Kelly criterion for trade optimization and risk management

EPAT Final Project

Hiroshi Nomura (Dec 2021)

Project Abstract

Firstly, we briefly discuss an overview of the Kelly criterion, with illustrative examples using the coin toss and the more complex dice roll bet. Next, we will apply this to real life markets using 15 year historical data of the SPY ETF and back-test with various strategies and scenarios. We highlight our findings both positive and negative, and how to further optimize the strategy so that it can be applied in practice. Lastly, we will discuss the limitations and how we can make best use of the Kelly criterion in practice as an upper limit for sizing/leverage for our trades. Suggestions for further investigation are also made for future projects.

Introduction

It is widely accepted in trading, that finding *positive expected value* (EV) trades and finding your *edge* is very important. However, this is only half the challenge as how you size the trade from initiation all the way to exit (hopefully with a profit!) is key and arguably even more difficult.

Many seem to focus on trade ideas and overlook the importance of position sizing. Firstly, one must ask herself, how much of my total capital should I deploy for each trade. Many beginners make the mistake of betting too large, while many experienced traders make the mistake of not betting big enough when they should on higher conviction trades.

As Charlie Munger once said; *"The wise ones bet heavily when the world offers them that opportunity. They bet big when they have the odds. And the rest of the time, they don't. It's just that simple."*

Talking from personal experience, I was aware of the Kelly criterion but I never actually applied it to my trading during my time as a professional trader/market-maker as the main task was to accommodate hundreds of customer flows and either offload the risk safely or warehouse the risk. Therefore, I assumed that the multivariate nature would not be suited or would not be applicable to what I understood of the Kelly criterion and that it considers only univariate cases which is typically the case in gambling. It is very interesting to investigate in this project if I can apply a method that can utilize Kelly criterion which can be effectively used to optimize portfolio performance.

So, what is the Kelly criterion? Well to briefly explain, the Kelly criterion was developed by John Kelly, who at the time worked as a scientist at AT&T Bell labs. What makes it so fascinating is that the Kelly criterion can tell us what is the optimal bet size while maximizing the *expected growth rate* and the *median of the terminal wealth*. It is important to take a moment to realize the significance of achieving the maximum median and not the mean, especially if you think about situations where the distributions of returns are highly skewed.

Without going into the mathematical formula yet, to help understand intuitively, the Kelly criterion can simply be defined as:

Edge / Odds = Fraction of capital that should be allocated

Where:

- The *edge* is calculated by the total expected value (EV), obtained by adding up the multiplication of each scenario's possible outcome by its corresponding probability.
- The *odds* are directly the positive outcome that can be obtained.

To understand this intuitively, it is important to acknowledge the key difference between *probability* and *expected value*. "Probability" is simply a representation of the chance that a given outcome will happen, whereas "expected value" represents the average outcome of a series of random events with identical odds being repeated over a long period of time.

For instance, say you flip a coin 10 times in a bet where you wager $1 per flip, and if you get heads, I'll double your bet; tails you lose your $1. The EV of this bet should tell you make (2 + -1) / 2 = $0.50 per flip (*see python coded simulation*). This can also be referred to as the *arithmetic EV* or *arithmetic mean*. However, what we are more interested in is the *geometric EV*, which is the *expected return on our capital per bet*. The higher this number is, the faster our wealth will compound and the greater our final capital ("*terminal value*") will be. We'll refer to this as the *geometric holding period return* or *GHPR*.

So how much should you allocate to this bet? Well you know that your EV or “edge” is 0.50, and your “odds” are 2, so the fraction of capital that should be allocated to each bet is 0.50/2 = 0.25 or 25%.

Alternatively in a more mathematical form, the Kelly criterion can also be expressed as:

$$K\% = W - \frac{(1 - W)}{R}$$

where:

$K\%$ = The Kelly percentage

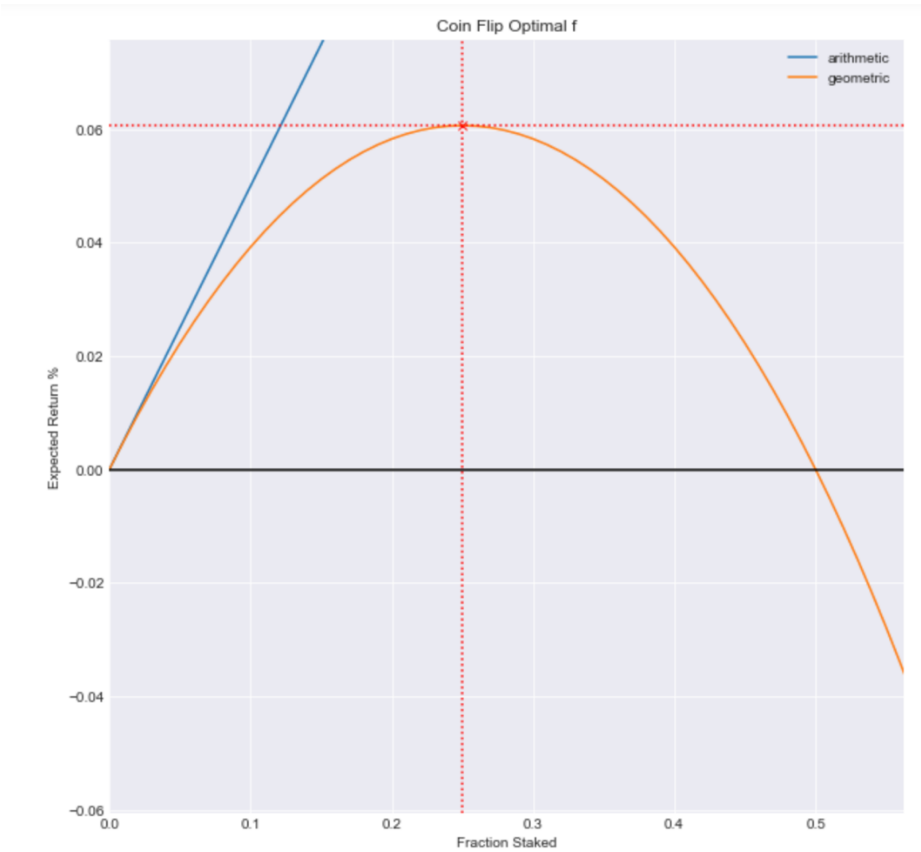
W = Winning probability

R = Win/loss ratio

Now going back to arithmetic and geometric EV, let’s see what happens if we bet 100% of our bankroll on each coin flip. As you might have guessed, if we play the game long enough, eventually the coin will come up tails and you end up losing everything. So even though the arithmetic EV is 50%, the geometric EV is -100%.

If we iterate through all possible fractions, you will find that the GHPR peaks at 6.06% under 25% capital and deteriorates as we further increase the bet size until GHPR 0% at 50% capital and negative thereafter. So betting big to win big is only true up until a certain point!

This approach will generate a curve, and the peak of this curve represents the optimal fraction to bet. Ralph Vince has termed this value the *optimal f*.



Optimal f: 0.25
Expected Value (Arithmetic): 12.5%
Expected Value (Geometric): 6.066%

You will notice that this method agrees with the optimal fraction of 25% as the Kelly criterion formula gave us. So why don't we just use the Kelly formula every time?

Let us now experiment with a scenario where possible outcomes are non-binary, like a dice roll. Again, we will skew the EV in your favour, where if an *even number* comes up, you *win that multiple* of your bet. If an *odd number* comes up, you *lose that multiple*.

Also note that in this example, we will need to *bound the returns* to avoid bankruptcy. The reason for this is that the maximum you could possibly bet is 1/5 of your capital, since you will lose five times your bet if you roll a five.

Now if we compute the Kelly fraction using the formula, we get a value of 2.5% whereas, if we plot the chart to get the optimal f, we get 3.4%. In other words, the Kelly formula underestimates the optimal bet size!

Next, let's apply this to real life financial markets situations which are more complex than the simple gambling examples we have discussed so far.

Data Mining

Data for this project was collected from *Yahoo Finance,* and we will use the adjusted closing prices for the SPY ETF which tracks the S&P500 and is the world's largest, most liquid ETF. The total sample data set spans a 15 year period from 28 November 2006 to 17 November 2021.

