

# Optimal portfolio allocation with uncertain covariance matrix

Maxime Markov\* and Vladimir Markov

November 14, 2023

## Abstract

In this paper, we explore the portfolio allocation problem involving an uncertain covariance matrix. We calculate the expected value of the Constant Absolute Risk Aversion (CARA) utility function, marginalized over a distribution of covariance matrices. We show that marginalization introduces a logarithmic dependence on risk, as opposed to the linear dependence assumed in the mean-variance approach. Additionally, it leads to a decrease in the allocation level for higher uncertainties. Our proposed method extends the mean-variance approach by considering the uncertainty associated with future covariance matrices and expected returns, which is important for practical applications.

## 1 Introduction

Portfolio allocation is a classical problem in finance. The two dominant methodologies for deriving portfolio weights are either by maximizing the expected utility (EU) [1] or by focusing on the mean-variance (MV) of the portfolio in relation to those weights [2]. While the MV approach is favored for its conceptual simplicity and analytical tractability, the utility function method offers a more comprehensive perspective. When using a Constant Absolute Risk Aversion (CARA) utility function with Gaussian returns, both methods yield identical results. However, because real-world returns are not strictly Gaussian, the utility function method, which can accommodate arbitrary return distributions, is often preferred.

All modern portfolio optimization models use expectations of future returns and variances estimated from the past data. These estimates are imperfect and can be very far from reality during market regime changes. Comprehensive modeling and optimization should explicitly incorporate the fact that future returns and variances are unknown, utilize predictive models of future returns and variances, and study the implications of these uncertainties. Therefore, investors would be better off not focusing on detailed

---

\*Corresponding Author: markov@theory.polytechnique.fr

estimation and forecasting of parameters but instead on averaging over all possible scenarios or estimating the worst-case scenario. This can be achieved in the EU approach to portfolio allocation.

In our prior work [3], we examined the portfolio allocation problem with uncertain expected returns for both Gaussian and Asymmetric Laplace-distributed (ALD) returns. Our findings illustrate that uncertainty in expected returns leads to shrinkage in optimal portfolio weights, while skewness and fat tails alter the risk-term dependence. This paper focuses on the portfolio allocation problem involving an uncertain covariance matrix. Our methodology is loosely inspired by the construction of Bayesian models. In this context, the model parameters — expected returns and covariance matrix — are treated as random variables with a specified distribution and corresponding hyperparameters. To obtain observables, we marginalize (integrate) over these random parameters.

The distribution of parameters, which is external to the model, reflects the modeler’s perspective on future parameter values and their associated uncertainty. This uncertainty primarily arises from two factors: statistical error due to finite sample estimation and, more significantly, prediction error in a non-stationary environment. Such non-stationarity makes it impossible to precisely estimate future parameter values, even with the most advanced forecasting algorithms. In this paper, we propose three covariance matrix noise models that capture both types of errors in different scenarios.

This paper is structured as follows: First, we will briefly compare the expected utility and mean-variance approaches to portfolio allocation. Second, we will discuss the marginalization over the variance distribution in the univariate case. Third, we compute the expected value of the utility function and corresponding allocation weights by marginalizing the covariance matrix using three analytically solvable noise models: the Wishart distribution, an equivariance model with a block structure, and a two-state model that considers the non-zero probability of a market crash. Finally, we will discuss the practical application of the results.

## 2 Mean-variance vs. Expected utility approaches

In the standard economic approach, asset allocations are derived by maximizing the expected value of the investor’s utility function. The utility non-linearly transforms the investor’s wealth and encodes his aversion to risk. The most commonly used utility model in finance is the CARA [4, 5]. The utility functions  $U_a(x, a)$  for CARA with respect to the investment outcome (return)  $x$  and the risk aversion parameter  $a$  can be expressed as follows:

$$U_a(x, a) = \begin{cases} \frac{1 - \exp(-ax)}{a}, & \text{if } a \neq 0 \\ x, & \text{if } a = 0 \end{cases}, \quad (1)$$

To determine the optimal portfolio weights, we maximize the expected value of the utility function  $U(x, a)$  with respect to  $w$ :

$$w = \arg \max_w E_P[U_a(x, a)] = \arg \max_w \int dx U_a(x, a)P(x) \quad (2)$$

where  $P(x)$  is the distribution of outcome  $x$ .

An investor using the MV approach optimizes the following function:

$$w^* = \arg \max_w [E[x] + \lambda \text{Var}(x)] \quad (3)$$

While the two approaches yield identical results in the case of a normally distributed outcome, they diverge when the outcome distribution cannot be fully characterized by only its first two moments. Obviously, the expected utility approach makes it straightforward to incorporate distributions with skewness and excess kurtosis. Some distributions, like the Asymmetric Laplace Distribution (ALD), offer complete analytical tractability, which is no more complex than that of the Gaussian distribution [3].

If the future covariance matrix  $\Sigma$  and the expected returns  $\mu$  are random variables, the optimal allocation weights are determined by integrating over all possible realizations of the covariance matrix  $\Sigma$  with probability distribution  $P(\Sigma)$  and the expected returns  $\mu$  with probability distribution  $P(\mu)$ . The optimal weights  $w^*$  are given by:

$$w^* = \arg \max_w \int d\Sigma \int d\mu \int dx U_a(x, a) P(x; \mu, \Sigma) P(\mu) P(\Sigma) \quad (4)$$

where  $P(\mathbf{x})$  is the distribution of portfolio returns  $\mathbf{x}$ .  $P(\mathbf{x})$  can be assumed to follow a Gaussian distribution, as the fat tails of the actual return distribution can be attributed to the integration over variance. The integration over  $\Sigma$  is a basic operation in multivariate statistics and involves integrating over all  $\frac{N(N+1)}{2}$  independent components of  $\Sigma$ . If  $P(\Sigma)$  has a discrete distribution, the integration is substituted by a summation over the corresponding probabilities.

For multivariate Gaussian returns  $r_t \sim N(\mu, \Sigma)$ , the expected utility is given by:

$$E_N[U_a(\mu, \Sigma)] = \int dx U_a(x, a) P(x; \mu, \Sigma) \sim (-1) e^{\frac{1}{2} a^2 \mathbf{w}^T \Sigma \mathbf{w} - a \mu^T \mathbf{w}} \quad (5)$$

As one can see, the terms related to  $\Sigma$  and  $\mu$  are separated and can be integrated independently. Correspondingly, we are interested in:

$$\int d\Sigma e^{\frac{1}{2} a^2 \mathbf{w}^T \Sigma \mathbf{w}} P(\Sigma) = E_{P(\Sigma)} [e^{\text{Tr}(\mathbf{W} \Sigma)}] = M(\mathbf{W}), \quad (6)$$

where  $\mathbf{W} = \frac{a^2}{2} \mathbf{w}^T \mathbf{w}$  is the Hadamard product of weights  $\mathbf{w}$  and  $M(\mathbf{W})$  is the momentum generating function (mgf) for the distribution  $P(\Sigma)$ . Here, we used the identity  $\mathbf{w}^T \Sigma \mathbf{w} = \text{Tr}(\mathbf{W} \Sigma)$ .

The integration over the expected returns  $\mu$  can be done using the following formal analogy:

$$\int d\mu e^{-a \mu^T \mathbf{w}} P(\mu) = E_P [e^{itx}]|_{t=ia\mathbf{w}} \quad (7)$$

The last term is a characteristic function (cf)  $E_P [e^{itx}]$  of the probability distribution  $P(\mu)$ , well known for many statistical distributions.

The goal of this paper is to examine the impact of uncertainties in future variances and correlation coefficients on optimal portfolio allocation. For a general distribution  $P(\Sigma)$ , the marginalization over  $\Sigma$  cannot be done analytically. Therefore, we investigate three analytically tractable models of covariance matrix distribution that may be relevant for practical applications. First, we model  $\Sigma$  as being distributed according to the Wishart distribution. Second, we study equivariance block diagonal covariance matrices where the variance of each block is distributed based on the shifted gamma distribution, and the correlation matrix remains fixed. The shift accounts for minimal future volatility, which is an important feature of the stock market. In the third model, we calculate the expected value of the utility function in two scenarios: with probability  $p$ , we assume the future market to be in a normal regime with parameters  $(\mu_n, \Sigma_n)$ , and with probability  $(1 - p)$ , the market is under stress with parameters  $(\mu_s, \Sigma_s)$ . This approach allows for the construction of a robust portfolio if a modeler anticipates a non-zero probability of a market crash.

It is important to note that marginalization over covariance matrix  $\Sigma$  produces a trivial result in the MV approach, which is linear in the risk term  $R \sim \Sigma$ . In this scenario, integration over all possible realizations of  $\Sigma$  yields the expected value of  $E[\Sigma]$ , and the parameter controlling the uncertainty of  $\Sigma$  is absent in the final result. Regarding the uncertainty of expected returns, the Black-Litterman model serves as the primary means to account for it in the MV approach [6]. Despite 30 years of development, the practical use of the model remains limited due to a slew of non-observable and challenging-to-guess parameters. We contend that it is more natural and computationally efficient to directly incorporate future market perspectives by treating both  $\mu$  and  $\Sigma$  as distributions of random variables and marginalizing (averaging) over their realizations in the EU approach.

