Stockholms universitet

The Entropy Pooling Approach: Incorporating Views on Forecast Distributions

Jan Alexandersson

Masteruppsats i matematisk statistik
Master Thesis in Mathematical Statistics

Masteruppsats 2021:7
Matematisk statistik
Juni 2021

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Mathematical Statistics
Stockholm University
Master Thesis **2021:7**
http://www.math.su.se

The Entropy Pooling Approach: Incorporating Views on Forecast Distributions

Jan Alexandersson*

June 2021

Abstract

An economic scenario generator can be described as a tool to simulate future scenarios of financial markets and provide a density forecast containing relevant macro-economic and financial variables used to make financial decisions. Economists are increasingly requiring the possibility to incorporate their own subjective views on the future market and in this thesis we study the entropy pooling approach, based on work laid out by Atillo Meucci, which allows views to be incorporated on general non-normal markets. The entropy pooling approach alter the forecast density to satisfy the views, while minimizing the change in the distribution, with regards to relative entropy.

We walk through the theoretical foundation of this method and present an analytical solution with the assumption of normality. However, without the assumption of normality we need to resort to a computational approach of the entropy pooling method. In the computational approach, the forecast density is represented by simulated sample points and the density is adjusted by assigning a weight to each sample point. The computational approach, however, causes a loss in convergence and we contribute to the current literature by proposing a method to obtain a small set of sample points, with increased convergence properties, which is useful in situations where significant computational limitations are present.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: janalexandersson97@gmail.com. Supervisor: Pieter Trapman.

Acknowledgements

This master's thesis was written in collaboraltion with Kidbrooke Advisory. I would like to thank Edvard Sjögren and Björn Bergstrand at Kidbrooke Advisory for the extremely valueable discussions and their entusiastic encouragement in the making of this thesis. I would also like to thank my supervisor Pieter Trapman for his guidance throughout the process of writing this thesis.

Contents

1. Introduction

An economic scenario generator (ESG) can be described as, in [PCC+16], as a model or software tool used to produce simulations of future scenarios of the financial market and economic variables. Historically, analytical methods have been the standard in solving risk-management problems. However, analytical methods are only tractable for problems where the underlying distributions are known or can be estimated. A lot of the information gained about the economy is obtained from empirical data. By using ESGs and simulation we are able to overcome many restrictions which occurs with using more analytical methods and allows for analysis of more complex dynamic financial systems. Due to the complexity of the interaction of different risk factors in the market over an extended time horizon, an ESG gives you a tool which is useful to understand the range of possible outcomes and likelihood of different scenarios. For these reasons, economic scenario generators have now become an important tool in solving risk management problems. An ESG has many uses and the tasks range from simulating the impact on European equities from a change in US inflation rate to investigating how a portfolio allocation performs over time under different financial landscapes.

Economists are increasingly demanding the possibility to impose their own outlooks and beliefs on the dynamics of the future market and its development. These subjective views can be stated in the shape of expected values, standard deviations, correlations, tail dependency, expected annualized returns and more. Furthermore, these views can be imposed on both short- and long-term horizons.

Views do not necessarily need to be subjective beliefs on the future, they can also be expectations on the future, imposed by independent experts, which pension companies and financial institutions must satisfy when calculating pension projections and return expectations for their customers. This thesis will focus on views being imposed in this context. The difference between personal views and views imposed as a requirement is that one should factor in the confidence the practitioner has in those views, while views imposed as a requirement can be regarded with full confidence. More specifically, the type of views we aim to satisfy are set by The Council for Return Expectations in Denmark [Rå20].

How views can be incorporated in the ESG may depend on the model used and its applications. Ideally, we would like to optimize the parameters of our model maximize the likelihood of the historical data while minimizing the deviance from the imposed views. However, if this approach is tractable or not depends heavily on the model and the shape of the views. Since there are typically many parameters present in an ESG, analytical expressions for time-varying moments are generally very complicated or may not even be possible to express analytically. Therefore, to incorporate the views by adjusting the parameters in the model it would require brute-force and full re-simulations at each step in the optimization, which is computationally costly, or a solution which satisfies the views which may not even exist.

By allowing for views to be incorporated in the model we can take into account how current macroeconomic events affects the future market. By incorporating the views into the model we can also investigate how views imposed on e.g. a five year horizon from now affects potential outcomes in 60 years.

In this thesis we investigate a method proposed by Meucci [Meu10], which does not change the initial model at all. The method instead converts existing simulated scenarios, from the ESG, to scenarios which satisfies the views. This method generalises to all models which are capable of simulating future scenarios and the idea is to assign a weight to each scenario which represents the probability of obtaining that scenario assuming that the model would satisfy the views. The weights need to be assigned so that the corresponding weighted set of scenarios both satisfy the views while being as close as possible to the original set of scenarios, produced by the ESG.

2. Incorporating views using the entropy pooling approach

2.1 The entropy pooling approach

Meucci [Meu10] presents a method which incorporates views on the market within a prior risk model calibrated to the historical data. This is a generalized Bayesian approach referred to as the entropy pooling (EP) approach. The EP approach is not dependent on the specific model to be used and supports any arbitrary market model, referred to as the prior. Meucci's entropy pooling method blends the unadjusted model (prior) with subjective views to obtain a posterior distribution, which can be used, for instance, for risk management and portfolio optimization. As in the well–known Black-Litterman model [BL90], we interpret the views as opinions that distorts the market, i.e., the prior distribution, and the objective is to obtain the posterior in a way that this distortion imposes the least extra spurious structure as possible. In case of the EP approach the posterior distribution minimizes the relative entropy to the prior. The relative entropy, also known as the Kullback-Leibner divergence, is a natural measure of structure because of the information-theoretical properties of this measure. In information theory, this measure can be interpreted as the information gained when the prior is updated to the posterior, and statistically how easily the posterior can be discriminated from the prior with statistical tests.

The idea behind the EP approach is to shift and adjust the probability mass in the prior distribution so that the imposed views are satisfied. This can be done non-parametrically, i.e., without any parametric assumptions, by representing the prior and posterior in terms of pairs of weights and sample points. The weights are probabilities, which paired with the sample points, represent the posterior as a weighted histogram. This can also be done parametrically by making assumptions on the prior and the posterior, such as assuming normality. However, imposing views under the assumption of normality is already possible in the Black-Litterman model but, according to Meucci [Meu06b], the main problems of the Black-Litterman model is just that of the assumption of normality on the market prior and the views, as well as estimating the parameters on the prior from a non-normal distribution.

2.2 Theoretical foundation

Once again, the objective is to find a posterior distribution which is perfectly aligned with our views while still being as close as possible to the unadjusted density forecast, the prior. The distance between the posterior and the prior is quantified by the relative entropy, defined as

$$\varepsilon\big(\tilde{f}, f\big) = \int \tilde{f}(y)\Big[\ln \tilde{f}(y) - \ln f(y)\Big]dy, \tag{1}$$

where $\tilde{f}(y)$ denotes the posterior distribution and f the prior. In the discrete case the relative entropy is defined as

$$\varepsilon\big(\tilde{f}, f\big) = \sum \tilde{f}(y)\Big[\ln \tilde{f}(y) - \ln f(y)\Big]. \tag{2}$$

However, it should be noted that the relative entropy is not a true measure of distance since it is not symmetric, that is, $\varepsilon\big(p, q\big) \neq \varepsilon\big(q, p\big)$.

Furthermore, let V be the set containing all possible distributions which satisfies the constraints imposed by the views. The target posterior will thus be contained in V, assuming V is non-empty. That is

$$\tilde{f} \in V.$$

For example, given views consisting of some set values for expected values, standard deviations and correlations, V would contain all possible distributions which satisfies these constraints. The posterior $\tilde{f}$, given by the minimum relative entropy distribution, solves

$$\tilde{f} = \underset{g \in V}{\operatorname{argmin}}\ \varepsilon\big(g, f\big). \tag{3}$$

The posterior follows $\tilde{f}$ granted that the practitioner has full confidence in his views. If not, the posterior needs to be reduced towards the prior distribution set by the market. This is achieved by opinion-pooling the prior model and the full-confidence posterior by

$$\tilde{f}^c \equiv (1-c)f + c\tilde{f},$$

where the pooling parameter $c \in [0, 1]$ corresponds to the level of confidence in the views. There are many different ways to specify confidence as described in [Meu10]. For example, different confidence in individual views, or multiple practitioners with equal or different levels in their views. However, our focus will be on only the full-confidence posterior due to the context of views being imposed according to requirements which must be satisfied.

2.3 Analytical solution for multivariate normal distributions

If a normal distribution is assumed for the prior, then the analytic solution for the problem set up in Section 2.2 was derived in [Meu10], where the views were set on equalities with regards to expected values and covariances. More general views for this case, i.e when assuming normality, was studied later in [MAK13]. We will now demonstrate the EP approach, with the assumption of normality, with views on linear combinations on the variables with regards to expected values and covariances.

Consider the multivariate normal prior

$$\boldsymbol{X} \sim N(\mu, \boldsymbol{\Sigma}),$$

where μ is a mean vector and $\boldsymbol{\Sigma}$ is a covariance matrix.

Furthermore, consider imposing views on the expectations and covariances of arbitrary linear combinations $\boldsymbol{QX}$ and $\boldsymbol{GX}$ respectively, that is

$$V : \begin{cases} \tilde{E}[\boldsymbol{QX}] = \tilde{\mu}_{\boldsymbol{Q}} \\ \tilde{C}ov[\boldsymbol{GX}] = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}, \end{cases}$$

where $\boldsymbol{Q}$, $\boldsymbol{G}$, $\tilde{\mu}_{\boldsymbol{Q}}$, $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}$ are conformable matrices and vectors. That is, if μ is a $n \times 1$ vector, then $\boldsymbol{Q}$ and $\boldsymbol{G}$ are matrices of size $q \times n$ and $g \times n$ respectively, where q and g is the number of views on the expected values and covariances. The resuting mean vector $\tilde{\mu}_{\boldsymbol{Q}}$ and covariance matrix $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}$ are thus of size $q \times 1$ and $g \times g$ respectively.

As shown in [Meu10], the posterior distribution is normally distributed with

$$\tilde{\mu} = \mu + \boldsymbol{\Sigma}\boldsymbol{Q}^T(\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^T)^{-1}(\tilde{\mu}_{\boldsymbol{Q}} - \boldsymbol{Q}\mu), \tag{4}$$

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\boldsymbol{G}^T\Big((\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^T)^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}(\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^T)^{-1} - (\boldsymbol{G}\boldsymbol{\Sigma}\boldsymbol{G}^T)^{-1}\Big)\boldsymbol{G}\boldsymbol{\Sigma}. \tag{5}$$

Note that if we let Q and G be identity matrices, then (4) and (5) simply reduces to

$$\tilde{\mu} = \mu + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}(\tilde{\mu}_{\boldsymbol{Q}} - \mu) = \tilde{\mu}_{\boldsymbol{Q}},$$

$$\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\Big(\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\Big)\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}}.$$

An example where this formula is applied is given in Section 3.1 and illustrated in Figure 4, where the analytical solution is compared with the solution given by the computational approach.

3.1 Panel representation and linear constraints

As discussed in the previous section, an analytical solution to the optimization problem can generally not be found, except for special cases, e.g. when assuming normality. Moreover, we may not even have an analytical expression for the prior distribution available. In general, one therefore needs to resort to a computational approach.

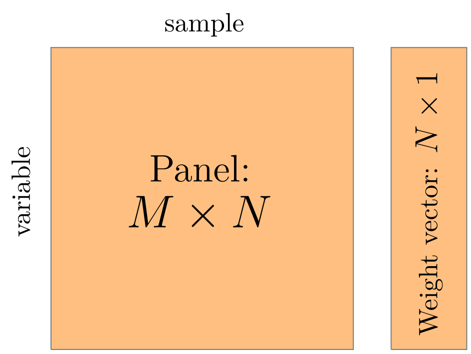


Figure 1: Illustration of the panel structure of the sample points and the corresponding weight vector.

Assume that we can obtain a sample from the forecast distribution which is represented by sample points S_{ij} where $i = 1, \dots, M$ denote the variable and $j = 1, \dots, N$ denote the sample point. That is, N is the sample size. In the prior distribution, each sample point is assumed to be equally relevant and thus each sample point is assigned equal weight, $w_j = 1/N$. An illustration of this structure and of the sample points are shown in Figure 1. The panel data and the weight vector then can be combined into corresponding weighted histograms. To clarify, the weight is the same for each variable in the sample point.

Now also assume the the posterior distribution, $\tilde{f}$, can be represented by the same sample points, but with adjusted weights, $\tilde{\boldsymbol{w}}$. The relative entropy between the prior and the posterior can now we expressed as

$$\varepsilon\big(\tilde{\boldsymbol{w}}, \boldsymbol{w}\big) := \sum_{j=1}^{N} \tilde{w}_j \big[\ln \tilde{w}_j - \ln w_j \big], \tag{6}$$

which is an approximation of the true relative entropy between $\tilde{f}$ and f, since we only have samples of the forecast distribution available.

