Exploring the Application of Kelly's Criterion in Portfolio Optimization

Introduction

When managing a portfolio, the fundamental problem is to find investments with excess risk-adjusted expected rates of return. Once the manager has identified some favorable opportunities and is ready to assemble the portfolio, they still need to decide how much capital, as a fraction of the value of assets under management, to allocate to each single trade. Assigning weights to each investment in a portfolio is what position sizing is about and the related decisions can significantly impact the long run returns of the portfolio. Empirical evidence shows that most traders tend to base their portfolio position sizing on heuristic rules, subject to behavioral psychological biases, that, obviously, do not mathematically guarantee performance maximization.

Mathematical Theory

With the goal of finding a mathematical theory able to determine the optimal position sizing for a portfolio composed of investments with positive risk-adjusted rates of return, we now introduce the Kelly Criterion, described for the first time by J. L. Kelly Jr in 1956. Originally invented to optimize wager sizes for gambling games that offered a positive expected value, the mathematics behind the Kelly Criterion have been extended to handle multiple correlated positive expected value bets, like stock portfolios. Nowadays, the Kelly Criterion has been implemented in many trading and investing strategies, to the point that even world-renowned investors such as Warren Buffet and Bill Gross reported to use the Kelly method in one of its many variations.

Before diving into portfolio management applications, it is worth briefly tackling the Criterion from a general angle, and to explain the mathematics behind it. For ease of understanding, we illustrate using a simple case, coin tossing, but the underlying concepts can be easily generalized.

Consider a coin, which guarantees, at each trial, a winning probability of p, where $p > \frac{1}{2}$, and therefore a losing probability of $q = 1 - p, q < \frac{1}{2}$. Also, the payoff in the event of winning is equal to +1, while in case of loss is equal to -1.

Since the game has a positive expected return, in order to maximize $E(X_n)$, the expected value of the wealth after n trials, we have to maximize $E(B_k)$, the expected value of the bet at the kth trial. Thus, to maximize expected gain the player should bet all their resources at each trial, which is extremely risky and almost guarantees ruin after few coin flips. At the same time, deciding the wager aiming at minimizing the ruin risk is an undesirable criterion as well, since it would require betting nothing at every trial, making no profit at all. This suggests that the optimal fraction of the bankroll to bet at each trial f, according to Kelly, is an intermediate value with respect to the two strategies described above, or rather a compromise between $f = 1$, which maximizes $E(X_n)$ and assures ruin, and $f = 0$, which minimizes the probability of ruin but does not allow the player to profit from the game.

After this consideration, we introduce Kelly's Criterion goal, which is to maximize the expected value of the natural logarithm of wealth, where the logarithmic function represents the value of money in terms of utility1.

If S is the number of successes and F the number of failures, then the capital after n trials is: $X_n = X_0(1+f)^S(1-f)^F$ where $S + F = n$ and $0 < f < 1$.

Note that since: $\dfrac{X_n}{X_0} = e^{n \, ln(\frac{X_n}{X_0})^{\frac{1}{n}}}$

Then: $G(f) = \ln\left(\dfrac{X_n}{X_0}\right)^{\frac{1}{n}} = \dfrac{S}{n}\ln(1+f) + \dfrac{F}{n}\ln(1-f)$ is the exponential growth rate per trial.

Turn now to the expected growth rate:

$$g(f) = E\left[\ln\left(\dfrac{X_n}{X_0}\right)^{\frac{1}{n}}\right] = E\left[\dfrac{S}{n}\ln(1+f) + \dfrac{F}{n}\ln(1-f)\right] = p\ln(1+f) + q\ln(1-f)$$

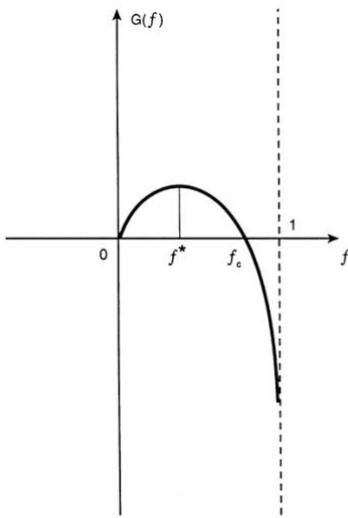As stated above, Kelly's Criterion consists in maximizing $E[\ln(X_n)]$, which, for some n fixed, being

$g(f) = \dfrac{1}{n}E[\ln(X_n) - \dfrac{1}{n}E[\ln(X_0)]$, is equivalent to maximize $g(f)$.

Consider now: $g'(f) = \dfrac{p}{1+f} - \dfrac{q}{1-f} = \dfrac{p-pf-q-qf}{(1+f)(1-f)} = \dfrac{p-q-f}{(1+f)(1-f)}$, setting it equal to 0, we find the stationary point $f^* = p - q$

Analyzing the function further, we find that, since the second derivative of g is always negative, f^* is the unique global maximizer for $g(f)$, whose corresponding maximum is $g(f^*) = p\ln(p) + q\ln(q) + \ln(2) > 0$. Moreover, the nature of the function (shown in the following graph) shows that there exists a unique $f_0 > 0$ such that $g(f_0) = 0$.

