Kelly criterion for multivariate portfolios: a model-free approach

Vasily Nekrasov*†

October 23, 2013

Abstract

The Kelly criterion is a money management principle that beats any other approach in many respects. In particular, it maximizes the expected growth rate and the median of the terminal wealth. However, until recently application of the Kelly criterion to multivariate portfolios has seen little analysis. We briefly introduce the Kelly criterion and then present its multivariate version based only on the first and the second moments of the asset excess returns. Additionally, we provide a simple numerical algorithm to manage virtually arbitrarily large portfolios according to so-called fractional Kelly strategies.

Keywords: Kelly criterion, money management, multivariate portfolios, fractional Kelly strategies, analytic and numerical approximation, portfolio optimization on GPUs, CUDA.

The Kelly criterion is well known among the gamblers as a formula to calculate the optimal bet size in games with a positive expected return. Assume a gambler tosses a biased coin so that the probability p to get a tail is known and larger than 0.5. After each bet a gambler loses or doubles the money at stake. In order to maximize his/her expected wealth s/he should put at stake all his/her capital on each bet. However this is too risky because each bet looms the danger to lose everything and as the number n of bets goes to ∞ a gambler will eventually go bankrupt. This "paradox" is due to the fact that by placing maximal bets there is only one "lucky" path: n wins on n bets. Since $p > 0.5$ the expected wealth $\mathbb{E}[X_n] = X_0(2p)^n$ still grows with each bet though the probability of the "lucky" path decreases to zero as n grows.

This, however, would be little comfort for a bankrupt gambler. Thus even in a favorable game one should put at stake only a fraction of the available capital. Kelly [1956] suggested to bet a fraction, which maximizes the *expected growth*

*University of Duisburg-Essen, Dept. of Economics, Chair of Energy Trading and Financial Services, finanzmaster@gmx.net

†The author gratefully acknowledges hardware donation from NVIDIA Academic Partnership Program.

rate. Most likely he wrote his seminal paper together with Claude Shannon but signed it alone because AT&T did not want to announce that bookies represented a large part of its customers (see Case [2006]).

Kelly pointed out two crucial assumptions: reinvested winnings and a large number of bets. This is not always the case in gambling but in portfolio management it usually is. Kelly also mentioned that his strategy is equivalent to the maximization of the logarithmic utility but he explicitely avoided any argumentation based on utility functions. His main rationale is that his approach eventually overtakes any other approach, which bets a different but still a constant wealth fraction. However, the equivalence with logarithmic utility turned out to be very practical: Bellman and Kalaba [1957] and Mossin [1968] proved that for a series of bets the optimal fraction remains constant if and only if one maximizes the expected power or logarithmic utility. Since the optimal fraction remains constant, it does not depend on the number of trades. Respectively, such utility functions are called *myopic* because one optimizes each trade as if it were the last trade.

Breiman [1961] proved that Kelly's approach X^* beats *any* other money management approach X in the sense[1] that $\mathbb{E}[X/X^*] \leq 1$. Moreover, Ethier [2004] showed that under fairly general conditions the Kelly criterion maximizes the median of terminal wealth. It is a very important property because (esp. in case of highly skewed distributions) the median as a measure of central tendency is often preferred to the mean. Indeed, maximizing the median we maximize what we get by a *typical* game or market scenario. Note that in this context "typical" does not mean "the most probable" (otherwise it would be a mode, not a median). But it is typical in the sense that a median represents neither bad luck nor good luck but, so to say, a "middle luck".

In spite of all its nice properties the Kelly criterion is suprisingly little known among portfolio managers. Probably this is due to the fact that portfolio management is essentially multivariate but (with a few exceptions) the literature on Kelly criterion considers only the univariate case. Breiman [1961] did consider a multivariate game but his appoach is purely theoretical since one needs to know the joint distribution of the outcomes. Maslov and Zhang [1998] started with a general approach, which is similar to ours but ended up with a formula for uncorrelated assets only. Their idea was enhanced in Laureti et al. [2010] where a solution for the case of correlated assets is, in principle, provided. However, for this case they just briefly sketch their approach in the appendix and neither reduce it to a well-known approach like e.g. quadratic programming, nor run any numerical trials, nor solve the optimization problem for the fractional Kelly strategies. Rising and Wyner [2012] is very close to our study but they had to assume that the variance of a portfolio return is approximately equal

[1] To prevent a confusion we emphasize that in general $\mathbb{E}[X/X^*] \neq \mathbb{E}[X]/\mathbb{E}[X^*]$. Breiman's result does not mean that the Kelly criterion overtakes any other strategy in expectation but it loosely means the following: if X performs well then X^* will likely also do (maybe a little bit worse). But if X performs (extremely) badly then X^* will probably be also bad but not as disastrous as X.

to its second (non-centered) moment, which is not the case if the expected asset returns are large. Moreover, they interpret fractional Kelly strategies as *proportional* reduction of risky asset fractions. As we demonstrate in Exhibit 3, such straightforward interpetation may be suboptimal if the asset returns are not jointly Gaussian. Additionally, neither Maslov and Zhang [1998], nor Laureti et al. [2010], nor Rising and Wyner [2012] analyze the quality of the approximation by Taylor series (see Appendix 1). Cover [1984] suggested a simple and very fast algorithm to maximize the log investment return but it is unsuitable for the fractional Kelly strategies. The recent paper by Davis and Lleo [2012] is very interesting but also very technical and not so easily applicable in practice since they deal with possibly unobservable factors.