Also assume that the constraints which is introduced by the views can be expressed as linear equalities or inequalities, that is

$$V = \{\tilde{\boldsymbol{w}} : \boldsymbol{A}\tilde{\boldsymbol{w}} = \boldsymbol{a}, \boldsymbol{B}\tilde{\boldsymbol{w}} \leq \boldsymbol{b}\}. \tag{7}$$

By this assumption, we obtain a convex optimization problem for the adjusted weights $\tilde{w}$, with linear constraints, which can be written as

$$\begin{cases} \min_{\tilde{\boldsymbol{w}}} \tilde{\boldsymbol{w}}^T(\log \tilde{\boldsymbol{w}} - \log \boldsymbol{w}), \\ \tilde{\boldsymbol{w}} > 0, \\ A\tilde{\boldsymbol{w}} = a, \\ B\tilde{\boldsymbol{w}} \leq b, \end{cases} \tag{8}$$

where the first row in $\boldsymbol{A}$ and first element in $\boldsymbol{a}$ represent the constraint of the sum of the weights:

$$\sum_{j=1}^{N} \tilde{w}_j = 1.$$

It is of importance that the imposed views can be written as linear constraints since then the optimization problem can be solved very efficiently, as shown in Section 3.2. However, the restriction of only allowing linear constraints on the weights limit us to only impose views on statistics which can be expressed as linear constraints. While we focus on views as linear constraints for efficiency purposes, this is not a requirement.

It is common to have views on the expected value, variance and correlations and it will now be shown how views on these statistics can be written as linear constraints. That is, how to construct $\boldsymbol{A}$ and $\boldsymbol{a}$ in (7). More views which can be written as linear constraints, e.g. the conditional value at risk, are discussed in Section 6.

When the distribution is represented by sample points, the expected value of the posterior distribution can be written as

$$E[\tilde{\boldsymbol{X}}] = \boldsymbol{S}\tilde{\boldsymbol{w}},$$

where $\boldsymbol{S}$ is the matrix containing the sample points and $\tilde{\boldsymbol{X}}$ is a random variable which is represented by $\tilde{f}$. We can now impose views on the expected values by setting the mean vector to some value $\boldsymbol{u}$ by introducing the linear constraint

$$\boldsymbol{S}\tilde{\boldsymbol{w}} = \boldsymbol{u}. \tag{9}$$

So in this case $\boldsymbol{A} = \boldsymbol{S}$ and $\boldsymbol{a} = \boldsymbol{u}$. More generally, we can set constraints on the expected value of a function of the view-adjusted forecast $\tilde{\boldsymbol{X}}$ as

$$E[g(\tilde{\boldsymbol{X}})] = g(\boldsymbol{S})\tilde{\boldsymbol{w}},$$

where $g : \mathbb{R}^M \to \mathbb{R}^K$, $M \geq K$, and g is applied to each column of $\boldsymbol{S}$.
The variance of $\tilde{X}_i$ can be expressed as

$$Var(\tilde{X}_i) = E[\tilde{X}_i^2] - E[\tilde{X}_i]^2 = (\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\tilde{\boldsymbol{w}} - (\boldsymbol{S}_{i\bullet}\tilde{\boldsymbol{w}})^2,$$

where $\boldsymbol{S}_{i\bullet}$ denotes the matrix which only consists of the row i in $\boldsymbol{S}$ and $\circ$ denotes element wise multiplication, known as Hadamard notation. To clarify, $\boldsymbol{S}_{i\bullet}$ is a $1 \times N$ matrix. This expression is clearly not linear in $\tilde{\boldsymbol{w}}$, so to set the variance to some value v, according to the view, we first need to fix the expected value which gives the linear constraints

$$\boldsymbol{S}_{i\bullet}\tilde{\boldsymbol{w}} = \boldsymbol{S}_{i\bullet}\boldsymbol{w}, \tag{10a}$$

$$(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\tilde{\boldsymbol{w}} = (\boldsymbol{S}_{i\bullet}\boldsymbol{w})^2 + v, \tag{10b}$$

where $\boldsymbol{w}$ is the vector of unadjusted weights all taking value $1/N$. In this case we fixed the expected value to be the same expected value as for the prior. If we want to impose views on both the expected value and the variance for $\tilde{X}_i$ to fix values u and v respecively, we write the constraints as

$$\boldsymbol{S}_{i\bullet}\tilde{\boldsymbol{w}} = u, \tag{11a}$$

$$(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\tilde{\boldsymbol{w}} = u^2 + v. \tag{11b}$$

In this case we construct A and a by stacking the constraints as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{S}_{i\bullet} \\ \boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet}, \end{pmatrix} \quad \text{and } \boldsymbol{a} = \begin{pmatrix} u \\ u^2 + v \end{pmatrix}.$$

The covariance between $\tilde{X}_i$ and $\tilde{X}_j$ is given by

$$Cov(\tilde{X}_i, \tilde{X}_j) = (\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{j\bullet})\tilde{\boldsymbol{w}} - (S_{i\bullet}\tilde{\boldsymbol{w}})(S_{j\bullet}\tilde{\boldsymbol{w}}).$$

Similarly as for the variance, we can write the view for the covariance as a linear constraint by again first fixing the expected value. Setting the covariance to a value c according to the views gives the constraints

$$\boldsymbol{S}_{i\bullet}\tilde{\boldsymbol{w}} = \boldsymbol{S}_{i\bullet}\boldsymbol{w}, \tag{12a}$$

$$\boldsymbol{S}_{j\bullet}\tilde{\boldsymbol{w}} = \boldsymbol{S}_{j\bullet}\boldsymbol{w}, \tag{12b}$$

$$(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\tilde{\boldsymbol{w}} = (\boldsymbol{S}_{i\bullet}\boldsymbol{w})(\boldsymbol{S}_{j\bullet}\boldsymbol{w}) + c. \tag{12c}$$

Furthermore, by also fixing the variance we can also obtain a linear constraint for the correlation. In this case we set the expected values and the variances to be unchanged, compared to the prior. We set the correlation to some value v, according to the views, by the following equalities

$$\boldsymbol{S}_{i\bullet}\tilde{\boldsymbol{w}} = \boldsymbol{S}_{i\bullet}\boldsymbol{w}, \tag{13a}$$

$$\boldsymbol{S}_{j\bullet}\tilde{\boldsymbol{w}} = \boldsymbol{S}_{j\bullet}\boldsymbol{w}, \tag{13b}$$

$$(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\tilde{\boldsymbol{w}} = (\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\boldsymbol{w}, \tag{13c}$$

$$(\boldsymbol{S}_{j\bullet} \circ \boldsymbol{S}_{j\bullet})\tilde{\boldsymbol{w}} = (\boldsymbol{S}_{j\bullet} \circ \boldsymbol{S}_{j\bullet})\boldsymbol{w}, \tag{13d}$$

$$\begin{aligned}(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{j\bullet})\tilde{\boldsymbol{w}} &= (\boldsymbol{S}_{i\bullet}\boldsymbol{w})(\boldsymbol{S}_{j\bullet}\boldsymbol{w}) \\ &\quad + v\sqrt{(\boldsymbol{S}_{i\bullet} \circ \boldsymbol{S}_{i\bullet})\boldsymbol{w} - (\boldsymbol{S}_{i\bullet}\boldsymbol{w})^2}\sqrt{(\boldsymbol{S}_{j\bullet} \circ \boldsymbol{S}_{j\bullet})\boldsymbol{w} - (\boldsymbol{S}_{j\bullet}\boldsymbol{w})^2}.\end{aligned} \tag{13e}$$

In order to give an idea on how the weights are adjusted, given different types of views, consider the prior to be an uncorrelated bivariate standard normal distribution, that is

$$\boldsymbol{X} \sim N(\mu, \boldsymbol{\Sigma}), \qquad \text{where } \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

We can see in Figure 2 that a view with increased mean for X_1 results in larger weights for large observations, increased variance of X_1 and X_2 results in larger weights in the tails of each distribution and increased correlation results in larger weights along the diagonal.

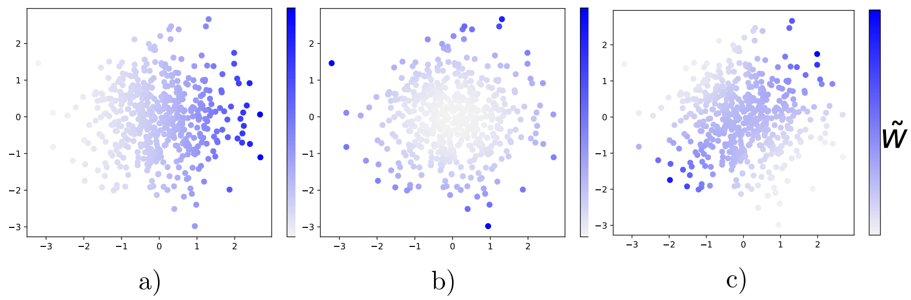


Figure 2: Illustrations on how different views affect the weight allocation for each scenario. Figure a) has a view of increased mean, figure b) increased variance and figure c) increased correlation. Darker color implies larger weight.

In Figure 3 we can we see the difference between the prior and the posterior distribution for X_1, i.e. the weighted histogram with weights $\boldsymbol{w}$ (equally weighted) and the weighted histogram with weights $\tilde{\boldsymbol{w}}$. We can see the clear similarities in the relative height of the bins in the two histograms, as a result of keeping the distributions as alike as possible by minimizing the relative entropy. As expected when imposing a view of a higher expected value, the positive outcomes are assigned more weight. However, the posterior is not obtained by simply moving the prior to the right. Instead, the probability mass are rearranged to represent the views and keep the domain of the prior distribution. One problem which can occur is if we want to impose extreme views, e.g. imposing a view of very high expected value. The view might be outside the range of the existing scenarios and thus a solution does not exist using the available scenarios. However, for the example of imposing a view of large expected value, the posterior distribution turns out to be approximately normally distributed, which was proved in [Sme16]. It was proved that all minimum relative entropy distributions with a large mean are asymptotically normally distributed.

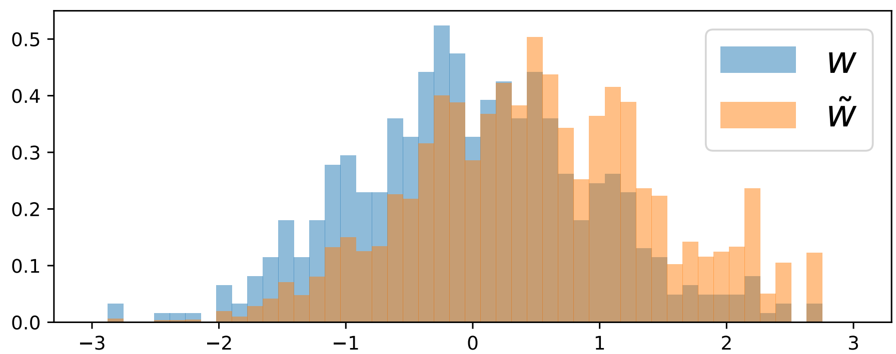


Figure 3: Histograms of the prior and the posterior distribution. In the prior distribution, each scenario is weighted equally, i.e by $\boldsymbol{w}$, and in the posterior distribution each scenarios is weighted by $\tilde{\boldsymbol{w}}$.

To demonstrate how well the computational approach works we recreate the example displayed in [Meu10], which gives a comparison between the analytical posterior and the numerical posterior. Consider the following:

Let

$$\boldsymbol{X} \sim N\left(\mu, \boldsymbol{\Sigma}\right), \qquad \text{where } \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}. \tag{14}$$

Furthermore, let $\boldsymbol{Q} = (1, 0)$, $\boldsymbol{G} = (1, 0)$, $\tilde{\mu}_{\boldsymbol{Q}} = 0.5$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{G}} = 0.1$. That is, impose the views $E[X_1] = 0.5$ and $Var(X_1) = 0.1$. The analytical solution is obtained by inserting the numbers into (4) and (5) which results in

$$\tilde{\mu} = \begin{pmatrix} 0.5 \\ 0.4 \end{pmatrix} \text{ and } \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} 0.01 & 0.008 \\ 0.008 & 0.3664. \end{pmatrix}. \tag{15}$$

In Figure 4 we can see that numerical approach follow the theoretical when we go from the prior to the posterior, even with extreme views such as dividing the standard deviation by a factor of 10.