Using the SPY data, this time we see an even bigger discrepancy between the two methods. The Kelly formula gives a fraction of 57.7% whereas optimal f gives 292.4%, a huge difference! Also the optimal f is suggesting that we should bet more capital than we have and employ leverage, either by borrowing money or via use of futures/options. To maintain a constant leverage ratio, we would have to rebalance every day, which in reality would incur high transaction costs but we will choose to ignore it here for simplification.

Clearly we need a different modified Kelly formula which can be applied to the financial markets. Assuming our trading asset follows normally distributed returns and our funding rate for borrowing/leverage is the risk-free rate (we will use 3 month US T-bills), then the Kelly criterion formula takes on the following form (also referred to as the *"Kelly Capital Growth Investment Criterion"*):

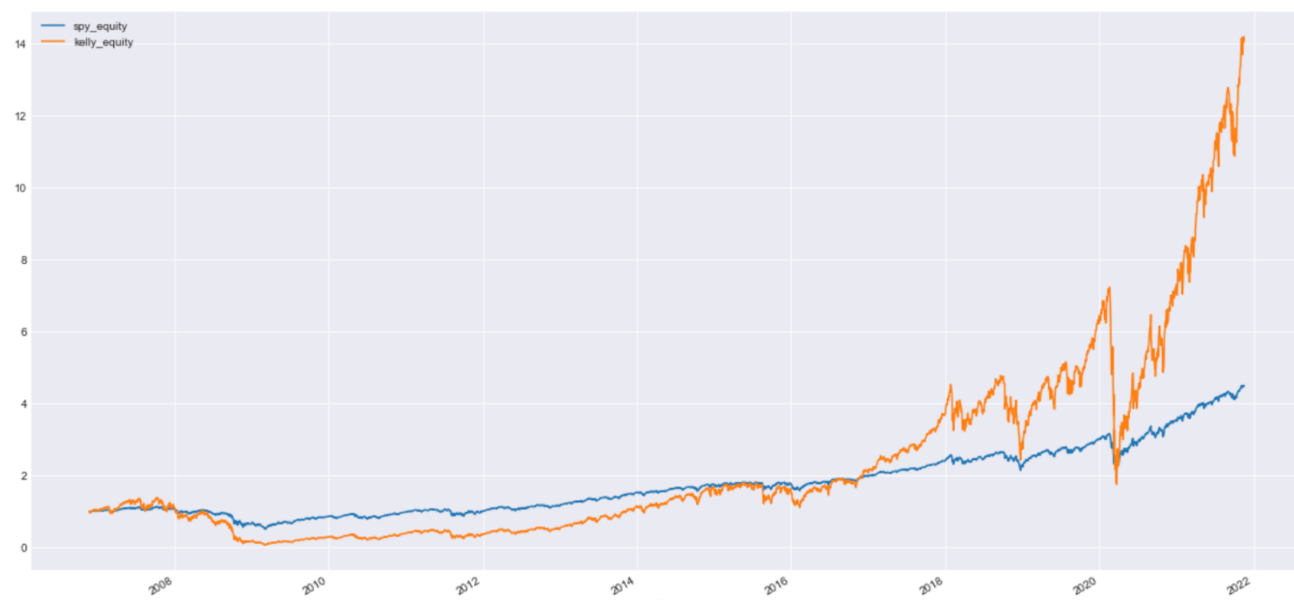
$$f^* = \frac{\mu - r}{\sigma^2}$$

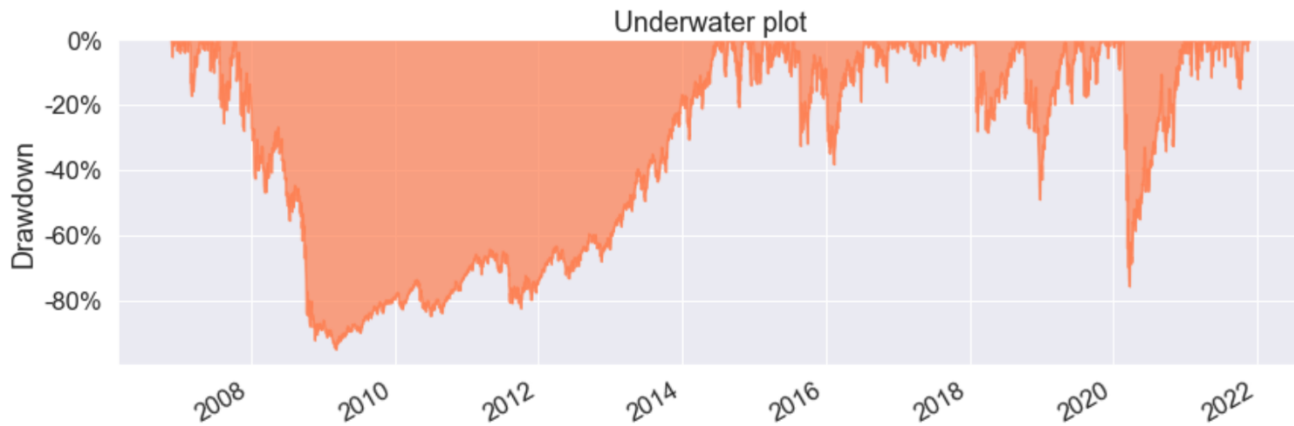
Where: Mu (μ) = average returns r = risk free rate sigma² (σ^2) = variance

Testing if our new formula works, using the SPY example, it returns an optimal f value of 297% which is very close to what we got earlier. However, it is important to note that technically, the Kelly capital growth investment formula assumes normally distributed (*Gaussian*) returns, and we know that returns on financial assets are notoriously non-normal!

When we back-test SPY with Kelly leverage, we can observe an impressive annualized return of 19% vs. 10.5% for buy-and-hold with no leverage. However, the issue here is a frightening annual volatility of 60% and not to mention an insane max drawdown of -95%!

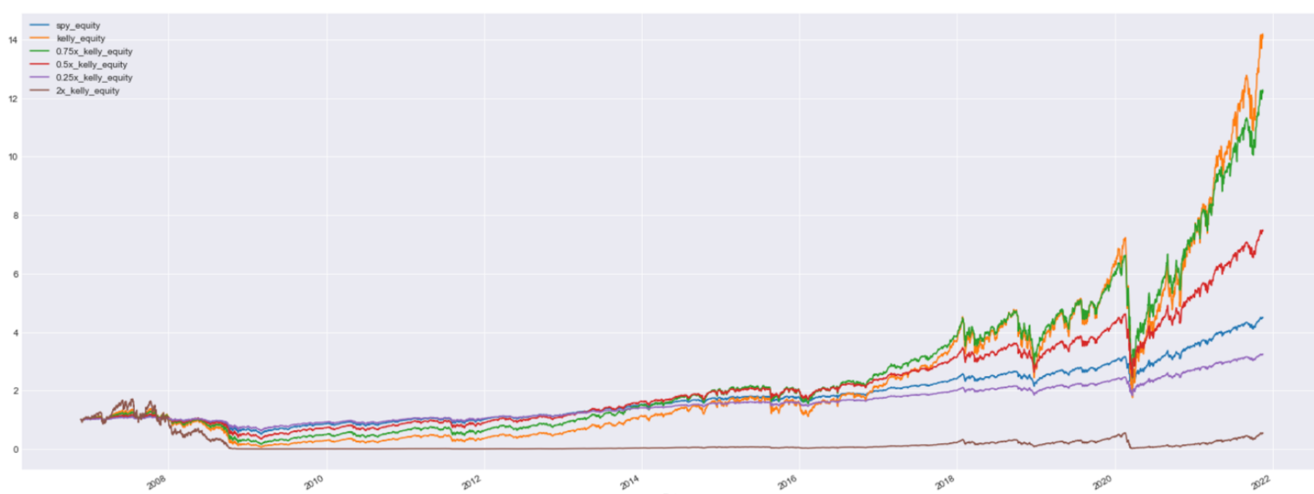
Start date	2006-11-21
End date	2021-11-17
Total months	179
	Backtest
Annual return	19.3%
Cumulative returns	1308.6%
Annual volatility	59.9%
Sharpe ratio	0.60
Calmar ratio	0.20
Stability	0.71
Max drawdown	-95.0%
Omega ratio	1.13
Sortino ratio	0.84
Skew	NaN
Kurtosis	NaN
Tail ratio	0.88
Daily value at risk	-7.4%





