Pension Funds Should Never Rely on Correlation

Ronald Lagnado†
Nassim Nicholas Taleb*† ‡
†Universa Investments
‡Tandon School of Engineering, New York University
November 2021
Forthcoming, *Journal of Alternative Investments*

INTRODUCTION/ABSTRACT

The central decision for a pension fund is the allocation between stocks and bonds, often relying, for intellectual backup, on metrics and methods from Modern Portfolio Theory (MPT). We show how, historically, such an "optimal" portfolio is in effect the *least* optimal one, as it fails to protect against tail risk and under-allocates to the high-returning asset class. MPT fails in both risk control and real-world investment optimization.

MODERN PORTFOLIO THEORY

Modern Portfolio Theory (MPT) originated about 70 years ago with the publication of "Portfolio Selection" , by Harry Markowitz [1]. The mean-variance tools of analysis accompanying the Markowitz framework was a cornerstone supporting the Capital Asset Pricing Model (CAPM), which intended to explain the relationship between systemic risk and asset returns. In spite of its adoption in business schools, CAPM has never been supported by empirical evidence –in fact changing assumptions about joint asset returns makes the mathematical framework crumble. Accordingly, it is not our intent to make the case here. Instead, we consider the pitfalls of MPT when applied by pension funds in asset allocation and portfolio management, focusing specifically on the role of correlation.

Why is this of concern? Blind deference to MPT, which we repeat made neither empirical nor theoretical sense, caused most institutional investors to misallocate assets. The average funded status of US state-managed pension plans dropped from almost 100% in 2001 to about 74% in 2019. With combined liabilities of over $4 trillion, the shortfall stands grimly at over $1 trillion despite the longest economic expansion in U.S. history spanning over a decade from 2009 to 2020.

MPT with its central tool of mean-variance analysis supposedly provides a formal argument for the benefits of diversification in investing. It leads investors seeking to determine the so-called "optimal" allocation to broad asset classes in top-down portfolio construction. But there are two major fallacies.

The first fallacy is that it is possible to forecast asset returns, volatilities, and correlations with sufficient accuracy ex ante

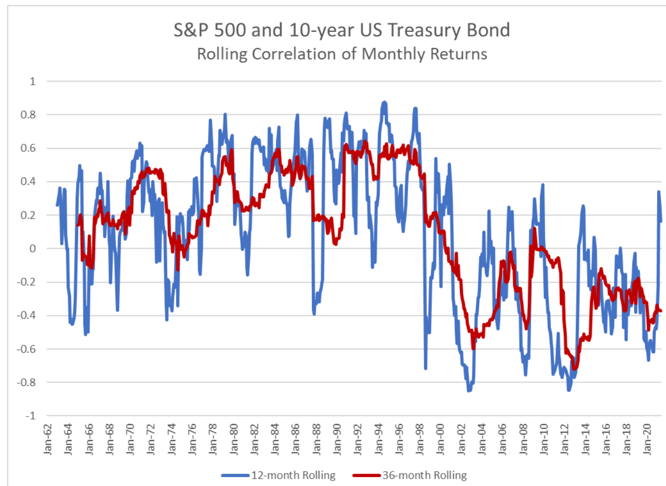

Fig. 1. Historical correlation swings between -1 and 1; there is less than 10^{-12} chance that these variations would come from simple sampling error under standard assumptions. The technical appendix explains why this is sufficient to invalidate the risk-reduction claims of MPT.

to facilitate the optimal allocation and the maximum risk-adjusted return ex post.

Comment 1: First Fallacy of MPT
Future returns and correlation —the two central parameters in MPT —are not observable in the real world.

The second misconception is that attaining the objective of maximizing risk-adjusted returns, i.e., maximizing returns subject to a constraint on volatility, is beneficial. Does suppression of volatility across all market environments provide cost effective protection against tail risk and ultimately better long-run compounded returns? Clearly, it does not.