Another reason for the obscurity may be the absence of the Kelly criterion in university curricula[2]. This is, in turn, due to the criticism by prominent economists like Samuelson [1971][3]. In this sense it is even more remarkable that some of Samuelson's counter-arguments actually speak in favour of the Kelly criterion. Samuelson [1971] wrote:

> Let the gambler-investor face a choice between investing completely in safe cash or completely in a "security" that yields for each dollar invested, \$2.70 with probability 1/2 or only \$0.30 with probability 1/2. To maximize the geometric mean, one must stick only to cash, since $[(2.7)(.3)]^{\frac{1}{2}} = .9 < 1$. But, Pascal will always put all his wealth into the risky gamble.

This is certainly true but who can force market players to invest *only* either in cash or in a security?! We write "security" without quotation marks, since suchlike securities do exist, consider e.g. Nokia or solar stocks. In Samuelson's case a Kelly investor will put 42% of his/her capital in a security and yield $[(1+1.7\cdot 0.42)(1-0.7\cdot 0.42)]^{\frac{1}{2}} = 1.100$, i.e. 10% growth rate in a typical market scenario. Moreover, a risk-averse person will likely bet even less than the Kelly fraction (one half of it is a common practice). Exhibit 1 shows two outcomes of "Kelly and half-Kelly vs. Pascal investment"[4].

Application to multivariate portfolios

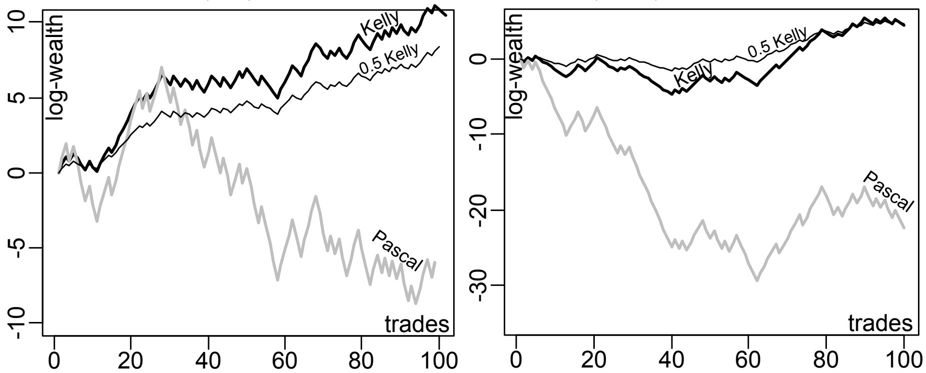
Consider a market with n correlated risky assets $S_k, k = 1, .., n$ and a riskless bond. Denote a random return on asset S_k with r_k and the return on bond with r. Without loss of generality let investor start with initial capital equal to \$1.0 and invest a wealth fraction u_k in S_k. The residual capital $(1 - \sum_{k=1}^{n} u_k)$

[2] E.g. in Germany, to our knowledge, the Kelly criterion is systematically taught only at the University of Greifswald and Frankfurt School of Finance and Management.

[3] Samuelson strongly criticized the Kelly criterion but actually got rich due to his investment in Warren Buffets Berkshire Hathaway Inc. Interestingly, Buffet seems to act as full Kelly investor. See Wenzel [2011] and Ziemba [2003].

[4] The readers can reproduce Exhibit 1 and Exhibit 2 with author's program in R language: http://www.yetanotherquant.de/kelly/kellyR.zip

Exhibit 1: Samuelson's example: Kelly and half-Kelly vs. Pascal in a typical market scenario (left) and in an unfortunate case (right).



is invested in risk-free bond. After the first trade one yields

$$\sum_{k=1}^{n} u_k\,(1+r_k) + (1+r)\,(1-\sum_{k=1}^{n} u_k) = (1+r) + \sum_{k=1}^{n} u_k(r_k - r) \tag{1}$$

According to the Kelly criterion we need to find a vector of fractions $\mathbf{u} = (u_1, \ldots, u_n)^T$ that maximizes

$$\mathbb{E}\left[\ln\left((1+r) + \sum_{k=1}^{n} u_k(r_k - r)\right)\right] \tag{2}$$

