

PORTFOLIO RISK IN MULTIPLE FREQUENCIES

Mustafa U. Torun, Ali N. Akansu, and Marco Avellaneda

Portfolio risk, introduced by Markowitz in 1952, and defined as the standard deviation of the portfolio return, is an important metric in the Modern Portfolio Theory (MPT). A popular method for portfolio selection is to manage the risk and return of a portfolio according to the cross-correlations of returns for various financial assets. In a real world scenario, estimated empirical financial correlation matrix contains significant level of intrinsic noise that needs to be filtered prior to risk calculations. In this paper, we present basic concepts of risk engineering in finance applications. Then, we extend our discussion to the eigenfiltering of measurement noise for hedged portfolios. Moreover, we extend risk measurement methods for trading in multiple frequencies. Finally, three novel risk management methods are proposed as an independent overlay of the underlying investment decision mechanism, i.e. the trading strategy. We highlight performance and merit of the risk engineering techniques introduced by presenting the back-testing results of an investment strategy for the stocks listed in the NASDAQ 100 index. It is shown in the paper that managing portfolio risk more intelligently may offer advantages for improved return on investment.

INTRODUCTION

Portfolio is a collection of investments in various financial assets (e.g. stocks, futures, options, commodities, bonds, etc.) with their allocated investment capitals. Two important aspects of a portfolio are its return and its risk. The main goal of a portfolio manager is to keep the ratio of portfolio return over portfolio risk as high as possible. The standard deviation of a portfolio return is a widely used metric for risk measurement. In Markowitz's celebrated MPT [1], a portfolio is formed, i.e. capital allocation in each asset is decided, by minimizing the risk under the constraint that the expected portfolio return is constant. The portfolio with the minimum risk for a given expected return is called *efficient portfolio*, and the set of efficient portfolios form the *Efficient Frontier* introduced by Markowitz. The main objective in MPT is to keep the portfolio balanced according to cross-correlations of asset pairs. As an example, being simultaneously long (buying the assets with the assumption that the prices will increase in time) in two assets that are positively and highly correlated is a highly risky bet and it must be avoided. However, portfolio size, i.e.

the number of assets in the portfolio, is much larger than just two in reality. In order to measure the risk of an N -asset portfolio, $N(N - 1)/2$ unknown asset pair correlations need to be estimated to create the N -by- N empirical financial correlation matrix, \mathbf{C} . Naturally, \mathbf{C} contains significant amount of inherent noise that needs to be removed. There are several methods in the quantitative finance and signal processing fields, including Principal Component Analysis (PCA), offering efficient noise filtering prior to risk calculations [2–4].

We present PCA based noise filtering of \mathbf{C} , followed by risk measurement for the case of hedged portfolios. In a hedged portfolio, each asset is traded against another asset, usually against an index-tracking asset such as an Exchange Traded Fund (ETF), in order to protect the investment from the volatility (fluctuations) of the market. Next, we argue that, in contrast to the traditional view, an investor may re-balance the portfolio (change the investment amount in each asset), measure and manage the risk at different sampling time intervals or time resolutions. Therefore, a novel extension of the traditional risk metric for trading in multiple frequencies is introduced. Then, three simple-to-implement and yet powerful risk management methods are proposed. The proposed methods can be embedded in any investment decision mechanism, i.e. trading strategy, since they work as independent components. The merit of the proposed risk engineering methods is illustrated by the Profit and Loss (PNL) curves for the back-testing of stocks listed in NASDAQ 100 index.

PORTFOLIO RETURN AND RISK

Consider a portfolio comprised of only two assets. The return of a two-asset portfolio is expressed as

$$R_p(n) = q_1(n)R_1(n) + q_2(n)R_2(n) \quad (1)$$

where n is the discrete time variable, q_i is the amount of capital invested in the i^{th} asset, and R_i is the return of the i^{th} asset defined as

$$R_i(n) = [P_i(n)/P_i(n - 1)] - 1 \quad (2)$$

and $P_i(n)$ is the market price of that asset at discrete time n . The investment amount, q_i in Eq. 1, can be dimensionless or its unit may be a currency. This choice reflects itself into the unit of risk. The time index n is omitted in further discussions

noting that each variable in an equation is a function of time. Expected return of the two-asset portfolio is calculated as

$$\mu_p = E[R_p] = q_1 E[R_1] + q_2 E[R_2]$$

where $E[\cdot]$ denotes the expectation operator. The variance of the portfolio return is expressed as

$$\sigma_p^2 = q_1^2 \sigma_1^2 + 2q_1 q_2 \sigma_1 \sigma_2 \rho_{12} + q_2^2 \sigma_2^2$$

where σ_p is the standard deviation of the portfolio return, i.e. the risk of the portfolio, σ_i is the volatility, i.e. standard deviation of the returns of the i^{th} asset, and ρ_{ij} is the correlation coefficient between the returns of i^{th} and j^{th} assets. It is straightforward to generalize this concept to a portfolio consisting of N assets. The return of the N -asset portfolio is expressed as

$$R_p = \mathbf{q}^T \mathbf{R} = \sum_{i=1}^N q_i R_i \quad (3)$$

where superscript T is the matrix transpose operator, \mathbf{q} is a size N -by-1 investment vector defined as

$$\mathbf{q} = [q_1 \quad q_2 \quad \cdots \quad q_N]^T \quad (4)$$

and \mathbf{R} is an N -by-1 vector comprised of the returns of assets in the investment universe expressed as

$$\mathbf{R} = [R_1 \quad R_2 \quad \cdots \quad R_N]^T$$

Hence, from Eq. 3, expected return of the portfolio is calculated as

$$\mu_p = E[R_p] = \mathbf{q}^T E[\mathbf{R}] = \mathbf{q}^T \boldsymbol{\mu} \quad (5)$$

where $\boldsymbol{\mu}$ is an N -by-1 vector whose elements are expected returns of individual assets. Similarly, the variance of an N -asset portfolio is obtained as

$$\sigma_p^2 = E[R_p^2] - \mu_p^2 = \mathbf{q}^T \boldsymbol{\Sigma}^T \mathbf{C} \boldsymbol{\Sigma} \mathbf{q} = \sum_{i=1}^N \sum_{j=1}^N q_i q_j \rho_{ij} \sigma_i \sigma_j \quad (6)$$

where $\boldsymbol{\Sigma}$ is an N -by- N diagonal matrix with its elements corresponding to the volatilities of individual assets, and \mathbf{C} is the correlation matrix where $[C]_{ij} = \rho_{ij}$. Note that all elements on the main diagonal of \mathbf{C} are equal to one. Furthermore, \mathbf{C} is a symmetric and positive definite matrix.

PORTFOLIO OPTIMIZATION WITH RISK

MPT [1] suggests a method to create efficient portfolios with the minimized risk for a given expected return by optimally allocating the amount of capital invested in each asset of the portfolio. More formally, in MPT, portfolio optimization is achieved by minimizing the portfolio risk, σ_p , given in Eq. 6

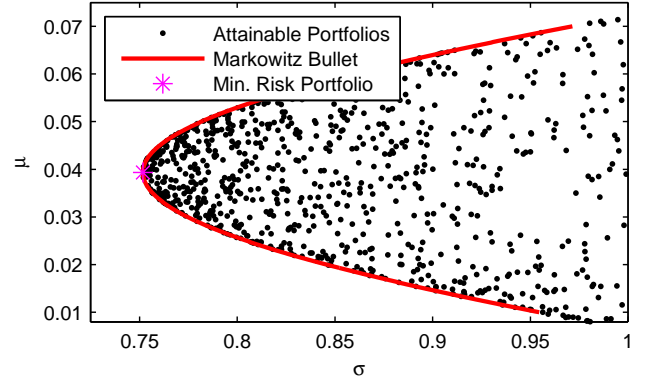


Fig. 1. Markowitz bullet. All the portfolios satisfying the constraint given in Eq. 8 (attainable portfolios) lie on or on the right of the frontier. Portfolios that lie on the upper-half of the Markowitz bullet are called *efficient*. Minimum risk portfolio is located at the far left tip of the bullet. In this example $\rho_{12} = 0.6$, $\rho_{13} = 0.2$, $\rho_{23} = 0.3$, $\mu_1 = 0.07$, $\mu_2 = 0.03$, and $\mu_3 = 0.02$. Some of the attainable portfolios are illustrated as black dots where investment vectors of those portfolios are drawn from a Gaussian joint-probability density function, i.e., $\mathbf{q} \sim N(\mathbf{0}, \mathbf{I})$, and only the ones that satisfy the constraint given in Eq. 8 are kept.