---

[1] 1 The nature of the logarithmic function optimally models the utility of money: the slowing pace of increase as x goes to +∞ well represents the decreasing marginal utility of the marginal dollar, while, as x goes to 0, it plummets rapidly to -∞, which symbolizes the disutility associated with going bankrupt.

The conclusion of the problem is that, as $n \to +\infty$, the player wealth unboundedly grows for every $0 < f < f_0$, but the optimal rate of growth is achieved as $f = f^*$ (Kelly strategy).

On the other hand, whenever $f > f_0$, ruin is almost certain[2]; while when $f = f_0$, player's wealth oscillates randomly between 0 and $+\infty$.

The simplest version of Kelly's formula, that we use in binary games like coin tossing, can easily be generalized to non-binary games, allowing for partial wins and partial losses. That is:

$$f^* = \frac{p}{a} - \frac{q}{b}$$

Where b is the fraction of capital lost in case of negative outcome and a is the fraction of capital gained in case of positive outcome.

Kelly Criterion for Portfolio Optimization

In order to implement the Kelly Criterion in the realm of portfolio optimization, one must consider a variable of the formula which takes into account continuous probability distributions. This is because for a financial asset there are an infinite number of outcomes to every possible bet that can be placed.

In this case, the formula that would be derived to obtain the optimal Kelly fraction/percentage would be:

$$f^* = \frac{\mu - r}{\sigma^2}$$

where μ is the mean return of the asset, r is the risk-free return, and σ is volatility.

A portfolio can be optimized under the Kelly Criterion in order to form a Kelly portfolio. A Kelly portfolio maximizes the expected return of any given combination of assets in the long run, by maximizing the geometric growth rate of the wealth, which can be expressed by:

$$g_\infty(f^*) = \frac{(\mu - r)^2}{2\sigma^2} + r$$

---

[2] Since f<1, it is technically impossible for the player to reach total ruin. Henceforth, here ruin's meaning is that for some $\varepsilon > 0$ arbitrary small: $\lim_{n \to +\infty} [p(X_n \leq \varepsilon)] = 1$

This means Kelly portfolio will have larger positions for bets which have a higher expected return.
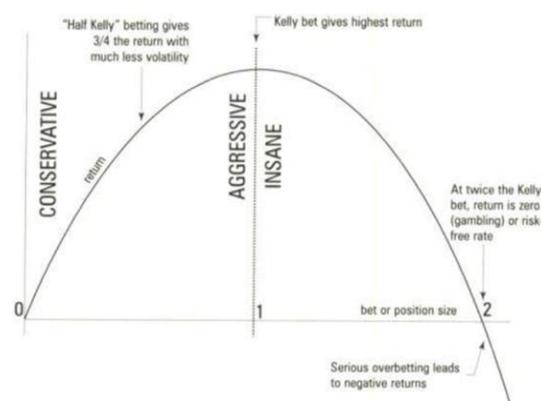
However, a Kelly portfolio has a number of limitations, which prevent it from being the clear-cut best optimization method in most cases. These have cause numerous debates on the effectiveness of its application and will be discussed in this paper.

Firstly, it should be noted that the first criterion to evaluate whether you should or should not use this method of optimization is determining your number one priority. In most cases, an institutional investor's number one priority is often maintaining a certain risk-level, not maximizing returns regardless of risk level, in order to avoid withdrawal of funds for example. A Kelly portfolio does not do this in any way, which reduces its utility at an institutional level. Also, Thorp (1971) proved that Kelly portfolios were not always mean-variance efficient, which again might discourage institutional investors from utilizing the criterion. This makes it a strategy more applicable to retail investors, who might be more willing to accept potential large short-term losses, knowing that they are maximizing wealth in the long-term.

Also, Kelly portfolios are often notoriously concentrated due to the nature of calculation in the criterion, thus making them riskier in this respect, than other optimization methods such as Markowitz, or tangent portfolios.

Furthermore, Samuelson (1971, 1979) showed that a Kelly portfolio is asymptotically optimal, but not optimal in finite time, as the high leverage or position size often leads to large drawdowns, which obviously hinder performance, by adding short-term risk. This is why Kelly portfolios work better in the long run. Another reason why Kelly portfolios work better in the long run is simply because of Bernoulli's Law of Large Numbers. This has been proven by Carta and Conversano (2020) with Monte Carlo simulations which show Kelly portfolios' performance support the properties of the Kelly Criterion much better at 40,000 trades than at 1,000 or 100 trades.

One way that these risks and disadvantages can be mitigated, is by implementing fractional Kelly as the criterion for the portfolio optimization. Fractional Kelly optimization is a variation of the traditional Kelly optimization, where the Kelly value obtained is divided, usually by 2 or by 3 in order to obtain more conservative bet sizes. These ensure your portfolio stays in the left (safe) side of the Kelly curve even if parameters are misestimated slightly, which minimizes your drawdowns, albeit reducing your expected returns. By reducing position sizes to fractions of Kelly positions, one can include more positions in a portfolio, and thus increase diversification.

