Comparison of Markowitz and Mean Absolute Deviation (MAD) Models in Portfolio Optimization

Eugene Han E.HAN@YALE.EDU
Yale University
New Haven, CT 06520, USA

Raymond Lee RAYMOND.LEE@YALE.EDU
Yale University
New Haven, CT 06520, USA

Abstract

This paper delves into the fundamental principles of portfolio optimization by exploring the Markowitz model introduced in 1952 and its subsequent extension by Konno and Yamazaki in 1991, known as the Mean Absolute Deviation (MAD) model. The Markowitz model focuses on minimizing variance while attaining a given expected return. Building upon this, Konno and Yamazaki developed the Mean Absolute Deviation (MAD) model using absolute deviation instead of variance-covariance, simplifying risk estimation and sidestepping assumptions about return distributions. We find that under the condition of jointly normal asset returns, Markowitz and Mean Absolute Deviation (MAD) models produce equivalent portfolios but different portfolios under multivariate lognormal returns, especially with a large number of assets. We also find that modern solvers seem to have mitigated the historical concerns about computational load: for both large numbers of assets (N) and observations (T), the Markowitz model tends to perform similarly or even faster in terms of computation time compared to the MAD model. When applied to real market data, portfolios constructed using Markowitz and MAD models perform similarly in most cases. However, in certain scenarios, MAD portfolios may show more robustness to market regime shifts and could outperform Markowitz portfolios. Both models are sensitive to the quantity and time range of data used for portfolio construction.

Keywords: Markowitz, Mean-variance model, mean absolute deviation, linear programming, quadratic programming

1 Background

Portfolio optimization calculates the best collection and amount of assets to own with respect to two key metrics: maximizing the expected return and minimizing the variance of the portfolio. Maximizing returns aligns with an investor's core motivation of offering their own capital to boost productivity and output, ultimately increasing their object of investment's value and thus the value of their own share as well. Meanwhile, minimizing variance is naturally motivated by an investor's aversion towards uncertainty, especially the tail-end probability of losing significant portions of investment. With these two objectives in mind, the problem of portfolio optimization can be formulated in two ways: the portfolio achieving a given level of expected return with minimum variance, or the portfolio achieving a given level of variance with maximum expected return. We recognize that these two approaches are mathematically analogous and proceed to discuss the minimization of variance at a given expected return level.

Markowitz (1952)'s paper represents a breakthrough in portfolio optimization theory, specifically introducing the idea of portfolio diversification. By using a collection of investments with low correlations with each other, it becomes possible to achieve lower risk for a given level of return. This thesis remains crucial to even modern portfolio creation for eliminating systematic risk (that is, risks associated with individual stocks rather than the entire market) as observed in the growing popularity of funds. Exchange-traded funds (ETFs), mutual funds, and hedge funds are all means to pool capital together to invest in multiple assets at a time to take advantage of diversification of risk at different levels. Markowitz (1952)'s impact can also be seen in economic literature: Sharpe (1964), Ross (1976), and Black and Litterman (1991) all build upon Markowitz (1952)'s diversification thesis to alternative methods for determining the fair price of assets.

Konno and Yamazaki (1991) builds upon the Markowitz Model by addressing its computation run time and applicable situations. Indeed, the Markowitz Model may be rendered impractical due to the computation complexity of using variance-covariance for estimating risk, which is a quadratic optimization problem. To address this, Konno and Yamazaki (1991) simplified risk estimation by using Mean Absolute Deviation (MAD) instead of variance-covariance, thus only requiring linear programming. Furthermore, the Markowitz model implicitly assumes normally distributed returns on assets, yet this assumption is not always met especially during volatile time periods. To address this, the MAD model does not assume normality or any distribution shape. With these two optimizations, Konno and Yamazaki (1991) claims to generate comparable results to the Markowitz model in low-dimensional problems with normally distributed returns while also being applicable to problems considering a greater number of stocks with nonnormal returns.

2 Problem Formulation

We now discuss the optimization problem formulations as adapted from Markowitz (1952) and Konno and Yamazaki (1991), using a common set of notations. To fairly compare the two models, we will focus on risk minimization subject to an expected return constraint. For simplicity, we will only consider portfolios consisting of long positions exclusively.

We begin by defining some notation common to the two formulations. Suppose we have N securities observed over T contiguous points in time, with security i having return R_i, which is a random variable. We can make a simple estimate of the expected return of the i-th security $\mathbb{E}[R_i]$ as

$$\mu_i = \frac{1}{T}\sum_{t=1}^{T} r_{it}$$

where r_{it} is a random variable equal to the rate of return of security i over some period at time point t. Let $\mathbf{w} \in \mathbb{R}^N$ such that $\mathbf{1}^T\mathbf{w} = 1$ and $\mathbf{w} \succeq 0$, where w_i is the proportion of money allocated to security i. Finally, let R be the random variable characterizing the return of our portfolio consisting of N securities with weights $\mathbf{w}$, $R = w_1R_1 + w_2R_2 + \cdots + w_NR_N$. Then, by linearity of expectation, the (estimated) expected return of the portfolio is

$$\mathbb{E}[R] = \mathbf{w}^T\boldsymbol{\mu}$$

This will serve as a proxy for our expected return constraint so that for some minimum target return $r_{\min}$ we require $\mathbf{w}^T\boldsymbol{\mu} \geq r_{\min}$.