with the constraint that the expected portfolio return, μ_p , of Eq. 5 is equal to a constant, i.e.

$$\mu_p = \mathbf{q}^T \boldsymbol{\mu} = \sum_{i=1}^N q_i \mu_i = \mu \quad (7)$$

where \mathbf{q} is the portfolio investment vector, Eq. 4, and $\boldsymbol{\mu}$ is the expected return vector of the portfolio, Eq. 5. There might be additional constraints such as constant investment capital of portfolio, i.e.

$$\mathbf{q}^T \mathbf{1} = \sum_{i=1}^N q_i = 1 \quad (8)$$

where $\mathbf{1}$ is an N -by-1 vector, i.e. $\mathbf{1} = [1 \quad 1 \quad \cdots \quad 1]^T$. The risk minimization problem to create an efficient portfolio subject to the constraints given in Eqs. 7 and 8 can be solved by introducing two Lagrangian multipliers. Hence, the optimum investment allocation vector providing the minimum risk for a given expected return subject to the constraint of a constant portfolio investment is the solution to the minimization problem stated as

$$\mathbf{q}^* = \underset{\mathbf{q}}{\operatorname{argmin}} \frac{1}{2} \mathbf{q}^T \mathbf{C} \mathbf{q} + \lambda_1 (\mu - \mathbf{q}^T \boldsymbol{\mu}) + \lambda_2 (1 - \mathbf{q}^T \mathbf{1}) \quad (9)$$

where \mathbf{C} is the correlation matrix defined in Eq. 6. Solution to Eq. 9 is given as

$$\mathbf{q}^* = \frac{\begin{vmatrix} \mu & \mathbf{1}^T \mathbf{C}^{-1} \boldsymbol{\mu} \\ 1 & \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} \end{vmatrix} \mathbf{C}^{-1} \boldsymbol{\mu} + \begin{vmatrix} \boldsymbol{\mu}^T \mathbf{C}^{-1} \boldsymbol{\mu} & \mu \\ \boldsymbol{\mu}^T \mathbf{C}^{-1} \mathbf{1} & 1 \end{vmatrix} \mathbf{C}^{-1} \mathbf{1}}{\begin{vmatrix} \boldsymbol{\mu}^T \mathbf{C}^{-1} \boldsymbol{\mu} & \mathbf{1}^T \mathbf{C}^{-1} \boldsymbol{\mu} \\ \boldsymbol{\mu}^T \mathbf{C}^{-1} \mathbf{1} & \mathbf{1}^T \mathbf{C}^{-1} \mathbf{1} \end{vmatrix}} \quad (10)$$

where $|\cdot|$ is the matrix determinant operator, and \mathbf{C}^{-1} is the inverse of \mathbf{C} that always exists since \mathbf{C} is a symmetric and positive definite matrix. Note that, optimum investment vector given in Eq. 10 is not a linear function of the expected returns of the assets in a portfolio. The set of optimum portfolios each satisfying the constraints of Eqs. 7 and 8, for $-\infty < \mu < \infty$ form a curve in the (σ, μ) plane. This curve is called the ‘‘Markowitz Bullet’’ and depicted in Fig. 1 for the case of a three-asset portfolio. Note that all of the (σ_p, μ_p) pairs for *attainable* portfolios (portfolios that satisfy the constraint of Eq. 8) are ‘‘on’’ or ‘‘to the right’’ of the bullet. The portfolios that are ‘‘on’’ the upper-half of the bullet are classified as *efficient* and they form the ‘‘Efficient Frontier.’’ Furthermore, only one of the efficient portfolios has the minimum risk, and therefore, it is called as the ‘‘Minimum Risk (or Variance) Portfolio.’’ The investment vector for the minimum risk portfolio is calculated by solving the following quadratic programming equation

$$\mathbf{q}_{min} = \underset{\mathbf{q}}{\operatorname{argmin}} \frac{1}{2} \mathbf{q}^T \mathbf{C} \mathbf{q} + (1 - \mathbf{q}^T \mathbf{1})$$

The trivial solution is found as

$$\mathbf{q}_{min} = \frac{\mathbf{C}^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1} \mathbf{1}}$$

Note that $(\sigma_{min}, \mu_{min})$ pair corresponding to \mathbf{q}_{min} is located at the far left tip of the Markowitz bullet as highlighted by an asterisk in Fig. 1. The minimum risk portfolio is unique. It has the minimum attainable risk, σ_p ; however, its expected return, μ_p , is not the best possible one.

NOISE IN ESTIMATED CORRELATION MATRIX

Financial correlation matrix \mathbf{C} as defined in Eq. 6 is estimated by using the available market data. *Sample correlation matrix* can be used as an estimate of \mathbf{C} given as

$$\hat{\mathbf{C}} = \frac{1}{M} \mathbf{R} \mathbf{R}^T \quad (11)$$

where \mathbf{R} is the N -by- M asset return matrix, N is the number of assets in the portfolio, and M is the number of available return samples per asset. Each element of \mathbf{R} , namely R_{ij} , is the normalized return of the i^{th} asset at the j^{th} discrete time instance; $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$. Normalization is done such that the return time series of each asset has zero

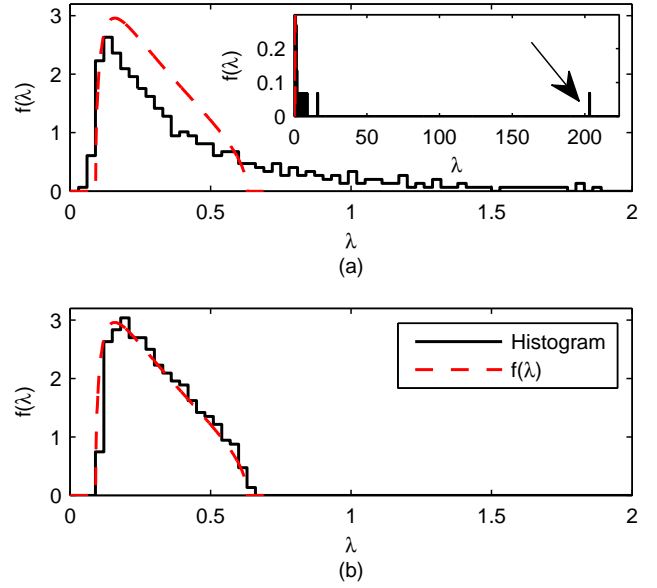


Fig. 2. (a) Histogram of the eigenvalues of the empirical financial correlation matrix, \mathbf{C} , (black solid line) along with the limiting probability density function of the eigenvalues of a random matrix expressed in Eq. 13 (red dashed line). The number of assets utilized is 494, and the number of samples per asset used in estimation is 2444. Since 95% of the eigenvalues are less than 2, the region bounded by $0 \leq \lambda \leq 2$ is zoomed and the full histogram is shown on the top right corner. The largest eigenvalue is marked with an arrow and its value is 203.24. The ‘‘bulk’’ corresponding to the noise in \mathbf{C} , and the deviating eigenvalues related to the valuable correlation information are clearly observed from the figure. (b) Histogram of the eigenvalues of an empirical random matrix constructed by using Eq. 12 along with the probability density function of its eigenvalues in the limit.

mean and unit variance. If the return process were stationary, choosing M as large as possible would be the best approach to improve the estimation. However, in financial processes, anything is hardly stationary. Second, some assets may not have long history. For example, centuries-long historical data for cotton might be available whereas Internet-based investment instruments have been around for only about 15 years. Third, an investor might want to exploit the short-term impacts of certain crisis periods where choosing a long estimation time window may wipe out time local events. Therefore, choosing the observation time window, M , is not a trivial task, and its value depends on the application scenario under consideration.