Comment 2: Second Fallacy of MPT
Volatility of returns (as expressed in portfolio standard deviation) does not map to risk.

CORRELATION AND RISK PARITY

The notion of stable and estimable expected returns is the easiest aspect of MPT to debunk. Nevertheless, extensions of

Corresponding author: N N Taleb. Thanks: Chitpuneet Mann, Brandon Yarckin, Mark Spitznagel, George Martin.

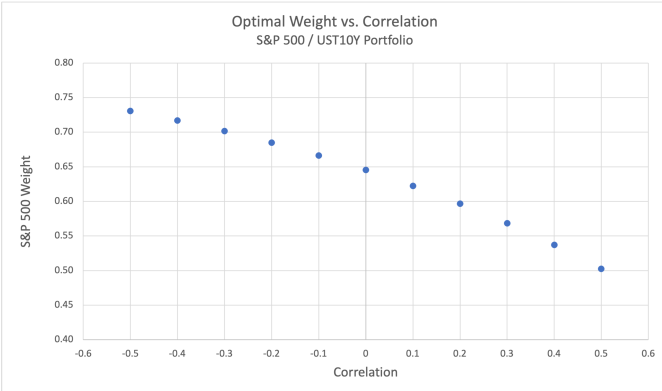

Fig. 2. Optimal weights.

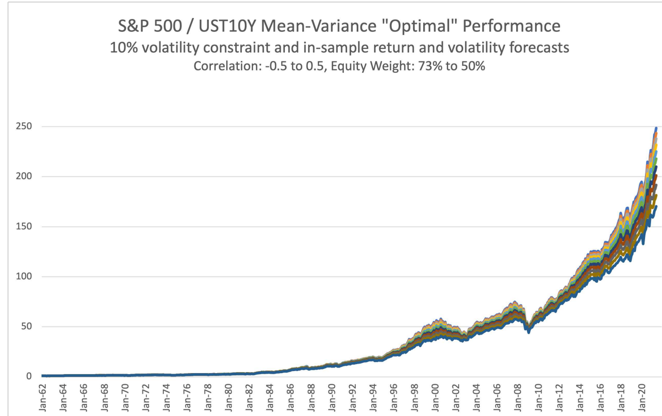

Fig. 3. Resulting Performance.

the theory to risk parity strategies obviate the estimation of expected return but still require accuracy in estimating risk parameters.

The key risk parameter – correlation –is neither robust nor reliable a measure of dependence when asset returns have fat-tailed distributions or when the relationship between the variables is not linear —both cases seem to apply to stocks and bonds. Let's never ignore that stocks and bonds have patently non-Gaussian distributions, and the relationship between the two variables, when regressed, exhibits a severe nonlinearity.

Comment 3: Correlation and association

Correlation, even if predictable (it's not), is a poor measure of association between nonGaussian variables or in the presence of nonlinear dependence.

To illustrate the problems associated with application of correlation in MPT, consider a risk averse investor who constructs a portfolio with allocation to the S&P 500 index and a U.S 10-year Treasury Bond index. Over the approximately 60-year period from January 1962 to June 2021, the annualized total return and volatility for the S&P 500 was 11% and 14.9%, respectively. The return and volatility of the constant maturity U.S. 10-year Treasury Bond were 7.1% and 7.8%, respectively. (Note that since we believe that, owing to the nonGaussian attributes of the returns, standard deviation is not a fit measure of expected future variations, we will only use it in this exercise insofar as it is illustrative of that specific shortcoming.)

Figure 1 shows the correlation of monthly returns over this period with 6, 12, and 36-month rolling windows. First, we can verify that correlation is far from being constant, violating the requirements for hedging and risk reduction. Second, more technically, it does appear to have a random structure, making the joint distribution between the two variables non-elliptical, thus violating a central assumption at the foundation of portfolio theory. The mathematical consequence is that correlation can be deemed, at best, as a non-informative metric and an unreliable indicator of the association. See Taleb (2020) [2] for the mathematical derivations.

