Tail Risk Constraint and Portfolio Optimization Methods with an Introduction to Maximum Entropy

- Tanishq Chauhan (tanishq6991@gmail.com)
- Junaid Ahmad (junayd03@gmail.com)

Abstract

We will be modeling the VaR and CVaR and examine the impact of different portfolio techniques/ methodologies on VaR and CVaR at different confidence levels (95% and 99%) of a Markowitzian portfolio for a varying range of optimization methodologies. We will also be presenting a reasonable discussion on the methods of curtailing the tail losses of a given portfolio distribution using various methods including the maximum entropy distribution.

Keywords: *Portfolio Optimization, VaR, CVaR, Maximum Entropy*

Introduction and Literature Review

Starting with a Markowitzian model a multisector asset portfolio will be considered within Markowitz's mean-variance framework of the modern portfolio theory which is widely used as the basis of constructing a portfolio. VaR and CVaR will be taken as standalone measures of portfolio selection, as well as constraints for optimization problems discussed in the following section. We have discussed theoretical insights of about seven optimization methods, of which we will be demonstrating computationally four methods.

The most important aspect of our work is to understand the effect of using entropy and its comparison to standard deviation on the overall assessment of risk. Our main effort is to draw attention to the idea of risk as an exposure to the overall portfolio based on the loss appetite of the investor, rather than wrongly defining risk as some variable (such as standard deviation, or confusing it with volatility). Diversification itself is not a great idea, as suggested by Peter Lynch it leads to what he calls “diworsification”. To summarize, diversification has many times led to the overall Compound Annual Growth Rate (CAGR) being decreased while not reducing the risk exposure of the portfolio. Also, our argument is that if the investor has to focus on the portfolio exposure then forecasting and prediction of the prices of assets included within the portfolio should be of only tertiary concern. To further our argument, we would like to borrow some ideas from the insurance industry, especially the concept of survival analysis. These we shall be discussing in the final sections of the document after going through our results.

Brief about VaR and CVaR

VaR in simple terms is a maximum potential loss pertinent to the portfolio under a given confidence level. Technically it is the measure that tells how much one can lose under a given confidence level (95% and 99%) over a given time horizon. This can further be explained with an example of a portfolio of USD 1 million having a 10-day VaR at 95% equal to USD 10000, which means the portfolio cannot lose more than USD 10000 over a period of 10 days with a confidence level of 95%. VaR assumes asset returns follow normal distribution.

There are mainly three types of VaR analysis viz; parametric VaR, Historical simulation based VaR and Monte Carlo simulation based VaR. The main constituents of VaR analysis are portfolio mean, standard deviation and alpha (confidence level). As VaR answers the maximum loss we can incur under a certain confidence level then there arises another question like "what will be the loss if actual loss exceeds the number obtained from VaR analysis". Then CVaR (conditional VaR) comes into picture answering such questions by taking the average of losses exceeding the VaR number in analysis. CVaR is also called expected shortfall and is used as an extension to the VaR.

Motivation

To understand the motivation behind the entire exercise we begin by presenting a brief refresher of the mean-variance setting. Please refer to equations 1 - 6 in the Appendix section on *Motivation* for a mathematical treatment.

Using this setting as the base of our discourse, the density function $f(x)$ and its behaviour subject to VaR and CVaR constraints are the topic of interest which is discussed at length either completely or partially in the following literature we have analyzed. Due to the lack of available guidelines to thoroughly analyze literature at hand (due to it being a combination of mathematics, statistics, and finance), we have adapted the method as previously discussed in our literature review.

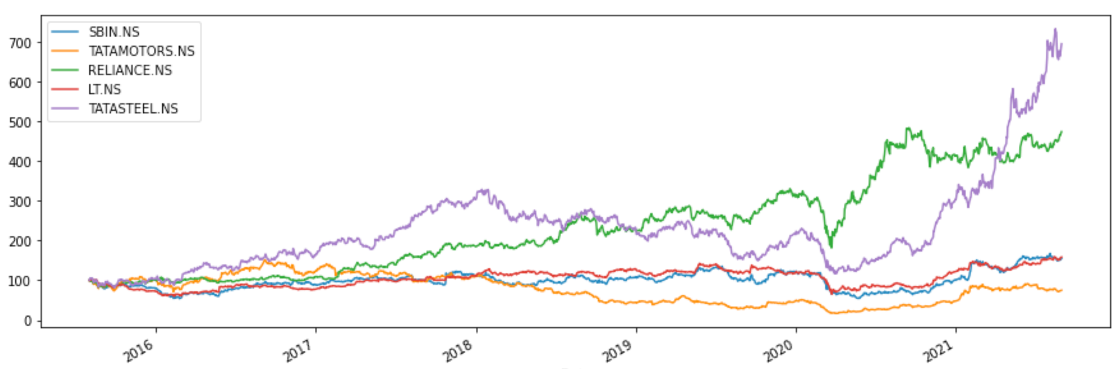
Once the data is selected it'll be processed and exploratory analysis will be conducted on local volatility to compute the local volatility of the data index over time. After characterizing the return distribution, we would see if it fits the normal distribution and the student's-t distribution. Since VaR and CVaR are considered as an alternate selection criterion for portfolio the statistical skewness and fitness of the portfolio data will also be worked upon.

Among the three popular types of VaR analysis our focus in this study will mainly be on historical simulation based VaR, the reason being different weights obtained from different optimization techniques can be fitted efficiently in historical simulation based VaR analysis for the optimization methodologies.

After getting the weights from different optimization techniques we will use those weights in out VaR and CVaR analysis to see the overall impact of portfolio optimization on overall reduction in VaR and CVaR. The process follows a stepwise procedure where we will first calculate VaR and CVaR based on the equal weightage (amount will be invested equally in all stock that is 20% each). Once we are ready with equally weighted portfolio VAR and CVAR then we will use different weights derived from different optimization techniques to see which optimization techniques increased and decreases the overall VaR and CVaR amount.

We will be selecting the following stocks given their stability and long-withstanding strong financial records in the indian market. The following are top conglomeratres with extensive global presence.

Stock	Symbol
State Bank of India	SBIN
Tata Motors	TATAMOTORS
Reliance Industries Limited	RELIANCE
Larsen and Toubro	LT
Tata Steel	TATASTEEL



Efficient frontier

Efficient frontier also known as portfolio frontier was developed by the American Economist Harry Markowitz in the year 1952. This model is based on a premise of giving the highest return for a minimal level of risk. The main components of efficient frontier are expected return, standard deviation (risk component) and covariance of underlying assets. Efficient frontier emphasis on the portfolio diversification and the merits associated with it.