Now, let us consider an N -by- N random matrix, i.e. the elements of the matrix are random variables, constructed as

$$\mathbf{K} = \frac{1}{M} \mathbf{W} \mathbf{W}^T \quad (12)$$

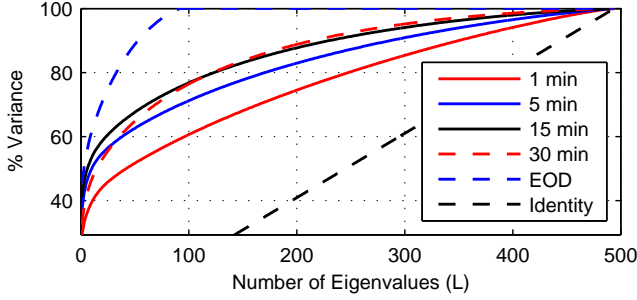


Fig. 3. Scree plot displaying the number of eigenvalues versus percentage of the represented total variance for different sampling intervals and for the $\mathbf{C} = \mathbf{I}$ case, i.e. no correlation between assets. As the sampling interval decreases, eigenspectrum of the correlation matrix becomes more spread. Thus, more eigenvalues are required to represent a certain percentage of the total variance. For the case of 15 min sampling interval, the largest 60 of the 494 eigenvalues, 13%, represent about 70% of the total variance in noisy financial correlation matrix. EOD stands for End of Day sampling rate, i.e. price data is sampled at the market closing of each day.

where \mathbf{W} is an N -by- M matrix comprised of uncorrelated elements drawn from a Gaussian distribution with zero mean and variance σ^2 , i.e. $[W]_{ij} \sim N(0, \sigma^2)$, for $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, M$. Note that \mathbf{K} belongs to the family of Wishart matrices as referred in the multivariate statistical theory. Statistics of random matrices such as \mathbf{K} are extensively studied in the literature [4]. It was shown that the distribution of the eigenvalues of the random matrix \mathbf{K} in the limit is formulated as [5]

$$f(\lambda) = \frac{M}{2\pi\sigma^2 N} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda} \quad (13)$$

where $f(\cdot)$ is the probability density function, $N \rightarrow \infty$, $M \rightarrow \infty$ with the ratio M/N is fixed, and λ_{max} and λ_{min} are the maximum and minimum eigenvalues of the matrix \mathbf{K} , respectively in the same limiting case, with the values defined as

$$\lambda_{max,min} = \sigma^2 \left(1 + \frac{N}{M} \pm 2\sqrt{\frac{N}{M}} \right) \quad (14)$$

If $\hat{\mathbf{C}}$ were a random matrix, its eigenvalues would be samples drawn from the distribution given in Eq. 13. In order to be able to check whether $\hat{\mathbf{C}}$ is consistent with this distribution or not, we estimate a financial correlation matrix $\hat{\mathbf{C}}$ using the estimator expressed in Eq. 11. Then, we decompose it into its eigenvectors with the corresponding eigenvalues, and we calculate the histogram of its eigenvalues [2, 3]. In this example, the time interval of return data is chosen to be 15 minutes, and the time window for the estimation is defined between Jan. 4, 2010 and May 18, 2010. Hence, $M = 2444$

(94 business days and 26 data samples per day). The number of assets in the investment universe is 494 of 500 stocks listed in S&P 500 index, $N = 494$ and $M/N = 4.95$. The return series of individual stocks are normalized to have zero mean and unit variance. The histogram of the eigenvalues of the estimated financial correlation matrix is displayed in Fig. 2.a along with the probability density function of the eigenvalues of a random matrix expressed in Eq. 13, and calculated with $M/N = 4.95$, $\sigma^2 = 0.3$, $\mu = 0$, $\lambda_{max} = 0.63$, and $\lambda_{min} = 0.091$ according to Eq. 14. It is inferred from the figure, by setting the parameter $\sigma^2 = 0.3$ in Eq. 13, that a reasonable fit to the empirical data for eigenvalues smaller than 0.63 is achievable. This suggests that about 30% of the energy in matrix \mathbf{C} is random. Hence, eigenvalues smaller than 0.63 can be considered as noise. The number of eigenvalues, L , versus the percentage of the total variance represented is displayed in Fig. 3 for different sampling intervals. For the case of 15 min sampling interval, the largest 60 of the 494 eigenvalues, 13%, represent about 70% of the total variance in noisy matrix \mathbf{C} . Moreover, only the largest 4 eigenvalues represent 50% of the total energy. The maximum and minimum eigenvalues of the estimated correlation matrix $\hat{\mathbf{C}}$, $\lambda_{max}^{\hat{\mathbf{C}}}$ and $\lambda_{min}^{\hat{\mathbf{C}}}$, are equal to 203.24 and 0.044, respectively. Note that the largest eigenvalue is approximately 322 times larger than its counterpart in a random matrix. For the purpose of validation, a random matrix \mathbf{K} of Eq. 12 is generated with the parameters $M/N = 4.95$, $\sigma^2 = 0.3$, and $\mu = 0$. The histogram of the eigenvalues of \mathbf{K} along with the probability density function of the eigenvalues for a random matrix as defined in Eq. 13 using the same parameters is displayed in Fig. 2.b. It is observed from the figure that empirical histogram fits quite nicely to the theoretical limiting distribution.

The histogram of the eigenvalues for the empirical financial correlation matrix has two major clusters as displayed in Fig. 2. Namely, there is a “bulk” of eigenvalues that are strongly related to the noise, and the remaining relatively small number of eigenvalues deviating from the bulk representing the “valuable information.” In the next section, we employ Principal Component Analysis (PCA) to filter out the noise component of the empirical financial correlation matrix. Then, we utilize the filtered correlation matrix in the calculation of portfolio risk, Eq. 6, and we also emphasize its effects on portfolio optimization and performance.

EIGENFILTERING OF NOISE IN ESTIMATED CORRELATION MATRIX

Let us decompose the estimated correlation matrix into its eigenvectors. We express the correlation matrix \mathbf{C} as

$$\mathbf{C} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \quad (15)$$

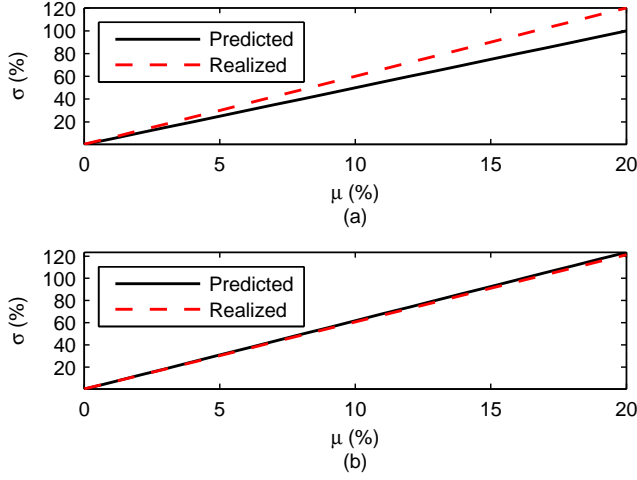


Fig. 4. (a) Predicted and realized risk functions versus target portfolio returns for the set of efficient portfolios. In this case, noisy empirical financial correlation matrices used without any filtering. The Root Mean Square (RMS) value of the error defined in Eq. 20 is $\sim 16.8\%$. (b) Empirical financial correlation matrices are filtered prior to risk calculation according to Eq. 16. The number of eigenvalues is set to $L = 4$ (It accounts for 52% of the total variance). The RMS value of the error is significantly reduced to $\sim 2.3\%$ by employing PCA.

where the k^{th} element of the diagonal matrix $\mathbf{\Lambda}$ is the k^{th} eigenvalue λ_k

$$\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$$

and $\lambda_k \geq \lambda_{k+1}$. In Eq. 15, \mathbf{V} is an N -by- N matrix with N eigenvectors as its columns as defined

$$\mathbf{V} = [\mathbf{v}_1 \quad \mathbf{v}_2 \quad \dots \quad \mathbf{v}_N]$$

where \mathbf{v}_k is an N -by-1 eigenvector corresponding to the k^{th} eigenvalue, λ_k . Note that all eigenvalues are positive, i.e. $\lambda_k > 0 \forall_k$ due to the positive definite property of the matrix \mathbf{C} . Now, we substitute Eq. 15 into Eq. 6, and the variance of the portfolio is formulated as

$$\sigma_p^2 = \mathbf{q}^T \mathbf{\Sigma}^T \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T \mathbf{\Sigma} \mathbf{q}$$

