Using Kelly Capital Growth Strategy

Dr. William T. Ziemba
Alumni Professor
Sauder School of Business
University of British Columbia

Introduction to the Kelly Capital Growth Criterion and Samuelson’s Objections to it

The Kelly capital growth criterion, which maximizes the expected log of final wealth, provides the strategy that maximizes long run wealth growth asymptotically for repeated investments over time. However, one drawback is found in its very risky behavior due to the log's essentially zero risk aversion; consequently it tends to suggest large concentrated investments or bets that can lead to high volatility in the short-term. Many investors, hedge funds, and sports bettors use the criterion and its seminal application is to a long sequence of favorable investment situations.

Edward Thorp was the first person to employ the Kelly Criterion, or “Fortune’s Formula” as he called it, to the game of blackjack. He outlines the process in his 1960 book Beat the Dealer and his findings changed the way this game was played once he had demonstrated that there was a winning strategy. As applied to finance, a number of note-worthy investors use Kelly strategies in various forms, including Jim Simons of the Renaissance Medallion hedge fund.

The purpose of this paper is to explain the Kelly criterion approach to investing through theory and actual investment practice. The approach is normative and relies on the optimality properties of Kelly investing. There are, of course, other approaches to stock and dynamic investing. Besides mean-variance and its extensions there are several important dynamic theories. Many of these are surveyed in MacLean and Ziemba (2013). An interesting continuous time theory based on descriptive rather than normative concepts with arbitrage and other applications is the stochastic portfolio theory of Fernholz and colleagues, see for example, Fernholz and Shay (1982), Fernholz (2002), and Karatzas and Fernholz (2008). They consider the long run performance of portfolios using specific distributions of returns such as lognormal. The Kelly approach uses a specific
utility function, namely log, with general asset distributions.

**What is the Kelly Strategy and what are its main properties?**

Until Daniel Bernoulli’s 1738 paper, the linear utility of wealth was used, so the value in ducats would equal the number of ducats one had. Bernoulli postulated that the additional value was less and less as wealth increased and was, in fact, proportional to the reciprocal of wealth so,

$$u'(w) = 1/w \quad \text{or} \quad u(w) = \log(w)$$

where \(u\) is the utility function of wealth \(w\), and primes denote differentiation. Thus concave log utility was invented.

In the theory of optimal investment over time, it is not quadratic (one of the utility function behind the Sharpe ratio), but log that yields the most long-term growth asymptotically. Following with an assessment of that aspect, the Arrow-Pratt risk aversion index for \(\log(w)\) is:

$$R^*_t(w) = -\frac{u''}{u'} = \frac{1}{w}$$

which is essentially zero. Hence, in the short run, log can be an exceedingly risky utility function with wide swings in wealth values.

John Kelly (1956) working at Bell Labs with information theorist Claude Shannon showed that for Bernoulli trials, that is win or lose 1 with probabilities \(p\) and \(q\) for \(p+q=1\), the long run growth rate, \(G\), namely

$$G = \lim_{t \to \infty} \log \left( \frac{w_t}{w_0} \right)^{1/t}$$

where \(t\) is discrete time and \(w_t\) is the wealth at time \(t\) with \(w_0\) the initial wealth is equivalent to \(\max E \{ \log w \}\)

Since \(w_t = (1+f)^t \ \ (1-f)^{t-M}\) is the wealth after \(t\) discrete periods, \(f\) is the fraction of wealth bet in each period and \(M\) of the \(t\) trials are winners.

Then, substituting \(W_t\) into \(G\) gives

$$G = \lim_{t \to \infty} \left[ \frac{M}{t} \log(1+f) + \frac{t-M}{t} \log(1-f) \right] + p \log(1+f) + q \log(1-f)$$

and by the strong law of large numbers

$$G = E \left[ \log w \right]$$

Thus the criterion of maximizing the long run exponential rate of asset growth is equivalent to maximizing the one period expected logarithm of wealth. So an optimal policy is myopic in the sense that the optimal investments do not depend on the past or the future. Since