2.1 Markowitz

To quantify the risk associated with this portfolio, Markowitz proposed using the variance of the return, $\mathrm{Var}(R)$. If we let $\mathbf{\Sigma}$ denote the covariance matrix of the returns $R_1, \ldots, R_N$ with entries $\sigma_{ij} = \mathrm{Cov}(R_i, R_j)$, which we again estimate from the data over T time points, then

$$\begin{aligned}
\mathrm{Var}(R) &= \mathrm{Var}(w_1 R_1 + \cdots + w_N R_N) \\
&= \sum_{i=1}^{N} w_i^2 \mathrm{Var}(R_i) + 2 \sum_{1 \le i < j \le N} w_i w_j \sigma_{ij} \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij} \\
&= \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}
\end{aligned}$$

Minimizing this measure of risk is then equivalent to finding $\mathbf{w}$ satisfying the following optimization problem:

$$\begin{aligned}
\min \quad & \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \\
\text{s.t.} \quad & \mathbf{w}^T \boldsymbol{\mu} \ge r_{\min} \\
& \mathbf{1}^T \mathbf{w} = 1 \\
& \mathbf{w} \succeq 0
\end{aligned}$$

This defines a quadratic program.

Although Markowitz never mentions any distribution assumptions for the returns, the original paper implicitly assumes that returns follow a multivariate normal distribution, that is $(R_1, R_2, \ldots, R_n) \sim \mathrm{MVN}(\boldsymbol{\mu}, \mathbf{\Sigma})$. This is because the model relies on expected values and covariance to measure risk and return, which fail to capture important information if the assumption doesn't hold. For instance, covariance exhaustively characterizes co-dependence between securities only when returns are jointly normal.

2.2 Mean Absolute Deviation

In addition to the implicit normality assumption, one criticism of the Markowitz model was its reliance on computing a potentially large dense covariance matrix, which requires $\binom{N}{2}$ calculations, and dealing with this $N \times N$ matrix in a quadratic program. This was an especially relevant issue at the time since algorithms and computers were not as sophisticated as they are today.

To address these concerns, Konno and Yamazaki proposed using the mean absolute deviation of the return as a measure of risk, defined and estimated as

$$\begin{aligned}
\mathrm{MAD}(R) &= \mathbb{E}\left[\left|\sum_{i=1}^{N} w_i R_i - \mathbb{E}\left(\sum_{i=1}^{N} w_i R_i\right)\right|\right] \\
&= \frac{1}{T} \sum_{t=1}^{T} \left| \sum_{i=1}^{N} w_i (r_{it} - \mu_i) \right|
\end{aligned}$$

This yields the following optimization problem:

$$\begin{aligned}
\min \quad & \frac{1}{T}\sum_{t=1}^{T}\left|\sum_{i=1}^{N} w_i(r_{it}-\mu_i)\right| \\
\text{s.t.} \quad & \mathbf{w}^T\boldsymbol{\mu} \geq r_{\min} \\
& \mathbf{1}^T\mathbf{w} = 1 \\
& \mathbf{w} \succeq 0
\end{aligned}$$

Reformulating the above problem by introducing a new set of variables $y_1, y_2, \ldots, y_T$ yields

$$\begin{aligned}
\min \quad & \frac{1}{T}\sum_{t=1}^{T} y_t \\
\text{s.t.} \quad & y_t + \sum_{i=1}^{N} w_i(r_{it}-\mu_i) \geq 0, \quad t = 1, 2, \ldots, T \\
& y_t - \sum_{i=1}^{N} w_i(r_{it}-\mu_i) \geq 0, \quad t = 1, 2, \ldots, T \\
& \mathbf{w}^T\boldsymbol{\mu} \geq r_{\min} \\
& \mathbf{1}^T\mathbf{w} = 1 \\
& \mathbf{w} \succeq 0
\end{aligned}$$

This defines a linear program and does not involve $\mathbf{\Sigma}$.

3 Summary of Theoretical Claims

3.1 Markowitz (1952)

For the convenience of geometric illustration, Markowitz considers portfolio optimization over three securities. Our problem objectives and constraints become:

$$\min \quad \mathbf{w}^T\mathbf{\Sigma}\mathbf{w} = \sum_{i=1}^{3}\sum_{j=1}^{3} w_i w_j \sigma_{ij}$$

$$\begin{aligned}
\text{s.t.} \quad & (1) \sum_{i=1}^{3} w_i\mu_i \geq r_{\min} \\
& (2) \sum_{i=1}^{3} w_i = 1 \\
& (3) \mathbf{w} \succeq 0
\end{aligned}$$

Indeed, (3) allows us to assert that $w_3 = w_1 + w_2$, thus giving us two degrees of freedom and allowing us to utilize 2D plots.

Central to Markowitz's claims are that iso-mean curves, the set of portfolios with the same expected returns, lie on a line. When we replace our r_{min} inequality constraint in (1) with an equality at a given return level r_0, we indeed see a linear relationship between μ_1 and μ_2:

$$\begin{aligned} r_0 &= \sum_{i=1}^{3} w_i \mu_i \\ &= w_1\mu_1 + w_2\mu_2 + (1 - w_1 - w_2)\mu_3 \implies \\ w_2 &= \frac{r_0 - \mu_3}{\mu_2 - \mu_3} - \frac{\mu_1 - \mu - 3}{\mu_2 - \mu - 3} w_1 \end{aligned}$$