We follow the approach proposed in [3] and define eigenfiltered correlation matrix as

$$\tilde{\mathbf{C}} = \sum_{k=1}^L \lambda_k \mathbf{v}_k \mathbf{v}_k^T + \mathbf{E} \quad (16)$$

where L is the number of selected factors (eigenvalues) satisfying $\lambda_k \geq \lambda_{max}$, and $L \ll N$. Identifying L is not a trivial task. Although Eq. 14 offers a framework to calculate its theoretically optimum value, in practice, investors might create

a screen plot as displayed in Fig. 3 to decide on the value of L with a corresponding fraction of the total variance. Alternatively, back-testing with different values of L might also be found useful. The diagonal noise matrix \mathbf{E} is introduced in Eq. 16 in order to preserve the total variance in the calculations. The elements of \mathbf{E} are defined as

$$[E]_{ij} = \epsilon_{ij} = \begin{cases} 1 - \sum_{k=1}^L \lambda_k v_i^{(k)} v_j^{(k)} & i = j \\ 0 & i \neq j \end{cases} \quad (17)$$

where $v_i^{(k)}$ is the i^{th} element of the k^{th} eigenvector. Note that the addition of matrix \mathbf{E} is equivalent to setting the diagonal elements $[\tilde{\mathbf{C}}]_{ii} = 1$, and it is required to keep $\text{Tr}(\mathbf{C}) = \text{Tr}(\tilde{\mathbf{C}})$. From Eqs. 16 and 17 we express the elements of the noise filtered correlation matrix as follows

$$[\tilde{\mathbf{C}}]_{ij} = \tilde{\rho}_{ij} = \sum_{k=1}^L \lambda_k v_i^{(k)} v_j^{(k)} + \epsilon_{ij} \quad (18)$$

Now, we substitute the filtered version of ρ_{ij} , Eq. 18, into Eq. 6, and we obtain

$$\tilde{\sigma}_p^2 = \sum_{k=1}^L \lambda_k \left(\sum_{i=1}^N q_i v_i^{(k)} \sigma_i \right)^2 + \sum_{i=1}^N \epsilon_{ii} q_i^2 \sigma_i^2 \quad (19)$$

Next, we investigate the impact of using noise filtered estimated variance Eq. 19 rather than noisy estimated variance Eq. 6 for risk calculations with an example following similar steps provided in [2, 3]. In this example, the data set consists of return time series for 494 stocks listed in S&P 500 index, and there are two time periods. First and second time periods include the business days from Jan 4, 2010 to May 18, 2010, and from May 19, 2010 to Sep 30, 2010, respectively. Suppose that, on the morning of May 19, 2010, i.e. the first day of the second time period, a risk manager has the perfect prediction of the future returns; i.e. the expected return vector of the second period, $\boldsymbol{\mu}_2$ in Eq. 5, is exactly known. On that morning, the risk manager is asked to create a portfolio for a given target portfolio return, μ . First, the risk manager calculates the sample correlation matrix for the first time period, $\hat{\mathbf{C}}_1$, using Eq. 11. Then, the manager obtains the investment vector, \mathbf{q}^* , using using Eq. 10 with $\hat{\mathbf{C}}_1$, $\boldsymbol{\mu}_2$, and μ . Next, the predicted risk of the portfolio for the second time period, $\hat{\sigma}_2$, using Eq. 6 with \mathbf{q}^* and $\hat{\mathbf{C}}_1$ is calculated. Finally, on Sep 30, 2010, i.e. the last day of the second time period, the empirical correlation matrix for the second time period, $\hat{\mathbf{C}}_2$, and the realized risk of the portfolio for the second time period, σ_2 , using Eq. 6 with \mathbf{q}^* and $\hat{\mathbf{C}}_2$ are found. Obviously, in this setup, the investment vector calculated via Eq. 10 is a function of only the target portfolio return, μ , and the correlation matrix, \mathbf{C} . Moreover, portfolio risk calculated via Eq. 6 is a function of the investment vector, \mathbf{q}^* , and the correlation matrix, \mathbf{C} . Hence, portfolio risk is a function of the

target portfolio return, μ , and the correlation matrix, \mathbf{C} . We formalize it as follows

$$\begin{aligned}\hat{\sigma}_2(\mu) &= f(\hat{\mathbf{C}}_1, \mu) \\ \sigma_2(\mu) &= f(\hat{\mathbf{C}}_2, \mu)\end{aligned}$$

where $f(\cdot)$ defines a function that involves the calculation of the optimum investment vector defined in Eq. 10, and using it in Eq. 6 in order to obtain the portfolio risk. Plots of $\hat{\sigma}_2(\mu)$ and $\sigma_2(\mu)$ as a function of μ are given in Fig. 4.a with black-solid and red-dashed lines, respectively. It is also possible for the risk manager to check how large the risk was under- or over-estimated by defining an error function as given

$$\epsilon(\mu) = [\hat{\sigma}_2(\mu) - \sigma_2(\mu)] / \sigma_2(\mu) \quad (20)$$

Root Mean Square (RMS) value of this error function is ~16.8% for the case displayed in Fig. 4.a. Now, suppose the risk manager does everything the same but this time, the filtered empirical correlation matrices are used, $\hat{\mathbf{C}}_1$ and $\hat{\mathbf{C}}_2$, as defined in Eq. 16, prior to creating the portfolio and calculating the corresponding risk functions, $\hat{\sigma}_2(\mu)$ and $\sigma_2(\mu)$. The results for this case are given in Fig. 4.b. The RMS value of the error defined in Eq. 20 is ~2.3%. It is observed that noise-free correlation matrix is more stable and it lets the risk manager predict the portfolio risk better. Although the example given in this section compares only two time periods and it is not 100% realistic due to the assumption that the future expected return vector, μ_2 , is known, it is sufficient to emphasize the significance of built-in noise in the correlation matrix and the importance of filtering it out. In practice, the parameters of the system, such as the number of eigenvalues, L , must be back-tested over several time periods to build a level of confidence.

EXTENSION OF NOISE FILTERING AND RISK MEASUREMENT TO HEDGED PORTFOLIOS

In finance, it is a common practice to *hedge* an investment to protect it from the volatility of the market. Hedging means to trade an asset against another asset. In general, the return of an asset is regressed on the return of another asset (commonly a financial instrument that tracks an index, e.g., SPY tracking S&P 500, QQQQ tracking NASDAQ 100) such that

$$R = \alpha + \beta Y + \xi$$

where R and Y are the returns of the asset and the relevant index-tracking asset, respectively, α is the excess return of the former over the latter, β is the *hedging factor* representing the proportion of asset's return that moves with the index, and ξ is the regression noise. Note that α , β , and ξ are also commonly

referred to as drift, systematic component, and idiosyncratic component, respectively. The decision here is to buy the asset with \$1 and short-sell (sell the borrowed assets due to the expectation that the prices will decrease) the index-tracking asset with β , or vice versa depending on the expectation of the investor. Therefore, the return of this investment for the former is expressed as follows

$$R_{inv} = R - \beta Y = \alpha + \xi \quad (21)$$

By hedging, the impact of the market on the stock is (ideally) removed as shown in Eq. 21. This investment is profitable if $\alpha > 0$ and $E[\xi] \geq 0$. Indeed, $E[\xi] = 0$, almost by definition, due to the Least Squares (LS) algorithm employed in regression provided that the random variables are stationary within the estimation time window. The caveat in hedging is the fact that entering and exiting from a trade require four transactions that might result in a higher transaction cost per trade than the achieved return.