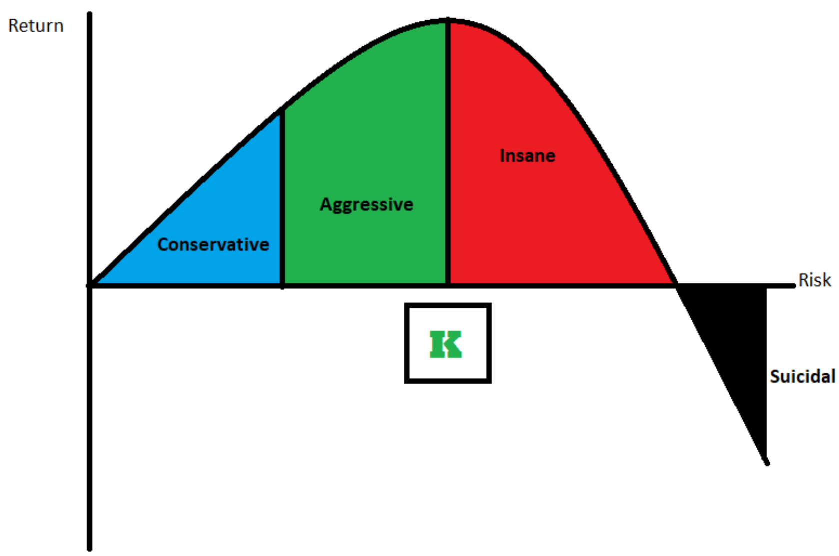
Also this “back-test” is totally cheating as it suffers from *look-ahead bias*. Not to mention I have not accounted for *transaction costs*, which would be significant with daily rebalancing.

If you were like most investors, you cannot possibly stomach the horrific volatility and drawdowns which come with a Kelly-sized leverage strategy. Institutional investors would almost certainly have a mandated maximum drawdown (let’s say 15-20%). That is why in practice, it makes much more sense to size trades at a fraction of the Kelly (“*Fractional-Kelly*”).



From this chart we can see that there is clearly a trade-off between leverage/high performance and higher risks/volatility. With fractions below the full Kelly, the GHPR drops rapidly and the gradient of the curve flattens aggressively as we see a significant reduction in the standard deviation. When reduced to ¼ Kelly the equity curve drops below the zero-leverage buy-and-hold curve. On the other hand, when we exceed the full Kelly, such as the 2x Kelly, the collapse in the GHPR to zero occurs almost immediately and fails to recover throughout the sample period.

This explains why in practice, many traders who employ Kelly Criterion often tend to apply the ½ Kelly strategy. As a quick and dirty rule, if you bet ½ the Kelly, you get about ¾ of the return with ½ the volatility. Not bad! The full Kelly should only be considered as an upper bound of leverage where when exceeded, could easily lead to ruin due to the non-Gaussian distribution of returns.



Source: https://rhsfinancial.com/2017/06/20/line-aggressive-crazy-leverage/

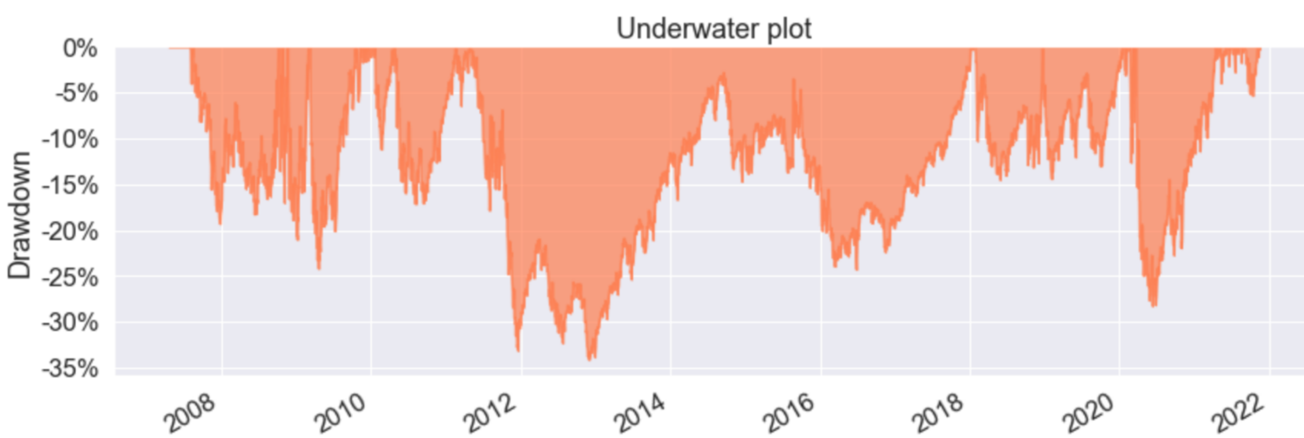
Taking this further, we will investigate the impact of Kelly and fractional-Kelly sizing on GHPR for 3 popular strategies: a simple momentum strategy, a momentum strategy using exponential moving averages (EMAs) and, a simple mean reversion strategy using Bollinger Bands. The same 15 years of historical data for SPY ETF was used.

For the simple momentum strategy, following a quick parameter optimization, we use a 10 day vs 100 day simple moving average crossover.


checking if positions are generated properly
ma_short
ma_long
position (right)
400
300
200
100
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
2008
2010
2012
2014
2016
2018
2020
2022
Date


Buy & hold' vs 'Momentum strategy' cumulative returns
bnh_returns
strategy_returns
1.5
1.0
0.5
0.0
−0.5
2008
2010
2012
2014
2016
2018
2020
2022
Date

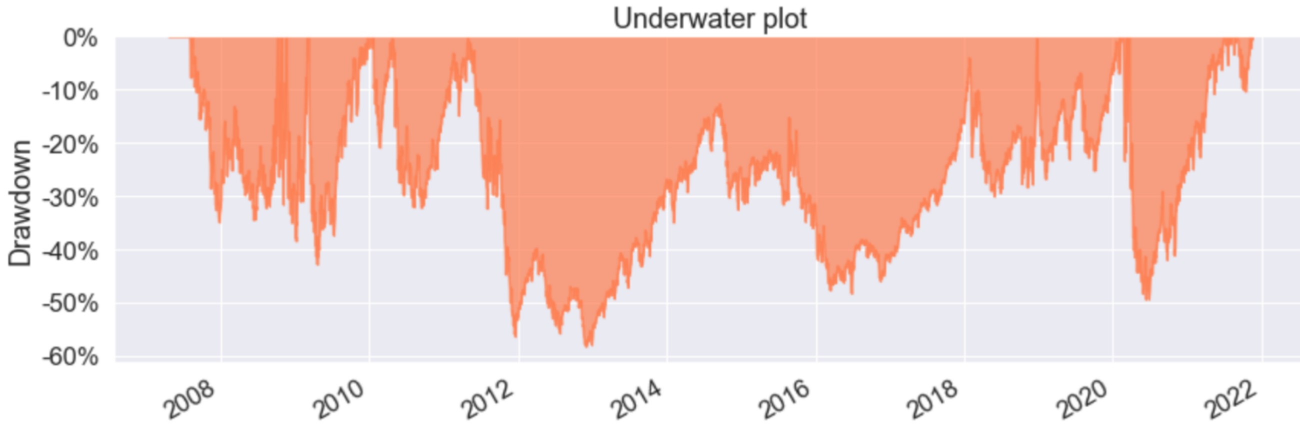
Start date	2007-04-19
End date	2021-11-17
Total months	174
	Backtest
Annual return	6.1%
Cumulative returns	136.8%
Annual volatility	20.3%
Sharpe ratio	0.39
Calmar ratio	0.18
Stability	0.60
Max drawdown	-34.2%
Omega ratio	1.08
Sortino ratio	0.56
Skew	NaN
Kurtosis	NaN
Tail ratio	0.94
Daily value at risk	-2.5%



The strategy initially performed well but was over-taken and underperformed the simple buy-and-hold from the latter half (2016-) and finished lower. The annualized volatility was 20% with a maximum drawdown of 34% proving to be a rather risky strategy. It will be interesting to see if Kelly sizing will improve the strategy performance. Using the formula, the optimal f was calculated to be 1.94 or 194% leverage.