### 3 Univariate case: Marginalizing over variance and expected return

In this section, we derive the optimal allocation weight after marginalizing over a one-dimensional covariance matrix, which is the variance. Although this is a simplified case, it offers full analytical tractability of the problem. Many features of this solution can also be extended to the multidimensional case.

#### 3.1 Shifted gamma distribution as a noise model for variance

Assume that returns are distributed according to the normal distribution  $r_t \sim N(0, \sigma^2)$ . Then, the sample variance  $s^2$  of  $n$  observations follows the chi-squared distribution  $\chi_n^2(\sigma^2)$ :

$$s^2 \sim \frac{1}{n} \sum_{i=1}^n r_i^2 = \frac{1}{n} \chi_n^2(\sigma^2) \quad (8)$$

The chi-squared distribution  $\chi_n^2(\sigma^2)$  is a one-dimensional version of the Wishart distribution, which will be discussed in the subsequent section. Additionally, it has a

relationship with the gamma distribution, given by the identity  $\chi_n^2(\sigma^2) = \Gamma(\frac{n}{2}, 2\sigma^2)$ . An important feature of the equity market is the presence of minimal variance that is not accounted for by the  $\chi_n^2(\sigma^2)$  distribution. In other words, variance can be divided into a minimal variance deterministic part and a stochastic part. As a result, we employ a gamma distribution shifted by minimal variance as the model for the future variance distribution. The shifted gamma distribution is flexible enough to capture both finite sample uncertainty and forecasting power limitations by treating  $n$  as a model parameter.

The probability density function of the three-parameter gamma distribution is given by:

$$\Gamma(x, \tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}) = \frac{(x - \tilde{\gamma})^{\tilde{\alpha}-1} e^{-\frac{(x-\tilde{\gamma})}{\tilde{\beta}}}}{\tilde{\beta}^{\tilde{\alpha}} \Gamma(\tilde{\alpha})}, \quad x > \tilde{\gamma}, \tilde{\alpha} > 0, \tilde{\beta} > 0 \quad (9)$$

The standard two-parameter (shape-scale) parametrization corresponds to  $\tilde{\gamma} = 0$ , and the exponential distribution corresponds to parameters  $\tilde{\alpha} = 1, \tilde{\gamma} = 0$ . The momentum-generating function is given by:

$$M_x(t) = E[e^{tx}] = \frac{e^{\tilde{\gamma}t}}{(1 - \tilde{\beta}t)^{\tilde{\alpha}}} \quad (10)$$

We model the future variance using the following distribution:

$$s^2 \sim \Gamma\left(\frac{\alpha}{2}, \frac{2\sigma^2}{\alpha}, \sigma_{min}^2\right), \quad (11)$$

with the shape parameter  $\frac{\alpha}{2}$ , the scale parameter  $\frac{2\sigma^2}{\alpha}$ , and  $\sigma_{min}^2$  shifting the distribution by a minimum variance value  $\sigma_{min}^2$ . The mean is given by  $E[s^2] = \sigma_{min}^2 + \sigma^2$ . We use this equation to estimate  $\sigma$ . The uncertainty parameter  $\alpha$  can be estimated from  $Var[s^2] = \frac{2\sigma^4}{\alpha}$ . The larger values of  $\alpha$  correspond to a lower variance around the target value  $E[s^2]$ .

Approximate values of the parameters  $\sigma_{min}$  and  $\alpha$  can be estimated from historical market data. The forward-looking 30-day volatility, as measured by the VIX index, reached a minimum value of 9.5 between the years 2006 and 2023. This value can be used to determine the approximate value  $\sigma_{min} \approx 10\%$ . The volatility of volatility is measured by the VVIX Index, which represents the expected volatility (standard deviation) of the 30-day forward price of the VIX Index. The typical value of the VVIX index is in the 80 percent range. This corresponds to a very high level of uncertainty in volatility, with  $\alpha < 1$ , which makes the mean value of volatility almost useless when applying the two-sigma rule.

In Figure 1, we show the distribution of annualized volatility in the model for high, medium, and low uncertainties of  $\sigma^2$  for  $\alpha = 10$ ,  $\alpha = 100$ , and  $\alpha = 1000$ , respectively. For this simulation, we set the parameters  $\sigma_{min} = 0.1$  (10% annualized) and  $\sigma = 0.15$  (15% annualized). Subsequently,  $N = 10^5$  random variances were generated according to Eq. 11.

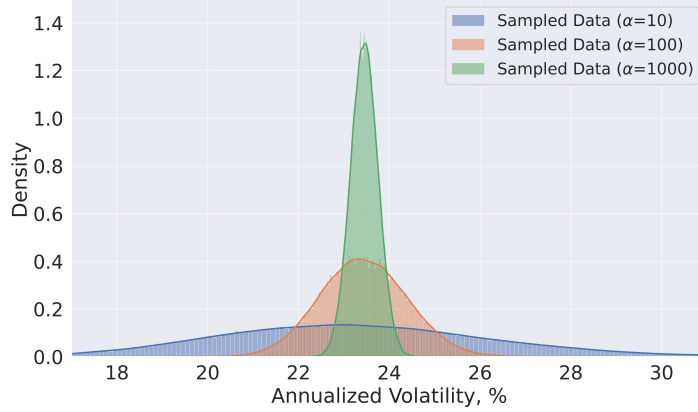


Figure 1: Noise model for volatility distribution given by Eq. 11 for different values of  $\alpha$

### 3.2 Marginalization over variance and expected return

To develop intuition, we first discuss the MV framework, which assumes Gaussian returns. Given that the portfolio return  $x$  follows a normal distribution  $x \sim N(\mu, \sigma)$ , the expected utility function  $E_N[U_a(x, a)]$  can be expressed as:

$$E_N[U_a(x, a)] = \int dx U_a(x, a) N(\mu, \sigma) = \frac{1 - e^{\frac{1}{2}a^2\sigma^2 - a\mu}}{a} \quad (12)$$

where we examine the CARA utility function  $U_a(x, a)$  from Eq. 1. According to Eq. 2, the optimal portfolio maximizes the expected utility function in Eq. 12. This is equivalent to maximizing the expression  $\mu - \frac{a}{2}\sigma^2$  or MV in Eq. 3.

The location parameter  $\mu$  is often unpredictable. To model it, we assume that  $\mu$  follows a normal distribution  $\mu \sim N(\mu_0, \sigma_0)$  with parameters  $\mu_0$  and  $\sigma_0$ . We then marginalize (integrate) Eq. 2 over the location parameter  $\mu$ . The optimal weight  $w^*$  is given by:

$$w^* = \arg \max_w \int d\mu E_N[U_a(x, a)] N(\mu; \mu_0, \sigma_0) = \arg \max_w E_\mu[U_a(x, a)] \quad (13)$$

To derive the optimal weight after marginalization over  $\mu$ , we rely on the following identity:

$$\int_{-\infty}^{\infty} d\mu e^{-aw\mu} N(\mu; \mu_0, \sigma_0) = e^{\frac{1}{2}a^2\sigma_0^2w^2 - a\mu_0w} \quad (14)$$

The optimal weight  $w_N^*$  after marginalization over  $\mu$  is given by [3]:

$$w_N^* = \arg \max_w \left[ (-1)e^{-a(\mu_0 - r_0)w + \frac{a^2}{2}w^2(\sigma^2 + \sigma_0^2)} \right] \quad (15)$$

Taking the logarithm of Eq. 15, we arrive at:

$$w_N^* = \arg \max_w \left[ w(\mu_0 - r_0) - \frac{a}{2}w^2(\sigma^2 + \sigma_0^2) \right] = \frac{\mu_0 - r_0}{a(\sigma^2 + \sigma_0^2)} \quad (16)$$

here,  $r_0$  is a risk-free return. In order to avoid trivial cluttering, we assume later in the text that an investor chooses between stock and cash ( $r_0=0$ ).

