Humboldt-Universität zu Berlin

Ladislaus von Bortkiewicz Chair of Statistics

Master Thesis by

Niels Wesselhöfft (MN: 553363)

The Kelly Criterion:

Implementation, Simulation and Backtest

In partial fulfillment of the requirements for the degree: Master in Statistics (M.Sc.)

First Advisor:	Prof. Dr. Brenda Lopez Cabrera
Second Advisor:	Prof. Dr. Wolfgang K. Härdle

February 28, 2016

Contents

List of Figures

List of Tables

Abstract

In this thesis the Kelly growth-optimum criterion, as one strand of portfolio theory, besides the widely used mean-variance approach, is implemented and tested in a simulation study and on empirical basis. The main objective of Kelly (1956) is the maximization of the expected logarithm of growth, leading to, as Breiman (1961) proves, the asymptotically optimal strategy in an i.i.d. world, which can be extended to arbitrary and time-dependent returns.

Under different parametric distribution assumptions for the outcomes, closed-form solutions for the growth-optimum strategy will be presented. Within a simulation study it will be shown that, sampling from the assumed data generating process naturally supports the asymptotic outperformance. As the assumption of the known process is loosened and the Kelly strategy needs to be implemented upon past, limited data, draw-down risks are increased and the portfolio maximizing the expected logarithm of end wealth is shifted to fractional Kelly bets. This holds for the empirical out-of-sample test. As the main statistical focus remains the improvement of the moment estimates in terms of errors, conditional moments are estimated by econometric time-series models. Setting the conditional mean forecast to zero if the conditional volatility forecasts surpasses the unconditional volatility, leads to a cancellation of positions in times of high uncertainty, which, one the one hand, decreases errors in the mean estimate and on the other hand, decreases portfolio draw-downs substantially.

Introduction

Given a set of investment opportunities, how should the investment weights should be chosen, in order to have more wealth than anyone else at the end of the investment period, assuming equal initial endowments?

The Kelly growth-optimum strategy is a betting scheme for an investor or a gambler, who seeks to asymptotically maximize his growth rate of capital. This evolutionary stable strategy outperforms any other significantly different strategy. In contrast to the theoretical feasibility, I will show for the example of financial market returns that, in the presence of finite data, the strategy exhibits exhibits significant risks in the short and medium term. Especially draw-down probabilities are not acceptable for risk-averse investors.

After motivating the Kelly Criterion for standard parametric assumptions for the outcomes of the bets, such as Bernoulli, Uniform, Normal or Student-T, I will test the criterion in a simulation study in order to examine the riskiness of the strategy given finite data, with and without knowledge of the parameters of the underlying stochastic process. Accordingly, the strategy is tested for empirical markets. In order to improve the results under the assumption of stationarity, I replace maximum likelihood estimators for the first two moments with time-series estimators, leading to improved out-of-sample performance for the tested markets. Furthermore, a two-regime approach for the conditional mean estimator is proposed, leading to a cancellation of long positions in times of high conditional volatility, being able to reduce mean square errors of the mean estimator in the out-of-sample test.

In literature, following Roll (1973), there are two main strands dealing with the management of risks, thus, the allocation of wealth into a portfolio. On the one hand the famous and wide-spread two moments, mean-variance approach of Markowitz (1952), Tobin (1958), Sharpe (1964) and Lintner (1965) and on the other hand, the Kelly growth-optimum approach by e.g. Kelly (1956), Breiman (1961) and Thorp (1971).

Originally, Kelly (1956) showed, in the context of information theory, that the highest asymptotical growth rate of capital equals the rate of transmission over a channel. Statistically, given a series of bets with Bernoulli outcomes, betting the Kelly strategy implies betting the edge, outperforming any significantly different strategy in the long run. This result can be extended to arbitrary (non-) stationary distributions (Algeot and Cover, 1988). Although Kelly was initially understood in the context of information theory and later gambling, Latané (1959), independent of Kelly, widened the field of view to the inter-temporal investment problem. Breiman (1961) proves, in a general, multivariate i.i.d. setting that no strategy, significantly differing from the Kelly-strategy, can asymptotically outperform the Kelly growth-optimum strategy.

Consequently, Thorp (1971) gathers the results from Kelly (1956) and Breiman (1961) and applies those to gambling as well as basic investment opportunities. Independently, Hakansson (1971) develops an inter-temporal investment-consumption model, not significantly different from Kelly, through which he shows that the fraction vector does not depend on the wealth level itself and he proposes, under the log-normal utility assumption, that a serial correlation of returns does not constrain the optimality of the solution. In the last part of this chapter, the results of Roll (1973) are gathered, who tested the log-optimal model on an empirical basis, relating it to the famous Sharpe-Lintner model.

From the view-point of asymptotic optimality, the theorems of Breiman (1961) are extended with rising generality: Finkelstein and Whitley (1981) expand the optimality the Kelly Criterion to arbitrary distributed returns. Subsequently, Barron and Cover (1988) clarify, in the theoretical context of information theory, the bounds on additional information and Algeot and Cover (1988), along with Thorp (2006), prove the asymptotic optimality for arbitrary, time-dependent returns, representing the highest form of generality in literature. Starting from there, the Kelly Criterion is examined in the context of portfolio theory, including further risk constraints. In contrast, I aim to test the strategy in its original intention.

For numerical computations and visualizations the software Matlab is going to be used. The empirical data-set in daily frequency is downloaded from yahoo finance and covers eleven different markets with a time span from 2005 to 2015. Covering the financial and the European dept crises, the asset prices underlie rapid changes in value.

The first chapter of the Master Thesis starts to implement the Kelly Criterion under different parametric distribution assumptions. Following MacLean, Thorp, Zhao, and Ziemba (2010), the special cases of the Bernoulli distribution and the normal distribution are analyzed in a simulation study in chapter two. Over different betting horizons, I will show that the Kelly bet outperforms asymptotically, as long as the true parameters of the data generating process are known. Within the simulation framework, it will be seen that the major favorable properties of the Criterion are stressed when past-dependent maximum-likelihood estimators, given finite data are, used.

In the third chapter, this result will hold for the out-of-sample test in financial markets. In a mean square error analysis for the mean estimators for different markets, the hypothesis of Louis Bachelier, that the best predictor for the value tomorrow is the value today, in line with weak market efficiency, will be tested. As financial markets are not perfect, statistical and econometric methods are utilized to estimate future moment forecasts of the according return distribution. Subsequently I will introduce the modelling of (squared) returns in a time-series framework with according forecasts for the first two moments. As the Kelly Criterion performs only well if the underlying process and especially the first moment can be forecasted sufficiently, a two-regime-approach is proposed. If the conditional volatility forecast exceeds a scaled unconditional volatility estimate, I assume the market to be in a risky state, in which the whole position will be squared.

Chapter 1

Methodology

In order to implement the Kelly growth-optimum strategy, closed-form solutions given a parametric assumption of the outcomes will be presented. The dealt with solutions are full Kelly solutions, meaning that this is the original strategy, which may assign to put the total wealth or even more in one market, excluding additional risk constraints. Risk constraints as part of the general optimization problem will be dealt with in a separate chapter, leading to partial Kelly solutions.

Shortly reviewing the central result of Kelly (1956), the situation under binary events is extended by including fair odds and a minimum bet in the expected growth rate Thorp (1984). Whereas Bicksler and Thorp (1973) extend the approach of Kelly to uniform returns, Merton (1992) derives, under the classic assumption of the geometric Brownian Motion for price changes, a continuous life time portfolio strategy. Due to the fact that log-normal prices are assumed, the first two moments of the returns need to be estimated. This leads, as the true distribution parameters are not known, inevitable to errors, which are examined by Chopra and Ziemba (1993) in a relative manner. Thorp (2006) shows that the results hold for the investment fraction when the outcome is symmetrically distributed and the wealth process is approximated by a second-order Taylor approximation. As especially financial returns are not normally distributed, Osorio (2008) generalizes the result of Merton (1992) for the Student-T distribution and distributions, which are not heavy-tailed.

1.1 Bernoulli trials (Kelly, 1956)

Kelly (1956) introduced the log-utility function in the context information theory and implicitly gambling. Given Bernoulli-trials (Binary Channel) he shows that the use of the logarithmic utility function maximizes the long run growth rate. But also in the short-term, he exhibits that the utility function is myopic by period-by-period maximization, only upon the current value of initial capital (MacLean, Thorp, and Ziemba., 2011).

Suppose a favorable game with winning probability $\frac{1}{2} < p \leq 1$ and outcome 1 and losing probability $q = 1 - p$ with outcome -1 and even odds, starting with initial wealth W_0. The wealth after n trials, betting fraction f of the initial capital (in %), is given by

$$W_n = W_0(1+f)^m(1-f)^{n-m}. \tag{1.1}$$

The exponential rate of asset growth per trial, equalling the logarithm of the geometric mean, can be restated according to formula 1.1 by:

$$\begin{aligned} G_n(f) &= \log\left(\frac{W_n}{W_0}\right)^{\frac{1}{n}} = \log\left[(1+f)^{\frac{m}{n}}(1-f)^{\frac{n-m}{n}}\right] \\ &= \left(\frac{m}{n}\right)\log(1+f) + \left(\frac{n-m}{n}\right)\log(1-f). \end{aligned} \tag{1.2}$$

Consequently, the expected growth rate coefficient is given by

$$\mathrm{E}[G_n(f)] = g(f) = p \cdot \log(1+f) + q \cdot \log(1-f) = \mathrm{E}[\log(W)], \tag{1.3}$$

where p is the winning probability and q the losing probability. Maximizing $g(f)$ with respect to f leads to

$$\begin{aligned} g'(f) &= \left(\frac{p}{1+f}\right) - \left(\frac{q}{1-f}\right) = \left[\frac{p-q-f}{(1+f)(1-f)}\right] = 0 \\ &\Leftrightarrow f = f^* = p - q, \quad p \geq q > 0. \end{aligned} \tag{1.4}$$

The second derivative according to f shows that $f = f^*$ is the unique maximum of the function $g(f^*) = p \cdot \log(p) + q \cdot \log(q) + \log(2) > 0$.

$$g''(f) = -\left[\frac{p}{(1+f)^2}\right] - \left[\frac{q}{(1-f)^2}\right] < 0 \tag{1.5}$$

In conclusion Kelly states that

Theorem 1.1.1 (Kelly)**.** *the optimal fraction, under Bernoulli trials, which should be invested per trial, is $f^* = p - q$, the edge. This fixed fraction strategy maximizes the expected value of the logarithm of capital at each trial (Kelly, 1956).*

As Thorp (1971) points out later, maximizing the expected logarithm of wealth $\mathrm{E}[\log(W_t)]$ is equivalent to maximizing the exponential rate of growth per time period $g(f)$.

Even odds - Kelly (1956)

Figure 1.1 plots the surface of the expected growth rate as a function of the fraction and the winning probability. The green parts of the surface are fraction-probability combinations through which $g(f,p)$ remains positive. The four data tips represent the optimal sets of portfolios for $p = [0.5, 0.6, 0.8, 0.95]$. The optimally invested fraction is a linear function on the winning probability.

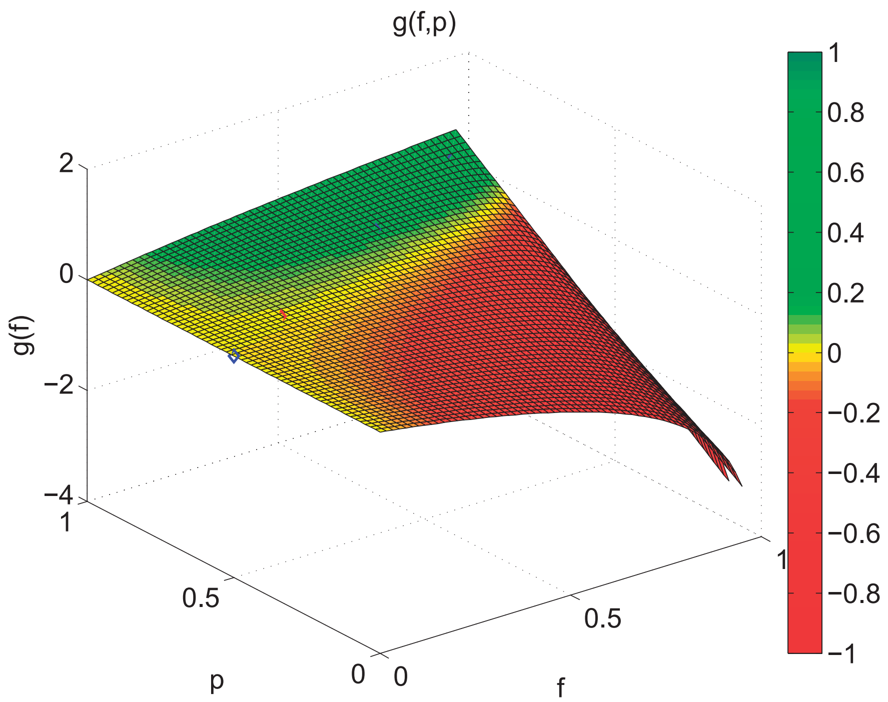


Figure 1.1: Logarithm of the geometric growth rate depending on fraction and winning probability

Plotting the expected growth rate as a function of the fraction f with known winning probability, the surface $g(f,p)$ is reduced to the function $g(f)$ for $p = 0.6$, which shall be maximized (Figure 1.2). The optimization algorithms in Matlab and in general aim to minimize convex or non-convex functions, but the location of the maximum of $g(f)$ is the location of the minimum of -$g(f)$. For illustration purposes the original $g(f)$ was plotted and the maximum was market with a data tip.

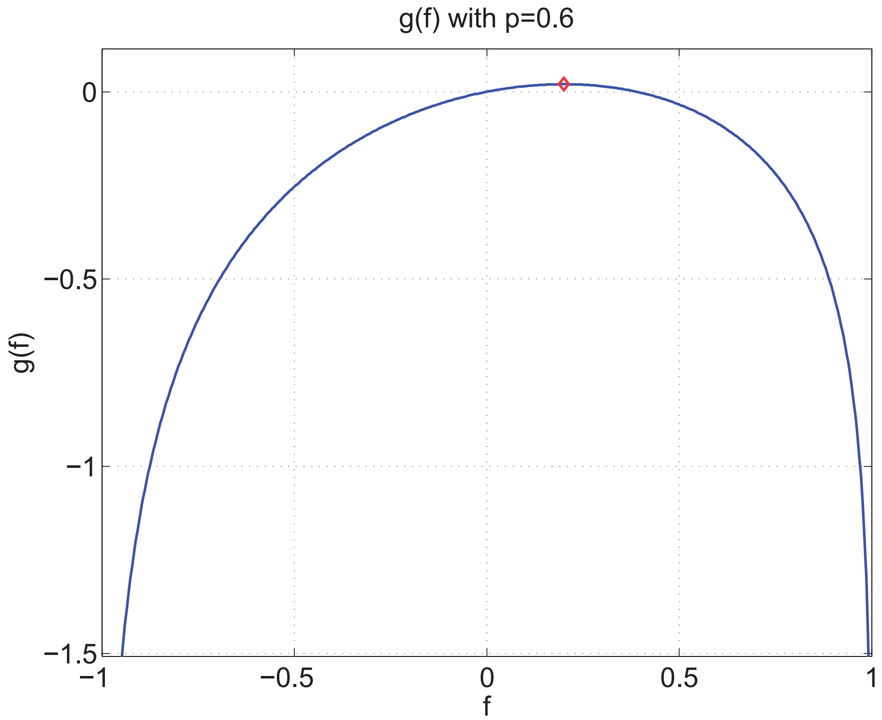


Figure 1.2: Logarithm of the geometric growth rate depending on fraction with fixed winning probability at 60%

Of course, the numerical solution $f^* = 0.2$ does not deviate from the analytical solution. Moreover, Figure 1.2 indicates that the expected growth rate cannot be negative if $f \leq f^*$ for $f \geq 0$. But if the winning probability is significantly overestimated, $g(f)$ becomes negative. This is also a starting point for risk-averse partial Kelly solutions.

Uneven odds - Thorp (1984)

Now assume that the odds are not even anymore, so $o \in \mathbb{R}^+$.[1] Thereafter, a game is favorable if $po - q > 0$, leading to a variation of the logarithm of the geometric growth rate

$$go(f, o) = p\log(1 + of) + q\log(1 - f), \tag{1.6}$$

which is maximized, using ordinary calculus, by

$$f^* = \frac{op - q}{o} = \frac{edge}{odds}. \tag{1.7}$$

From the analytical solution, $\partial f^*/\partial o > 0$ and $\partial^2 f^*/\partial o^2 < 0$. In order to examine the effect of changing odds on the fraction, the surface of the function $go(f, o)$ with $p = 0.6$ is plotted in Figure 1.3. The five data tips represent the optimal sets of portfolios for $o = [1 : 5]$. The invested fraction is not linear in the odds.

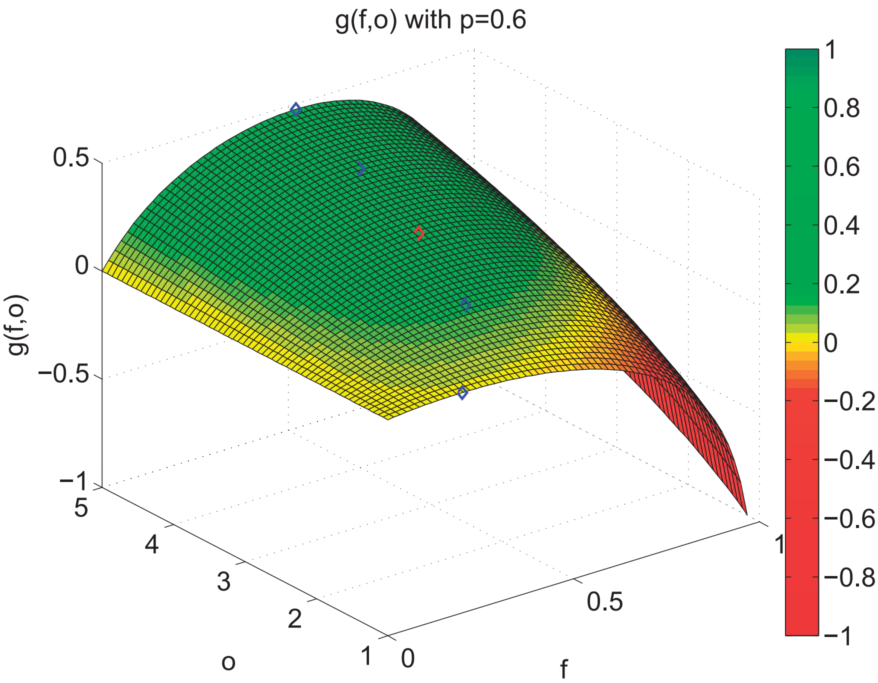


Figure 1.3: Logarithm of the geometric Growth Rate depending on fraction and odds with fixed winning probability at 60%

[1] If the odds for winning are 1:1, the probabilities for both events, winning and losing, are even, so is the payoff. If the odds are fair and for example 2:1, the probability of losing is two times the probability of winning, so the payoff for winning is two times the payoff for losing.

Thus, holding the odds constant at three leads to the maximization of $g(f)$ given $o = 3$ and $p = 0.6$ (Figure 1.4). The observation from Figure 1.2, that solely overaggressive betting leads to a negative expected growth rate, holds.

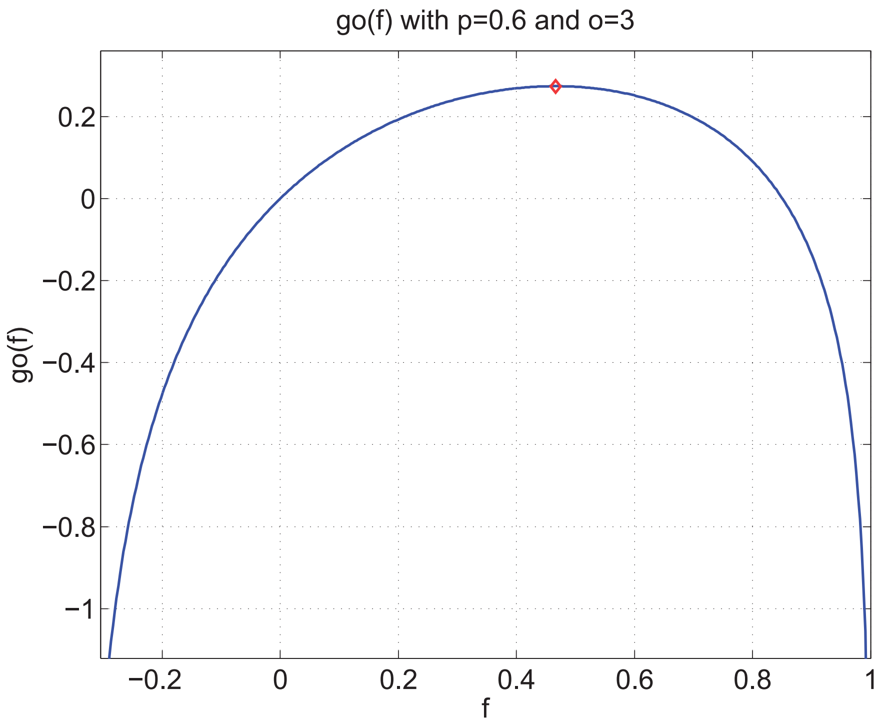


Figure 1.4: Logarithm of the geometric growth rate depending on fraction with fixed odds three and winning probability 60%

Even odds and a minimum bet - Thorp (2006)

There are many games, such as Blackjack or Poker, in which it would be seen curious, or where it is just not possible, if one would play solely favorable situations. Hence, there is often a minimum bet $a \in [0, 1]$ involved. Let f be the bet on the favorable situation, where an edge is given, and af be the bet on unfavorable situations. Using $\mathrm{P}(x)$ as notation for the probability for the favorable game situation, Thorp (2006) modifies the Kelly growth rate in the following way:

$$ga(f, a, \mathrm{P}(x)) = \mathrm{P}(x)\,[p\log(1+f) + q\log(1-f)] + [1-\mathrm{P}(x)]\,[q\log(1+af) + p\log(1-af)] \quad (1.8)$$

The analytical solution will not be handy. Therefore the focus will lie on the numerical results. To visualize the function which shall be optimized, we restrict on $\mathrm{P}(x)$ and the winning probability, which we assume is given in a certain game. Changing the expected growth rate from Thorp (2006) by

i. approximating $\mathrm{P}(x)$ with $(1/\#players)$,

ii. assuming two players, so that $\mathrm{P}(x) = 0.5$ and

iii. an edge of 20%

leads to the maximization of the expected growth rate

$$ga(f, a) = 0.5\,[0.6\log(1 + f) + 0.4\log(1 - f)] + 0.5\,[0.4\log(1 + af) + 0.6\log(1 - af)]\,. \tag{1.9}$$

In the first place, the surface for the expected growth rate was plotted as a function of the fraction and the minimum bet as a percentage of the fraction (Figure 1.5). The four data tips represent the optimal fractions for $a = [0, 0.05, 0.2, 0.5]$. The Kelly gambler reduces the invested fraction, starting with no minimum bet, from 0.2 to zero as a tends to one (Thorp, 2006). From the practical point of view, under approximation of $\mathrm{P}(x)$, the fractions with varying a should be calculated beforehand, due to the fact that the maximization $ga(f, a)$ has to be done numerically. For $a = 0.05$, the surface $ga(f, a)$ reduces to the function $ga(f)$ with $p = 0.6$, which was plotted with a data tip at the maximum of the function (Figure 1.6).