The optimization of the logarithmic utility is myopic. Thus in order to maximize the expected terminal logarithmic wealth we just maximize (2) on each trade[5]. Since we make no assumption on the probability distribution of returns, we can hardly find a closed form solution for (2). But expanding it as the Taylor series about $\mathbf{u_0} = (0, \ldots, 0)$ we obtain

$$\mathbb{E}\left[\ln(1+r) + \sum_{k=1}^{n} \frac{u_k(r_k - r)}{1+r} - \frac{1}{2}\sum_{k=1}^{n}\sum_{j=1}^{n} u_k u_j \frac{(r_k - r)(r_j - r)}{(1+r)^2}\right] \tag{3}$$

or equivalently in matrix notation

$$\mathbb{E}\left[\ln(1+r) + \frac{1}{1+r}(\mathbf{r} - \mathbf{1}r)^T\mathbf{u} - \frac{1}{2}\frac{1}{(1+r)^2}\mathbf{u}^T\mathbf{\Sigma}\mathbf{u}\right] \tag{4}$$

This is a *quadratic optimization problem* and the [unconstrained] solution (with estimated parameters $\widehat{\mathbf{r}}$ and $\widehat{\mathbf{\Sigma}}$) is

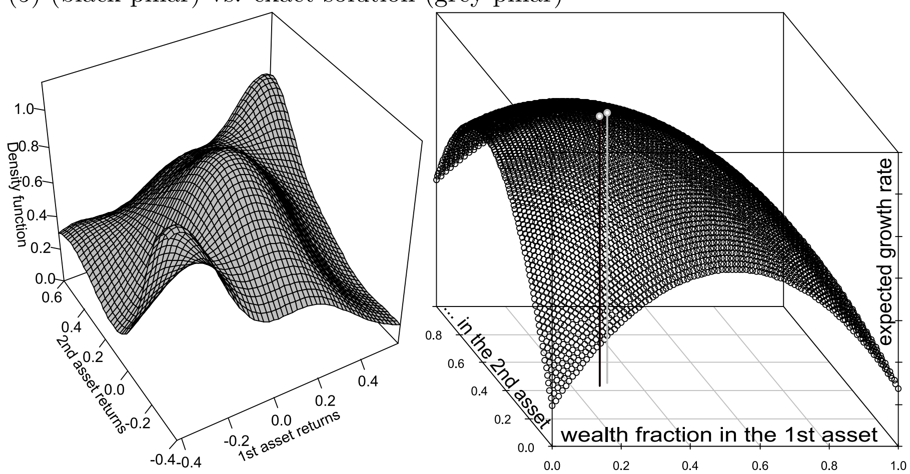
$$\mathbf{u}^\star = (1+r)(\widehat{\mathbf{\Sigma}})^{-1}(\widehat{\mathbf{r}} - \mathbf{1}r) \tag{5}$$

[5]In case of a numerical solution it may be better to simulate many trades per path in order to make the terminal results more distinguishable.

In practice a fund manager must usually adhere to the no leverage and no short selling constraints. In this case the solution does exist as well but must be found numerically, e.g. with 'quadprog' R library[6] or Maple.

In Appendix 1 we analytically evaluate the error of the approximation of (2) by (3) under some constraints that are, however, common in practice. Additionally, we ran various numerical trials that have verified the quality of the approximation. Exhibit 2 shows such a case: the returns are drawn from a rather irregular distribution (asymmetric, multimodal, heavy-tailed), which can hardly be exhaustively characterized by its first and second moments. Still the approximative solution of the portfolio optimization problem is very good.

Exhibit 2: Probability density of returns and the approximation according to (5) (black pillar) vs. exact solution (grey pillar)



Note that we made no assumption on the distribution of returns. All we need to know is just the expectation of the *excess* returns and the matrix of their second (noncentral) mixed moments. These parameters can be readily estimated from historical asset prices. However, the parameter estimation errors can drastically distort the solution. Thus a portfolio manager should not rely merely on the point estimates. Instead s/he should ponder the confidence intervals of the parameters and proceed from the *lower* bound for the expected returns and the *upper* bound of covariance. This will lead to underbetting so that the expected gain and the risk both decrease, whereas in case of overbetting the risk grows but the expected gain still decreases (see Thorp [2006]).

Algorithm for the fractional Kelly strategies

The strategies that bet less than the Kelly fraction are called fractional Kelly strategies. As we just pointed out, they are less profitable but also less risky. In

[6] http://cran.r-project.org/web/packages/quadprog/index.html

the case of a univariate portfolio we just decrease the capital fraction invested in the risky asset and shift the released capital to the riskless bond. But for a multivariate portfolio such an approach is ambiguous. Indeed, an investor usually cares about the *total* fraction of risky assets in his/her portfolio. And s/he can at first proportionally descrease the fractions of all risky assets in $\mathbf{u}^\star$ but then additionally redistribute the capital among the risky assets. If the returns are *jointly* Gaussian then we shall just proportionally reduce the fractions of the risky assets. Moreover, in this case the strategy with Kelly fraction $x := 1/(1-\alpha)$ is equivalent to the maximization of the (negative) power utility x^α/α, $\alpha < 0$. But in general this nice property does not hold true (see MacLean et al. [2010] and Davis and Lleo [2012]). For instance[7], even if the marginal returns are Gaussian but the marginals are paired by Clayton copula, then the proportional fraction reduction is suboptimal. Imposing additional constraints $0 \leq u_i \leq x \; \forall i$ and $\sum_{i=1}^n u_i \leq x$ we can still use (3) to find the optimal portfolio for a given x. However, (3) may be a bad approximation for the leveraged portfolios or the portfolios that contain assets with possibly very high returns[8] (see Appendix 1). Additionally, practitioners usually like having two *independent* optimization methods: if both of them get similar results then they are plausible.

