NORTHFIELD | modeling financial markets worldwide

Evaluation of Investment Performance when Returns are Not Normally Distributed

Dan diBartolomeo

PMAR

May 25th, 2023

Motivation

Beyond the routine measurement and reporting of investment returns, it is essential that financial organizations analyze performance outcomes to understand their economic implications and statistical significance.

Most metrics of investment performance (e.g. Sharpe Ratio) silently embed the same assumptions that are present in established financial theories (e.g. MPT, CAPM).

In using such metrics in their usual form, we are assuming that returns are normally distributed, random (or almost so), and that transaction costs are so small as to be inconsequential.

Many pension funds use unrealistically high return assumptions for actuarial calculations. Relative to these expectations realized returns will have higher moments.

Presentation Outline

IN THIS PRESENTATION, WE FIRST PROVIDE ILLUSTRATIONS OF THE CRITICAL LIMITATIONS OF OUR PERFORMANCE METRICS WHEN THESE ASSUMPTIONS FAIL.

SECONDLY, WE PROVIDE SOME BASIC METHODS TO TEST WHETHER SUCH ASSUMPTIONS ARE JUSTIFIED FOR A PARTICULAR SET OF RETURN OBSERVATIONS, OR WHETHER ADJUSTMENTS IN ANALYTICAL METHODS ARE REQUIRED.

FINALLY, WE WILL PRESENT SOME ALGEBRAIC ADJUSTMENTS TO TRADITIONAL INVESTMENT PERFORMANCE METRICS THAT WILL ALLOW THEIR VALUES TO BE INTERPRETED CORRECTLY.

WE WILL ALSO DEMONSTRATE HOW THE VOLATILITY EQUIVALENT APPROACH CAN BE USED IN AN EX-ANTE FASHION TO SENSIBLY EVALUATE RETURN DISTRIBUTIONS WHERE A LARGE EVENT MAY BE EXPECTED TO SOMETIME OCCUR IN THE FUTURE, EVEN WHEN NO SUCH LARGE RETURN EVENT IS CONTAINED WITHIN THE DATA SAMPLE BEING ANALYZED.