We model variance with the model in Eq. 11:  $s^2 \sim \Gamma(\frac{\alpha}{2}, \frac{2\sigma^2}{\alpha}, \sigma_{min}^2)$ . The optimal weight  $w^*$  is given by:

$$w^* = \arg \max_w \int ds^2 E_N[U_a(x, a)] \Gamma(s^2; \frac{\alpha}{2}, \frac{2\sigma^2}{\alpha}, \sigma_{min}^2) = \arg \max_w E_{\sigma^2}[U_a(x, a)] \quad (17)$$

Using the mgf of the gamma distribution in Eq. 10 and the identity in Eq. 14, the expected utility function  $E_{\mu, \sigma^2}[U_a(x)]$  after marginalization over variance  $s^2$  and expected return  $\mu$  is given by:

$$E_{\mu, \sigma^2}[U_a(x, a)] \sim (-1)e^{\frac{a^2(\sigma_{min}^2 + \sigma_0)w^2 - a\mu_0}{2}} \left(1 - \frac{a^2}{\alpha}w^2\sigma^2\right)^{-\frac{\alpha}{2}} \quad (18)$$

Maximization leads to a cubic equation:

$$a^3(\sigma_{min}^2 + \sigma_0^2)\sigma^2w^3 - a^2\mu_0\sigma^2w^2 - a\alpha(\sigma^2 + \sigma_{min}^2 + \sigma_0^2)w + \mu_0\alpha = 0 \quad (19)$$

The solution is a cumbersome expression given by the Cardano formula. The asymptotics for  $\mu_0 \rightarrow 0$  ( $\mu_0 \ll \sigma_{min}^2 + \sigma_0^2 + \sigma^2$ ) is given by:

$$w^* = \frac{\mu_0}{a(\sigma_{min}^2 + \sigma_0^2 + \sigma^2)} - \frac{\sigma^4\mu_0^3}{a\alpha(\sigma_{min}^2 + \sigma_0^2 + \sigma^2)^4} + O(\mu_0^4) \quad (20)$$

The asymptotics for  $\mu_0 \rightarrow \infty$  ( $\mu_0 \gg \sigma_{min}^2 + \sigma_0^2 + \sigma^2$ ) is given by:

$$w^* = \frac{\sqrt{\alpha}}{a\sigma} - \frac{\alpha}{2a\mu_0} + O(\mu_0^{-2}) \quad (21)$$

The asymptotics for  $\alpha \rightarrow \infty$  (small variance uncertainty) is given by:

$$w^* = \frac{\mu_0}{a(\sigma_{min}^2 + \sigma_0^2 + \sigma^2)} - \frac{\sigma^4\mu_0^3}{a\alpha(\sigma_{min}^2 + \sigma_0^2 + \sigma^2)^4} + O(\alpha^{-\frac{3}{2}}) \quad (22)$$

The leading term corresponds to the MV solution but with additional regularization (shrinkage) due to the expected return uncertainty  $\sigma_0^2$ . We note that an increase in  $\alpha$  (a decrease in variance uncertainty) leads to an increase in the allocation to the risky asset  $w^*$ .

The asymptotics for  $\alpha \rightarrow 0$  (large variance uncertainty) is given by:

$$w^* = \frac{\sqrt{\alpha}}{a\sigma} - \frac{\alpha}{2a\mu_0} + O(\alpha^{\frac{3}{2}}) \quad (23)$$

The leading term corresponds to the inverse volatility allocation.

The asymptotics for  $\sigma_0^2 \rightarrow \infty$  (large expected return uncertainty  $\mu_0 \ll \sigma_0^2$ ) is given by:

$$w^* = \frac{\mu_0}{a\sigma_0^2} - \frac{\mu_0(\sigma_{min}^2 + \sigma^2)}{a\sigma_0^4} + O(\sigma_0^{-6}) \quad (24)$$

The asymptotics for  $\sigma_0^2 \rightarrow 0$  and  $\sigma_{min} = 0$  (small expected return uncertainty  $\mu_0 \gg \sigma_0^2$ ) is given by:

$$w^* = \frac{-\sigma\alpha + \sqrt{\alpha(4\mu_0^2 + \sigma^2\alpha)}}{2a\mu_0\sigma} + O(\sigma_0^2) \quad (25)$$

We discuss the behavior of this case as a function of  $\alpha$  and the square of the Sharpe ratio  $\frac{\mu_0^2}{\sigma^2}$  in the next section.

## 4 Multivariate case: Marginalization over covariance matrices

The integration over the covariance matrix distribution  $P(\Sigma)$  in Eq. 4 cannot be performed for an arbitrary distribution. Therefore, in this section, we study three models of covariance matrix distributions that allow analytical tractability of the problem.

### 4.1 Model 1: Marginalization over Wishart distribution of covariance matrices

The Wishart distribution  $\mathbf{S} \sim W_N(n, \Sigma)$  is commonly used to model sample covariance matrices of multivariate Gaussian data  $\mathbf{X} \sim N(0, \Sigma)$ , where  $\Sigma \in R^{N \times N}$  [7]. The multivariate distribution can be seen as an extension of the chi-squared distribution, and it naturally emerges when analyzing the sample covariance matrix of multivariate normal data. When we sample  $n$  observations from the distribution, the sample covariance matrix  $\mathbf{X}$  follows a Wishart distribution with a probability density function given by:

$$f(\mathbf{X}|\Sigma, n) = \frac{|\mathbf{X}|^{\frac{n}{2} - \frac{N+1}{2}} e^{-\frac{1}{2}\text{tr}(\Sigma^{-1}\mathbf{X})}}{2^{\frac{n}{2}N} |\Sigma|^{\frac{n}{2}} \Gamma_N\left(\frac{n}{2}\right)} \quad (26)$$

where  $|\mathbf{X}|$  is the determinant of matrix  $\mathbf{X}$  and  $\Gamma_N(\alpha)$  is multivariate gamma function.

The expected value is  $E[\mathbf{X}] = n\Sigma$ , and the momentum-generating function is  $M_W(\Sigma) = E[e^{\text{tr}(\mathbf{W}\Sigma)}] = |\mathbf{I} - 2\mathbf{W}\Sigma|^{-\frac{n}{2}}$ . Here,  $\Sigma$  is an empirical covariance matrix derived from historical data, and  $\mathbf{I}$  is an identity matrix.

Consequently, we study the following noise model for the covariance matrix  $\Sigma$ :

$$\mathbf{S} \sim W_N\left(\alpha, \frac{\Sigma}{\alpha}\right) \quad (27)$$

where  $\alpha$  is a parameter that controls the noise level. For large  $\alpha$  values, most of the sample matrices  $\mathbf{S}$  will closely resemble  $\Sigma$ , while for small  $\alpha$  values, the elements of the sample matrices  $\mathbf{S}$  can deviate significantly from the desired covariance matrix  $\Sigma$ . The expected value of  $\mathbf{S}$  is given by the target covariance matrix  $E[\mathbf{S}] = \Sigma$ , and the momentum-generating function is given by:

$$M_W(\Sigma) = \left| \mathbf{I} - \frac{2}{\alpha} \mathbf{W}\Sigma \right|^{-\frac{\alpha}{2}} \quad (28)$$

In Appendix B, we present the distribution of variances and correlation coefficients of two-dimensional matrices, as described by the model in Eq. 27, for various values of the parameter  $\alpha$ .



## 4.2 The expected utility function and optimal weights

In the case of multivariate Gaussian returns  $\mathbf{r}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , the expected utility function to be maximized is proportional to:

$$E_N[U_a(\boldsymbol{\mu}, \boldsymbol{\Sigma})] \sim (-1)e^{\frac{a^2}{2}\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}-a\boldsymbol{\mu}^T\mathbf{w}} \quad (29)$$

Correspondingly, the optimization problem to maximize  $E_N[U_a(\boldsymbol{\mu}, \boldsymbol{\Sigma})]$  is given by:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left[ \boldsymbol{\mu}^T \mathbf{w} - \frac{a}{2} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \right], \quad (30)$$

which has the following classical MV solution:

$$\mathbf{w}^* = \frac{1}{a} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} \quad (31)$$

The effect of the transaction cost can also be taken into account<sup>1</sup>. The minimum variance portfolios with normalized weights can be obtained from MV portfolios Eq. 30 by setting all elements of  $\boldsymbol{\mu}$  to 1. In this case, the term  $\boldsymbol{\mu}^T \mathbf{w}$  becomes a constant, since the sum of the weights  $\sum_{i=1}^n w_i$  is constant if the weights are normalized. This constant term is irrelevant for optimization.

Marginalization of the expected utility function  $E_N[U_a(\boldsymbol{\mu}, \boldsymbol{\Sigma})]$  over the expected returns  $\boldsymbol{\mu}$ , which follow a normal distribution  $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ , can be done analytically:

$$\int_{-\infty}^{\infty} d\boldsymbol{\mu} e^{-a\mathbf{w}^T\boldsymbol{\mu}} N(\boldsymbol{\mu}; \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) = e^{\frac{1}{2}a^2\mathbf{w}^T\boldsymbol{\Sigma}_0\mathbf{w}-a\boldsymbol{\mu}_0^T\mathbf{w}} \quad (32)$$

where  $\boldsymbol{\Sigma}_0$  is a diagonal matrix with elements equal to the variances of the individual components of  $\boldsymbol{\mu}$ .