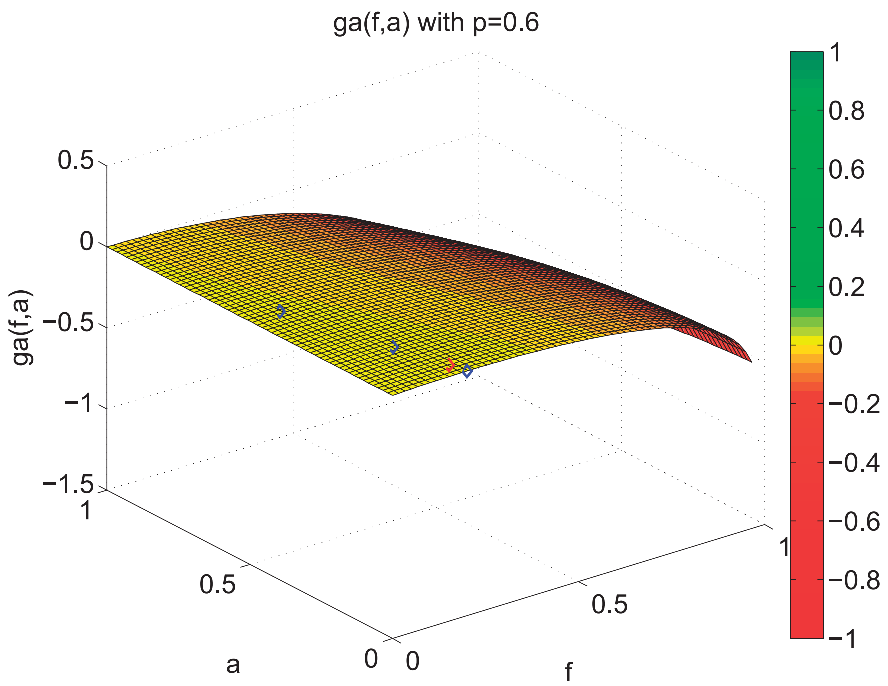


Figure 1.5: Logarithm of the geometric growth rate depending on fraction and minimum bet with fixed winning probability 60%

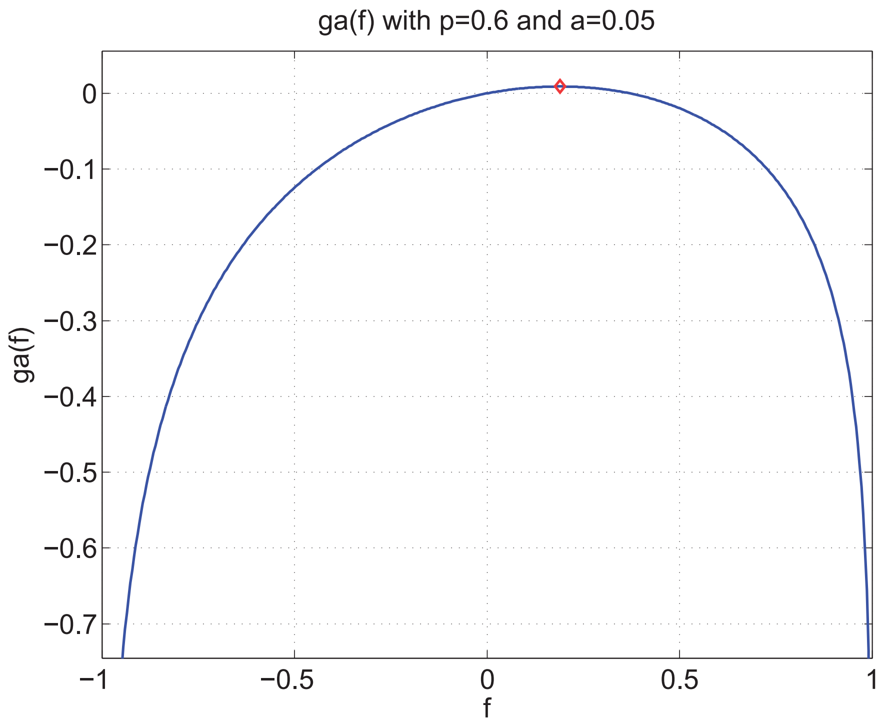


Figure 1.6: Logarithm of the geometric growth rate depending on fraction with fixed minimum bet 5% and winning probability 60%

Moreover, it is important to indicate that the Kelly fraction further reduces when $\mathrm{P}(x)$ decreases, respectively when the number of players increases (see Table 1.1).

Players	a	0	0.2	0.4	0.6	0.8	1
2	f^*	0.2	0.155	0.104	0.059	0.024	≈ 0
3	f^*	0.2	0.112	0.03	≈ 0	≈ 0	≈ 0
4	f^*	0.2	0.072	≈ 0	≈ 0	≈ 0	≈ 0

Table 1.1: Optimally invested fraction depending on the number of players and the minimum bet

1.2 Uniform returns (Bicksler and Thorp, 1973)

Presume a market of one risky asset, where the return is uniformly distributed on lower bound a and upper bound b, so $x \sim U(a,b)$. Additionally, the investor can buy a risk free asset with constant return r. So, the wealth W_n can be phrased as

$$W_n = W_0\left[1 + r + f(x - r)\right]. \tag{1.10}$$

The exponential growth rate for this opportunity is

$$G(f) = \log\left(\frac{W_n}{W_0}\right) = \log\left[1 + r + f(x - r)\right] \tag{1.11}$$

and the maximization problem can be formulated as

$$\begin{aligned} \max_f \mathrm{E}\left[G(f)\right] &= \max_f g(f) \\ \Leftrightarrow \max_f \mathrm{E}\left\{\log\left[1 + r + f(x - r)\right]\right\}. \end{aligned} \tag{1.12}$$

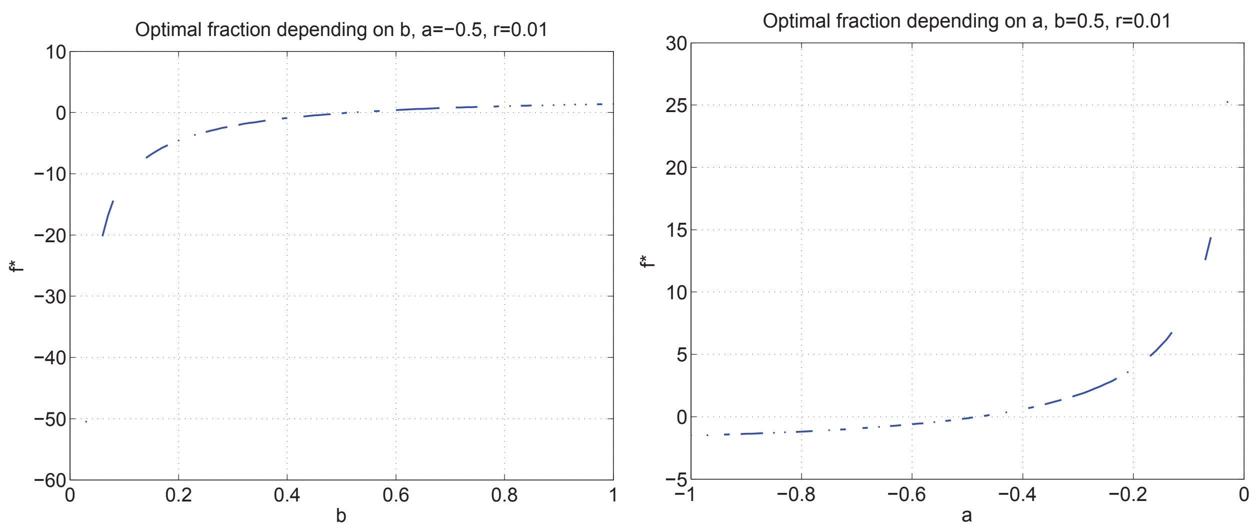
The first order condition is given and rewritten by

$$\begin{aligned} \int_a^b \left[\frac{x - r}{1 + r + f(x - r)}\right] \times \left[\frac{1}{b - a}\right] dx &= 0 \\ \Leftrightarrow f(b - a) &= (1 + r)\log\left[\frac{1 + r + f(b - r)}{1 + r + f(a - r)}\right] \\ \Leftrightarrow \left[\frac{1 + r + f(b - r)}{1 + r + f(a - r)}\right]^{1/f} &= \exp\left[\frac{b - a}{1 + r}\right]. \end{aligned} \tag{1.13}$$

As seen, there is no suitable closed form solution, so the optimal fraction f has to be calculated using numerical procedures as Newton Raphson or Bisection method.

For exemplary analysis I set the bounds of the uniform r.v. x as $[a, b] = [-0.5, 0.5]$ and the risk free rate to r to 0.01. Following the Kelly strategy for those parameters, would imply to short the risky asset by 12.12% of the initial capital, $f^* = -0.1212$. Note, as the risk free rate increases, the optimally invested fraction in the risky asset decreases.

Letting the upper bound b of the r.v. increase from 0.01 to 1, the optimal fraction increases non-linearly. Letting the lower bound a of the r.v. increase from 0.01 to 1, the optimal fraction decreases non-linearly (Figure 1.7). Missing values indicate numerical instabilities.



(a) (b)

Figure 1.7: Optimally invested fractions under uniform with changing bounds [a,b]

1.3 Log-normal prices (Merton, 1969/1992)

The goal of this section is to derive a closed-form solution for the optimal fraction under lognormal prices P_j for assets j to k, Gaussian log-returns X_j with μ_j and σ_j. The optimization problem is

$$\begin{aligned} &\max_f \mathrm{E}\left[G(f)\right] = \max_f g(f) \\ \Leftrightarrow &\max_f \mathrm{E}\left\{\log\left[1 + r + f(X - r)\right]\right\}. \end{aligned} \tag{1.14}$$

The classic continuous-time solution of the inter-temporal investment-consumption problem is mainly due to Merton (1969) and has been extended by many scientists, e.g. Browne (1997) and Browne (2000). The focus primary lies on chapter four of his book "Continuous time finance" (Merton, 1992). The crucial assumption, in order to derive the following results, is that the logarithm of the price ratio, $\log\left(\frac{P_{j,t+1}}{P_{j,t}}\right) = \log(X_{j,t})$, follows a Geometric Brownian Motion (GBM), also called Itô-process. So, the price of the risky asset j satisfies the stochastic differential equation

$$dP_{j,t} = \mu_{j,t} P_{j,t} dt + \sigma_{j,t} P_{j,t} dZ_{j,t}, \tag{1.15}$$

where $Z_{j,t}$ are standard Brownian Motions, which can be dependent. Additionally, a risk free asset with price R and risk free return $0 \leq r < \mu_j$ is assumed, evolving according to

$$dR_t = rR_t dt. \tag{1.16}$$

Consistent with the Black-Scholes-Merton approach, the parameters μ_j, σ_j and r are supposed to be constants - fixed over time - to attain one time-constant solution. The continuous wealth process, depending on the consumption C in period t, can be described as

$$dW_t = \left[\sum_{j=1}^{k} f_{j,t} \mu_j W_t\right] dt + \sum_{j=1}^{k} f_{j,t} \sigma_j W_t dZ_{j,t}. \tag{1.17}$$

For the univariate case, one risky and one risk-free asset, the wealth dynamics can be rewritten in the following form

$$dW_t = \left[(f\mu + (1-f)r)W_t - C_t\right] dt + f\sigma W_t dZ_t, \tag{1.18}$$

and the lifetime objective function, as Merton (1992) calls it, is given by

$$I[W_t, t] = \max_{C,f} \mathrm{E}\left[\int_0^T e^{-\rho t} U(C_t) dt + B(W_T, T)\right], \tag{1.19}$$

with impatience factor ρ and the Bequest valuation function at time T, concave in wealth at T. Using a Taylor approximation at t and taking expectations

$$\begin{aligned} 0 = \max_{C,f} \Bigg\{ e^{-\rho t} U(C_t) + \frac{\partial I(W_t,t)}{\partial t} + \frac{\partial I(W_t,t)}{\partial W} \left[f_t((\mu - r) + r) W_t - C_t \right] \\ + \frac{1}{2} \frac{\partial^2 I(W_t,t)}{\partial W^2} f_t^2 \sigma^2 W_t^2 \Bigg\} \equiv \phi \end{aligned} \tag{1.20}$$

with first order conditions

$$\begin{aligned} \phi_c &= e^{-\rho t} U'(C^*) - \frac{\partial I(W_t,t)}{\partial W} = 0, \\ \phi_w &= (\mu - r) W \frac{\partial I(W_t,t)}{\partial W} + \frac{\partial^2 I(W_t,t)}{\partial W^2} f^* W^2 \sigma^2 = 0. \end{aligned} \tag{1.21}$$

The solution to ϕ, the live time objective function is not trivial; hence, it needs to be simplified. Assume that

$$J(W_t,t) = e^{-\rho t} I(W_t,t). \tag{1.22}$$

Letting $T \to \infty$ the Bequest function at T, $B(W_t,T)$, falls out. The new objective function can be written as

$$J[W_t] = \max_{C,f} \mathrm{E} \left[\int_0^\infty e^{-\rho t} U(C_t) dv \right], v \in [0,\infty]. \tag{1.23}$$

Thus, the Partial Differential Equation (PDE) simplifies to the Ordinary Differential Equation (ODE)

$$\begin{aligned} 0 = \max_{C,f} \Bigg\{ U(C_t) - \rho J(W) + \frac{\partial J(W_t,t)}{\partial W} \left[f_t((\mu - r) + r) W_t - C_t \right] + \\ \frac{1}{2} \frac{\partial^2 J(W_t,t)}{\partial W^2} f_t^2 \sigma^2 W_t^2 \Bigg\}, \end{aligned} \tag{1.24}$$

which is no longer a function of time, due to the fact that dt fell out. Poon (2010) describes this step as a "key development in solving the life time consumption decision".

Remembering Thorp (1971), the main aim is to present optimal portfolio strategies under the log-utility function in a normative way: For the case of CRRA (Constant Relative Risk Aversion). The isoelastic marginal utility is given by

$$U(C) = \frac{1}{\gamma}C^{\gamma}, \tag{1.25}$$

with relative risk aversion (RRA)

$$RRA = -\frac{U''(C)}{U'(C)}C = \frac{-(\gamma-1)C^{\gamma-2}}{C^{\gamma-1}}C = 1-\gamma, \tag{1.26}$$

which is a constant, therefore, constant RRA. If $U(C) = \log(C)$, then $\gamma = 0$ and $RRA = 1$. For the isoelastic case, substituting the RRA into the the FOC ϕ_c gives

$$\begin{aligned} e^{-\rho t}(C^*)^{\gamma-1} = I'(W) \Leftrightarrow C^* = \left[e^{-\rho t} I'(W)\right]^{\frac{1}{\gamma-1}}, \\ f^* = -\left(\frac{\mu-r}{\sigma^2}\right) W \frac{J'(W)}{J''(W)}. \end{aligned} \tag{1.27}$$

For the $T \to \infty$, the optimal decision rules can be rewritten as

$$C^* = J'(W)^{\frac{1}{\gamma-1}}, \tag{1.28}$$

$$f^* = -\left(\frac{\mu-r}{\sigma^2}\right) W \frac{J'(W)}{J''(W)}. \tag{1.29}$$

Following the solution of J(W) with $\frac{v^{\gamma-1}}{\gamma}W^{\gamma}$ the following theorem can be stated.

Theorem 1.3.1. *Univariate Solution*
Assuming that the infinitesimal price changes follow a GBM under an isoelastic marginal utility $U(C) = \frac{1}{\gamma}C^{\gamma}$ with CRRA, Merton (1969) shows that the optimal consumption and investment rules in the infinite time case are given by

$$C^*_{\infty,t} = \left[\frac{\rho}{1-\gamma} - \gamma\left(\frac{(\mu-r)^2}{2\sigma^2(1-\gamma^2)} + \frac{r}{1-\gamma}\right)\right] \tag{1.30}$$

$$f^*_{\infty} = \left(\frac{\mu-r}{\sigma^2}\right)[1-\gamma]. \tag{1.31}$$

The optimal consumption and investment strategy is, consistent with e.g. Hakansson (1970) or Hakansson (1971), independent of wealth and consumption. Owing to the fact, that μ, σ and r are supposed to be constants, the optimal fraction f^*_{∞} solely depends on the risk aversion parameter γ.

Corollary 1.3.1. *Log-utility*
Assuming the logarithmic utility, so $\gamma = 0$, theorem 1 simplifies to

$$C^*_{\infty,t} = (\rho + r)W_t \tag{1.32}$$

$$f^*_{\infty} = \left(\frac{\mu-r}{\sigma^2}\right). \tag{1.33}$$

Consumption becomes a linear function of wealth and the invested fraction exhibits a close relationship to the Sharpe-Ratio (Poon, 2010).

Theorem 1.3.2. *Multivariate Solution*
Consequently, Merton (1992) extends the univariate solution to the multivariate analogue, given by

$$f_{\infty}^{*} = \Sigma^{-1}(\mu - r1), \tag{1.34}$$

with the fixed fraction vector $f_{\infty}^{*} = \begin{bmatrix} f_{j,\infty}^{*} \\ \vdots \\ f_{k,\infty}^{*} \end{bmatrix}$, the covariance matrix $\Sigma = \begin{bmatrix} \sigma_j^2 & \dots & \sigma_{j,k} \\ \vdots & \ddots & \vdots \\ \sigma_{j,k} & \dots & \sigma_k^2 \end{bmatrix}$, containing the variances of assets $j : k$ in the diagonal and the covariances around, the mean vector $\mu = \begin{bmatrix} \mu_j \\ \vdots \\ \mu_k \end{bmatrix}$ and the risk free rate r multiplied by an appropriate vector of ones $1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$.

A crucial fact to consider is that the results under the GBM-assumption, starting from Breiman (1961), up to Algeot and Cover (1988), hold also for Merton's continuous time solution, although Merton never relates to the names Kelly or Breiman. On the one hand, Merton rests his solution on the maximization of the expected logarithm of terminal capital and therefore indirectly follows Kelly (1956), but on the other hand he widens the field of utility functions and provides more general solutions.
In 2009, on the foundation of Breiman (1961) and Thorp (2006), Lv and Meister derive an identical multivariate, continuous time solution as in the presented theorem from Merton. But, they prove the existence of this "self-financing trading strategy" under multiple Ornstein-Uhlenbeck processes in complete markets (Lv and Meister, 2009). Owing to the fact that the solution to the optimal fraction is identical to Merton (1992), the details will be omitted.

1.4 A continuous approximation (Thorp, 2006)

In order to apply the Kelly Criterion for security markets, continuous distributions are relevant. As reasoned before, the goal is to maximize $g(f) = \mathrm{E}[\log(1 + fx)] = \int \log(1 + fx)d\,\mathrm{P}(x)$ with $\mathrm{P}(x)$ as probability measure for the outcomes and f as invested fraction of capital, which we aim to optimize. Constraints are $1 + fx > 0$, so log is defined and $\sum f_j = 1$. If we let outcomes x be a symmetric r.v. around $\mathrm{E}(x) = \mu$ with $Var(x) = \sigma^2$, the wealth W can be described as

$$W(f) = W_0[1 + (1 - f)r + fx] = V_0[1 + r + f(x - r)], \tag{1.35}$$

with r as return of the risk free asset. Consequently,

$$g(f) = \mathrm{E}[G(f)] = \mathrm{E}\log[W(f)/W_0] = \mathrm{E}\log[1 + r + f(x - r)]. \tag{1.36}$$

For a subdivided time interval with T independent steps

$$\frac{W_T(f)}{W_0} = \prod_{t=1}^{T}[1 + (1 - f)r + fx_t]. \tag{1.37}$$

Taking expectation and natural logarithm on both sides leads to $g(f)$. This result derives from a second order Taylor-approximation:

$$g(f) = r + f(\mu - r) - \sigma^2 f^2/2 + \mathcal{O}(n^{-1/2}). \tag{1.38}$$

For $t \longrightarrow \infty$, $\mathcal{O}(n^{-1/2}) \longrightarrow 0$, leading to

$$g_\infty(f) = r + f(\mu - r) - \sigma^2 f^2/2. \tag{1.39}$$

Differentiating $g(f)$ according to f leads to

$$\frac{\partial g_\infty(f)}{\partial f} = \mu - r - \sigma^2 f = 0 \Leftrightarrow f^* = \frac{\mu - r}{\sigma^2}. \tag{1.40}$$

The result holds for any bounded r.v. with the first two moments μ and σ^2. Note that this is the same simplified fraction as derived in Merton (1992), assuming the log-utility function in the normative sense. Thorp (2006) observes that as $t \longrightarrow \infty$, the wealth tends to a log-normal diffusion process with an underlying security with drift μ and variance rate σ^2. So, $g_\infty(f)$ is the instantaneous growth rate of depending on fraction f:

$$g_\infty(f) = r + f(\mu - r) - \sigma^2 f^2/2. \tag{1.41}$$

Betting the optimal fraction f^* leads to growth rate

$$g_\infty(f^*) = \frac{(\mu - r)^2}{2\sigma^2} + r. \tag{1.42}$$

For the given approximation, $g_\infty(f)$ is parabolic around f^* with range $0 \leq f^* \leq 2f^*$.
From the log-normality of $W(f)/W_0$ it follows that

$$\begin{aligned}\log\left(\frac{W(f)}{W_0}\right) &\sim N(\tilde{\mu}\cdot t, \tilde{\sigma}^2 \cdot t), \\ &\sim N(g_\infty(f)\cdot t, \sigma^2 f^2 \cdot t).\end{aligned} \tag{1.43}$$

Furthermore, Thorp solves for the expected portfolio growth with scaling k as $t_k g_\infty = k t_k^{1/2} Std(G_\infty(f))$ to get

$$\begin{aligned} t_k g_\infty &= \frac{k^2\sigma^2 f^2}{g_\infty} \\ \Longleftrightarrow t_k &= \frac{k^2\sigma^2 f^2}{g_\infty^2}. \end{aligned} \tag{1.44}$$

Thorp (2006) points out that the moments of x and also the risk free rate r are changing over time, leading to a changing optimal fraction f^*. Without further detail he proposes to re-estimate the optimal fraction periodically.
The multivariate case is derived analogously, giving, in accordance with Merton (1992):

$$f^* = \Sigma^{-1}(\mu - r1) \tag{1.45}$$

$$g_\infty(f^*) = r + \frac{f^{*\top}\Sigma f^*}{2}. \tag{1.46}$$

1.5 Student-T returns (Osorio, 2008)

Osorio (2008) argues that especially stock prices are not log-normally distributed as argued in Merton (1992). Excess kurtosis and skewness cannot be sufficiently captured. Having a daily return x, coming from the probability measure $\mathrm{P}(x)$, the Kelly bet implies betting the fraction f^*, which maximizes $\mathrm{E}[\log(1+fx)]$.