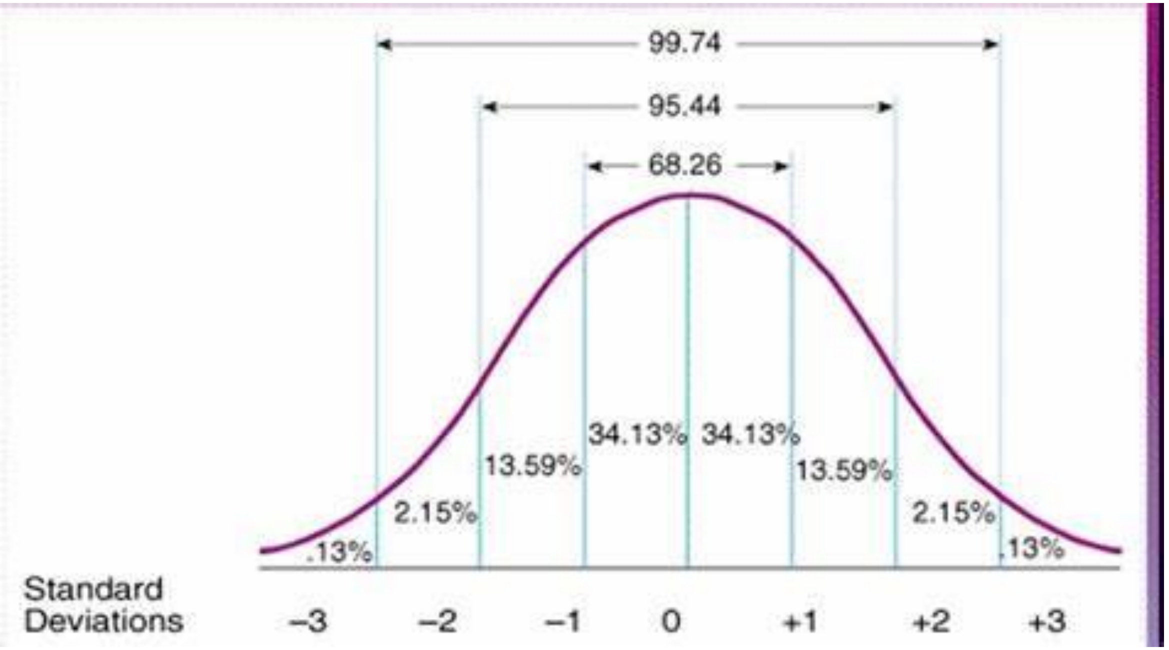
Abuse of the CLT

The assumption of normality comes from Central Limit Theorem of statistics. We know that when we add a sum up a large number of *independent* observations (e.g. returns), the distribution of the sum will be normally distributed irrespective of distributions of the individual observations.

We also know that if transaction costs aren't zero portfolios returns over different periods aren't completely independent, so you need a larger number of observations to get close to normality.

If we tested these assumptions as formal hypotheses, we would find that in many cases we would reject these notions as untrue, and therefore cannot no longer rely on the validity of the performance metrics.

In his book Iceberg Risk, finance researcher Kent Osband aptly named one of the chapters "The Abnormality of Normality".



A Scary, But Real Example

Consider investment returns that outperformed the risk-free asset by 8% per year with a standard deviation of 10% per year. The realized Sharpe Ratio (see Sharpe, 1966) is **.8** (8/10). The expectation of the geometric mean excess return is **7.5%** per annum (see Messmore, 1995).

Now let's consider what happens when we add a rare, but large negative event. Let's assume that with a 2% annual likelihood (1 year in 50), our investment result will be a loss 90% below the risk-free return. The distribution of returns will now be expected to have negative skew and positive excess kurtosis.

The expected arithmetic mean return (numerator) is 6.04% and the volatility equivalent is 33% for a Sharpe ratio of **.18,** *less than a quarter of the earlier value.* The expectation of the geometric mean active return is just **60 basis points (.6%) per year**, *less than a tenth of previous value*.

Volatility Equivalence

In the above example, we redefined the denominator of the Sharpe Ratio as a "volatility equivalent" in presence of skew and kurtosis.

We assert that the single concept of defining a volatility equivalent will suffice to correct other performance metrics such as "Information Ratio" (Rudd and Clasing, 1982) and "mean/variance risk-adjusted return" (Levy and Markowitz, 1979).

Our definition of volatility equivalence will also address other assumptions that often fail in the real world such that transaction costs are so low that we can consider different return periods as independent observations.

In addition, the volatility equivalent method also allows for traditional measures of statistical significance (e.g. T statistics and P values) to be utilized in the context of evaluating agent manager skill.

Some Selected Literature

1

To deal with "risk adjusted return" you need to understand investor risk aversion: Litzenberger and Rubinstein (1976), Wilcox (2000, 2003), diBartolomeo (2021)

2

To consider the statistical significance of Sharpe and Information Ratios we have Jobson and Korkie (1981), Lo (2002), Memmel (2003), Ledoit and Wolf (2008), Bertrand and Protopescu (2010).

3

To adjust observed returns when assets are illiquid, we highlight Geltner (1991) and Lo, Getmansky and Makarov (2008).

- General issues of calculating standard deviation in the presence of serial correlation go back to Bloch (1968) and Brugger (1969).