Suppose the investor has perfect foresight with respect to future returns and volatilities, and both assets follow the 1962-2021 trajectories over the subsequent 60-year period. The initial allocation (with full investment and no leverage) is determined by maximizing expected return subject to a typical 10% volatility constraint; the portfolio is rebalanced monthly with zero cost. The optimal weight for the S&P 500 as a function of a correlation estimate ranging from -0.5 to $+0.5$ is shown in Figure 2. Depending on the ex-ante assumption for correlation the investor would have selected a portfolio ranging from 73/27 to 50/50 in composition with the conviction that the "optimal" choice had been made.

Table 1 presents the range in realized return over the subsequent 60-year period shown in Figure 3; we can see extremely widespread outcomes. The accumulated wealth would have been about 1.5 times greater if the correlation had been assumed to be -0.5 rather than 0.5. Assuming that bonds are consistently negatively correlated to stocks would have resulted in a higher stock allocation of 73% versus 50% (and a better outcome). However, the average realized correlation between stocks and bonds during the decades prior to the DotCom Bubble collapse in 2001-2002 was about 0.3. There would have been no empirical grounds to support an assumption of negative correlation, and optimization would have resulted in an under-allocation to stocks.

TABLE I
VARYING CORRELATION

Assumed Correlation	Equity Weight	Compound Return	Annualized Return	Vol.	Ret/Vol
-0.5	0.73	247.63	0.1	0.11	0.86
-0.4	0.72	242.57	0.1	0.11	0.87
-0.3	0.7	237.03	0.1	0.11	0.88
-0.2	0.68	230.95	0.1	0.11	0.9
-0.1	0.67	224.29	0.1	0.1	0.91
0.	0.65	216.98	0.09	0.1	0.93
0.1	0.62	208.96	0.09	0.1	0.94
0.2	0.6	200.19	0.09	0.1	0.96
0.3	0.57	190.65	0.09	0.09	0.98
0.4	0.54	180.37	0.09	0.09	1.01
0.5	0.5	169.44	0.09	0.09	1.03

To summarize, all portfolio allocations represented in Table 1 are mean-variance "optimal" for different assumed correlations. Choosing a low correlation of -0.5 enabled the investor

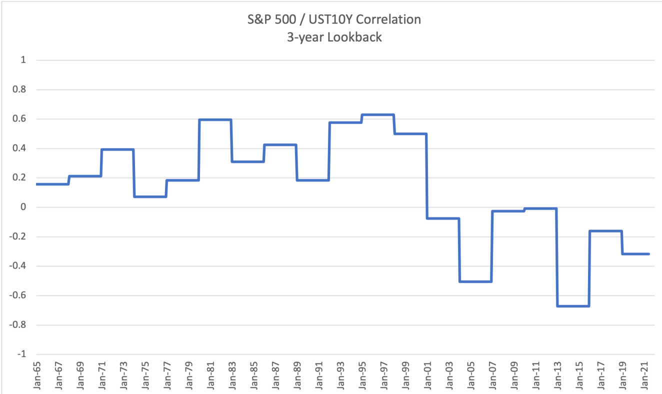

Fig. 4. Lookback correlation levels.

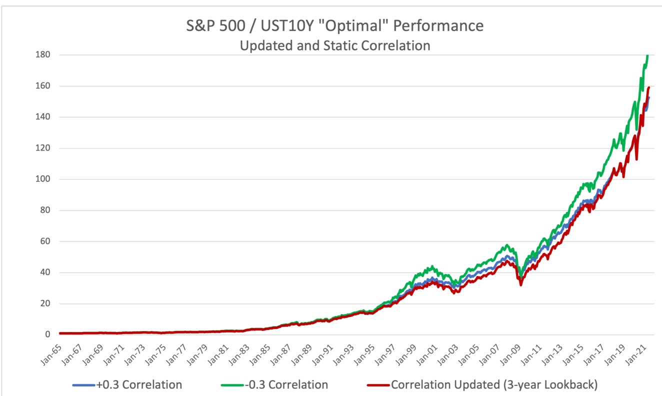

Fig. 5. "Optimal Performance"