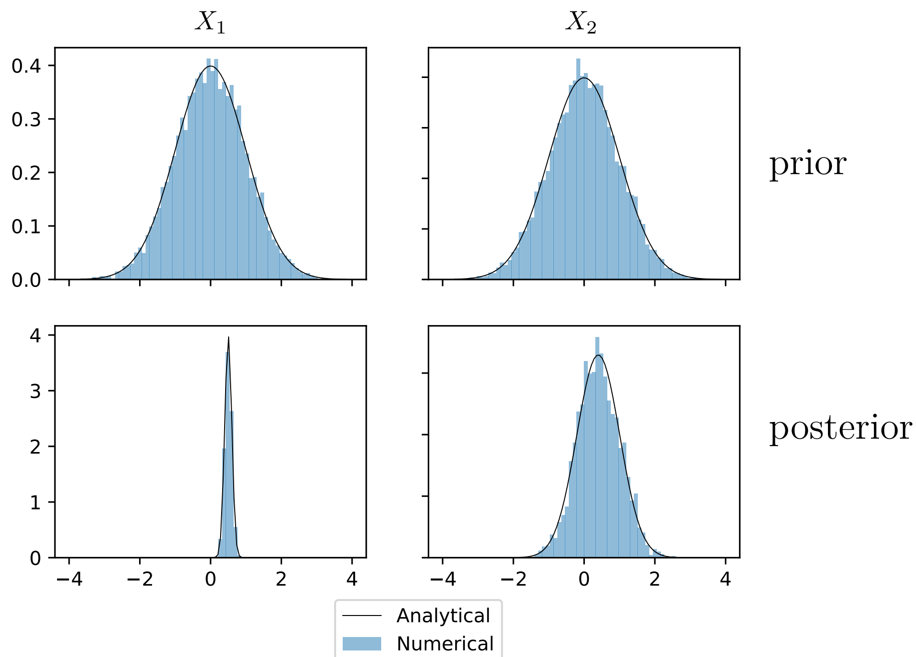


Figure 4: Numerical approach follows the analytical solution for normal distributions.

3.2 Finding the solution to the optimization problem

When the matrix and vector pair $\boldsymbol{A}$ and $\boldsymbol{a}$ consisting of the collected equality constraints and $\boldsymbol{B}$ and $\boldsymbol{b}$ the inequality constraints, the entropy minimization reads as

$$\tilde{\boldsymbol{w}} = \underset{\substack{\boldsymbol{Ax}=\boldsymbol{a} \\ \boldsymbol{Bx}\leq\boldsymbol{b}}}{\operatorname{argmin}} \left(\sum_{j=1}^{N} x_j (\ln x_j - \ln w_j) \right). \tag{16}$$

We will later see that the required inequality constraint $x \geq 0$ will be automatically satisfied and therefore does not need to be specified.
However, since $w_j = 1/N$ for $j = 1, ..., N$ we have that

$$\sum_{j=1}^{N} x_j (\ln x_j - \ln w_j) = \sum_{j=1}^{N} x_j \ln x_j - \sum_{j=1}^{N} x_j \ln \frac{1}{N} = \sum_{j=1}^{N} x_j \ln x_j + \ln N,$$

where we used that $\sum_{j=1}^{N} x_j = 1$. We see that minimizing

$$\sum_{j=1}^{N} x_j(\ln x_j - \ln w_j)$$

is the same as minimizing

$$\sum_{j=1}^{N} x_j \ln x_j,$$

commonly known as the negative entropy.

This is a convex optimization problem and these types of optimization problems have been studied extensively. Various entropy optimization problems are shown in Chapter 5 in [BV04] and we will solve this optimization problem accordingly by deriving the Lagrange dual function making use of the conjugate function.

Let $f_0(x)$ denote the function we want to minimize and let f_0^* denote the conjugate to f_0, defined as

$$f_0^*(y) = \sup_{x \in \mathrm{dom}\ f_0} \left(y^T x - f_0(x) \right). \tag{17}$$

In our case we have that $f_0(x) = \sum_{j=1}^{N} x_j \ln x_j$ and we find the conjugate by

$$\begin{aligned} f_0^*(y) &= \sup_{x \in \mathrm{dom}\ f_0} (y^T x - f_0(x)) \\ &= \sup_{x \in \mathrm{dom}\ f_0} \left(\sum_{j=1}^{N} x_j y_j - \sum_{j=1}^{N} x_j \ln x_j \right) \\ &= \sum_{j=1}^{N} \sup_{x \in \mathrm{dom}\ f_0} \left(x_j y_j - x_j \ln x_j \right). \end{aligned} \tag{18}$$

Let $g(x_j) = x_j y_j - x_j \ln x_j$ and $x_j^* = \underset{x_j}{\mathrm{argmax}}\ g(x_j)$. By the first order conditions for x_j we have

$$0 = g'(x_j^*) = y_j - \ln x_j^* - 1 \iff x_j^* = e^{y_j - 1}. \tag{19}$$

Substituting this into (18) gives

$$f_0^*(y) = \sum_{j=1}^{N} y_i e^{y_j - 1} - e^{y_j - 1} \ln e^{y_j - 1} = \sum_{j=1}^{N} e^{y_j - 1}. \tag{20}$$

We can now set up the Lagrangian

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\lambda,\nu) &= \boldsymbol{x}^T \ln \boldsymbol{x} + \lambda^T(\boldsymbol{B}\boldsymbol{x}-\boldsymbol{b}) + \nu^T(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{a}) \\ &= \boldsymbol{x}^T \ln \boldsymbol{x} + (\boldsymbol{B}^T\lambda + \boldsymbol{A}^T\nu)^T\boldsymbol{x} - \boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu.\end{aligned} \tag{21}$$

and find the dual function using the conjugate function we derived above by

$$\begin{aligned}\mathcal{G}(\lambda,\nu) &= \inf_x \left(\boldsymbol{x}^T \ln \boldsymbol{x} + (\boldsymbol{B}^T\lambda + \boldsymbol{A}^T\nu)^T\boldsymbol{x} - \boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu\right) \\ &= -\boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu + \inf_x \left(\boldsymbol{x}^T \ln \boldsymbol{x} + (\boldsymbol{B}^T\lambda + \boldsymbol{A}^T\nu)^T\boldsymbol{x}\right) \\ &= -\boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu - \sup_x \left((-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu)^T\boldsymbol{x} - \boldsymbol{x}^T \ln \boldsymbol{x}\right) \\ &= -\boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu - f_0^*(-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu) \\ &= -\boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu - \sum_{j=1}^{N} e^{-\boldsymbol{B}^T_{\bullet j}\lambda - \boldsymbol{A}^T_{\bullet j}\nu - 1},\end{aligned} \tag{22}$$

where $\boldsymbol{B}^T_{\bullet j}$ and $\boldsymbol{A}^T_{\bullet j}$ denotes the transpose of the j-th column in the matrices respectively.

The dual function can also be found by first identifying the first order conditions for $\boldsymbol{x}$, given by

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = \ln \boldsymbol{x} + \mathbf{1} + \boldsymbol{B}^T\lambda + \boldsymbol{A}^T\nu \equiv 0, \tag{23}$$

with the solution given by

$$\boldsymbol{x}(\lambda,\nu) = e^{-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu - \mathbf{1}}, \tag{24}$$

as shown in [Meu10].

We can now derive the dual function as

$$\begin{aligned}\mathcal{G}(\lambda,\nu) &= \mathcal{L}(\boldsymbol{x}(\lambda,\nu),\lambda,\nu) \\ &= \left(e^{-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu - \mathbf{1}}\right)^T \left(-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu - \mathbf{1}\right) \\ &\quad + (\boldsymbol{B}^T\lambda + \boldsymbol{A}^T\nu)^T \left(e^{-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu - \mathbf{1}}\right) - \boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu \\ &= -\boldsymbol{b}^T\lambda - \boldsymbol{a}^T\nu - \left(e^{-\boldsymbol{B}^T\lambda - \boldsymbol{A}^T\nu - \mathbf{1}}\right)^T \mathbf{1}.\end{aligned} \tag{25}$$

The optimal Lagrange multipliers are found by maximizing the dual function numerically

$$(\lambda^*, \nu^*) = \underset{\lambda \geq 0, \nu}{\operatorname{argmax}} \ \mathcal{G}(\lambda, \nu). \tag{26}$$

Note that the Lagrangian dual function $\mathcal{G}(\lambda, \nu)$ should be maximized whereas the Lagrangian should be minimized. The solution is finally be obtained by

$$\tilde{\boldsymbol{w}} = \boldsymbol{x}(\lambda^*, \nu^*). \tag{27}$$

Important to note is that the optimization is performed for limited number of parameters equal to the number of views and not directly on the weights, which are the variables of interest. The number of weights, equal to the number of simulated scenarios, could be very large and optimization may not be numerically feasible if performed directly on the weights. Instead, by performing the optimization on the Lagrange multipliers we guarantee numerical feasibility. To add to this, we can also compute the gradient of $\mathcal{G}(\lambda, \nu)$ as

$$\begin{aligned}
\frac{\partial \mathcal{G}(\lambda, \nu)}{\partial \lambda_i} &= -b_i - \sum_{j=1}^{N} -B_{ij} e^{-\boldsymbol{B}_{\bullet j}^T \lambda - \boldsymbol{A}_{\bullet j}^T \nu - 1}, \\
\frac{\partial \mathcal{G}(\lambda, \nu)}{\partial \nu_k} &= -a_k - \sum_{j=1}^{N} -A_{kj} e^{-\boldsymbol{B}_{\bullet j}^T \lambda - \boldsymbol{A}_{\bullet j}^T \nu - 1},
\end{aligned}$$

and the Hessian, given by

$$\begin{aligned}
\frac{\partial^2 \mathcal{G}(\lambda, \nu)}{\partial \lambda_i^2} &= -\sum_{j=1}^{N} B_{ij}^2 e^{-\boldsymbol{B}_{\bullet j}^T \lambda - \boldsymbol{A}_{\bullet j}^T \nu - 1}, \\
\frac{\partial^2 \mathcal{G}(\lambda, \nu)}{\partial \nu_k^2} &= -\sum_{j=1}^{N} A_{kj}^2 e^{-\boldsymbol{B}_{\bullet j}^T \lambda - \boldsymbol{A}_{\bullet j}^T \nu - 1}, \\
\frac{\partial^2 \mathcal{G}(\lambda, \nu)}{\partial \lambda_i \nu_k} &= -\sum_{j=1}^{N} A_{kj} B_{ij} e^{-\boldsymbol{B}_{\bullet j}^T \lambda - \boldsymbol{A}_{\bullet j}^T \nu - 1}.
\end{aligned}$$

The gradient and the Hessian is used to speed up the numerical optimization of the dual function. With the gradient available we can find the minimum using gradient descent and with the Hessian also available we can use Newton's method in optimization to even further increase the speed of the optimization. The gradient and Hessian can also be written as

$$\mathcal{G}'(\lambda,\nu) = \begin{pmatrix} -\boldsymbol{b} + \boldsymbol{B}\cdot\boldsymbol{x}(\lambda,\nu) \\ -\boldsymbol{a} + \boldsymbol{A}\cdot\boldsymbol{x}(\lambda,\nu) \end{pmatrix} \tag{30}$$

and

$$\mathcal{G}''(\lambda,\nu) = \begin{pmatrix} -\boldsymbol{B}\cdot((\boldsymbol{x}(\lambda,\nu)\cdot\mathbb{1})\circ\boldsymbol{B}^T) & -\boldsymbol{B}\cdot((\boldsymbol{x}(\lambda,\nu)\cdot\mathbb{1})\circ\boldsymbol{A}^T) \\ -\boldsymbol{A}\cdot((\boldsymbol{x}(\lambda,\nu)\cdot\mathbb{1})\circ\boldsymbol{B}^T) & -\boldsymbol{A}\cdot((\boldsymbol{x}(\lambda,\nu)\cdot\mathbb{1})\circ\boldsymbol{A}^T) \end{pmatrix}. \tag{31}$$

4.1 Introducing scenarios

The EP approach presented in [Meu10] is sufficient if the practitioner is only interested in a certain time point in the future. That is, how his views at the h-step forecast affects the h-step forecast distribution. However, a natural question which arises is, for example, how the views on the 5-step forecast affect the 10-step forecast, or the 60-step forecast. It may also be desirable to have views on multiple time periods at the same time.

Meucci and Nicolosi (2016) presents a theoretical foundation of a time-dependent method of Meucci's entropy pooling approach. This was later extended by [vdS19] with a computational approach for the time-dependent setting.

In the time-dependent setting we, instead of sample points, consider scenarios

$$S_{ij}(t),$$

where the time dimension is added. Previously each sample point had a corresponding weight, but in the time-dependent setting the weights $\tilde{w}_j$ are assigned to each scenario. That is, the weight for scenario j is the same for each variable and for each time point t. An illustration of this representation is shown in Figure 5. In this setting, the optimization problem remains the same as before, i.e. the weights are determined by minimizing the relative entropy subject to linear constraints. Notably, this method does not introduce any new scenarios, only changing the weights of the existing scenarios, analogous to how no new sample points were introduced in the one-period setting.