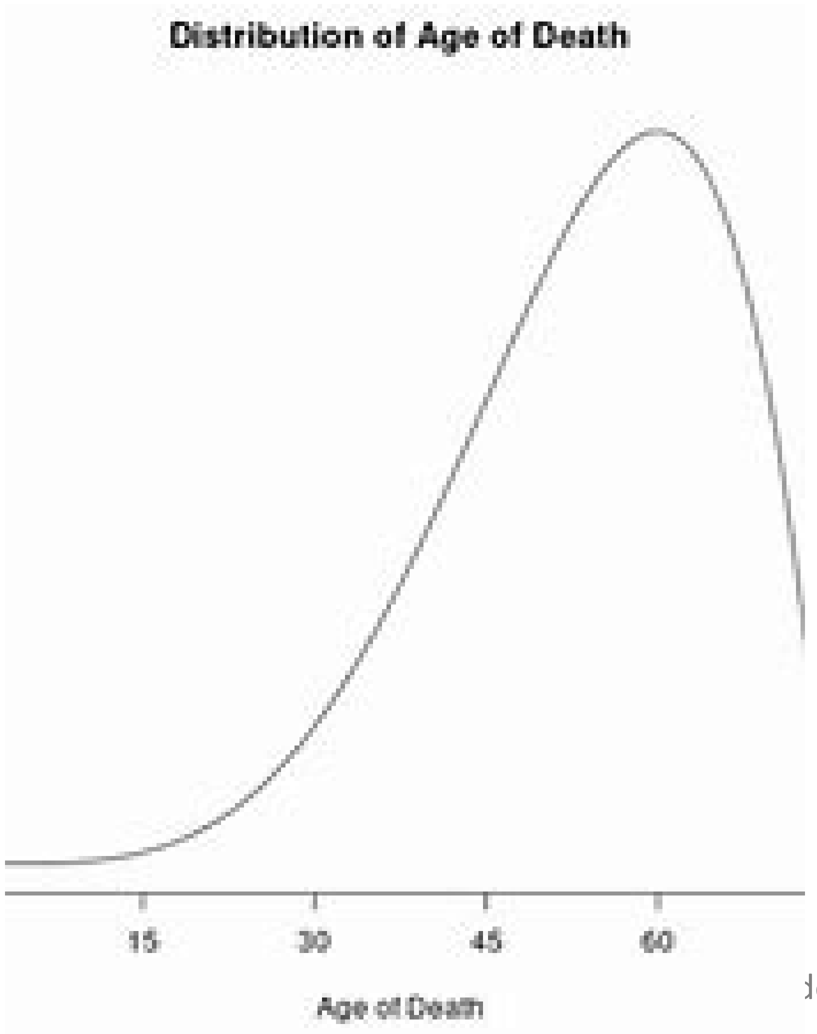
More Literature

The basics of economic interpretation of investment return distributions with skew and kurtosis is provided in Satchell and Hall (2013).

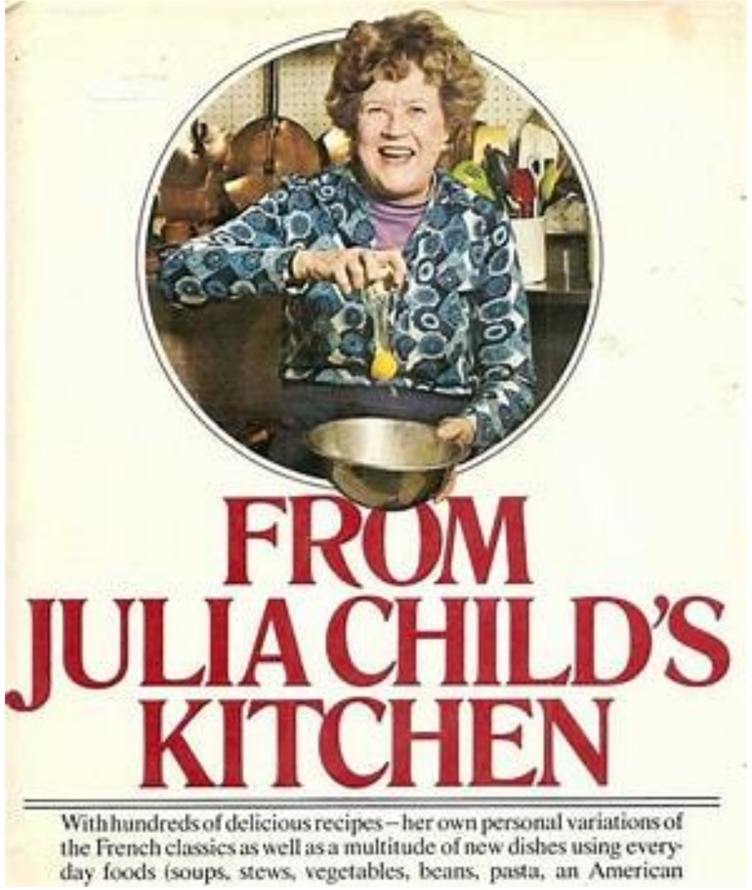
The impact of “rare but large” return events on the empirical evaluation of common equity investment strategies (e.g. “value”, “momentum”) is explored in diBartolomeo and Kantos (2020).

diBartolomeo (2007) provides a theoretical link between “price sensitive” investment strategies and option replicating strategies such as “constant proportion portfolio insurance” as proposed by Perold (1986).