Efficient frontier is based on few assumptions like:

- Investors are rational and do posses all market knowledge pertinent to the asset in a portfolio.
- All investors are risk averse and their goal is to avoid risk
- No market player is so big to impact market holistically.
- Investors can borrow to a greater extent
- Lending and borrows has risk free rate
- Markets are efficient
- Normal distribution is followed by asset returns
- Expected return and standard deviation are the main premises on which investors make decisions

For a mathematical treatment, please refer to the equations 7 – 14s in the Appendix section on *Efficient Frontier*.

The referred equations are conditional on the fact that the relationship between portfolio weights is linear in nature and hence can be ascertain by use of correlation, also the portfolio variance and portfolio weight must show quadratic relationship (so that the set of portfolios form a parabolic shape, hence being quadratic mathematically) and demonstrates concavity.

Let us use Markowitz's quadratic programming for identifying the efficient frontier (Appendix section *Efficient Frontier* equations 15 - 16), which is a simple optimization problem. We shall be discussing different variations of this basic relationship.

Sharpe ratio optimization output analysis:

TICKER	WEIGHT
RELIANCE.NS	0.56966
SBIN.NS	0
LT.NS	0
TATAMOTORS.NS	0
TATASTEEL.NS	0.43034

Expected annual return	33.7%
Annual volatility	28.1%
Sharpe Ratio	1.13

A Sharpe Ratio of 1.13 is high due to the high volatility of 28.1%. However, contemporarilty a high Sharpe ratio is good, the higher it is the better it is. Thus, a Sharpe ratio optimized portfolio allows only to hold two assets of the listed five, these include Reliance Industries Limited and Tata Steel. Both firms are global conglomerate giants with stable financials and consistent growth with their investment diversified throughout sectors.

Based on the suggestion from Sharpe ratio optimization, we consider a portfolio of value INR.1,00,00,000. Using a discrete allocation method, we get the following allocation of stocks given our portfolio value.

RELIANCE.NS	TATASTEEL.NS
2509	2987

The following is the result from our VaR, CVaR computations.

Details	Equally Weighted	Efficient Frontier	ERC	HRP	Max Entropy
Amount in Mn	Amount in Mn	Amount in Mn	Amount in Mn	Amount in Mn	
Portfolio Value	10.000	10.000	10.000	10.000	10.000
Daily VaR @ 95%	0.253	0.250	0.232	0.224	0.291
Monthly VaR@ 95%	1.188	1.172	1.086	1.052	1.364
Daily VaR@ 99%	0.477	0.433	0.429	0.414	0.505
MonthlyVaR @ 99%	2.236	2.032	2.014	1.943	2.366
Daily CVAR @ 95	0.389	0.382	0.369	0.348	0.436
Daily CVAR@ 99%	0.674	0.646	0.648	0.623	0.721

Max Values			Min Values		
Daily VaR @ 95%	0.291	Max Entropy	Daily VaR @ 95%	0.224	HRP
Monthly VaR@ 95%	1.364	Max Entropy	Monthly VaR@ 95%	1.052	HRP
Daily VaR@ 99%	0.505	Max Entropy	Daily VaR@ 99%	0.414	HRP
MonthlyVaR @ 99%	2.366	Max Entropy	MonthlyVaR @ 99%	1.943	HRP
Daily CVAR @ 95	0.436	Max Entropy	Daily CVAR @ 95	0.348	HRP
Daily CVAR@ 99%	0.721	Max Entropy	Daily CVAR@ 99%	0.623	HRP

Interpreting the impact of different optimization methodologies onVaR and CVaR

As it can be seen the VaR amount for an equally weighted portfolio of USD 10 million as different confidence levels reflects that the portfolio cannot lose more than USD 0.25 million at a confidence level 95% on any given day and monthly VaR cannot exceed $1.18 million at same confidence level.

Similarly, for 99% confidence level VaR amount cannot exceed 0.47 million and monthly VaR cannot be more than $2.23 million for an equally weighted portfolio of USD 10 million at the same confidence level

After calculating the VaR there arises a question what if actual losses excel the predefined VaR amount. For such queries expected shortfall comes into picture. The numbers and analysis can be interpreted as "the conditional VaR for an equally weighted portfolio of USD 10 million cannot exceed 0.38 million at any given day at a confidence level of 95% and $0.34 million at 99% confidence level.

Likewise VaR and CVaR has also been presented for the same portfolio of USD 10 million where the weights for computation has been derived from different optimization techniques. It can be seen the VaR and CVaR values are higher for the portfolio where maximum entropy optimization technique has been used. This is mainly due to the imbibition of addition disturbance in terms of variance and standard deviation which eventually resulted in higher values for VaR and CVaR the reason being standard deviation is main component of VaR analysis. It can also be seen VaR and CVaR values are minimum for the portfolio where weights have been derived from HRP the reason being HRP uses Hierarchical cluster formation technique for optimising the portfolio. Where returns of assets are clustered together on the basis of correlation (distance) between then and then ranked accordingly.

Minimum Variance (Haugen, Baker 1991)

Maximizing returns for minimum while concentrating exclusively on minimizing volatility.

This is financial application of a basic mathematical optimization problem where the aim is to minimize the variance of the equation representing the portfolio. Please refer Appendix section on *Minimum Variance* (equations 17 – 26), our further discussion will be based on the results of the computed weights discussed in these equations.

Now let us evaluate heuristically eq.(26). Let us say that the market usually has a high volatility which is denoted by the idiosyncratic variable $\sigma^2_{\varepsilon,i}$. Also, let us assume the investor would like to aim for higher returns (which are complemented by higher market volatility) which points to high market beta. As we can see, this invariably push the eq.(26) to zero. Thus, we can straightaway conclude that many of the securities in our portfolio will either have smaller weights or none at all.

The minimum variance method lets us disqualify putting efforts into securities which in an overall sense are not investable for our portfolio. This helps us concentrate on exploring other financial metrics for the selected securities.

Maximum Diversification (Choueifaty, Coignard 2008)

To summarize the maximum diversification method introducted by Choueifaty et. al. (2008), investments in risk-efficient markets will produce returns in proportion to their total risk.

To measure how diverified a portfolio is, we take a simple ratio with the weighted average assets volatility in the numerator to portfolio volatility in denominator. This ratio is maximized as a part of the basic optimization problem. The maximum diversification method takes (all) asset's marginal contribution (to the portfolio) and equalizes it for a change in asset's weight.

We defined the diversification ratio (being maximized) as follows;

$$\text{Max:}D(\mathbf{w}) = \frac{\mathbf{w}'\boldsymbol{\sigma}}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}, \qquad (27)$$

where the asset's volatility is denoted by a $N \times 1$ vector $\boldsymbol{\sigma}$, and the diversification ratio is denoted by the variable $D(\mathbf{w})$. Please refer to Appendix section equations 28 – 30 for mathematical treatment.