For the continuous return distribution $\mathrm{P}(x)$ in the $(-\infty,\infty)$ domain, the aim is to maximize

$$U(f)=\int_{-\infty}^{\infty}\mathrm{P}(x)\log(1+fx)dx. \tag{1.47}$$

In order to avoid that the log function becomes zero or negative, the integral requires a lower bound $x_1 > -1/h$. For distributions decaying sufficiently fast to zero, the truncation does not imply to cut off significant probability mass. But for heavy-tailed, slower decaying distributions this approximation may not be sufficient (Osorio, 2008).

Osorio (2008) alternates the problem formulation for unbounded probability in two steps:

i. Specify a small number $\delta << 1$, representing the tail area, which should be neglected by the log-utility. x_1 and x_2 are the according thresholds in the sense of

$$\int_{-\infty}^{x_1}\mathrm{P}(x)dx=\int_{x_2}^{\infty}\mathrm{P}(x)dx=\delta. \tag{1.48}$$

Neglecting a part of the left tail should remove the divergence in the integrand at $x=-1/h$. The right tail is truncated accordingly to make the utility function 'fair'.

ii. Optimize the optimal fraction on the (x_1,x_2) domain

$$\max_f\int_{x_1}^{x_2}\mathrm{P}(x)\log(1+fx)dx. \tag{1.49}$$

By choosing δ accordingly, we try to solve the trade-off of neglecting areas under $\mathrm{P}(x)$ (choose δ small enough) and having $x_1 > -1/f^*$ for the optimal fraction (choose δ large enough). The optimal fraction is derived by differentiating the integral with respect to fraction f:

$$\int_{x_1}^{x_2}\mathrm{P}(r)\left[\frac{\partial\log(1+fx)}{\partial f}\right]_{f^*}dx=\int_{x_1}^{x_2}\mathrm{P}(r)\left[\frac{x}{1+f^*x}\right]dx=0 \tag{1.50}$$

The factor $\frac{x}{1+f^*x}$ of the integrand on (x_1,x_2) can be approximated by a Taylor Expansion of second order, $1-f^*x$. According to Osorio (2008) this approximation holds for most practical applications including the case of the Student-T distribution with three degrees of freedom, having infinite kurtosis.

If we assume further that the neglected areas are small,

$$\begin{aligned} \left| \int_{-\infty}^{x_1} \mathrm{P}(r)x(1-fx)dx \right| << 1 \\ \left| \int_{x_2}^{\infty} \mathrm{P}(r)x(1-fx)dx \right| << 1 \end{aligned} \tag{1.51}$$

and the functions $\mathrm{P}(x)x(1-hr)$ and $\mathrm{P}(x)x/(1+hr)$ are close to each other between x_1 and x_2

$$\int_{x_1}^{x_2} \mathrm{P}(x)x(1-f^*x)dx \approx \int_{x_1}^{x_2} \mathrm{P}(x)\frac{x}{1+f^*x}dx \tag{1.52}$$

the original derivative can be approximated by

$$\int_{\infty}^{-\infty} \mathrm{P}(x)x(1-f^*x)dx \approx 0. \tag{1.53}$$

Thus, the solution for the optimal Kelly fraction can be formulated as

$$f^* = \frac{\langle x \rangle}{\langle x^2 \rangle} = \frac{\mu}{\mu + \sigma^2}, \tag{1.54}$$

where $\langle \ldots \rangle$ denotes the average for the probability density function with mean μ and variance σ^2. This result holds for all probability density functions decaying fast enough, so that the probability mass below x_1 and above x_2 is neglectable and $f^* < 1/|x_1|$ to avoid singularity in the integrand (Osorio, 2008). Empirically, $\mu << \sigma^2$, simplifying the given solution to the already given solution $f^* = \mu/\sigma^2$. For the Student-T distribution Osorio shows that the approximation, using the second order Taylor approximation, breaks down at $\nu < 4$ as kurtosis tends to infinity.

Chapter 2

Simulation study

Whereas chapter 1 gave an overview over the parametric solutions to the inter-temporal investment problem of maximizing the expected logarithm of wealth, the aim of this section is to test if the Kelly strategy Λ outperforms risk-averse and -seeking analogues in a simulation study under different distribution assumptions with and without knowledge of generally unknown distribution moments (In-sample and Out-of-sample). Hereby, I aim to test if the Kelly Criterion has favourable short- and mid-term properties, additionally to the two asymptotic theorems of Breiman (1961).

2.1 Asymptotic optimality in discrete time - Breiman (1961)

Inspired by the paper of Kelly (1956), Breiman (1961) proves three asymptotic results for the Kelly strategy of maximizing the expected logarithm of wealth in a general discrete time setting, in which returns are assumed to be inter-temporally independent and stationary (i.i.d.) in discrete time (MacLean, Thorp, and Ziemba., 2011).

Define an investment strategy Λ as the investment fractions f (in % of initial wealth) from time t to T and opportunities, say stocks, j to k as $\begin{bmatrix} f_{j,t} & \cdots & f_{j,T} \\ \vdots & \ddots & \vdots \\ f_{k,t} & \cdots & f_{k,T} \end{bmatrix}$, in vector notation $[f_{j,t}, \cdots, f_{j,T}]$.

Introducing the security price vector p_t of the assets with $\begin{bmatrix} P_{j,t} \\ \vdots \\ P_{k,t} \end{bmatrix}$, the return per unit invested x_t is given by $\begin{bmatrix} \frac{P_{j,t}}{P_{j,t-1}} \\ \vdots \\ \frac{P_{k,t}}{P_{k,t-1}} \end{bmatrix}$. Consequently, the wealth of the investor in period t can be expressed as

$$W_t = \left[f_t^\top x_t \right] W_{t-1} \tag{2.1}$$

For betting systems, W_t increases exponentially. Therefore, maximizing $\mathrm{E}[\log(W_t)]$ maximizes the rate of growth. Breiman calls a game favorable if there is a strategy, $\lim_{t\to\infty} W_t = \infty$ almost surely. The first objective of Breiman is to minimize the time to reach wealth goal $g > 1$, $T(g)$ in the form of the smallest t, such that the wealth $W_n \geq g$. $T^*(g)$ is the number of games needed to reach a wealth level larger that g given Kelly strategy Λ^*.

Theorem 2.1.1 (Breiman)**.** *In a given i.i.d. setting, there is a constant α, which is independent of Λ and g, so that*

$$\mathrm{E}[T^*(g)] - \mathrm{E}[T(g)] \leq \alpha, \tag{2.2}$$

where α is non-negative iff Λ^* is not essentially different from Λ.

Therefore, investment strategy Λ^* asymptotically minimizes the time to reach goal g, following the proposed model assumptions.

Theorem 2.1.2 (Breiman)**.** *If investor A bets according to maximize* $\mathrm{E}[\log(W_t)]$*, strategy Λ^*, and investor B bets, upon the same information set, an essentially different strategy Λ,*

$$\mathrm{E}[\log(W_t|\Lambda^*)] - \mathrm{E}[\log(W_t|\Lambda)] \longrightarrow \infty, \tag{2.3}$$

then Breiman proves that

$$\lim_{t\to\infty} \frac{W_t|\Lambda^*}{W_t|\Lambda} = \infty \tag{2.4}$$

almost surely.

Heuristically speaking, Breiman proves that, in this discrete i.i.d. setting, the investment strategy Λ^* is asymptotically optimal.

Theorem 2.1.3 (Breiman)**.** *Assuming a fixed set of opportunities, the optimal fraction set exists and is independent of time, thus, constant.*

This theorem will be crucial, as the assumptions behind it are not fulfilled. In the chapter 2, it will be shown, that e.g. estimators for unknown parameters vary over time, and therefore, the set is not fixed.

2.2 Portfolio simulations under Bernoulli trials

2.2.1 The Kelly bet from the short term to the long term, constant edge

A discrete random variable has a Bernoulli distribution if, for two possible states, the first state has probability p and the second state has the probability $q = 1 - p$. Our initial wealth is given with 100. Assuming an edge of $p - q = 4\%$, the question is how much of our initial wealth (in %) should the individual bet on the first state. Imagine a blackjack game in which the individual is able to estimate the edge precisely.[1]

The game is played $100, 1000, 10,000$ and $100,000$ times. Those different scenarios are run with $10,000$ trajectories each. The solution, which has been presented already by Kelly (1956) implies betting the edge itself, as it is optimal in the sense of maximizing the expected logarithm of wealth. This strategy shall be called full Kelly strategy. Half Kelly, being more risk averse, implies betting half the original Kelly bet and double Kelly means betting double the amount of the full Kelly bet, representing overaggressive betting. According to MacLean, Ziemba, and Blazenko (1992) overbetting is not advised since long run growth rate and also security fall.

[1] For a more comprehensive discussion around Blackjack, see Thorp (1966) and Thorp (2006)

Edge of 4%, 100 games

The descriptives of the end-wealth after 100 trials of the 10,000 trajectories are given in table 2.1. The first exemplary trajectory for each betting scheme can be seen in figure 2.1. For the scenario of 100 trials with an edge of 4% I observe that with rising betting size up to double Kelly (see table 2.1)

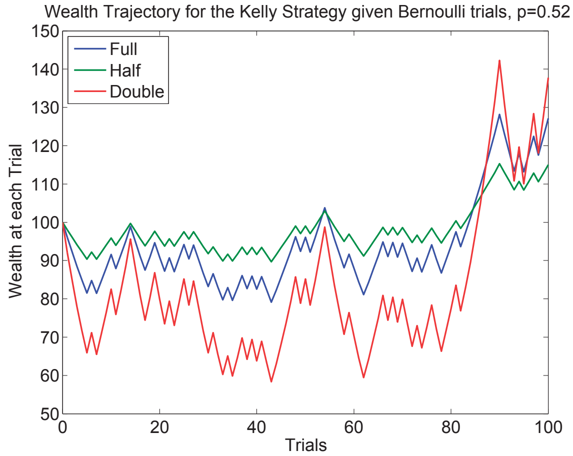


Figure 2.1: First trajectory, 100 Bernoulli trials

- the resulting wealth on average increases (line 1), although the full Kelly strategy always has the highest mean of the final logarithmic wealth,
- the standard deviation of the final wealth increases exponentially (line 2),
- the probability of ending below the initial, half initial and one tenth of initial wealth increases (line 3-5),
- the probability to hit the wealth goals of 200 and 1000 at some trial increases (line 6,8) and
- the average time to reach that goal, for those trajectories reaching the goal, decreases (line 7,9).

4% edge, 100 trials	Half Kelly	full Kelly	Double Kelly
$Mean(W_T)$	108.03	116.62	135.39
$Std(W_T)$	21.56	47.73	122.61
P$(W_T < 100)$	0.38	0.46	0.54
P$(W_T < 50)$	0	0.0270	0.1865
P$(W_T < 10)$	0	0	0.001
P$(W_t > 200)$	0.001	0.1	0.35
$Mean(T : 200)$	92.11	73.08	50.74
P$(W_t > 1000)$	0	0	0.002
$Mean(T : 1000)$	NaN	NaN	86.68

Table 2.1: Descriptives for the wealth trajectories after 100 Bernoulli trials

Edge of 4%, 1000 games

The descriptives of the end-wealth after 1000 trials of the 10,000 trajectories are given in table 2.2. The first exemplary trajectory for each betting scheme can be seen in figure 2.2. The differences to the simulations with 100 trials are, that as the number of games increases

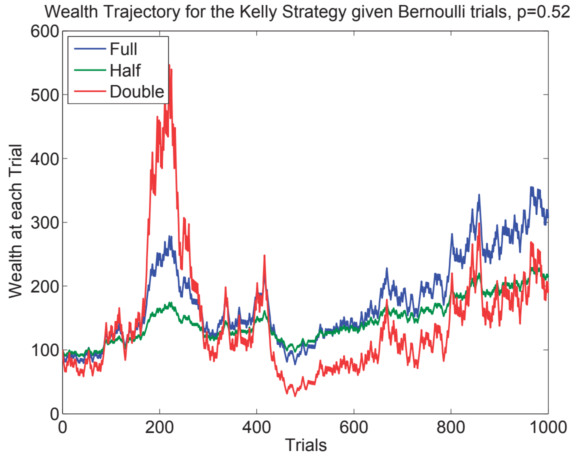


Figure 2.2: First trajectory, 1000 Bernoulli trials

- the mean-variance trade-off in end-wealth intensifies, although the full Kelly strategy always has the highest mean of the final log-wealth,
- the probability of ending below the initial wealth decreases, whereas the probability of ending up with less than half or one tenth increases and
- the probability reaching double or ten times the initial wealth increases.

4% edge, 1000 trials	Half Kelly	full Kelly	Double Kelly
$Mean(W_T)$	222.41	487.15	1980.74
$Std(W_T)$	152.17	859.54	15699.32
$\mathrm{P}(W_T < 100)$	0.18	0.27	0.51
$\mathrm{P}(W_T < 50)$	0.02	0.12	0.38
$\mathrm{P}(W_T < 10)$	0.0001	0.008	0.18
$\mathrm{P}(W_t > 200)$	0.6	0.76	0.77
$Mean(T : 200)$	540.38	342.18	205.82
$\mathrm{P}(W_t > 1000)$	0.005	0.18	0.35
$Mean(T : 1000)$	887.32	716.07	504.59

Table 2.2: Descriptives for the wealth trajectories after 1000 Bernoulli trials

Edge of 4%, 10,000 games

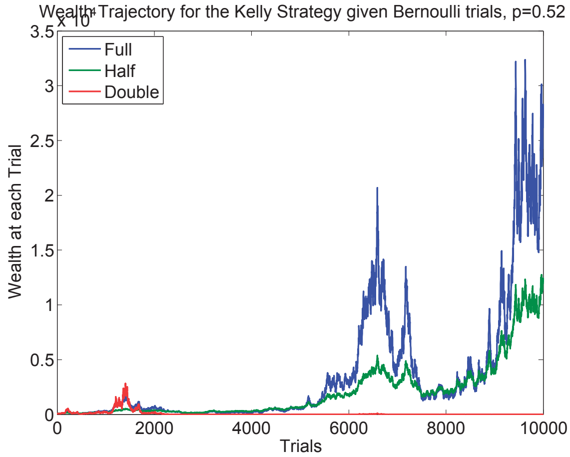


Figure 2.3: First trajectory, 10,000 Bernoulli trials

The descriptives of the end-wealth after 10,000 trials of the 10,000 trajectories are given in table 2.3. The first exemplary trajectory for each betting scheme can be seen in figure 2.3. The differences to the simulations with 1000 trials are, that as the number of games increases

- the mean-variance trade-off in end-wealth intensifies, although the full Kelly strategy always has the highest mean of the final log-wealth,
- the probability of ending below the initial wealth decreases fast for the half and full Kelly betting scheme whereas for the double Kelly strategy the downside risk increases and
- the probability reaching the specified goals increases, but it increases faster for half Kelly, then full Kelly, then double Kelly, meaning that a risk-averse strategy leads to the highest probability of reaching a goal, but needing the largest average time.

4% edge, 10,000 trials	Half Kelly	full Kelly	Double Kelly
$Mean(W_T)$	3.35e+05	2.1e+09	2.09e+13
$Std(W_T)$	3.36e+06	1.35e+11	2.02e+15
P($W_T < 100$)	0.001	0.02	0.49
P($W_T < 50$)	0.0005	0.015	0.45
P($W_T < 10$)	0	0.005	0.38
P($W_t > 200$)	0.99	0.99	0.93
$Mean(T:200)$	1151.8	822.1	684.97
P($W_t > 1000$)	0.98	0.97	0.77
$Mean(T:1000)$	3703.8	2587.2	2086.8

Table 2.3: Descriptives for the wealth trajectories after 10,000 Bernoulli trials

Edge of 4%, 100,000 games

The descriptives of the end-wealth after 100,000 trials of 2,000 trajectories are given in table 2.4. The first exemplary trajectory for each betting scheme can be seen in figure 2.4. The differences to the simulations with 10,000 trials are, that as the number of games increases

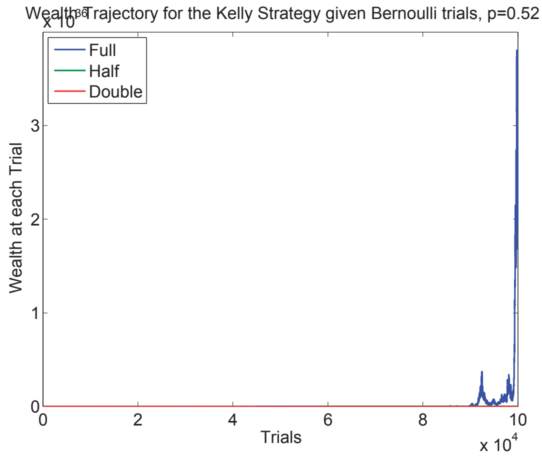


Figure 2.4: First trajectory, 100,000 Bernoulli trials

- the full Kelly portfolio outperforms its risk-averse and risk-seeking counterpart,
- the probability of ending below the initial wealth is zero for the half and full Kelly betting scheme whereas for the double Kelly strategy remains with a probability of around one half that the capital will be below a tenth of the original wealth and
- the probability reaching double or ten times the initial wealth is one for half and full Kelly; the smallest average time to reach those goals is given by the full Kelly strategy.

4% edge, 100,000 trials	Half Kelly	full Kelly	Double Kelly
$Mean(W_T)$	4.7311e+34	1.7076e+53	1.4887e+38
$Std(W_T)$	1.9787e+36	7.62e+54	6.6577e+39
$\mathrm{P}(W_T < 100)$	0	0	0.5145
$\mathrm{P}(W_T < 50)$	0	0	0.5
$\mathrm{P}(W_T < 10)$	0	0	0.47
$\mathrm{P}(W_t > 200)$	1	1	0.97
$Mean(T : 200)$	1118.9	876.32	2286.3
$\mathrm{P}(W_t > 1000)$	1	1	0.92
$Mean(T : 1000)$	3837.3	2808.9	6530.5

Table 2.4: Descriptives for the wealth trajectories after 100,000 Bernoulli trials

2.2.2 Influence of the edge-size

The edge size influences the results of the portfolio simulations. As the edge increases

- the mean-variance trade-off intensifies for given trials, although the full Kelly strategy always has the highest mean of the final log-wealth,
- draw-down probabilities decrease when betting below full Kelly and draw-down probabilities increase when over-betting,
- the probability of reaching a specific target is, also for a lower number of trials, optimal for the full Kelly portfolio,
- the time to reach that specific target is minimized for a significantly lower horizon in contrast to having the lower edge,
- the out-performance of the full Kelly strategy in terms of the mean of the end wealth occurs at a significantly lower number of trials.

Exemplary, the summary statistics are given for an edge of 20% in table 2.5. The invested fraction is $f = p - q = 0.2$. The results use the given 10,000 trajectories for each strategy for the number of trials 100.

20% edge, 100 trials	Half Kelly	full Kelly	Double Kelly
$Mean(W_T)$	708.24	4756.5	1.61e+05
$Std(W_T)$	859.07	24899	7.87e+06
$\mathrm{P}(W_T < 100)$	0.06	0.18	0.54
$\mathrm{P}(W_T < 50)$	0.02	0.09	0.46
$\mathrm{P}(W_T < 10)$	0	0.02	0.3
$\mathrm{P}(W_t > 200)$	0.89	0.91	0.81
$Mean(T : 200)$	38.41	25.07	17.36
$\mathrm{P}(W_t > 1000)$	0.27	0.57	0.52
$Mean(T : 1000)$	79.4	57.77	37.54

Table 2.5: Descriptives for the wealth trajectories after 100 Bernoulli trials

Results

The short-term results given the Bernoulli trials for 100 days, respective three months, indicate that there is a trade-off between safety and return. Although the full Kelly strategy has the highest mean of the logarithm of final wealth, over-betting leads to better absolute results on average, with a significantly higher risk of severe draw-downs. Nevertheless, if an individual has the objective to double its initial wealth within three months, a risk-seeking strategy is not avoidable.

The mid-term results for 1000 days, respective three years, indicate that the trade-off between safety and returns remains. Whereas draw-down probabilities for the risk-averse (seeking) strategy decrease (increase) overall, the probability to double the initial wealth within three years is not significantly higher for the risky strategy. If the goal is to increase the wealth by a factor of ten, the risky strategy is again, not avoidable, taking severe draw-down risks.