$$\max G(f) = p \log(1+f) + q \log(1-f)$$

the optimal fraction to bet is the edge \(f^* = p - q\). The edge is the expected value for a bet of one less the one bet. These bets can be large. For example, if \(p=0.99\) and \(q=0.01\), then \(f^* = 0.98\), that is 98% of one’s fortune. Some real examples of very large and very small bets appear later in the paper. If the payoff odds are \(+B\) for a win and -1 for a loss, then the edge is \(Bp - q\) and

$$f^* = \frac{Bp - q}{B} = \frac{\text{edge}}{\text{odds}}$$

So the size of the investments depend more on the odds, that is to say, the probability of losing, rather than the mean advantage. Kelly bets are usually large and the more attractive the wager, the larger the bet. For example, in the trading on the January turn-of-the-year effect with a huge advantage, full Kelly bets approach 75% of initial wealth. Hence, Clark and Ziemba (1988) suggested a 25% fractional Kelly strategy for their trades, as discussed later in this article.

Latane (1959, 1978) introduced log utility as an investment criterion to the finance world independent of Kelly’s work. Focusing, like Kelly, on simple intuitive versions of the expected log criteria, he suggested that it had superior long run properties. Chopra and Ziemba (1993) have shown that in standard mean-variance investment models, accurate mean estimates are about twenty times more important than covariance estimates and ten times variances estimates in certainty equivalent value. But this is risk aversion dependent with the importance of the errors becoming larger for low risk aversion utility functions. Hence, for \(\log w\) with minimal risk aversion, the impact of these errors is of the order 100:3:1. So bettors who use \(E \log\) to make decisions can easily over bet.

Leo Breiman (1961), following his earlier intuitive paper Breiman (1960), established the basic mathematical properties of the expected log criterion in a rigorous fashion. He proved three basic asymptotic results in a general discrete time setting with intertemporally independent assets.

Suppose in each period, \(N\), there are \(K\) investment opportunities with returns per unit investe \(X_{1,M}\). Let \(\Lambda = (\Lambda_1, \ldots, \Lambda_K)\) be the fraction of wealth invested in each asset. The wealth at the end of period \(N\) is

$$w_N = \left( \sum_{i=1}^{K} \Lambda_i X_i \right) w_{N-1}.$$  

In each time period, two portfolio managers have the same family of investment opportunities, \(X\), and one uses a \(\Lambda\) which maximizes \(E \log w_N\), whereas the other uses an essentially different strategy, \(\Lambda^*\), so they differ infinitely often, that is,

$$E \log w_N \Lambda^* - E \log w_N (\Lambda) \to \infty.$$  

Then

$$\lim_{N \to \infty} \frac{w_N (\Lambda^*)}{w_N (\Lambda)} \to \infty$$  

So the wealth exceeds that with any other strategy by more and more as the horizon becomes more distant. This generalizes the Kelly Bernoulli trial setting to intertemporally independent and stationary returns.

The expected time to reach a preassigned goal \(A\) is asymptotically least as \(A\) increases with a strategy maximizing \(E \log w_N\). Assuming a fixed opportunity set, there is a fixed fraction strategy that maximizes \(E \log w_N\), which is independent of \(N\).
Consider the example described in Exhibit 1. There are five possible investments and if we bet on any of them, we always have a 14% advantage. The difference between them is that some have a higher chance of winning than others. For the latter, we receive higher odds if we win than for the former. But we always receive 1.14 for each 1 bet on average. Hence we have a favorable game. The optimal expected log utility bet with one asset (here we either win or lose the bet) equals the edge divided by the odds. So for the 1-1 odds bet, the wager is 14% of one's fortune and at 5-1 its only 2.8%. We bet more when the chance that we will lose our bet is smaller. Also, we bet more when the edge is higher. The bet is linear in the edge so doubling the edge doubles the optimal bet. However, the bet is non-linear in the chance of losing our money, which is reinvested so the size of the wager depends more on the chance of losing and less on the edge.