We consider the portfolio optimization under the no leverage and no short selling constraints. A Monte Carlo gradient algorithm we employ is as follows:

1. Assume we want to invest x of our capital in risky assets[9], $x \in [0\%, 100\%]$. Choose an initial portfolio so that all fractions of the risky assets are between 0.0 and x and their sum is equal to x. Distributing the capital over all assets equally may be a good start.
2. (Gradient step): modify the fractions at random so that the no short selling and no leverage conditions still hold. If the new portfolio yields higher expected growth rate, set it as initial portfolio, otherwise leave initial portfolio unchanged.
3. Repeat Step 2 until portfolio performance stops increasing.

We can readily simulate a uniform distribution on the set of admissible portfolios, i.e. those conforming to the no leverage and no short selling constraints. In case of n risky assets let us simulate $z_i = \mathbb{U}[0,1]$ for $i \in [0,..,n]$, define $r := \sum_{i=1}^n z_i$ and set $u_i := xz_i/r$. Then $0 \leq u_i \leq x, \forall i$ and $\sum_{i=1}^n u_i = x$.
This algorithm should also be used to validate the approximation (5). If the solution from (5) is similar to that by simulation then it is plausible. Since we optimize the logarithmic or, in case of a fractional Kelly strategy, a power utility[10], there is a unique extremum and the algorithm will (sooner or later)

[7] See http://www.yetanotherquant.de/kelly/kellyR.zip, script 3a and 3b

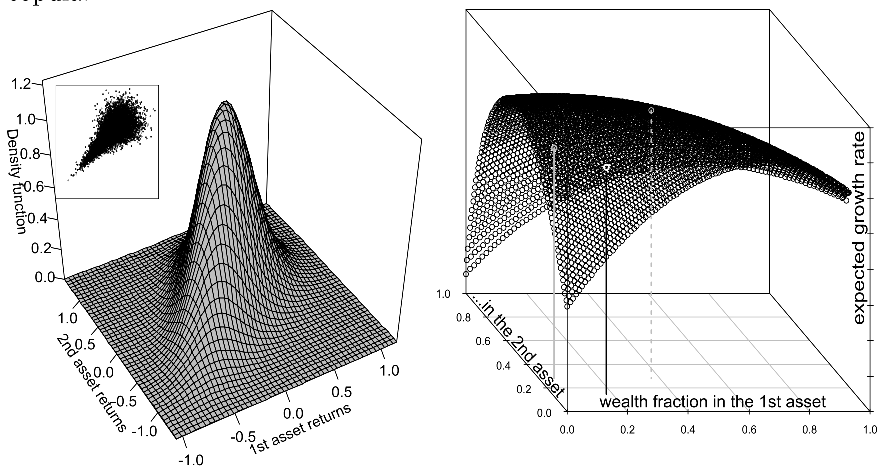
[8] These may be e.g. the call options.

[9] A risk-averse investor will choose x less than in a full Kelly strategy. But the algorithm is also correct for risk-seeking strategies that bet more than the Kelly fraction. These strategies may be relevant for short-term speculators, see Browne [2000].

[10] As we do not know the number of future trades, we have to optimize myopically. This implies, that we optimize a power utility function. However, in case of non-Gaussian returns

converge.

Exhibit 3: Expected growth rate by full Kelly strategy (grey dashed pillar), proportionally reduced fractions(black pillar) & reduced and restructured fractions(grey pillar). Marginal Gaussian returns are paired by the Clayton copula.



The algorithm is very flexible since all we need is just to be able to simulate the returns. Moreover, we can even stay model-free and sample with replacement from the historical returns. Although the algorithm is really primitive, it works well. In contrast to Cover [1984] we can also optimize fractional Kelly strategies.

The only problem we encountered is an enormous computational intensity. As a matter of fact we need to simulate very many paths to assure convergence. Moreover, a path should be long enough to let the Kelly strategy prevail for sure. This problem can, however, be efficiently mitigated by means of GPUs (graphics processing units) [11]. We have tested the algorithm with a portfolio of seven stocks from DAX[12]: Adidas, Bayer, BMW, Lufthansa, Fresenius, RWE and Siemens. We assume daily portfolio rebalance. This is certainly not very common in practice but our primary goal is to test the numerical convergence. For this purpose the daily returns are particularly good: the difference is much less pronounced than e.g. by annual returns thus the tested algorithm is really challenged. For every stock we simulate 26214400 paths for each iteration, a path consists of 10000 trades and the number of iterations is set to 300. So the total computational complexity is about 10^{14}, i.e. 100 trillions operations.