The long-term results for 10,000 and 100,000 days, respective thirty and 300 years indicate that the average final wealth of the full Kelly strategy starts to outperform the Kelly variants. Strategies below or at full Kelly exhibit low draw-down probabilities at the end of the trials and have the highest probabilities to reach specified goals.

Asymptotically, the paper of Breiman (1961) has already proven that the full Kelly strategy is optimal in outperforming any other significantly different strategy and needing the minimal time to reach a certain goal. Hence, over-betting is not advised as it decreases long term growth and increases draw-down probabilities. Risk-averse individuals on the other hand, should bet below full Kelly as it minimizes draw-downs, although decreasing long-term growth. Additionally, if the true winning probability is estimated incorrectly, a risk-averse strategy is not only beneficial but elementary.

2.3 Portfolio simulations for one risky and a risk-free asset

2.3.1 Gaussian simulations, given moments

In this chapter, I am going to extend the papers of Ziemba and Hausch (1986) as well as MacLean, Thorp, Zhao, and Ziemba (2010). Besides increasing the amount of trajectories and covered time span, I will simulate from the non-parametric density of stock returns and drop the assumption that we know the past moments of stock returns. This leads to varying fractions over time.

The portfolio simulations start with wealth 100. Assuming the risky asset to be the S&P500, according prices following a Geometric Brownian Motion, the first two moments are calculated by Maximum Likelihood Estimates under the Normal with $\hat{\mu} = 0.00019959$ and $\hat{\sigma}^2 = 0.00016444$ (Annualized: $\hat{\mu}_{p.a.} = 0.0503$, $\hat{\sigma}^2_{p.a.} = 0.0414$). Using the daily estimates, the stock is simulated $100, 1000$ and $10,000$ times. Those different scenarios are run with $10,000$ trajectories each. In this chapter I assume the correct estimates to be given. Distribution assumption and parameter knowledge will be dropped in the following subsections.

Assuming the existence of a risk free asset with constant annual interest rate 0.5%, the model allows for leveraged portfolios. There are no constraints for short-selling. The full Kelly bet, under the knowledge of the parameters from which will be simulated, implies betting $f = \frac{\mu - r}{\sigma^2} = 1.0931$ times the initial capital on the risky asset and short the risk free asset by 0.0931% of initial capital as it is optimal in the sense of maximizing the expected logarithm of wealth (Merton, 1992; Thorp, 2006; Osorio, 2008). The famous rule of thumb to bet half Kelly would imply to invest $\approx 54.6\%$ in the risky asset and $\approx 45.4\%$ in the risk-free asset.

S&P500, 100 days, Gaussian

Following the notation of Breiman (1961), the wealth in period t is given by

$$W_t = W_{t-1} \times \left(f^\top x_t \right), \tag{2.5}$$

where $x_t = P_t / P_{t-1}$ is the price ratio vector and f the according fraction vector. Given the existence of the risk free asset, the constant portfolio fraction vector has in our univariate example, dimensions 2×1, $f \in \mathbb{R}^2$ with $\sum_{i=1}^{2} f_i = 1$.

Given the Maximum Likelihood Estimates, the according univariate distribution was simulated for 100 trials over 10,000 trajectories. The results were summarized in Table 2.6. Line one will always give the mean wealth of the final period over the different trajectories. Each mean is calculated over the different multiples of the original Kelly strategy. 1/4 for example implies betting a quarter of the original risky proportion, being more risk-averse. Line 2,3 and 4 will give you the according standard deviations, skewness and kurtosis for the different strategies in T. Line 5 to 7 gives you the probability that the end wealth is smaller than 100, 50 or 10 for the different strategies. Line 8, 10 and 12 will give you the probability that the wealth process is above level 200, 400 or 1600 at any point in time. Line 9,11 and 13 will give you the according average time of those trajectories to reach that goal.

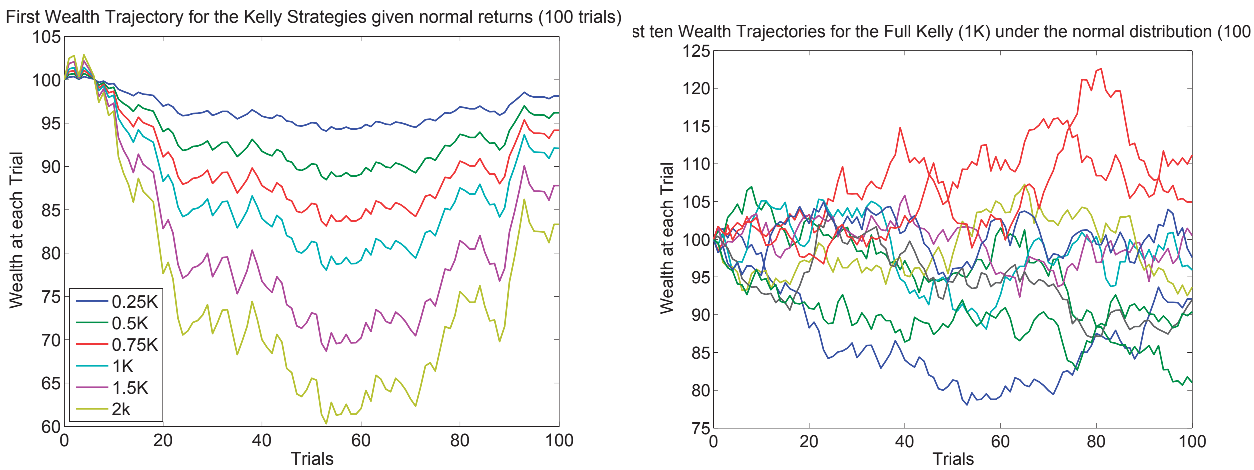
Under the Gaussian simulations for 100 trials, Table 2.6 shows that with increasing betting size in terms of Kelly fractions

- the four moments of the final wealth are increasing,
- the probability of ending below given draw-down levels in the final period remains relatively constant but is slightly increasing in betting size and
- although the probability to reach double the initial wealth in 100 trials is close to zero for all strategy, it is highest for the risk-seeking strategies.

Gaussian, 100 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	100.52	101.04	101.55	102.06	103.07	104.07
$Std(W_T)$	3.44	6.92	10.44	14.02	21.34	28.93
$Skew(W_T)$	0.07	0.17	0.27	0.38	0.59	0.81
$Kurt(W_T)$	3.01	3.05	3.13	3.26	3.63	4.21
P$(W_T < 100)$	0.44	0.45	0.45	0.46	0.47	0.49
P$(W_T < 50)$	0	0	0	0	0.0005	0.006
P$(W_T < 10)$	0	0	0	0	0	0
P$(W_t > 200)$	0	0	0	0	0.0007	0.01
$Mean(T : 200)$	NaN	NaN	NaN	NaN	87.57	83.5
P$(W_t > 400)$	0	0	0	0	0	0
$Mean(T : 400)$	NaN	NaN	NaN	NaN	NaN	NaN
P$(W_t > 1600)$	0	0	0	0	0	0
$Mean(T : 1600)$	NaN	NaN	NaN	NaN	NaN	NaN

Table 2.6: Descriptives for the wealth trajectories after 100 Gaussian trials

Under the knowledge of the true parameters of the distribution, from which was simulated, each strategy is profitable on average. The first trajectory of each of the six played strategies for 100 periods is shown in figure 2.5a and the first ten full Kelly trajectories are given in figure 2.5b. The mean-variance trade-off is given as well, amplified by increasing skewness and kurtosis of the final wealth distribution, presented as kernel density in figure 2.6. As betting size increases the mean of the distribution lies right of the median/modus of the distribution. Consequently, the final wealth is not normally distributed. Instead, the final wealth distributions is log-normally distributed for all betting sizes, as the logarithm of the final wealth is normally distributed (see figure 2.7a). This is expected from the geometric growth process itself as the log-normal is the exponential of the normal distribution. In terms of optimization, knowing the true distribution, the highest average in terms of logarithmic wealth is given by the full Kelly strategy, coming from the optimization of maximizing the expected logarithm of wealth (see figure 2.7b).



(a) First wealth trajectory for the Kelly strategies given normal returns (100 trials)

(b) First ten wealth trajectories for the full Kelly (1K) under the normal (100 trials)

Figure 2.5: Wealth trajectories under the Normal (100 trials)

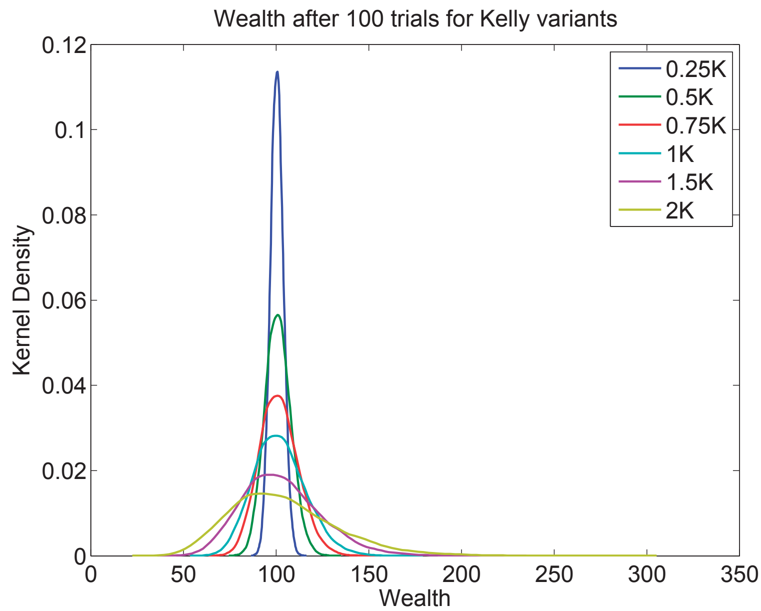
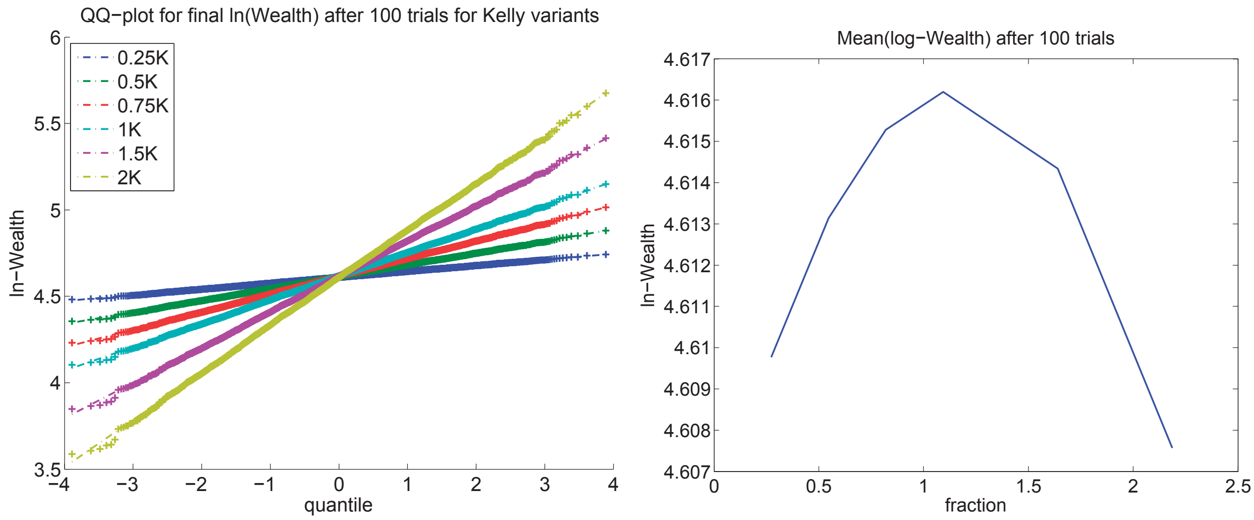


Figure 2.6: Wealth after 100 Gaussian trials for Kelly variants



(a) QQ-plot for final log(Wealth) after 100 trials for Kelly variants

(b) Mean(log-Wealth) after 100 trials

Figure 2.7: QQ-plot and Mean(log-Wealth) after 100 Gaussian trials

S&P500, 1000 days, Gaussian

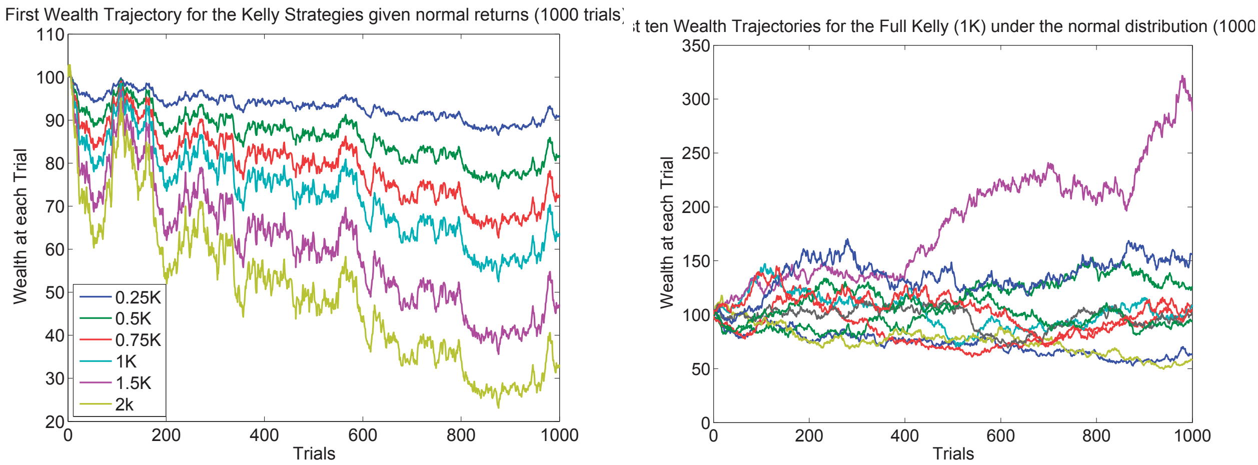
In contrast to the simulations under the Gaussian for 100 trials, the simulation results for 1000 trials in Table 2.7 show that

- the four moments of the final wealth keep increasing,
- the probability of ending below given draw-down levels in the final period decreases for betting sizes for/below full Kelly and increases for strategies over-betting and
- the probabilities to reach wealth goals in 1000 trials increases for risk-seeking strategies. The riskiest strategies need the least time to reach those goals.

Gaussian, 1000 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	105.05	110.34	115.88	121.7	134.18	147.89
$Std(W_T)$	11.62	24.66	39.47	56.48	99.30	158.86
$Skew(W_T)$	0.34	0.68	1.05	1.45	2.43	3.78
$Kurt(W_T)$	3.16	3.75	4.84	6.59	13.33	27.98
$P(W_T < 100)$	0.34	0.371	0.39	0.41	0.45	0.49
$P(W_T < 50)$	0	0	0.01	0.03	0.12	0.21
$P(W_T < 10)$	0	0	0	0	0	0.005
$P(W_t > 200)$	0	0.005	0.06	0.16	0.32	0.42
$Mean(T : 200)$	NaN	860.19	767.95	677.01	545.69	453.05
$P(W_t > 400)$	0	0	0	0.0039	0.05	0.11
$Mean(T : 400)$	NaN	NaN	NaN	861.85	763.61	677.24
$P(W_t > 1600)$	0	0	0	0	0	0.002
$Mean(T : 1600)$	NaN	NaN	NaN	NaN	NaN	876.21

Table 2.7: Descriptives for the wealth trajectories after 1000 Gaussian trials

When using 1000 periods for the 10,000 trajectories each strategy is again profitable on average. The first trajectory of each of the six played strategies for 1000 periods is shown in figure 2.8a and the first ten full Kelly trajectories are given in figure 2.8b. The mean-variance trade-off is intensified, amplified by further increasing skewness and kurtosis of the final wealth distribution. As betting size increases the modus of the distributions tends to go faster to zero the higher the betting size gets. The final wealth distributions is again log-normally distributed for all betting sizes. The full Kelly strategy has again the highest average of the logarithm of wealth.



(a) First wealth trajectory for the Kelly strategies given normal returns (1000 trials)

(b) First ten wealth trajectories for the full Kelly (1K) under the normal (1000 trials)

Figure 2.8: Wealth trajectories under the Normal (1000 trials)

S&P500, 10,000 and 100,000 days, Gaussian

As the sample size tends to even larger periods, the median of the full Kelly final wealth is the highest among the other betting strategies. When over-betting, the modus of the distribution goes to zero and the probabilities to reach the wealth goals decrease. The asymptotic theorems of Breiman (1961) are beginning to kick in.

2.3.2 Non-parametric simulations, given moments

The results of Merton (1992) / Thorp (2006) rest on the assumption of the Geometric Brownian Motion / the second order Taylor approximation of the expected growth rate. This gives, under the knowledge of the true parameters and simulations under the normal, an averagely increasing wealth, which is log-normally distributed. This may not be realistic, as neglected outliers may influence results in the short and long run. Hence, I vary the simulation method for the returns by simulating from the Kernel density of the empirical non-parametric distribution instead of a parametric distribution, such as the Gaussian.
The portfolio fractions for the risky asset is again assumed to be time constant, estimated as in the subsection before. I aim to test the hypothesis if the statistics of the portfolio-wealth simulations are changing significantly. In other words, I aim to test if especially the third and the fourth moment offer potential for a more general portfolio solution, not resting on the presented approximations, leading to the already presented paper of Osorio (2008) and the general stochastic optimization problem.

S&P500, 100 days, non-parametric

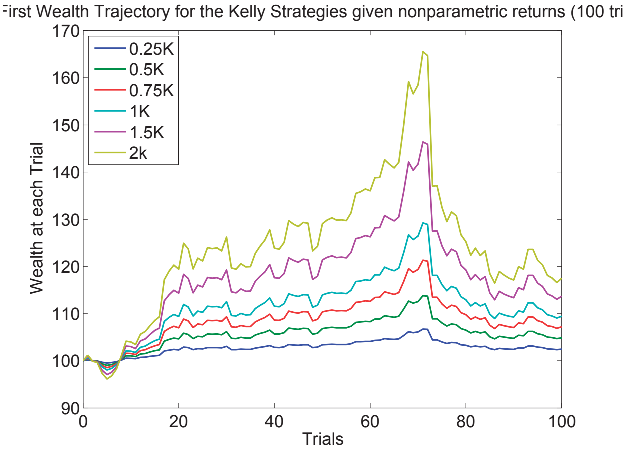
Comparing the portfolio results from the non-parametric simulations (see table 2.8) with the simulations under the Gaussian for 100 trials (see table 2.6),

- the first four moments are slightly increased, the average of the logarithmic final wealth is still maximal for the full Kelly strategy,
- the draw-down probabilities are slightly decreased and
- the probabilities to reach the wealth goals are slightly increased.

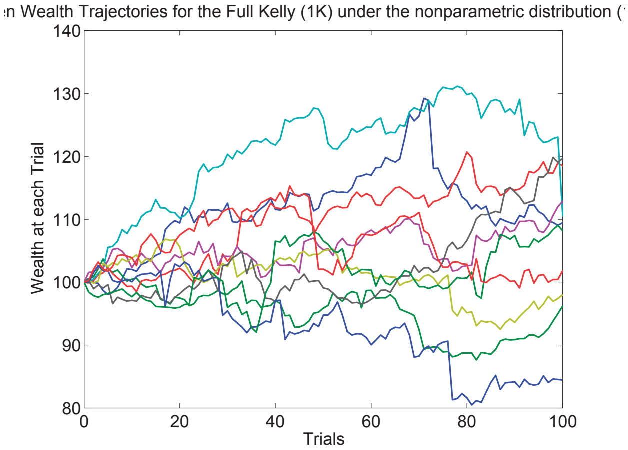
The wealth trajectories for the scaled Kelly variants are given in figure 2.9a/b. Although the empirical distribution of the S&P500 is significantly different from the one under the Gaussian, using for example the Jarque-Bera test, the differences in final wealth after 100 trials are negligible in absolute terms of Table 2.8. I aim to test whether this holds for larger periods of time.

nonpara, 100 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	100.54	101.09	101.63	102.18	103.28	104.38
$Std(W_T)$	3.53	7.11	10.73	14.42	21.99	29.87
$Skew(W_T)$	0.07	0.18	0.28	0.39	0.61	0.84
$Kurt(W_T)$	3.13	3.16	3.25	3.37	3.76	4.34
$\mathrm{P}(W_T < 100)$	0.44	0.44	0.45	0.46	0.47	0.48
$\mathrm{P}(W_T < 50)$	0	0	0	0	0.001	0.01
$\mathrm{P}(W_T < 10)$	0	0	0	0	0	0
$\mathrm{P}(W_t > 200)$	0	0	0	0	0.001	0.011
$Mean(T : 200)$	NaN	NaN	NaN	NaN	87.09	78.89
$\mathrm{P}(W_t > 400)$	0	0	0	0	0	0
$Mean(T : 400)$	NaN	NaN	NaN	NaN	NaN	NaN
$\mathrm{P}(W_t > 1600)$	0	0	0	0	0	0
$Mean(T : 1600)$	NaN	NaN	NaN	NaN	NaN	NaN

Table 2.8: Descriptives for the wealth trajectories after 100 non-parametric trials



(a) First wealth trajectory for the Kelly strategies given non-parametric returns (100 trials)



(b) First ten wealth trajectories for the full Kelly (1K) given non-parametric returns (100 trials)

Figure 2.9: Wealth trajectories under the non-parametric distribution (100 trials)

Indeed, the final wealth densities under the non-parametric simulations (see figure 2.10) are not significantly different from the Gaussian simulations for all betting strategies. Using the non-parametric two-sample Kolmogorov-Smirnov test, the null of equality of the according CDFs cannot be denied with a p-value of one. Note that the S&P500 is a broad stock market index, exhibiting rarely drastic draw-downs.