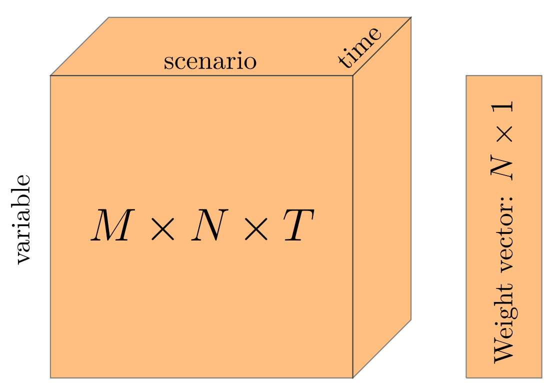


Figure 5: Illustration of the structure of the scenarios and the corresponding weight vector.

Even though we now consider a time-dependent setting, the statistics we impose views on should be written as functions of samples, as in the one-period setting. This method still supports all linear constraints that can be written as $\boldsymbol{A}\tilde{\boldsymbol{w}} = \boldsymbol{a}$ and $\boldsymbol{B}\tilde{\boldsymbol{w}} \leq \boldsymbol{b}$. As an example, suppose that the following view is imposed: the mean of variable i at time point $t = 5$ is v. The corresponding constraint is thereby written as $\boldsymbol{S}_{i\bullet}(5)\tilde{\boldsymbol{w}} = v$.

4.2 Effective number of scenarios

In the computational approach, extreme views may result in large weights for only some of the scenarios while other scenarios are assigned very small weights. A natural question needs to be addressed: how does the changing of weights affect the statistical significance of statistics derived from the re-weighted scenarios? In this section we define the *effective number of scenarios*, see [Meu12], as a measure on how much the statistical significance is affected as the weights are adjusted.

First, let $\tilde{N}$ denote the effective number of scenarios, to be defined below. In the simple case where all scenarios are equally weighted, $\tilde{w}_j = 1/N, j = 1, ..., N$, we rely on all N available scenarios equally, that is $\tilde{N} = N$. However, in the extreme case where one scenario is assigned all weight and the remaining scenarios are assigned zero weight, any statistic computed relies only on one scenario, i.e. $\tilde{N} = 1$.

Meucci defines the effective number of scenarios (ENS) as the exponential of the entropy of the weights

$$\tilde{N} = e^{-\sum_{j=1}^{N} \tilde{w}_j \ln \tilde{w}_j}. \tag{32}$$

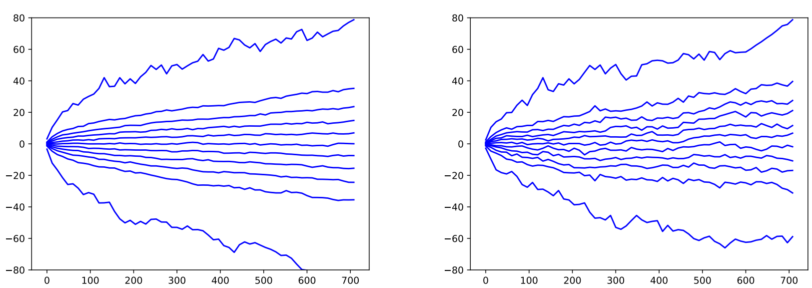
The entropy measure is a natural choice due to it's interpretation. It is a measure of the concentration of the probability mass in the weights and in information theory the entropy can be described as the uncertainty of the possible outcomes of a random variable. The effective number of scenarios will be a value between 1 and N. If the adjustment made to the density forecast is small, the measure will be close to N. If the adjustment is extreme the effective number will be closer to 1, granted that the adjustment needed to satisfy the view is possible. Depending on the application, the minimum effective number of scenarios varies. For example, an effective number of 50 might be sufficient to estimate the mean of the adjusted forecast distribution, but is likely to be insufficient to give an estimate of the 99.5% value at risk which is accurate.

A tempting idea would be to increase the lower bound of the weights to a constant larger than zero, with the idea is that the weights would have a lower variation, i.e higher entropy. If this was the case we would gain more effective number of scenarios while sacrificing accuracy, with regards to the relative entropy. However, increasing the lower bound of the weights has the opposite effect. Scenarios with high weights are assigned even more weight and therefore have an even higher impact than before. In fact, the optimization problem which yields the maximum effective number of scenarios is equivalent to the minimization of the relative entropy, since

$$
\begin{aligned}
\tilde{N} \equiv e^{-\sum_{j=1}^{N} \tilde{w}_j \ln \tilde{w}_j} &= e^{-\sum_{j=1}^{N} \tilde{w}_j \ln \tilde{w}_j - \tilde{w}_j \ln w_j + \tilde{w}_j \ln w_j} \\
&= e^{-\varepsilon(\tilde{\boldsymbol{w}},\boldsymbol{w}) + \sum_{j=1}^{N} \tilde{w}_j \ln \frac{1}{N}} = e^{-\varepsilon(\tilde{\boldsymbol{w}},\boldsymbol{w}) + \ln \frac{1}{N}} = \frac{1}{N} e^{-\varepsilon(\tilde{\boldsymbol{w}},\boldsymbol{w})},
\end{aligned}
\tag{33}
$$

where we used that the $\sum_{j=1}^{N} \tilde{w}_j = 1$ and $w_j = \frac{1}{N}$. We see that minimizing $\varepsilon(\tilde{\boldsymbol{w}}, \boldsymbol{w})$ also minimizes $\tilde{N}$.

To illustrate the importance of the effective number of scenarios we simulate 1000 uncorrelated bivariate Gaussian random walks to act as our scenarios. In Figure 6a we can see the deciles of the marginal distribution over time for the first variable, where the imposed view adjusted the correlation between the two vaiables to 0.1 at time point 120 and in Figure 6b the correlation was adjusted to 0.7. In Figure 6b the variance for the second marginal distribution was also adjusted to 0.75 from 1. For a), the effective number of scenarios for the forecast distribution declined from 1000 to 997.7 and for b) to 92.7. Although, for b), the view is rather unrealistic but the example clearly showcases the importance of the effective number of scenarios. We can see that the deciles of the forecast distribution fluctuate significantly more over time in b) and therefore are less reliable. For statistics such as percentiles and other statistics sensitive to a low number of sample points, e.g. value at risk, the effective number of scenarios is too low.



a) b)

Figure 6: Deciles for the marginal distribution of a bivariate Gaussian random walk when the correlation is adjusted from 0 to 0.1 in a) and to 0.8 in b), as well as the variance of the corresponding second marginal distribution from 1 to 0.75.

4.3 Discrete weights

When one has obtained the weights which minimizes the relative entropy, the weights have to be taken into account in every application thereafter, e.g. when calculating expected future cash flow or optimal investing strategies. However, many calculations in portfolio allocation applications and risk management are made with the assumption of equally weighted scenarios. It is therefore desirable to be able to produce a density forecast which can be represented by scenarios which are equally weighted rather than modifying each application to take the weights into account. If such representation can be obtained, the scenarios can be used as previously in each application, without the need of adjusting the computations to account for the weights.

The objective is to convert the weights $\tilde{w}_j$, $j = 1, ..., N$, to discrete integer weights $\hat{w}_j$. Instead of summing up to one, the discrete weights sum to a specified integer $\hat{N}$. The multivariate posterior density forecast can then be represented by a scenario set $\hat{S}_{ik}(t)$, where $k = 1, ..., \hat{N}$, i.e. containing $\hat{N}$ scenarios, where each scenario $\boldsymbol{S}_{\bullet j}(t)$ are copied $\hat{w}_j$ times. In this representation some of the scenarios are identical but each scenario has the same weight. Thus, the scenario set $\hat{S}_{ik}(t)$ is suitable for calculations where equally weighted scenarios is assumed.

Ideally, one could modify the optimization problem and include the constraint that the weights should be integers which sums to $\hat{N}$. However, this introduces some problems. First, it might not be possible to satisfy all constraints with integer weights. Secondly, integer programming problems are NP-hard and fast solvers does not exist [PS13]. The integer weights thus need to be approximated, which in turn sacrifices some accuracy with regards to satisfying the views, i.e. the views are no longer guaranteed to be satisfied exactly.

The weights $\tilde{w}$ can be converted to discrete weights $\hat{w}$ by first rounding $\hat{N}\tilde{w}_j$ to the nearest integer. However, the resulting weights after the rounding does not necessarily sum up to $\hat{N}$ as specified. Thus, if the weights $\hat{w}$ sum to a number less than $\hat{N}$, increase the weight $\hat{w}_j$, with the largest negative rounding error, by 1 and repeat until the weights sum to $\hat{N}$. Likewise, if the weights $\hat{w}$ sum to a number greater than $\hat{N}$, decrease the weight $\hat{w}_j$, with the largest positive rounding error, by 1.

It should be emphasized that converting the weights to discrete weights may decrease the effective number of scenarios and make the constraints no longer be satisfied exactly. However, with a sufficiently large $\hat{N}$ this effect is small. In fact, $\hat{\boldsymbol{w}}/\hat{N} \to \tilde{\boldsymbol{w}}$ as $\hat{N} \to \infty$.

To demonstrate the importance of the choice of $\hat{N}$, consider a sample of size 5000 from the standard normal distribution, as our prior. Imposing a view for the expected value to be 1 leads to an effective number of scenarios of 3039, using the continuous weights $\hat{\boldsymbol{w}}$. In Figure 7 we can see the effective number of scenarios when using $\hat{\boldsymbol{w}}$ for different choices of $\hat{N}$, in the lower axis. We see that we, in this example, need $\hat{N}$ to be roughly 4 times larger than N to converge to the original effective number of scenarios.

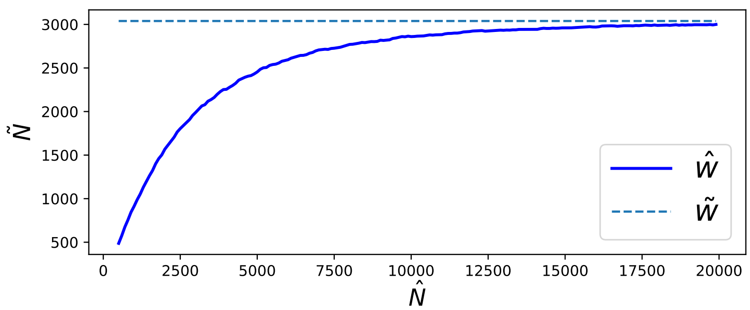


Figure 7: Effective number of scenarios for different choices of $\hat{N}$ when converting continuous weights into discrete integer weights.

A heuristic is suggested in [vdS19] which tries to satisfy the equality constraints as well as possible when using integer weights. The algorithm is provided in Algorithm A1. However, if it is possible, the continuous weights should be used in the succeeding applications of the scenarios to obtain the most accurate results.

5.1 The weighted sampling method

Situations may occur in practice where the number of scenarios representing the posterior distribution has limitation due to computational time. That is, we need the posterior distribution to be represented by $K \leq N$ scenarios. However, applying the EP method to a set of scenarios of size K results in a lower effective number of scenarios compared to using N scenarios. In practice there are also limitations in how low the effective number of scenarios can be to reach convergence in some applications, as discussed in Section 4.2. To clarify, we have an upper limit in the number of scenarios representing the posterior distribution and a lower limit in effective number of scenarios. In some cases, both limitations can not be satisfied simultaneously when applying the EP method as described previously. Our goal is to find a set of scenarios of size K, representing the posterior, with an effective number of scenarios as close to K as possible. The idea, inspired by importance sampling [KM53], is to make use of weighted sampling using the weight vector from the EP approach.

Let N be the number of scenarios representing the prior and let $K \leq N$ be the upper limit of the number of scenarios representing the posterior. Perform the EP approach as previously and obtain the weight vector $\tilde{\boldsymbol{w}}$ of size $N \times 1$. Now, sample K scenarios from the N available scenarios with replacement, where the probability for scenario j to be drawn is equal to the corresponding weight $\tilde{w}_j$. The resulting set of K scenarios now consist of equally weighted scenarios representing the posterior. However, since randomness is introduced, the views will no longer be satisfied exactly but are satisfied on average since we sample from a distribution which satisfies the views exactly by construction. The posterior obtained from the weighted sampling will thus still satisfy the views in expectation. We now need to show that this representation of the posterior has better convergence properties than simply performing the EP approach using K scenarios. We refer to the posterior obtained from weighted sampling as the *weighted sampling posterior* and the posterior represented by the weight vector as the *EP posterior*.

Since there are possibilities of repetitions of scenarios due to sampling with replacement, the effective number of scenarios is not necessarily equal to K, even though in this representation all scenarios are equally weighted. We can compute the effective number of scenarios in the weighted sampling posterior by letting the number of repetitions of each scenario correspond to the weight used in the computation of the effective number of scenarios (32), e.g. if a scenario is repeated twice, the corresponding weight is $2/K$. This gives the effective number of scenarios as

	Weighted sampling posterior	EP posterior
Prior	N	K
Posterior	K	K
ENS	$\tilde{K}_{WS}$	$\tilde{K}$

Table 1: Number of scenarios in the representation of the prior and posterior in the weighted sampling and standard representation as well as the effective number of scenarios for both methods.

$$\exp\left\{-\sum_{j=1}^{N}\frac{x_j}{K}\ln\left(\frac{x_j}{K}\right)\right\}, \tag{34}$$

where x_j is the number of repeated draws of scenario j. In fact, the distribution of the of the number of repetitions is given by the multinomial distribution, conditioning on the weights $\tilde{\boldsymbol{w}}$.

