Zorc Finance Whitepaper

O. Hernandez[1] and W. Baumgart[1]
Zorc Finance Holding GmbH

(*Electronic mail: oscar-hernandez@zorc-finance.com)

(Dated: 2 July 2021)

This white paper outlines the algorithmic components of Zorc Finance software that solve various problems in mathematical porfolio theory. Our software is specialized in solving these problems when the key distributional parameters of the underlying portfolio, such as the mean expected returns and covariance matrix, are unknown and must be estimated from historical data. Parameter estimation results in uncertainties that must be carefully handled. Ignoring estimation errors and their sensitivity to input models results in incorrect estimates and poorly performing optimizations. These issues are tackled by Zorc Finance software through Bayesian algorithms. This statistical framework addresses the critical issues related to estimation uncertainties through the construction of the posterior distributions for the relevant parameters. The posterior distributions are also used to conduct rigorous statistical comparisons required for portfolio backtesting, or for performance comparisons against selected benchmarks. Furthermore, we apply this methodology to known portfolio optimization algorithms to quantify and test the robustness for the end users. As the scope of this work is large we proposed benchmarks for each software component to be addressed in the future.

I. INTRODUCTION

Zorc Finance GmbH is a software company aiming to improve the speed and quality of portfolio modelling and optimization leveraging Bayesian uncertainty analysis and experimental hardware such as quantum computing. There are four main components to the software that are in need of validation. The portfolio construction, modelling, optimization and comparison modules. The purpose of this article is to detail each component, their claimed value propositions and their proposed benchmarks for the future.

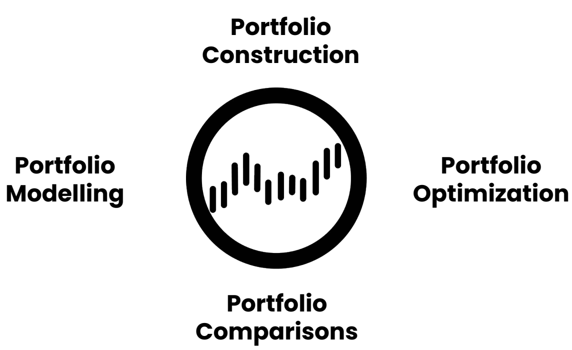

FIG. 1: An overview of the four main components of the Zorc Finance software tools: the portfolio constructor, the portfolio optimizer, the portfolio comparison and portfolio modelling tools.

In this introduction we discuss the data sources and datasets used, Bayesian time series analysis, the statistical model, and a discussion of outlier removal. This is is followed by Section II that describes in detail the portfolio construction problem and its implementation as an optimization. Section III describes in detail the portfolio modelling and comparison tools. In the final Section IV the various portfolio optimization methods are outlined.

A. Datasets and data providers

In this section we discuss the data sets that are used to conduct the benchmarks along with the current data provider. The data provider that is used by Zorc Finance is Leeway.tech which uses eodhistoricaldata.com. The available data includes

- 60+ stock exchanges,
- 120,000+ tickers,
- 200 indices,
- 120+ crypto currencies,
- 150+ forex pairs,
- 20+ years of fundamental data,
- 30+ years of end-of-day data,
- 20,000 US mutual funds with equity funds as well as balanced and bond-based mutual funds and
- 6000 ETFs across different exchanges and countries.

The full list of supported ETFs and exchanges can be found on their website.

B. Bayesian time series analysis

The algorithms at Zorc Finance generate the predictive posterior distributions of key parameters (volatilities, correlations, expected returns and Sharpe ratios) based on observed historical data using Bayesian inference. The main advantages of this approach is that:

- It provides a systematic way of combining prior information with current data. When new data is available, previous estimates can be updated consistently with Bayes' theorem.

- It is possible to conduct exact inference based on the data without need for approximations. This method can estimate any parametric function without reliance on previous parameter estimates.
- The method provides interpretable answers using credible intervals.
- Complex models are computationally tractable with current sampling approaches.

However, there are also several disadvantages that must be kept in mind

- There is no rule for selecting priors. Priors must be selected based on reasonable assumptions.
- Posterior distributions can be sensitive to priors, when little data is available.
- High computational cost.

The following subsection provides a detailed theoretical explanation of how this approach is implemented in practice along with the priors used.

C. Statistical model

Suppose we have a set of p price time series denoted as D. The first processing step that is carried out is to convert the time series data into a stationary time series. This can be carried out in many different ways, although here we use the fractional returns

$$r_i(t) = \frac{S_i(t) - S_i(t-1)}{S_i(t-1)}, \tag{1}$$

where $r_i(t)$ are the returns at time t of the ith price time series $S_i(t)$. Based on this data we want to determine the posterior distributions of distributional parameters $P(\boldsymbol{\theta}|)$ conditioned on the observed data D. For example, this can be the average (or expected) returns of the stock ($\boldsymbol{\mu}$) as well as the covariance matrix $\mathrm{Cov}(r_i, r_j)$. These distributions are derived from Bayes theorem as

$$P(\theta_k|D) \propto \int d\boldsymbol{\theta}_k \, P(D|\boldsymbol{\theta}) P(\boldsymbol{\theta}), \tag{2}$$

where $P(D|\boldsymbol{\theta})$ is the likelyhood of the data given the parameters $\boldsymbol{\theta}$ of the distribution. The integration vector $d\boldsymbol{\theta}_k = \frac{1}{d\theta_k}\prod_{j=0}^{J} d\theta_j$ where J is the total number of distributional parameters and $P(\boldsymbol{\theta})$ are the priors of the parameters. There are many types of priors that can be used. However, we will choose the forms given in Table I.

Name	Formula $P(\theta)$
Jeffreys prior	$\frac{1}{\theta}H(\theta)$,
Uniform prior	$\frac{1}{\Delta_x}H(\theta-\theta_{\mathrm{min}})H(\theta_{\mathrm{max}}-\theta)$.