Integration over all possible realizations of  $\boldsymbol{\Sigma}$ , as given by the noise model in Eq. 27, of the expected utility  $E_N[U_a(\boldsymbol{\mu}, \boldsymbol{\Sigma})]$  in Eq. 29, is provided by:

$$E_{\boldsymbol{\Sigma}}[U_a(\boldsymbol{\mu}, \boldsymbol{\Sigma})] = \int d\boldsymbol{S} E_N[U_a(\boldsymbol{\mu}, \boldsymbol{S})] f(\boldsymbol{S}|\alpha, \frac{\boldsymbol{\Sigma}}{\alpha}) \sim (-1)e^{-a\boldsymbol{\mu}^T\mathbf{w}} |\mathbf{I} - \frac{a^2}{\alpha} \mathbf{W}\boldsymbol{\Sigma}|^{-\frac{\alpha}{2}} \quad (33)$$

Using the identity  $\ln \det(\mathbf{X}) = \text{Tr} \ln(\mathbf{X})$ , the definition of matrix logarithm  $\ln(\mathbf{I} - \mathbf{X}) = \sum_{n=1}^{\infty} (\mathbf{I} - \mathbf{X})^n / n$ , and  $(\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}) = \text{Tr}((\mathbf{w}^T \mathbf{w}) \boldsymbol{\Sigma})$ , we obtain the expected utility after marginalization over random covariance matrices  $\boldsymbol{S}$ :

$$E_{\boldsymbol{\Sigma}}[U(\boldsymbol{\mu}, \boldsymbol{\Sigma})] \sim (-1)e^{-a\boldsymbol{\mu}^T\mathbf{w} - \frac{\alpha}{2} \ln \left[ 1 - \frac{a^2}{\alpha} (\mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}) \right]} \quad (34)$$

---

<sup>1</sup>Additionally, one can take into account the transaction costs by including a term proportional to the turnover,  $e^{\frac{\eta}{2}(\mathbf{w}-\mathbf{w}_0)^2}$ , in the utility function in Eq. 29. Here,  $\mathbf{w}_0$  represents the target weights. Since the form is quadratic in  $\mathbf{w}$ , it leads to a redefinition of  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$ :

$$\boldsymbol{\mu}' = \boldsymbol{\mu} + \frac{\eta}{a} \mathbf{w}_0, \quad \boldsymbol{\Sigma}' = \boldsymbol{\Sigma} + \frac{\eta}{a^2} \mathbf{I}$$

Combining this result with marginalization over the expected returns  $\boldsymbol{\mu}$ , we arrive at:

$$E_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}[U(\boldsymbol{\mu}, \boldsymbol{\Sigma})] \sim (-1)e^{-a\boldsymbol{\mu}_0^T \boldsymbol{w} + \frac{a^2}{2}\boldsymbol{w}^T \boldsymbol{\Sigma}_0 \boldsymbol{w} - \frac{\alpha}{2} \ln \left[ 1 - \frac{a^2}{\alpha} (\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}) \right]} \quad (35)$$

Later in the text, we assume that the parameters  $a$  and  $\alpha$  are set in such a way that the logarithm is real in order to avoid unnecessary notation clutter.

The optimal weights  $\boldsymbol{w}^*$ , considering the uncertain expected return and covariance matrix, are given by:

$$\boldsymbol{w}^* = \arg \max_{\boldsymbol{w}} \left[ \boldsymbol{\mu}_0^T \boldsymbol{w} - \frac{a}{2} \boldsymbol{w}^T \boldsymbol{\Sigma}_0 \boldsymbol{w} + \frac{\alpha}{2a} \ln \left[ 1 - \frac{a^2}{\alpha} (\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}) \right] \right] \quad (36)$$

The optimization problem is convex and can be solved with constraints by a numerical optimizer. In the limit of a large  $\alpha$  (representing a small noise level in the covariance matrix  $\boldsymbol{\Sigma}$ ), the Taylor expansion of Eq. 36 can be expressed as:

$$\boldsymbol{w}^* = \arg \max_{\boldsymbol{w}} \left[ \boldsymbol{\mu}_0^T \boldsymbol{w} - \frac{a}{2} \boldsymbol{w}^T (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_0) \boldsymbol{w} \right] \quad (37)$$

The solution is given by:

$$\boldsymbol{w}^* = \frac{1}{a} (\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_0)^{-1} \boldsymbol{\mu}_0 \quad (38)$$

For the case of zero uncertainty in the expected returns, with  $\boldsymbol{\Sigma}_0 = 0$ , we have the following optimization problem:

$$\boldsymbol{w}^* = \arg \max_{\boldsymbol{w}} \left[ \boldsymbol{\mu}_0^T \boldsymbol{w} + \frac{\alpha}{2a} \ln \left[ 1 - \frac{a^2}{\alpha} (\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}) \right] \right] \quad (39)$$

Maximizing Eq. 39 yields the following system of quadratic equations with respect to  $\boldsymbol{w}$ :

$$\boldsymbol{\mu}_0 \left( 1 - \frac{a^2}{\alpha} \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w} \right) - a \boldsymbol{\Sigma} \boldsymbol{w} = 0 \quad (40)$$

The solution is given by:

$$\boldsymbol{w}^* = \frac{1 - \frac{a^2}{\alpha} d}{a} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 = g_W(q, \alpha) \times \left( \frac{1}{a} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 \right) \quad (41)$$

where  $d = \boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}$ ,  $q = \boldsymbol{\mu}_0^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0$ , and  $g_W(q, \alpha) = 1 - \frac{a^2}{\alpha} d$ . By substituting the solution (41) into the definition of  $d$ , we obtain a quadratic equation with the following solution:

$$d = \frac{\alpha}{2a^2 q} \left( 2q + \alpha - \sqrt{\alpha(4q + \alpha)} \right) \quad (42)$$

The solution provided by Eq. 41 coincides with Eq. 25 in the univariate case.

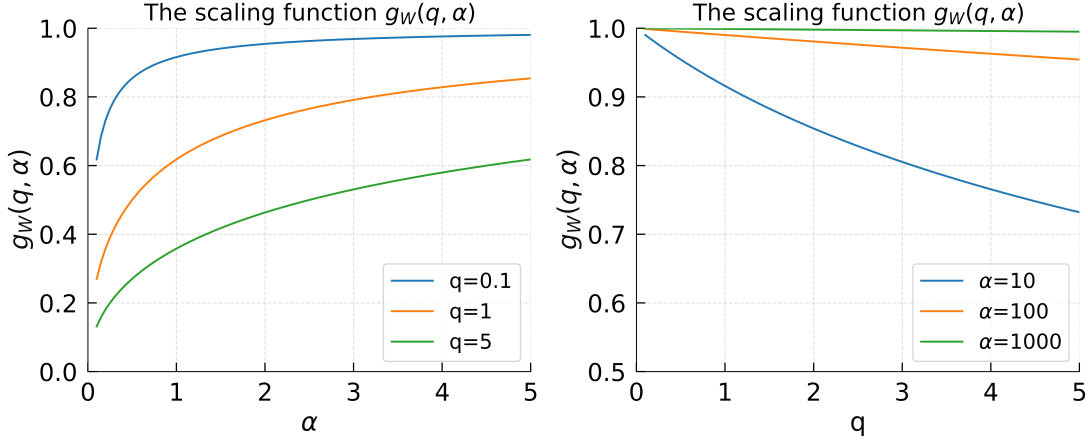


Figure 2: Scaling function  $g_W(q, \alpha)$

We note that the solution in Eq. 41 appears similar to the optimal weights calculated in [3] for outcomes that follow a Laplace distribution:

$$\mathbf{w}_{LD}^* = g_{LD}(q) \times \left( \frac{1}{a} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_0 \right) \quad (43)$$

where

$$g_{LD}(q) = 1 - \frac{a^2 \tilde{d}}{2}, \quad \tilde{d} = \frac{2(1+q - \sqrt{1+2q})}{a^2 q} \quad (44)$$

The scaling function  $0 < g_W(q, \alpha) \leq 1$  does not depend on the risk aversion parameter  $a$ . It represents the decrease in the absolute allocation level due to non-zero uncertainty regarding the future covariance matrix in comparison to the standard MV solution  $\mathbf{w}^* = \frac{1}{a} \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}$ . Thus,  $g_W(q, \alpha)$  can be used to compare allocation levels for different return-to-risk parameters  $q$  and varying levels of covariance matrix noise  $\alpha$ . We present the function  $g_W(q, \alpha)$  in Figure 2 for different values of  $q$  and  $\alpha$ . The scaling function  $g_W(q, \alpha)$  is positive and decreases as the noise level  $\alpha$  increases. This suggests that as uncertainty about future  $\boldsymbol{\Sigma}$  decreases, the allocation level also decreases. In the conventional MV approach, weights are normalized to one  $\sum_{i=1}^N w_i = 1$ , and only relative weights are considered. While using the EU approach, we provide insights into both the absolute level and relative weights as functions of model parameters due to the nature of absolute risk in the utility function. In other words, accounting for uncertainty in the covariance results in a decrease in the absolute level while maintaining the same relative allocation.

