. Improving grid-based methods for estimating value at risk of fixed-income portfolios. *Board of Governors of the Federal Reserve System, Finance and Economics Working Paper*, (25), 2000.

[23] Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

[24] Darryll Hendricks. Evaluation of value-at-risk models using historical data. *Economic policy review*, 2(1), 1996.

[25] John Hull and Alan White. Incorporating volatility updating into the historical simulation method for value-at-risk. *Journal of risk*, 1(1):5–19, 1998.

[26] Farshid Jamshidian and Yu Zhu. Scenario simulation: Theory and methodology. *Finance and stochastics*, 1(1):43–67, 1996.

[27] Oskar Kallur. On the use of value-at-risk based models for the fixed income market as a risk measure for central counterparty clearing, 2016.

[28] Paul Kupiec. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*, 3(2), 1995.

[29] Jose A Lopez. Methods for evaluating value-at-risk estimates. *Economic Review-Federal Reserve Bank of San Francisco*, (2):3, 1999.

[30] Benoit B Mandelbrot. The variation of certain speculative prices. In *Fractals and scaling in finance*, pages 371–418. Springer, 1997.

[31] Alexander J McNeil and Rüdiger Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of empirical finance*, 7(3-4):271–300, 2000.

Electronic copy available at: https://ssrn.com/abstract=3889826

[32] AH Murphy and H Daan. Forecast evaluation. probability, statistics, and decision making in the atmospheric sciences. *Chapter*, 10:379–437, 1985.

[33] Allan H Murphy. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, 105(7):803–816, 1977.

[34] Georg Ch Pflug. Some remarks on the value-at-risk and the conditional value-at-risk. In *Probabilistic constrained optimization*, pages 272–281. Springer, 2000.

[35] Simona Roccioletti. *Backtesting value at risk and expected shortfall*. Springer, 2015.

[36] Pilar Abad Romero, Sonia Benito Muela, and CARMEN LÓPEZ MARTÍN. Role of the loss function in the var comparison. *Documentos de Trabajo FUNCAS*, 756(1), 2014.

[37] Mandira Sarma, Susan Thomas, and Ajay Shah. Selection of value-at-risk models. *Journal of Forecasting*, 22(4):337–358, 2003.

[38] Leonard J Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[39] Emrah Şener, Sayad Baronyan, and Levent Ali Mengütürk. Ranking the predictive performances of value-at-risk estimation methods. *International Journal of Forecasting*, 28(4):849–873, 2012.

[40] Giorgio Szegö. Measures of risk. *Journal of Banking & finance*, 26(7):1253–1272, 2002.

[41] William Thomson. Eliciting production possibilities from a well-informed manager. 1978.

[42] Ruodu Wang and Ričardas Zitikis. An axiomatic foundation for the expected shortfall. *Management Science*, 2020.

Electronic copy available at: https://ssrn.com/abstract=3889826

Appendix

A An Overview of the VaR Calculation Methods

This section briefly reviews all the VaR calculation methods used in this paper, classified into three groups: non-parametric, parametric, and semi-parametric. The first category, non-parametric methods, assumes no specific distributions for the return and directly uses historical data. Parametric approaches use distributions such as normal and student's-t. Semi-parametric methods employ parametric and non-parametric methods with different techniques.

A.1 Non-parametric Methods

A.1.1 Historical Simulation

This method uses historical scenarios to estimate the expected distribution of losses. VaR is a specific quantile of this distribution.

A.1.2 Monte Carlo Simulation

In the Monte Carlo framework for measuring risk, we simulate the value of financial assets based on a specific rule, resulting in a series of portfolio PnL values. Similar to the historical simulation approach, considering the quantile of this portfolio PnL series, we can estimate VaR or other risk measures.

A.2 Parametric Methods

A.2.1 Variance-Covariance

In this method, we assume a specific distribution for portfolio returns. To obtain the variance of this distribution, we use the covariance matrix of assets in the portfolio. Thus, considering normal returns with zero mean, the VaR at time t can be calculated as follows:

$$\mathrm{VaR}_t = \phi^{-1}(\Theta)\sigma_t, \tag{23}$$

where ϕ is the cumulative standard normal distribution, σ_t is the calculated variance of the portfolio, and Θ is the confidence level.