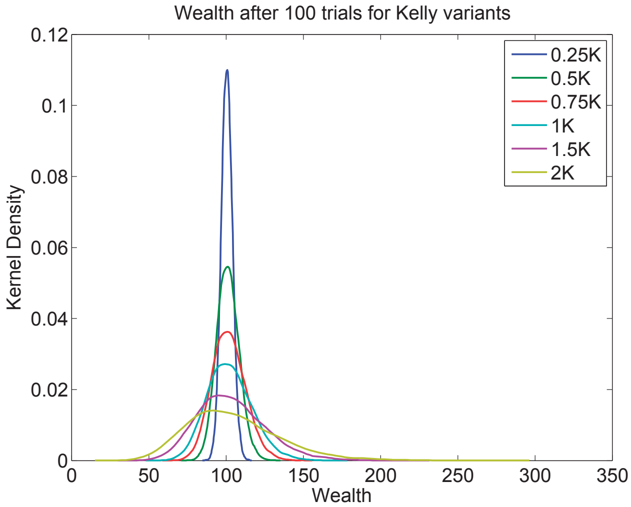
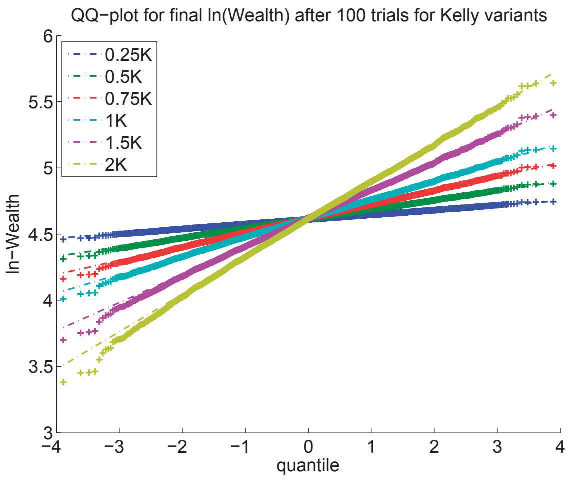
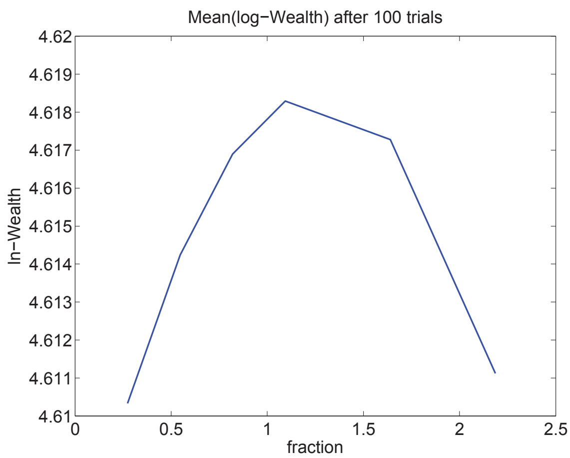


Figure 2.10: Wealth after 100 non-parametric trials for Kelly variants

The qq-plot from figure 2.11a indicates that the final wealth distributions come from the log-normal distribution. But, as the betting size increases, the log-normal distribution fails in capturing the left tail of the distribution. Consequently, the Jarque-Bera-test rejects the null of normality of the final log-wealth with a p-value of 0.0095 for the quarter Kelly strategy. This results holds for all betting sizes and also increasing trials, therefore the normality plots and also the expected logarithm of wealth plots are consequently omitted as the full Kelly strategy always has the highest log-wealth (see 2.11b). If this also holds for a longer time frame, this means that, as long as the first two moments are known, the optimum growth portfolio outperforms significantly different strategies asymptotically.



(a) QQ-plot for final log(Wealth) after 100 trials for Kelly variants



(b) Mean(log-Wealth) after 100 trials

Figure 2.11: QQ-plot and Mean(log-Wealth) after 100 non-parametric trials

S&P500, 1000 days, non-parametric

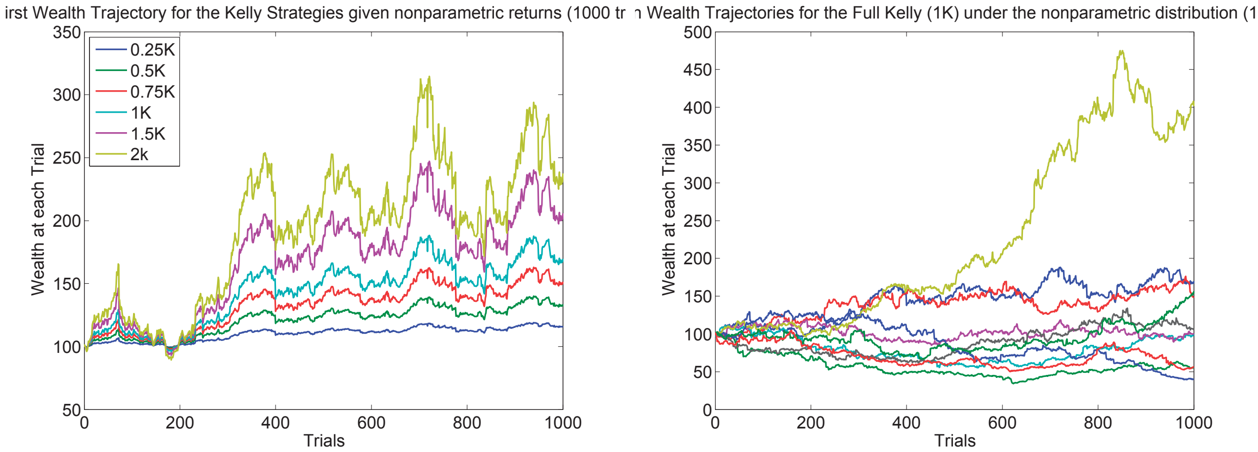
Comparing the portfolio results from the non-parametric simulations (see table 2.9) with the simulations under the Gaussian for 1000 trials (see table 2.7),

- the first four moments increase steadily, the average of the logarithmic final wealth is still maximal for the full Kelly strategy,
- the draw-down probabilities decrease insignificantly and
- the probabilities to reach the wealth goals increase additionally.

nonpara, 1000 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	105.71	111.75	118.11	124.82	139.36	155.51
$Std(W_T)$	11.8	25.16	40.49	58.31	104.17	170.53
$Skew(W_T)$	0.31	0.66	1.05	1.48	2.63	4.53
$Kurt(W_T)$	3.22	3.84	5.05	7.18	17.24	48.64
$\mathrm{P}(W_T < 100)$	0.32	0.34	0.36	0.38	0.43	0.48
$\mathrm{P}(W_T < 50)$	0	0.0003	0.009	0.036	0.12	0.2
$\mathrm{P}(W_T < 10)$	0	0	0	0	0.0002	0.005
$\mathrm{P}(W_t > 200)$	0	0.005	0.07	0.16	0.33	0.44
$Mean(T:200)$	NaN	860.15	773.07	673.37	545.32	457.03
$\mathrm{P}(W_t > 400)$	0	0	0.0001	0.003	0.05	0.12
$Mean(T:400)$	NaN	NaN	820	866.91	776.95	680.95
$\mathrm{P}(W_t > 1600)$	0	0	0	0	0.0001	0.002
$Mean(T:1600)$	NaN	NaN	NaN	NaN	869	845.72

Table 2.9: Descriptives for the wealth trajectories after 1000 non-parametric trials

The wealth trajectories for the scaled Kelly variants are given in figure 2.12a/b. Table 2.9 indicates that the differences in final wealth after 1000 trials are not negligible anymore.



(a) First wealth trajectory for the Kelly strategies given non-parametric returns (1000 trials)

(b) First ten wealth trajectories for the full Kelly (1K) given non-parametric returns (1000 trials)

Figure 2.12: Wealth trajectories under the non-parametric distribution (1000 trials)

In contrast to the simulations for 100 periods, the final wealth distributions under the non-parametric simulations are significantly different from the non-parametric for all betting strategies. Using the non-parametric two-sample Kolmogorov-Smirnov test, the alternative of difference of the according CDFs cannot be denied with a p-value of 0.00014 for the quarter Kelly strategy. The results for the riskier betting strategies are equivalent. This statistically significant difference leads us to the conclusion that I will use the non-parametric simulations technique for subsequent chapters.

S&P500, 10,000 days, non-parametric

Comparing the portfolio results from the non-parametric simulations with the simulations under the Gaussian for 10,000 trials, the differences are identical to the differences when using 1000 trials. Especially for betting sizes larger than 3/4-Kelly, the Kurtosis of final wealth returns increases even more. According figures are omitted. I will simulate from the according non-parametric density.

2.3.3 Non-parametric simulations, estimated moments

In the last two subsections the first two moments of the empirical S&P500 returns were given. This knowledge is dropped as, in a realistic scenario, parameters of the probability distribution have to be estimated.

Consequently, the first two moments are estimated via MLE using the data given up to day $t + h$, where h is the minimum of data which has to be used to construct the estimates. So, if h=50, there is no investment in the risky asset for the first 51 days. From day 51 on, parameters are estimated upon the past data to construct the fraction for the risky asset. In order to use the moments under the Gaussian, the estimators from day 52 are using the past 51 returns instead of a rolling window estimate of 50 returns. As time passes by, more data will be used, as, in the MLE-sense, it decreases the standard error of the mean. The returns are simulated under the non-parametric distribution, as the wealth distributions have significantly differed from the simulations under the Gaussian assumption.

S&P500, 100 days, non-parametric, time varying parameters

For the 10,000 portfolio simulations for 100 days using, $h = 50$, the wealth trajectories underlie rapid changes as the horizon is short. As the data span is extremely short, parameters are fluctuating rapidly and the daily changes in portfolio fractions are extreme. Especially in the context of transaction costs, which are neglected here, a strategy based on such a short data span is not advised.

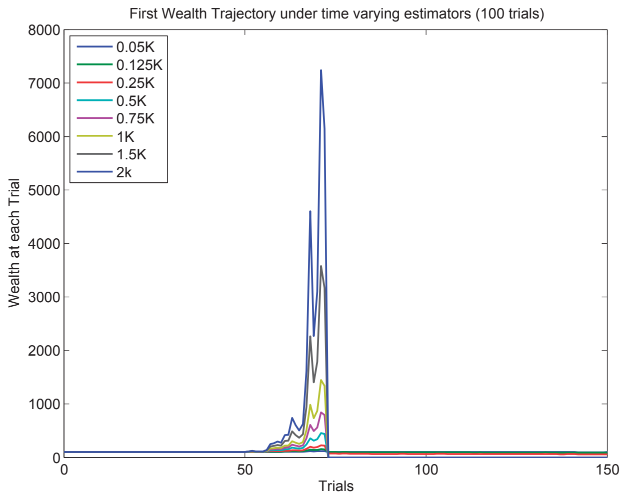
Comparing the results of table 2.10 using time-varying estimators with the results from table 2.8,

- the average wealth is larger up to full Kelly, but decreases strictly for the over-betting strategies
- but, for the risk-averse strategies up to full Kelly, the second-, third- and fourth-order moments are increased in an extreme manner, not justifying the higher mean,
- the draw-down probabilities are significantly increased, even for the risk-averse quarter Kelly strategy the probability of ending below the initial wealth is $\approx 69\%$ whereas
- the probability of hitting wealth goals is increased, but not that much to compensate for the extreme draw-down risks.

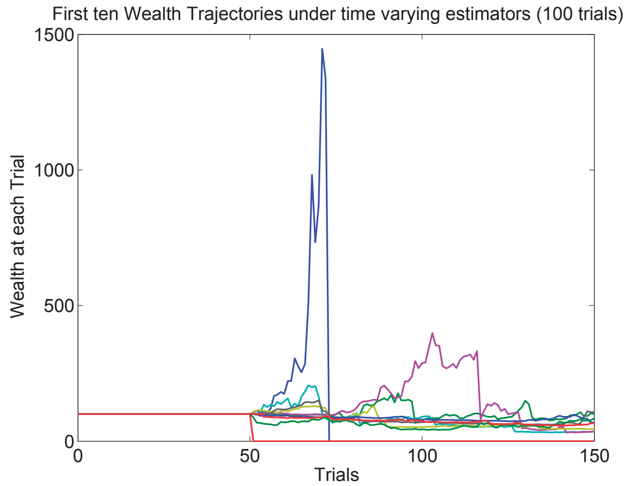
timevary, 100 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	100.94	102.09	103.34	104.72	91.52	70.27
$Std(W_T)$	41.77	138.79	390.23	917.35	1211.4	1377.8
$Skew(W_T)$	7.04	22.16	49.7	73.79	66.55	74.25
$Kurt(W_T)$	101.77	820.73	3333.4	6369.1	5328.2	6266.1
$\text{P}(W_T < 100)$	0.69	0.74	0.79	0.83	0.89	0.93
$\text{P}(W_T < 50)$	0.008	0.1	0.27	0.46	0.72	0.84
$\text{P}(W_T < 10)$	0	0.01	0.05	0.12	0.32	0.51
$\text{P}(W_t > 200)$	0.05	0.16	0.23	0.29	0.35	0.38
$Mean(T : 200)$	101.33	92.3	85.84	82.32	77.42	73.56
$\text{P}(W_t > 400)$	0.005	0.04	0.08	0.11	0.16	0.19
$Mean(T : 400)$	116.67	102.03	95.18	92.07	85.23	80.01
$\text{P}(W_t > 1600)$	0	0.002	0.01	0.02	0.04	0.05
$Mean(T : 1600)$	NaN	115.05	108.68	102.17	93.64	89.32

Table 2.10: Descriptives for the wealth trajectories after 100 non-parametric trials, time

The first wealth trajectory in figure 2.13a shows how volatile the trajectories are. The first ten full Kelly trajectories in figure 2.13b support the result that the wealth is far away from monotonically increasing. Instead, relatively few trajectories lead to the outperforming result of the optimum growth portfolio.



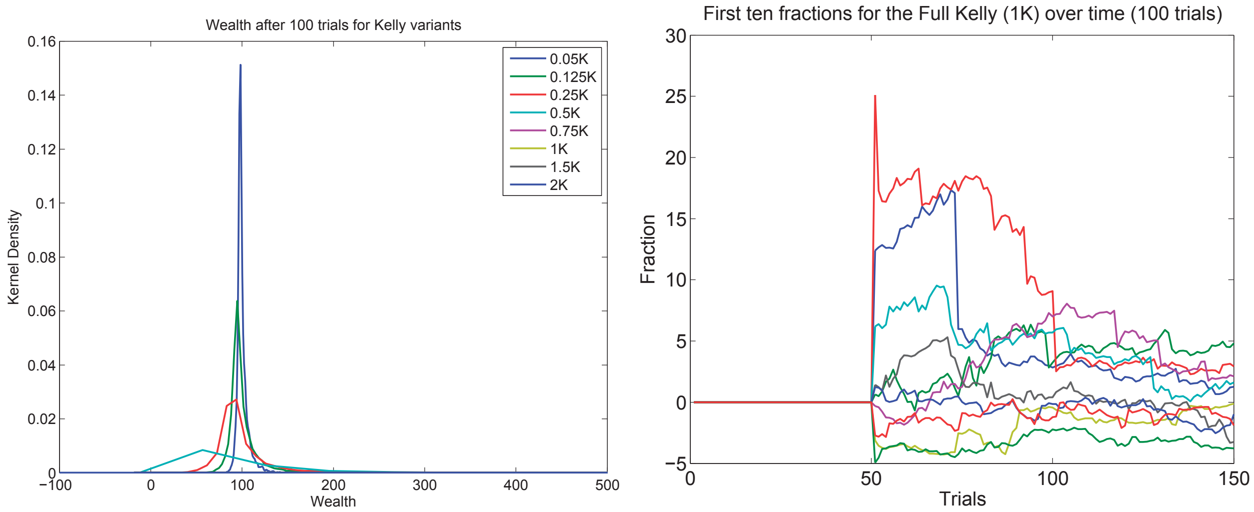
(a) First wealth trajectory for the Kelly strategies given time-varying estimators (100 trials)



(b) First ten wealth trajectories for the full Kelly (1K) given time-varying estimators (100 trials)

Figure 2.13: Wealth trajectories under the non-parametric distribution, estimated moments (100 trials)

The final wealth distributions in figure 2.14a show that as the betting size increases, the modus and median of the distribution tend to zero, as derived in Hakansson and Ziemba (1995). Around 29% of full Kelly trajectories end below one tenth of the initial wealth. The distributions for the over-betting strategies are not plotted as the probability mass is only visual around zero. The reason for the high volatility in the wealth paths is that the short data span in the start leads to extreme portfolio fractions. The first trajectory starts investing 24-times the initial wealth in the risky asset. As time passes by, more data is used for the ML-estimators leading to more smooth fractions in the end (see figure 2.14b). Remember, the in-sample fraction given the same approximation is 1.0931.

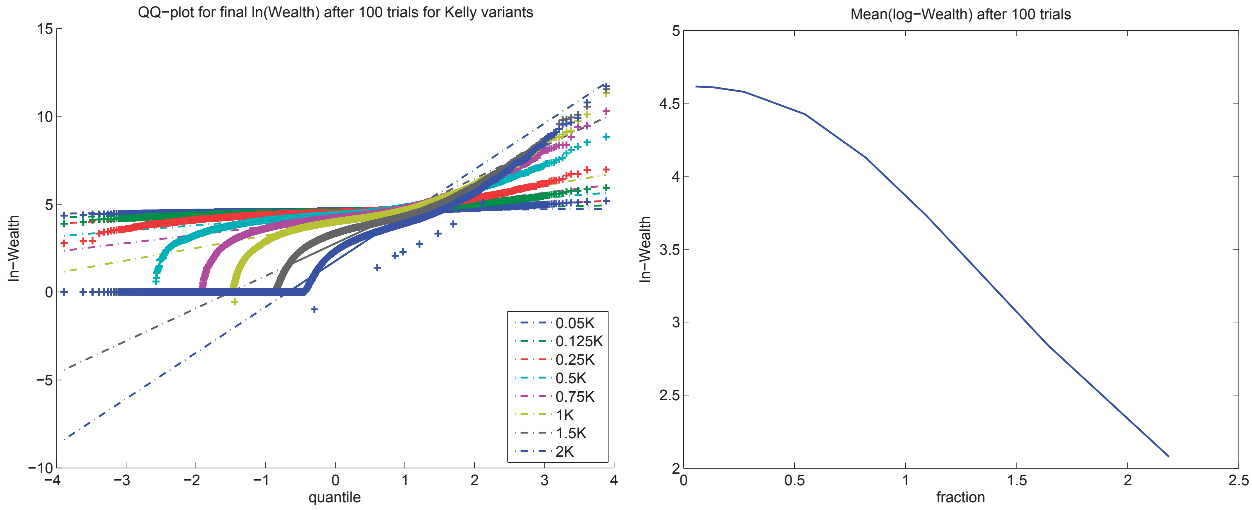


(a) Wealth after 100 trials for Kelly variants given time-varying estimators

(b) First ten fractions for the full Kelly (1K) over time (100 trials)

Figure 2.14: Wealth return density and first ten fractions for full Kelly, 100 non-parametric trials, estimated moments

Additionally, the wealth is by no means log-normally distributed anymore (see figure 2.15a). The riskier the bets, the more the final wealth distribution tends away from the log-normal. Important, from figure 2.15b it is observable that the average of the logarithmic final wealth is not maximal for full Kelly, instead it increases as the fraction decreases. The foundation of the Kelly Criterion is that the expected logarithm of wealth should be maximized. This does not hold for the given portfolio simulations for a short time period of 100 days where the parameters are time-varying ML-estimators. Hence, I will increase the time span for the simulations and the data span for the estimators to verify if this result holds true for a larger time horizon. As the time span for the loading tends to infinity, the wealth trajectories tend to the in-sample result, assuming that the data generating process is not changing over the data span.



(a) QQ-plot for final log(Wealth) after 100 trials for Kelly variants given time-varying estimators

(b) Mean(log-Wealth) after 100 trials

Figure 2.15: QQ-plot and Mean(log-Wealth) after 100 non-parametric trials, estimated moments

S&P500, 1000 days, non-parametric, time varying parameters

For the 10,000 portfolio simulations for 1000 days using 150 days as starting data span, $h = 150$, the final wealth distributions still underly rapid changes as parameters are fluctuating heavily and the daily changes in portfolio fractions are extreme.

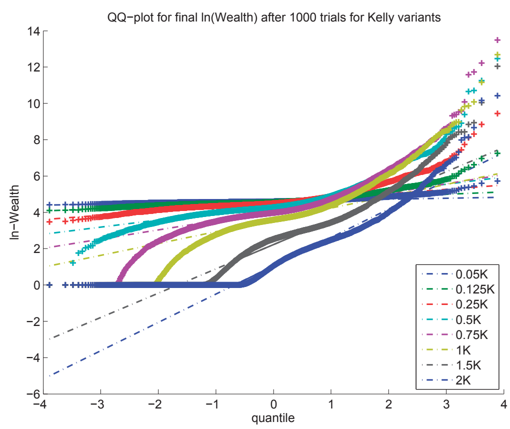
Comparing the results of table 2.11 using time-varying estimators with the results from table 2.9, the results are equivalent to the comparison at 100 days of simulation.

- The average wealth is larger up to full Kelly, but decreases strictly for the over-betting strategies
- but, for the strategies smaller or equal than full Kelly, the higher-order moments are increased in an extreme manner, not justifying the higher mean,
- the draw-down probabilities are significantly increased, even for the risk-averse quarter Kelly strategy the probability of ending below the initial wealth is $\approx 69\%$ whereas
- the probability of hitting wealth goals is increased, but not that much to compensate for the extreme draw-down risks.