TABLE I: The priors currently used in the Zorc Finance software.

In Table I, $H(\theta)$ is the Heaviside step function. Jeffreys prior is a scale-invariant prior which will be used for variances. The uniform prior is a step function normalized over the region between the minimum value θ_{min} and the maximum value θ_{max}. Next, we examine the form of the likelyhood function. One key assumption is that the returns are independent and identically distributed random variables (i.i.d)[1]. This allows the factorization of the likelyhood function as

$$P\left(D = \{\boldsymbol{r}_0, \boldsymbol{r}_1, ..., \boldsymbol{r}_N\}|\boldsymbol{\theta}\right) = \prod_{k}^{N} P(\boldsymbol{r}_k|\boldsymbol{\theta}), \tag{3}$$

where $P(\boldsymbol{r}_k|\boldsymbol{\theta})$ is the probability distribution of $\boldsymbol{r}_k$ given $\boldsymbol{\theta}$ in Table II. The evaluation of Eq. (3) is carried out with Markov Chain Monte Carlo (MCMC)[1] that samples the logarithm of the likelyhood.

Multivariate distribution	$P(\boldsymbol{r}\|\boldsymbol{\theta})$
Gaussian	$\frac{1}{\sqrt{(2\pi)^p\|\Sigma\|}} \times$ $\mathrm{Exp}\left(-\frac{1}{2}(\boldsymbol{r}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{r}-\boldsymbol{\mu})^T\right)$,
t-distribution	$\frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{p/2}\|\Sigma\|^{1/2}} \times$ $\left(1+\frac{1}{\nu}(\boldsymbol{r}-\boldsymbol{\mu})\Sigma^{-1}(\boldsymbol{r}-\boldsymbol{\mu})^T\right)^{-\frac{p+\nu}{2}}$.

TABLE II: The multivariate probability distributions used for modelling return distributions. Here Σ is a $p \times p$ matrix, where p is the number of time series considered.

For the multivariate Gaussian distribution we can relate the parameters of the distribution to the covariance and mean of the data through

$$\langle \boldsymbol{r} \rangle = \boldsymbol{\mu}, \tag{4}$$

$$\mathrm{Cov}(\boldsymbol{r}) = \Sigma. \tag{5}$$

For the multivariate t-distribution the relation between the parameters of the distribution and the covariance and mean are

$$\langle \boldsymbol{r} \rangle = \boldsymbol{\mu}, \tag{6}$$

$$\mathrm{Cov}(\boldsymbol{r}) = \frac{\nu}{\nu-2}\Sigma. \tag{7}$$

D. Approximation of the likelyhood function

The likelyhood function given in Eq. (3) involves J parameters. For p-assets, the value of J is related to p as

$$J = \frac{1}{2}p(p+3) + \lambda, \tag{8}$$

where $\lambda = 0$ for a multivariate Gaussian and $\lambda = 1$ for a t-distribution. For a portfolio of 30 assets this corresponds to approximately 500 variables. This can still be carried out by evaluating the log-likelyhood of Eq. (3), however, in practice, it was found to be more stable to simplify the problem by exploiting the properties of the distributions used and the fact that correlations are generally small. We begin by breaking up the data set into pairs of time-series $D_{ij} = \{\boldsymbol{r}_i, \boldsymbol{r}_j\}$. Using the basic conditional probability rule, we have

$$
\begin{aligned}
P(D|\boldsymbol{\theta}) = P(D_{12}|\boldsymbol{\theta}, D_{13}, D_{14}..., D_{21}, ..., D_{p-1,p})\times \\
P(D_{13}|\boldsymbol{\theta}, D_{14}, ..., D_{21}, ...)\times \\
P(D_{14}|\boldsymbol{\theta}, ..., D_{21}, ...)\times \\
\ldots\times \\
P(D_{p-1,p}|\boldsymbol{\theta}).
\end{aligned} \tag{9}
$$

To a leading order expansion in powers of the correlation coefficients, each of the conditional probability distributions can be expanded as

$$
\begin{aligned}
&P(D_{12}|\boldsymbol{\theta}, D_{13}, ..., D_{p-1,p}) = P(D_{12}|\theta_1, \theta_2, \rho_{12}) \\
&+ \sum_{(i,j)\neq(1,2)}^{(p-1,p)} \rho_{ij} \frac{\partial P(D_{12}|\boldsymbol{\theta}, D_{13}, ..., D_{p-1,p})}{\partial \rho_{ij}}.
\end{aligned} \tag{10}
$$

In the above expression θ_1, θ_2 are distributional parameters that correspond to $\boldsymbol{r}_1$ or $\boldsymbol{r}_2$, respectively, and ρ_{ij} are the correlation coefficients between assets (i, j). We have used the fact that the distributions used only have pair-wise correlations between different time series and do not consider higher moments. Therefore, to leading order in correlations we have

$$
\begin{aligned}
P(D|\boldsymbol{\theta}) = \prod_{ij} P(D_{ij}|\theta_i, \theta_j, \rho_{ij}) \\
+ \sum_{ij} \rho_{ij} \frac{\partial P(D_{ij}|\boldsymbol{\theta}...)}{\partial \rho_{ij}}.
\end{aligned} \tag{11}
$$

For practical calculations the correlation coefficient $|\rho_{ij}| < 1$. Hence, the likelyhood function is approximated as

$$
P(D|\boldsymbol{\theta}) \approx \prod_{(i,j);\ i>j}^{(p,p-1)} P(D_{ij}|\theta_i, \theta_j, \rho_{ij}). \tag{12}
$$

E. Outlier removal

In some cases it may be appropriate to improve the estimates of distributional parameters $\boldsymbol{\theta}$ by removing outliers in the data sets. The method used for this purpose is the interquartile range rule described below. Recall that the q-th percentile of a distribution is the location p_q where $q\%$ percent of the data are below this point. Algorithm 1 describes this process below.