---

we cannot, in general, relate its risk-aversion parameter to the Kelly fraction, see MacLean et al. [2010].

[11] Additionally, we are developing an accelerated proprietary version of the algorithm, enhanced both technically and mathematically.

[12] German equivalent of the Dow Jones Industrial Average. We have chosen the least correlated stocks.

We chose seven stocks because on the one hand, it is sufficient to make optimization by a brute-force enumeration of the portfolio fractions impossible, on the other hand the correlation matrix is still managable. It is given by:

	Adidas	*Bayer*	*BMW*	*Lufth.*	*Fresen.*	*RWE*	*Siem.*
Adidas	1.00000	0.16548	0.16272	0.38367	0.86897	0.13664	0.22027
Bayer	0.16548	1.00000	0.43397	0.42181	0.11369	0.47749	0.49384
BMW	0.16272	0.43397	1.00000	0.53210	0.05055	0.45504	0.61104
Lufth.	0.38367	0.42181	0.53210	1.00000	0.29718	0.43157	0.55309
Fresen.	0.86897	0.11369	0.05055	0.29718	1.00000	0.08171	0.10376
RWE	0.13664	0.47749	0.45504	0.43157	0.08171	1.00000	0.50391
Siem.	0.22027	0.49384	0.61104	0.55309	0.10376	0.50391	1.00000

We set the interest accrued overnight to $0.04/365 = 0.00011$ and estimate the daily mean returns, they are:

Adidas	*Bayer*	*BMW*	*Lufth.*	*Fresen.*	*RWE*	*Siem.*
0.000380	0.000361	0.000263	0.000093	0.000428	0.000209	0.000254

However, solving the optimization problem with these data according to (5) yields negative fractions for Adidas and Lufthansa. Since we want to test under no short selling and no leverage conditions, we first test under these constraints with original data and additionally run a test with the estimation of the mean returns "adjusted" as follows:

Adidas	*Bayer*	*BMW*	*Lufth.*	*Fresen.*	*RWE*	*Siem.*
0.000456	0.000253	0.000263	0.000279	0.000385	0.000209	0.000254

This let us consistently compare the results from (5) and from our numerical algorithm. Futher we assume that the returns are normally distributed. Of course we may assume any distribution as long as the 1st and the 2nd statistical moments of the simulated data are equal to those of historical data. Hence we can even stay distribution-free and sample with replacement from the historical data. But for the normally distributed returns there is an exact closed form solution by Merton [1969], so it would be useful to compare this solution with the approximation according to (5) and with the solution via numerical simulation. The results for the case of the "adjusted" mean returns are as follows:

	Adidas	*Bayer*	*BMW*	*Lufth.*	*Fresen.*	*RWE*	*Siem.*
Merton	0.01207	0.15903	0.24826	0.13879	0.2469	0.02839	0.06981
Approx.	0.01212	0.15892	0.24820	0.13896	0.2468	0.02839	0.06977
Numeric	0.01898	0.15455	0.23513	0.15101	0.2410	0.03150	0.06482

As we can readily see, Merton's exact solution and the approximation (5) are nearly equivalent and the numerical simulation yields very similar results as well.

It takes about ten hours to solve the problem with a single NVIDIA Tesla K20 card[13] but one can install up to four cards in a modern PC with a suitable

[13]Source code in CUDA: http://www.yetanotherquant.de/kelly/exampleInCUDATeslaK20.zip

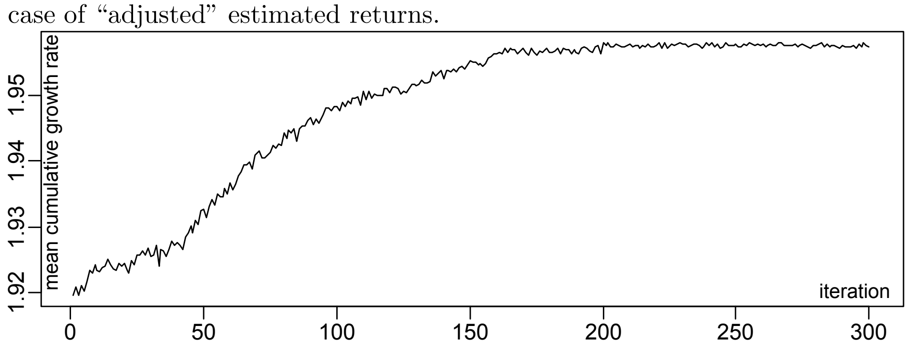
mainboard[14]. Since our method is not affected by the curse of dimensionality, it may be applied to virtually arbitrarily large portfolios[15]. From Exhibit 4 we conclude that about 200 iterations are sufficient to achieve convergence in this case.