We can thus see that maximizing eq.(29) will yield us the Sharpe ratio of our portfolio.

An advantage of using maximum diversification method is that the weights tend to be less concentrated in our portfolio (compared to minimum variance), however the method suffers from the same drawbacks as the minimum variance.

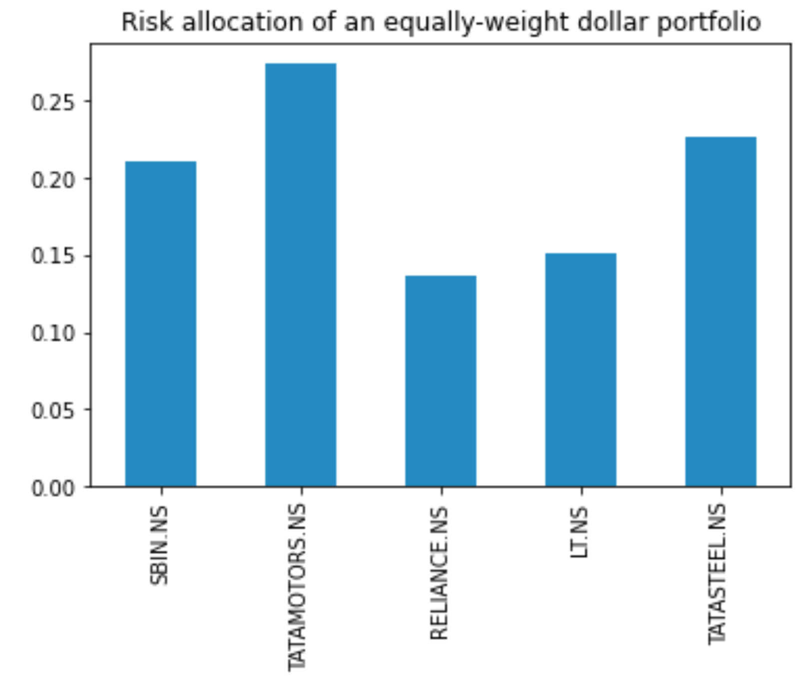
Maximum Decorrelation (Christoffersen et al. 2010)

Primary principle is all assets have similar returns and risks but heterogeneous correlations.

Maximum decorrelation is fairly simple, the portfolio is optimal if all the assets somehow have equal expected return and equal volatilities, bar heterogenous correlations. Please refer to Appendix section equations 31 – 36 for mathematical derivation.

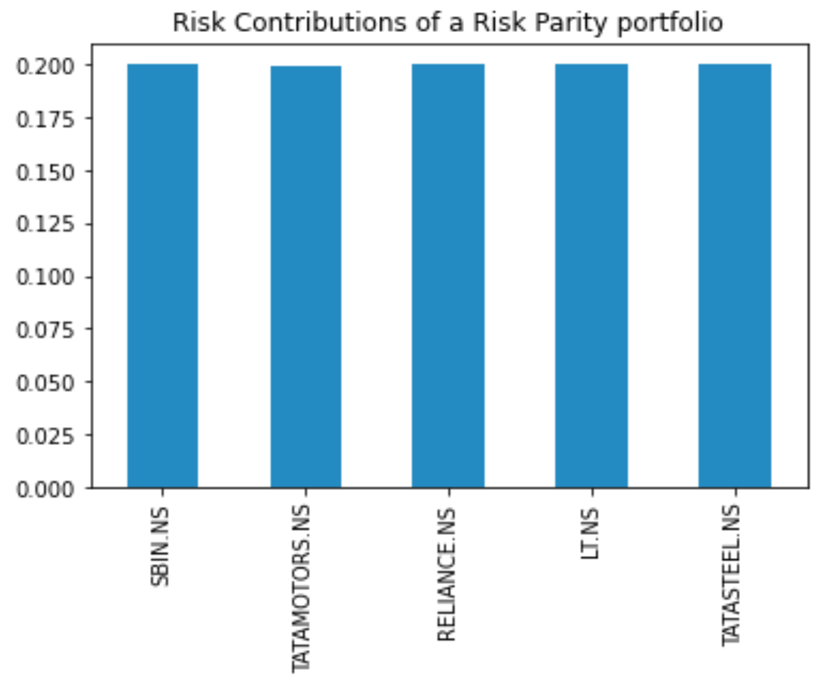
Equal Risk Contribution (a.k.a Risk Parity Contribution) (Mailard, Roncalili, Teiletche 2008)

As the name suggests the method is about equalizing the risk from each asset in the portfolio. This method of equal risk contribution is a compromise between other methods wherein the goal of optimization is equal volatility for all assets.



We can see that by allocating equal weights (input all weights) the optimization problem then solves for the risk and the result is the above figure.

Risk contribution of a risk parity (equal risk contribution) portfolio



By setting equal risk for all assets (input risk), the optimization problem then solves for the weight output as follows.

Optimal weight output as per risk parity method

TICKER	WEIGHT (in %)
RELIANCE.NS	18.217026
SBIN.NS	14.781808
LT.NS	25.599156
TATAMOTORS.NS	24.162237
TATASTEEL.NS	17.239773

Larsen and Toubro (LT) along with Tata Motors has around 25% allocation, while the others having allocation in the range of 14% - 18%.

Hierarchical Risk Parity (Lopez de Prado 2016)

We perform the clustering with the help of a distance matrix. Following this the below two steps are performed:

1. Let the cluster be represented by individual objects. We then take two closest clusters.

2. The clusters from step 1 are then merged.

Above step 1 and step 2 are excuted in a loop until all the clusters are merged together in the aforementioned manner.

The length of the straight line is used to compute the distance between 2 clusters. This distance is alternatively called the Euclidean distance. The overall procedure invovled is called the measure of distance, alternatively its is also called similarity.

To determine the initial point from where we shall compute the distance (only after the distance metric is selected) we will use a simple heuristic. Suppose we have 2 closest clusters as described above and they are similar in nature, we can then compute the distance between them. We can use a number of criterias which can range from similarity, dissimilar cluster, etc. This is referred to as the Linkage Criteria.

Definition: Linkage Matrix

Let there be a $(N-1) \times 4$ dimensional matrix such that its structure is of the form

$$Y = \{(y_{m,1}, y_{m,2}, y_{m,3}, y_{m,4})\}_{m=1,\dots,N-1} \qquad (37)$$

meaning a set of quadriplets per cluster, wherein the consituents of the set are represented by $(y_{m,1}, y_{m,2})$. The other consituent $(y_{m,3})$ represents the distance between the former two constituents. Whereas $(y_{m,4})$ represents the total original items encompassed in cluster m.