to reap the greatest reward on this path through time despite the worst risk-adjusted return and a volatility exceeding the target of 10%. Choosing a high correlation of 0.5 (supported by so-called empirical evidence) resulted in the best "risk adjusted" return but the worst cumulative gain. This sensitivity to an unpredictable parameter — correlation — highlights the fallacy of the label portfolio "optimization" and utter uselessness of MPT.

Chasing Correlation

Would a real pension fund manager simply allocate based on a point estimate of correlation and maintain the allocation indefinitely? As preposterous as that may sound, it has been the actual practice at some funds that subscribe to the 60/40 approach to portfolio construction. Alternatively, it would seem sensible to update correlation and other parameters periodically and recompute the "optimal" allocation.

Pension funds typically perform a review of asset allocation and revise the apportionment every 2 to 5 years. Suppose instead of maintaining a constant assumed correlation, the investor estimates correlation using historical data in a 3-year lookback period and recomputes and maintains the mean-variance optimal allocation for the subsequent 3-year period. The history of the estimated correlation over the period from 1965 to 2021 is shown in Figure 4. Figure 5 then compares the performance of the dynamic "optimal" portfolio using 3-year lookback correlation with static portfolios based on assumptions of $+0.3$ and -0.3 correlation —average values in sub-periods before and after the year 2000. The performance of the dynamic portfolio is no better and, in fact, much worse than would be attained under assumptions leading to higher equity allocations.

Conclusion

The single most consequential investment decision for a pension is the allocation between stocks and bonds (or more generally between growth and income investments). The influence of MPT on pension fund managers is apparent in that a 60/40 stock-bond portfolio is representative of pension funds on average. Correlation is the driving force behind the allocation decision — while both the mathematical and empirical justifications of such an application have been lacking.

We believe that the deficiencies of MPT and the reliance on correlation estimates have been responsible for tepid performance of pension funds since the Global Financial Crisis. It is never possible to reliably construct the "optimal" portfolio that will deliver the best possible risk-adjusted return. Even if it were possible, it would not be beneficial. Following this course limits portfolio volatility in benign market environments over the short term while making huge sacrifices in long-run performance. The so-called "optimal" portfolio is in effect the worst of all worlds. It offers scant protection against tail risk and, at the same time, achieves an under-allocation to the riskier assets with higher returns in the long periods of economic expansion such as the past decade.

References

[1] H. Markowitz, "Portfolio selection*," *The journal of finance*, vol. 7, no. 1, pp. 77–91, 1952.

[2] N. N. Taleb, *The Statistical Consequences of Fat Tails*. STEM Academic Press, 2020.

[3] K.-T. Fang, "Elliptically contoured distributions," *Encyclopedia of Statistical Sciences*, 2006.

[4] J. Owen and R. Rabinovitch, "On the class of elliptical distributions and their applications to the theory of portfolio choice," *The Journal of Finance*, vol. 38, no. 3, pp. 745–752, 1983.

[5] M. Hamada and E. A. Valdez, "Capm and option pricing with elliptically contoured distributions," *Journal of Risk and Insurance*, vol. 75, no. 2, pp. 387–409, 2008.

[6] R. Chicheportiche and J.-P. Bouchaud, "The joint distribution of stock returns is not elliptical," *International Journal of Theoretical and Applied Finance*, vol. 15, no. 03, 2012.

[7] N. N. Taleb, "The meta-distribution of standard p-values," *arXiv preprint arXiv:1603.07532*, 2016.

Technical Appendix

What is the failure of Ellipticality?