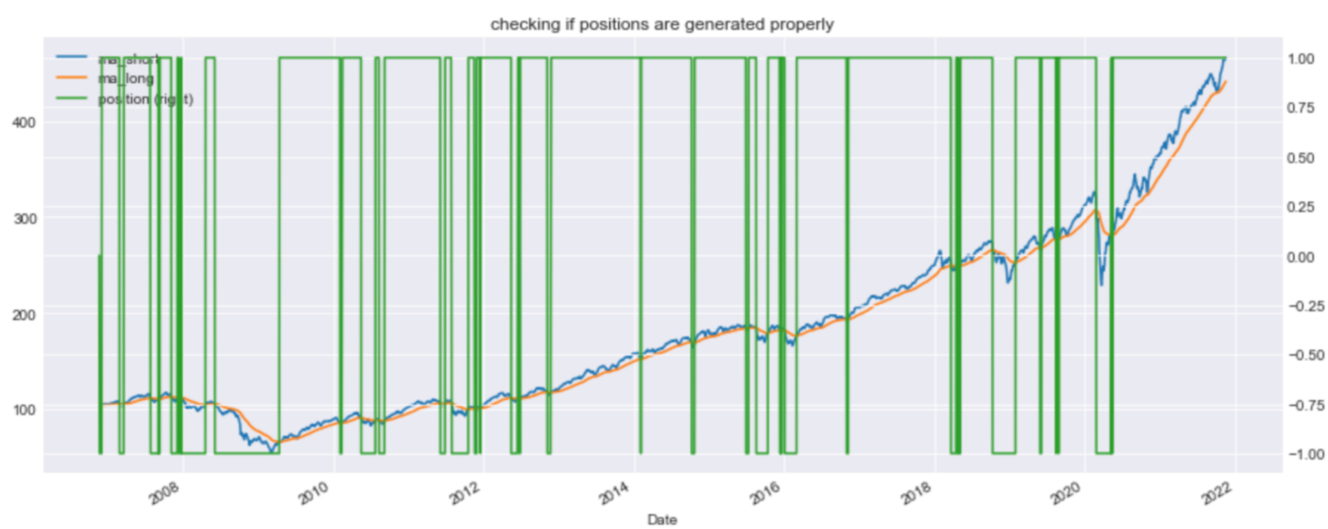
The chart above shows that the full-Kelly leverage sharply boosted the lagging performance of the strategy from 2nd half of 2020 to the end of 2021 (end of sample period) to finish almost the same as the buy-and-hold. The fractional half-Kelly was almost identical to the momentum strategy while slightly under-performing.

Start date	2007-04-19
End date	2021-11-17
Total months	174
	Backtest
Annual return	8.0%
Cumulative returns	206.8%
Annual volatility	39.4%
Sharpe ratio	0.39
Calmar ratio	0.14
Stability	0.38
Max drawdown	-58.2%
Omega ratio	1.08
Sortino ratio	0.56
Skew	NaN
Kurtosis	NaN
Tail ratio	0.94
Daily value at risk	-4.9%



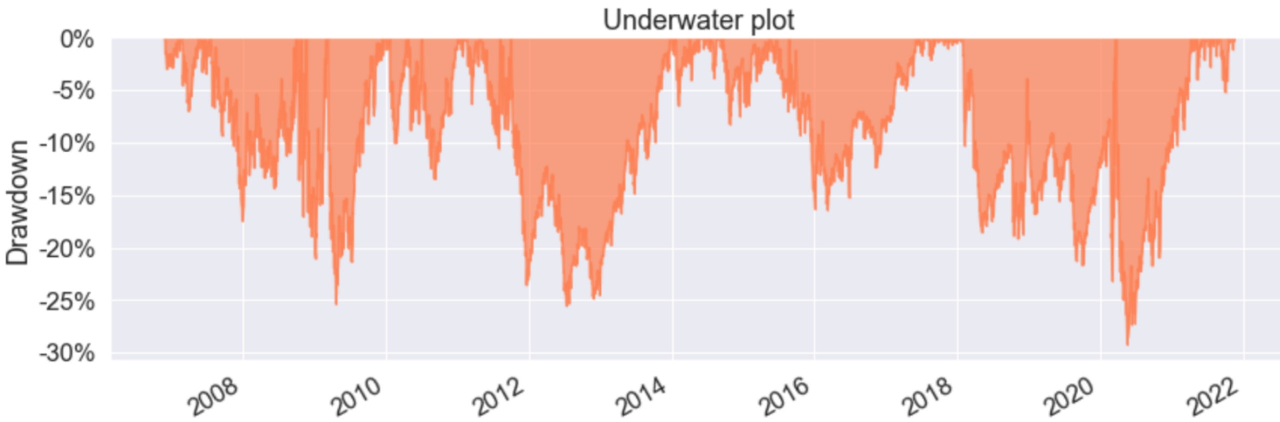
Ofcourse as expected, the full-Kelly does not come without its price and suffered a painful maximum drawdown of -58% with an annual volatility of 39%.

Next we look at a momentum strategy using exponential moving averages (EMAs) and again, following a quick parameter optimization, we use a 5 day vs 100 day EMA crossover.



Start date	2006-11-22
End date	2021-11-17
Total months	179

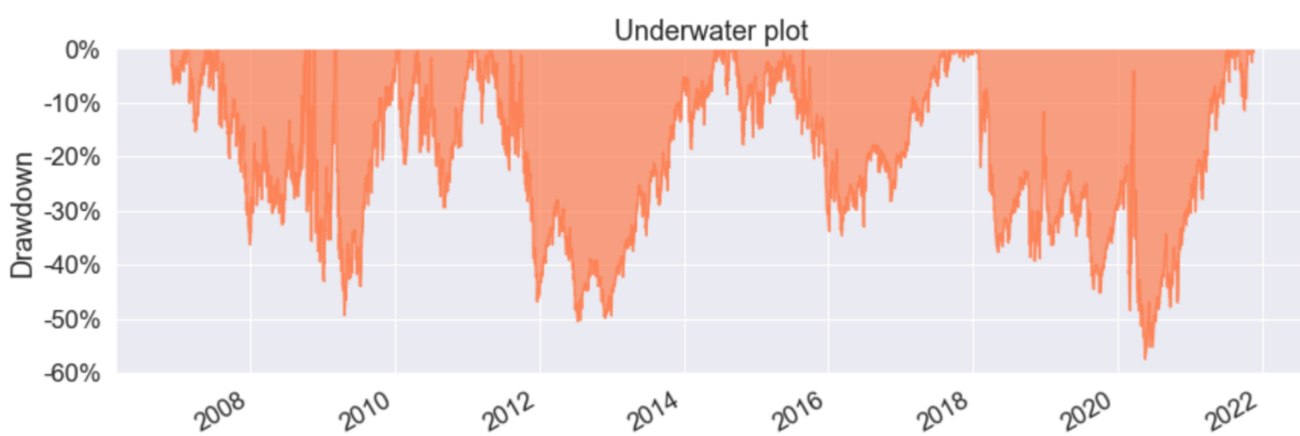
	Backtest
Annual return	7.3%
Cumulative returns	186.3%
Annual volatility	20.2%
Sharpe ratio	0.45
Calmar ratio	0.25
Stability	0.80
Max drawdown	-29.2%
Omega ratio	1.09
Sortino ratio	0.64
Skew	NaN
Kurtosis	NaN
Tail ratio	0.91
Daily value at risk	-2.5%



The strategy performed better than the simple momentum strategy and the buy-and-hold as well for most of the sample period. From late-2020 to the end-2021, the buy-and-hold began to overtake and outperformed. The annualized volatility was 20% (same as the simple momentum strategy) but the maximum drawdown albeit still bad was a 5% better -29%. Using the formula, the optimal f was calculated to be 2.22 or 222% leverage.