The ellipsoid shape of iso-variance can be proved using spectral decomposition of $\mathbf{\Sigma}$. Specifically because the covariance matrix $\mathbf{\Sigma}$ is real and symmetric, we can perform eigen-decomposition to rewrite as $\mathbf{\Sigma} = \mathbf{X}\mathbf{Q}\mathbf{X}^{\mathbf{T}}$, where the *ith* column of $\mathbf{X}$ is the *ith* eigenvectors of $\mathbf{\Sigma}$, and $\mathbf{Q}$ is a diagonal matrix with each diagonal element $\mathbf{Q}_{\mathbf{ii}} = \lambda_i$, which is the *ith* eigenvalue corresponding to the eigenvector in the *ith* column of x. Let $\mathbf{y} := \mathbf{X}^{\mathbf{T}}\mathbf{w}$, then for a fixed portfolio risk level c, we yield:

$$c = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} = \mathbf{w}^{\mathbf{T}} \mathbf{X}\mathbf{Q}\mathbf{X}^{\mathbf{T}} \mathbf{w} = \mathbf{y}^{\mathbf{T}} \mathbf{Q} \mathbf{y} = \sum_{i}^{3} y_i^2 \lambda_i \implies \sum_{i}^{3} \frac{y_i^2}{\lambda_i c} = 1$$

where we are able to proceed forward with dividing by our eigenvalues because our covariance matrix should not include linear dependencies due to the fact we are using real data (i.e. it is extremely rare for perfect linear dependencies to exist between stock prices). Indeed, we recognize that the form of the equation matches the form of an equation of an ellipsoid (or hyperellipsoid in higher dimensions).

With this equation, we see that as our target risk level c decreases, our ellipsoid shrinks (because the denominator decreases) while the center stays the same (because our numerator stays the same). We denote our point of minimum risk c_0, which also represents the center of concentric iso-variance ellipses. Below is an illustration of iso-mean and iso-variance curves:

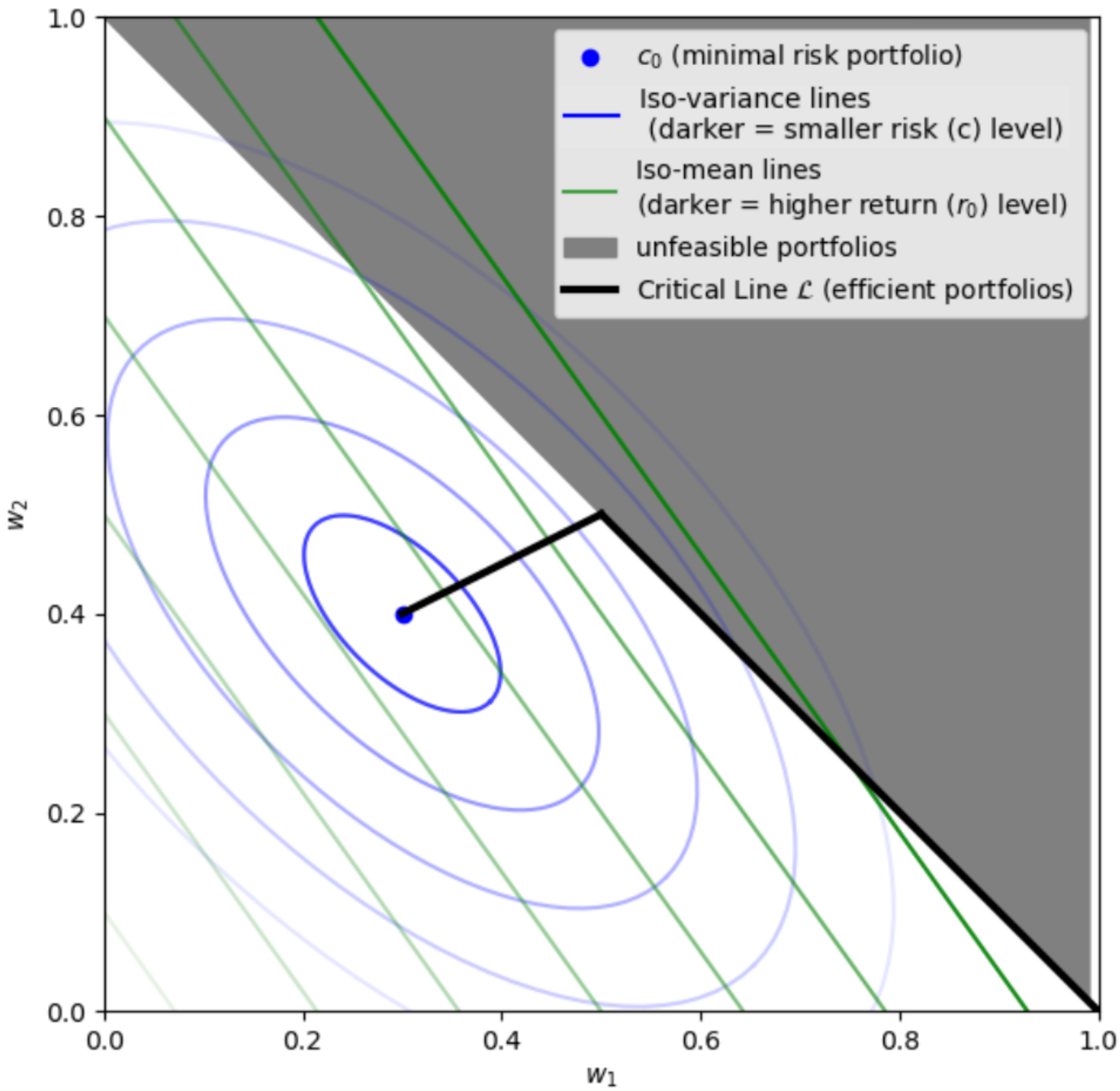


Figure 1: Iso-mean and Iso-variance visualization for three security space

Now let's consider our efficient frontier, that is the minimum variance at each return level. Using a geometric argument, we are looking for the point on each iso-mean curve that intersects the smallest iso-variance curve. Due to the convex nature of our iso-variance ellipsoids, this is indeed just the point on each iso-mean curve that lies tangent to some iso-variance curve, whose corresponding variance is exactly the smallest variance achievable at that level of return.