Quasi-Diagonalization

To ensure that the largest of the value are aligned with the diagonal, the covariance matrix is rearranged. The covariance matrix quasi-diagonalization procedure ensures that based on the above described similarity cluster principle, that the most similar investments are clubbed while the least similar ones are not.

Algorithm for quasi-diagonalization procedure -

1. We know that each row of the linkage matrix megers 2 branches into 1. We replace clusters in $(y_{N-1,1}, y_{N-1,2})$ with their constituents recursively, until no clusters remain.
2. These replacemnets preserve the order of the clustering.
3. The output is a sorted list fo original items.

Algorithm for quasi-diagonalization procedure -

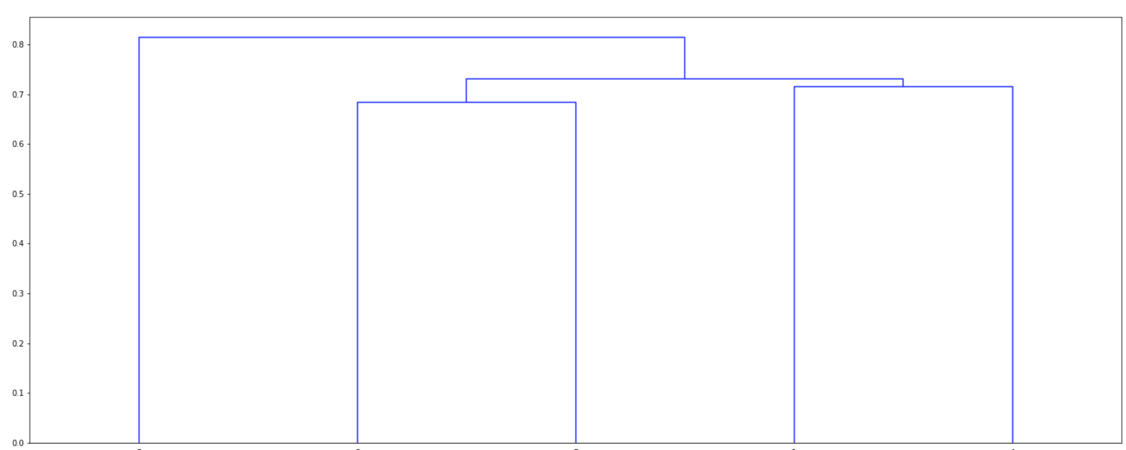
1. As we are clubbing together two clusters based on the linkage criteria, each cluster must merge into a single branche. This procedure keeps repeating itself and soon all clusters will cease to exist. To mathematically generalize the constituents replace clusters in $(y_{N-1,1}, y_{N-1,2})$ reiterating the procedure.
2. The clusters' arrangement order is preseved by these constituents.
3. Original items are then listed out as outputs.

Recursive Bisection

Now that we have a quasi-diagonal matrix, for optimality we prefer inverse variance. We have a look at the following two procedures;

1. bottom up, the variace of a continuous subset is first defined as the variance of inverse variance allocation.
2. top down, The allocations are split adjacently in proportions inverse to their variances (aggregatedly).

following is a resultant dendogram



we thus have the following weights via the HRP

TICKER	WEIGHT
RELIANCE.NS	0.317518
SBIN.NS	0.202151
LT.NS	0.302281
TATAMOTORS.NS	0.072657
TATASTEEL.NS	0.105393

Reliance and Larsen & Toubro (L&T) have almost similar allocation, whereas SBI has 20%, with Tata Motors having the lowest.

Maximum Entropy (Geman et al. 2015)

Some introduction to shannon entropy / how is entropy different from standard deviation. Effect of using entropy instead of std / variance on portfolio.

Maximum Entropy Portfolio result interpretation

For maximum entropy we have the following;

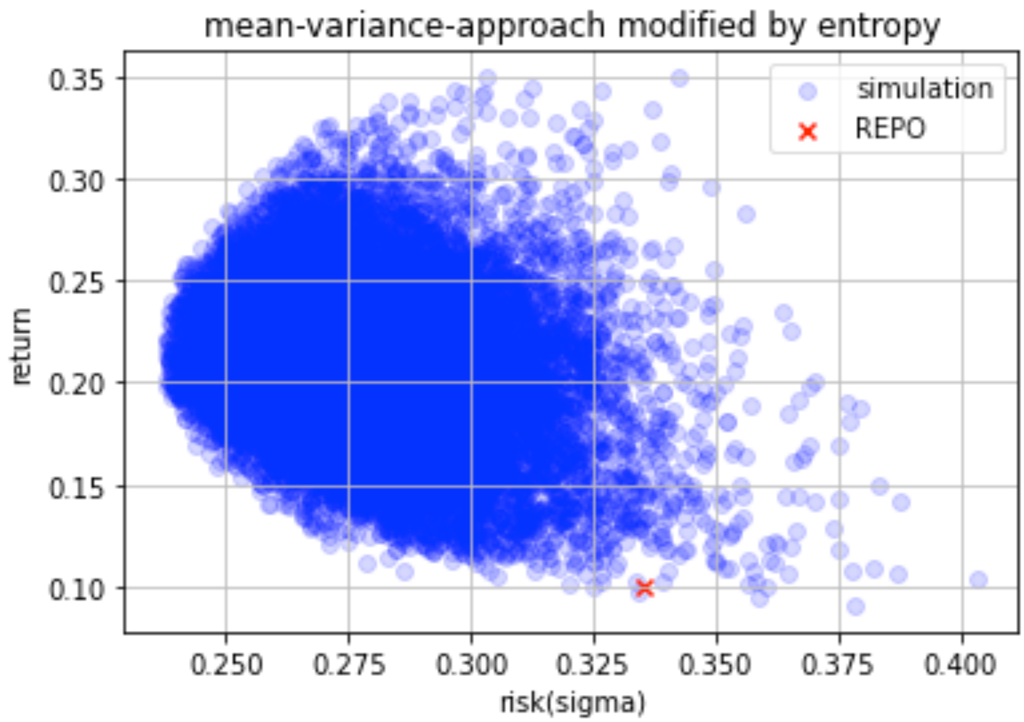
Weight No.	Weight
RELIANCE.NS	0.18365
SBIN.NS	0.5293
LT.NS	0.0215
TATAMOTORS.NS	0.2595
TATASTEEL.NS	0.0058

We can see that Reliance Industries Limited has around 18.3% weight, while State Bank of India (SBI) has the most weight of 52.93%, half of it is Tata Motors with 25.95%, whereas Tata Steel barely has any allocation with 0.58%. On a note, SBI is one of the largest public sector government owned bank in the world.