Negative skew in returns for hedge funds arising from being “short volatility” has been well documented in papers such as Weisman (2002), Bondarenko (2004), and Fung and Hsieh (2004).



Ingredients to Today's Process



A method for explicitly forecasting skew and kurtosis in a return distribution conditional on the probability of a "large event" is provided in Blackburn, diBartolomeo, and Zieff (2022) which relies on the "mixture of normal distribution" process first proposed by Robertson and Fryer (1969).

There are many tests for non-normality in the statistics literature which are summarized in Thode (2002).

For illustration we will use the popular JB statistic as proposed by Jarque and Bera (1980).

Our method for converting to the volatility equivalent from a four-moment distribution was first proposed by Cornish and Fisher (1938).

Several papers have proposed algebraic refinements of the Cornish Fisher method including Chernozuckov, Fernandez-Val, and Galichon (2010), and Martin and Arora (2017).

Testing for Imperfect Liquidity

Imperfect liquidity typically manifests as positive serial correlation in returns.

We will first consider the simple case of whether a particular return series of returns exhibits serial correlation. The relationship between the return in the prior period and return in the current period is expressed in a simple univariate equation.

As the indication of statistical significance, practitioners typically treat a T statistic with absolute value between 1.5 and 3 as sufficient to justify concluding significance. If the serial correlation property of the series is significantly different from zero, we apply the relevant adjustment to the volatility of the returns, s_t.

The most widely used adjustment is presented. This formulation makes it easy to calculate with standard spreadsheet statistical functions.

Adjusting for Serial Correlation

$$R_t = a + \rho_{t,t-1} (\sigma_{t/} \sigma_{t-1}) R_{t-1} + e_t$$

$$T = \rho_{t,t-1} (n-2)^{.5} / (1-\rho_{t,t-1}{}^2)$$

$$V_t = ((1 + \rho_{t,t-1})/(1- \rho_{t,t-1}))^{.5} \sigma_t$$

n = number of data points
R_t is the return observed in period t
e_t is the error term of the relationship in period t
$\rho_{t,t-1}$ is the correlation between the return series elements and the returns of the prior period
σ_t is the standard deviation of the returns in the sample period
Vt = the volatility equivalent for the returns in the sample period

Testing for Non-Normality with the JB Statistic

Our next task is to test for and where necessary define volatility equivalence for a set of past returns.

We will first test using the Jarque-Bera (JB) statistic as it is easily done in a spreadsheet, but many other statistical tests for normality exist.

If the value of JB is found to be statistically significant, we can reject the hypothesis that the distribution is normal, and thereby justify the need for the computation of volatility equivalence.

- For samples with large numbers of data points, the probability value of the JB statistic being indicating significance can be computed as a Chi-Square distribution with two degrees of freedom (i.e. a statistical function in most spreadsheets).
- For smaller samples, a table from Wikipedia estimated via Monte Carlo simulation has been provided.

Chi-square test

- Pearson's chi-square (χ^2) test is the best-known of several chi-square tests. It is mostly used to assess the tests of goodness of fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

- The usual rule of thumb is that the chi-square test is not suitable when the expected values in any of the cells of the table, given the margins, is below 5, the sampling distribution of the test statistic that is calculated is only approximately equal to the theoretical chi-squared distribution.

	case	control	
Allele 1	a	b	a+b
Allele 2	c	d	c+d
	a+c	b+d	n=a+b+c+d

The Jarque Bera Statistic

Calculated *p*-values equivalents to true alpha levels at given sample sizes

True α level	20	30	50	70	100
0.1	0.307	0.252	0.201	0.183	0.1560
0.05	0.1461	0.109	0.079	0.067	0.062
0.025	0.051	0.0303	0.020	0.016	0.0168
0.01	0.0064	0.0033	0.0015	0.0012	0.002

$JB = (n/6)\ (S^2 + .25\ (K-3)^2)$

S = the sample skew of the returns

K = the sample kurtosis of the returns (*note many spreadsheets report "excess kurtosis" with the value 3 already subtracted)*

Converting Four Moments to Two

Once we have concluded that the a given sample of returns is not normally distributed, we can use the Cornish Fisher (CF) expansion method to compute volatility equivalence.