As to the test with original data, we obtain (under no leverage and no short selling constraints) the following optimal portfolio:

	Adidas	*Bayer*	*BMW*	*Lufth.*	*Fresen.*	*RWE*	*Siem.*
Approx.	0.0	0.56517	0.14144	0.0	0.29339	0.0	0.0
Numeric	0.0	0.57785	0.12905	0.0	0.29311	0.0	0.0

The simulated results are again very close to the analytical approximation, yet in this case we needed about 600 iterations. Namely, at first we have simulated with a relatively large gradient step and took the best portfolio from 300 iterations as the starting point for the second simulation with a smaller gradient step.

Exhibit 4: Convergence of the solution for a portfolio of seven DAX stocks, the case of "adjusted" estimated returns.



Obviously, we may need more iterations for other cases, especially when the number of assets is large. But for the optimization with respect to fractional Kelly strategies there is an efficient rule of thumb:

1. Find the "optimal" portfolio for the fractional Kelly strategy analytically under the assumption of jointly Gaussian returns (i.e. proportional reduction of the risky asset fractions).
2. Tune the simulation parameters (number of paths, gradient step, number of iterations) so that the numerical solution converges to the analytical approximation.
3. Apply the simulation algorithm with these parameters to find the optimal portfolio for the fractional Kelly strategy. (In this step we do not assume jointly Gaussian returns anymore and draw the returns from a realistic distribution or from historical data with replacement).

[14]For instance, ASUS Z9PE-D8 WS has 7 PCI-Ex16 slots and costs less than $600. Tesla K20 is more expensive but still affordable even for a private investor.

[15]Computational overhead will still grow with the number of assets just because we need to simulate paths for each asset. But this is merely a linear growth.

4. If there is no clear sign of convergence like that at Exhibit 4 after the 200th iteration, halve the gradient step and continue simulation with as many iterations as were done so far.

References

Bellman, R. and Kalaba, R. (1957). On the role of dynamic programming in statistical communication theory. *Information Theory, IRE Transactions on*, 3(3):197 –203.

Breiman, L. (1961). Optimal gambling systems for favourable games. In *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1. Univ. Calif. Press, Berkeley, CA.

Browne, S. (2000). Can you do better than Kelly in the short run? *Finding the Edge, Mathematical Analysis of Casino Games*, pages 215–231.

Case, J. (2006). Book review: Fortunes formula: The untold story of the scientific betting system that beat the casinos and Wall street. *SIAM News*, 39(3):754–755.

Cover, T. M. (1984). An algorithm for maximizing expected log investment return. *IEEE Transactions on Information Theory*, 30(2):369–373.

Davis, M. and Lleo, S. (2012). Fractional Kelly strategies in continuous time: Recent developments. In MacLean, L. C. and Ziemba, W., editors, *Handbook of the Fundamentals of Financial Decision Making*. World Scientific Publishing.

Ethier, S. N. (2004). The Kelly system maximizes median fortune. *Journal of Applied Probability*, 41(5):1230 – 1236.

Kelly, J. L. (1956). A new interpretation of information rate. *Bell System Technical Journal*, pages 917–926.

Laureti, P., Medo, M., and Zhang, Y.-C. (2010). Analysis of Kelly-optimal portfolios. *Quantitative Finance*, 10(7):689–697.

MacLean, L., Thorp, E., and Ziemba, W. T. (2010). Good and bad properties of the Kelly criterion. In *The Kelly Capital Growth Investment Criterion: Theory and Practice*. Scientific Press, Singapure.

Maslov, S. and Zhang, Y.-C. (1998). Optimal investment strategy for risky assets. *International Journal of Theoretical and Applied Finance*, 1(3):377 – 387.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *The review of Economics and Statistics*, 51(3):247–257.

Mossin, J. (1968). Optimal multiperiod portfolio policies. *The Journal of Business*, 41(2):215–229.

Rising, J. K. and Wyner, A. J. (2012). Partial Kelly portfolios and shrinkage estimators. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium*, pages 1618–1622. IEEE.

Samuelson, P. (1971). The "fallacy" of maximizing the geometric mean in long sequences of investing or gambling. *Proceedings of the National Academy of Sciences*, 68(10):2493–2496.

Thorp, E. O. (2006). The Kelly criterion in blackjack sports betting and the stock market. In Zenios, S. A. and Ziemba, W. T., editors, *Handbook of Asset and Liability Management: Theory and Methodology*, volume 1 of *Handbook of Asset and Liability Management*, pages 385–428. Elsevier Science.

Wenzel, R. (2011). How Paul Samuelson really got rich. `http://www.economicpolicyjournal.com/2011/01/how-paul-samuelson-really-got-rich.html`. [Online; accessed 19-April-2013].

Ziemba, W. T. (2003). The stochastic programming approach to asset, liability, and wealth management. *Research Foundation Publications*, 2003(3):1–192.