Algorithm 1: Interquartile range outlier removal

Result: Data set X with removal of outliers
Choose outlier threshold f (default =1.5) ;
Compute the 25th and 75th percentile of the data;
p_{25} = percentile(0.25);
p_{75} = percentile(0.75);
$IQR = p_{75} - p_{25}$;
$x_{\max} = p_{75} + f \cdot IQR$;
$x_{\min} = p_{25} - f \cdot IQR$;
while x_k *in* X **do**
  instructions;
  if $(x_k < x_{\min})$ *or* $(x_k > x_{\max})$ **then**
    remove x_k from X;
  else
end

Once this outlier removal is carried out the distributional parameters are estimated. In practice, due to the heavy-tailed nature of the return distributions of securities, this IQR method is not appropriate when fitting these distributions. Therefore, we do not apply Algorithm 1 during the fitting of return distributions.

Questions to validate

1. How often to statistical projections for volatility and expected returns fall within the predicted confidence intervals for 3-month, 6-month, 1-year time frames?
2. How do the correlations,volatilities and expected return distributions change over different time periods? 3-month, 6-month, 1-year?
3. How sensitive are the distributions to different priors?
   - Uniform prior
   - Jeffreys prior
4. What is the difference in the results between Gaussian based error models and t-distributed errors?

II. PORTFOLIO CONSTRUCTION

The portfolio construction tool is a constrained combinatorical optimization algorithm that frames portfolio construction as a quadratic unconstrained binary optimization (QUBO) problem[2]. The D-Wave hybrid solver is used to solve the resulting QUBO with the constraint that a fixed amount of securities be chosen. This allows the user to construct a portfolio from scratch with K securities from a universe of P total securities. As an example, P will be number of securities available in the exchanges chosen by the user (P > 5000) while the desired size of the users portfolio is typically K< 50. In mathematical terms, the portfolio constructor is a mapping $f : \mathbb{R}^P \rightarrow \mathbb{R}^K$, from a vector space of dimensionality P to a subspace with dimensionality K with $K \ll P$.

Furthermore, the original portfolio construction problem has been extended to allow the user to begin with a pre-specified portfolio to which the user desires to add new securities in an optimal manner. In the following section, the mathematical details of the portfolio construction procedure will be described.

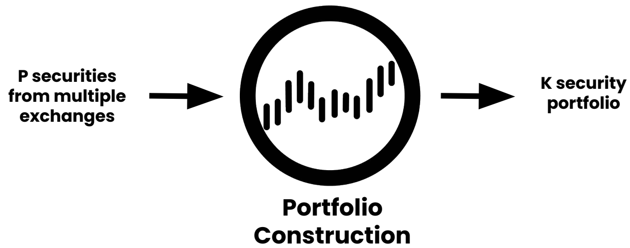

FIG. 2: A simple schematic of the input and output of the portfolio construction module.

A. Mathematical theory

The Hamiltonian that describes the mean-variance portfolio utility function[3] that is to be minimized is[4,5]

$$\mathscr{H} = -(1-\lambda)\,\boldsymbol{\mu}\cdot\mathbf{x} + \lambda\mathbf{x}\Sigma\mathbf{x}, \tag{13}$$

where λ is the risk aversion parameter $0 \le \lambda \le 1$, $\boldsymbol{\mu}$ is the vector of expected returns of a security and Σ is the covariance matrix of the securities. A value of $\lambda = 1$ signifies maximum aversion to risk, while $\lambda = 0$ is the most aggressive portfolio which maximizes returns at any cost. Here, $\boldsymbol{x}$ is a vector of length P with its entries represented as integer variables. These integer variables can be represented in binary form as

$$x_k = \sum_{n=0}^{N_x} 2^n b_{k,n}, \tag{14}$$

where $b_{k,n}$ are the binary coefficients. For simplicity, we consider the case where $N_x = 1$ so $x_k \in \{0,1\}$. This case corresponds to determining the optimal, equal, price-weighted portfolio. The minimization of Eq. (13) may result in a portfolio with any number of assets. For most practical purposes, we want the optimal portfolio with K securities. This investment requirement is imposed by adding the quadratic constraint function

$$C = A\left(K - \sum_i x_i\right)^2, \tag{15}$$

to the Hamiltonian in Eq. (15) and where A is a Lagrange multiplier to be determined numerically. The value K is also expanded in binary variables z_n as

$$K = \sum_{n=1}^{M} y_n z_n, \tag{16}$$

where y_n are the expansion coefficients given by

$$y_n = \begin{cases} 2^n & \text{if } n < M \\ K + 1 - 2^M & \text{if } n = M, \end{cases} \tag{17}$$

with the maximum index M

$$M = \lceil \log_2(K) \rceil - 1. \tag{18}$$

In order to map the Hamiltonian described above into a QUBO problem the Hamiltonian must assume the form[6]

$$\mathscr{H} = \sum_i Q_i x_i + \sum_{i<j} Q_{ij} x_i x_j. \tag{19}$$

To place the Hamiltonian in this form we first expand the covariance term in Eq. (13) and separate it into linear and quadratic terms

$$\begin{aligned} \mathbf{x}\Sigma\mathbf{x} &= \sum_{ij} x_i x_j \Sigma_{ij} \\ &= \sum_i x_i^2 \Sigma_{ii} + \sum_{ij} x_i x_j \Sigma_{ij} \\ &= \sum_i x_i \Sigma_{ii} + \sum_{i<j} x_i (2\Sigma_{ij})\, x_j. \end{aligned} \tag{20}$$