The simulation results shown in Exhibit 2 assume that the investor’s initial wealth is $1,000 and that there are 700 investment decision points. The simulation was repeated 1,000 times. The numbers in Exhibit 2 are the number of times out of the possible 1,000 that each particular goal was reached. The first line is with log or Kelly betting. The second line is half Kelly betting. That is, you compute the optimal Kelly wager but then blend it 50-50 with cash. For lognormal investments $\alpha$-fractional Kelly wagers are equivalent to the optimal bet obtained from using the concave risk averse, negative power utility function, $-w^{\beta}$, where $\alpha = \frac{1}{1 + \beta}$. For non lognormal assets this is an approximation (see MacLean, Ziemba and Li, 2005 and Thorp, 2010, 2011). For half Kelly ($\alpha=1/2$), $\beta=-1$ and the utility function is $-w^{-1} = -1/w$. Here the marginal increase in wealth drops off as $w^2$, which is more conservative than log’s $w$. Log utility is the case $\beta \rightarrow -\infty$, $\alpha = 1$ and cash is $\beta \rightarrow -\infty$, $\alpha = 0$.

A major advantage of full Kelly log utility betting is the 166 in the last column. In fully 16.6% of the 1,000 cases in the simulation, the final wealth is more than 100 times as much as the initial wealth. Also in 302 cases, the final wealth is more than 50 times the initial wealth. This huge growth in final wealth for log is not shared by the half Kelly strategies, which have only 1 and 30, respectively, for these 50 and 100 times growth levels. Indeed, log provides an enormous growth rate but at a price, namely a very high volatility of wealth levels. That is, the final wealth is very likely to be higher than with other strategies, but the wealth path will likely be very bumpy. The maximum, mean, and median statistics in Exhibit 2 illustrate the enormous gains that log utility strategies usually provide.

Let us now focus on bad outcomes. The first column provides the following remarkable fact: one can make 700 independent bets of which the chance of winning each one is at least 19% and usually is much more, having a 14% advantage on each bet and still turn $1,000 into $18, a loss of more than 98%. Even with half Kelly, the minimum return over the 1,000 simulations was $145, a loss of 85.5%. Half Kelly has a 99% chance of not losing more than half the wealth versus only 91.6% for Kelly. The chance of not being ahead is almost three times as large for full versus half Kelly. Hence to protect ourselves from bad scenario outcomes, we need to lower our bets and diversify across many independent investments.

Exhibit 3 shows the highest and lowest final wealth trajectories for full, $\frac{1}{3}$, $\frac{1}{2}$, $\frac{3}{4}$ and $\frac{2}{3}$ Kelly strategies for this example. Most of the gain is in the final 100 of the 700 decision points. Even with these maximum graphs, there is much volatility in the final wealth with the amount of volatility generally higher with higher Kelly fractions. Indeed with $\frac{2}{3}$ Kelly, there were losses from about decision points 610 to 670.

The final wealth levels are much higher on average, the higher the Kelly fraction. As you approach full Kelly, the typical final wealth escalates dramatically. This is shown also in the maximum wealth levels in Exhibit 4.
There is a chance of loss (final wealth is less than the initial $1,000) in all cases, even with 700 independent bets each with an edge of 14%.

If capital is infinitely divisible and there is no leveraging, then the Kelly bettor cannot go bankrupt since one never bets everything (unless the probability of losing anything at all is zero and the probability of winning is positive). If capital is discrete, then presumably Kelly bets are rounded down to avoid overbetting, in which case, at least one unit is never bet. Hence, the worst case with Kelly is to be reduced to one unit, at which point betting stops. Since fractional Kelly bets less, the result follows for all such strategies. For levered wagers, that is, betting more than one’s wealth with borrowed money, the investor can lose much more than their initial wealth and become bankrupt.

Selected Applications

In this section, I focus on various applications of Kelly investing starting with an application of mine. This involves trading the turn-of-the-year effect using futures in the stock market. The first paper on that was Clark and Ziemba (1988) and because of the huge advantage at the time suggested a large full Kelly wager approaching 75% of initial wealth. However, there are risks, transaction costs, margin requirements, and other uncertainties which suggested a lower wager of 25% Kelly. They traded successfully for the years 1982/83 to 1986/87 - the first four years of futures in the TOY; futures in the S&P500 having just begun at that time. I then continued this trade of long small cap minus short large cap measured by the Value Line small cap index and the large cap S&P500 index for ten more years with gains each
Using Kelly Capital Growth Strategy

The plots and tables describing these trades for these 14 years from 1982/83 to 1995/ are in Ziemba (2012).