From the standard definition, [3], $\mathbf{X}$, a $p \times 1$ random vector is said to have an elliptical (or elliptical contoured) distribution with location parameters μ, a non-negative matrix Σ, and some scalar function Ψ if its characteristic function φ is of the form

$$\varphi(t) = \exp(it'\mu)\Psi(t\Sigma t'). \tag{1}$$

There are equivalent definitions focusing on the density; consider that the central attribute is that Ψ is a function of *a single* covariance matrix Σ.

The main property of the class of elliptical distribution is that it is closed under linear transformation. Intuitively, it means (in a bivariate situation) that tail effects are less likely to come from one than two marginal deviations.

This closure under linear transformation leads to attractive properties in the building of portfolios, and in the results of portfolio theory (in fact one cannot have portfolio theory without ellipticality of distributions). For under ellipticality, all portfolios can be characterized completely by their location and scale and any two portfolios with identical location and scale (in return space) have identical distributions returns.

Accordingly, ellipticality (under the condition of finite variance) allows the extension of the results of modern portfolio theory (MPT) under the so-called "nonnormality", an argument initally discovered by [4], also see [5]. However it appears (from those of us who work with stochastic correlations) that returns are not elliptical by any conceivable measure, see Chicheportiche and Bouchaud [6] and simple visual graphs of stability of correlation as in Fig.1.

In other words, the characteristic function under stochastic correlation becomes:

$$\varphi(t)=\sum_{j}^{n}\omega_j \exp(it'\mu)\Psi(t\Sigma_j t'). \tag{2}$$

where Σ_j represents different covariance matrices indexed by the state j, and ω_j are the weights of the distributions under consideration mapping to the frequencies of states. So if, as we saw in Fig. 1, one needs to average across correlations, the condition in equation 2 fails and, given that market returns have fat tails, and there is no mathematical justification for MPT.

Distinguishing between sample error and random correlation

Next we derive the distribution of the second moment of a sample of correlation coefficients under the null assumption that variations are due to sampling error off a constant correlation. We focus on non-overlapping observations from a sample size of 713 months of underlying pairs (of stocks and bonds).

Let X and Y be n independent Gaussian variables centered to a mean 0 and scaled to a unitary variance. Let $\rho_n(.)$ be the operator.

$$\rho_n(\tau)=\frac{X_\tau Y_\tau + X_{\tau+1}Y_{\tau+1}\ldots + X_{\tau+n-1}Y_{\tau+n-1}}{\sqrt{(X_\tau^2+X_{\tau+1}^2\ldots+X_{\tau+n-1}^2)(Y_\tau^2+Y_{\tau+1}^2\ldots+Y_{\tau+n-1}^2)}}. \tag{3}$$

First, we consider the distribution of the Pearson correlation for n observations of pairs assuming $\mathbb{E}(\rho)\approx 0$ (the mean is of small relevance as we are focusing on the second moment, with tracking of $O\left(\frac{1}{n^2}\right)$):

$$f_n(\rho)=\frac{\left(1-\rho^2\right)^{\frac{n-4}{2}}}{B\left(\frac{1}{2},\frac{n-2}{2}\right)}, \tag{4}$$

with characteristic function:

$$\chi_n(\omega)=2^{\frac{n-1}{2}-1}\omega^{\frac{3-n}{2}}\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)J_{\frac{n-3}{2}}(\omega),$$

where $J_{(.)}(.)$ is the Bessel J function.

We can assert that, for n sufficiently large:

$$2^{\frac{n-1}{2}-1}\omega^{\frac{3-n}{2}}\Gamma\left(\frac{n}{2}-\frac{1}{2}\right)J_{\frac{n-3}{2}}(\omega)\approx e^{-\frac{\omega^2}{2(n-1)}},$$

the corresponding characteristic function of the Gaussian.