We consider a portfolio comprised of N assets and M index-tracking assets where $N \geq M$. The return of this index-hedged portfolio is

$$R_p = A + I = \sum_{i=1}^N q_i R_i + \sum_{j=1}^M g_j Y_j \quad (22)$$

where A and I are the total returns on the investment for assets and index-tracking assets, respectively; R_i and Y_j are the returns of the i^{th} asset and the j^{th} index-tracking asset, respectively; and q_i and g_j are the amounts of capital invested in the i^{th} asset and the j^{th} index-tracking asset, respectively. Hence, assuming that the returns of the assets in the portfolio have zero mean, the variance of the portfolio return, is expressed as

$$\sigma_p^2 = E[R_p^2] = E[A^2] + 2E[AI] + E[I^2] \quad (23)$$

It follows from Eqs. 22 and 23 that one needs to estimate cross-correlations of asset returns in order to measure the risk as follows

$$E[A^2] = \sum_{i=1}^N \sum_{j=1}^N q_i q_j E[R_i R_j] \quad (24)$$

and the cross-correlations between returns of assets and index-tracking assets are written as

$$E[AI] = \sum_{i=1}^N \sum_{j=1}^M q_i g_j E[R_i Y_j] \quad (25)$$

Similarly, cross-correlations of index-tracking asset returns are defined as

$$E[I^2] = \sum_{i=1}^M \sum_{j=1}^M g_i g_j E[Y_i Y_j] \quad (26)$$

Following the steps explained earlier, we calculate the estimated portfolio variance by using filtered version of the empirical correlation matrix for the given hedged portfolio as follows

$$\begin{aligned} \tilde{\sigma}_p^2 = & \sum_{k=1}^L \lambda_k \left(\sum_{i=1}^N q_i v_i^{(k)} \sigma_i + \sum_{j=1}^M g_j \beta_{j,k} \right)^2 \\ & + \sum_{i=1}^N \epsilon_{ii} q_i^2 \sigma_i^2 + \sum_{j=1}^M \xi_j^2 g_j^2 \end{aligned} \quad (27)$$

where λ_k is the k^{th} eigenvalue, $v_i^{(k)}$ is the i^{th} element of the k^{th} eigenvector, σ_i is the volatility of the asset, $\beta_{j,k}$ is the hedging factor for the j^{th} index-tracking asset and k^{th} eigenportfolio (defined in Appendix), and ξ_j^2 is the variance of the idiosyncratic component of the returns for the j^{th} index-tracking asset regressed on L eigenportfolios. Proof of Eq. 27 is given in Appendix.

TRADING IN MULTIPLE FREQUENCIES

In finance, *frequency* in general means the speed of rebalancing a portfolio, i.e. changing the investment amounts invested in each asset in the portfolio. In most of the literature and in our prior discussions, it is assumed that the frequency of trading for each asset is the same. However, trading in multiple frequencies, i.e. trading different assets at different times, may be desirable for the investors due to several reasons including a) Liquidity, i.e. the availability of the asset in the market, may not be the same for all the assets in the portfolio. Therefore, the investor may want to trade certain assets faster or slower than others, b) Different assets may reveal certain aspects of the market the investor is looking for such as a trend, a relative-value etc., at different sampling frequencies, c) A high frequency investor may want to keep the portfolio diverse and balanced in terms of risk, and the traditional methods for measuring and managing risk require relatively high correlations between the assets. However, cross-correlation of asset returns is reduced as the sampling frequency increases due to the well-known phenomenon in finance called the Epps effect [6]. This is also observed in Fig. 3, as the sampling frequency increases the eigenspectrum of the correlation matrix becomes more spread, and it suggests lower correlations among the assets. Thus, more eigenvalues are required to represent a certain percentage of the total variance. Therefore, in order to accommodate the novel concept of trading in multiple frequencies, the risk definition of Eq. 6 must be properly modified. Now, we define the log-return of an asset, $G(n)$, as follows

$$S(n) = S(n-1) + G(n) \quad (28)$$

where $S(n) = \ln P(n)$ is the natural logarithm of the asset price at discrete time index n . Note that the return defined in Eq. 2 is an approximation to the log-return due to the Taylor series expansion of the logarithm, i.e.

$$G(n) = \ln [P(n)/P(n-1)] \simeq [P(n)/P(n-1)] - 1 = R(n)$$

It follows from Eq. 28 that the next sample of the log-price is expressed as

$$S(n+1) = S(n) + G(n+1)$$

Similarly, the k^{th} sample of the log-price is written as follows

$$S(kn) = S(0) + \sum_{i=1}^k G(i) \quad (29)$$

If the sampling period in Eq. 28 is T_s , then the sampling period of the sub-sampled log-price at rate k is kT_s , and the return of the sub-sampled log-price is equal to

$$G_k(l) = S_k(l) - S_k(l-1) \quad (30)$$

where $l = kn$ is the sample index at the new time resolution, and $S_k(l)$ is equal to $S(kn)$ as defined in Eq. 29. By substituting Eq. 29 into Eq. 30 we obtain

$$G_k(l) = \sum_{i=1}^k G(i)$$

Although the Epps effect [6] is not explained by the Black-Scholes pricing model [7], for the sake of simplicity, let us assume that the asset log-prices follow a geometric Brownian motion path [7]. Then, according to Eq. 28, the return is a white Gaussian noise with mean μ and standard deviation (volatility) σ , i.e. $G(n) \sim N(\mu, \sigma^2)$. Since sum of k Gaussian random variables is also a Gaussian random variable, the return process of the sub-sampled log-price process given in Eq. 29 is also distributed as Gaussian with mean $k\mu$ and standard deviation $\sqrt{k}\sigma$, i.e.

$$G_k(l) \sim N(k\mu, k\sigma^2) \quad (31)$$

It follows from Eq. 31 that volatilities estimated at different sampling frequencies have the following relationship

$$\sigma_1 = \sqrt{k_1/k_2} \sigma_2 = m \sigma_2 \quad (32)$$

where σ_1 and σ_2 are the volatilities estimated at sampling intervals $k_1 T_s$ and $k_2 T_s$, respectively, and T_s is the base sampling interval. Hence, it is possible to measure portfolio risk at a certain time interval, and manage risk of assets by rebalancing the individual investment allocations at different time intervals, by expanding the original risk formula of Eq. 6 as follows

$$\sigma_p^2 = \mathbf{q}^T \Sigma^T \mathbf{M}^T \mathbf{C} \mathbf{M} \Sigma \mathbf{q} \quad (33)$$

where $\mathbf{M} = \text{diag}(m_1, m_2, \dots, m_N)$ and m_i is the scaling factor of Eq. 32 provided that Σ and \mathbf{C} matrices are estimated at $k_2 T_s$ time intervals.

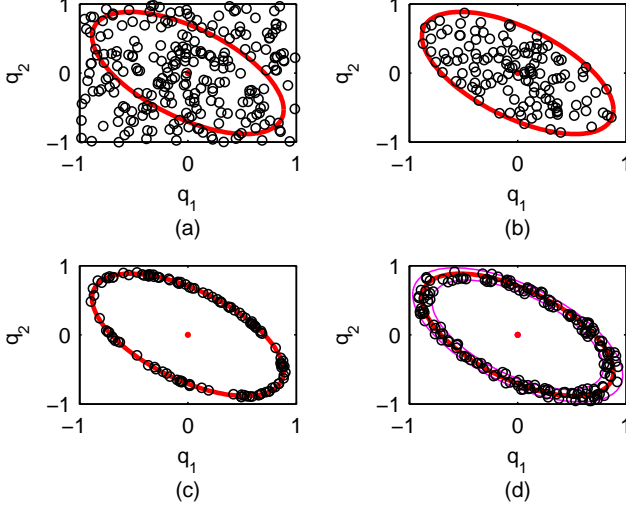


Fig. 5. (a) Possible risk locations of a two-asset portfolio (circles) and the risk ellipsoid (red and solid line) with $\sigma_{THR} = \sqrt{0.5}$ (See Eq. 36), $\rho_{12} = 0.6$, and $\sigma_1 = \sigma_2 = 1$. Risk of a non-managed portfolio may be in, out of, or on the risk ellipsoid. (b) Stay in the ellipsoid (SIE), (c) Stay on the ellipsoid (SOE), and (d) Stay around the ellipsoid (SAE) with $\Delta = \sqrt{0.1}$ risk management methods. Note that $\mathbf{q} \sim N(\mathbf{0}, \mathbf{I})$ in this example.