We will first choose which percentile P of the return distribution that we will use as the critical value from which to infer volatility equivalence. Given that we are interested in estimating volatility as a metric of risk we are interested in the left tail of the return distribution.

Much as we would for choosing a critical value for statistical significance, percentile choices of 1 to 5% are sensible with 5% being the most common. Practitioners should note that the choice of P will slightly influence the value of the volatility equivalent. When comparing multiple funds or time periods the same value of P should be used.

Definition [edit]

For a random variable ***X*** with mean μ, variance σ², and cumulants κ_n, its quantile y_p at order-of-quantile p can be estimated as $y_p \approx \mu + \sigma w_p$ where:[3]

$$
\begin{aligned}
w_p = \; & x + [\gamma_1 h_1(x)] \\
& + [\gamma_2 h_2(x) + \gamma_1^2 h_{11}(x)] \\
& + [\gamma_3 h_3(x) + \gamma_1 \gamma_2 h_{12}(x) + \gamma_1^3 h_{111}(x)] \\
& + \cdots \\
x &= \Phi^{-1}(p) \\
\gamma_{r-2} &= \frac{\kappa_r}{\kappa_2^{r/2}};\ r \in \{3, 4, \ldots\} \\
h_1(x) &= \frac{\mathrm{He}_2(x)}{6} \\
h_2(x) &= \frac{\mathrm{He}_3(x)}{24} \\
h_{11}(x) &= -\frac{[2\mathrm{He}_3(x) + \mathrm{He}_1(x)]}{36} \\
h_3(x) &= \frac{\mathrm{He}_4(x)}{120} \\
h_{12}(x) &= -\frac{[\mathrm{He}_4(x) + \mathrm{He}_2(x)]}{24} \\
h_{111}(x) &= \frac{[12\mathrm{He}_4(x) + 19\mathrm{He}_2(x)]}{324}
\end{aligned}
$$

where He_n is the n^{th} probabilists' Hermite polynomial. The values γ_1 and γ_2 are the random variable's skewness and (excess) kurtosis respectively. The value(s) in each set of brackets are the terms for that level of polynomial estimation, and all must be calculated and combined for the Cornish–Fisher expansion at that level to be valid.

Recognizing Fat Tails

Given the four moments of the distribution (mean, standard deviation, skew, and kurtosis) we can compute the return value that corresponds to the chosen value of P.

To compute the tail weight parameter, Wp we begin with the value Z that corresponds to the P percentile of the normal distribution (e.g. Z for P = 5% is -1.645). The computation of Wp involves a class of math functions known as Hermite polynomials but the approximation is sufficient for most cases.

Let's compute the example of a distribution with mean 8%, standard deviation 10%, sample skew -4, and sample kurtosis 5 (excess kurtosis = 2), using the 5th percentile as the critical value (Z =-1.645). In this case, Wp = -2.44 and our "volatility equivalent" (V_t) = 14.83 as compared to the original standard deviation of 10.

Cornish Fisher Computation

$V_t = \sigma_t W_p / Z_p$

$W_p = Z_p + (S/6) * (Z_p^2 - 1) + Z_p * ((K-3)/24) * (Z_p^2 - 3) - Z_p * (S^2/36) * (2Z_p^2 - 5)$

W_p = the Cornish Fisher "tail weight" parameter

How about the Dog that Didn't Bark or Bite (yet)?

Financial market returns can be severely impacted by rare, "large" events such as wars, pandemics, and market crashes.

Similarly, individual investments are sometimes subject to rare, generally negative extreme "tail" events such as the credit default of a bond, or the collapse of a levered hedge fund due to margin calls.

Each of these kinds of event are infrequent so it is likely that in most series of observed returns the "big, bad, event" hasn't happened yet.