The expansion of the quadratic constraint term results in

$$\begin{aligned} C &= A\left(\sum_n y_n z_n - \sum_i x_i\right)^2, \\ &= A\left(\sum_{n_1 n_2} y_{n_1} y_{n_2} z_{n_1} z_{n_2} - 2\sum_{ni} y_n x_i z_n + \sum_{ij} x_j x_i\right), \\ &= A\left(\sum_{n=1} y_n^2 z_n + \sum_{n_1<n_2} 2 y_{n_1} y_{n_2} z_{n_1} z_{n_2} \right. \\ &\left. -2\sum_{ni} y_n x_i z_n + \sum_i x_i + \sum_{i<j} 2 x_j x_i\right). \end{aligned} \tag{21}$$

Substituting Eq. (20) and Eq. (21) into the full Hamiltonian gives

$$\begin{aligned} \mathscr{H} &= \sum_i \left(-(1-\lambda)\,\mu_{0,i} + \lambda\Sigma_{ii} + A\right) x_i + \sum_n A y_n^2 z_n \\ &+ \sum_{i<j} (2\lambda\Sigma_{ij} + 2A)\, x_i x_j + \sum_{n_1<n_2} (2A y_{n_1} y_{n_2})\, z_{n_1} z_{n_2} \\ &+ \sum_{ni} (-2A y_n) x_i z_n. \end{aligned} \tag{22}$$

The first line in Eq. (22) contain the linear terms, the second are the quadratic terms, while the last line represents the interaction terms.

The QUBO in Eq. (22) represents the Hamiltonian in the general case and the minimization of the Hamiltonian will select the best combination of K assets from a universe of P securities. However, in certain cases, a user may already have an existing portfolio and wishes to simply add more securities to it. In this case, Eq. (22) will need to be modified. To understand how to conduct the optimization under these conditions, we take our original P-dimensional vector and partition it into two components. Let the set $\mathscr{P} = \{k : k = 0, \ldots, P\}$ be the set of indices of the original vector $\boldsymbol{x}$. We separate this set into the two subsets sets

$$\mathscr{P} = S + I, \tag{23}$$

where S is the set of indices corresponding to the previously selected securities ($x_s = 1$) and I is the set of indices for all other securities. Furthermore,

$$K = K_0 + \Delta K \tag{24}$$

where K_0 is the number of securities in the current portfolio $K_0 = \sum_{s \in S} x_s$ and ΔK is the desired number of additional securities that the user wants to add to their portfolio. The value ΔK is expanded with binary variables z_n as

$$\Delta K = \sum_{n}^{\Delta M} c_n z_n, \tag{25}$$

where the coefficients c_n and ΔM are defined analogously to Eq. (17) and Eq. (18). The full Hamiltonian is given by

$$\begin{aligned}\mathscr{H} = &\sum_{i \in I} \left(-(1-\lambda)\,\mu_{0,i} + \lambda \Sigma_{ii} + A + \lambda \sum_{s \in S, s<i} (2\Sigma_{si}) \right) x_i \\ &+ \sum_n A c_n^2 z_n \\ &+ \sum_{(i,j) \in I \times I, i<j} (2\lambda \Sigma_{ij} + 2A)\, x_i x_j \\ &+ \sum_{n_1 < n_2} (2A c_{n_1} c_{n_2})\, z_{n_1} z_{n_2} \\ &+ \sum_{(n,i), i \in I} (-2A c_n) x_i z_n . \end{aligned} \tag{26}$$

The minimization of Eq. (22) and Eq. (26) falls within the NP-hard class of computational problems and may be solved using classical optimization methods such as simulated annealing, or through quantum annealing. For the latter case, we use quantum annealing hardware from D-Wave. The architecture of the hardware allows for problems to be solved with up to 64 fully-connected qubits on the 2000Q or up to 126 qubits in the new D-Wave Advantage system[7]. For a single stock exchange consisting of thousands of listings, the portfolio optimization problem is too large to be mapped onto this hardware. To overcome this limitation, the D-Wave hybrid solver is used. This solver uses both classical and quantum algorithms to solve problems much larger than would be allowed by the hardware and is the principle D-Wave solver that we use for the portfolio construction problem.

Questions to validate

1. What is the speed of the optimization vs problem size for the D-Wave hybrid solver vs the classical optimizer?
   - How does the increase in bit-precision of the solution effect the solution and stability qualities?
   - Is the QUBO solution better than the continuous variable formulation of the problem?
2. What is the quality of the solutions from the D-Wave hybrid annealer in comparison to classical solvers?
   - How do the energies of the solutions compare?
   - How do the stability of the solutions compare?
3. What are the costs of the D-Wave machine for the problem sizes considered in comparison to classical computations?
4. How well do the solutions generalize into the future for the D-Wave and the classical solutions?

III. PORTFOLIO MODELLING AND COMPARISONS

The purpose of these modules are to determine the key metrics of an existing portfolio in a rigorous statistical manner and to simulate the effects of changes in its composition that may be caused through customer actions such as (selling and buying) or to compare two different portfolios. The functions performed by the modelling and comparison tools are:

- to determine the key parameters of a specified portfolio,
- compare the key metrics of different portfolios using confidence intervals,
- and continuously update determined metrics based on new data.

For convenience, the portfolio modelling and comparison tools are separated into two distinct functions. A schematic of the portfolio modelling and comparison modules are given in Fig. 3 and Fig. 4, respectively. Details of the mathematical formulation of these Bayesian algorithms are provided in the next section.