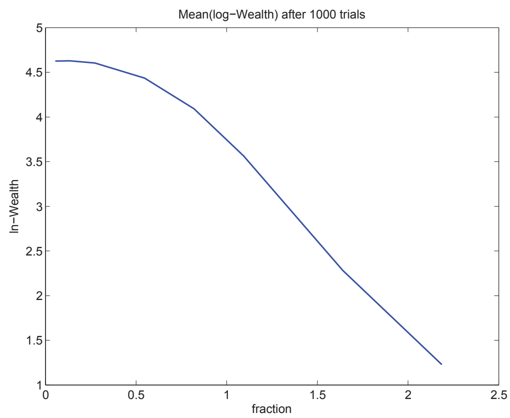
Increasing the starting data span to 150 days and increasing the horizon has still equivalent effects on the wealth paths. Although having higher returns on average than under the simulations, where the parameters are known, the strategies under time varying estimators are too risky. Omitting consequent figures, it is crucial that the maximum logarithmic wealth is neither normally distributed (see figure 2.16a) nor peaks at the growth optimum portfolio. Instead, it decreases with rising betting size (see figure 2.16b).

timevary, 1000 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	110.77	158.87	225.71	125.25	50.461	16.266
$Std(W_T)$	168.27	2786.6	7638	3358.4	1719.4	434.94
$Skew(W_T)$	50.956	83.55	85.19	91.51	96.01	68.15
$Kurt(W_T)$	3419.7	7579.6	7815.8	8775	9440.6	4880.5
$\mathrm{P}(W_T < 100)$	0.66	0.74	0.81	0.87	0.95	0.98
$\mathrm{P}(W_T < 50)$	0.005	0.12	0.45	0.71	0.91	0.97
$\mathrm{P}(W_T < 10)$	0	0.001	0.02	0.1	0.45	0.82
$\mathrm{P}(W_t > 200)$	0.12	0.26	0.34	0.39	0.43	0.45
$Mean(T : 200)$	553.54	443.85	390.63	355.93	307.23	268.27
$\mathrm{P}(W_t > 400)$	0.02	0.09	0.15	0.18	0.22	0.23
$Mean(T : 400)$	663.82	562.49	484.73	431.61	363.44	309.63
$\mathrm{P}(W_t > 1600)$	0.001	0.01	0.03	0.04	0.06	0.07
$Mean(T : 1600)$	734.55	665.78	594.98	528.46	430.16	360.53

Table 2.11: Descriptives for the wealth trajectories after 1000 non-parametric trials, time



(a) QQ-plot for final log(Wealth) after 1000 trials



(b) Mean(log-Wealth) after 1000 trials

Figure 2.16: QQ-plot and Mean(log-Wealth) after 1000 non-parametric trials, estimated moments

S&P500, 1000 days, non-parametric, increasing the data span

Increasing the starting data span to 500 days for the first two moments and hence the optimal fraction, leads, in contrast to the use of a starting data span of 150 days to

- less volatile fractions over time and therefore also significantly less volatile wealth paths,
- increasing results on average starting with 3/4-Kelly,
- significantly smaller third and fourth moment for the resulting wealth,
- smaller draw-down probabilities but also smaller probabilities to reach certain goals.

timevary, 1000 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	109.05	129.99	189.44	319.83	315.92	66.282
$Std(W_T)$	67.87	615.67	4318.2	16366	19917	1087.3
$Skew(W_T)$	18.58	68.59	92.74	98.89	99.39	63.24
$Kurt(W_T)$	674.48	5535.6	8962.1	9848.6	9918.1	4502
P($W_T < 100$)	0.62	0.67	0.73	0.78	0.87	0.93
P($W_T < 50$)	0.002	0.05	0.17	0.34	0.7	0.85
P($W_T < 10$)	0	0.0002	0.006	0.03	0.18	0.39
P($W_t > 200$)	0.08	0.21	0.29	0.34	0.4	0.43
$Mean(T : 200)$	1028.9	904.03	843.54	796.71	740.33	702.12
P($W_t > 400$)	0.01	0.06	0.11	0.15	0.2	0.22
$Mean(T : 400)$	1167.9	1028.2	956.51	902.13	827.81	779.14
P($W_t > 1600$)	0.0003	0.006	0.02	0.03	0.05	0.06
$Mean(T : 1600)$	1216.3	1182.1	1079	1015.2	930.55	858.49

Table 2.12: Descriptives for the wealth trajectories after 1000 non-parametric trials, time (2)

In essence the smoothing of the wealth paths by starting to estimate with two years of data leads to overall better results, especially in terms of draw-downs. Still, after three years of simulations the full Kelly strategy with the highest average result wealth, has around 80% probability of ending below the initial wealth.

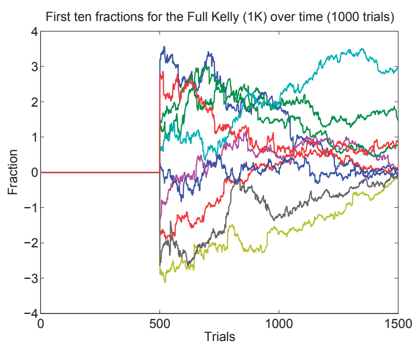


Figure 2.17: First ten fractions for the full Kelly (1K) over time (1000 trials)

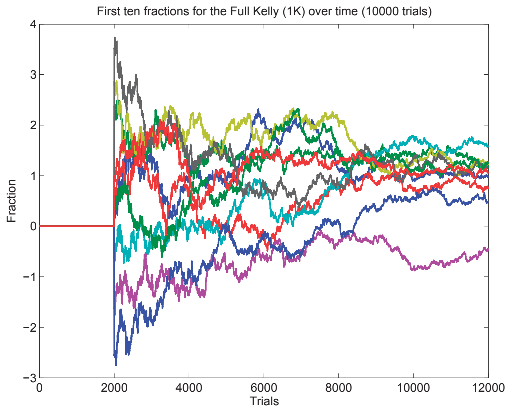
S&P500, 10,000 days, non-parametric, time varying parameters

For 10,000 portfolio simulations for 10,000 days using 2000 days (8 years) as starting data span, $h = 2000$, the final wealth distributions still underlie significant changes although the parameter changes become smaller as time passes by. On average, the full Kelly outperforms risk-averse and -seeking portfolios. As the Kelly portfolio aims to maximize the median wealth process at each time point, for this simulation study, the half Kelly strategy maximizes the median final wealth. Although the probability of ending below (half) initial wealth is 35.7% (3.4%), potential upside gains are large: More than half of the trajectories double its initial wealth after 30 years. Over one fourth quadruple the initial wealth and every tenth trajectory multiplies the initial wealth by factor 16.

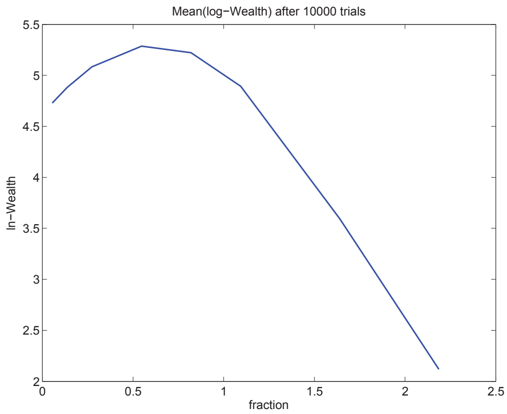
timevary, 10000 trials	1/4	1/2	3/4	1	3/2	2
$Mean(W_T)$	239.28	1222.2	4491	4920.8	954.16	225.37
$Std(W_T)$	697.9	24493	1.17e+05	1.03e+05	12502	3436.3
$Skew(W_T)$	35.15	57.34	49.42	34.21	30.16	34.69
$Kurt(W_T)$	1729.2	3631.3	2749.7	1264.9	1104	1414.4
$\mathrm{P}(W_T < 100)$	0.29	0.35	0.42	0.52	0.74	0.89
$\mathrm{P}(W_T < 50)$	0.002	0.034	0.14	0.31	0.63	0.85
$\mathrm{P}(W_T < 10)$	0	0.0004	0.007	0.03	0.24	0.58
$\mathrm{P}(W_t > 200)$	0.39	0.56	0.62	0.65	0.66	0.65
$Mean(T : 200)$	6562	5435.7	4884.5	4511.7	4027	3684.7
$\mathrm{P}(W_t > 400)$	0.14	0.33	0.42	0.46	0.46	0.44
$Mean(T : 400)$	8067.4	6774.4	6058.5	5628.7	5008.1	4397.3
$\mathrm{P}(W_t > 1600)$	0.01	0.1	0.16	0.21	0.21	0.19
$Mean(T : 1600)$	9338.7	8156.5	7505.4	6964.8	6132.9	5324.7

Table 2.13: Descriptives for the wealth trajectories after 10000 non-parametric trials

As more data are being used, the closer tend the fractions to the theoretical parameters from the data generating process. Figure 2.18a gives the first ten fractions of the portfolio simulations. Remember, the in-sample fraction, which should be optimal is 1.0931. In contrast to the section before, the changes in fraction are reducing furthermore, yet are converging, due to the non-parametrical simulation method, slower as expected under the Gaussian. As the loading period would tend to infinity, the fraction would tend to its in-sample analogue. For empirical application this is not appropriate, as I assume the data generating process of a financial market to change constantly. Therefore, limited data have to be used, leading inevitable to risk-averse betting schemes.



(a) First ten fractions for the full Kelly (1K) over time (10000 trials)



(b) Mean(log-Wealth) after 10,000 trials

Figure 2.18: First ten fractions for the full Kelly and Mean(log-Wealth) for 10,000 non-parametric trials, estimated moments

Although the wealth is not log-normally distributed (see Figure 2.18b), the average logarithm of wealth peaks for the half Kelly strategy. Remembering the results when the parameters were given beforehand, by optimization, the full Kelly strategy should lead to the maximization of the log-wealth, which is not the case here. The growth optimum portfolio cannot asymptotically outperform every other strategy, which is significantly different as the assumptions made in Breiman (1961) are not met.

Chapter 3

Empirical backtest: Univariate financial market returns

In chapter 1, closed-form solutions for the Kelly growth-optimum Criterion under different distribution assumptions have been derived. In contrast to the simulation study from chapter 2, the Criterion will be tested on an empirical basis as well. Without knowledge of the future distribution, the smooth results of the simulation chapter, as already shown via time-varying, past-dependent estimators cannot be held.

The depicted daily price series $\{P_{t,j}\}_{t=1}^{T}$ with $j = 1, \ldots, 11$ as assets are S&P 500, DAX 30, Nikkei 225, Schweizer MI, DJ Commodities, PHLX Gold/Silver Sector, iShares US Real Estate, Vanguard Bond Market Index, CBOE 10 Year Interest, Bershire (B) and Deutsche Bank in order to evaluate if the Kelly bet performs well over different, positive and negative performing assets. This will happen in a univariate way, testing each market separately. The data span is ten years: from t, the first January 2005 till T, the first January of 2015. Instead of deleting data points in order to have price vectors of the same length, different amount of daily data points are used. There was no data cleaning in order to get matrices to vectorize operations as in chapter 2. Instead struct- and cell- arrays are used in Matlab, in order to deal with vectors of different length. The according benchmarks are the asset itself, large stock indices and the risk free asset. Betting schemes are derived and discussed in chapter 1. Starting with Kelly (1956) for Bernoulli trials the Kelly portfolio is generalized for normal returns (Merton, 1969) and further continuous approximations (Osorio, 2008).

3.1 In-sample backtest: i.i.d.-assumption

The data span from 2005 to 2015 is chosen on purpose. As to be seen in figure 3.1, ten out of the eleven normalized assets underlie, due to the financial crisis, drastic changes in value. The bond-market on the other hand is the sole market, performing steady without larger draw-downs.

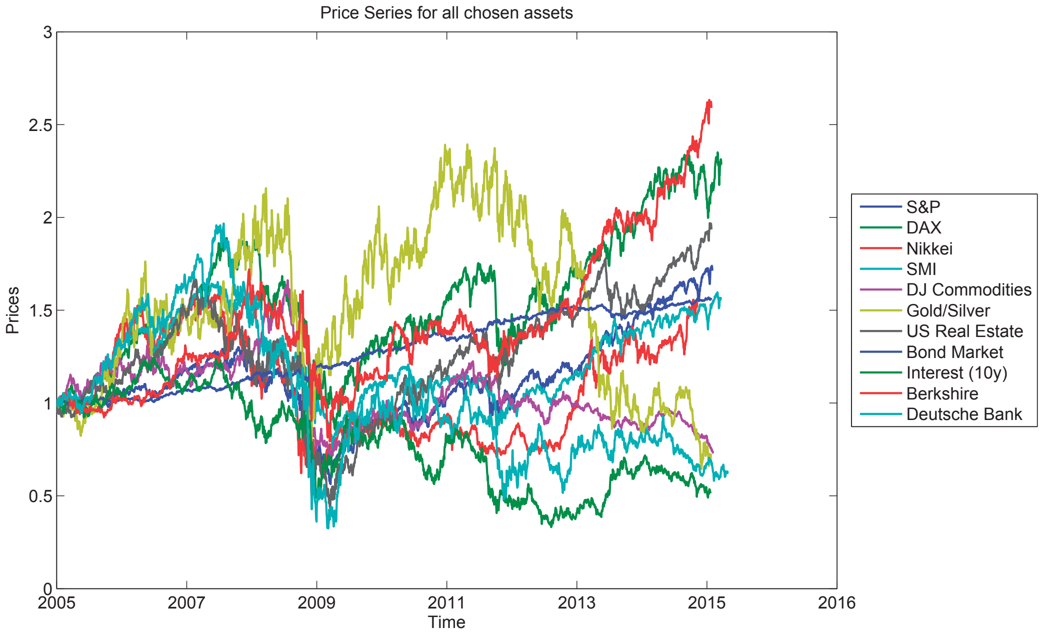


Figure 3.1: Normalized price series for all chosen assets

The descriptives for prices and consequently wealth paths are given in the same structure as in table 3.1. Line one will present the days of data for each asset. Although different, trading days for ten years lie between 2464 (SMI) and 2582 (DBK). Line two and three show the start and end price of the asset, whereas line four and five contain the minimum and maximum price. Thus, the annualized arithmetic mean and standard deviation are given in lines six and seven. The third and fourth moment of the returns are shown in line eight and nine. As performance measurement the Sharpe- and Sortino-ratio are to be seen in line ten and eleven. Additionally, the minimum and maximum return are shown in line twelve and thirteen. The assets itself will serve as benchmark for the upcoming wealth paths. From the descriptive point of view the leptokurtic behaviour of the return distributions is obvious, leading to an overall rejection of the null that the assets are normally distributed.

Descriptives	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
Days of data	2521	2555	2464	2542	2523	2516	2517	2517	2514	2517	2582
Start Price	1202.1	4291.5	11518	5768.7	143.23	95.65	39.71	6.94	4.22	57.98	40.14
End Price	2058.9	9805.5	17451	8983.4	104.33	68.78	76.84	10.87	2.17	150.15	24.99
Min Price	676.53	3666.4	7055	4307.7	102	61.63	17.35	6.84	1.4	46	13.02
Max Price	2090.6	10087	18262	9531.5	237.96	228.95	78.2	10.88	5.25	152.67	78.95
Mean Return (p.a.)	0.055	0.085	0.055	0.045	-0.031	-0.032	0.068	0.046	-0.0645	0.099	-0.045
Std Return (p.a.)	0.2	0.22	0.25	0.18	0.18	0.41	0.34	0.04	0.33	0.23	0.41
Skewness	-0.33314	0.03	-0.56	-0.02	-0.28	0.09	-0.47	0.13	-0.14	0.83	0.32
Kurtosis	14.03	9.95	11.28	11.65	5.69	9.44	17.36	4.45	6.74	19.33	12.88
Sharpe Ratio (p.a.)	0.27	0.38	0.17	0.25	-0.17	-0.07	0.19	1.13	-0.19	0.43	-0.11
Sortino Ratio (p.a.)	0.37	0.54	0.23	0.34	-0.23	-0.11	0.27	1.71	-0.27	0.65	-0.15
Min Return (p.a.)	-0.1	-0.07	-0.12	-0.08	-0.06	-0.18	-0.23	-0.02	-0.17	-0.11	-0.17
Max Return (p.a.)	0.11	0.1	0.13	0.11	0.05	0.23	0.15	0.01	0.09	0.17	0.21

Table 3.1: Descriptives for the different assets, 2005-2015

3.1.1 Investment fractions: Utilizing Merton (1969) and Osorio (2008)

The crucial assumption in order to derive the results of Merton (1992) is that the changes in the price of the risky asset follow a Geometric Brownian Motion (GBM), also called Itô-process. So, the price of the risky asset P_j satisfies the stochastic differential equation

$$dP_{j,t} = \mu_{j,t} P_{j,t} dt + \sigma_{j,t} P_{j,t} dZ_{j,t}, \tag{3.1}$$

where $Z_{j,t}$ are standard Brownian Motions. Consistent with the Black-Scholes-Merton approach, the parameters μ_j and σ_j are supposed to be constants - fixed over time - to attain one time-constant solution.

The fractions under the Normal assumption of the individual assets are given in the first line of Table 3.2. In this subsection the parameters under the assumed distributions assumption are given, hence, the investment fraction is indeed constant over time (Breiman, 1961). Betting $f_{S\&P} = 1.2879$ for the S&P 500 implies betting 128.79% of the initial capital in the risky asset and shorting the risk-free asset by 28.79% in order to leverage the portfolio. This is the full Kelly strategy, equivalent to $\gamma = 0$ from the isoelastic marginal utility function $U(C) = \frac{1}{\gamma} C^{\gamma}$, which is the natural log in this case. Betting half Kelly implies halfing the fraction for the risky asset (Merton, 1992). This implies $\gamma = 0.5$, closer to the empirically relevant, risk-averse values. Note that the investment fraction for the bond-market is highly levered, due to the fact that the estimated volatility (4.06% p.a.) of the asset return is small compared to the other assets, still giving 4.6% p.a. over the series. The Standard and Poor's 500 for example gave an annual return around 5.5% with according standard deviation of 20%.

The calculated fractions under the Student-T- and GEV-assumption in lines two to four in table 3.2 utilize the result of Osorio (2008). The fractions differ significantly from the fractions under the Gaussian. Especially, the centre of the distributions is shifted, leading to e.g. higher fractions for the stock index investments. The fractions under the non-parametrical distribution assumptions, where the moments are calculated numerically, are close to the fractions under the Gaussian. I will test whether the wealth trajectories are significantly different. As the portfolio strategy will rest on the first two moments without further restriction in the optimization, the non-parametric strategy, not using higher moments, should be close to the strategy under the Gaussian.

Fractions	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	REstate	Bond	Int. (10y)	Berkshire	DBK
Gaussian	1.28	1.68	0.67	1.37	-0.96	-0.19	0.55	27.27	-0.62	1.83	-0.27
Student-T	1.71	3.46	2.48	3.99	0.78	-0.08	0	24.89	-1.14	0.24	-0.04
GEV	6.25	5.19	4.74	6.81	2.15	2.37	3.74	32.11	1.31	6.79	2.46
Nonparametric	1.26	1.64	0.66	1.34	-0.93	-0.18	0.54	26.41	-0.6	1.8	-0.26

Table 3.2: Fractions for the assets, given different distribution assumptions for Osorio (2008)

3.1.2 Wealth dynamics, Gaussian assumption

The eleven wealth paths $W_{t,j}$ start with wealth 100 in $t = 0$. The wealth dynamics can be described as

$$W_{t,j} = \left[f_j^\top x_t \right] W_{t-1,j} \tag{3.2}$$

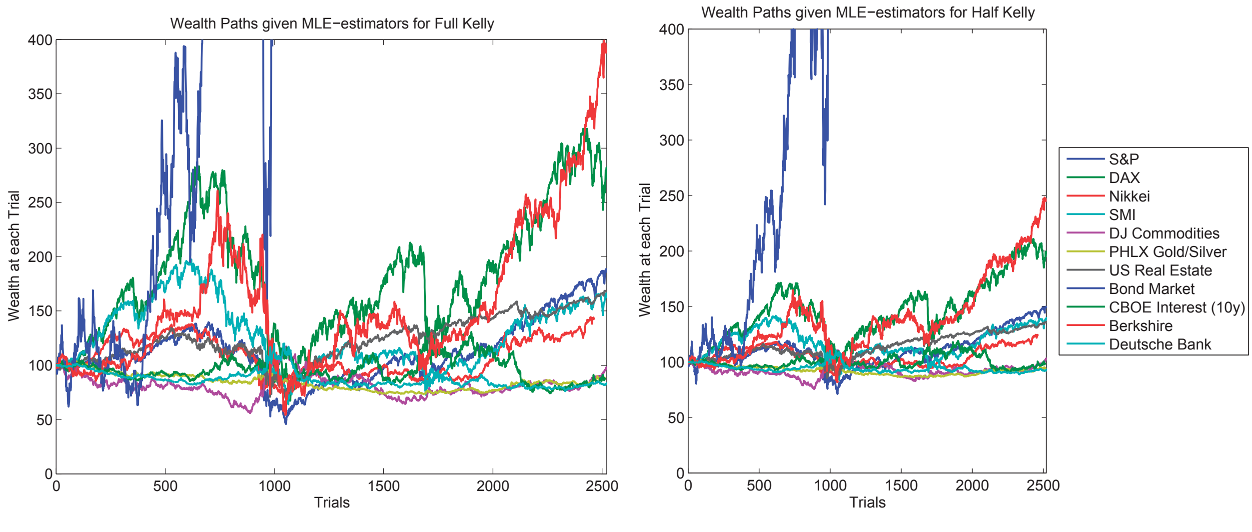
giving in the end of the investment period

$$W_{T,j} = W_0 \prod_{t=1}^{T} \left[f_j^\top x_{t,j} \right] \tag{3.3}$$

In this subsection the fractions upon the Normal distribution assumption are used, hence the time index for the fraction can be dropped. The wealth trajectories for the full Kelly strategy, $\gamma = 0$, are given in figure 3.2a with according statistics in table 3.3. Note, that the wealth path for the bond-market is beyond the wealth paths of the other ten investment strategies. The final wealth for the full Kelly strategy reaches from the Deutsche Bank investment giving 83.98 (-1.69% p.a.) to the bond-market investment paying off with 57180 (88.7% p.a.) in the end. This gives an average of 5323 but median wealth for the Real Estate investment of 166.91 (5.26% p.a.).

The wealth trajectories for the half Kelly strategy, $\gamma = 0.5$, are given in figure 3.2b with according statistics in table 3.4. The final wealth for the half Kelly bet reaches from the Deutsche Bank investment giving 93.155 (-0.689% p.a.) to the bond-market investment paying off with 11027 (60.09% p.a.) in the end. The average wealth trajectory gives 1128.7 after ten years of investing. Still the median trajectory gives 135.19 (3.06% p.a.) over the assets (Real Estate investment).