If the uncertainty of the expected returns is not zero  $\boldsymbol{\Sigma}_0 \neq 0$ , the optimal weights  $\mathbf{w}^*$  are given by:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left[ \boldsymbol{\mu}_0^T \mathbf{w} - \frac{a}{2} \mathbf{w}^T \boldsymbol{\Sigma}_0 \mathbf{w} + \frac{\alpha}{2a} \ln \left[ 1 - \frac{a^2}{\alpha} \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w} \right] \right] \quad (45)$$

Eq. 45 can be solved by a convex numerical optimizer, such as CVXPY [8], even when additional constraints, such as long-only, maximum number of positions, turnover, and other constraints, are introduced. The unconstrained solution of Eq. 45 can be obtained by taking the derivative with respect to  $\mathbf{w}$ . Thus, we arrive at the following system of cubic equations:

$$(\boldsymbol{\mu}_0 - a\boldsymbol{\Sigma}_0\mathbf{w}) \left(1 - \frac{a^2}{\alpha}\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}\right) - a\boldsymbol{\Sigma}\mathbf{w} = 0 \quad (46)$$

The solution is given by:

$$\mathbf{w}^* = \frac{(1 - \frac{a^2}{\alpha}d)\boldsymbol{\mu}_0}{a(\boldsymbol{\Sigma} + (1 - \frac{a^2}{\alpha}d)\boldsymbol{\Sigma}_0)} \quad (47)$$

where  $d = \mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$ . By substituting the solution from Eq. 47 into the definition of  $d$ , we obtain a nonlinear equation for  $d$  that can be solved numerically. It is worth noting that the uncertainty  $\boldsymbol{\Sigma}_0$  of the  $\boldsymbol{\mu}$  plays a similar role to the regularization (shrinkage) approach detailed in [9]. However, the EU approach provides a clearer interpretation, and there is no need to estimate a shrinkage constant.

In the case of non-zero uncertainty in expected returns, i.e.,  $\boldsymbol{\Sigma}_0 \neq 0$ , both the analytical solution in Eq. 47 and the numerical simulations indicate that the solution of Eq. 45 can deviate (sometimes significantly) from the MV solution  $\mathbf{w}_{MV}^* = \frac{1}{a}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0$ . In this case, the uncertainties of expected returns and the covariance matrix interfere with each other.

### 4.3 Model 2: Marginalization over variance using block covariance matrices

In [3], we proposed deriving the covariance matrix from a precision matrix with a graphical structure determined by conditional independence. The resulting matrix has a block structure. The graphical structure reduces the number of nonzero elements in the covariance matrix, thereby improving the condition number and making the matrix inversion operation more stable. In this section, we investigate allocation rules for two analytically solvable cases of covariance matrices,  $\boldsymbol{\Sigma}_{m1}$  and  $\boldsymbol{\Sigma}_{m2}$ . In the first model, each block has an equivariance structure, but there are no cross-correlation terms between blocks:

$$\boldsymbol{\Sigma}_{m1} = \begin{bmatrix} (\sigma_{min;1}^2 + \sigma_1^2)R_1 & & & \\ & (\sigma_{min;2}^2 + \sigma_2^2)R_2 & & \\ & & \ddots & \\ & & & (\sigma_{min;K}^2 + \sigma_K^2)R_K \end{bmatrix}, \quad (48)$$

In the second model, all blocks have the same variance, but the correlation structure is encoded in equicorrelation blocks:

$$\boldsymbol{\Sigma}_{m2} = (\sigma_{min}^2 + \sigma^2)R_{2m} = (\sigma_{min}^2 + \sigma^2) \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1K} \\ R_{21} & R_{22} & \cdots & R_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ R_{K1} & R_{K2} & \cdots & R_{KK} \end{bmatrix} \quad (49)$$

We marginalize the variance of each block, assuming that the future variance follows a shifted gamma distribution.

If a covariance matrix  $\Sigma$  has a block equivariant structure, the bilinear form  $\mathbf{w}\Sigma_{1m}\mathbf{w}$  is given by the sum over  $K$  blocks:

$$\mathbf{w}\Sigma_{m1}\mathbf{w} = \sum_{i=1}^K \sigma_i^2 \mathbf{w}_i \mathbf{R}_i \mathbf{w}_i \quad (50)$$

where  $\sigma_i^2$ ,  $\mathbf{R}_i$ , and  $\mathbf{w}_i$  represent the  $i$ -th block variance, correlation matrix, and weights, respectively. Assuming that each block variance  $\sigma_i$  follows a shifted gamma distribution, we derive the expected utility function by marginalizing over variance and expected returns:

$$E_{\mu, \Sigma}[U(x)] \sim (-1)e^{-a\mathbf{w}^T \mu + \frac{a^2}{2} \sum_{i=1}^K \sigma_{min;i}^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i} \prod_{i=1}^K \left(1 - \frac{a^2 \sigma_i^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i}{\alpha_i}\right)^{-\frac{\alpha_i}{2}} \quad (51)$$

This leads to the following optimization problem:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left[ \mu_0^T \mathbf{w} - \frac{a}{2} \mathbf{w}^T \Sigma_0 \mathbf{w} - \frac{a}{2} \sum_{i=1}^K \sigma_{min;i}^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i + \sum_{i=1}^K \frac{\alpha_i}{2a} \ln \left[1 - \frac{a^2 \sigma_i^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i}{\alpha_i}\right] \right] \quad (52)$$

Eq. 52 can be solved numerically using a convex optimizer, with additional constraints if needed. In the limit of large  $\alpha$  (small noise level), we have the standard MV accompanied by shrinkage:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left[ \mu_0^T \mathbf{w} - \frac{a}{2} \mathbf{w}^T (\Sigma_{min} + \Sigma + \Sigma_0) \mathbf{w} \right] \quad (53)$$

where  $\mathbf{w}^T \Sigma \mathbf{w} = \sum_{i=1}^K \sigma_i^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i$  and  $\mathbf{w}^T \Sigma_{min} \mathbf{w} = \sum_{i=1}^K \sigma_{min;i}^2 \mathbf{w}_i^T \mathbf{R}_i \mathbf{w}_i$ .

In the second model, each block  $\mathbf{R}_{ij}$  of the covariance matrix  $\Sigma_{2m}$  is represented by an equicorrelation matrix:

$$\mathbf{R}_{ij} = \begin{bmatrix} 1 & \rho_{ij} & \cdots & \rho_{ij} \\ \rho_{ij} & 1 & \cdots & \rho_{ij} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{ij} & \rho_{ij} & \cdots & 1 \end{bmatrix} \quad (54)$$

At first glance, the model  $\Sigma_{2m}$  might seem like a crude approximation. However, in practice, sample estimates may not align with the population parameters of the covariance matrix, making precise modeling unwarranted. Furthermore, instead of the  $N(N+1)/2$  parameters in the original covariance matrix, the matrix  $\Sigma_{2m}$  requires fewer parameters to estimate. Specifically, it only requires a single univariate variance  $\sigma^2$  and  $K(K+1)/2$  correlation coefficients  $\rho_{ij}$ , which can be taken as averages per block [10]. The block assumption ensures that the matrix  $\Sigma_{2m}$  is invertible. In [11], advanced statistical methods were introduced to estimate matrices with such block structures. The

global industry classification standard (GICS) groups emerged as the best model for block selection based on minimizing the Bayesian Information Criterion.

Thus, the optimization problem is as follows:

$$\mathbf{w}^* = \arg \max_{\mathbf{w}} \left[ \boldsymbol{\mu}_0^T \mathbf{w} - \frac{a}{2} \mathbf{w}^T \boldsymbol{\Sigma}_0 \mathbf{w} - \frac{a}{2} \sigma_{min}^2 \mathbf{w}^T \mathbf{R}_{2m} \mathbf{w} + \frac{\alpha}{2a} \ln \left[ 1 - \frac{a^2 \sigma^2}{\alpha} \mathbf{w}^T \mathbf{R}_{2m} \mathbf{w} \right] \right] \quad (55)$$

Based on the symmetry of the problem (assuming that the expected return parameters  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Sigma}_0$  follow the same block structure), stocks within each  $i$ th block have equal weight allocation:

$$\mathbf{w}_i = [w_i, w_i, \dots, w_i] \quad (56)$$

and the optimization in Eq. 55 only needs to find  $K$  block weights  $w_i$ , leading to more stable weights.

#### 4.4 Model 3: Two-state scenario optimization

Stock markets do not grow in a linear trend. There are multiple periods when (almost) all assets experience severe and prolonged declines. In this section, we aim to calculate portfolio allocation rules for situations where a modeler assigns a non-zero probability of a market crash or correction occurring during the expected holding period of the portfolio.