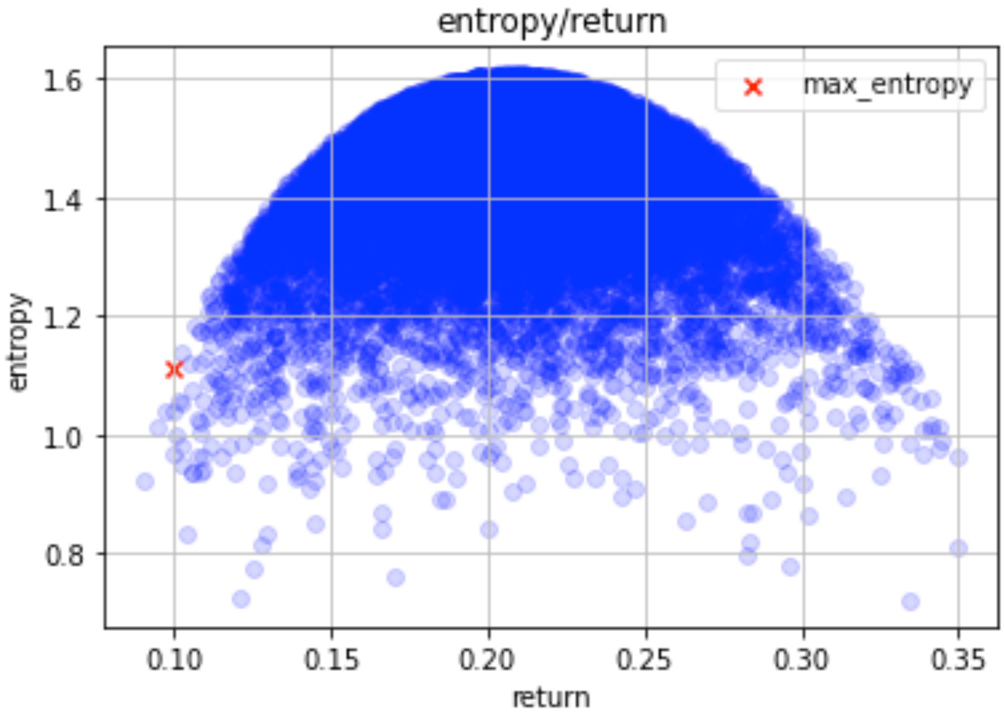
Portfolio output:

Portfolio risk	0.3351
Portfolio return	0.0999
Portfolio entropy	1.110

For the given stocks, we have a 33.51% standard deviation for a portfolio return of 9.99%.



As we can see, the entropy modified mean-variance approach has given us plenty of choices where the returns are higher and the standard deviation is low.



We can see that a higher entropy does not imply higher return, which is reasonable. Unlike the high-risk high return theory which is extremely dangerous.

Discussion on Entropy vs Standard Deviation

Of the methods we have discussed, in three of them we have used standard deviation as a measure of risk assessment, while in the maximum entropy we have used entropy as a measure of risk assessment. Let us first understand the fundamental difference between standard deviation and entropy.

Let us do a simple comparison of the formulas for standard deviation and entropy. The standard deviation is

$$\sigma(\mathrm{X}) = \sqrt[2]{\mathbb{E}[(\mathrm{X}\text{-}\mathbb{E}(\mathrm{X}))^2]}$$

, we can clearly see that the standard deviation is dependent on the values of the random variable X. The variance clearly gives an idea about the spread around the mean, which we call as the deviation, but that is it.

The entropy of any given random variable X is given as

$$H(X) = -\sum_x p(x) \log p(x) = -\mathbb{E}[\log p(X)]$$

, again, it is clear that entropy does not concern itself with the values the random variable X takes, the only concern is that of the distribution. Also, entropy has the property of maximizing whenever (in the process) the outcomes occur with equal probability. This means entropy induces more uncertainty while minimizing in case of single outcome processes, meaning lesser or no uncertainty.

This is a significant difference as the investor in case of using entropy as a measure can focus less on prediction and forecasting of the values the variable X could take, and put a greater focus on the exposure of the portfolio. In realistic terms, we mostly would encounter not unimodal but either bimodal or generally speaking multimodal distributions because our portfolio of assets will not comprise merely of publicly traded equities, but also other assets such as bonds, and options. The standard deviation in multimodal distribution will concentrate around the peaks of the distribution, thus its effectiveness will diminish with each additional node, and will be unable to take into consideration moments of the process in the tails.

Example: Let random variable X take values 1,…,N such that the distribution p(x) is non-uniform. Say we randomly change the order of the variables (moving larger values around to center, or left tail, or the right tail for instance) the standard deviation will change accordingly. The entropy however remains unaffected.

Entropy has the advantage of supporting more varieties of data because of its use of probability distribution than the data itself, thus incorporating more statistical information than standard deviation. Most importantly, entropy is not restricted to a specific type of distribution, making the models more realistic. For more in depth study of information theory, please refer to Cover et. al. (2006) [40].

Conclusion

We have used several methods to analyze the output return and volatility as per the weights suggested by the various techniques. On an initial look some techniques such as HRP outplays Maximum Entropy (and the rest of the other methods). However, one should always keep in mind that we are merely back testing our methods on "historical" data, which is merely one of the many ways in which an event could have unfolded. Hence, it is always recommended to take historical data as merely a base point and then do several Monte-Carlo simulations to check for the best and worst-case scenarios. A major advantage of maximum entropy technique is that it introduces randomness in the signals "historical data in our case" such that the entropic random variable (causing the most disturbance) is maximized as per optimization method, and hence we have the advantage of the basic portfolio optimization as well as the randomness of Monte-Carlo simulation.

s

References

1. Roger Clarke, Harindra De Silva, and Steven Thorley. Risk parity, maximum diversification,and minimum variance: An analytic perspective. Journal of Portfolio Management, 39(3):39,2013.

2. Roger Clarke, Harindra De Silva, and Steven Thorley. Minimum-variance portfolio composition.Journal of Portfolio Management, 37(2):31, 2011.

3. Yves Choueifaty and Yves Coignard. Toward maximum diversification. Journal of Portfolio Management, 35(1):40–51, 2008.

4. Dowd, K. (1998). Beyond Value at Risk: The New Science of Risk Management.

5. Rockafellar, R. (2007). Coherent Approaches to Risk in Optimization Under Uncertainty.

6. Rockafellar, R., & Uryasev, S. (2001). Conditional Value-at-Risk for General Loss Distributions. Corporate Finance and Organizations eJournal.