In many cases the value of an asset or portfolio could go to zero, so the impact of these events cannot be observed in the past because they would constitute the end of the existence of the asset (e.g. SVB)

Forward Looking Tail Probabilities

We can certainly estimate what a sensible volatility equivalent might be given a historical standard deviation of returns, and some assumptions of about *large, negative events could happen but have not happened yet.*

A hedge fund investor might look at hedge fund return databases and observe that each year 3% of the hedge funds that were in the data in the prior year dropped out of the database.

There is an extensive literature on survivorship bias in hedge fund returns such as Liang (2000), Lo (2001), and Posthuma and Van der Sluis (2003).

From any of many studies investors could frame expectations about the *future probability* of a "hedge fund blow- up" and the range of loss severity when such funds suffer difficulties.

Even for unlevered portfolios such as passive index funds there is always some non-zero probability of a severe market decline due to the onset of a war, pandemic, or financial crisis.

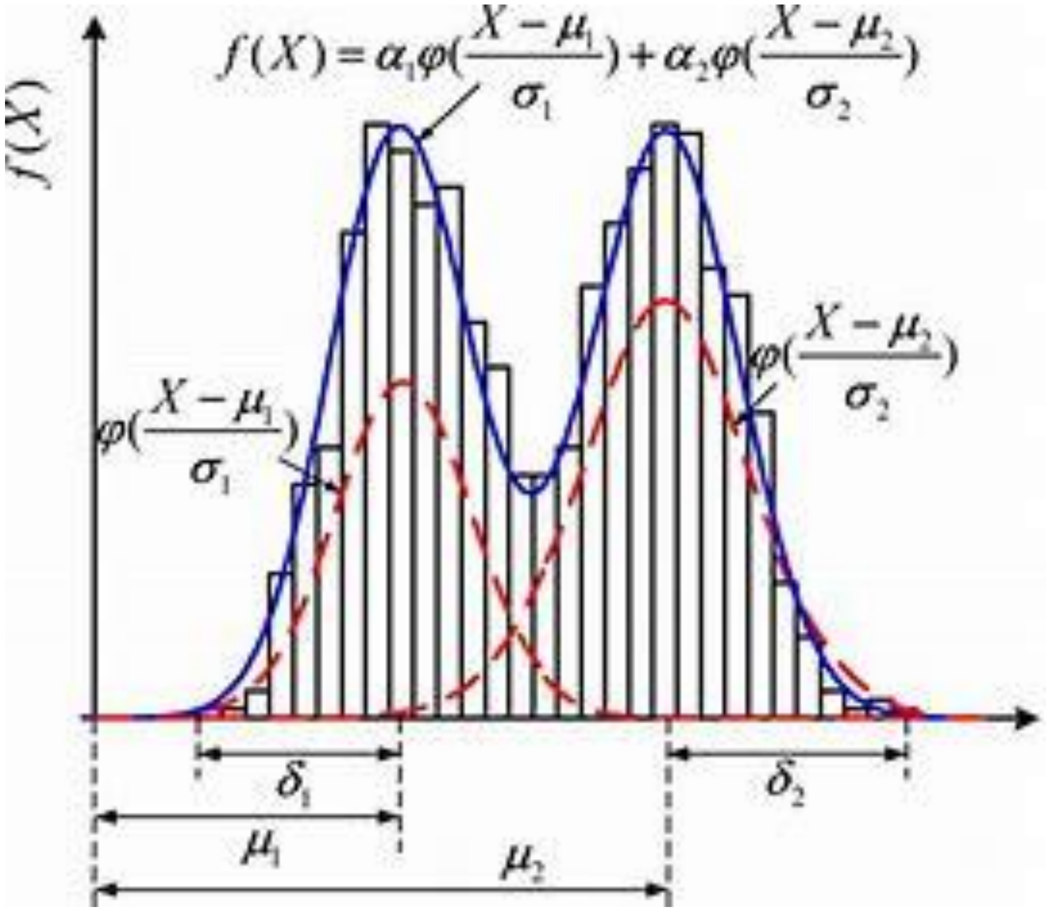
Mixture Distributions

For any kind of "large event" we will introduce the forward-looking annual probability M and the conditional distribution of loss L which has its own degree of uncertainty as to the severity.