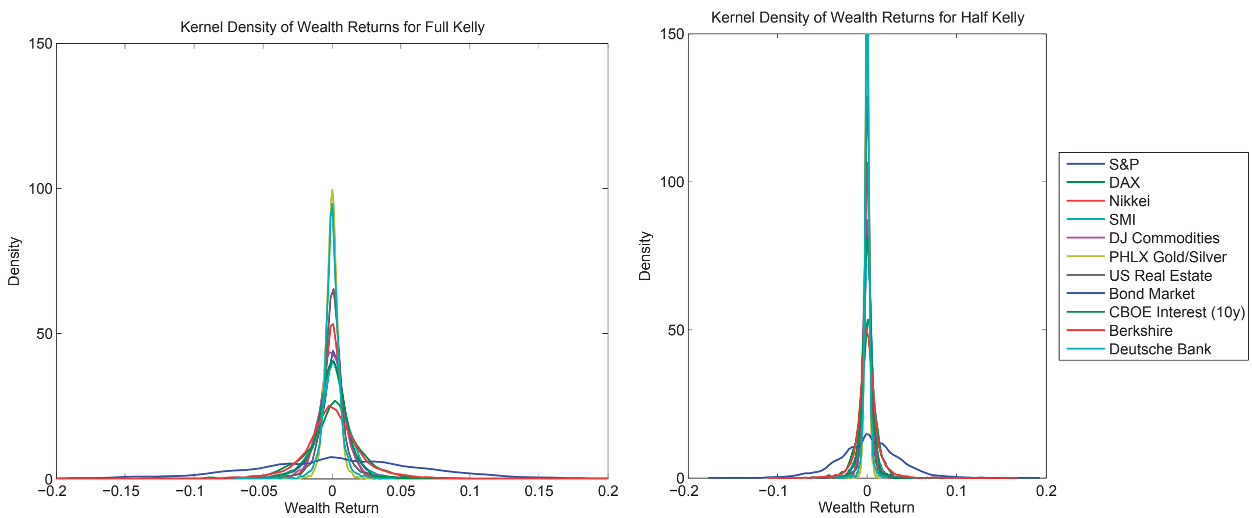
I would like to emphasis that the investments are done separately, without knowledge of the other one. If the true first two moments would be known, a confidence interval according to ML-theory could be set up easily to test whether the mean is statistically different from zero. If this would have been done with a large alpha (type one error) of 40%, only three (DAX, bond-market and Berkshire) of the eleven assets would be traded, solely long positions. Nevertheless, utilizing the Kelly Strategy without any constraint through the series of ten years implies, although the first two moments are known, to sustain severe draw-downs through the financial crisis of 2008.



(a) (b)

Figure 3.2: Wealth paths for full (left) and half (right) Kelly, in-sample MLE

Although the average final wealth is reduced due to betting below the full Kelly strategy, the standard deviation and according draw-downs are reduced heavily, as in the simulation chapter 2. The wealth returns are visualized by Kernel densities in figure 3.3 for both strategies. The more risk-averse the betting strategy is, the more narrow is the probability mass around the average return. The mean-variance trade-off is existent again. Although the full Kelly strategy offers a substantially higher final wealth, the risk-averse investor would certainly give off a substantial amount of financial gains in order to reduce draw-downs over the series. For the average wealth trajectory, the Swiss Market Index, the wealth reduces by 17.16% for the half Kelly strategy, but the difference in minimal wealth at each day is 24% (from 63.925 down to 83.525 for half Kelly).



(a) (b)

Figure 3.3: Kernel density of wealth returns for full (left) and half (right) Kelly, in-sample MLE

Wealth paths (in-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	185.04	302.98	141.77	169.03	99.51	87.51	166.91	57180	88.25	386.93	83.98
Min Wealth	45.59	67.05	74.23	63.92	55.61	73.25	69.34	61.71	73.88	54.16	75.99
Max Wealth	188.71	318.45	144.39	197.38	117.28	103.58	168.52	60126	138.57	399.02	119.93
Mean Return (p.a.)	0.063	0.115	0.036	0.053	-0.0005	-0.013	0.052	0.887	-0.012	0.145	-0.0168
Std Return (p.a.)	0.26	0.37	0.16	0.24	0.17	0.08	0.19	1.1	0.2	0.41	0.11
Skewness	-0.4	-0.08	-0.48	-0.09	0.12	-0.51	-0.23	-0.2	-0.13	0.53	-0.93
Kurtosis	14.02	9.66	11.21	11.5	5.62	11.2	16.6	4.45	6.13	17.64	14.96
Sharpe Ratio	0.24	0.31	0.21	0.21	-0.003	-0.16	0.27	0.8	-0.06	0.34	-0.14
Sortino Ratio	0.32	0.43	0.29	0.3	-0.004	-0.23	0.39	1.13	-0.08	0.51	-0.2
Min Return	-0.12	-0.12	-0.08	-0.11	-0.05	-0.05	-0.12	-0.34	-0.06	-0.22	-0.07
Max Return	0.14	0.17	0.09	0.14	0.06	0.03	0.08	0.32	0.09	0.3	0.04

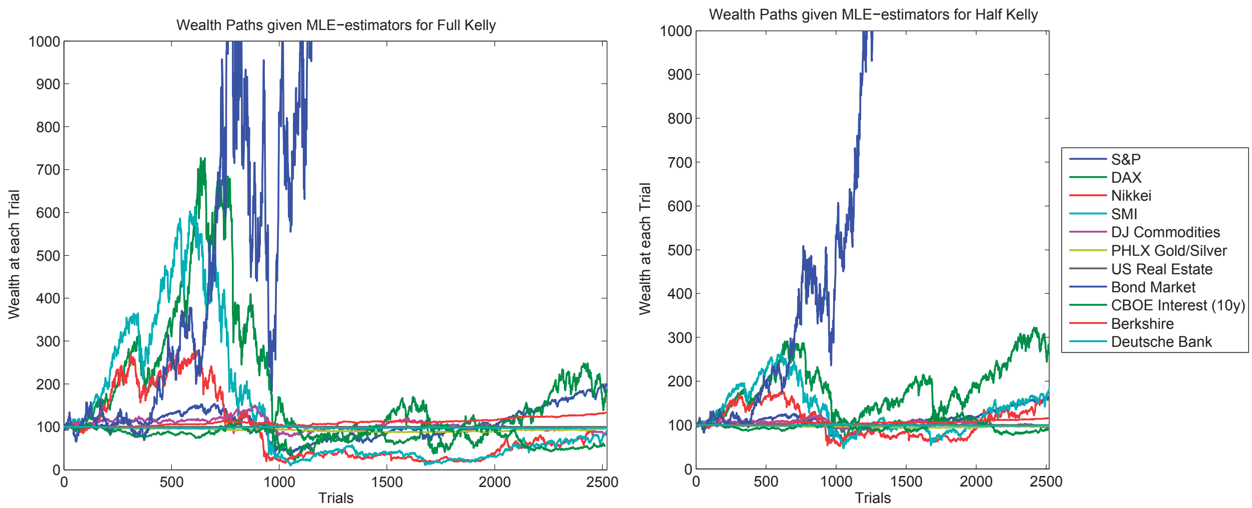
Table 3.3: Descriptives of the full Kelly wealth paths, in-sample MLE

Wealth paths (in-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	148.35	207.12	123.33	140.26	103.62	94.3	135.19	11027	98.92	244.33	93.155
Min Wealth	71.01	88.96	87.35	83.52	75.47	86.09	85.54	80.36	90.25	81.58	88.56
Max Wealth	149.82	212.06	124.45	142.68	110.54	101.78	135.84	11225	119.04	248.1	110.24
Mean Return (p.a.)	0.04	0.074	0.022	0.034	0.004	-0.006	0.031	0.601	-0.001	0.094	-0.007
Std Return (p.a.)	0.13	0.18	0.08	0.12	0.09	0.04	0.09	0.55	0.1	0.21	0.06
Skewness	-0.24	0.06	-0.41	0.03	0.16	-0.47	-0.09	-0.03	-0.08	0.86	-0.86
Kurtosis	14.07	10.04	11.18	11.82	5.63	10.97	16.29	4.37	6.2	19.54	14.63
Sharpe Ratio	0.31	0.4	0.25	0.27	0.04	-0.14	0.32	1.08	-0.01	0.44	-0.12
Sortino Ratio	0.42	0.56	0.35	0.38	0.06	-0.2	0.46	1.59	-0.01	0.67	-0.16
Min Return	-0.06	-0.06	-0.04	-0.05	-0.03	-0.03	-0.06	-0.16	-0.03	-0.11	-0.03
Max Return	0.07	0.09	0.05	0.07	0.03	0.02	0.04	0.17	0.05	0.16	0.02

Table 3.4: Descriptives of the half Kelly wealth paths, in-sample MLE

3.1.3 Wealth dynamics, Student-T assumption

In this subsection the fractions upon the Student-T distribution assumption are used (Osorio, 2008), implying that $x_j \sim Student(\mu_j, \sigma_j^2, \nu_j)$. Although the Student-T distribution captures the tails more appropriately, the centre of the distribution deviates too much from the values under the normal, leading to deviating fractions (see Table 3.2) in relation to the fractions under the normal and on average worse portfolio results. For the stock indices for example, the fractions are larger, leading to more severe draw-downs than under the normal. The results hold for the fractions and according wealth paths under the GEV-distribution. Therefore, for the out-of-sample test, the Student-T assumptions for the portfolio fraction will not be assessed.



(a) (b)

Figure 3.4: Wealth paths for full (left) and half (right) Kelly, in-sample MLE, Student-T

3.1.4 Wealth dynamics, non-parametric

Choosing the fractions according to Osorio (2008) with the according moments under the non-parametric distribution, the deviations to the fractions under the Normal, as argued, are small as only the first two moments are used for the estimation of the fraction. Using the two-sample Kolmogorov-Smirnov test to test whether the wealth returns under the Gaussian are the same as under the non-parametric moments, the null hypotheses that both distributions are equal, cannot be denied for p-values ≈ 0 for all assets. Hence, reducing computational complexity and therefore improving speed for the out-of-sample testing, I will consequently use the fraction under the Gaussian, as derived in Merton (1992).[1]

[1] As this chapter aims to test the approximations for the Kelly bet, instead of implementing the stochastic optimization problem, the first two moments should capture the main portfolio drivers (Chopra and Ziemba, 1993).

3.2 Out-of-sample backtest: i.i.d. with limited memory

In the preceding subsection 3.1, the first two unconditional moments of the underlying asset distributions have been known. As future moments are not known, the moments under the Gaussian are estimated upon past data.

The fractions $f_j(\mu_t, \sigma_t^2 | \mathcal{F}_{t-1})$ are estimated upon the filtration in t, covering the closing prices for the according series. After collecting four years of data it is assumed, that the investor forgets price data. Hence, in order to make his decision, four years of daily data are being used, rolling over the data period of ten years. In this chapter the Gaussian moments is chosen as first choice as it has the smallest variance of all unbiased estimators, for the Normal as approximation. In order to make the results comparable, the loading period (four years) is added to the data set (before the period of 2005-2015).

3.2.1 Investment fractions

Accordingly, figure 3.5 shows fractions, means and variances over the whole time period. After the loading phase from 2002-2005 the majority of the estimated means is positive, leading to long positions in most of the assets. The financial crisis, starting 2007/2008 led to plummeting prices, although the fractions first remained positive. The relatively long data-span for the estimators succeeds in slow adjustments of the mean estimates. In 2009 most mean estimates are negative, leading to overall short positions. Again, the estimates are lagging. From 2012 on the mean estimates and hence fraction begin to increase. The volatility estimates fail to capture the increased risk during the financial crisis, as the long data span effects the volatility estimate too slowly. The only exception is the bond market, whose mean and variance remain relatively constant over the whole data span, leading to a heavily levered long position. In the following chapter, the autocorrelated structure of squared returns will be used to improve the modelling of the second moment.

3.2.2 Wealth dynamics

The eleven wealth paths start with wealth 100. The wealth dynamics from formula 3.2 change to

$$W_{t,j} = \left[f_{t,j}^{\top} x_t \right] W_{t-1} \tag{3.4}$$

as the moment estimates are varying over time. In this subsection the fractions upon four years as rolling data window are used. The wealth trajectories for the full Kelly strategy, $\gamma = 0$, are given in figure 3.6a with according statistics in table 3.5. Note, that the wealth path for the bond-market is

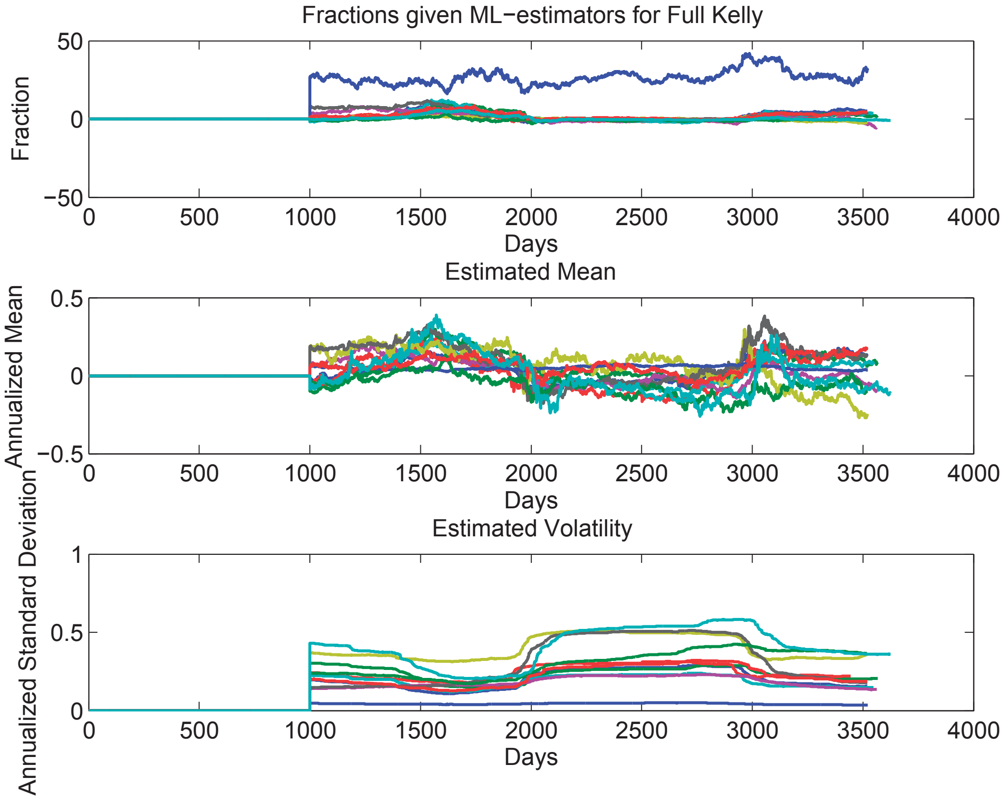
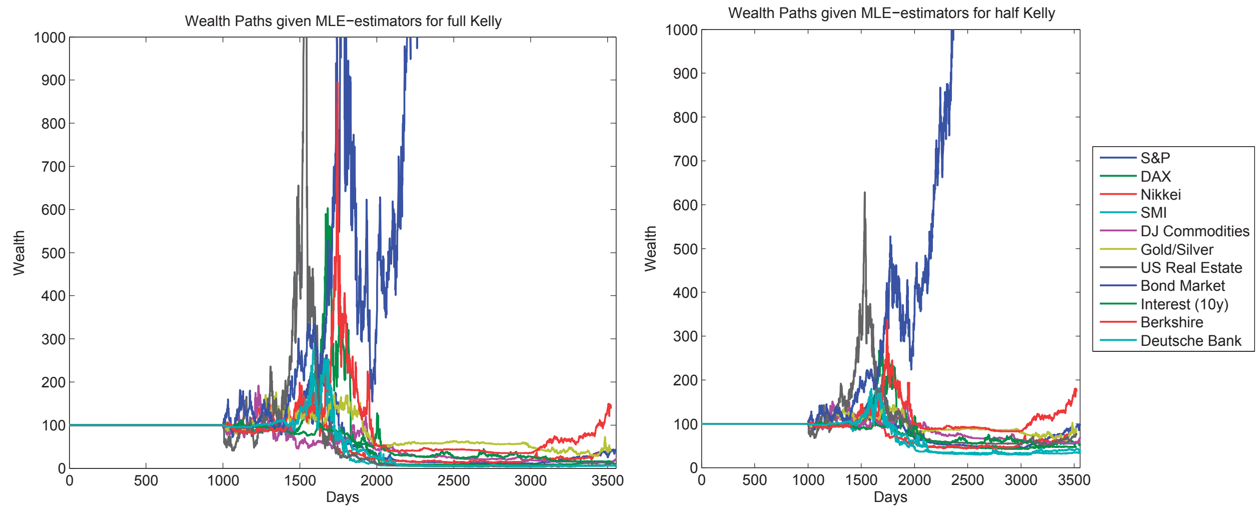


Figure 3.5: Fractions, means and variances over time, out-of-sample MLE, Gaussian

beyond the wealth paths of the other ten investment strategies. The final wealth for the full Kelly strategy reaches from the Deutsche Bank investment giving 5.6 (-18.16% p.a.) to the bond-market investment paying off with 13783 (42.2% p.a.) in the end, leading to an average of 1282.5 but median of 26.117 (-9.06% p.a.) over the assets.

The wealth trajectories for the half Kelly strategy, $\gamma = 0.5$, are given in figure 3.6b with according statistics in table 3.6. The final wealth for the half Kelly bet reaches from the Deutsche Bank investment giving 33.69 (-7.29% p.a.) to the bond-market investment paying off with 4826.3 (31.96% p.a.) in the end. Hence, the average wealth over the assets is 507.04. But, the median trajectory, the Real Estate investment, gives 73.893 (-2.14% p.a.). Decreasing the values for γ furthermore to 1/4 and 1/8 of full Kelly leads again to a lower mean of wealth, but higher median over the assets. Still, for the 1/8 Kelly strategy median wealth is 98.871. Consequently, risk averse betting in a setting of unknown future market returns is definitely advised. Nevertheless, past-dependent ML-estimates fail to capture recent changes in financial market prices. Till the financial market crisis the majority of the asset investment were able to generate enormous profits, as the markets went up steadily. The US Real Estate market, DAX, Berkshire and the bond-market were able to generate multiples of initial wealth within three years. Still, in contrast to the in-sample results, the wealth swings are significantly larger, producing losses on average. As the fractions through the financial crises remained positive, severe draw-downs are the consequence. Except the bond-market investment, most assets clearly fell below initial wealth, the Deutsche Bank investment suffered draw-downs up to 95.2% (69%) for the full (half) Kelly strategy.

In essence, leveraged portfolios with wrong parameter choice is deadly. The Kelly Criterion performs well out-of-sample if the moment estimates are robust over time, as seen for the bond-market investment. As the distribution for financial market returns is indeed not i.i.d., but time-varying, the financial crisis puts a spoke in the investors wheel.



(a) (b)

Figure 3.6: Wealth paths for full (left) and half (right) Kelly, out-of-sample MLE

The wealth paths (see tables 3.5 and 3.6) significantly differ from the wealth returns of the last chapter, when the in-sample moments have been known. In essence, for the out-of-sample back-test using ML-estimators from the past,

- the mean of the wealth returns is significantly smaller, indeed negative,
- the standard deviation has increased,
- the skewness of the wealth return becomes significantly negative and
- the kurtosis increases furthermore.

Investment strategies, which have been profitable, when the true moments have been known, lead to heavy draw-downs in wealth due to the financial crisis, as estimates are not capturing rapidly changing parameters. Still, risk averse betting reduces draw-downs on average and increase the median wealth accordingly.

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	39.97	10.13	31.31	7.2	26.11	40.87	8.16	13783	15.99	137.68	5.6012
Min Wealth	9.48	6.03	13.03	3.64	14	28.52	3.84	71.66	15.36	34.26	4.82
Max Wealth	266.36	497.86	192.09	290.82	191.48	175.96	1731	30601	127.61	894.78	254.6
Mean Return (p.a.)	-0.063	-0.149	-0.081	-0.168	-0.09	-0.062	-0.164	0.422	-0.123	0.0231	-0.181
Std Return (p.a.)	0.48	0.52	0.39	0.51	0.46	0.36	0.73	0.89	0.31	0.47	0.44
Skewness	-2.77	-3.21	-1.58	-2.44	-0.14	-0.78	-1.45	-0.42	-1.07	-1.47	-4.4
Kurtosis	38.03	57.46	19.22	35.95	15.38	15.01	17.42	6.65	31.01	36.87	81.24
Sharpe Ratio	-0.13	-0.28	-0.21	-0.32	-0.19	-0.17	-0.22	0.46	-0.38	0.048	-0.41
Sortino Ratio	-0.16	-0.35	-0.26	-0.41	-0.26	-0.22	-0.28	0.65	-0.51	0.07	-0.5
Min Return	-0.48	-0.63	-0.26	-0.51	-0.21	-0.21	-0.46	-0.43	-0.24	-0.45	-0.56
Max Return	0.19	0.25	0.15	0.2	0.29	0.17	0.29	0.24	0.2	0.28	0.2

Table 3.5: Descriptives of the full Kelly wealth paths, out-of-sample MLE

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	94.92	51.86	72.94	42.8	75.371	80.48	73.89	4826.3	47.75	173.26	33.68
Min Wealth	43.03	38.54	43.32	29.18	54.61	63.148	49.59	86.92	46.71	81.92	31.04
Max Wealth	181.14	248.2	144.56	180.68	150.69	142.79	628.73	5780.1	117.38	337.42	170.09
Mean Return (p.a.)	-0.003	-0.045	-0.023	-0.057	-0.019	-0.015	-0.021	0.319	-0.051	0.04	-0.072
Std Return (p.a.)	0.23	0.25	0.19	0.25	0.23	0.18	0.36	0.44	0.15	0.23	0.21
Skewness	-2.13	-2.14	-1.29	-1.77	0.17	-0.55	-0.95	-0.19	-0.63	-0.73	-3.17
Kurtosis	30.5	40.06	17.07	28.22	16.16	14.48	15	6.12	29.3	31.49	59.87
Sharpe Ratio	-0.016	-0.17	-0.11	-0.22	-0.084	-0.085	-0.059	0.71	-0.325	0.17	-0.33
Sortino Ratio	-0.02	-0.22	-0.15	-0.29	-0.11	-0.11	-0.078	1.02	-0.43	0.24	-0.429
Min Return	-0.21	-0.26	-0.12	-0.22	-0.1	-0.09	-0.2	-0.19	-0.11	-0.2	-0.24
Max Return	0.1	0.13	0.08	0.1	0.15	0.09	0.15	0.12	0.1	0.15	0.1

Table 3.6: Descriptives of the half Kelly wealth paths, out-of-sample MLE

3.3 Out-of-sample backtest: non-stationarity with limited memory

3.3.1 Autocorrelation in financial returns

As Maximum Likelihood Estimates under the Gaussian for the mean, based upon past data, are not sufficiently forecasting future market returns, the standard time-series approach to capture serial correlation is fitting models of the ARIMA(i,p,q)-family for the according price series P_j. A comprehensive analysis of univariate time-series is omitted, for reference see for example Tsay (2002).