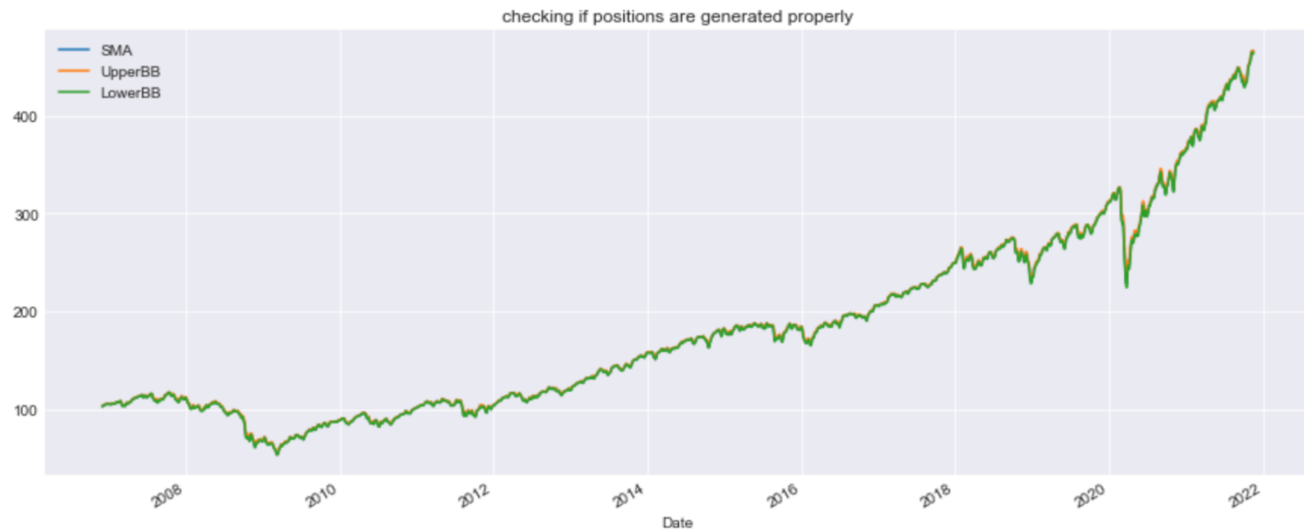
The chart above shows that the full-Kelly leverage resulted in quite a rollercoaster but managed to stay above buy-and-hold for most of the sample period, before screaming higher into end-2021 to finish at a terminal wealth of 4.47 vs. buy-and-hold (3.3) and EMA strategy (2.86). The fractional half-Kelly marginally outperformed the EMA strategy but slightly underperformed buy-and-hold, finishing at a terminal wealth of 3.1.

Start date	2006-11-22
End date	2021-11-17
Total months	179
	Backtest
Annual return	10.5%
Cumulative returns	347.4%
Annual volatility	44.8%
Sharpe ratio	0.45
Calmar ratio	0.18
Stability	0.69
Max drawdown	-57.4%
Omega ratio	1.09
Sortino ratio	0.64
Skew	NaN
Kurtosis	NaN
Tail ratio	0.91
Daily value at risk	-5.6%

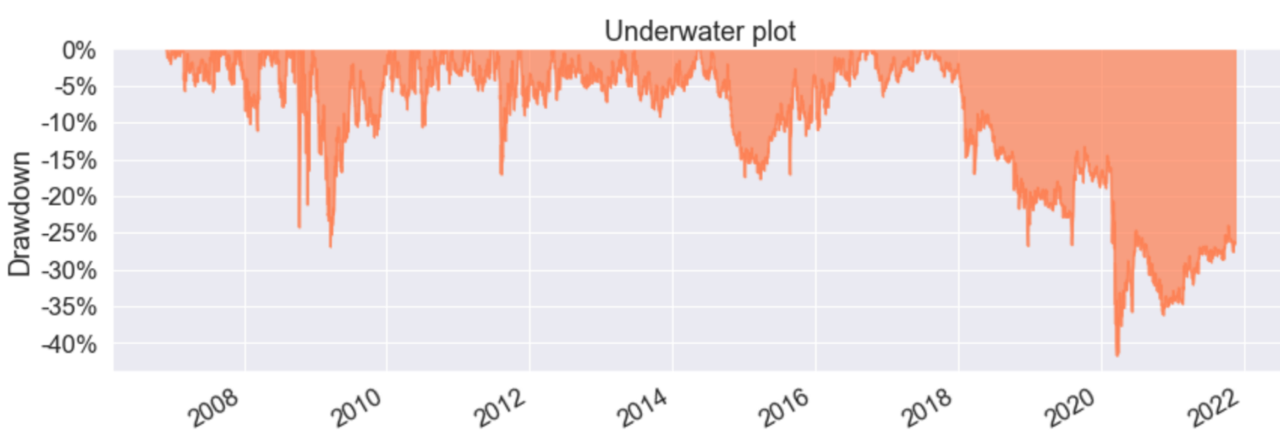


The large risks that come with the full-Kelly was again no joke, as it suffered a maximum drawdown of -57% with an annual volatility of a frightening 45%.

Finally we look at a simple mean reversion strategy using Bollinger Bands and again, following a quick parameter optimization, we use a 5 day SMA with 0.5 standard deviation bands.



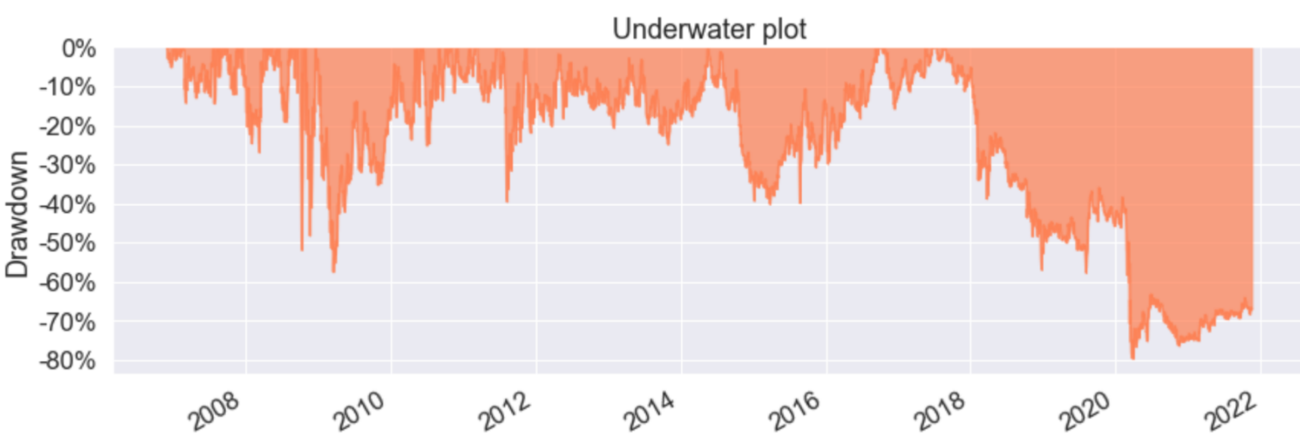
Start date	2006-11-28
End date	2021-11-17
Total months	179
	Backtest
Annual return	7.9%
Cumulative returns	211.8%
Annual volatility	19.3%
Sharpe ratio	0.49
Calmar ratio	0.19
Stability	0.54
Max drawdown	-41.7%
Omega ratio	1.11
Sortino ratio	0.73
Skew	NaN
Kurtosis	NaN
Tail ratio	1.17
Daily value at risk	-2.4%



The mean-reversion strategy outperformed the buy-and-hold by quite a significant margin for most of the time before being over-taken right before the end of the sample period. The annualized volatility was 19% (similar to the momentum strategies) but the maximum drawdown was a horrifying -42% caused by the covid crash of early 2020. Using the formula, the optimal f was calculated to be 2.53 or 253% leverage.

The chart above shows that the full-Kelly leverage wildly outperformed all strategies especially during late 2017 where it was approximately 17x buy-and-hold! However from thereafter, it literally fell off a cliff at a frightening pace, although it still finished at a solid terminal wealth of 6.05 vs. buy-and-hold (3.34). The fractional half-Kelly also managed to outperform buy-and-hold and the mean reversion strategy, finishing at a terminal wealth of 3.84.