7. Rockafellar, R., Uryasev, S., & Zabarankin, M. (2005). Equilibrium With Investors Using a Diversity of Deviation Measures. Capital Markets: Asset Pricing & Valuation.

8. Rockafellar, R., Uryasev, S.P., & Zabarankin, M. (2004). Generalized Deviations in Risk Analysis. Risk Management.

9. Rockafellar, R., Uryasev, S., & Zabarankin, M. (2004). Master Funds in Portfolio Analysis with General Deviation Measures. Mutual Funds.

10. Grechuk, B., Molyboha, A., & Zabarankin, M. (2009). Maximum Entropy Principle with General Deviation Measures. Math. Oper. Res., 34, 445-467.

11. Liu, C., Chang, C., & Chang, Z. (2020). Maximum Varma Entropy Distribution with Conditional Value at Risk Constraints. Entropy, 22.

12. Rockafellar, R., Uryasev, S., & Zabarankin, M. (2006). Optimality conditions in portfolio analysis with general deviation measures. Mathematical Programming, 108, 515-540.

13. Rockafellar, R., & Uryasev, S. (2000). Optimization of conditional value-at risk. Journal of Risk, 3, 21-41.

14. Kalinchenko, K. (2012). Optimization with generalized deviation measures in risk management.

15. Stuart, A., & Markowitz, H. (1959). Portfolio Selection: Efficient Diversification of Investments. A Quarterly Journal of Operations Research, 10, 253.

16. Rockafellar, R. (1970). Convex Analysis. Princeton Landmarks in Mathematics and Physics.

17. Uryasev, S., Theiler, U., & Serraino, G. (2010). Risk-return optimization with different risk- aggregation strategies. The Journal of Risk Finance, 11, 129-146.

18. Zabarankin, M., & Uryasev, S. (2014). Statistical Decision Problems.

19. Geman, D., Geman, H., & Taleb, N. (2015). Tail Risk Constraints and Maximum Entropy. Entropy, 17, 3724-3737.

20. Basak, S., & Shapiro, A. (1999). Value-at-Risk Based Risk Management: Optimal Policies and Asset Prices. ERN: Pricing (Topic).

21. Sarykalin, S., Serraino, S., & Uryasev, S. (2008). Value-at-Risk vs. Conditional Value-at- Risk in Risk Management and Optimization.

22. Milhomem, D.A., & Dantas, M.J. (2020). Analysis of new approaches used in portfolio optimization: a systematic literature review. Production Journal, 30.

23. JP Morgan and Reuters: Risk Metrics TM—Technical Document - 1996

24. Chopra, Vijay K., and William T. Ziemba. 1993. "The Effect of Errors in Means, Variances, and Covariances on Optimal Portfolio Choice." Journal of Portfolio Management 19 (2): 6–11.

25. Choueifaty, Yves, and Yves Coignard. 2008. "Toward Maximum Diversification." Journal of Portfolio Management 35 (1). http://www.tobam.fr/inc/uploads/2014/12/TOBAM-JoPM- Maximum-Div-2008.pdf: 40–51.

26. Choueifaty, Yves, Tristan Froidure, and Julien Reynier. 2012. "Properties of the Most Diversified Portfolio." Journal of Investment Strategies 2 (2). http://www.qminitiative.org/UserFiles/files/FroidureSSRN-id1895459.pdf: 49–70.

27. Christoffersen, P., V. Errunza, K. Jacobs, and X. Jin. 2010. "Is the Potential for International Diversification Disappearing?" Working Paper. https://ssrn.com/abstract=1573345.

28. DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2007. "Optimal Versus Naive Diversification: How Inefficient is the 1/N Portfolio Strategy?" http://faculty.london.edu/avmiguel/DeMiguel-Garlappi-Uppal-RFS.pdf: Oxford University Press.

29. Fama, Eugene, and Kenneth French. 1988. "Permanent and Temporary Components of Stock Prices." Journal of Political Economy 96. https://teach.business.uq.edu.au/courses/FINM6905/files/module-2/readings/Fama: 246–73.

30. Haugen, R., and N. Baker. 1991. "The Efficient Market Inefficiency of Capitalization- Weighted Stock Portfolios." Journal of Portfolio Management 17. http://dx.doi.org/10.3905/jpm.1991.409335: 35–40.

31. Lopez de Prado, Marcos. 2016. "Building Diversified Portfolios that Outperform Out of Sample." Journal of Portfolio Management 42 (4): 59–69.

32. Maillard, Sebastien, Thierry Roncalli, and Jerome Teiletche. 2008. "On the properties of equally-weighted risk contributions portfolios." http://www.thierry-roncalli.com/download/erc.pdf.

33. Poterba, James M., and Lawrence H. Summers. 1988. "Mean Reversion in Stock Prices: Evidence and Implications." Journal of Financial Economics 22 (1). http://www.nber.org/papers/w2343: 27–59.

34. Spinu, Florin. 2013. "An Algorithm for Computing Risk Parity Weights." SSRN. https://ssrn.com/abstract=2297383.

35. Yu, J., Lee, W., & Chiou, W.P. (2014). Diversified portfolios with different entropy measures. Appl. Math. Comput., 241, 47-63.

36. Shannon, Claude E.. "A mathematical theory of communication." Bell Syst. Tech. J. 27 (1948): 379-423.

37. J. Ke, C. Zhang, Study on the optimization of portfolio based on entropy theory and mean-variance model, in: IEEE International Conference on Service Operations and Logistics, and Informatics, IEEE/SOLI 2008. 2008, pp. 2668–2672.

38. Y. Zheng, M. Zhou, G. Li, Information entropy based fuzzy optimization model of electricity purchasing portfolio, in: Power & Energy Society General Meeting, PES ’09, IEEE, 2009, pp. 1–6.

39. Y. Jiang, S. He, X. Li, A maximum entropy model for large-scale portfolio optimization, in: International Conference on Risk Management & Engineering Management, ICRMEM ’08, 2008, pp. 610–615.