Moments of order p become:

$$M(p)=\frac{((-1)^p+1)\,\Gamma\left(\frac{n}{2}-1\right)\Gamma\left(\frac{p+1}{2}\right)}{2B\left(\frac{1}{2},\frac{n-2}{2}\right)\Gamma\left(\frac{1}{2}(n+p-1)\right)} \tag{5}$$

where $B(.,.)$ is the Beta function. The standard deviation is $\sigma_n=\sqrt{\frac{1}{n-1}}$ and the kurtosis $\kappa_n=3-\frac{6}{n+1}$.

This allows us to treat the distribution of ρ as Gaussian, and given infinite divisibility, derive the variation of the components, again of $O(\frac{1}{n^2})$ (hence simplify by using the second moment in place of the variance):

$$\rho_n\sim\mathcal{N}\left(0,\sqrt{\frac{1}{n-1}}\right),$$

To test how the second moment of the sample coefficient compares to that of a random series, and thanks to the assumption of a mean of 0, define the squares for nonoverlapping correlations:

$$\Delta_{n,m}=\frac{1}{m}\sum_{i=1}^{\lfloor m/n\rfloor}\rho_n^2(i\,n),$$

where m is the sample size and n is the correlation window.

Now we can show that:

$$\Delta_{n,m}\sim\mathcal{G}\left(\frac{p}{2},\frac{2}{(n-1)p}\right),$$

where $p=\lfloor m/n\rfloor$ and $\mathcal{G}$ is the gamma distribution with PDF:

$$f(\Delta)=\frac{2^{-\frac{p}{2}}\left(\frac{1}{(n-1)p}\right)^{-\frac{p}{2}}\Delta^{\frac{p}{2}-1}e^{-\frac{1}{2}\Delta(n-1)p}}{\Gamma\left(\frac{p}{2}\right)},$$

and survival function:

$$S(\Delta)=Q\left(\frac{p}{2},\frac{1}{2}\Delta(n-1)p\right),$$

which allows us to obtain p-values below, using $m=714$ observations (and using the leading order $O(.)$:

n	Sample Δ_n	p-values $O(.)$
12	0.2049	10^{-8}
15	0.1735	10^{-8}
20	0.1535	10^{-9}
25	0.1324	10^{-9}
36	0.1370	10^{-13}

Such low p-values exclude any controversy as to their effectiveness [7].

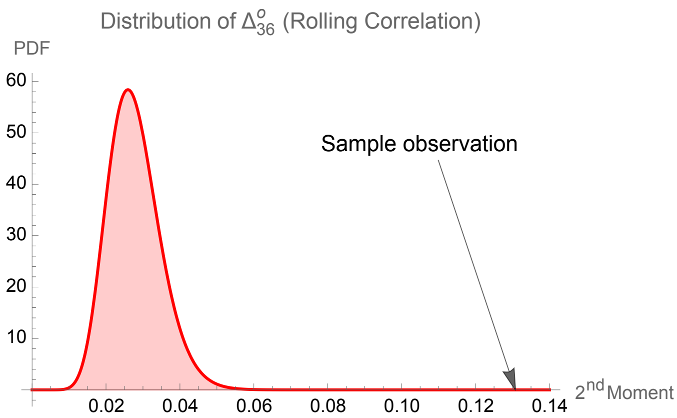

Fig. 6. The Monte Carlo distribution of rolling correlation for 36 months ($n = 10^6$). We get astonishingly low p-values of order 10^{-17}.

We can also compare rolling correlations using a Monte Carlo for the null with practically the same results (given the exceedingly low p-values). We simulate $\Delta^o_{n,m}$ with overlapping observations:

$$\Delta^o_{n,m} = \frac{1}{m} \sum_{i=1}^{m-n-1} \rho_n^2(i),$$

Rolling windows have the same second moment, but a mildly more compressed distribution since the observations of ρ over overlapping windows of length n are autocorrelated (with, we note, an autocorrelation between two observations i orders apart of $\approx 1 - \frac{1}{n-i}$). As shown in Figure 6, for $n = 36$ we get exceedingly low p-values of order 10^{-17}.