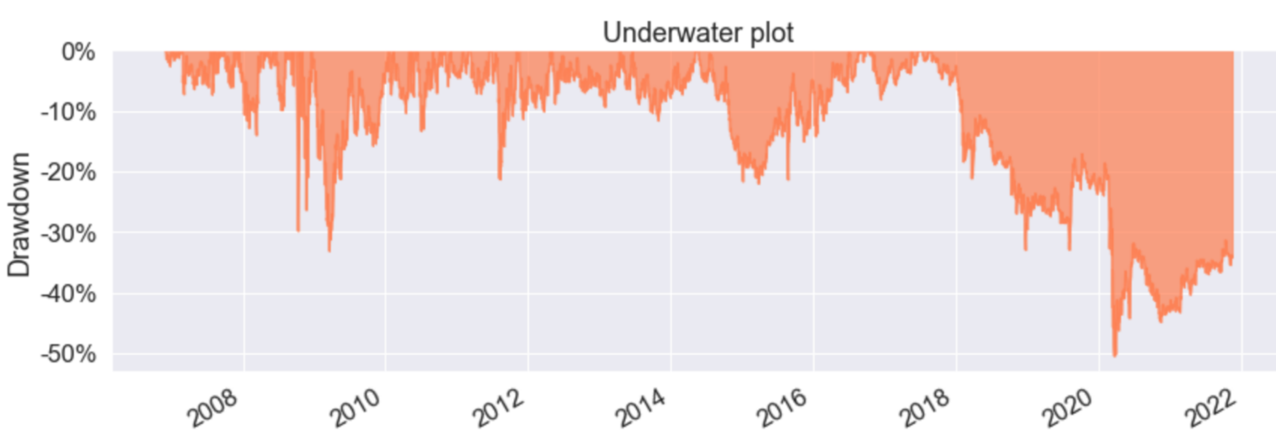
Start date	2006-11-28
End date	2021-11-17
Total months	179
	Backtest
Annual return	12.8%
Cumulative returns	504.7%
Annual volatility	48.9%
Sharpe ratio	0.49
Calmar ratio	0.16
Stability	0.36
Max drawdown	-79.6%
Omega ratio	1.11
Sortino ratio	0.73
Skew	NaN
Kurtosis	NaN
Tail ratio	1.17
Daily value at risk	-6.1%



Again we can see that the full-Kelly is not for the faint-hearted, as it records an annual volatility of 49% and suffered a maximum drawdown of -80% hit during the covid crash of early 2020!

Let us also take a closer look at the fractional half-Kelly strategy.

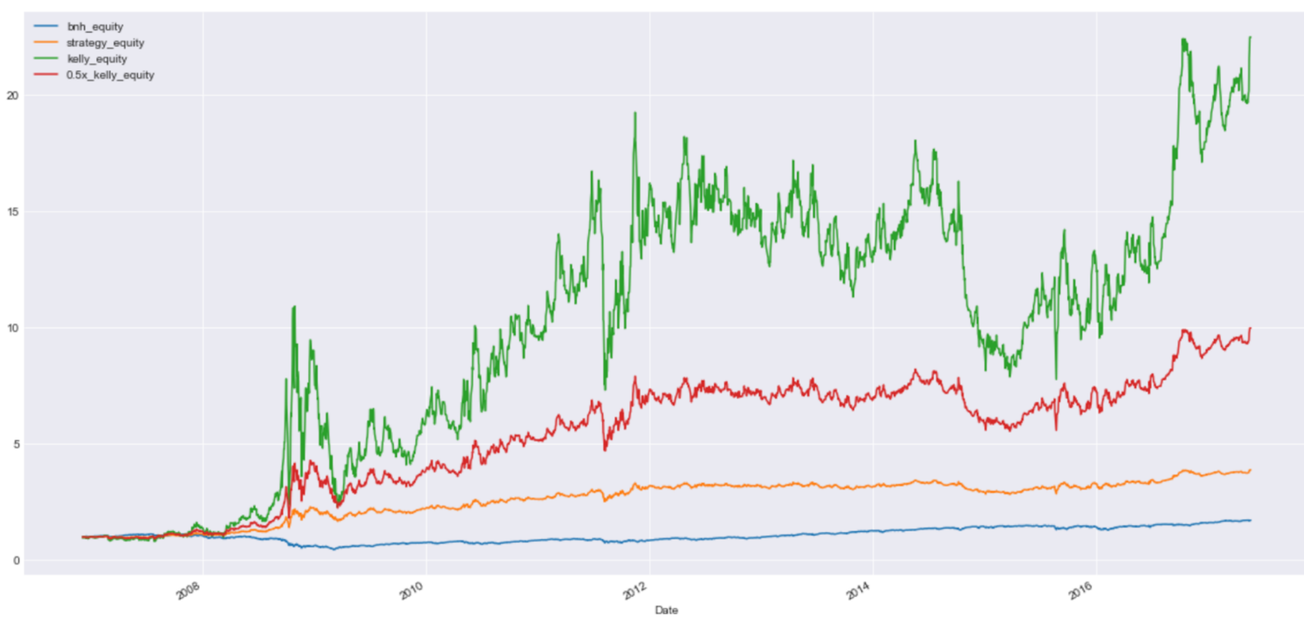
Start date	2006-11-28
End date	2021-11-17
Total months	179
	Backtest
Annual return	9.4%
Cumulative returns	284.0%
Annual volatility	24.5%
Sharpe ratio	0.49
Calmar ratio	0.19
Stability	0.52
Max drawdown	-50.4%
Omega ratio	1.11
Sortino ratio	0.73
Skew	NaN
Kurtosis	NaN
Tail ratio	1.17
Daily value at risk	-3.0%



We can see that with the half-Kelly the annual volatility is roughly halved to 24% with a maximum drawdown of -50% which is still unacceptable. However despite this, there is no disputing that the performance is impressive. Although this strategy would probably not be possible with an institutional investor, it may be considered for say a retail investor with "diamond hands" who is not bothered by the drawdowns and is only focused on maximizing GHPR and terminal value.

As with any back-test, past performance is not a reflection of future returns. This is even more true if the back-test suffered from *data-snooping bias*. Therefore we will now proceed to perform an *over-fitting test*. We will use the first 70% of the data as *in-sample* data to calculate the optimal f value, and the remaining 30% as the *out-of-sample* data to test.

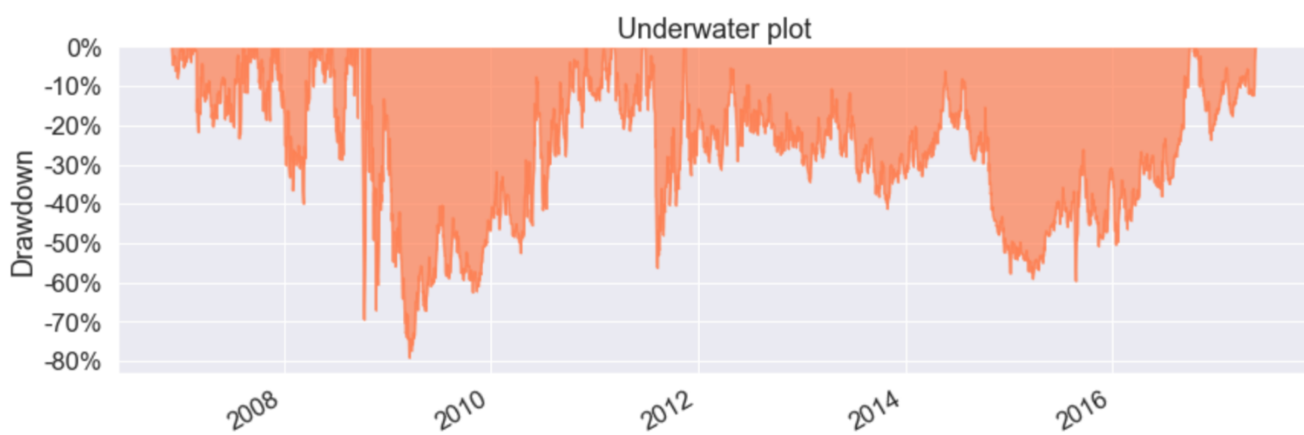
Using the in-sample data set (*28 Nov 2006 – 23 May 2017*), the optimal f comes out as 3.94 or 394% leverage which is much higher than the 253% we calculated previously. If we apply this new optimal f the performance is truly staggering.