A. Theory

In what follows, we assume that the user has a portfolio specified by a weight vector $\boldsymbol{w}$ that represents their percentage holdings $w_i > 0$ and where $\sum w_i = 1$. The daily returns R_i, of the portfolio is

$$R_i = \sum_{k=1}^{N} w_k r_{k,i}, \tag{27}$$

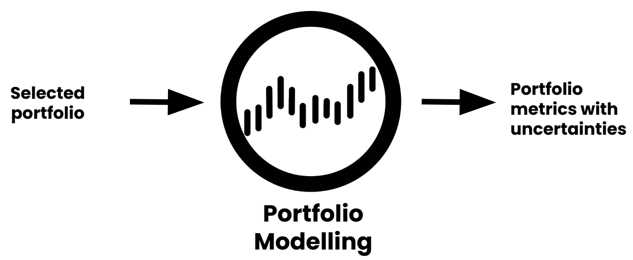

FIG. 3: A simple schematic of the input and output of the portfolio modelling module.

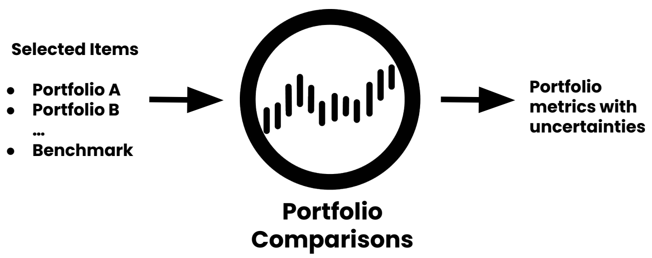

FIG. 4: A simple schematic of the input and output of the portfolio comparison module.

where $r_{k,i}$ is the returns of the kth security at the ith time step and N is the total number of assets in the portfolio. The data set is given by $D = \{R_0, R_1, \ldots R_T\}$. As before, we are interested in determining the distributional parameters $\boldsymbol{\theta}$ that correspond to the data generated by this portfolio. Applying Bayes theorem we have that the posterior distribution of the parameters $\boldsymbol{\theta}$ is given by

$$P(\boldsymbol{\theta}|D) \propto \int d\boldsymbol{\theta}\; P(D|\boldsymbol{\theta})P(\boldsymbol{\theta}). \tag{28}$$

Once these parameters are estimated, the return distribution is the marginalized distribution over the parameters $\boldsymbol{\theta}$.

$$\begin{aligned} P(R|D) &= \int d\boldsymbol{\theta}\; P(R|\boldsymbol{\theta}, D)P(\boldsymbol{\theta}|D), \\ &= \int d\boldsymbol{\theta}\; P(R|\boldsymbol{\theta})P(\boldsymbol{\theta}|D). \end{aligned} \tag{29}$$

In the second line, the independence of R and D was used. This is the preferred method to obtain forecasts of the distributions of the portfolios. In Figure 5 the distribution $P(R|\boldsymbol{\theta}_0)$ for $\boldsymbol{\theta}_0 = \text{median}\{P(\boldsymbol{\theta}|\boldsymbol{D})\}$ is shown for two different statistical models in comparison to the sample data.

The figure indicates that while the Gaussian distribution is a reasonable approximation to the data, the t-distribution yields a much better fit and is therefore the preferred data model to use. The next sections of this article describe how to apply this formalism in order to perform Bayesian updating, comparisons and forecasting.

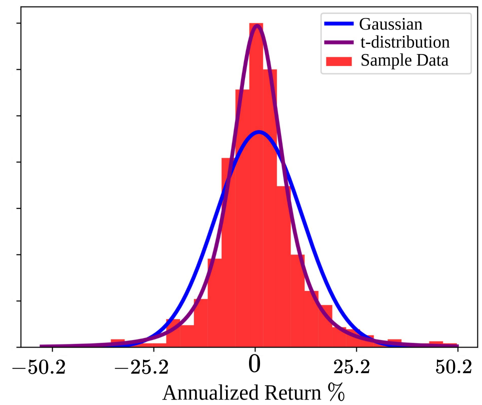

FIG. 5: The annualized return distribution for a sample portfolio along with different fits to the data. The t-distribution in purple and the Gaussian distribution in blue for the $p = 1$ case from Table II.

1. *Bayesian updating*

The Bayesian methodology provides a clear prescription on how to update the estimated parameters $\boldsymbol{\theta}$ when a new data point D_{k+1} is available. Let the set $\boldsymbol{D}_t = \{D_0, \ldots, D_t\}$ denote the data collected up to time t. At time $t+1$, the data point D_{t+1} is collected. The posterior distribution of the parameters $P(\boldsymbol{\theta}|\boldsymbol{D}_k)$ will then be updated according to the following prescription

$$\begin{aligned} P(\boldsymbol{\theta}|\boldsymbol{D}_{t+1}) &= P(\boldsymbol{\theta}|D_{t+1}, \boldsymbol{D}_t) \\ &= \frac{P(D_{t+1}|\boldsymbol{\theta}, \boldsymbol{D}_k)P(\boldsymbol{\theta}|D_t)}{P(D_{t+1}|\boldsymbol{D}_t)}, \\ &= \frac{P(D_{t+1}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{D}_t)}{P(D_{t+1}|\boldsymbol{D}_t)}. \end{aligned} \tag{30}$$

The distribution $P(\boldsymbol{\theta}|\boldsymbol{D})$ is assumed to be i.i.d. The denominator $P(D_{t+1}|\boldsymbol{D}_t)$ is the normalizing factor given by

$$P(D_{t+1}|\boldsymbol{D}_t) = \int d\boldsymbol{\theta}\; P(D_{t+1}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{D}_t). \tag{31}$$

The procedure in Eq. (30) allows the forecasts and other parameters to adapt to incoming information efficiently using particle filter algorithms[8].