Now, consider this optimal point for each level of expected return. These optimal points form the critical line $\mathcal{L}$, which is the set of admissible portfolios (i.e. the lowest risk for each return level). Note that our minimal risk portfolio c_0 is a point on $\mathcal{L}$ because it is the global risk minimizer and thus necessarily minimizes risk for its return level. When $\mathcal{L}$ reaches the edge of feasible sets, it follows along the feasible set boundary in the direction of increasing returns. In the 3-security case, this means down if the iso-mean curve slope is greater than 1 (as in the illustration), and up otherwise.

Though Markowitz (1952) does not outline a specific methodology for solving in higher dimension cases, his later paper Markowitz (1956) details a Critical Line Algorithm (CLA) for efficient portfolios. While modern quadratic solvers involve faster algorithms, the CLA remains important in the context of the paper to understand motivations for further optimizations made later by Konno and Yamazaki (1991). In essence, the CLA uses the Lagrangian to integrate our constraints into the objective function, thus forming an unconstrained optimization problem. CLA begins setting all assets' weights to their lower bound, and iteratively increasing the weights of each asset to their upper bound until weights sum above one; the asset we are adjusting when this condition is violated becomes a "free asset" while all other asset weights are "bounded" (because they are either at their upper or lower bound), thus creating a one-variable problem which we can use the Lagrangian to find its variance minimizer. The solution to this subproblem is denoted a "turning point," and after iterating this method across all assets, we create a set of turning points and we know that our efficient portfolios must be a convex combination of these turning points.

3.2 Konno and Yamazaki (1991)

One of the key points made by Konno and Yamazaki is the equivalence of using their proposed mean absolute deviation and Markowitz's variance as measures of risk under multivariate normality of asset returns. Suppose the returns $(R_1, \ldots, R_N) \sim \mathrm{MVN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which can be written in the form $\boldsymbol{\mu} + \mathbf{CZ}$ where $\boldsymbol{\Sigma} = \mathbf{CC}^T$ by Cholesky decomposition and the Z_i's are i.i.d. standard normal random variables. Then,

$$\sum_{i=1}^{N} w_i R_i = \mathbf{w}^T \boldsymbol{\mu} + \mathbf{w}^T \mathbf{CZ}$$

This is just a linear combination of i.i.d. standard normals, which is a normal random variable with mean $\mathbf{w}^T \boldsymbol{\mu} = \sum_{i=1}^{N} w_i \mu_i$ and variance $\mathbf{w}^T C(\mathbf{w}^T C)^T = \mathbf{w}^T \boldsymbol{\Sigma} \mathbf{w}$. Recall the mean absolute deviation risk measure was defined as

$$\mathrm{MAD}(R) = \mathbb{E}\left[\left|\sum_{i=1}^{N} w_i R_i - \mathbb{E}\left(\sum_{i=1}^{N} w_i R_i\right)\right|\right] = \mathbb{E}\left[\left|\sum_{i=1}^{N} w_i R_i - \sum_{i=1}^{N} w_i \mu_i\right|\right]$$

If we let $X = \sum_{i=1}^N w_i R_i - \sum_{i=1}^N w_i \mu_i$, then $X \sim N(0, \mathbf{w}^T \mathbf{\Sigma} \mathbf{w})$ and

$$\begin{aligned}
\mathrm{MAD}(R) &= \mathbb{E}(|X|) \\
&= \frac{1}{\sqrt{2\pi \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}} \int_{-\infty}^{\infty} |x| \exp\left\{-\frac{x^2}{2 \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}\right\} dx \\
&= \frac{2}{\sqrt{2\pi \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}} \int_{0}^{\infty} x \exp\left\{-\frac{x^2}{2 \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}\right\} dx \\
&= \frac{2}{\sqrt{2\pi \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}}} \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w} \int_{0}^{\infty} e^{-u} du && \text{let } u = \frac{x^2}{2 \cdot \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}} \\
&= \sqrt{\frac{2}{\pi}} \sqrt{\mathbf{w}^T \mathbf{\Sigma} \mathbf{w}} && \text{integrates to 1} \\
&= \sqrt{\frac{2}{\pi}} \sqrt{\mathrm{Var}(R)}
\end{aligned}$$

Since $\sqrt{\frac{2}{\pi}}$ is a positive constant, $\mathrm{Var}(R) = \mathbf{w}^T \mathbf{\Sigma} \mathbf{w}$ is a convex function, and $g(t) = \sqrt{t}$ is strictly increasing, minimizing $\mathrm{MAD}(R)$ is equivalent to minimizing $\mathrm{Var}(R)$ when returns follow a multivariate normal distribution.

The paper also highlights some theoretical advantages of the MAD formulation that translate to practical gains. For instance, let's consider when the number of observations, T, is fixed. If the number of securities, N, is varied, then the number of constraints remains the same and so the complexity of the problem grows linearly as $O(N)$. In contrast, the Markowitz model grows as $O(N^3)$ since it involves solving a system of linear equations. Moreover, we note that if a linear programming problem is feasible, then the optimal value must occur at one of the vertices of the feasible set. Vertices themselves are extreme points, so it follows that the number of nonzero values in the optimal weights will be at most the number of constraints, which is determined by T. The Markowitz model, on the other hand, could output a large number of small nonzero weights. At the time, Konno and Yamazaki argued that this "sparsity" offered by the MAD model was advantageous to the investor because it meant fewer transaction fees due to many small stock purchases, fewer issues dealing with minimum transaction sizes, and a smaller number of securities to manage in one's portfolio. With fractional shares and commission-free trading now being offered by the majority of brokerages, the concerns over transaction fees and rounding up or down are largely irrelevant. Still, it holds merit with respect to preventing over-diversification.