The chart above shows that during the in-sample period, the full-Kelly leverage tremendously outperformed all strategies with a whopping terminal wealth of 22.46 vs. buy-and-hold (1.71). The fractional half-Kelly also massively outperformed finishing at a terminal wealth of about 10. This looks very promising!

Start date	2006-12-01
End date	2017-05-22
Total months	125

	Backtest
Annual return	34.7%
Cumulative returns	2146.2%
Annual volatility	76.5%
Sharpe ratio	0.76
Calmar ratio	0.44
Stability	0.66
Max drawdown	-79.1%
Omega ratio	1.18
Sortino ratio	1.20
Skew	NaN
Kurtosis	NaN
Tail ratio	1.22
Daily value at risk	-9.4%



The drawdown and volatility statistics look frightening as expected but let us assume we are that retail investor who does not care about anything but the GHPR and terminal value.

Now for the moment of truth, we will test on the out-of-sample data set (*24 May 2017 – 17 Nov 2021*) and see how it performs.

The chart above shows how not only did the mean-reversion strategy and Kelly sizing underperform buy-and-hold but they all lost money! During this period the full-Kelly strategy generated a -34% annualized return!

This is a classic example of *'over-fitting'* where we get stellar performance in the in-sample data but get the complete opposite when tested with the out-of-sample test data. When we apply this into practice, we need to be extremely careful and cognizant of the dangers as it can lead us to immense losses and even bankruptcy.

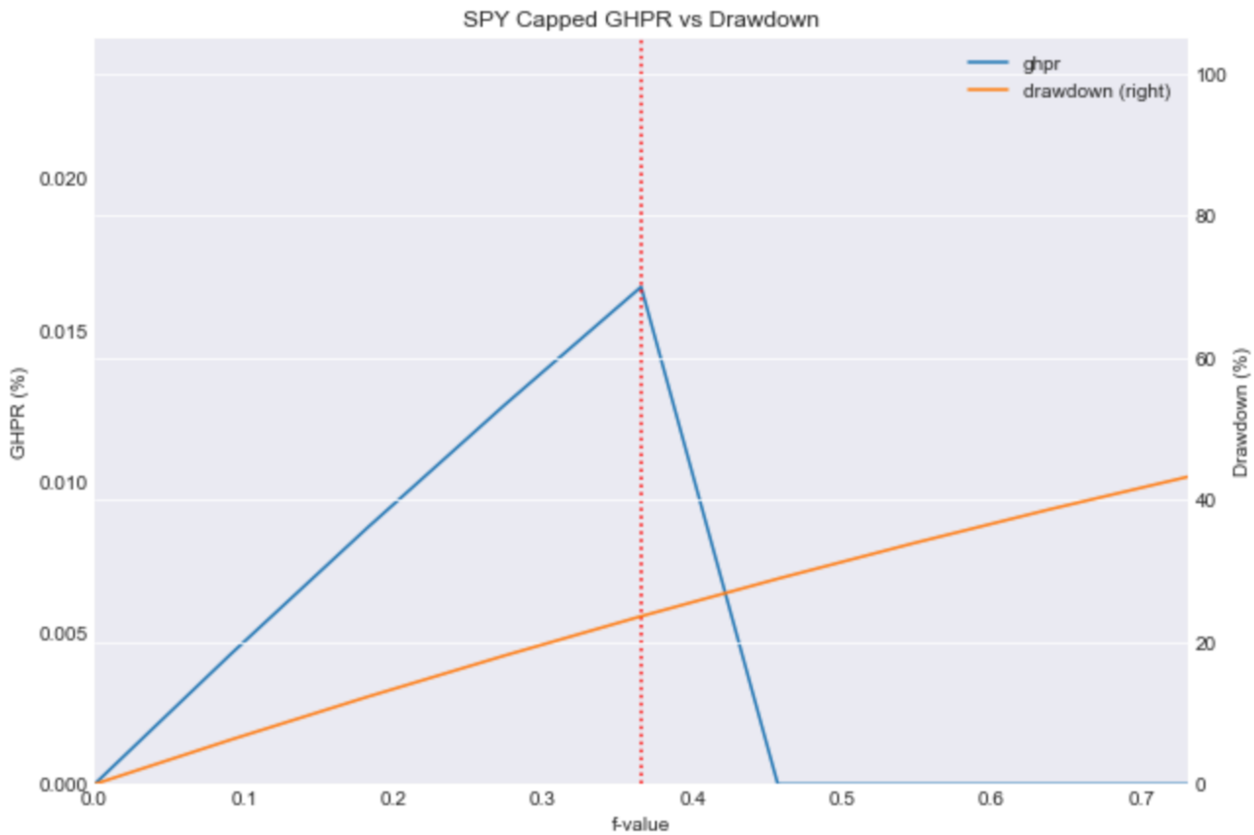
So far, we have demonstrated through back-testing with different strategies, that use of leverage at the optimal f-value can significantly boost returns. However, clearly the issue is the path volatility and the gut-wrenching drawdowns of 80-95% !

This is clearly not practical and bad risk management. There must be an f-value where the maximum drawdown is capped at an acceptable level while optimizing the GHPR.

We will investigate this by plotting two curves: f-value vs. drawdowns, and f-value vs. GHPR. Going back to our initial SPY buy & hold strategy with optimal Kelly-sizing, f = 2.97 (or 297% leverage), let's recall that the maximum drawdown was -95% during our back-testing period.

Let's see where our f-value will be capped when we limit the allowed maximum drawdown to -25%. We can already guess that this value will be above 0 and below 1. To avoid confusion, we will call this f-value the *"capped f".*

After running 100 simulations, using different f-values and optimizing for maximum GHPR while staying below maximum drawdown limit of 25%, we get the results below.