RISK MANAGEMENT

Consider a real world scenario where an independent investment strategy constantly rebalances a given portfolio and its details are not necessarily known by the risk manager. Here, the goal is to manage the portfolio risk by filtering the decisions of the underlying investment strategy based on a predetermined risk limit. The locus of q_1, q_2, \dots, q_N satisfying Eq. 6 for a fixed value of σ_p is an ellipsoid centered at the origin. The ellipsoids are nested since as σ_p increases, the ellipsoids become larger. The shape of the ellipsoid is defined by the asset return correlation matrix \mathbf{C} . The risk ellipsoid for the trivial case of two-asset portfolio is displayed in Fig. 5.a. Depending on the investment vector \mathbf{q} , the risk of a non-managed portfolio, depicted by black circles in Fig. 5, may be in, out of, or on the risk ellipsoid. We propose three novel risk management methods. They are called “stay in the ellipsoid (SIE)”, “stay on the ellipsoid (SOE)”, and “stay around the ellipsoid (SAE)” methods, respectively. Although the first method is the simplest one to implement, performances of the second and third ones are superior to the first. Note that, in all methods, we assume that once a *signal to enter* a new position is rejected by the risk manager, the underlying strategy does not create another signal until the *signal to exit* is generated.

STAY IN THE ELLIPSOID METHOD

Stay in the ellipsoid (SIE) method wants to keep the portfolio risk anywhere inside the predefined risk ellipsoid by checking the risk of the target portfolio, and rejecting any new investment position that violates this requirement. SIE risk management method is expressed as

$$\mathbf{q}_{t+\Delta t} \leftarrow \begin{cases} \mathbf{q}_{t+\Delta t} & \sigma_{t+\Delta t} < \sigma_{MAX} \\ \mathbf{q}'_{t+\Delta t} & \sigma_{t+\Delta t} \geq \sigma_{MAX} \end{cases} \quad (34)$$

where Δt is the time interval between the two consecutive rebalances of the portfolio, σ_{MAX} is a predetermined maximum allowable risk level, and $\mathbf{q}'_{t+\Delta t}$ is the modified capital investment vector achieved according to the new investment allocation rules

$$[q'_{t+\Delta t}]_i = \begin{cases} 0 & q_{t,i} = 0 \text{ and } |q_{t+\Delta t,i}| > 0 \\ q_{t+\Delta t,i} & \text{otherwise} \end{cases} \quad (35)$$

where $q_{t,i}$ is the investment amount in the i^{th} asset at time t . Note that Eq. 35 employs an *all-or-none* approach in order to expose the portfolio to each asset in approximately equivalent amounts. It is observed from Eqs. 34 and 35 that the proposed method rejects any new investment position in the target portfolio whenever it generates a target risk higher than the maximum allowable risk. Risk locations of possible two-asset portfolios are displayed in Fig. 5.b along with the limiting risk ellipsoid. Note that any re-balancing act of the portfolio taking the risk beyond any target risk point outside of the risk ellipsoid is not permitted in the SIE risk management method. However, it is still possible for the portfolio to move out of the risk ellipsoid due to the abrupt, unavoidable changes in the returns of the assets that are invested in.

STAY ON THE ELLIPSOID METHOD

Stay on the ellipsoid (SOE) risk management method tries to keep the portfolio risk not only inside the risk ellipsoid but also as close to it as possible. The difference between the SIE and SOE methods is observed from Figs. 5.b and 5.c for the case of a two-asset portfolio. The latter maximizes the utilization of the allowable risk limits, and it is formulated as follows

$$\mathbf{q}_{t+\Delta t} \leftarrow \begin{cases} \mathbf{q}_{t+\Delta t} & \sigma_{t+\Delta t} < \sigma_{THR} \\ \mathbf{q}'_{t+\Delta t} & \sigma_{t+\Delta t} \geq \sigma_{THR} \end{cases} \quad (36)$$

where σ_{THR} is the desired risk threshold, and $\mathbf{q}'_{t+\Delta t}$ may be obtained by employing various optimization algorithms to minimize the risk distance as expressed

$$\mathbf{q}'_{t+\Delta t} = \underset{\mathbf{q}}{\operatorname{argmin}} |\sigma_{THR} - \sigma(\mathbf{q})| \quad (37)$$

$\sigma(\mathbf{q})$ is the calculated risk for the investment allocation vector \mathbf{q} given that its elements are limited to

$$q_i \in \begin{cases} \{0, q_{t+\Delta t, i}\} & q_{t, i} = 0 \text{ and } |q_{t+\Delta t, i}| > 0 \\ \{q_{t+\Delta t, i}\} & \text{otherwise} \end{cases} \quad (38)$$

where the notation $\{\cdot\}$ defines a set of numbers. Eqs. 36, 37, and 38 suggest to search for a specific combination of signals (flags generated by the underlying decision mechanism to enter a trade) to open new investment positions targeting the portfolio risk level as close to its limits as possible. The intuition here is to maintain a relatively diverse portfolio while keeping the risk within a desired limit. For the two-asset portfolio case, the solution for the optimization problem is trivial. However, the optimization problem of an N -asset portfolio might become computationally intensive, particularly when N is large.

STAY AROUND THE ELLIPSOID METHOD

The idea behind the stay around the ellipsoid (SAE) risk management method is similar to the one for SOE. However, SAE introduces more flexibility by defining a risk *ring* with the help of the two risk ellipsoids located at a fixed distance around the target risk ellipsoid. The difference between SOE and SAE methods is observed from Figs. 5.c and 5.d for the two-asset portfolio case. In SAE, it is less likely for a candidate portfolio to be rejected due to the increased flexibility. SAE allows new positions only if the target portfolio risk stays inside the ring. It is expressed as

$$\mathbf{q}_{t+\Delta t} \leftarrow \begin{cases} \mathbf{q}_{t+\Delta t} & \sigma_{MIN} < \sigma_{t+\Delta t} < \sigma_{MAX} \\ \mathbf{q}'_{t+\Delta t} & \text{otherwise} \end{cases} \quad (39)$$

where $\sigma_{MIN} = \sigma_{THR} - \Delta$ and $\sigma_{MAX} = \sigma_{THR} + \Delta$ are the minimum and maximum allowable risk levels defining the risk ring, and Δ is the distance to the target risk. The modified investment vector $\mathbf{q}'_{t+\Delta t}$ in Eq. 39 may be obtained by solving the multi-objective minimization problem defined as

$$\mathbf{q}'_{t+\Delta t} = \underset{\mathbf{q}}{\operatorname{argmin}} [f(\mathbf{q}), g(\mathbf{q})]^T \quad (40)$$

where objective functions f and g are defined as

$$\begin{aligned} f(\mathbf{q}) &= \sigma(\mathbf{q}) - \sigma_{MIN} \\ g(\mathbf{q}) &= \sigma_{MAX} - \sigma(\mathbf{q}) \end{aligned}$$

where $\sigma(\mathbf{q})$ is the calculated risk for investment allocation vector \mathbf{q} . Its elements are defined in Eq. 38. Eq. 40 may be obtained by creating an aggregate objective function or via various multi-objective optimization algorithms such as successive Pareto optimization [8], evolutionary algorithms [9], and others.

BACK-TESTING AND PNL PERFORMANCE

In finance, an experimental performance result for a proposed investment method or algorithm is commonly obtained via *back-testing*. The most common figure of merit used in back-testing is the performance of a Profit and Loss (PNL) curve. We define the instantaneous PNL of an investment strategy as

$$\begin{aligned} E_n &= E_{n-1} + r_n E_{n-1} + \sum_{i=1}^N q_{i, n-1} R_{i, n} \\ &\quad - \sum_{i=1}^N q_{i, n-1} r_n - \sum_{i=1}^N |q_{i, n} - q_{i, n-1}| I_{i, n} \varepsilon \end{aligned} \quad (41)$$

where E_n is the equity at discrete time index n with $E_0 = \$100$, r_n is the interest rate at n , $q_{i, n}$ is the dollar amount invested in i^{th} asset at n with $q_{i, 0} = 0 \forall i$, $R_{i, n}$ is the return of i^{th} asset at n , ε is the friction parameter that includes the transaction cost and slippage, and $I_{i, n}$ is the indicator function with the value of 1 if there is a transaction (buying or selling) on the i^{th} asset at time n , and 0 otherwise. Investment in an asset evolves in time as $q_{i, n} = q_{i, n-1} (1 + R_{i, n})$. Performance of the PNL expressed in Eq. 41 is calculated with the average return and the volatility (risk) of the PNL, defined as $\mu_E = E[R_E]$ and $\sigma_E = \operatorname{var}[R_E]$, respectively. A PNL with high average return and low volatility is desired for any investment strategy. Therefore, risk-adjusted return (also known as the Sharpe ratio) is defined as $SR = (\mu_E - r) / \sigma_E$. It is commonly used to quantify the performance of PNLs for various competing investment strategies.