2. *Bayesian comparisons for portfolios and backtesting*

A user may want to compare two portfolios characterized by weights $\boldsymbol{w}_A$ and $\boldsymbol{w}_B$, respectively. The comparison of the distributional parameters $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_A - \boldsymbol{\theta}_B$ is carried out as

$$P(\Delta\boldsymbol{\theta}) = \int d\boldsymbol{\theta}_A \int d\boldsymbol{\theta}_B \, P(\Delta\boldsymbol{\theta}|\boldsymbol{\theta}_A, \boldsymbol{\theta}_B)P(\boldsymbol{\theta}_A)P(\boldsymbol{\theta}_B),$$
$$= \int d\boldsymbol{\theta}_A \int d\boldsymbol{\theta}_B \, \delta(\Delta\boldsymbol{\theta} - \boldsymbol{\theta}_A + \boldsymbol{\theta}_B)P(\boldsymbol{\theta}_A)P(\boldsymbol{\theta}_B). \quad (32)$$

The evaluation of Eq. (32) is carried out using MCMC samples. The main advantage of this method is that it allows the comparison of portfolio metrics in a statistically rigorous manner with confidence intervals. In Fig. 6 the posterior distributions $P(\sigma_A|\boldsymbol{D})$ and $P(\sigma_B|\boldsymbol{D})$ of the volatilties for two sample portfolios is presented using this approach. This figure demonstrates the utility of Bayesian uncertainty analysis in the context of portfolio modelling. In contrast to a simple point estimate of the volatility, Bayesian modelling reproduces the full set of values consistent with the statistical model and confidence intervals may be easily extracted.

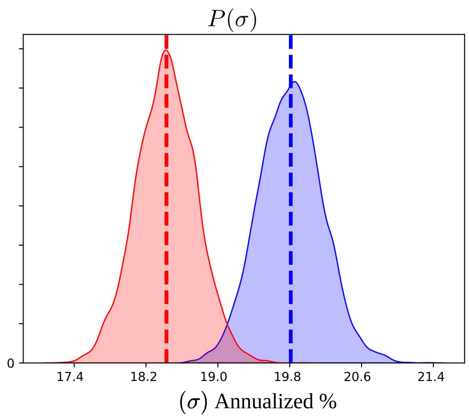

FIG. 6: The distributions of the volatility estimates for two sample portfolios during a chosen time frame obtained from Bayesian parameter estimation.

In Fig. 7 the posterior distribution of the difference $\Delta\sigma$ of volatility for two sample portfolios is plotted with different confidence intervals shown in dashed lines. The generation of this posterior distribution allows for a rigorous comparison between the two sample portfolio metrics based on available data which is more general than traditional parametric tests for statistical significance.

During backtests, the time series data is split into training and testing sets according to a specified fraction which is typically 70-80% training data. The training/testing data is denoted as $D_{\text{train}}/D_{\text{test}}$, respectively. The best performing statistical models will have equal performances in both the D_{train} and D_{test} data sets. The portfolio comparison module is suited to conduct this type of comparison test. The key metrics of the portfolio $\boldsymbol{\theta}_{\text{train}}$ and $\boldsymbol{\theta}_{\text{test}}$ are calculated and compared in a statistically rigorous way as $P(\Delta\boldsymbol{\theta})$ from Eq. (32).

3. *Bayesian forecasting*

Once the posterior distributions of a given portfolio $P(R|D)$ have been calculated, it is straight-forward to use this distribution to propagate the price or value of the portfolio forward in

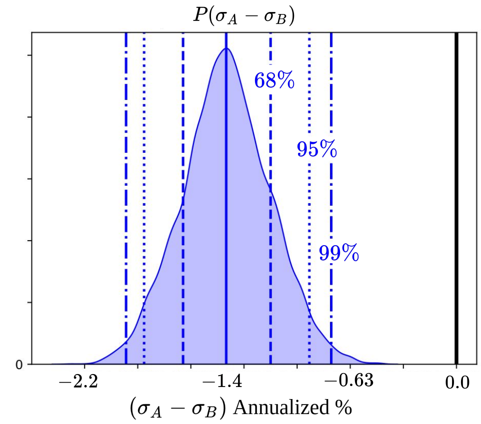

FIG. 7: The posterior distribution of the differences $\sigma_A - \sigma_B$ for the two sample portfolios obtained from Bayesian parameter estimation. The dashed lines indicate the 68%, 95% and 99% confidence intervals of the differences. The vertical solid black indicates the location of origin.

time. The price or value of an asset at time $S(t+1)$ is related to the the returns at time $R(t+1)$ and the price or value today $S(t)$ through

$$S(t+1) = (1 + R(t+1))S(t), \quad (33)$$

where $R(t+1)$ is the return sampled from the stationary distribution

$$P(R) = \int d\boldsymbol{\theta} \, P(r|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{D}). \quad (34)$$

The density distribution of the price $P(S(T)|D)$ is generated through Monte-Carlo (MC) simulations with the initial condition $S(t=0)$ being the price of the asset at the start of the simulation, or set to $S(t=0) = 1$ if only the relative increase of the value is needed.

An example of a Bayesian forecast $P(S|D)$ for a sample portfolio based on a training set D is provided in Fig. 8 for five forward time-steps obtained through the MC procedure above. The dark blue and light blue band indicates the 95% and 68% confidence intervals, respectively.

Questions to validate

1. How often do the statistical portfolio projections fall within the predicted confidence intervals?
2. How do the projections compare, in the 3-month, 6-month, 1-year periods?
3. Can anomalies in market such as during market crashes or sell-offs be detected using this methodology?
4. How well can VAR and CVAR be calculated?
5. What are the β-values of the portfolios?