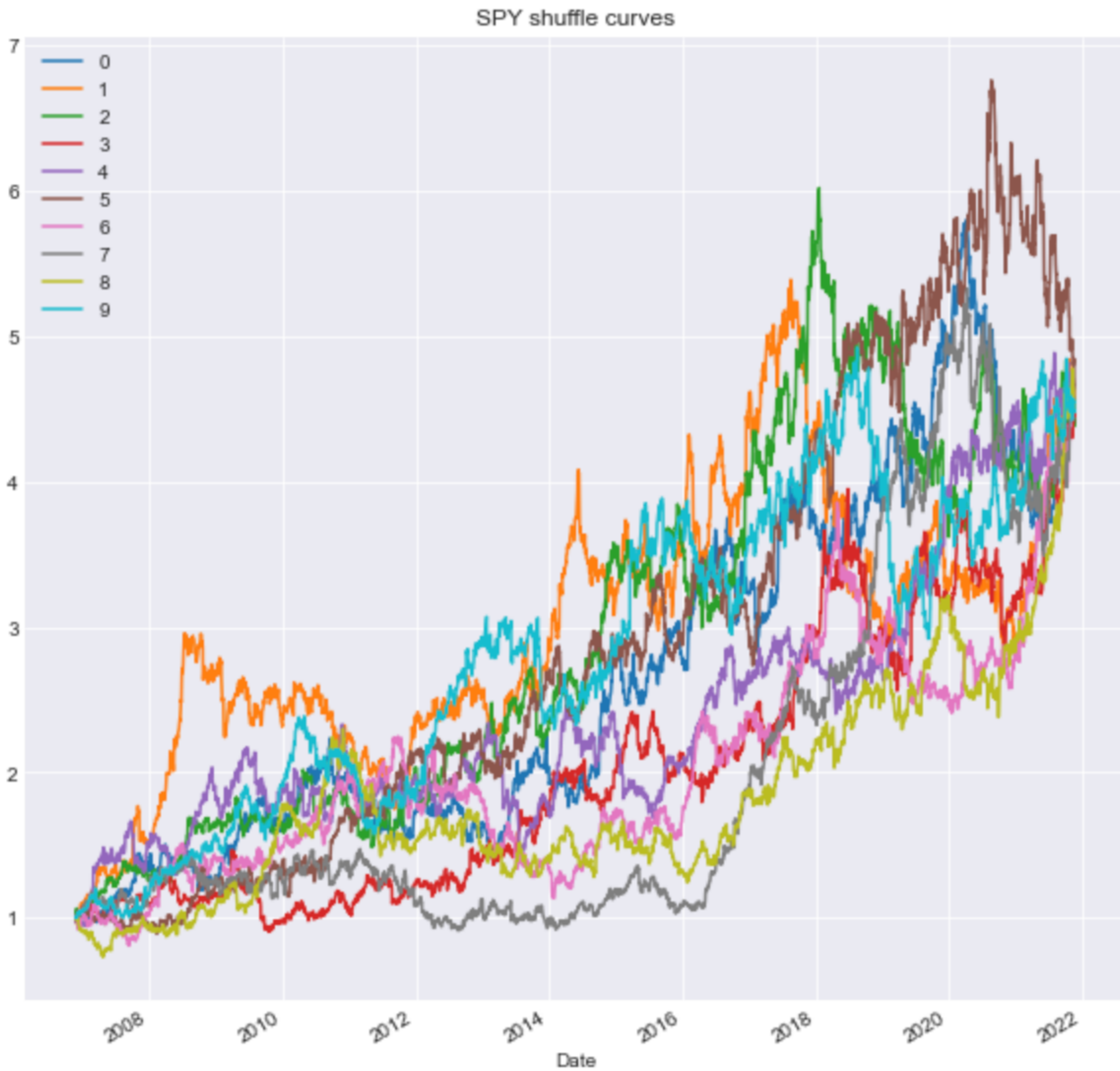


Capped f: 0.366
GHPR: 0.016%
Max Drawdown: 23.63%

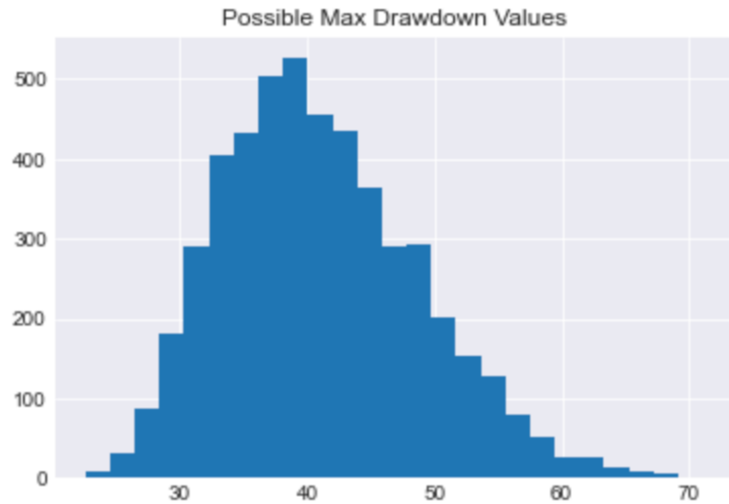
Our capped f-value gives 0.366 with a maximum drawdown within the 25% limit at 23.63%. Notice that the f-value has been massively scaled back by 8x compared to our optimal Kelly leverage of 2.97! This illustrates the steep trade-off between risk and returns. However if you are a fund manager with drawdown limit constraints, there really is no alternative.

To make things even more complicated, there is a caveat to using this methodology. The difference between your high watermark and your worst point on your equity curve (i.e your drawdown) depends on the *order* in which your returns are realized, in other words it is *path dependent*.

We will investigate how significant this impact of path dependency is on drawdown, and see whether we can dismiss it. To do this, we generate 10 randomly shuffled equity curves that all begin and end at the same point but with varying paths to get there.



We can visually see from the curves that the difference between the largest and smallest drawdowns experienced is significant! Using higher leverage will only serve to further amplify this effect. Now we know that the path dependency cannot be ignored. To investigate further, instead of generating only 10 curves, we will now generate 5000 curves and plot a histogram.



Minimum Drawdown: 22.574%
Maximum Drawdown: 71.1%

What we notice immediately is that there is an enormous 50% difference between the minimum and maximum drawdowns in this simulation! This shows that if we were to use drawdown as a risk metric, we would need to deal with its inherent uncertainty. Ofcourse financial markets are notorious for uncertainty with highly unlikely *Black Swan* events occurring more often than we would like/expect. We could however take from this simulation that we can say, with some level of confidence, that the maximum drawdown experienced won't exceed our threshold seen here. But the bottom line is that if we want to avoid breaching our drawdown limit with a high level of confidence, at least statistically, then we would need to scale down our position size by more than you might think to avoid over-betting the optimal f.

The key to successfully applying the Kelly strategy and maximizing the rate of growth of capital and the terminal value while staying within risk constraints, depends on your ability to correctly predict the returns distribution of your trading strategy.

Ofcourse no one can accurately and consistently predict this, but we can take steps to minimize the error between the estimated and actual distributions. Machine learning methods such as neural networks can be used to make better predictions of the distribution. This would be worth investigating in future work.

Challenges/Limitations

First and foremost, it is clear that when applying the Kelly Criterion to gambling bets with binary outcomes where the odds are in her favour, it works fantastically. However, we have found that it is difficult to apply to investing in markets. Not to forget, it only holds true for positive EV trading strategies. But with some modifications and assumptions, we can say that it is still useful in maximizing GHPR and terminal wealth. We also identified the inherent risk when sizing our trades under the full-Kelly optimal f leverage. Using fractional-Kelly like a ¼ or ½ Kelly can reduce the risk although sacrificing a large portion of the returns.

Also in practice, we need to maintain a constant leverage ratio meaning it requires constant rebalancing, ideally on a daily basis. This incurs large trading costs depending on the trading asset and we have omitted this here for simplicity. If we factor these in to make it realistic, then the returns are likely to suffer a big hit.

In our back-testing results we observed a classic example of over-fitting and look-ahead bias. This results in underestimating risk and overestimating the GHPR. This is very dangerous as it will leave you massively over-leveraged and quickly exceed your risk tolerance. Over-fitting can obviously be overcome with *walk-forward testing* where possible via actual live trading or at least paper trading.

Finally, it is important to understand that the example we used (SPY ETF) and other equity index returns tend to exhibit both *volatility clustering* and *autocorrelation*. Hence this goes against our model assumption that the individual returns are independent from previous returns. What we could potentially do to overcome this, is to take these into account for our algorithm inputs in future.

Conclusion

Assuming we enjoy the luxury of focusing solely on maximizing GHPR and our investment terminal value without any risk constraints, there is no other simple yet powerful strategy like the Kelly criterion if you have a positive EV (profitable) trade. However, in practice a full-Kelly sized strategy is too dangerous especially under high leverage, and we need to mitigate some of the risk. As we have demonstrated, there is a consistent trade-off between GHPR and the optimal f value. Also a drawdown limit approach when calculating the optimal f value is very difficult and depends largely on the path of returns. Therefore, the Kelly criterion can be used to set an upper limit on the sizing or leverage that should be used for a trade which once exceeded can lead to bankruptcy with high confidence.

Due to the uncertainty of large drawdowns with a Kelly-leveraged strategy, it would make sense to diversify your portfolio with uncorrelated assets. However, optimizing the portfolio with the Kelly criterion is still a valid strategy for high risk-tolerance investors and the returns can be fantastic, especially under a bullish market regime.

Annexure/Codes

Please see separate file: *EPAT_Final_Project_2021.ipynb*

Bibliography

Maclean, L., Thorp, E. and Ziemba, W.T. (2010) *Long-Term Capital Growth: The Good and Bad Properties of the Kelly Criterion. Quantative Finance,* 10, 681-687. https://doi.org/10.1080/14697688.2010.506108

Davis , M. and Lleo, S. (2010) *Fractional Kelly Strategies for Benchmarked Asset Management.* https://www.researchgate.net/publication/228860076_Fractional_Kelly_Strategies_for_Benchmarked_Asset_Management

Thorp, E. (2007) *The Kelly Criterion in blackjack sports betting, and the stock market* http://www.eecs.harvard.edu/cs286r/courses/fall12/papers/Thorpe_KellyCriterion2007.pdf

Vince, R. (2009) *The leverage space trading model: reconciling portfolio management strategies and economic theory*. Vol. 425. John Wiley and Sons.