To clarify the number of scenarios represented in the prior and posterior for each method see Table 1. Note that while $\tilde{K}$ is deterministic conditioned on the scenarios, $\tilde{K}_{WS}$ is a random variable due to the weighted sampling.

It can easily be argued that when $N = K$ we have that

$$E[\tilde{K}_{WS}] = \tilde{K},$$

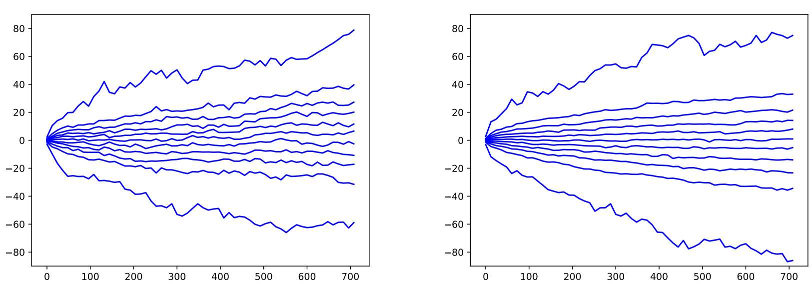
since scenario j in the weighted sampling posterior have expected number of repetitions $K\tilde{w}_j$, which is the expected value of the multinomial distribution. Inserting $x_j = K\tilde{w}_j$ in (34) yields $\tilde{K}$. Furthermore, when letting $N \to \infty$ the probability of having repeating scenarios in the weighted sampling posterior tend to zero and thus $E[\tilde{K}_{WS}] \to K$, since all scenarios are equally weighted and no repetitions are present. This argument assumes that when $N \to \infty$ the number of non-zero weights also tend to infinity, which is the case for most applications in practice.

It is of importance that the sampling is made with replacement to obtain a set of scenarios representing the posterior. Consider the simple case where $N = K$. In this case, sampling without replacement would make the weights $\tilde{\boldsymbol{w}}$ irrelevant since we would always end up with the same set of scenarios, the prior scenarios. Sampling without replacement would introduce a bias towards the prior and the obtained posterior would therefore not satisfy the views.

As an example we again consider the bivariate Gaussian random walk as in Figure 6. We impose drastic views to obtain a low effective number of scenarios to better illustrate the advantage of the weighted sampling method. The setting for this example is shown in Table 2 together with the resulting effective number of scenarios. We can see that we have achieved a significant increase in effective number of scenarios in the weighted sampling posterior than in the EP posterior. We can also visualize this effect by plotting the deciles over time for both methods, which can be found in Figure 8. We see that the deciles fluctuate less in the weighted sampling case which showcases the improved convergence properties when using this method.

	Weighted sampling posterior	EP posterior
Prior	$N = 10000$	$K = 1000$
Posterior	$K = 1000$	$K = 1000$
ENS	$\tilde{K}_{WS} \approx 688$	$\tilde{K} \approx 174$

Table 2: Effective number of scenarios for the weighted sampling method and the standard method. In this example the posterior is represented by 1000 scenarios for both methods.



a) b)

Figure 8: Deciles of the posterior scenarios using the EP method a) and the weighted sampling method b).

The advantage of using the weighted sampling method is, as we have shown, the increased convergence properties. However, another advantage is that it is not necessary to incorporate the weight vector in the evaluations of the posterior scenarios and the scenarios can be used directly, without the need to change the computations to account for the weight vector, as in the case with discrete weights.

The disadvantages are that it is not possible to compute the exact relative entropy between the prior and the posterior since they are represented by sets of scenarios with different sizes. However, this can be approximated by using the histograms of both distributions, where the data in both histograms are grouped into the same bins. Then, the probability mass in each bin is used in the computation of the relative entropy. Another disadvantage is that the views are not satisfied exactly, as in the EP approach, and are instead only satisfied in expectation.

	Weighted sampling posterior	EP posterior	EP posterior
Prior	$N = 10000$	$K = 10000$	$K = 5000$
Posterior	$K = 5000$	$K = 10000$	$K = 5000$
ENS	$\tilde{K}_{WS} \approx 1365$	$\tilde{K} \approx 1767$	$\tilde{K} \approx 901$

Table 3: Effective number of scenarios for the weighted sampling method and the standard method. In this example the posterior is represented by 1000 scenarios for both methods.

To further demonstrate these methods, consider the same example given by (14) and (15). The resulting effective number of scenarios for different choices of N and K, for the weighted sampling method and the regular EP method are given in Table 3. We see that the weighted sampling method increase the effective number of scenarios significantly when comparing the two cases where the posterior distribution is represented by 5000 scenarios. We can also see the effective number of scenarios with the EP approach when keeping all 10000 scenarios also in the posterior. Furthermore, we can see the respective histograms in Figure 9, and see that the weighted sampling posterior correctly follows the EP posterior as well as the analytical solution. In Figure 10 we take a closer look at the posterior of X_2 in the case where the posterior is represented by 5000 scenarios using both methods.

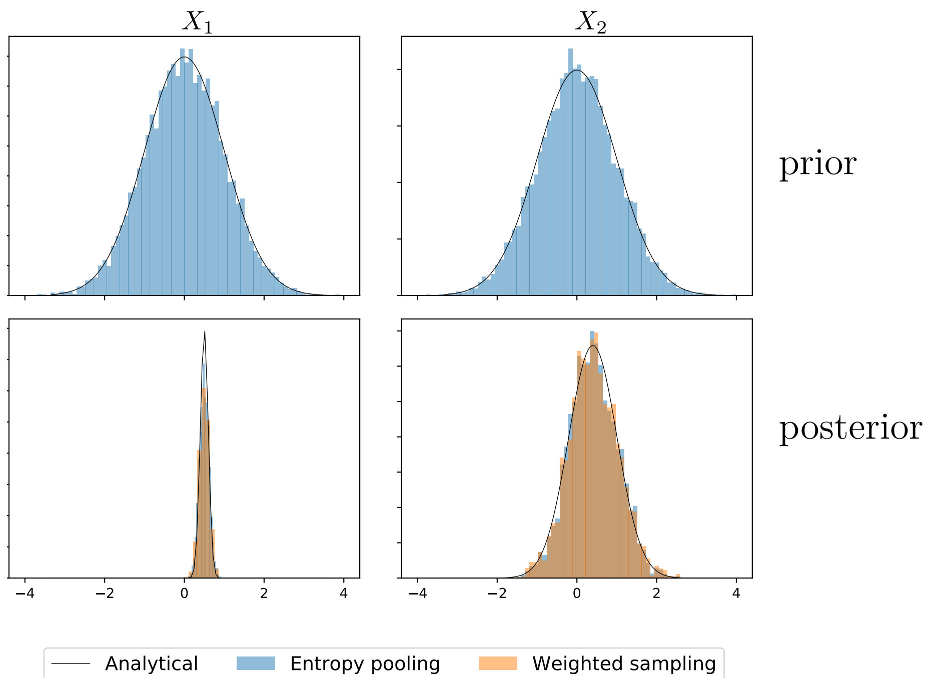


Figure 9: Prior and posterior for the EP approach and the weighted sampling approach. The EP posterior is represented by 10000 scenarios and the resulting effective number of scenarios is 1767. The weighted sampling posterior is represented by 5000 scenarios and the resulting effective number of scenarios is 1365.

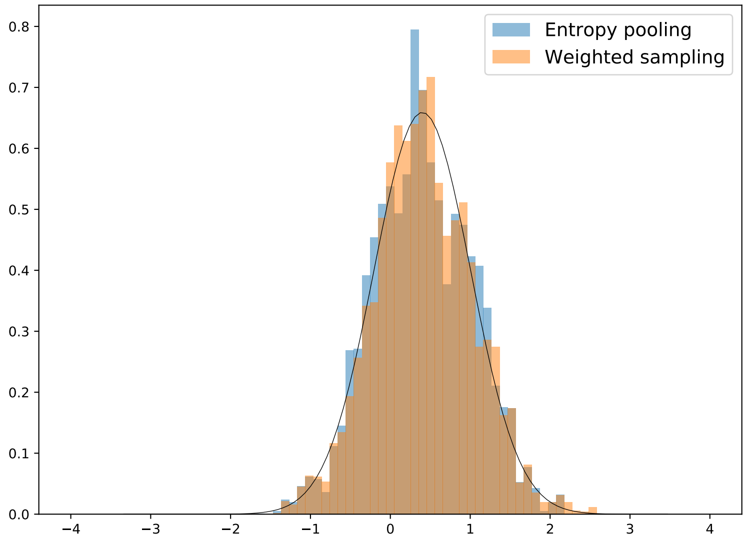


Figure 10: Posterior for the EP approach and the weighted sampling approach. The EP posterior is represented by 5000 scenarios and the resulting effective number of scenarios is 901. The weighted sampling posterior is represented by 5000 scenarios and the resulting effective number of scenarios is 1365.

6. More views

So far we have shown how to impose views on the mean, variance and correlations as linear constraint, which is only a small subset of possible features which the practitioner can impose views on. Examples of other possible views which can be expressed in terms of linear constraints includes views on:

- Median.
- Mean, variance and correlations between cumulative returns over a certain time horizon.
- Mean, variance and correlations between geometrically and arithmetically annualized returns over a certain horizon.
- Value at risk and conditional value at risk.
- Tail behaviour and tail codependence between variables.
- Marginal distributions, by matching moments up to a given order.
- Copula.

In this section we showcase how to impose views on geometrically annualized returns, views on the conditional value at risk as well as discuss how to take rebalancing into account when the views are set.

6.1 Include rebalancing in views

A common procedure in portfolio management is to realign the weightings of the assets in the portfolio to keep a desired level of asset allocation or risk. For example, assume an asset allocation of 50% stocks and 50% bonds. If the stocks perform well and increase to a weighting of 75% in the portfolio, the investor may decide to sell stocks and buy bonds to retain the original asset allocation of 50% stocks and 50% bonds. This procedure is called rebalancing.

It is therefore desirable to be able to impose views on how the portfolio performs in the future, when including the rebalancing process. This can easily be done by introducing a new variable corresponding to the total value of the portfolio where each future scenario for this new variable is constructed by applying the rebalancing procedure to all scenarios.

6.2 Views on geometrically annualized returns

Suppose we want to be able to state views on the geometrically (or arithmetically) annualized returns over some specified horizon. To be aligned with the computational approach we need to be able to express the views as linear constraints and to express this constraint as linear we first need to convert the scenarios to annualized returns.

First, the return for variable i and scenario j over the time horizon T_0 to T is expressed by $\frac{S_{ij}(T)}{S_{ij}(T_0)}$. Thus, we can set the expected return over the horizon to a value u, according to the view, by

$$\left(\boldsymbol{S}_{i\bullet}(T) \oslash \boldsymbol{S}_{i\bullet}(T_0)\right) \tilde{\boldsymbol{w}} = u,$$

where $\oslash$ denotes the Hadamard division (element-wise). For ease of notation, let

$$\boldsymbol{R}_{i\bullet}(T_0, T) = \boldsymbol{S}_{i\bullet}(T) \oslash \boldsymbol{S}_{i\bullet}(T_0)$$

denote the return over the period (T_0, T). Furthermore, the geometrically annualized return for variable i and scenario j, over the horizon, is given by $R_{ij}(T_0, T)^{1/(T-T_0)}$ and thus we can set the view of the annualized return, to a value u, by

$$\boldsymbol{R}_{i\bullet}(T_0, T)\tilde{\boldsymbol{w}} = u^{(T-T_0)}. \tag{35}$$

Notice that it is of importance which operation we do first, annualize or averaging. That is, the difference between the annualized average return

$$\left(\boldsymbol{R}_{i\bullet}(T_0, T)\boldsymbol{w}\right)^{1/(T-T_0)} \tag{36}$$

and the average annualized return

$$\boldsymbol{R}_{i\bullet}(T_0, T)^{1/(T-T_0)}\boldsymbol{w}. \tag{37}$$

To motivate the choice of the former, consider the example where we have two variables and two scenarios. The first variable has over a two-year horizon the returns 1.2 and 0.8 respectively for each scenario and the second variable has 1.1 and 0.9.