- If we assume the probability of the negative event is M per year, then probability that conditions will remain stable, and the event will not occur is (1-M).

We now have two mutually exclusive states of the future, one in which conditions remain as they have been historically (the bad event has happened yet), and one where the large, negative event takes place. Each of these two states can be represented as a normal distribution with some expectation of the mean u_I and standard deviation s_i.

- From a small amount of algebra, we obtain the four moments of the ex-ante distribution. We can simply repeat the Cornish-Fisher exercise presented above to calculate a volatility equivalent *inclusive of the potential for a "large event" that has not yet taken place*, conditional on the probability and likely range of severity of such an event.



Robertson and Fryer (1969)

- To estimate the “volatility equivalent” appropriate to this situation, we will first combine the two distributions.

$$\mu = \Sigma_{i=1}^{2} m_i \mu_i$$

$$\sigma^2 = \Sigma_{i=1}^{2}[m_i\left(\sigma_i^2 + \mu_i^2\right) - \mu_i^2 \ ]$$

$$S = \frac{1}{\sigma^3}\{\Sigma_{i=1}^{2} m_i (\mu_i - \mu)[3\sigma_i^2 + (\mu_i - \mu)^2]\}$$

$$K = \frac{1}{\sigma^4}\{\Sigma_{i=1}^{2} m_i [3\sigma_i^4 + 6(\mu_i - \mu)^2\sigma_i^2 + (\mu_i - \mu)^4]\}$$

where

m_i = the probability of state *i* , μ = mean return of the combined distribution

σ = standard deviation of the combined distribution

μ_i = the expected return in state *i* , σ_i = the expected volatility in state *i*

S = skew of the combined distribution, K = kurtosis of the combined distribution (raw)

Conclusions

- The typical process of analyzing investment performance has always been rooted in capital market theories that make **false assumptions about normality and randomness.**

- Our usual metrics such as Sharpe Ratio, Information Ratio, and "risk adjusted returns" are not valid unless the underlying assumptions cannot be rejected as false.

- If the underlying assumptions do not hold, whether from non-normality or autocorrelation, we must modify our analytical framework to accommodate the practical realities rather than make decisions on distorted information.

- In this presentation, we have provided both statistical tests and tractable algebraic adjustments to allow investment performance to be fairly evaluated when the distribution of returns is serially correlated, non-normal, or is expected to be non-normal.

References

Osband, Kent, Iceberg Risk, Texere Publishing Limited, London, 2002.

Sharpe, William F., “Mutual Fund Performance”, *Journal of Business*, Volume 39-1, pp. 119-138, 1966.

Messmore, Tom, “Variance Drain”, *Journal of Portfolio Management*, Volume 21-4, pp. 104-110, 1995.

Rudd, Andrew and Henry Clasing, Modern Portfolio Theory, Dow-Jones Irwin, Homewood, Illinois, 1982.

diBartolomeo, Dan, “Measuring Investment Skill Using the Effective Information Coefficient”, *Journal of Performance Measurement*, Volume 13-1, pp. 45-52, 2008.

Levy, Haim, and Harry Markowitz, “Approximating Expected Utility a Function of Mean and Variance”, *American Economic Review*, Volume 69-3, pp. 308-317, 1979.

Litzenberger, Robert and Mark Rubinstein, “The Strong Case for the Generalized Logarithmic Utility as the Premier Model of Financial Markets”, *Journal of Finance*, Volume 31-2, pp. 551-571, 1976.

Wilcox, Jarrod, “Better Risk Management”, *Journal of Portfolio Management*, Volume 26-4, pp. 53-64, 2000.

References

Wilcox, Jarrod, “Harry Markowitz and the Discretionary Wealth Hypothesis”, *Journal of Portfolio Management*, Volume 29-3, pp. 58-65, 2003.

diBartolomeo, Dan, “Simplified Investment Performance Evaluation”, *Journal of Performance Measurement*, Volume 25-3, 2021.

Jobson, J.D. and Bob Korkie, “Performance Hypothesis Testing with the Sharpe and Treynor Measures”, Journal of Finance, Volume 36-4, pp. 889-908, 1981.