Appendix 1. Evaluation of the approximation error by Taylor expansion

It turns out that setting the risk-free rate r to zero we can reduce our problem to the univariate case, simplifying both reasoning and notation. Assign

$$t(\mathbf{u}) := r + \sum_{k=1}^{n} u_k(r_k - r) = \sum_{k=1}^{n} u_k r_k \tag{6}$$

Then we can rewrite (2) as

$$\mathbb{E}\left[\ln(1 + t(\mathbf{u}))\right] \tag{7}$$

and we want to maximize this expression. According to our approach we replace the objective function with its Taylor approximation about $\mathbf{u_0} = (0, \ldots, 0)$. But for $r = 0$ it follows that $t(\mathbf{u_0}) = 0$ thus we expand $\ln(1 + t(\mathbf{u}))$ about $t = 0$ and maximize

$$\mathbb{E}\left[t(\mathbf{u}) - \frac{1}{2}t^2(\mathbf{u})\right] \tag{8}$$

Note that current risk-free rates are extremely low, occasionally they are even negative[16] thus the assumption $r = 0$ is rather a stylized fact than a mathematical trick.

[16] For instance, some German bonds were issued with negative yield, see http://www.ft.com/cms/s/0/b414c77a-abcf-11e1-a8a0-00144feabdc0.html

We could have tried to estimate

$$\left\| \max_{\mathbf{u}} \mathbb{E}\left[\ln\left(1+t(\mathbf{u})\right)\right] - \max_{\mathbf{u}} \mathbb{E}\left[t(\mathbf{u}) - \frac{1}{2}t^2(\mathbf{u})\right] \right\| \tag{9}$$

where $\|.\|$ may be any feasible norm, e.g. the supremum norm, which represents the approximation error in the worst case. However, in the context of the portfolio optimization it is *not* our goal, since we search for $\mathbf{u}^\star$ that maximizes (8) and then put it into (7) in hopes to maximize the expected portfolio growth rate. Let $\mathbf{u}^{\star\star}$ be the point in which (7) achieves its *true* maximum. Our problem is that in general $\mathbf{u}^\star \neq \mathbf{u}^{\star\star}$ hence we need to analyse

$$\left|\mathbb{E}\left[\ln\left(1+t(\mathbf{u}^{\star\star})\right)\right] - \mathbb{E}\left[\ln\left(1+t(\mathbf{u}^{\star})\right)\right]\right| \tag{10}$$

First of all we assume that $\mathbb{E}[\ln(1+t(\mathbf{u}))]$ exists and is finite for each $\mathbf{u}$ s.t. $u_k \in [0,1]\ \forall k$. Further we note that $\ln(1+t(\mathbf{u}))$ can be expanded in a Taylor series only on the interval $(-1,1]$. Additionally, Exhibit 5 shows that the Taylor approximation up to $O(t^3)$ is more or less accurate only for relatively small $|t|$.

Let us assume that $|r_k| \leq R\ \forall k$, where R is fixed. Under no leverage and no short selling conditions it follows that $|t| \leq R$. Indeed, for arbitrary but fixed $u_k \in (0,1]\ \forall k$, t is increasing in r_k. The worst case (both in the sense of the porfolio yield and the approximation error) takes place if $r_1 = r_2 = \ldots = r_n = -R$ and we are fully invested in risky assets, i.e. $\sum_{k=1}^n u_k = 1$. But in this worst case it immediately follows from (6) that

$$t = r + \sum_{k=1}^{n} u_k r_k - r\sum_{k=1}^{n} u_k = \sum_{k=1}^{n} u_k r_k = -R \tag{11}$$

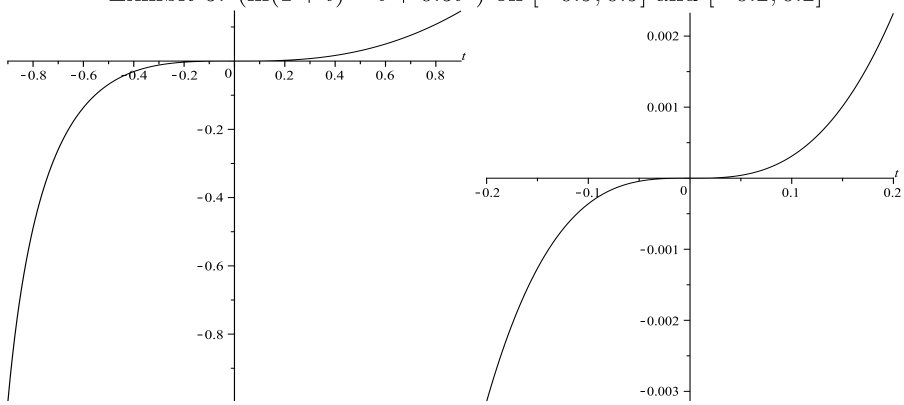
For instance, take $R = 20\%$, which is not implausible for the case of the medium-term trading in blue chips. From Exhibit 5 we see that in this case the approximation error $l < 0.0035$. Analogously we can conclude that $t = 0.2$ in the best case[17], i.e. when $r_1 = r_2 = \ldots = r_n = 0.2$ and $\sum_{k=1}^n u_k = 1$

So far we have considered only a single realization of the random returns but it was the worst case. Hence by any other realization it still holds that the approximation error $l \leq 0.0035$.