Using the former, $\left(\boldsymbol{R}_{i\bullet}(T_0, T)w\right)^{1/(T-T_0)}$, we get for the two variables the annualized return

$$\left(\frac{1.2 + 0.8}{2}\right)^{0.5} = 1,$$
$$\left(\frac{1.1 + 0.9}{2}\right)^{0.5} = 1$$

while when using $\boldsymbol{R}_{i\bullet}(T_0,T)^{1/(T-T_0)}w$ we get

$$\frac{1.2^{0.5}+0.8^{0.5}}{2}=0.9949,$$
$$\frac{1.1^{0.5}+0.9^{0.5}}{2}=0.9987.$$

Now suppose that we invested 1 unit in each variable. After two years the value would still be 1 unit on average, for both variables. It is reasonable to assume that the average annualized return therefore would be equal for both variables over this horizon, which is the motivation for using the first method.

Furthermore, we can add views on the variance of the annualized return by first considering the variance of the unanualized return. If v is the view of the variance of the annualized return over the horizon (T_0,T), then $(T-T_0)v$ is the variance of the return over the horizon (T_0,T).

Let X denote a random variable of the annualized return over the period (T_0,T) and let Y be the corresponding random variable of the return. We can set the second moment of Y written in terms of $E[X]=u$ and $Var(X)=v$ as

$$\begin{aligned}E[Y^2] &= E[Y]^2+Var(Y)\\ &= \left(E[X]^{(T-T_0)}\right)^2+(T-T_0)Var(X)\\ &= u^{2(T-T_0)}+(T-T_0)v.\end{aligned}$$

We see that $E[Y^2]$ is determined by the mean and variance of X and we can use this to express the linear constraint to set the variance of X. We can write the constraints as

$$\boldsymbol{R}_{i\bullet}(T_0,T)\tilde{\boldsymbol{w}}=u^{(T-T_0)}, \tag{40a}$$
$$\left(\boldsymbol{R}_{i\bullet}(T_0,T)\circ\boldsymbol{R}_{i\bullet}(T_0,T)\right)\tilde{\boldsymbol{w}}=u^{2(T-T_0)}+(T-T_0)v. \tag{40b}$$

With similar arguments as above we can express the view on the correlation of the annualized return between two variables to a value c, by the following constraints

$$\boldsymbol{R}_{i\bullet}(T_0,T)\tilde{\boldsymbol{w}}=u_i^{(T-T_0)}, \tag{41a}$$
$$\boldsymbol{R}_{j\bullet}(T_0,T)\tilde{\boldsymbol{w}}=u_j^{(T-T_0)}, \tag{41b}$$
$$\left(\boldsymbol{R}_{i\bullet}(T_0,T)\circ\boldsymbol{R}_{i\bullet}(T_0,T)\right)\tilde{\boldsymbol{w}}=u_i^{2(T-T_0)}+(T-T_0)v_i, \tag{41c}$$
$$\left(\boldsymbol{R}_{j\bullet}(T_0,T)\circ\boldsymbol{R}_{j\bullet}(T_0,T)\right)\tilde{\boldsymbol{w}}=u_j^{2(T-T_0)}+(T-T_0)v_j, \tag{41d}$$
$$\left(\boldsymbol{R}_{i\bullet}(T_0,T)\circ\boldsymbol{R}_{j\bullet}(T_0,T)\right)\tilde{\boldsymbol{w}}=u_i^{(T-T_0)}u_j^{(T-T_0)}+c\cdot(T-T_0)\sqrt{v_i}\sqrt{v_j}. \tag{41e}$$

6.3 Views on CVaR

Views on conditional value at risk (CVaR), also known as the expected shortfall (ES), is not possible to state as a linear constraint on the posterior, however a method to impose views on CVaR, and the value at risk (VaR), is proposed by Meucci in [MAK12].

The view on the CVaR is stated as

$$\tilde{E}[X|X \leq \tilde{q}_\gamma] = \tilde{v}_\gamma, \tag{42}$$

where $\tilde{v}_\gamma$ is the $(1-\gamma)$-CVaR of the view of the posterior, where $\gamma \in (0,1)$. Furthermore, $\tilde{q}_\gamma$ denotes the γ-quantile of the posterior

$$\int_{-\infty}^{\tilde{q}_\gamma} \tilde{f}(x)dx = \gamma. \tag{43}$$

The quantile $\tilde{q}_\gamma$ is known as the value at risk. As previously stated, it is not possible to map the above constraint into a linear constraint since we do not know a-priori the value of the VaR. That is, we do not know how many scenarios which is falling below the VaR-quantile in advance. Thus, our objective is to find VaR, or the number of scenarios below the VaR-quantile, while having the minimum relative entropy between the posterior and the prior.

Assume that the scenarios for the variable of interest i, are sorted so that $S_{i1} \leq \ldots \leq S_{iN}$ and that the prior weights w_j are rearranged accordingly. Furthermore, for $s \in \{1, ..., N\}$, define the vectors of weights $\tilde{\boldsymbol{w}}^{(s)}$ by

$$\tilde{\boldsymbol{w}}^{(s)} = \underset{\boldsymbol{\omega} \in C_s}{\operatorname{argmin}}\ \varepsilon(\boldsymbol{\omega}, \boldsymbol{w}), \tag{44}$$

where C_s are the constraints given by

$$C_s = \begin{cases} S_{i1}\omega_1 + \ldots S_{is}\omega_s = \gamma \cdot \tilde{v}_\gamma \\ \omega_1 + \ldots + \omega_s = \gamma. \end{cases} \tag{45}$$

Since these constraints are linear in $\boldsymbol{\omega}$ we can now proceed as previously as described in Section 3.1. The pairs consisting of S_{ij} and $\tilde{w}_j^{(s)}$ now represent the minimum relative entropy posterior which satisfies the view of VaR equal to $\tilde{q}_\gamma$ and $CVaR$ equal to $\tilde{v}_\gamma$. The problem is now finding the ideal choice of s, which is done with regards to the relative entropy. We choose that s which among all $\tilde{\boldsymbol{w}}^{(s)}$ display the least amount of distortion overall, compared to the prior. In other words, s is chosen such that $\tilde{\boldsymbol{w}}^{(s)}$ gives the minimum relative entropy of all possible choices of s. That is, let $\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{w}}^{(\tilde{s})}$ where

$$\tilde{s} = \underset{s\in\{1,\ldots N\}}{\operatorname{argmin}} \ \varepsilon\left(\tilde{\boldsymbol{w}}^{(s)}, \boldsymbol{w}\right). \tag{46}$$

In practice, it is computationally intensive to compute all terms in (46) to find the ideal choice of s. Fortunately, the relative entropy $\varepsilon\left(\tilde{\boldsymbol{w}}^{(s)}, \boldsymbol{w}\right)$ is a concave function of s and the computation can therefore be made more efficient compared to computing the relative entropy for all choices of s. The algorithm used is a discrete counterpart to the well known Newton-Raphson method. First, define the empirical derivative of the relative entropy as

$$D\varepsilon\left(\tilde{\boldsymbol{w}}^{(s)}, \boldsymbol{w}\right) = \varepsilon\left(\tilde{\boldsymbol{w}}^{(s+1)}, \boldsymbol{w}\right) - \varepsilon\left(\tilde{\boldsymbol{w}}^{(s)}, \boldsymbol{w}\right). \tag{47}$$

Then initialize a value $\underline{\mathrm{s}} \in \{1, \ldots, N\}$ as the largest integer which satisfies

$$w_1 + \ldots w_{\underline{\mathrm{s}}} \leq \gamma, \tag{48}$$

where $w_1, \ldots, w_N$ are the prior weights.

The Newton-Raphson method is then applied by iterating the following steps until convergence is reached

$$\bar{s} = \text{RoundToClosestInteger}\left(\underline{\mathrm{s}} - \frac{D\varepsilon\left(\tilde{\boldsymbol{w}}^{(\underline{\mathrm{s}})}, \boldsymbol{w}\right)}{D^2\varepsilon\left(\tilde{\boldsymbol{w}}^{(\underline{\mathrm{s}})}, \boldsymbol{w}\right)}\right) \tag{49a}$$

$$\underline{\mathrm{s}} = \bar{s}. \tag{49b}$$

Numerical studies shows that typically only few iterations are required to reach convergence [MAK12].

We consider the same example given by Meucci to visualize the effect imposing views on the CVaR has on the posterior distribution. Let the prior be a sample from the standard normal distribution and impose the view $\tilde{E}[X|X \leq \tilde{q}_{0.2}] = -2$, which is a view on the 80%-CVaR. After finding the s which minimizes the relative entropy between the prior and posterior, we obtain the weight vector representing the posterior. The resulting histogram illustrated in Figure 11 shows the effect of the CVaR view on the posterior.

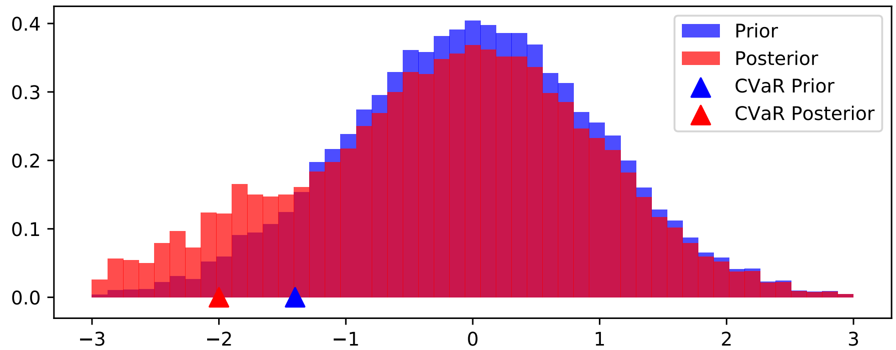


Figure 11: Histograms of the prior and posterior where the imposed view on the 80%-CVaR was set to -2.

The issues with imposing views on the VaR and the CVaR is that the views are on the tails of the distribution where the scenarios are sparse, and without enough scenarios these views may not be possible to cover. This obviously depend on the value of γ, and values of γ close to 0 or 1 can show to be problematic.

7.1 The model

Since the methods studied in this thesis are not model dependent the model used in this case study will not be specified in detail. However, the model used in this case study is a model which is used in the industry. The model originates from the Vector Autoregressive Moving Average (VARMA) model, which is a commonly used and studied model and a more extensive formulation of this model can be found in [RD13]. The focus will be on the EP approach and not the model generating the scenarios and thus we only briefly describe the VARMA model as background for this case study.

Consider modelling a multivariate time series $\{x_t\}$, where $\boldsymbol{e}_t \in \mathbb{R}^M$. A VARMA$(p, q)$ process is given by the following recursion

$$\boldsymbol{x}_t = \Phi_0 + \Phi_1 \boldsymbol{x}_{t-1} + \ldots + \Phi_p \boldsymbol{x}_{t-p} + \boldsymbol{e}_t + \Theta_1 \boldsymbol{e}_{t-1} + \ldots + \Theta_q \boldsymbol{e}_{t-q}, \tag{50}$$

where $\boldsymbol{e}_t = \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t$ denotes the forecast error, or innovation. It is assumed that the innovations are serially independent and normally distributed, $\boldsymbol{e}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t)$, where $\boldsymbol{\Sigma}_t \in \mathbb{R}^{M \times M}$. The parameters of the model $\Phi_0, \ldots, \Phi_p, \Theta_1, \ldots, \Theta_q$ are estimated through least squares and how this is done can be read about in [RD13]. When the parameters are estimated, future paths can be simulated by applying the recursion above. We will not go into detail about how p and q are chosen, as well as checking that the *stability* and *invertibility* conditions are satisfied, which can also be read about in [RD13].

7.2 Imposing views

In this case study we consider six variables, or risk factors, consisting of returns for different bonds, equities and real estate. The views to be imposed are on the same form as the ones imposed by Danish requirements [Rå20], which are views on the mean, standard deviation and correlations of annualized returns over a time horizon. That is, we impose the views described in Section 6.2 and do this on a 5-year horizon. The views are summarized in Table 4 and Table 5. It should be mentioned that while the scenarios are generated from real data and a model which is used in the industry, the numbers in the views are manually chosen and not stated by any official document. The views are chosen to be similar to the returns, standard deviations and correlations exhibited in the prior scenarios. That is, no drastic views are introduced.

		Return	Standard Deviation
1	Government and Mortgage Bonds	-0.2%	0.6%
2	Investment Grade Bonds	0.9%	0.8%
3	High Yield Bonds	1.5%	0.7%
4	Global Equities	4.9%	13.5%
5	Private Equity	8.7%	52.0%
6	Real Estate	5.1%	7.8%

Table 4: Views on the means and standard deviations of annualized returns for years 1-5.

	Correlations						
		1	2	3	4	5	6
1	Government and Mortgage Bonds		0.9	0.6	0.1	0.0	0.1
2	Investment Grade Bonds			0.7	-0.1	0.0	0.1
3	High Yield Bonds				-0.1	0.0	0.0
4	Global Equities					0.5	0.2
5	Private Equity						0.1
6	Real Estate						

Table 5: Views on the correlations of annualized returns for years 1-5.