Memmel, Christoph, “Performance Hypothesis Testing with the Sharpe Ratio”, https://ssrn.com/abstract=412588

Lo, Andrew, “The Statistics of Sharpe Ratios”, *Financial Analyst Journal*, Volume 58-4, pp. 36-52, 2002.

Bertrand, Philippe and Costin Protopopescu, “The Statistics of Information Ratios”, *SSRN Electronic Journal*, 2010.

Ledoit, Olivier and Michael Wolf, “Robust Hypothesis Testing with the Sharpe Ratio”, *Journal of Empirical Finance*, Volume 15, pp. 850-859, 2008.

Geltner, David, “Smoothing in Appraisal-based Returns”, *Journal of Real Estate Finance and Economics*, Volume 4-3, pp. 327-345, 1991.

References

Lo, Andrew, Mila Getmansky and Igor Makarov, National Bureau of Economic Research, Working Paper 9571, March 2003.

Bolch, Ben, “More on Unbiased Estimation of the Standard Deviation”, *The American Statistician*, Volume 22-3, pp. 27, 1968.

Brugger, Richard, “A Note on Unbiased Estimation on the Standard Deviation”, *The American Statistician*, Volume 23-4, 1969.

Satchell, Steven and Anthony Hall, “The Anatomy of Portfolio Skewness and Kurtosis”, *Journal of Asset Management*, Volume 14, pp.228-235, 2013.

diBartolomeo, Dan and Christopher Kantos, “How the Pandemic Taught Us to Turn Smart Beta into Real Alpha, *Journal of Asset Management*, Volume 21-7, pp. 581-590, 2020.

diBartolomeo, Dan. “Applications of Portfolio Variety”, in J. Knight and S. Satchell Editors, Forecasting Volatility, Butterworth-Heinemann. 2007.

Perold, Andre, “Constant Proportion Portfolio Insurance”, Harvard Business School Working Paper, 1986.

References

Weisman, Andrew, "Informationless Investing and Hedge Fund Performance", *Journal of Portfolio Management*, Volume 28-4, pp. 80-91, 2002.

Bondarenko, Oleg, "Market Price of Variance Risk and Performance of Hedge Funds", *SSRN Electronic Journal*, March 2004.

Fung, William and David Hsieh, "Hedge Fund Benchmarks: A Risk Based Approach", Financial Analyst Journal, Volume 60-5, pp. 65-80, 2004.

Blackburn, Thomas, Dan diBartolomeo, and William Zieff, "Assessment of Cryptocurrency Risk for Institutional Investors", *Global Commodity Applied Research Digest*, Summer 2022.

Robertson, C.A. and J.G. Fryer, "Some Descriptive Properties of Normal Mixtures", *Scandinavian Actuarial Journal*, pp. 137-146, 1969.

Thode, Henry, <u>Testing for Normality</u>, CRC Press, 2002.

Jarque, Carlos M. and Anil K. Bera, "Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals", *Economics Letters,* Volume 6-3, pp. 255-259. 1980.

References

Cornish, E.A. and Ronald Fisher, “Moments and Cumulants in the Specification of Distributions”, *Review of the International Statistical Institute*, Volume 5-4, pp. 307-320. 1938.

Chernozhuckov, Victor, Ivan Fernadez-Val, and Alfred Galichon, “Rearranging Edgeworth-Cornish-Fisher Expansions”, *Economic Theory*, Volume 42-2, pp.419-435, 2010.

Martin, Douglas and Rohit Arora, “Inefficiency and Bias of Modified Value at Risk and Expected Shortfall”, *Journal of Risk*, Volume 19-6, pp. 59-84, 2017.

Liang, Bing, “Hedge Funds: The Living and the Dead”, Journal of Financial and Quantitative Analysis, Volume 35-3, pp. 309-326, 2000.

Lo, Andrew. “Risk Management for Hedge Funds: Introduction and Overview”, https://ssrn.com/abstract=283308, 2001.

Posthuma, Nolke and Pieter J. van der Sluis, “A Reality Check on Hedge Funds Returns”, https://ssrn.com/abstract=438840, 2003.