Exhibit 5 has graphs of investing with the author’s money successfully in December 2009, 2010, and 2011, where the dots are the entries and the squares are the exits. The size of the position is 15% fractional Kelly. The profit on these trades can be seen in the three December periods in the graph. The January effect still exists in the futures markets, but now is totally in December contrary to the statements in most finance books such as Malkiel (2011). The fractional Kelly wager suggested in the much more dangerous market situation now is low. Programmed trading, high frequency trading and other factors add to the complexity, so risk must be lowered as one sees in the volatile 2011/12 graph.

These turn of the year bets are large, however, the Kelly wagers can be very small even with a large edge if the probability of winning is low. An example is betting on unpopular numbers in lotto games. MacLean, Ziemba, and Blazenko (1992) show that with an 82.7% edge, the full Kelly wager is only 65 $1 tickets per $10 million of one’s fortune. This is because most of the edge is in very low probability of winning the Jackpot and second prize. While there is a substantial edge, the chance of winning a substantial amount is small and indeed to have a high probability of a large gain requires a very long time, in the millions of years.

Kelly investing has several characteristics. It is not diversified but instead places large bets on the few very best assets. Hence, given the large bets, the portfolio can have considerable monthly losses. But the long run growth of wealth is usually high.

So, How Much Should You Bet?

The optimal Kelly bet is 97.5% of wealth and half Kelly is 38.75%. Obviously an investor might choose to go lower, to 10%, for example. While this seems quite conservative, other investment opportunities, miscalculation of probabilities, risk tolerance, possible short run losses, bad scenario Black Swan events, price pressures, buying in and exiting sometimes suggest that a bet much lower than 97.5% would be appropriate. Of course there are also many ways to blow up; see Ziemba and Ziemba (2013) for discussions of several hedge fund disasters, including Long Term Capital Management, Amaranth, and Societe Generale.

However, impressive gains are possible with careful risk management. During an interview in the Wall Street Journal (March 22-23, 2008) Bill Gross and Ed Thorp discussed turbulence in the markets, hedge funds, and risk management. Bill noted that after he read Ed’s classic *Beat the Dealer* in 1966, he ventured to Las Vegas to see if he could also beat blackjack. Just as Ed had done earlier, he sized his bets in proportion to his advantage, following the Kelly Criterion as described in the book, and he ran his $200 bankroll up to $10,000 over the summer. Bill ultimately wound up managing risk for Pacific Investment Management Company’s (PIMCO) investment pool of almost $1 trillion and stated that he was still applying lessons he had learned from the Kelly Criterion: “Here at PIMCO it doesn’t matter how much you have, whether it’s $200 or $1 trillion. Professional blackjack is being played in this trading room from the standpoint of risk management and that is a big part of our success.”
Conclusions
The Kelly capital growth strategy has been used successfully by many investors and speculators during the past fifty years. In this article I have described its main advantages, namely its superiority in producing long run maximum wealth from a sequence of favorable investments. The seminal application is to an investment situation that has many repeated similar bets over a long time horizon. In all cases one must have a winning system that is one with a positive expectation. Then the Kelly and fractional Kelly strategies (those with less long run growth but more security) provide a superior bet sizing strategy. The mathematical properties prove maximum asymptotic long run growth. But in the short term there can be high volatility.

However, the basic criticisms of the Kelly approach are largely concerned with over betting, the major culprit of hedge fund and bank trading disasters. Fractional Kelly strategies reduce the risk from large positions but then usually end up with lower final wealth. If properly used, the Kelly strategy can provide a superior long-term wealth maximizing technique.

Note
This article is a short version of a longer article entitled, "Understanding Using The Kelly Capital Growth Investment Strategy"

References


Author’s Bio

Dr William T. Ziemba
Alumni Professor
Sauder School of Business
University of British Columbia

Distinguished Visiting Associate
Systemic Research Centre London
School of Economics