The prior consisted of 5000 scenarios of monthly returns over 60 years. After carrying out the EP approach the resulting effective number of scenarios was 2642, which is roughly half the effective number of scenarios in the prior, which can be considered a significant decrease. Next, the weighted sampling method, discussed in Section 5, was applied to obtain a posterior consisting of 1000 scenarios. This smaller set of scenarios had an effective number of scenarios of 716, which is a higher percentage of effective number of scenarios, relative to the number of scenarios in the posterior, than simply using the EP approach. However, the posterior when using the weighted sampling method does not satisfy the views exactly, as in the case of the EP approach, but only on average. An effective number of scenarios of 716 was large enough for the posterior to converge to the values set by the views with regards to the average annualized returns and their standard deviations. However, the effective number of scenarios was not sufficiently large for the correlations to converge to the views, and thus the posterior does not satisfy those views well. The values of the average annualized returns, standard deviations and correlations for the prior, views and posterior, in this case study, can be found in Table A1 and Table A2.

7.3 Curse of clashing views

Even though the values in the views were close to the value exhibited in the prior, the resulting effective number of scenarios was low, compared to the prior. This is due to an effect which we choose to refer to as the *curse of clashing views*. This is an effect which often occurs when imposing a large number of different views. The curse of clashing views occurs when a combination of views, which individually are easily satisfied, are hard to satisfy when combined.

As an example, consider two highly positively correlated variables. We impose a view of increased mean for the first variable and a decreased mean for the second variable. When carrying out the EP approach on each view individually, the effect on the effective number of scenarios is low, however when the views are imposed simultaneously the effect on the effective number of scenarios is high. This is due to the high correlation of the two variables which causes the two views to clash with each other.

This was something that was also observed in this case study. Let $V_1, \ldots, V_6$ denote the views on the annualized returns for the six variables respectively, which was stated in Table 4. That is, V_1 denotes the view that Government and Mortgage Bonds have an annualized average return of -0.2 over a 5-year horizon.

In Figure 12 we can see the loss in effective number of scenarios when imposing the views individually and the loss when imposing the combined views. To clarify, when a view is imposed individually, we apply the EP approach with only that view. We can see that when imposing V_1, V_2 and V_3 individually, the sum of the respective loss in effective number of scenarios is around 600. However, when imposing all three views, V_1, V_2 and V_3, at the same time results in a significantly higher loss of effective number of scenarios, at around 1200. This is an effect of the curse of clashing views between V_3 and V_1 or V_2, or the combination of all three.

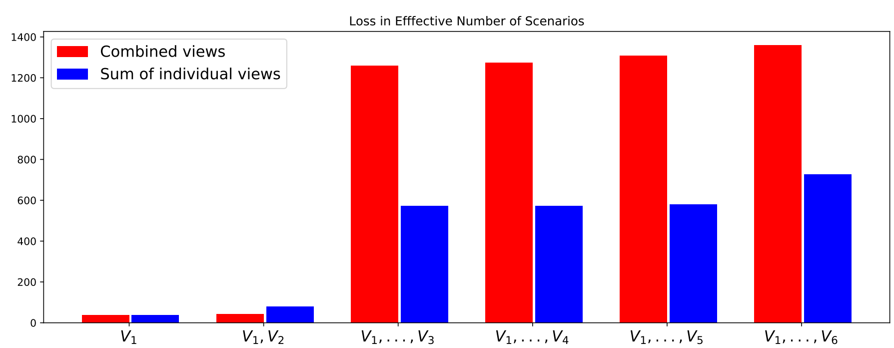


Figure 12: Loss of effective number of scenarios when the EP approach is applied separately (blue) or when all views are combined (blue).

7.4 Results

We checked above how well the views were satisfied and we also need to check how close we are to the historical data and the prior scenarios. We know that the EP approach uses the relative entropy as a measure of closeness and minimizes this between the prior and the posterior. However, we want to make sure that we keep other properties such as higher moments, e.g. skewness and kurtosis, as well. Since the effective number of scenarios is rather low for the posterior we can not obtain reliable estimates of skewness and kurtosis and we instead have to resort to "eye"-tests, that is, to investigate how properties have changed using plots.

By plotting the quantiles of the prior and posterior against each other, in a QQ-plot, we get some intuition on how similar the prior and posterior are, for some property. For example, we can see in Figure 13 the quantiles of the normalized annualized returns over a 10-year horizon, for Global Equities, which we imposed views on, but also US Equities, which we did not impose any views on. It is important to also check variables which do not have any views since these are affected too and if some important property of this variable is lost when the EP approach is applied one may need to consider adding a view which keeps that property. We observe that for both Global Equities and US Equities the QQ-plots display an approximately straight line indicating that the posterior distribution has kept the properties of the prior distribution and has not become heavier or light tailed, or more or less skewed.

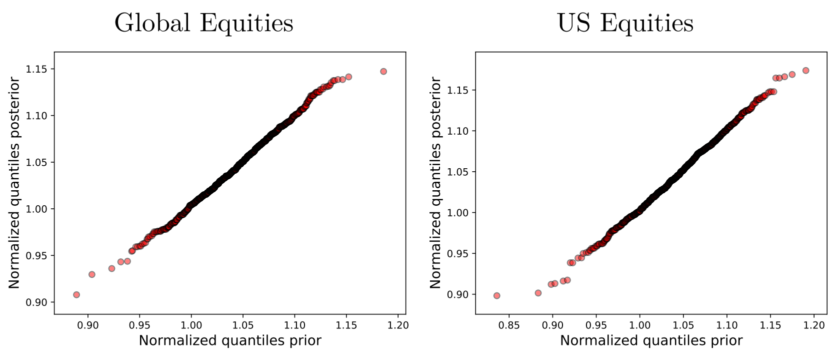


Figure 13: QQ-plots of annualized returns over a 10-year horizon for Global Equities and for US Equities.

In Figure 14 and Figure 15 see the simulated log returns in the prior and posterior distribution of Global Equities and US Equities, where the colors indicate quantiles of the scenarios over time. We can also see the mean of annualized returns over rolling windows of different horizons, for each scenario, and the respective standard deviation. That is, for each scenario we compute the annualized return over a set horizon from now. Then, the annualized return is computed again over the same horizon length, but moved one month forward.

Lastly, the mean of the rolling window returns is computed for each scenario and for different lengths of the window. This way we can see how the EP approach affects returns for different horizons. We see that the difference between the prior and the posterior is minimal in both figures.

It should be mentioned that the blue line in the bottom figures is computed from the historical data and the reason why the historical data deviates from the generated scenarios is because views already has been incorporated in the model generating the scenarios. That is, the scenarios is not generated from a pure VARMA model and views has been incorporated using other methods. The fact that the EP approach can be added on top of other methods imposing views makes it powerful. Furthermore, by combining multiple methods to add views we can reduce some of the issues with the EP approach, namely the decrease of the effective number of scenarios. For example, if views on the means can already be satisfied in the generation of the scenarios, the loss in effective number of scenarios will be lower compared to incorporating the views of the means when applying the EP approach.

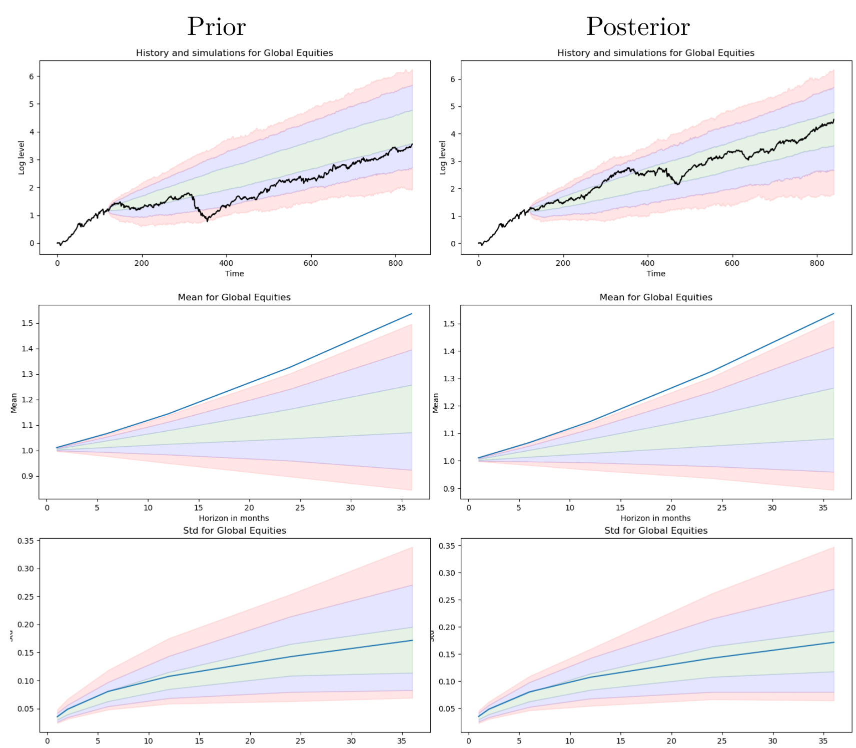


Figure 14: Simulated log-returns of Global Equities, mean of rolling windows of annualized returns for each scenario and the respective standard deviation. Black line in top figures represent one scenario and blue line in in bottom figures represents historical data.

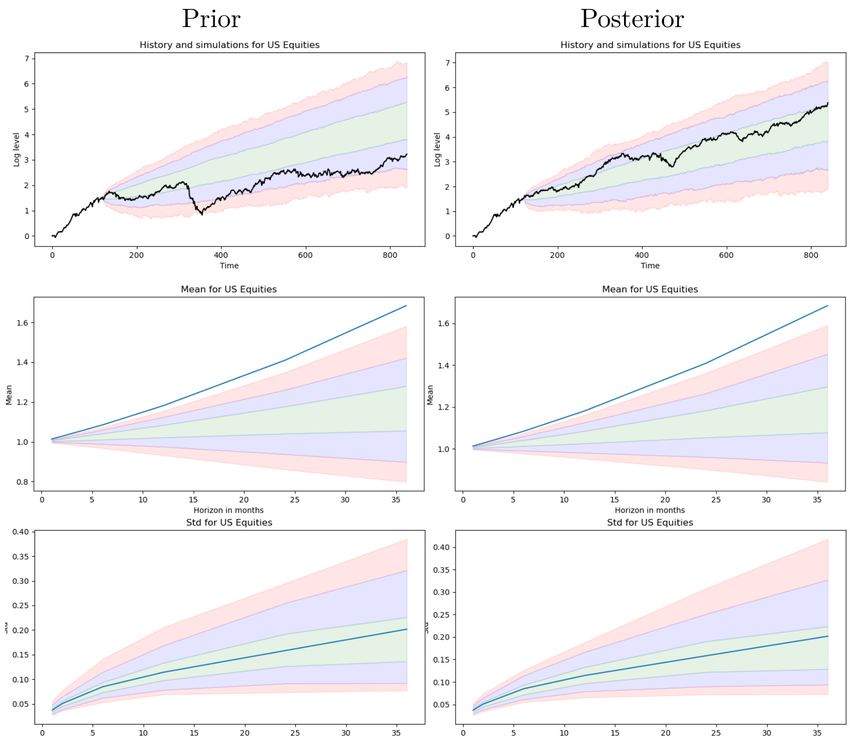


Figure 15: Simulated log-returns of US Equities, mean of rolling windows of annualized returns for each scenario and the respective standard deviation. Black line in top figures represent one scenario and blue line in in bottom figures represents historical data.

8.1 Issues and advantages of the entropy pooling approach

There are four main issues with the EP approach which has been discussed. Firstly, the posterior is restricted to being expressed by the prior scenarios which makes the method not suitable for imposing views where the concentration of scenarios in the prior is low, and it is not even possible to impose views outside the range of the prior views. This is not a problem considering the analytical solution discussed in theoretical solution discussed in Section 2, however, since one most often needs to resort to the computational approach in practice, this becomes an issue. The solution to most of the issues with this method is to generate a large number of scenarios as the prior, which would also solve this issue. With a large number of prior scenarios, the range of the scenarios is wider, which makes the posterior more flexible.

Secondly, the EP approach can significantly decrease the effective number of scenarios, which is the biggest issue with this method. The effective number of scenarios is a crucial number which needs to be considered when applying this method. A low effective number of scenarios can be detrimental and may result in a useless posterior. With a limited number of scenarios or limitations in computational time the effective number of scenarios may simply be too low for the desired application.

Thirdly, when imposing many views, the possibility of a large effect of the curse of clashing views occurring increases. This can have a very large impact on the effective number of scenarios or the views may not even be possible to satisfy with the available scenarios. This effect can be hard to predict in practice and if this effect occurs depend both on the available scenarios and the views.

Lastly, the computational approach is restricted to linear constraints to guarantee computational feasibility and while, as has been shown, many views which are natural to impose can be written as linear constraints, we are still limited by this restriction.