Financial correlation matrices have a unique property: the correlation between stocks increases with volatility. In the event of a market crash, volatility spikes, and almost all correlations approach one. Although it is possible to integrate the correlation coefficient over an appropriate distribution, operating in a discrete space is more practical. In this approach, the future is modeled as a two-state system: a normal regime with parameters  $\boldsymbol{\mu}_n$  and  $\boldsymbol{\Sigma}_n$ , and a stressed regime with parameters  $\boldsymbol{\mu}_s$  and  $\boldsymbol{\Sigma}_s$ . Within this context, the expected value of the utility function is a discrete sum, with probability  $p$  assigned to the normal regime and probability  $1 - p$  to the stressed regime. Consequently, the expected utility of the two-state system  $E_{2s}[U_a]$  is given by:

$$E_{2s}[U_a] = pE[U_a(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)] + (1 - p)E[U_a(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s)] \quad (57)$$

For conceptual clarity, we use the expected utility from Eq.29 instead of Eq.35. Taking the logarithm of Eq. 57, we have the following optimization problem:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} [\log(\exp u_n + \exp u_s)] \quad (58)$$

where  $u_n = \log(p) + \frac{a^2}{2} \mathbf{w}^T \boldsymbol{\Sigma}_n \mathbf{w} - a \boldsymbol{\mu}_n^T \mathbf{w}$  and  $u_s = \log(1 - p) + \frac{a^2}{2} \mathbf{w}^T \boldsymbol{\Sigma}_s \mathbf{w} - a \boldsymbol{\mu}_s^T \mathbf{w}$ . The optimization problem corresponds to the minimization of the LogSumExp (LSE) function [12] and is convex. This convexity can be inferred from the fact that both  $u_n$  and  $u_s$  are convex functions of  $\mathbf{w}$ , and the LogSumExp function is also convex. LogSumExp is an approximation to the maximum  $\max_{i \in \{n, s\}} u_i$  with the following bounds:

$$\max \{u_n, u_s\} \leq \log(\exp u_n + \exp u_s) \leq \max \{u_n, u_s\} + \log(2) \quad (59)$$

In the limit of high risk aversion parameter  $a \rightarrow \infty$ , the equation Eq. 58 simplifies to:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \max_{n,s} \left[ \frac{a}{2} \mathbf{w}^T \boldsymbol{\Sigma}_n \mathbf{w} - \boldsymbol{\mu}_n \mathbf{w}, \frac{a}{2} \mathbf{w}^T \boldsymbol{\Sigma}_s \mathbf{w} - \boldsymbol{\mu}_s \mathbf{w} \right] \quad (60)$$

Thus, we end up optimizing the MV utility function for the corresponding state.

In the limit of low risk aversion parameter  $a \rightarrow 0$ , the exponent in the expected value of  $\log E_{2s}[U_a]$  can be approximated by the Taylor expansion  $e^x \approx 1 + x + \frac{x^2}{2}$  with respect to the risk aversion parameter  $a$  and we have:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \left[ \frac{a}{2} \mathbf{w}^T \tilde{\boldsymbol{\Sigma}} \mathbf{w} - \tilde{\boldsymbol{\mu}}^T \mathbf{w} \right] \quad (61)$$

where  $\tilde{\boldsymbol{\mu}} = p\boldsymbol{\mu}_n + (1-p)\boldsymbol{\mu}_s$  and  $\tilde{\boldsymbol{\Sigma}} = p\boldsymbol{\Sigma}_n + (1-p)\boldsymbol{\Sigma}_s + p(1-p)(\boldsymbol{\mu}_n - \boldsymbol{\mu}_s)^T(\boldsymbol{\mu}_n - \boldsymbol{\mu}_s)$ . Thus  $\mathbf{w}^* = \frac{1}{a} \tilde{\boldsymbol{\Sigma}}^{-1} \tilde{\boldsymbol{\mu}}$ . In one dimensional case, the optimal weight is given by:

$$w^* = \frac{1}{a} \frac{p\mu_n + (1-p)\mu_s}{p\sigma_n^2 + (1-p)\sigma_s^2 + p(1-p)(\mu_n - \mu_s)^2} \quad (62)$$

The minimum variance portfolio with additional constraints  $w > 0$  and  $\sum_{i=1}^N w_i = 1$  can be obtained from Eq. 61 by setting all expected returns as equal ( $\mu_{n;i} = \text{const}_1$  and  $\mu_{s;i} = \text{const}_2$ ). In this scenario, both  $\boldsymbol{\mu}_n^T \mathbf{w}$  and  $\boldsymbol{\mu}_s^T \mathbf{w}$  become constants that are irrelevant for optimization. The optimal weights are then given by:

$$\mathbf{w}^* = \arg \min_{\mathbf{w}} \left[ \mathbf{w}^T \tilde{\boldsymbol{\Sigma}} \mathbf{w} + c \mathbf{w}^T \mathbf{w} \right], \quad \tilde{\boldsymbol{\Sigma}} = p\boldsymbol{\Sigma}_n + (1-p)\boldsymbol{\Sigma}_s \quad (63)$$

We observe that the resulting covariance matrix  $\tilde{\boldsymbol{\Sigma}}$  is shrunk towards the covariance matrix in the stressed regime  $\boldsymbol{\Sigma}_s$  with a shrinkage coefficient of  $1-p$  with added  $L^2$  regularization term  $\mathbf{w}^T \mathbf{w}$ . A similar problem involving the CARA utility function and a mixture of multivariate Gaussian returns was also studied in [13].

In theory, one can extend the formalism to an arbitrary number of market states to better capture the fat tails and skewness of real returns. However, in practice, we need to model future returns and the covariance matrix, both of which are known with significant uncertainty. In such cases, a two-state model appears optimal, and the hypothetical stressed state parameters  $\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s$  should be modeled using the simplest model possible.

Often, there is no particular benefit in precisely modeling the covariance matrix in the stressed regime,  $\boldsymbol{\Sigma}_s$ . For long-only equity strategies, this matrix can be represented in a simplified form as an equicorrelation and equivariance matrix with a high correlation coefficient value,  $\rho_s \approx 0.8$ . Its variance is denoted by  $\sigma_s^2$ . The matrix is defined as:

$$\boldsymbol{\Sigma}_s = \sigma_s^2 ((1 - \rho_s) \mathbf{I} + \rho_s \mathbf{1}\mathbf{1}^T) \quad (64)$$

where  $\mathbf{I}$  is an identity matrix and  $\mathbf{1}\mathbf{1}^T$  is a matrix of ones. In Appendix C, we provide a scatter plot of monthly average volatility vs. monthly average correlation between index constituents for the years 2016 and 2022, which correspond to the normal regime, and for the years 2008 and 2020, which correspond to the stressed regime. The historical values can be used to specify  $\boldsymbol{\Sigma}_s$ .

In general, there are three regimes for correlation coefficients in the stressed regime, and the specifics depend on asset classes and investment strategies. In equities, the correlation between assets tends to approach  $\rho_s = 1$  as all assets move in unison. For statistical arbitrage or relative value strategies, assets that are typically correlated can become anti-correlated, with  $\rho_s = -1$ . Furthermore, during a crash, the correlation between different asset classes might approach zero,  $\rho_s = 0$ , as price dynamics become chaotic. In such scenarios, price movements are largely influenced by liquidity demands or the need to cover margin calls.

## 5 Concluding Remarks

In both our previous work [3] and in this paper, we have investigated possible ways to go beyond the limitations of the classical MV approach. We found it useful to think of MV as a limiting case of the more general CARA expected utility maximization in the case of multivariate Gaussian returns with zero uncertainty of expected returns and the covariance matrix. We demonstrated that it is often technically easier to start the analysis by utilizing the formal definition of expected utility and its optimization.

In the EU approach, the expected value of the utility function can be calculated for returns distributed according to the ALD [3]. This allows for studying the effects of fat tails and skewness in the outcome distribution. We have shown that stock returns on daily, weekly, and monthly scales can be well approximated by the ALD with moderate skewness values. The skewness of the outcome becomes more important for quantitative strategies, like those with trend-following programs, which demonstrate positive skewness, or options or volatility selling, which exhibit negative skewness. Paradoxically, the expected value of the utility function with returns following a skewed normal distribution cannot be expressed as an elementary function. In this sense, the ALD offers the best of both worlds: it takes into account the properties of real-world return distributions and allows for full analytical tractability of the problem.

Uncertainty about the future values of model parameters is a cornerstone assumption in any financial modeling. To account for this, we model the expected returns and the covariance matrix as random variables. The corresponding expected values of the utility function are obtained by marginalization (integration) over all possible values of these random variables. In simple terms, the marginalization of expected returns results in the shrinkage of the covariance matrix. Taking into account the uncertainty of the covariance matrix leads to a decrease in the absolute level of allocation. This fact is missing when returns are normalized.

Finally, we demonstrate that the worst-case scenario allows for an analytical solution using the Karush–Kuhn–Tucker (KKT) method [3]. In the worst-case scenario, the maximum weight is allocated to the worst-performing asset. The expected utility of such a minimax (MM) portfolio has a convex form:

$$\arg \min_w \left[ |\mathbf{w}|_\infty + \frac{b}{2} \mathbf{w}^T \Sigma \mathbf{w} \right], \text{ s.t. } \sum_{i=1}^N w_i = 1, w_i \geq 0 \quad (65)$$



where the infinity norm  $|w|_\infty$  selects the maximum weight, defined as  $|w|_\infty = \max_i w$ . The solution is a combination of uniform and MV weights and does not suffer from the excessive concentration that is typical of the pure MV approach. The utility serves as an alternative to risk-parity, but it is convex and has a clear interpretation.