40. Cover, T. M.; Thomas, J. A. (2006). "Chapter 12, Maximum Entropy". Elements of Information Theory (2 ed.). Wiley.

Appendix

Motivation:

Let m single-period asset returns be denoted by

$$\vec{X} = (x_1, \dots, x_m), \qquad (1)$$

such that their mean returns are

$$\vec{\mu} = (\mu_1, \dots, \mu_m), \qquad (2)$$

and a $m \times m$ covariance matrix

$$\Sigma: \Sigma_{ij} = \mathbb{E}\left(X_i, X_j\right) - \mu_i \mu_j, \quad 1 \le i, j \le m, \qquad (3)$$

assuming that all the aforementioned variables can be estimated using data Then, we denote the return on the portfolio as

$$X = \sum_{i=1}^{m} w_i X_i \qquad (4)$$

where the portfolio weights are denoted by

$$\vec{w} = (w_1, \dots, w_m), \qquad (5)$$

which then has mean and variance

$$\mathbb{E} = \vec{w}\vec{\mu}^T, \quad V(X) = \vec{w}\Sigma\vec{w}^T \qquad (6)$$

Efficient Frontier:

Let us consider our portfolio of containing two stocks, A and B. Each stocks' mean expected return and standard deviation is

$$\mu_A, \mu_B \qquad (7)$$

and

$$\sigma_A, \sigma_B \qquad (8)$$

respectively. Now let the weights of the portfolio be denoted by w_A and w_B such that

$$w_A + w_B = 1 \qquad (9)$$

as the weights add upto a total of 100% or 1. Then with the application of basic econometric theory,

$$E(R_P) = w_A\mu_A + w_B\mu_B \qquad (10)$$

$$= w_A\mu_A + (1 - w_A)\mu_B \qquad (11)$$

$$= w_A(\mu_A - \mu_B) + \mu_B \qquad (12)$$

$$\sigma^2 = w_A^2\sigma_A^2 + w_B^2\sigma_B^2 + 2w_A w_B \rho\sigma_A\sigma_B \qquad (13)$$

$$= \sigma^2 = w_A^2\sigma_A^2 + (1 - w_A^2)\sigma_B^2 + 2w_A(1 - w_A)\rho\sigma_A\sigma_B < (w_A\sigma_A + (1 - w_A)\sigma_B)^2 \qquad (14)$$

The maximization problem then will have the following form;

$$\max \sum_{i=1}^{N} w_i\,\mu_i \qquad (15)$$

$$\text{s.t.} \;\; \sigma^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i\,w_j\sigma_i\sigma_j \qquad (16)$$

Minimum Variance:

$$\text{Min:}\sigma_p^2 = w'\Sigma\; w \qquad (17)$$

$$\text{s.t. } w'l = 1 \qquad (18)$$

where the covariance matrix of the portfolio assets is denoted as $N \times N$ and Σ represents the set consisting of prospective investments, w is a $N \times 1$ vector which includes the weight of each asset, finally l is a $N \times 1$ vector consisting of ones.

This optimization problem has the solution of the form

$$w_{MV} = \frac{\Sigma^{-1}l}{l'\Sigma^{-1}l} \qquad (19)$$

with Σ^{-1} representing an inverse matrix such that

$$\Sigma^{-1} = \text{Diag}(1/\sigma^2) - \frac{(\beta/\sigma^2)(\beta/\sigma^2)'}{\frac{1}{\sigma_M^2} + (\beta/\sigma^2)'\beta} \qquad (20)$$

De silva et. al.(2011,2013)[ref both papers w/ numbers] have shown that the returns on a given i^{th} asset can be modelled using the following basic CAPM equation as

$$r_i = \alpha_i + \beta_i r_M + \varepsilon_i \qquad (21)$$

where $\sigma_{\varepsilon,i}^2$ is the idiosyncratic random variable which is not correlated with all other factors within the above CAPM model, the error term ε_i has mean of zero.

This gives us an $N \times N$ covariance matrix of the form

$$\mathbf{\Sigma} = \boldsymbol{\beta}\boldsymbol{\beta}'\sigma_M + \mathrm{Diag}(\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^{\mathbf{2}}) \qquad (22)$$

where the variance within the CAPM model is denoted by σ_M, the $N \times 1$ vector of β_i is denoted by $\boldsymbol{\beta}$, $\mathrm{Diag}(\sigma_\varepsilon^2)$ being the diagonal matrix consisting of $\sigma_{\varepsilon,i}$ with null values populated in all places except the diagonal coordinates.

To find the weights for our portfolio allocation, we solve eq.(9) and eq.(12) whose computational solution is

$$w_i = \frac{\sigma_{MV}^2}{\sigma_{\varepsilon,i}^2}(1 - \frac{\beta_i}{\beta_{LS}}) \qquad (23)$$

such that the beta of the market (ex ante) is given by β_i. The beta threshold **Long-Short** β_{LS} is a determining variable (for the weights) in an unconstrainted portfolio optimization, and is represented by the implicit solution

$$\beta_{LS} = \frac{\frac{1}{\sigma_M^2} + \Sigma \frac{\beta_i^2}{\sigma_{\varepsilon,i}^2}}{\Sigma \frac{\beta_i^2}{\sigma_{\varepsilon,i}^2}}. \qquad (24)$$

Thus, from our above solution for individual weights eq.(13) we have two cases:

- Those securities which have $\beta_i < \beta_{LS}$ will have positive weights
- Those securities which have $\beta_i > \beta_{LS}$ will have negative weights

For the **long only** case we will discard the second case and only consider the first. In doing so we have a conditional expression as follows;

$$w_i = \begin{cases} \frac{\sigma_{LMV}^2}{\sigma_{\varepsilon,i}^2}\left(1 - \frac{\beta_i}{\beta_L}\right) & \text{if } \beta_i < \beta_L \\ 0, & \text{otherwise} \end{cases} \qquad (25)$$

here the long only MV-portfolio's variance (ex ante) is denoted by σ_{LMV}. The threshold beta for the long only portfolio is denoted by β_L which has the following expression;

$$\beta_{LS} = \frac{\frac{1}{\sigma_M^2} + \Sigma\beta_i < \beta_L \frac{\beta_i^2}{\sigma_{\varepsilon,i}^2}}{\Sigma\beta_i < \beta_L \frac{\beta_i^2}{\sigma_{\varepsilon,i}^2}}. \qquad (26)$$

Maximum Diversification:

Assumption: Let R be the returns on the assets such that $R(w)$ is the returns as a function of the weights w. Then we can assume that the risk is directly proportional to the expected additional returns on the asset. We mathematically formulate as follows;

$$E[R(\mathbf{w})] = k\mathbf{w}'\boldsymbol{\sigma}, \qquad (28)$$

where k being a proportionality constant.