The investment universe for back-testing performance reported in this paper is comprised of stocks listed in NASDAQ 100 index. The time span considered is from April 1, 2010 to May 28, 2010 with the time interval of 5 minutes. The data used is the reported price of the stocks for trades that are done through the NBBO, i.e. National Best Bid and Offer. Correlation matrix is estimated at each sample by looking at the returns of the past three days. A simple trading strategy generating about 50% long orders and 50% short orders in the course of a day is employed. At each entering point 4% of the capital is invested in a particular stock. The interest rate, r , and friction, ε , given in Eq. 41 are considered as 0 and 1.5 bps (0.015%), respectively.

PNL for the test strategy without any risk management method is displayed in Fig. 6.a (black line). Similarly, PNL curves for the risk managed cases are displayed in Fig. 6.a with blue, red, and magenta colored lines for SIE, SOE, and SAE methods, respectively. In all methods, risk threshold is set to 3 bps / sample (~25 bps / day). The day after the flash crash of May 6, 2010 [10] is of special interest to us since the risk managed strategies avoid the 1.8% draw-down while the strategy without any risk management suffered. The estimated risk values are displayed in Fig. 6.b for all the scenarios considered in this study. It is observed from the figures that

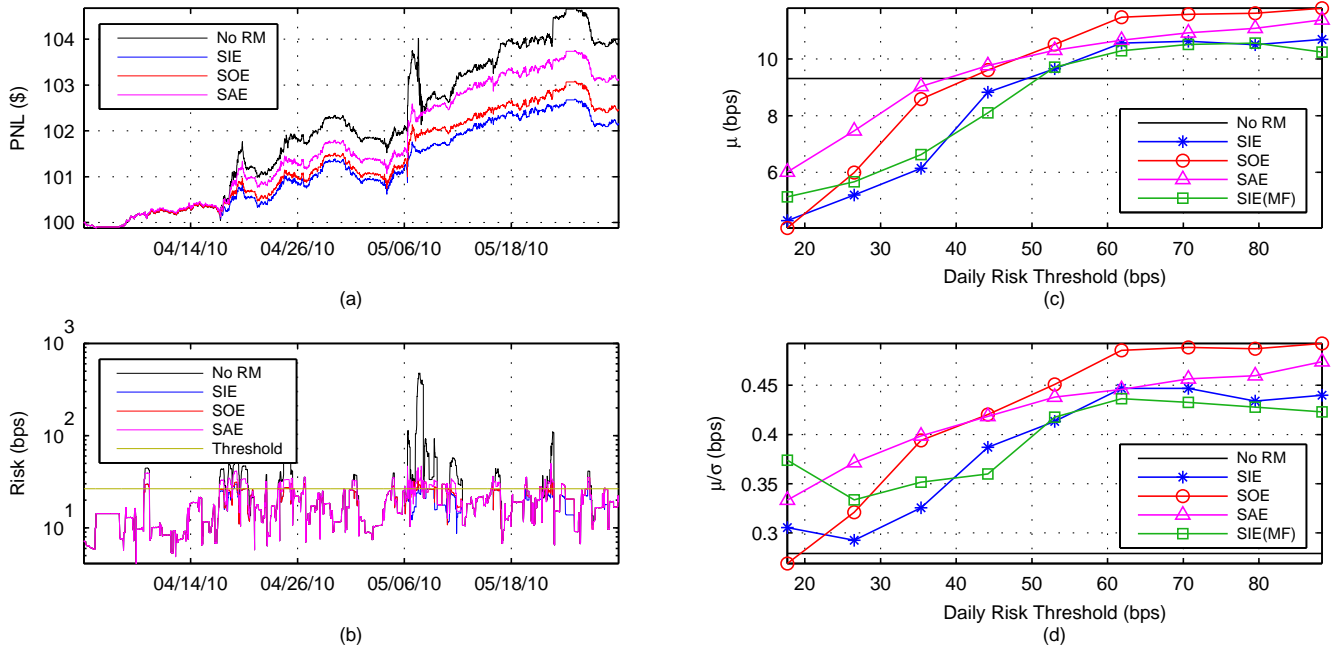


Fig. 6. (a) PNLs for no risk management case along with the proposed Stay in the Ellipsoid (SIE), Stay on the Ellipsoid (SOE), and Stay around the Ellipsoid (SAE) risk management methods. Their average daily returns are of 9.4 bps (0.094%), 5.2 bps, 5.9 bps, and 7.4 bps; daily volatilities are of 33.4 bps, 17.8 bps, 18.7 bps, and 20.1 bps; daily Sharpe ratios are of 0.28, 0.29, 0.32, and 0.37; and average numbers of transactions per day are of 25.3, 16.8, 17.8, and 19.7, respectively, (b) Corresponding estimated daily risk values calculated as σ_p/E (see Eqs. 6 and 41), (c) Average daily return versus daily risk threshold for SIE, SOE, SAE, and multiple frequency SIE methods along with the average daily return of no risk management case, and (d) Corresponding daily Sharpe ratios. Sharpe ratio for the non-managed risk case is the lowest one compared to the managed risk cases.

SAE method outperforms SIE and SOE methods in terms of average return while keeping the volatility at a desired level. All of the methods considered perform well in terms of keeping the portfolio risk bounded with the trade-off of reduced return. However, a less risk-averse investor may easily set the risk threshold to a higher level to increase the level of desired return.

This experiment is repeated by changing the risk threshold from 2 to 10 bps / sample (from ~17 bps / day to ~88 bps / day). Average daily return and daily Sharpe Ratio of the PNLs for non-managed risk case and for all managed cases are displayed in Fig. 6.c and Fig. 6.d, respectively. It is observed from the figures that the SAE and SOE methods yield significantly higher returns with a negligible increase in the volatility than the others for a given risk level. The PNL performance for the multiple frequency SIE method formulated in Eqs. 32 and 33 with the sampling rates of $k_1 = 1$, $k_2 = 3$, and $T_s = 5 \text{ min}$ is displayed in Fig. 6.c and Fig. 6.d. with green colored lines. Note that, in this scenario, all the assets in the portfolio are traded at the same frequency although the framework introduced in this paper allows investors to trade different assets at different frequencies. The trivial multiple

frequency trading results are presented to highlight the flexibility of the proposed framework. It is also evident from this study that the portfolio risk management and re-balancing may be performed at multiple frequencies by utilizing the novel framework forwarded in this paper.

CONCLUSIONS

We presented the basics and significance of portfolio risk in this paper. The inherent measurement noise in the empirical financial correlation matrix, and the eigendecomposition for noise cleaning prior to risk calculations are discussed in detail. Furthermore, we extended the popular risk engineering concepts to rebalancing of hedged portfolios and trading in multiple frequencies. We proposed three novel risk management methods. Then, we highlighted the merit of the proposed techniques by presenting their impact on back-testing performance of an investment strategy employed for the investment universe comprised of stocks listed in the NASDAQ 100 index. It is concluded that the portfolio risk management and re-balancing may be performed at multiple trading

frequencies with more flexibility by utilizing the framework introduced in this paper.

APPENDIX

Steps involved in deriving the eigenfiltered version of Eq. 24 are similar to the ones in Eq. 19. Hence, eigenfiltered version of Eq. 24 is expressed as

$$E[A^2] = \sum_{k=1}^L \lambda_k \left(\sum_{i=1}^N q_i v_i^{(k)} \sigma_i \right)^2 + \sum_{i=1}^N \epsilon_{ii} q_i^2 \sigma_i^2 \quad (42)$$

where λ_k is the k^{th} eigenvalue, L is the number of selected market factors (eigenvalues), $v_i^{(k)}$ is the i^{th} element of the eigenvector corresponding to the k^{th} eigenvalue, and σ_i is the volatility of the i^{th} asset. Note that, for the sake of brevity, we set $L = N$ in the following discussions where L is the number of selected market factors, and N is the number of assets, respectively. However, as stated in the paper, L is usually very small compared to N , i.e. $L \ll N$.