The diversity of trading styles, asset classes, and the use of leverage by portfolio managers justify the variety of objective functions available for optimization studied in this paper. Also, the exponential utility function is commonly used in many areas of economics and decision-making science. Extensions of MV to skewed, fat-tailed distributions (such as ALD), differentiating between mean optimization and worst-case optimization, and considerations of the uncertainty of outcome distribution parameters are recurring themes in this research. We hope that the results obtained will prove useful in both theoretical and practical contexts.

## Bibliography

- [1] William F. Sharpe. Expected utility asset allocation. *Financial Analysts Journal*, 63(5):18–30, 2007.
- [2] Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.
- [3] Maxime Markov and Vladimir Markov. Portfolio optimization rules beyond the mean-variance approach. *arXiv:2305.10403*, 2023. URL <https://doi.org/10.48550/arXiv.2305.08530>.
- [4] Kenneth J. Arrow. Aspects of the theory of risk-bearing. *Economica*, 33:251, 1966.
- [5] John W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964.
- [6] Fischer Black and Robert Litterman. Asset allocation: Combining investor views with market equilibrium. *Goldman Sachs Fixed Income Research*, 1990.
- [7] Theodore W. Anderson. An introduction to multivariate statistical analysis. *Wiley Interscience*, 2003.
- [8] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.
- [9] Olivier Ledoit and Michael Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621, 2003.
- [10] Michael Burke David M. Corey, William P. Dunlap. Averaging correlations: Expected values and bias in combined pearson rs and fisher’s z transformations. *The Journal of General Psychology*, 125(3):245–261, 1998.

- [11] Ilya Archakov and Peter R. Hansen. A canonical representation of block matrices with applications to covariance and correlation matrices. *The Review of Economics and Statistics*, page 1–39, 2022.
- [12] Wikipedia logsumexp. URL <https://en.wikipedia.org/wiki/LogSumExp>.
- [13] Eric Luxenberg and Stephen Boyd. Portfolio construction with gaussian mixture returns and exponential utility via convex optimization. *Optim Eng*, 2023.
- [14] Gunnar Taraldsen. The Confidence Density for Correlation. *Sankhya A* 85, 600–616, 2023.

## Appendix A: Statistical uncertainty in variance and correlation coefficients

In this section, we revisit well-established results concerning the finite sample uncertainty of variance and correlation coefficients, specifically in the idealized context of Gaussian returns. We anticipate that accounting for distribution with fatter tails would likely heighten the uncertainty estimates. Specifically, we focus on the conditional distribution of the population variance and correlation coefficient, given the sample variance and correlation coefficient. Although we do not apply the derived functional forms of uncertainty in the models discussed in this paper, we present them to illustrate how sample values may deviate from population parameters.

### 5.1 Sample uncertainty of variance

Assume that returns are i.i.d. and generated from a Gaussian distribution  $r_t \sim N(\mu, \sigma^2)$ . The joint distribution of  $\mu$  and  $\sigma^2$ , given data  $D = \{x_1, x_2, \dots, x_n\}$ , is:

$$p(\mu, \sigma^2 | D) \sim \frac{1}{\sigma^{n+2}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \quad (66)$$

Assuming the mean is unknown and the prior for  $\mu$  is  $p(\mu) \sim \text{const}$ , integrating over  $\mu$  yields the marginal distribution of variance  $\sigma^2$  given the sample variance estimate  $s^2$ :

$$p(\sigma^2 | s^2) \sim (\sigma^2)^{-\frac{n+1}{2}} e^{-\frac{(n-1)s^2}{2\sigma^2}} \quad (67)$$

This is equivalent to a scaled inverse chi-squared distribution  $\chi^2(\nu, s^2)$  with parameters  $\nu = n - 1$  and scale  $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ .

In Figure 3, for a commonly used values of  $n = 20, 60, 252$  days for volatility estimation, we present the distribution of the population volatility  $\sigma$  corresponding to sample volatilities  $s = 15\%, 25\%,$  and  $30\%$ . It can be observed that the variance distributions are broad and overlapping. Analytically, for  $x \sim \chi^2(\nu, s^2)$ , the mean is given by  $E[x] = s^2 \frac{n-1}{n-3}$  and the variance by  $Var[x] = s^4 \frac{2(n-1)^2}{(n-3)^2(n-5)}$ . For  $n = 20$ , the mean is  $\frac{19}{17}s^2$ , and the standard deviation is  $0.4s^2$ , with the ratio of the mean of the variance to its standard deviation being  $\frac{19}{17}/0.4 = 2.8$ .

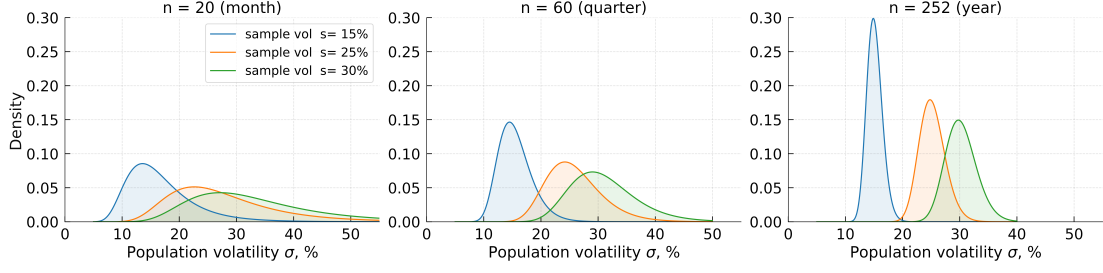


Figure 3: Conditional distribution of population volatility for a finite sample of observations

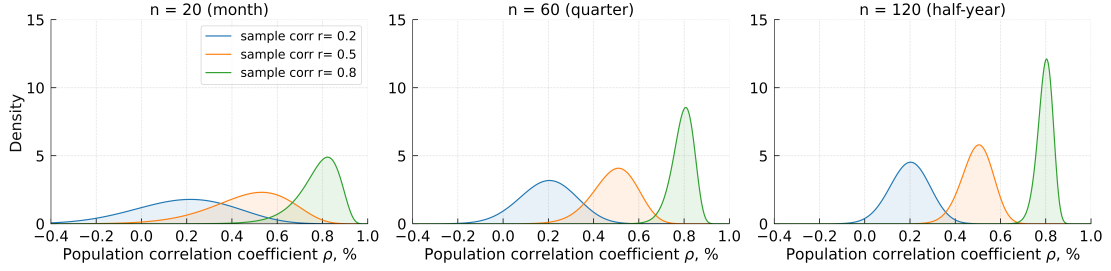


Figure 4: Conditional distribution of correlation coefficient for finite sample of observations

## 5.2 Sample uncertainty of correlation coefficient

The conditional distribution of the correlation coefficient  $\rho$ , given the sample estimation  $r$ , for a bivariate normal distribution was calculated in [14]. The formula for this distribution is as follows:

$$\pi(\rho|r) = \frac{\Gamma(\nu + 1)}{\sqrt{2\pi}\Gamma(\nu + \frac{1}{2})} (1 - r^2)^{\frac{\nu-1}{2}} (1 - \rho^2)^{\frac{\nu-2}{2}} (1 - r\rho)^{\frac{1-2\nu}{2}} {}_2F_1\left(\frac{3}{2}, -\frac{1}{2}, \nu + \frac{1}{2}; \frac{1+r\rho}{2}\right) \quad (68)$$

here,  ${}_2F_1$  is the Gauss hypergeometric function and  $\nu = n - 1$ .

In Figure 4, we present the distribution of the population correlation coefficient  $\rho$  for sample correlations of  $r = 0.2, 0.5$ , and  $0.8$  with sample sizes of  $n = 20, 60$ , and  $120$ .

In the financial industry, the convention is to use daily data with  $n = 20$  observations (a one-month window) for volatility estimation, and a year of data with  $n = 252$  observations for correlation and covariance matrix estimation. This approach can be seen as a practitioner's implicit trade-off between sample uncertainty and uncertainty caused by non-stationarity.

## Appendix B: Distribution of parameters of two-dimensional Wishart matrices

In this section, we simulate the distribution of two-dimensional Wishart matrices as described by the model in Eq. 27:

$$\mathbf{S} \sim W_2\left(\alpha, \frac{\boldsymbol{\Sigma}}{\alpha}\right)$$

with matrix  $\boldsymbol{\Sigma}$  given by:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_A^2 & \rho_{AB}\sigma_A\sigma_B \\ \rho_{AB}\sigma_A\sigma_B & \sigma_B^2 \end{bmatrix}$$

where the volatility of asset A is  $\sigma_A = 0.2$  (20% annualized), that of asset B is  $\sigma_B = 0.4$  (40% annualized), and the correlation coefficient is  $\rho_{AB} = 0.5$ . We simulate  $N = 10^5$  samples and display the distribution of volatilities and the correlation coefficient. The result of the simulation is shown on Figure 5.