Source: Seeking Alpha

These factors, in combination with the fact that the Kelly Criterion is so sensitive to parameter misestimation, are the reason why the Kelly Criterion should be seen as the upper bound of leverage to use in each position of your portfolio, as opposed to the target amount. If the Kelly Criterion is used as a target amount for leverage with every trade, ruin is a possibility.

This can be boiled down to the fact that the strategy returns are non-Gaussian. In other words, a normal distribution greatly underestimates the tail risk of a Kelly strategy, which again highlights why a Kelly strategy carries more short-term risk than what is usually anticipated. Tail risk can be modelled more effectively with a Generalized Pareto Distribution, or other fat-tail distributions.

Another potential issue with Kelly portfolios is that the stock market contains 'fluid odds', meaning that the parameters of the equation needed to calculate the Kelly Criterion vary every second of every trading day. Naturally, this leads us to the fact that a Kelly strategy works best with frequent rebalancing, in order to always be in the optimal position size or leverage window. The problem with constant rebalancing is that transaction costs end up eroding the returns the strategy yields. This is a problem which is often overlooked as most Monte Carlo simulations, and other methods used to back test Kelly strategies published online, assume no transaction-costs.

Back Test of Partial Kelly Portfolio and Results

In order to back test the strategy, we implemented a python script to back test the feasibility of Kelly's theory. The code's objective was to implement the Kelly criterion on a portfolio of all the components of the S&P500 and compare its performance to both the actual index as well as the SPXEW (the S&P500 index that just has equal weighting for all of its components). The formula was fed data from 2016 and the portfolio was tested on the period going from 2017 to the end of 2022.

To do this we first calculated the expected daily returns for each stock as well as their covariance based on their 2016 performance, moreover the risk-free rate was assumed to be 1.84% (the average yield of the 10-year treasury rate in 2016).

Based on the data we then found the Kelly weights for each component by optimizing the quadratic formula for the Kelly criterion (to do so we utilized a modified version of this (GitHub repository). After obtaining the weights we added further constraints by preventing the undertaking of any leverage, forcing the program to invest all its capital in the market and prohibiting the sale of any of the assets. We can call this strategy a partial Kelly. This results in one of the many variations which reduces potential returns but makes up for it by reducing risk more than proportionally. This allows for a strategy that mitigates the shortcomings of a full Kelly strategy, like the fractional Kelly explored earlier in the paper.

Since the trade is always open, there was no need to model turnover and transaction costs are negligible.

As we can see from the report (generate using quantstats), the long-only partial Kelly achieved a Compounded Annual Growth Rate of 17.4%, generating an alpha with respect to the SPX of more than 6.24% and 7.09% with respect to the SPXEW . It also boasted a more attractive Sharpe ratio (0.84) as well as a much shorter duration of

Drawdown days (at 275 days compared to 360 of the SPX and 359 of the SPXEW), however this comes at the cost of higher volatility and a bigger maximum drawdown (-41.65% compared to the SPX's -33.72%).

We expected the performance of the partial Kelly S&P500 to increase using a rolling strategy, possibly updating the sizing of each stock every quarter or even every few days in such a way to avoid having the index deviate too much from the optimal position sizing, however this would come at the cost of much higher transaction costs.

Analysis of Strategy vs SPXEW

Cumulative Returns vs Benchmark

Key Performance Metrics

Metric	Strategy	SPXEW
Risk-Free Rate	0.0%	0.0%
Time in Market	100.0%	100.0%
Cumulative Return	161.49%	79.99%
CAGR﹪	17.4%	10.31%
Sharpe	0.84	0.58
Prob. Sharpe Ratio	97.66%	91.84%
Smart Sharpe	0.75	0.52
Sortino	1.16	0.8
Smart Sortino	1.04	0.72
Sortino/√2	0.82	0.57
Smart Sortino/√2	0.73	0.51
Omega	1.18	1.18
Max Drawdown	-41.65%	-39.04%
Longest DD Days	275	359
Volatility (ann.)	22.17%	20.75%
R^2	0.95	0.95
Information Ratio	0.09	0.09
Calmar	0.42	0.26
Skew	-0.86	-0.58
Kurtosis	20.3	14.89

Worst 5 Drawdown Periods

Analysis of Strategy vs SPX

Cumulative Returns vs Benchmark

Key Performance Metrics

Metric	Strategy	SPX
Risk-Free Rate	0.0%	0.0%
Time in Market	100.0%	100.0%
Cumulative Return	161.49%	88.52%
CAGR%	17.4%	11.16%
Sharpe	0.84	0.63
Prob. Sharpe Ratio	97.66%	93.67%
Smart Sharpe	0.75	0.56
Sortino	1.16	0.87
Smart Sortino	1.04	0.78
Sortino/√2	0.82	0.62
Smart Sortino/√2	0.73	0.55
Omega	1.18	1.18
Max Drawdown	-41.65%	-33.72%
Longest DD Days	275	360
Volatility (ann.)	22.17%	19.86%
R^2	0.87	0.87
Information Ratio	0.05	0.05
Calmar	0.42	0.33
Skew	-0.86	-0.58
Kurtosis	20.3	12.59

Worst 5 Drawdown Periods