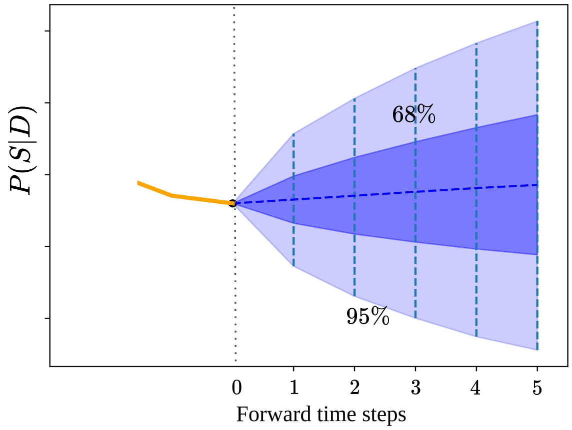

FIG. 8: The forecast of a sample portfolio into the future obtained from MC simulations using the posterior distribution in Eq. (34)

IV. PORTFOLIO OPTIMIZATION

The portfolio optimization module implements several types of optimization methods based in the literature[9,10]. The value that we add to this optimization is the ability to construct the confidence intervals and joint distributions of the optimization method that demonstrate the robustness of the method used. Testing the numerical stability of the solutions from the optimizers is crucial for finding portfolios that generalize well into the future[10,11]. The set of portfolio optimization strategies to which we apply the robust optimization procedure are

- maximum sharpe ratio portfolio,
- minimum volatility portfolio,
- inverse volatility portfolio,
- equally weighted portfolio,
- maximum-entropy to volatility ratio portfolio,
- hierarchical risk parity portfolio[11],
- the Kelly criterion portfolio[12],
- and the entropy regulated mean-variance portfolio customized by the customer.

A. Mathematical theory

Unlike the minimization of the the combinatorical Hamiltonian in Eq. (22), in this software component, the portfolio weights $\boldsymbol{w}$ that are optimized are continuous variables. Furthermore, it is assumed that the user has already chosen the set of stocks that they wish to hold and only requires to fine-tune the allocations and to study the numerics of the resulting portfolio in more detail. The objective of portfolio optimization is

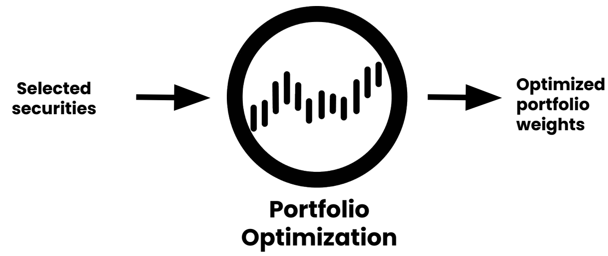

FIG. 9: A schematic of the portfolio optimization module.

to minimize the Hamiltonian given by

$$\mathscr{H} = -\lambda_\mu \boldsymbol{w}\cdot\boldsymbol{\mu} + \lambda_\sigma \boldsymbol{w}^T \Sigma \boldsymbol{w} - \lambda_S R(\boldsymbol{w}), \tag{35}$$

where the first term is the expected return of the portfolio and the second term is the volatility of the portfolio. The third term $R(\boldsymbol{w})$ is a regulator term that maximizes the concentration of the portfolio weight vector. We consider long-only portfolios where $\sum_i w_i = 1$. The parameters $\lambda_\mu, \lambda_\sigma, \lambda_S$ are chosen according to the preference of the user and parametrized so that they satisfy

$$0 \le \{\lambda_\mu, \lambda_\sigma, \lambda_S\} \le 1, \tag{36}$$
$$\lambda_\mu + \lambda_\sigma + \lambda_S = 1. \tag{37}$$

The following two forms of regulators are chosen

$$R_S(\boldsymbol{w}) = -\sum_{i=1}^{K} w_i \ln(w_i), \tag{38}$$
$$R_{w^2}(\boldsymbol{w}) = -\boldsymbol{w}\cdot\boldsymbol{w}. \tag{39}$$

The optimal portfolio $\boldsymbol{w}^*$ is then given by

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}}\{\mathscr{H}\}. \tag{40}$$

The posterior distribution of the optimal portfolio weights is calculated as

$$\begin{aligned} P(\boldsymbol{w}^*|D) &= \int d\boldsymbol{\theta}\, P(\boldsymbol{w}^*|\boldsymbol{\theta}) P(\boldsymbol{\theta}|D), \\ &= \int d\boldsymbol{\theta}\, \delta\left(\boldsymbol{w}^* - \min_{\boldsymbol{w}}\{\mathscr{H}\}\right) P(\boldsymbol{\theta}|D). \end{aligned} \tag{41}$$

The evaluation of Eq. (41) is carried out through the MC sampling of $P(\boldsymbol{\theta}|D)$. It is well-documented that minimizing the Hamiltonian in Eq. (35) suffers from numerical instabilities brought from estimation errors in $\boldsymbol{\theta}$[10,13,14]. Furthermore, our numerical experiments demonstrate that the stability of the minimization procedure is more sensitive to the errors from the estimation of $\boldsymbol{\mu}$ as opposed to Σ as discussed in Ref.[13]. Therefore, $\lambda_\mu = 0$ leads to numerically stable results. In the following subsections the various optimization methods implemented in the software are discussed.

1. Maximum Sharpe ratio portfolio

This portfolio is obtained by constructing the efficient frontier and finding the point along the frontier with the optimal Sharpe ratio r_S

$$r_S = \frac{\mu_P}{\sigma_P}, \tag{42}$$

where μ_P and σ_P is the expected return and volatility of the portfolio.

2. Minimum volatility portfolio

This portfolio is obtained by setting $\lambda_\mu = \lambda_S = 0$. This will construct the portfolio with the minimum volatility.

$$\boldsymbol{w}^* = \min_{\boldsymbol{w}}\{\boldsymbol{w}^T\Sigma\boldsymbol{w}\}, \tag{43}$$

subject to the normalization condition $\sum_i w_i = 1$.