In order to derive the filtered version of Eq. 26 we utilize the *eigenportfolios*. The return of k^{th} eigenportfolio F_k is defined as

$$F_k = \sum_{i=1}^N \frac{v_i^{(k)}}{\sigma_i} R_i \quad (43)$$

Note that correlation between two different eigenportfolio returns is zero, and the variance of a particular eigenportfolio return is equal to its corresponding eigenvalue, i.e.

$$E[F_i F_j] = \begin{cases} \lambda_i & i = j \\ 0 & i \neq j \end{cases} \quad (44)$$

One can regress the return of an index-tracking asset Y_j over L eigenportfolio returns as follows

$$Y_j = \sum_{k=1}^L \beta_{j,k} F_k + \zeta_j \quad (45)$$

where $\beta_{j,k}$ is the regression coefficient for the j^{th} index-tracking asset and k^{th} eigenportfolio, and ζ_j is the residual term. Now, we assume that the residual term, ζ_j , is orthogonal to all eigenportfolios and we substitute Eq. 45 in Eq. 26 such that we have

$$E[I^2] = \sum_{i=1}^M \sum_{j=1}^M g_i g_j \left(\sum_{k=1}^L \sum_{l=1}^L \beta_{i,k} \beta_{j,l} E[F_k F_l] + E[\zeta_i \zeta_j] \right)$$

It follows from Eq. 44 that

$$E[I^2] = \sum_{i=1}^M \sum_{j=1}^M g_i g_j \left(\sum_{k=1}^L \lambda_k \beta_{i,k} \beta_{j,k} + E[\zeta_i \zeta_j] \right) \quad (46)$$

Assuming the cross-correlation between residual terms is zero, i.e. $E[\zeta_i \zeta_j] = 0$ for $i \neq j$, Eq. 46 can be re-written as

$$E[I^2] = \sum_{k=1}^L \lambda_k \left(\sum_{j=1}^M g_j \beta_{j,k} \right)^2 + \sum_{i=1}^M g_i^2 \xi_i^2 \quad (47)$$

where $\xi_i^2 = \text{var}[\zeta_i]$. In a similar fashion, we assume that there is no correlation between asset returns and residual terms, i.e. $E[R_i \zeta_j] = 0$, and we express the cross-correlation between the i^{th} asset and the j^{th} index-tracking asset using Eq. 45 as follows

$$E[R_i Y_j] = \sum_{k=1}^L \beta_{j,k} E[R_i F_k] \quad (48)$$

Now, let us substitute Eq. 43 in Eq. 48, and we have

$$E[R_i Y_j] = \sum_{k=1}^L \sum_{l=1}^N \beta_{j,k} \frac{v_l^{(k)}}{\sigma_l} E[R_i R_l] \quad (49)$$

By substituting Eq. 18 in Eq. 49, due to the orthogonality of the eigenvectors, we have

$$E[R_i Y_j] = \sum_{k=1}^L \lambda_k v_i^{(k)} \beta_{j,k} \sigma_i$$

Hence, Eq. 25 is equal to

$$E[AI] = \sum_{k=1}^L \lambda_k \sum_{i=1}^N \sum_{j=1}^M q_i g_j v_i^{(k)} \beta_{j,k} \sigma_i \quad (50)$$

Finally, the substitution of Eqs. 42, 47, and 50 in Eq. 23, and re-arranging components yields Eq. 27.

AUTHORS

Mustafa U. Torun (mustafa.torun@njit.edu) received his B.S. and M.S. degrees from the Dokuz Eylul University (D.E.U.), Izmir, Turkey, in 2005 and 2007 respectively, both in electrical and electronics engineering. He was a research assistant in the Department of Electrical and Electronics Engineering at D.E.U. from 2005 to 2008. Since 2008, he has been a Ph.D. candidate in the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology, Newark, NJ. His research interests include high performance computing, data-intensive research in signal processing, multi-resolution signal processing, statistical signal processing, pattern classification, neural networks, genetic algorithms; and their applications in quantitative finance, electronic trading, digital communications, digital imaging, and biomedical engineering.

Ali N. Akansu (akansu@njit.edu) received his B.S. degree from the Technical University of Istanbul, Turkey, M.S.

and Ph.D degrees from the Polytechnic University, Brooklyn, New York, all in Electrical Engineering. Since 1987, he has been on the faculty of the Department of Electrical and Computer Engineering at the New Jersey Institute of Technology. He was a Founding Director of the New Jersey Center for Multimedia Research and NSF Industry-University Cooperative Research Center for Digital Video. Dr. Akansu was the Vice President for Research and Development of IDT Corporation. He was the founding President and CEO of PixWave, Inc., and Senior VP for Technology Development of TV.TV (IDT subsidiaries). He has been on the boards of several companies and an investment fund. He visited David Sarnoff Research Center, IBM T.J. Watson Research Center, GEC-Marconi Electronic Systems Corp., and Courant Institute of Mathematical Sciences at the New York University. Dr. Akansu has published numerous articles and several books on his research work over the last two decades. His current research interests include signal and transform theories, and applications in quantitative finance and electronic trading, wireless communications and RF engineering, data intensive discovery with high performance DSP for next generation Internet Engineering. Dr. Akansu is a Fellow of the IEEE.

Marco Avellaneda (avellane@cims.nyu.edu) has held positions at the New York University, Courant Institute of Mathematical Sciences since 1985, including the Director of Division of Financial Mathematics. He has worked within Derivative Products and Foreign Exchange at Morgan Stanley and Banque Indosuez, respectively. He is a founding partner of Finance Concepts. He has extensive advisory and consulting experience in the fields of volatility trading, relative-value trading, pricing and analysis, arbitrage and the OTC market. An established author, in the areas of quantitative modeling of derivative securities and quantitative analysis on financial markets, Dr. Avellaneda has written numerous research papers and he is the Managing Editor for International Journal of Theoretical and Applied Finance. Most recently, Dr. Avellaneda was named 2010 Quant of the Year by the Risk Magazine for his groundbreaking work on the effect of short-selling restrictions on price dynamics. Dr. Avellaneda received his Licenciado en Ciencias (Mathematics), from Buenos Aires University in 1981, and his Ph.D. Degree in Mathematics from University of Minnesota in 1985.

REFERENCES

- [1] H. M. Markowitz, *Portfolio selection: efficient diversification of investments*. Wiley, New York, 1959.
- [2] L. Laloux, P. Cizeau, M. Potters, and J.-P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and Applied Finance*, vol. 3, pp. 391–397, Jan 2000.
- [3] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, "Random matrix approach to cross correlations in financial data," *Phys. Rev. E*, vol. 65, pp. 066 126–1–066 126–18, Jun 2002.
- [4] J. P. Bouchaud and M. Potters, "Financial applications of random matrix theory: a short review," arXiv.org, Quantitative Finance Papers 0910.1205, Oct. 2009. [Online]. Available: <http://ideas.repec.org/p/arx/papers/0910.1205.html>
- [5] A. M. Sengupta and P. P. Mitra, "Distributions of singular values for some random matrices," *Phys. Rev. E*, vol. 60, no. 3, pp. 3389–3392, Sep 1999.
- [6] T. W. Epps, "Comovements in stock prices in the very short run," *Journal of the American Statistical Association*, vol. 74, no. 366, pp. 291–298, 1979.
- [7] F. Black and M. S. Scholes, "The pricing of options and corporate liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–54, Jun 1973.
- [8] D. Mueller-Gritschneider, H. Graeb, and U. Schlichtmann, "A successive approach to compute the bounded Pareto front of practical multiobjective optimization problems," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 915–934, 2009.
- [9] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, 2001.
- [10] A. A. Kirilenko, A. P. S. Kyle, M. Samadi, and T. Tuzun, "The flash crash: The impact of high frequency trading on an electronic market," *SSRN eLibrary*, Oct 2010. [Online]. Available: <http://ssrn.com/paper=1686004>