## 6 Appendix C: Dependence between volatility and correlation coefficient

In this section, we calculated the average volatility and average correlation coefficient for major stock indexes on a monthly scale for the years 2008, 2016, 2020, and 2022. We then plotted these values on a scatter plot Figure 6. Figures for all indexes are shown in <sup>2</sup>. The constituents of the index were taken from the first day of each respective year. This result can be used to model the covariance matrix in the stressed regime  $\boldsymbol{\Sigma}_s$  in equivariant and equicorrelation approximation.

---

<sup>2</sup>Figures for all indices: [link to the Supplementary Material page on GitHub](#).

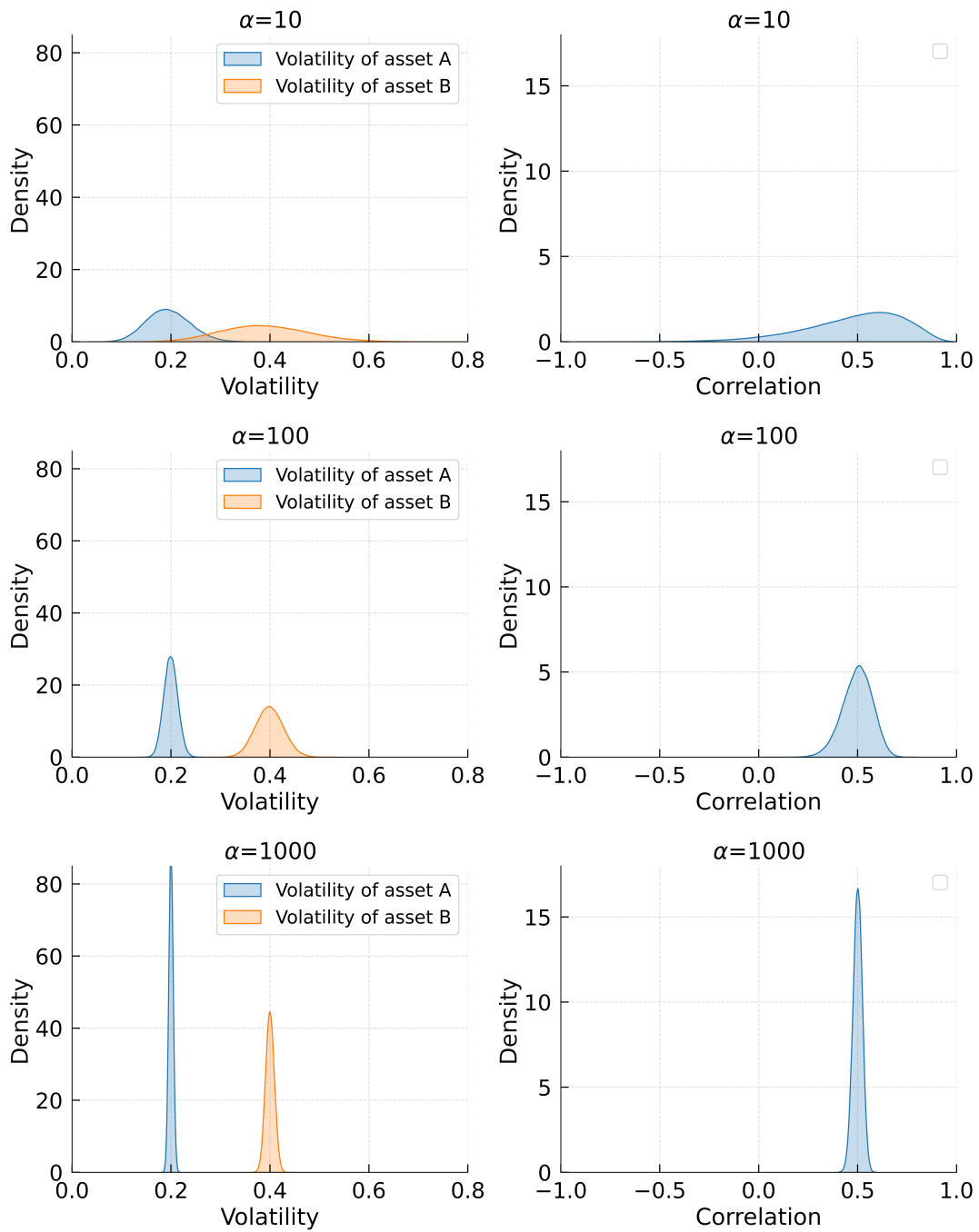


Figure 5: Volatility (left panels) and correlation coefficient (right panels) distribution for  $\alpha = 10$  (high noise level, top panels),  $\alpha = 100$  (medium noise level, middle panels), and  $\alpha = 1000$  (low noise level, bottom panels)

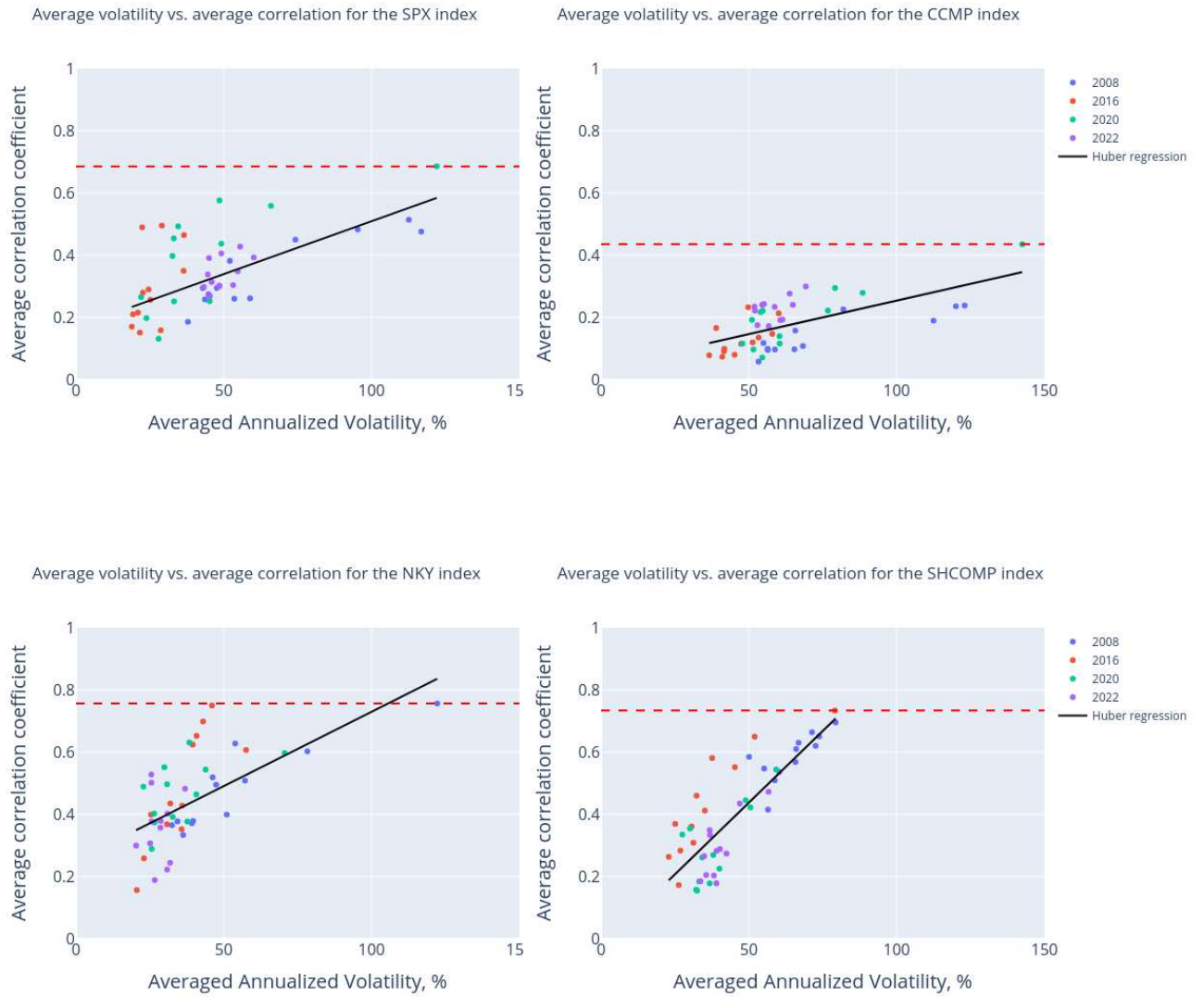


Figure 6: Scatter plot of average volatility versus average correlation coefficient for the SPX, CCMP, NKY, and SHCOMP indices. The red dashed line corresponds to the maximum correlation value, and the black solid line corresponds to the Huber regression curves.