3. Equally weighted portfolio

An equally weighted portfolio has weight vector

$$\boldsymbol{w}_K = \frac{1}{K}[1, \,\ldots\, 1], \tag{44}$$

where K is the total number of assets. This portfolio has the property of having maximum entropy, in the sense that for every $\boldsymbol{w}$ where $\sum_i w_i = 1$ and $w_i > 0$

$$S(\boldsymbol{w}) \leq S(\boldsymbol{w}_K), \tag{45}$$

where the entropy function $S(\boldsymbol{w})$ is defined as

$$S(\boldsymbol{w}) = -\sum_i w_i \ln(w_i) > 0, \tag{46}$$

and $S(\boldsymbol{w}_K) = K\ln(K)$. This portfolio is often used as a benchmark against other types of optimization strategies.

4. Maximum entropy to volatility portfolio

The objective of this optimization is to select the portfolio that minimizes the volatility of a portfolio while also maximizing the entropy of the portfolio weight vector $\boldsymbol{w}$. This is obtained by setting $\lambda_\mu = 0$ in Eq. (35) and minimizing the resulting Hamiltonian

$$\mathscr{H}(\lambda) = \lambda\boldsymbol{w}^T\Sigma\boldsymbol{w} - (1-\lambda)R(\boldsymbol{w}), \tag{47}$$

for different λ until the weights with the maximum ratio of $S(\boldsymbol{w})/\left(\boldsymbol{w}^T\Sigma\boldsymbol{w}\right)$ are obtained.

5. Inverse volatility

This portfolio assigns the weights w_i in proportion to the inverse of the volatility of each of the assets. The relative portfolio weight is

$$f_i = \frac{1}{\sigma_i^2}, \tag{48}$$

which is normalized so that $\sum w_i = 1$, resulting in

$$w_i = \frac{f_i}{\sum_i f_i}. \tag{49}$$

6. Hierarchical risk parity

This method implements the algorithm from Refs.[10,11] for portfolio optimization. This method applies clustering techniques to the covariance matrix of assets in a portfolio to construct well-diversified portfolios that perform out-of-sample.

7. Kelly allocation portfolio

The principle of this portfolio allocation strategy is based on the principle of maximizing the log-wealth of the investor. Suppose the total initial wealth of the investor is W_0 and the investor has the choice of placing a fractional amount of his wealth, f_i, into the risky asset such as a stock and the rest $1-f_i$ into a riskless asset, then the optimal fraction to invest maximizes the following

$$f_i = \mathrm{Max}_y\{\int dx_i \log(1+yx_i)P(x_i)\ \}, \tag{50}$$

where x_i is return of the ith risky asset and $P(x_i)$ is its distribution[12]. For certain distributions Eq. (50) admits closed forms. For example, if there are only two possible outcomes r_+ and r_- with probabilities p_+ and p_-, respectively, then the optimal fraction is

$$f_i = -\frac{p_+r_+ + p_-r_-}{r_-r_+}. \tag{51}$$

If the returns are normally distributed with means μ_i and volatilities σ_i, then the optimal fractions are

$$f_i = \frac{\mu_i}{\sigma^2}. \tag{52}$$

Due to the focus on long-only weights, we impose the additional constraint thats $f_i \geq 0$ and normalize the weights so $\sum_i f_i = 1$.

Questions to validate

1. For each of the portfolio optimization strategies in the previous section what is the comparison between the following metrics:

- volatility
- expected returns
- sharpe ratio
- max-draw downs
- VAR and CVAR?

2. How well do the above metrics generalize from training period, to testing periods for 3-months, 6-months, 1-year, 2-years?
3. How do each of these methods compare to market indices for the same periods?
   - S&P 500
   - DAX
   - Dow Jones
   - NASDAQ
   - MSCI World Index
   - Euro Stoxx 50
4. How do the returns from optimized portfolios compare to other assets such as bonds for multiple time periods?
5. How well can we detect and conduct rebalancing?
6. Does rebalancing improve the performance of the portfolios?

REFERENCES

[1] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support* (Cambridge University Press, 2005).

[2] F. Glover, G. Kochenberger, and Y. Du, "A tutorial on formulating and using qubo models," (2019), arXiv:1811.11538 [cs.DS].

[3] H. Markowitz, "Portfolio selection," The Journal of Finance **7**, 77–91 (1952).

[4] S. Mugel, C. Kuchkovsky, E. Sanchez, S. Fernandez-Lorenzo, J. Luis-Hita, E. Lizaso, and R. Orus, "Dynamic portfolio optimization with real datasets using quantum processors and quantum-inspired tensor networks," (2020), 2007.00017 [quant-ph].

[5] M. Hodson, B. Ruck, H. Ong, D. Garvin, and S. Dulman, "Portfolio rebalancing experiments using the quantum alternating operator ansatz," (2019), arXiv:1911.05296 [quant-ph].

[6] A. Lucas, "Ising formulations of many np problems," Frontiers in Physics **2**, 5 (2014).

[7] C. McGeoch and P. Farre, "The d-wave advantage system: An overview," Dwave Technical Report Series (2020).

[8] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," IEEE Transactions on Signal Processing **50**, 174–188 (2002).

[9] J. Baz and H. Guo, *An asset Allocation Primer: Connecting Markowitz, Kelly and Risk Parity* (PIMCO, 2017).

[10] M. L. de Prado, Journal of Portfolio Management (2016), https://dx.doi.org/10.2139/ssrn.2708678.

[11] M. Lopez de Prado, *Advances in financial machine learning* (Wiley, 2018).

[12] L. C. MacLean, E. O. Thorp, and W. T. Ziemba, *The Kelly Capital Growth Investment Criterion* (WORLD SCIENTIFIC, 2011) https://www.worldscientific.com/doi/pdf/10.1142/7598.

[13] D. P. S. M. F. Frank J. Fabozzi, Petter N. Kolm, *Robust Portfolio Optimization and Management*, Frank J Fabozzi Series (John Wiley Sons, Inc., 2007).

[14] R. Rebonato and A. Denev, *Portfolio Management under Stress: A Bayesian-Net Approach to Coherent Asset Allocation* (Cambridge University Press, 2014).