4 Empirical Results

For the purposes of this project, we have formulated the Markowitz and MAD models using CVXPY, which employs more general LP and QP solvers, so that we can compare the two frameworks in a practical setting.

4.1 Numerical Issues

In practice, numerical issues regarding precision can cause problems when computing the covariance matrix $\mathbf{\Sigma}$ for the Markowitz model, something that was not pointed out in the Konno and Yamazaki paper.

To be a valid quadratic programming problem, we require the objective function in the Markowitz model, $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$, to be convex which necessitates $\boldsymbol{\Sigma} \succeq 0$. By definition of the covariance matrix, this is always true *in theory*. However, based on our experimentation, we found that when some of the true eigenvalues of $\boldsymbol{\Sigma}$ are close to 0 or, in the case when $T < N$, exactly equal to 0, the calculated matrix would have small negative eigenvalues typically on the order of 10^{-11} or smaller. This caused our formulation of the Markowitz model in CVXPY to break because the covariance matrix wouldn't be positive semidefinite.

One way to mitigate this is to perturb the diagonals of the calculated covariance matrix. If $\widetilde{\boldsymbol{\Sigma}}$ is the calculated matrix, then we take $\boldsymbol{\Sigma} = \widetilde{\boldsymbol{\Sigma}} + \varepsilon I_N$ for some small $\varepsilon > 0$ which will guarantee that $\boldsymbol{\Sigma}$ is PSD if ε is chosen carefully. A more sound approach is to reformulate the problem by examining the objective function, $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} = \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij}$. We estimate σ_{ij} using the sample covariance, so the objective function becomes

$$\begin{aligned}
\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} &= \sum_{i=1}^N \sum_{j=1}^N w_i w_j \sigma_{ij} \\
&= \sum_{i=1}^N \sum_{j=1}^N w_i w_j \left[\frac{1}{T-1}\sum_{t=1}^T (r_{it}-\mu_i)(r_{jt}-\mu_j)\right] \\
&= \frac{1}{T-1}\sum_{t=1}^T \left[\left(\sum_{i=1}^N w_i(r_{it}-\mu_i)\right)\left(\sum_{j=1}^N w_j(r_{jt}-\mu_j)\right)\right] \\
&= \frac{1}{T-1}\sum_{t=1}^T \left(\sum_{i=1}^N w_i(r_{it}-\mu_i)\right)^2
\end{aligned}$$

Defining $m_t = \sum_{i=1}^N w_i(r_{it}-\mu_i)$ lets us then formulate the problem as

$$\begin{aligned}
\min \quad & \frac{1}{T-1}\sum_{t=1}^T m_t^2 \\
\text{s.t.} \quad & m_t = \sum_{i=1}^N w_i(r_{it}-\mu_i), \quad t = 1, 2, \ldots, T \\
& \mathbf{w}^T\boldsymbol{\mu} \geq r_{\min} \\
& \mathbf{1}^T\mathbf{w} = 1 \\
& \mathbf{w} \succeq 0
\end{aligned}$$

which doesn't rely on $\boldsymbol{\Sigma}$ and undermines some of Konno's and Yamazaki's criticisms of the Markowitz model regarding the covariance matrix.

4.2 Efficient Frontiers

In portfolio optimization, the efficient frontier is defined as the set of portfolios for a given universe of assets such that for any feasible level of risk, there exists no other portfolio that achieves an expected return greater than that of the "efficient" portfolio. In other words, it is those portfolios that achieve optimal returns given risk and vice versa.

As an example, we visualize this concept below by randomly sampling $T = 1000$ vectors of simulated returns from a MVN$(\mathbf{0}, \boldsymbol{I})$ distribution with $N = 5$ assets, generate random portfolio weights and compute the corresponding standard deviation, and obtain the optimal standard deviations using the Markowitz model over a range of minimum expected return values.

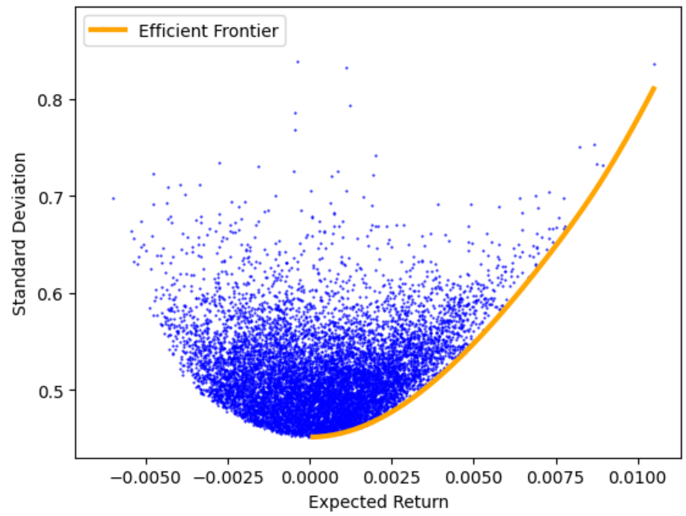
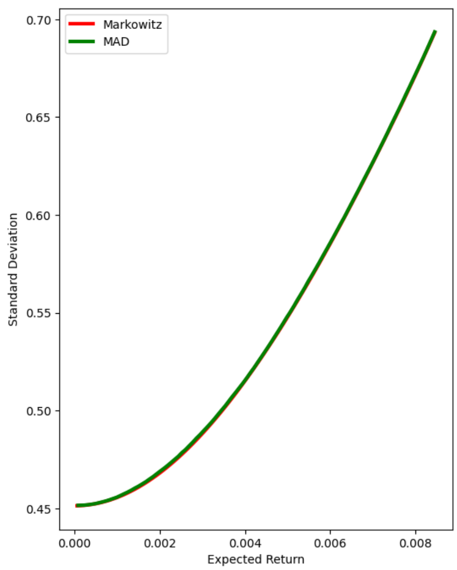
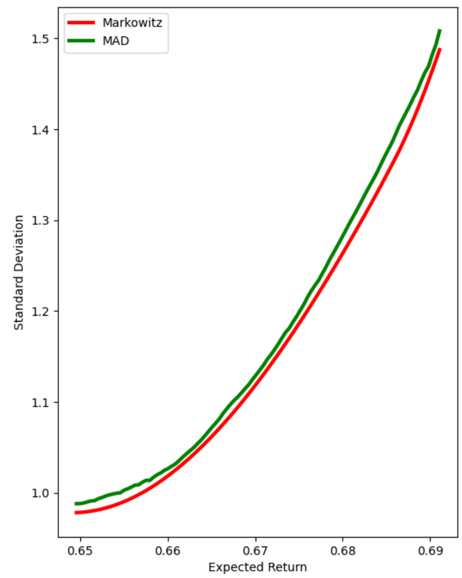