A.2.2 GARCH

The GARCH model uses previous information from the lagged variances and returns to predict the variance of future returns distribution. GARCH(p,q) calculates the variance as follows

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{24}$$

$$\varepsilon_t = \sigma_t \eta_t, \quad \eta_t \sim \phi(0,1) \quad i.i.d., \tag{25}$$

where α_i's and β_i's are constants, ε_t is the innovation term and σ_t^2 is the variance of ε_t. Then, VaR is calculated by using Equation (23) with the forecasted variance. It is worth noting that the Student's t-distribution can also be used for η_t.

Electronic copy available at: https://ssrn.com/abstract=3889826

A.2.3 EGARCH

As an extension of the GARCH family models, EGARCH has been developed to better model the volatility asymmetry in financial data. In this method, the forecasted variance is calculated as follows

$$g\left(Z_t\right)=\theta Z_t+\lambda\left(\left|Z_t\right|-E\left(\left|Z_t\right|\right)\right) \quad Z_t\sim\phi(0,1), \tag{26}$$

$$\log\sigma_t^2=\omega+\sum_{k=1}^{q}\beta_k g\left(Z_{t-k}\right)+\sum_{k=1}^{p}\alpha_k\log\sigma_{t-k}^2, \tag{27}$$

where θ, λ, ω, β and α are constant coefficients, and σ_t^2 is the conditional variance. Then, Equation (23) is used to estimate the VaR. A Student's t-distribution can also be assumed instead of normal distribution for Z_t.

A.3 Semi-parametric Methods

A.3.1 Filtered Historical Simulation

This method calculates the VaR by combining the Historical Simulation (HS) and GARCH methods. Since one of them is parametric and the other is non-parametric, the FHS falls into semi-parametric methods. The VaR estimation procedure in this method is similar to the HS. Nevertheless, instead of using the selected scenarios directly, we rescale them according to the predicted volatility using a GARCH process.

B An Overview of Back-testing approaches

In this section, we give a brief overview of the back-testing approaches used in this paper. We employ four widely used models to assess the performance of the VaR estimation methods from different viewpoints. We use Kupiec and Christoffersen tests for model validation and the Diebold-Mariano (DM) predictive ability test, and the Ranking Model to compare methods with each other. In the DM test, different loss functions are used to identify the superior method from two different perspectives of regulators and firms. The main concern of regulators is the realization of losses greater than the VaR, while firms' is the over-allocation of capital. Also, using [39] ranking model, different methods are ranked based on their accuracy in estimating the VaR.

B.1 Kupiec test

The Kupiec test examines the equality of the realized violation rate and the expected violation rate based on VaR estimation's confidence level. The null hypothesis is defined as follows

$$H_0: p=\hat{p}=\frac{x}{T}, \tag{28}$$

where x and T are the numbers of violations and observations, respectively, p is the expected violation rate corresponding to the VaR confidence level, and $\hat{p}$ is the realized violation rate. The test statistics

Electronic copy available at: https://ssrn.com/abstract=3889826

in this likelihood-ratio test is

$$LR_{uc} = -2\ln\left(\frac{(1-p)^{T-x}p^{x}}{\left[1-\left(\frac{x}{T}\right)\right]^{T-x}\left(\frac{x}{T}\right)^{x}}\right). \tag{29}$$

The test statistic LR_{uc} has a $\chi^2(1)$ distribution under the null hypothesis.

B.2 Christoffersen test

In addition to the number of violations, the Christoffersen test considers their status of independence. This test tries to examine whether the violations are bunching or spreading independently over time. The test statistics are defined as follows

$$LR_{cc} = -2\ln\left(\frac{(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right) + LR_{uc}, \tag{30}$$

where n_{ij} is the number of periods with state j, which is followed by a period with state i. Occurrence and non-occurrence of violation are shown by states 1 and 0, respectively, and the parameters π_0, π_1 and π are defined as follows

$$\begin{aligned}\pi_0 &= \frac{n_{01}}{n_{00}+n_{01}},\\ \pi_1 &= \frac{n_{11}}{n_{10}+n_{11}},\\ \pi &= \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}.\end{aligned} \tag{31}$$

The null hypothesis in this test is

$$H_0 : \pi_0 = \pi_1 = \alpha, \tag{32}$$

where α is the expected rate of violation corresponding to the VaR confidence level. Under this hypothesis, LR_{cc} has a $\chi^2(2)$ distribution.