We maximize the expected portfolio growth rate but the expectation (in case of discrete probability distribution) is just the sum of outcomes weighted by their probabilities. If the returns have a continuous distribution, we can always turn it to the discrete one approximating (arbitrarily precisely) the probability density by a step function. Both the logarithm and its Taylor approximation up to the 3rd term are convex functions. Thus their expectations will also be convex (as the sums of convex functions). Exhibit 6 demonstrates the simplest case: zero risk-free rate and only one risky asset whose return can take only two values (in this case 0.5 and -0.35) with equal probability 0.5. Denote the

[17] This case will really be the best in the sense of portfolio yield as long as the risk-free rate $r \leq 0.2$. In the sense of approximation error it is still a bad case but not as bad as when $r_1 = r_2 = \ldots = r_n = -0.2$.

Exhibit 5: $(\ln(1+t) - t + 0.5t^2)$ on $[-0.9, 0.9]$ and $[-0.2, 0.2]$



objective function with $f(t(\mathbf{u})) := \mathbb{E}\left[\ln\left(1 + t(\mathbf{u})\right)\right]$ and its Taylor approximation with $g(t(\mathbf{u})) := \mathbb{E}\left[t(\mathbf{u}) - \frac{1}{2}t^2(\mathbf{u})\right]$.
Let l be the maximal approximation error on the interval $[a, b]$, in our case $[a, b] = [-0.2, 0.2]$. In the trivial case when $\mathbf{u}^{\star\star} = \mathbf{u}^{\star}$ it is obvious that $|f(t(\mathbf{u}^{\star\star})) - g(t(\mathbf{u}^{\star\star}))| \leq l$.

In the general case the following combinations are possible:

$$
\begin{aligned}
f(t(\mathbf{u}^{\star\star})) \geq g(t(\mathbf{u}^{\star\star})) \qquad & f(t(\mathbf{u}^{\star})) \geq g(t(\mathbf{u}^{\star})) && (12)\\
f(t(\mathbf{u}^{\star\star})) \geq g(t(\mathbf{u}^{\star\star})) \qquad & f(t(\mathbf{u}^{\star})) \leq g(t(\mathbf{u}^{\star})) && (13)\\
f(t(\mathbf{u}^{\star\star})) \leq g(t(\mathbf{u}^{\star\star})) \qquad & f(t(\mathbf{u}^{\star})) \leq g(t(\mathbf{u}^{\star})) && (14)\\
f(t(\mathbf{u}^{\star\star})) \leq g(t(\mathbf{u}^{\star\star})) \qquad & f(t(\mathbf{u}^{\star})) \geq g(t(\mathbf{u}^{\star})) && (15)\\
& && (16)
\end{aligned}
$$

For the first combination we have

$$f(t(\mathbf{u}^{\star\star})) - g(t(\mathbf{u}^{\star\star})) \leq f(t(\mathbf{u}^{\star})) - g(t(\mathbf{u}^{\star})) \leq l \tag{17}$$

(since by definition $\mathbf{u}^{\star\star}$ maximizes f and $\mathbf{u}^{\star}$ maximizes g).

For the second combination it holds that $f(t(\mathbf{u}^{\star\star})) - g(t(\mathbf{u}^{\star\star})) \leq l$ and $g(t(\mathbf{u}^{\star})) - f(t(\mathbf{u}^{\star})) \leq l$ thus

$$f(t(\mathbf{u}^{\star\star})) - f(t(\mathbf{u}^{\star})) \leq [f(t(\mathbf{u}^{\star\star})) - f(t(\mathbf{u}^{\star}))] + [g(t(\mathbf{u}^{\star})) - g(t(\mathbf{u}^{\star\star}))] \leq 2l \tag{18}$$

The third and the fourth combinations are analogous to, resp., the first and the second. So we proved that even in the worst case the approximation error does not exceed $2l$. The assessment whether this is large or not depends on the maximum of the expected growth rate. E.g. in our case of $r_k \in [-20\%, 20\%]$ $\forall k$ we have that $2l \leq 0.007 = 0.7\%$. But in general we can say nothing about the expected growth rate of the optimal portfolio. For such returns it may readily be 7% or more and then the approximation error may be disregarded. But if the expected returns of all assets are very close to zero then so will be the maximum expected growth rate and the approximation error gets relatively large. However, in this case it is probably not worth trading at all.

Exhibit 6: Left: $\ln(1+0.5u)$ and $\ln(1-0.35u)$. Right: $0.5[\ln(1+0.5u)+\ln(1-0.35u)]$ and its Taylor approximation. (We write simply u instead of $t(u)$).