Figure 2: Markowitz's efficient frontier for MVN simulated returns

The same can be done by replacing the measure of risk and model with that of the Mean Absolute Deviation framework. Earlier, we showed that the Markowitz and MAD models lead to equivalent portfolios under the condition that asset returns are jointly normal. To see this empirically, we again sample $T = 1000$ vectors of simulated returns from a MVN$(\mathbf{0}, \boldsymbol{I})$ distribution with $N = 5$ assets. For a range of minimum expected returns, we obtain the portfolio weights using both the Markowitz and MAD models, compute the corresponding standard deviations (i.e. we use Markowitz's risk measure as a basis for comparison), and then superimpose their efficient frontier graphs. Furthermore, to see their differences, we repeat the exact same process drawing from a shifted multivariate lognormal distribution derived from a MVN$(\mathbf{0}, \boldsymbol{I})$ distribution.



(a) Multivariate normal returns



(b) Multivariate lognormal returns

Figure 3: Comparison of Markowitz and MAD portfolio standard deviations

As expected, for multivariate normal returns the two plots coincide perfectly up to some level of precision, whereas they deviate by a small but noticeable amount when returns follow a multivariate lognormal distribution. This difference appears to be exacerbated for larger N, as seen below for $N = 50$.

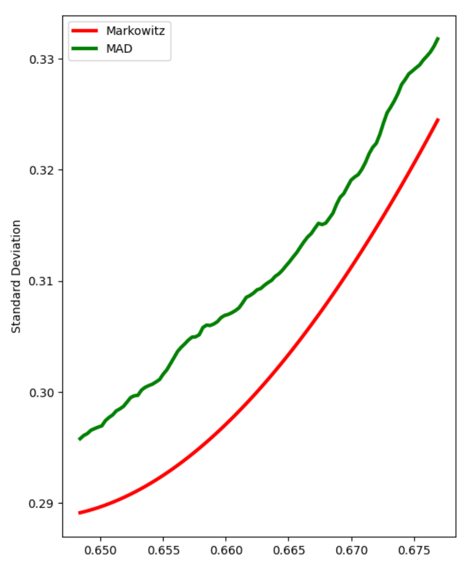


Figure 4: Multivariate lognormal returns, $N = 50$

4.3 Computation Time

Konno and Yamazaki argued that one of the biggest shortcomings of the Markowitz model was its heavy computational load at the time due to calculating a covariance matrix and solving a quadratic program. However, it isn't clear if this continues to be an issue today since modern solvers are much faster and more accessible. To investigate this, we explored the effects that the number of assets being considered, N, and the number of observations, T, have on compute times when using the Markowitz versus MAD models.

For varying N, we first generated fake returns data from a multivariate normal distribution with a random covariance matrix with $T = 150$ and 501 assets. For each N in $\{25, 50, \dots, 475, 500\}$, we randomly sampled N of the assets from the generated data and computed the time taken for the optimal portfolio weights to be determined for some fixed minimum expected return; this was repeated 10 times for each value of N so that we could use the mean and standard deviation of the times.

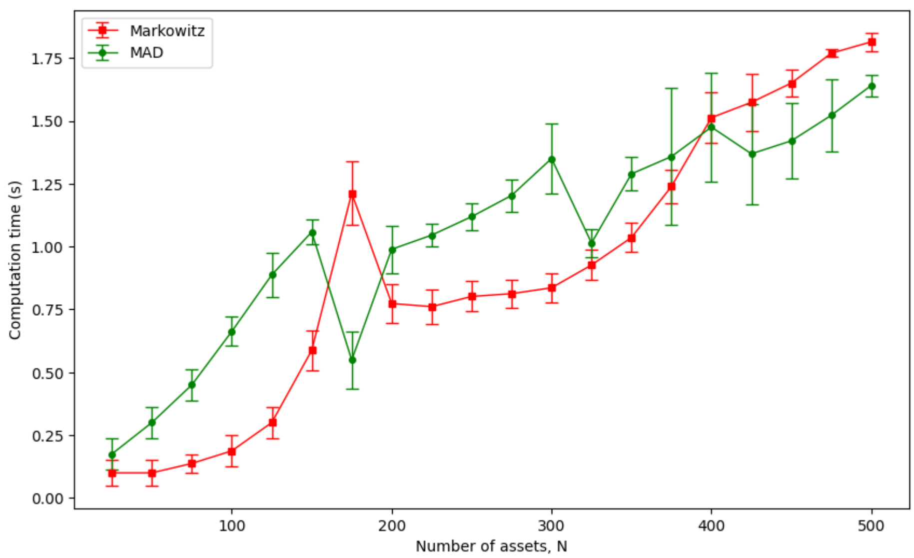


Figure 5: Computation time vs. number of assets

It isn't immediately clear from the figure if either model is significantly faster than the other for large N, with the computation times growing similarly for both. In fact, the Markowitz model tends to take less time more often than not. As a result, it appears that advances in technology, both hardware and software, have largely overcome the original concerns.