B.3 Diebold-Mariano test

This test compares the accuracy of two different methods in estimating a variable. The Diebold-Mariano test defines the loss series of method i by e_i, and the loss differential series by $d = g(e_i) - g(e_j)$, where g is a loss function. The null hypothesis and the test statistics under the null hypothesis is

$$H_0 : \mu_d = 0. \tag{33}$$

Assuming that the forecasts are h-step-ahead, the test statistics under the null hypothesis is

$$DM = \frac{\bar{d}}{\sqrt{\frac{\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k}{T}}}, \tag{34}$$

where T is the number of observations, and γ_k is defined as follows

$$\hat{\gamma}_d(k) = \frac{1}{T}\sum_{t=|k|+1}^{T}\left(d_t - \bar{d}\right)\left(d_{t-|k|} - \bar{d}\right). \tag{35}$$

Under the null hypothesis, the test statistics DM follows a standard normal distribution.

Electronic copy available at: https://ssrn.com/abstract=3889826

B.4 The Ranking Model

In the ranking model, the data is divided into violation and safe spaces. In the violation space, the size of violations and clusters formed by the succession of violations are considered. The first factor is defined as

$$\epsilon_t = (\mathrm{VaR}_t - x_t)\,, \tag{36}$$

and to combine the effect of successive violations, we assign the quantity C_i to the cluster number i with z_i violations,

$$C_i = \prod_{b=1}^{z_i} (1 + \epsilon_{b,i}) - 1. \tag{37}$$

Then, the interaction between clusters i and $i+m$ is defined as

$$C_i * C_{i+m} = \frac{1}{k_{i,i+m}} \left(\prod_{b=1}^{z_i} (1 + \epsilon_{b,i}) \prod_{b=1}^{z_{i+m}} (1 + \epsilon_{b,i+m}) - 1 \right), \tag{38}$$

where $k_{i,i+m}$ is the distance between clusters i and $i+m$. The penalty for the violation space is defined as

$$\begin{aligned} \Phi(x, \mathrm{VaR}) &= \sum_{i=1}^{\alpha-1} \sum_{m=1}^{\alpha-i} C_i * C_{i+m} \\ &= \sum_{i=1}^{\alpha-1} \sum_{m=1}^{\alpha-i} \frac{1}{k_{i,i+m}} \\ &\quad \times \left(\prod_{b=1}^{z_i} (1 + \epsilon_{b,i}) \prod_{b=1}^{z_{i+m}} (1 + \epsilon_{b,i+m}) - 1 \right). \end{aligned} \tag{39}$$

In the safe space, penalties are imposed only for deviations of the VaR from negative returns. The total penalty in this space is

$$\Psi(x, \mathrm{VaR}) = \sum_{t=1}^{T} \left[\mathbf{1}\left(x_t > \mathrm{VaR}_t | x_t < 0\right) \right] (x_t - \mathrm{VaR}_t)\,. \tag{40}$$

The total penalization measure (PM) is equal to the weighted average of the penalties associated to the two spaces. The weights depend on the confidence level used to calculate the VaR. Since only negative returns are considered in this model, we use a scaling factor, $\frac{1}{T^*}$, where T^* is the number of these observations. Therefore, the total penalization measure is calculated as

$$\mathrm{PM}(\theta, x, \mathrm{VaR}) = \frac{1}{T^*}[(1-\theta)\Phi(x, \mathrm{VaR}) + \theta\Psi(x, \mathrm{VaR})], \tag{41}$$

where θ is the VaR quantile. Finally, to compare the performance of several different VaR calculation methods, the ratio of the PM value for the j^{th} method to the sum of all PMs is calculated,

$$\mathrm{Ratio}_j = \frac{\mathrm{PM}_j}{\sum_{i=1}^{n} \mathrm{PM}_i}. \tag{42}$$

The lower the ratio obtained for each method, the better that method.

Electronic copy available at: https://ssrn.com/abstract=3889826