Substituting eq.(28) in eq.(27) we get,

$$\text{Max:}D(\mathbf{P}) = \frac{E[R(\mathbf{w})]}{\sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}}}, \qquad (29)$$

where

$$D(P) = \frac{\mathbf{w}}{k}. \qquad (30)$$

Maximum Decorrelation:

$$\text{Min:}_w \mathbf{w}^T\mathbf{C}\mathbf{w} \qquad (31)$$

$$\text{s.t. } \mathbf{1}^T\mathbf{w} = 1 \qquad (32)$$

Thus, we rely only on the correlation as a measure of variation in this method. Equal Risk Contribution (a.k.a Risk Parity Contribution) (Mailard, Roncalili, Teiletche 2008)

Equal Risk Contribution:

For a set of n risky assets let $w = (w_1, \ldots, w_n)$ represent a set of weights, then

$$c_i(w) = \frac{(\Sigma\mathbf{w})_i}{\sqrt{\mathbf{w}'\Sigma\mathbf{w}}} = \frac{\partial\sigma(w)}{\partial w_i} = \frac{\partial\sqrt{\mathbf{w}'\Sigma\mathbf{w}}}{\partial w_i} \qquad (33)$$

is the marginal volatility contribution of an asset i which represents the change in volatility which is a consequence of minor changes caused within the portfolio by the changes of the weights of an asset. Here the $(\Sigma(w))$ is a $N \times 1$ dimensional vector.

We can decompose the portfolio risk by using the following expression;

$$\sigma(\mathbf{w}) = \sum_{i=1}^{n} w_i\, c_i(\mathbf{w}) = \mathbf{w}' \frac{\Sigma \mathbf{w}}{\sqrt{\mathbf{w}'\Sigma\mathbf{w}}} = \sqrt{\mathbf{w}'\Sigma\mathbf{w}}, \qquad (34)$$

here the i^{th} assets' total risk contribution is denoted by $\sigma_i(\mathbf{w}) = w_i \times c_i(\mathbf{w})$.

By solving the following optimization problem, the weights can be found as follows;

$$\text{Min:} \sum_{i=1}^{n} \left(\sigma(\mathbf{w})/n - w_i c_i(\mathbf{w})\right)^2, \qquad (35)$$

$$\text{s.t.} \sum_{i=1}^{n} w_i = 1 \qquad (36)$$

Maximum Entropy:

Shannon Entropy

Say we have a discrete random variable X with a set of possible outcomes contained in the set $x_1, \dots, x_n$, with each outcome having a probability $P(x_1), \dots, P(x_n)$. Then the Shannon entropy for the random variable X is represented by the following expression,

$$H(X) = -\sum_{i=1}^{n} P\,(x_i)\log P(x_i) \qquad (38)$$

where the expression $H(X)$ gives the level of entropy (on average) in the random variables' outcomes.

We will be using a modified form of Shannon entropy where the probability outcome will be replaced by the weight w_i. We have borrowed this idea from J, Lee, Chiou (2014). The modified Shannon entropy is then of the form,

$$H = -\sum_{i=1}^{n} w_i \ln w_i, \quad \sum_{i=1}^{n} w_i = 1, \quad 1,2,\dots,n \qquad (39)$$

where there are n investable assets. Since $\sum_{i=1}^{n} w_i = 1$ $w_i = \frac{1}{n}$, we can thus see that this is an optimization problem with a constraint. Our aim is to maximize the entropy function H, meaning our observations (data) will relay more information. This also includes cases of extremity where the wight will either completely be one (100%) or have no relevant value at all (zero).

There have been additional modifications to the mean-variance model to comply with the construction using Shannon entropy as the focal point. Ke and Zhang(2008) have proposed the following modifications to the mean-variance procedure;

$$\text{Min: } w^T \Omega w + \sum_{i=1}^{n} w_i \ln w_i \qquad (40)$$

$$\text{s.t.} \sum_{i=1}^{n} r_i\, w_i = E, \qquad (41)$$

$$\sum_{i=1}^{n} w_i = 1 \qquad (42)$$

$$w: (w_1, \ldots, w_n)^T \qquad (43)$$

where the covariance matrix is denoted by Ω, the returns (on average) are denoted by r_i.

Zheng et al. (2009) have proposed to replace the variance within the model with entropy by using the following method;

$$\text{Max} - \sum_{i=1}^{n} w_i \ln w_i \qquad (44)$$

$$\text{s.t.} \sum_{i=1}^{n} r_i\, w_i = E \qquad (45)$$

$$\sum_{i=1}^{n} \sum_{t=1}^{m} w_i\, (r_{it} - \bar{r_i})^2 \leq \text{Var}, \qquad (46)$$

$$\sum_{i=1}^{n} w_i = 1 \qquad (47)$$

where the variance is Var and the return of the asset at time t is r_{it}.

For further discourse on the optimization of the lagrangian function using the above equations, please refer to Jiang et. al. (2008) for a more copious discussion, wherein the solution ultimately leads to the weights which are of the form;

$$L = -\sum_{i=1}^{n} w_i \ln w_i + \gamma_1 (\sum_{i=1}^{n} r_i\, w_i - E) + \gamma_2 (\sum_{i=1}^{n} \sum_{t=1}^{m} w_i\, (r_{it} - \bar{r_i})^2 - \text{Var}) + \gamma (\sum_{i=1}^{n} w_i - 1) \qquad (48)$$

which yields the solution

$$w_i = \frac{e^{\left[\gamma_1 \bar{r_i} + \gamma_2 \sum_{t=1}^{m} (r_{it} - \bar{r_i})^2\right]}}{\sum_{i=1}^{n} e^{\left[\gamma_1 \bar{r_i} + \gamma_2 \sum_{t=1}^{m} (r_{it} - \bar{r_i})^2\right]}}. \qquad (49)$$

we shall now discuss the proof in a simplified manner to retain the financial core of our exercise and not wither away in the mathematical depths of the subject.

Proof of Maximum Entropy

We shall take the above equations (refer equations and rephrase this statement) and solve the following lagrangian problem.

Proof: We start by

$$\text{Max: } -\sum_{i=1}^{n} w_i \ln w_i \qquad (50)$$

$$\text{s.t.} \sum_{i=1}^{n} w_i = 1 \qquad (51)$$

$$\text{Min: } w^T \Omega w + \sum_{i=1}^{n} w_i \qquad (52)$$

thus we have a lagrange equation of the form;

$$L(w, \lambda) = -\sum_{i=1}^{n} w_i \ln w_i + (\lambda + 1)\left(\sum_{i=1}^{n} w_i - 1\right) \qquad (53)$$

Imposing the first order condition;

$$\text{First Order Condition} = \begin{cases} \frac{\partial L}{\partial w_i} = (1 + \ln w_i) + (\lambda + 1) = 0 \\ \frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} w_i - 1 = 0, \end{cases} \qquad (54)$$

solving the first order condition on the lagrangian equation gives us

$$w_i = e^{\lambda} \qquad (55)$$

$$ne^{\lambda} = 1 \qquad (56)$$

which gives us $w_i = \frac{1}{n}$, same result as that from equation(a), meaning it is possible to maximize the entropic function and get the resultant solution as weights.