We repeated the same process for the number of observations except by fixing $N = 50$ and letting T vary in the set $\{50, 100, \dots, 950, 1000\}$.

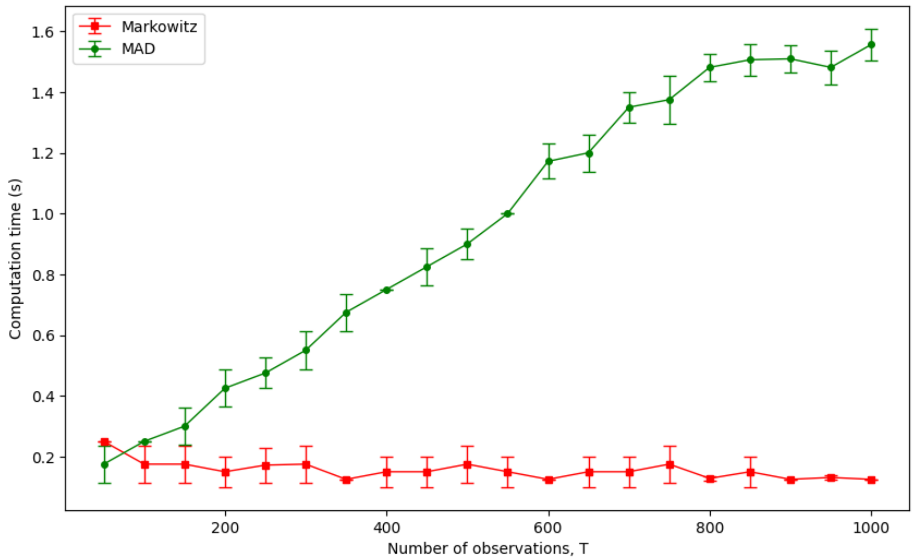


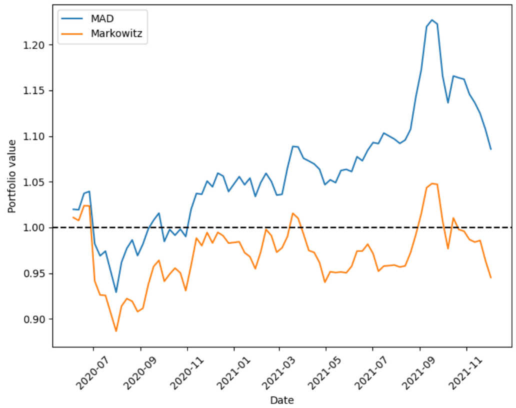
Figure 6: Computation time vs. number of observations

As expected, the computation time required for the Markowitz model remains roughly constant since the complexity of the problem only depends on the number of assets, N. On the other hand, computation time appears to grow roughly linearly for the MAD model; this growth makes sense since the number of constraints in the formulation increases with T. As a result, the MAD model fails to provide significant advantages in computation time when compared to the Markowitz model, given modern methods and computers, for both large N and T.

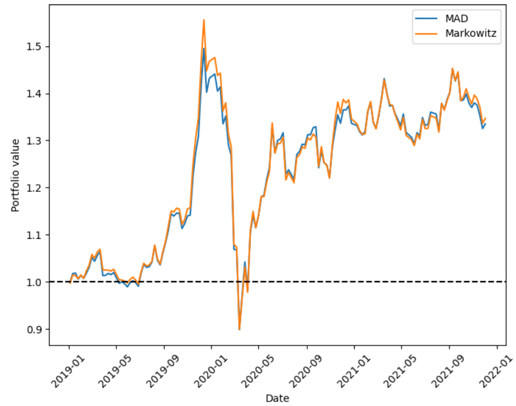
4.4 Performance on Real Market Data

Lastly, we explore how portfolios constructed using the Markowitz and MAD models would perform on actual stock markets, extending the original papers' work to high dimensions. In particular, we used weekly closing prices (from which weekly returns were later calculated) on Fridays for 1767 stocks from the Tokyo Stock Exchange between 01/06/2017 and 12/03/2021, and for 3151 stocks from AMEX, NYSE, and NASDAQ exchanges between 01/02/2015 and 02/14/2020.

For each of the Japanese and American stock collections, we tried two different cutoff dates for estimating portfolio weights versus backtesting, each with minimum weekly return $r_{\min} = 0.005$ or 0.5%. In general, the portfolios given by the Markowitz and Mean Absolute Deviation models behaved and performed similarly, with the exception of Figure 6(a) where the MAD portfolio performed considerably better than the Markowitz portfolio, confirming that both of these frameworks are very sensitive to the quantity and time range of data from which the portfolio weights are calculated. From these tests, it appears that MAD portfolios generally perform similarly to or better than Markowitz portfolios and are more robust to market regime shifts.