The EP approach can also be considered somewhat of as a "black-box" since we lose the interpretability of the parameters in the original model after the EP approach is applied. Properities and assumptions made in the original model may also be lost when the EP approach is applied. Therefore, a method which instead adjusts the parameters in the model would keep the interpretability and the motivation for why that model was chosen still holds, compared to the EP approach. However, depending on the flexibility of the model and the views it may not even be possible to find a set of values for the parameters which makes the model satisfy all views, and some tolerance for how well the views are satisfied may need to be introduced. There may also be a problem with a decreasing number of free parameters in the model when imposing further views. This method might be feasible for a low number of simple views but becomes exceedingly more complicated when imposing more views and more complex views, e.g. correlations. This is the major advantage of the EP approach, more complex views such as correlations can be incorporated with ease and the optimization is performed over only a limited number of parameters, equal to the number of views.

To summarise, the EP approach should be used carefully and should mainly be used for fine tuning and not for drastic views, if a very large number of prior scenarios is not available. The prior scenarios are generally not flexible enough to satisfy more extreme views, with a realistic number of scenarios. However, since the EP approach is applied on the scenarios, we can combine it with other methods which incorporates views in the model, prior to applying the EP approach, to overcome some of the issues with the EP approach.

Meucci [Meu10] presents a table to showcase the capabilities of the EP approach compared to other methods in Black and Litterman [BL90], Almgren and Chriss [AC06], Qian and Gorman [QG01], Pezier [Pez07], Meucci [Meu09], and the COP approach in [Meu06a]. This table is presented in Table 6.

	BL	AC	QG	P	M	COP	EP
Normal market and linear views	✓	·	✓	✓	✓	✓	✓
Scenario analysis	·	·	✓	✓	✓	✓	✓
Correlation stress-test	·	·	✓	✓	✓	·	✓
Trading desk: non-linear pricing	·	·	·	·	✓	✓	✓
External factors: macro, etc.	·	·	·	·	✓	✓	✓
Partial specifications	·	·	·	✓	·	·	✓
Non-normal market	·	·	·	·	·	✓	✓
Multiple users	·	·	·	·	·	✓	✓
Non-linear views	·	·	·	·	·	·	✓
Trading desk: costly pricing	·	·	·	·	·	·	✓
Lax constraints: ranking	·	✓	·	·	·	·	✓

Table 6: Comparison of the capabilities of different methods incorporating views, as shown in [Meu10].

8.2 Transforming views

In most situations, it is desirable or necessary to transform the time series data to remove trends and to satisfy assumptions of the model. Vast majority of models assume that the amount of variability is constant in time. This causes a problem if the views are stated in the pre-transformed context since the views therefore also needs to be transformed to be used in the model. However, this is not a problem in the EP approach.

Two commonly used transformations are differencing and the log-transformation. The dth differencing operator applied to a time series x_t creates a new time series $y_t = x_t - x_{t-d}$. This method is used to remove the dependence on time, like trends and seasonality. As an example, if the practitioner has views V_t of the expected value of variables before the difference transform is applied, then the transformed views, denoted V_t^*, can be written $V_t^* = V_t - V_{t-d}$.

This introduces the problem of incorporating views into the model since the views also needs to be transformed, which can be challenging. Consider the following example, suppose than there are views imposed in the 5-step ahead forecast, it is not obvious how the views also could be transformed if views for the 1- to 4-step ahead forecast is not available.

One of the advantages of the EP approach is that transformations of views are not required. Assume that a transformation has been applied to some input variable in the model, then the scenarios generated by the ESG will also be in the transformed context. Since the EP approach is only applied on the scenarios, we can first apply the inverse transform on the scenarios before we carry out the EP approach with the imposed views.

8.3 Conclusion

We described the theoretical foundation as well as the computational approach to the EP approach, which is a framework to incorporate views on the forecasted expected value, variance and correlations, as well as other more complicated views. It was showcased for both the one-period setting as well as the time-dependent setting, where scenarios were introduced. It was also shown that if the views can be written as linear constraints the computation of the weights is extremely efficient. We also contributed to the current literature with a method to obtain a posterior represented by a small set of scenarios for situations where computational limitations are present, while maintaining a high effective number of scenarios. The EP approach has some clear advantages compared to other methods which also incorporates views on the market with the main advantage being that the method can be applied without assuming normal markets.

The EP approach was first introduced by Meucci in 2010 and is thus a relatively new method, with areas for future research. The list of views which can be written as linear constraints can be further extended as well as finding efficient methods for views which can not be written as linear constraints. The area which needs most attention is to find ways to deal with the issues regarding the loss of effective number of scenarios, where an idea would be to sacrifice some accuracy, with regards to satisfying the views, to gain effective number of scenarios.

While it is desirable that the posterior only contains scenarios which are present in the prior, and thus only contains scenarios generated by the original model in the ESG, this is the reason for the loss in effective number of scenarios. It would therefore be desirable to obtain a posterior which we can sample from directly. However, the posterior can not be expressed analytically, but it may be possible to approximate the posterior.

References

[AC06] R. Almgren and N. Chriss, *Optimal portfolios from ordering information*, The Journal of Risk **9** (2006), no. 1, 1–47.

[BL90] F. Black and R. Litterman, *Asset allocation: Combining investor views with market equilibrium*, Goldman Sachs Fixed Income Research Note (1990).

[BV04] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, USA, 2004.

[KM53] H. Kahn and A. W. Marshall, *Methods of Reducing Sample Size in Monte Carlo Computations*, Operations Research **1** (1953), no. 5, 263–278.

[MAK12] A. Meucci, D. Ardia, and S. Keel, *Fully flexible extreme views*, 2012, pp. 39–49.

[MAK13] ———, *Fully Flexible Views in Multivariate Normal Markets*, Cahiers de recherche 1311, CIRPEE, 2013.

[Meu06a] A. Meucci, *Beyond black-litterman in practice: A five-step recipe to input views on non-normal markets*, Risk **19** (2006), 114–119.

[Meu06b] ———, *Beyond black-litterman: Views on non-normal markets*, Risk **19** (2006), 87—92.

[Meu09] ———, *Enhancing the black-litterman and related approaches: Views and stress-test on risk factors*, Journal of Asset Management **10** (2009), 89–96.

[Meu10] ———, *Fully flexible views: Theory and practice*, 2010.

[Meu12] ———, *Effective number of scenarios in fully flexible probabilities*, 2012, pp. 32–35.

[PCC+16] H. Pedersen, M. P. Campbell, S. Christiansen, S. Cox, D. Finn, K. Griffin, N. Hooker, M. Lightwood, S. Sonlin, and C. Suchar, *Economic scenario generators: A practical guide*, 2016.

[Pez07] J. Pezier, *Global portfolio optimization revisited : A least discrimination alternative to black-litterman*, ICMA Centre Discussion Papers in Finance (2007).

[PS13] C. Papadimitriou and K. Steiglitz, *Combinatorial optimization: Algorithms and complexity*, Dover Books on Computer Science. Dover Publications, 2013, (Cit. on p. 6).

[QG01] E. Qian and S. Gorman, *Conditional distribution in portfolio theory*, Financial Analysts Journal **57** (2001), no. 2, 44–51.

[RD13] R. Rebonato and A. Denev, *Portfolio management under stress : a bayesian-net approach to coherent asset allocation*, Cambridge University Press, Cambridge, 2013 (eng).

[Rå20] Rådet for Afkastforventninger, *Samfundsforudsætninger for 1. halvår 2021*, https://www.afkastforventninger.dk/media/1453/samfundsforudsaetninger-1-halvaar-2021-1.pdf.

[Sme16] M. Smerlak, *Minimum relative entropy distributions with a large mean are gaussian*, Physical Review E **94** (2016), no. 6.

[vdS19] M. van der Schans, *Entropy pooling with discrete weights in a time-dependent setting*, Computational Economics **53** (2019).

Appendices

Algorithm A1: Heuristic from [vdS19] to convert $\tilde{w}$ to discrete weights.

input : Continuous weights $\tilde{w}$ obtained after solving the optimization problem.
The required sum of the discrete weights $\hat{N}$.

output : Discrete weights $\hat{w}$.

Round $\tilde{w}$ to discrete weights: $\hat{w} \leftarrow \text{RoundToClosestInteger}(\hat{N}\tilde{w})$.

while *weights* $\hat{w}_j$ *do not sum to* $\hat{N}$ **do**

  Increase/decrease weight $\hat{w}_j$ with largest negative/positive rounding error.

for $number\_iterations <= max\_iter$ **do**

  if $||A\hat{w} - a|| < max\_tol$ **then**

    return $\hat{w}$

  Select an equality constraint k_s where constraint k is selected with probability

$$\frac{|(A\hat{w} - a)_k)|}{\sum |(A\hat{w} - a)_k)|}.$$

  Randomly select the number of weights to decrease/increase $n\_changes$ from $[1, ..., max\_changes]$.

  for $number\_tries <= max\_tries$ **do**

    Use weighted random sampling without replacement with weights

$$\exp(A_{k_s \bullet})$$

    to select $n\_changes$ weights $\hat{w}_j$ that have not been increased already by an amount max_diff. Use weighted random sampling without replacement with weights

$$\exp(-A_{k_s \bullet})$$

    to select $n\_changes$ weights $\hat{w}_j$ that have not been decreased already by an amount max_diff.

    if *increase/decrease of the selected weights by* 1 *decreases error* $||A\hat{w} - a||$ **then**

      Increase/decrease the selected weights with 1.

return $\hat{w}$

	Prior		View		Posterior	
	Return	S.d.	Return	S.d.	Return	S.d.
1	−0.1662%	0.6094%	−0.2%	0.6%	−0.1946%	0.5929%
2	0.9371%	0.6732%	0.9%	0.8%	0.9041%	0.7655%
3	1.3303%	0.8416%	1.5%	0.7%	1.5013%	0.7170%
4	4.7789%	14.1169%	4.9%	13.5%	4.7361%	13.2114%
5	9.3845%	46.7171%	8.7%	52.0%	8.2847%	53.3697%
6	5.9072%	9.2510%	5.1%	7.8%	5.0877%	7.7389%

Table A1: Prior, views and posterior values of the means and standard deviations of annualized returns for years 1-5.

Correlations

Prior

	1	2	3	4	5	6
1		0.8907	0.6256	0.0538	-0.0447	0.1011
2			0.7443	-0.0908	-0.052	0.0457
3				-0.1327	-0.0480	-0.0301
4					0.4497	0.1408
5						0.0958
6						

Views

	1	2	3	4	5	6
1		0.9	0.6	0.1	0.0	0.1
2			0.7	-0.1	0.0	0.1
3				-0.1	0.0	0.0
4					0.5	0.2
5						0.1
6						

Posterior

	1	2	3	4	5	6
1		0.8236	0.5563	0.0876	0.0014	0.1358
2			0.6657	-0.1263	-0.0111	0.1541
3				-0.1236	-0.0275	0.0259
4					0.4639	0.1851
5						0.0928
6						

Table A2: Prior, views and posterior values of the correlation between annualized returns for years 1-5.

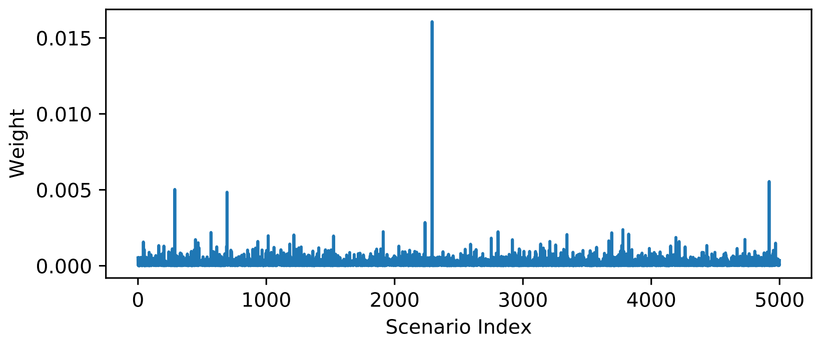


Figure A1: Weight vector $\tilde{\boldsymbol{w}}$ obtained in the case study.

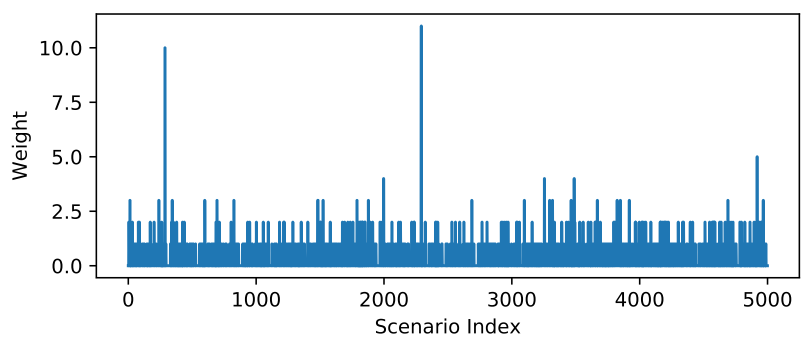


Figure A2: Discrete weights obtained from the weighted sampling method in the case study.