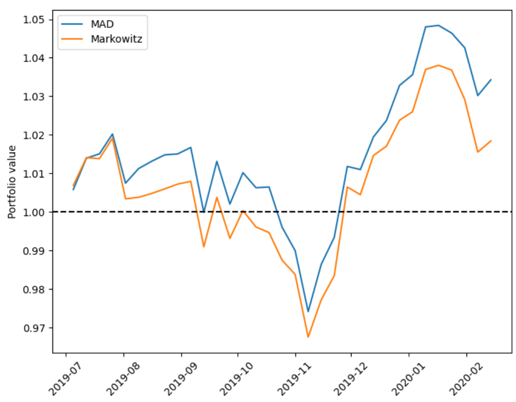


(a) Cutoff date of 06/01/2020

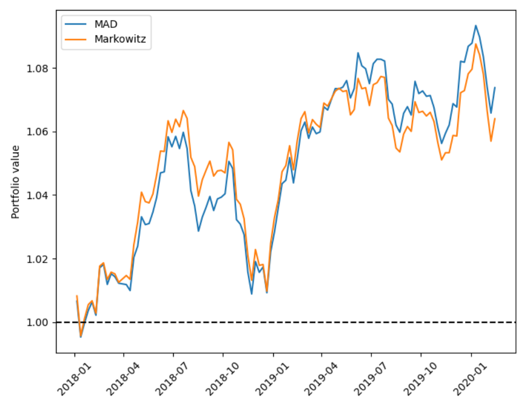


(b) Cutoff date of 01/01/2019

Figure 7: Markowitz vs. MAD portfolio returns for Tokyo Stock Exchange stocks



(a) Cutoff date of 07/01/2019



(b) Cutoff date of 01/01/2018

Figure 8: Markowitz vs. MAD portfolio returns for AMEX/NYSE/NASDAQ stocks

5 Extensions

Most notable of our discovery was that the computational improvements made by Konno and Yamazaki (1991) and the MAD model over the original Markowitz model are rendered negligible in modern solvers. As this conflicts with the theoretical run-time hypothesized in Konno and Yamazaki (1991), further investigations into why computational performance gains in quadratic programming have seemingly outpaced linear programming gains would be necessary to understand this contradiction. Furthermore, while Markowitz (1952) and Konno and Yamazaki (1991) utilized the L2 and L1 norm respectively to estimate security risk, it may be interesting to consider higher-degree norms as well to measure a portfolio's risk, and then also evaluate their computational efficiency using other nonlinear programming optimization techniques.

While Konno and Yamazaki (1991) circumvented distribution assumptions from Markowitz (1952) by using mean-absolute deviation instead of variance for risk estimation, this risk model nevertheless implicitly assumes a symmetry in an investor's response to upside and downside. Rom and Ferguson (1994) overcame the symmetric variance assumption in their Post-Modern Portfolio Theory (PMPT) framework by incorporating higher moments of asset return distribution, such as skew and kurtosis, to also quantify the difference between upside and downside risks. Though different authors have different implementations of higher moments in their models, the need for nonlinear programming techniques is still necessary for practical applications of these methods. As such, it would be interesting to also compare the runtime efficiencies of different PMPT adaptations while varying the number of securities and time periods.

As an alternative to specifying variance data in MPT and PMPT, Black and Litterman (1991) developed a model utilizing historical returns data and risk-free rate to arrive at implied volatility according to the Capital Asset Pricing Model from Sharpe (1964). In addition, Black and Litterman (1991) also allowed the user to input their own expected returns assumptions that differ from market expectations, to arrive at portfolios incorporating qualitative priors. From an optimization perspective, this model continues to use covariances matrices for modeling portfolio risk and thus would involve the same quadratic solvers utilized in the methods of this report.

With respect to the impractical assumptions of the Markowitz model, Merton (1969) extended the Markowitz model to both incorporate the ability to short sell, as well as the ability to borrow and lend at a specified risk-free rate. In addition, Miguel S. Lobo (2006) implemented both linear and fixed transaction costs within the Markowitz model while still maintaining the core nature of the problem as convex optimization. Finally, Merton (1973) extended Markowitz's mean-variance model to a continuous-time framework to consider the change in asset expected returns and risk across time. Indeed, these all demonstrate efforts to make the Markowitz model more realistic and ultimately practical for investors.

Appendix A.

All code used in the creation of this paper is original and open-source. The repository can be found here: `https://github.com/ehan03/MAD-vs-MW-Portfolio-Optimization`. The stock price data was downloaded from Kaggle and modified.

References

Fischer Black and Robert Litterman. Asset Allocation: Combining Investor Views with Market Equilibrium. *The Journal of Fixed Income*, 1(2):7–18, 1991. doi: https://doi.org/10.3905/jfi.1991.408013.

Hiroshi Konno and Hiroaki Yamazaki. Mean-Absolute Deviation Portfolio Optimization Model and Its Applications to Tokyo Stock Market. *Management Science*, 37(5):519–531, 1991.

Harry Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952.

Robert C. Merton. Lifetime Portfolio Selection under Uncertainty: the Contunous-Time Case. *The Review of Economics and Statistics*, 15(3):247–257, 1969. doi: https://doi.org/10.2307/1926560.

Robert C. Merton. An Intertemporal Capital Asset Pricing Model. *Econometrica*, 41(5):867–887, 1973. doi: https://doi.org/10.2307/1913811.

Stephen Boyd Miguel S. Lobo, Maryam Fazel. Portfolio optimization with linear and fixed transaction costs. *Annals of Operations Research*, 152(1):341–365, 2006. doi: https://doi.org/10.1007/s10479-006-0145-1.

Brian M. Rom and Kathleen W. Ferguson. Post-Modern Portfolio Theory Comes of Age. *Journal of Investing*, 3(3):11–17, 1994. doi: https://doi.org/10.3905/joi.3.3.11.

Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic THeory*, 13(3):341–360, 1976. doi: https://doi.org/10.1016%2F0022-0531%2876%2990046-6.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964. doi: https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.