As price series of financial market assets are often integrated of order one, the price series are transformed accordingly in order to utilize ARMA(p,q)-models. Notation-wise, moments of log-returns are going to be modelled, implying that $y_t = \log(P_t/P_{t-1})$. Thus the individual returns series $y_{t,j}$ are modelled as

$$y_t = \mu_t + \epsilon_t, \tag{3.5}$$

where $\mu_t = \mathrm{E}\left[y_t|\mathcal{F}_{t-1}\right]$ is the conditional mean given information set $\mathcal{F}$ in the last period $t-1$. In order to capture the low lag dependence in the data series, the conditional mean is modelled as

$$\mu_t = c + \sum_{p=1}^{P} \phi_p \mu_{t-p} + \sum_{q=1}^{Q} \theta_q \epsilon_{t-q}, \tag{3.6}$$

where ϕ_r is the autoregressive coefficient and θ_m is the moving average coefficient. As standard procedure, I will use AR(1)-forecasts, minimizing the mean square error, for all eleven markets. Varying lag lengths and ARMA(p,q) forms did not improve the out-of-sample performance of the conditional mean estimator on average. Finding one model structure for different series cannot be advised.

3.3.2 Comparing mean square error for mean estimators

From MLE-theory, the simple average, using all available data should perform best as it is the estimator, which has the smallest variances among all unbiased estimators. If there are significant autocorrelations, the AR(1) should outperform the ML-estimators in terms of root mean square error (RMSE).

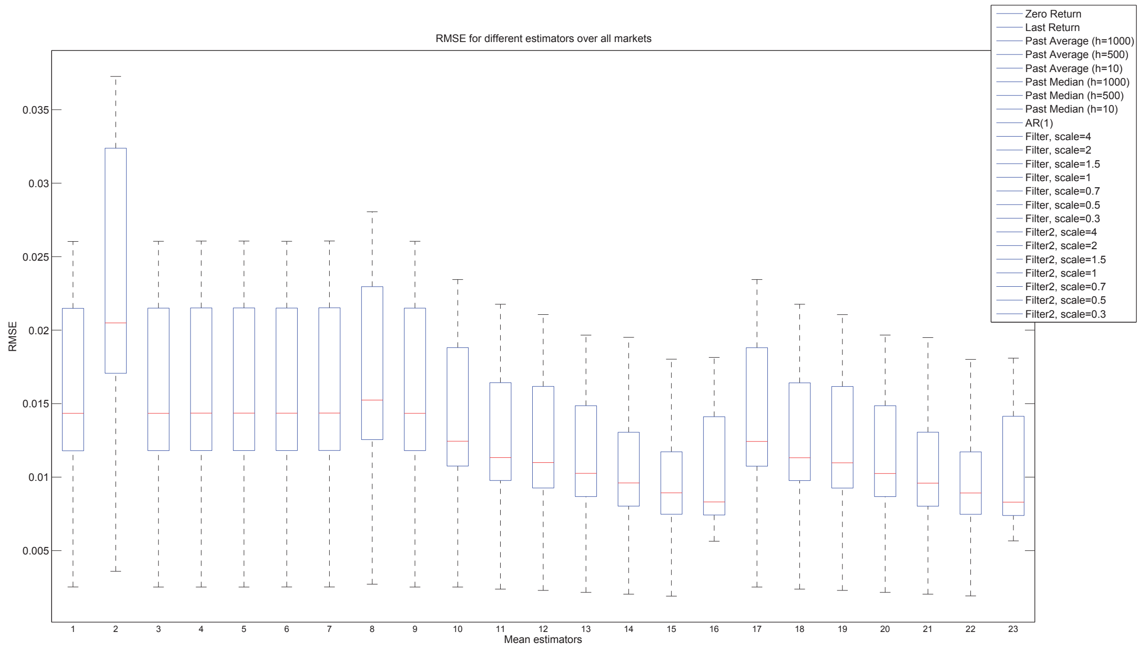


Figure 3.7: Root Mean Square Errors for different estimators over all markets

In order to compare some simple estimator choices for the mean, a short comparison in terms of Mean Square Error (RMSE) is presented in figure 3.7. The RMSEs for the estimators are calculated for the eleven assets, brought together in a box-plot. The simple estimator choices are as follows:

- The first estimator is setting the forecasted mean to zero.
- The second estimator uses the last return as forecast for the return today.
- The third to fifth estimators use the mean with rolling data windows (h) of 1000, 500 and 10.
- The sixth to eighth estimators use the median with rolling data windows (h) of 1000, 500 and 10.
- The ninth estimator makes use of the AR(1)-forecast.

Out of the the first nine estimators, the zero-return forecast has the smallest RMSE over the given markets. The only exception is the bond-market, exhibiting significant and also in absolute parameter values relevant autocorrelation structure, thus, improving the forecasting performance. The Mariano-Diebhold-Test supports this result over a large majority of tests. The straight forward consequence would be to invest solely in the bond-market as the other estimators are not outperforming the zero-return forecast. This is the crucial reasoning behind the results of section 3.2. Nevertheless, the investment would assume that the underlying parameters remain robust for the future.

3.3.3 Autocorrelation in squared returns

As Maximum Likelihood Estimates under the Gaussian for the variance based upon past data are not sufficiently forecasting future market uncertainty, the standard time-series approach to capture serial correlation is fitting models of the APARCH(p,q,δ) family. In the elementary paper of Engle (1982) the residual part of the ARMA-equation was rewritten as autoregressive conditional heteroscedastic (ARCH) process:

$$\epsilon_t = \sigma_t Z_t,\ \ Z_t \sim F(0,1). \tag{3.7}$$

The assumed conditional distribution is chosen to be Gaussian. The distribution assumption can nevertheless be widened to Student-T, Skew-T, GED, Skew-GED, Gram-Charlier and Cauchy distributions. The time varying conditional variance of the returns is $\sigma_t^2 = Var\left[y_t|\mathcal{F}_{t-1}\right]$, a function of the past residuals and further, past conditional variances (Bollerslev, 1990). The Asymmetric power ARCH model, introduced by Ding, Granger, and Engle (1993), offers a flexible form of the conditional variance equation, which incorporates several of the later presented ARCH-models:

$$\sigma_t^\delta = \omega + \sum_{p=1}^{P} \beta_p \sigma_{t-p}^\delta + \sum_{q=1}^{Q} \alpha_q \left(|\epsilon_{t-q}| - \gamma_q \epsilon_{t-q}\right), \tag{3.8}$$

where $\omega, \delta \in \mathbb{R}^+, \alpha_q, \beta_p \in \{\mathbb{R}^+, 0\}$ and $-1 < \gamma_q < 1$.
The GJR(1,1)-model (Glosten, Jagannathan, and Runkle, 1993) will be used ad-hoc for all markets, implying $\delta = 2$ and $0 \leq \gamma_j < 1$, in order to capture the leverage effect.

3.3.4 Investment fractions

Although the analysis using mean square errors has shown that the mean estimators upon the past average of returns are worse than setting the forecast to zero, except for the bond market, I aim to show that the conditional variance estimators from the GJR(1,1) offer improvements. Utilizing again the results of Merton (1992) and Osorio (2008), the according fractions are given in figure 3.8. The fractions utilizing the AR(1)-GJR(1,1)-forecasts for the different financial returns differ significantly from the fractions from the last section 3.2 when past-dependent ML-estimators have been chosen. Centrally, with exception for the time-series forecast for the bond-market, the forecasts for the financial returns are on average not giving improvements in terms of MSE (see figure 3.7). The crucial element in the rapidly changing fractions are the time-varying volatility estimates, leading to smaller (higher) investment fractions in times of high (low) anticipated volatility, especially in the financial crisis 2007/2008. [2]

[2]Prospectively, I will use high-frequency data for further research papers for the estimation of uncertainty in the market.

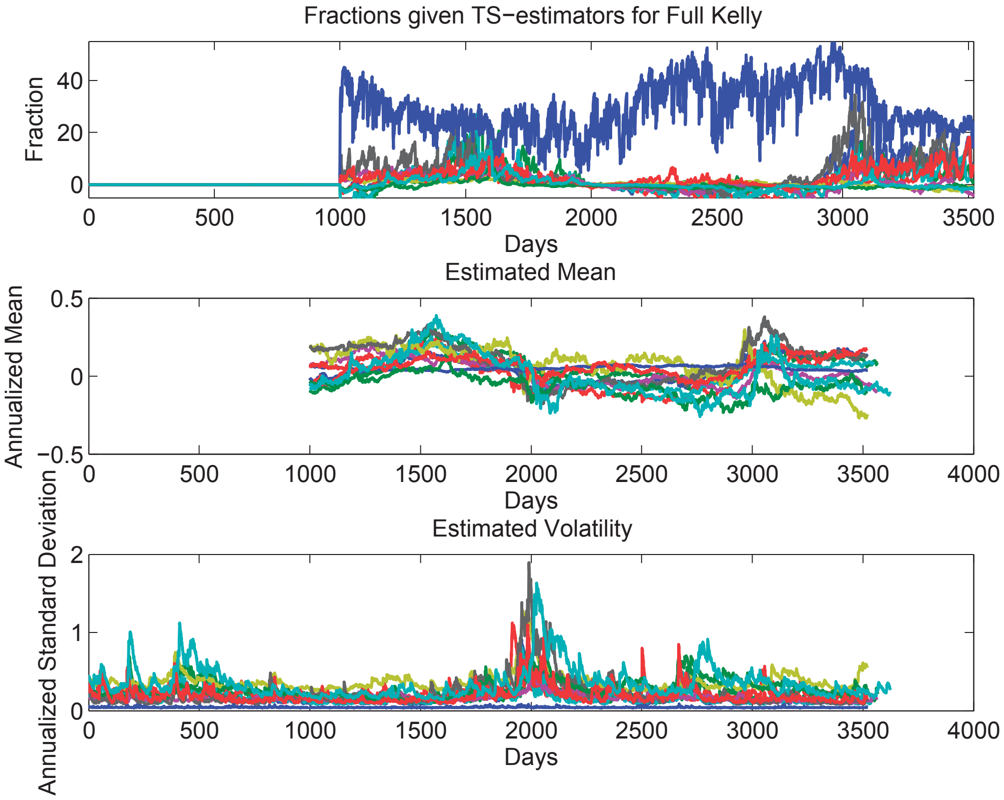
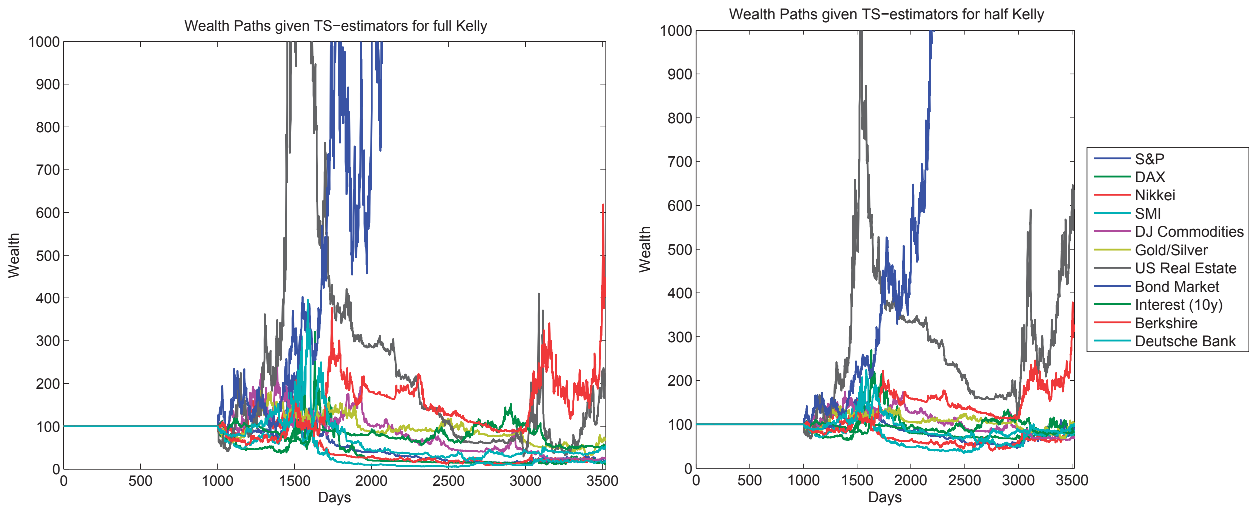


Figure 3.8: Fractions, means and variances over time, out-of-sample Time-Series

3.3.5 Wealth dynamics

The relevant wealth paths are given in figure 3.9 with according statistics for the full Kelly bet ($\gamma = 0$) in table 3.7 and for the half Kelly bet ($\gamma = 0.5$) in table 3.8. Using the presented time-series estimators, the full Kelly strategy reaches from the DAX-investments (-17.99% p.a.) to the bond-market investment giving 78905 (94.58% p.a.) in the end. Average over the markets, the average final wealth is 8271.9, whereas the median is 41.48 (-8% p.a.). For the risk averse half Kelly portfolio the Nikkei-investment realizes an end portfolio value of 67.9 (-3.9% p.a.) for the market performing worst and gives 14349 (64.17% p.a.) for the bond-market investment. The average wealth of the half Kelly bet is 1497.8 with according median of 87.26 (-1.3%). The half Kelly strategy performs better, in terms of mean, standard deviation and kurtosis than the full Kelly strategy, for nine out of eleven markets!



(a) (b)

Figure 3.9: Wealth Paths for full (left) and half (right) Kelly, out-of-sample Time-Series

The final wealth statistics in table 3.7 and 3.8 significantly differ from the returns using past-dependent ML-estimators. Over the different markets

- the mean of the wealth returns is increased on average for full and half Kelly,
- the standard deviation is increased on average and
- skewness and the kurtosis are neither coherently increasing nor decreasing.

In essence, conditional volatility estimates improve the results for the financial market investments significantly. Risk averse betting is again advised, as it reduces losses through the financial crises dramatically. Nevertheless, the mean estimator, which is lagging behind recent market changes, leads to wrong parameter estimates in and after the financial crises. Only the estimates for the Bond-market remain robust. Without robust parameters choice, the majority of the markets should not be traded.

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	17.51	13.24	19.59	16.62	38.67	67.67	193.39	78905	54.22	402.48	41.7
Min Wealth	10.25	10.52	8.31	4.8	21.37	32.18	28.72	74.01	45.03	55.68	19.01
Max Wealth	261.79	321.13	207.89	311.43	240.7	183.34	7391.4	1.635e+05	152.14	619	395.67
Mean Return (p.a.)	-0.15	-0.179	-0.15	-0.162	-0.089	-0.038	0.0681	0.945	-0.059	0.149	-0.081
Std Return (p.a.)	0.75	0.8	0.59	0.72	0.51	0.4	1.08	1.14	0.31	0.59	0.5
Skewness	-5.7	-6.33	-1.51	-5.22	-0.2	-0.39	-2.87	-0.25	-0.23	-0.59	-1.34
Kurtosis	109.01	135.08	17.432	114.47	9.15	7.87	42.56	4.49	10.46	11.71	24.08
Sharpe Ratio	-0.21	-0.22	-0.26	-0.22	-0.17	-0.09	0.063	0.83	-0.19	0.25	-0.15
Sortino Ratio	-0.25	-0.27	-0.33	-0.27	-0.23	-0.13	0.08	1.17	-0.26	0.35	-0.21
Min Return	-1.06	-1.19	-0.42	-1.04	-0.24	-0.2	-0.99	-0.4	-0.14	-0.32	-0.42
Max Return	0.18	0.24	0.14	0.243	0.22	0.14	0.41	0.28	0.16	0.19	0.22

Table 3.7: Descriptives of the full Kelly Wealth Paths, out-of-sample Time-Series

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	85.5	81.91	67.9	79.26	87.25	100.52	591.34	14349	83.18	310.2	90.02
Min Wealth	45.24	61.96	39.2	34.73	64.24	67.78	67.28	89.73	75.17	76	56.73
Max Wealth	176.58	269.47	156.24	207.17	174.82	141.69	1454.5	17611	133.86	378.23	224.67
Mean Return (p.a.)	-0.015	-0.019	-0.039	-0.022	-0.013	0.0005	0.194	0.642	-0.018	0.119	-0.01
Std Return (p.a.)	0.35	0.37	0.29	0.34	0.25	0.19	0.51	0.56	0.15	0.29	0.25
Skewness	-3.35	-3.27	-1.13	-2.68	-0.007	-0.27	-1.35	-0.07	-0.09	-0.3	-0.85
Kurtosis	50.05	56.53	13.59	50.82	9.03	7.49	22.82	4.32	10.5	10.77	20.06
Sharpe Ratio	-0.04	-0.05	-0.13	-0.06	-0.05	0.003	0.37	1.13	-0.11	0.41	-0.04
Sortino Ratio	-0.05	-0.06	-0.17	-0.08	-0.07	0.004	0.51	1.64	-0.162	0.58	-0.05
Min Return	-0.39	-0.42	-0.19	-0.39	-0.11	-0.09	-0.37	-0.18	-0.07	-0.14	-0.18
Max Return	0.09	0.13	0.07	0.13	0.12	0.07	0.22	0.15	0.08	0.1	0.12

Table 3.8: Descriptives of the half Kelly Wealth Paths, out-of-sample Time-Series

3.4 A two-regime approach

The portfolio back-tests using time-series estimators improved the results compared to the out-of-sample ML analysis. Still, there are major drawbacks:

- the AR(1)-forecasts are not coherently outperforming the naive random walk assumption and
- due to the delayed mean estimates, all position within the financial crisis are down-scaled long positions. Vice versa, after the financial crises, the majority of the mean estimators went negative, leading to up-scaled short-positions.

A starting point, in order to resolve the given problems, will be implemented as follows:

- The mean estimates will be fixed according to modified long-term estimates by Siegel (1998), implicating long positions only. This assumes a positive long-term drift. For further research I will use separate time-series analysis, comparable to the AR(1) for the bond-market, in order to verify markets, in which an edge can be gained. A real investment should only take place, given a robust edge in the aimed market.
- As investments will be solely on the long-side, the positions will be completely closed / hedged, if the conditional-volatility forecast from the GJR(1,1)-model lies above the scaled unconditional volatility from past data using MLE under the Normal.

Hence,

$$\begin{array}{ll} \text{if} \quad \sigma_{t|GJR} > \alpha\ \sigma_{t|MLE}, & f_t = 0, \\ \text{else} & f_t = f(\mu, \sigma_{t|GJR}), \end{array}$$

with $\sigma_{t|GJR}$ as forecast from the GJR(1,1)-model, $\sigma_{t|MLE}$ as ML estimator from the past and $\alpha \in \mathbb{R}^+$.

3.4.1 Mean square error - revisited

The forecasting analysis of different choices of mean estimators has shown on average that the random walk hypothesis cannot be rejected, except for the bond-investments. The two-regime model structure implies setting the mean-forecast to NaN, if the conditional forecast for the volatility from the GJR-model is larger that the unconditional expectation of the volatility given the past. Hence, we are not setting a forecast for the mean in times of high uncertainty.

Going back to figure 3.7, estimators ten to sixteen are chosen upon the regime, using past-dependent ML-estimators with a rolling period of 1000 days for different values of α. The two-regime approach leads to smaller mean square errors for the mean estimates. As scaling factor α decreases, the faster the MSE decreases, but the shorter are the investment periods. For estimators 17 to 23 from figure 3.7, the mean estimate vector is chosen according to long-term estimates, given by $\mu = [0.06\ 0.07\ 0.05\ 0.06\ 0.02\ 0.02\ 0.02\ 0.03\ 0\ 0.08\ 0.03]$ for the eleven assets. Over the different markets the mean estimators with the smallest MSE are estimators 14/15 and 22/23, using values of $\alpha < 1$. As the fixed mean estimators are not significantly different from the ML-estimators in terms of MSE, the fixed estimators are chosen for the purpose of robust fractions over time.

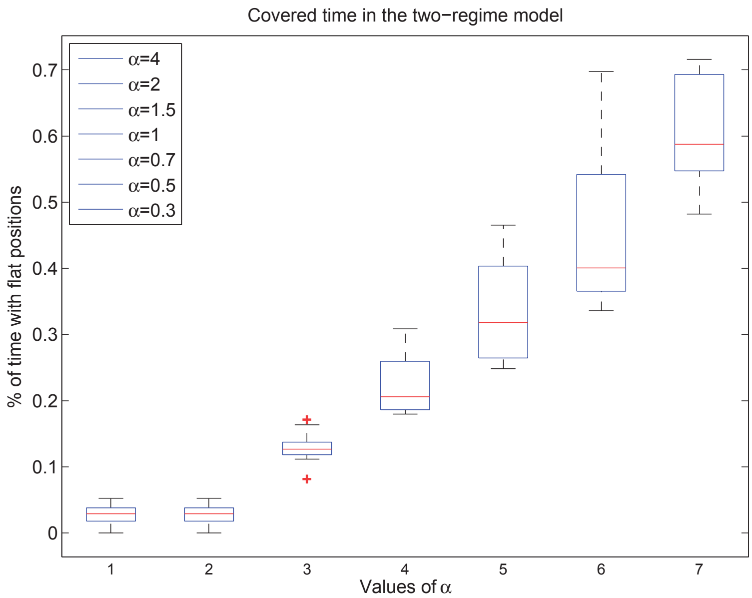


Figure 3.10: Covered time in the two-regime model

For the choice of alpha, the analysis in terms of MSE would suggest choosing $\alpha < 0.5$. Calculating the time period in % in which asset investments are covered / hedged, see figure 3.10, decreasing values of α lead to decreasing periods of times being invested. In order to maintain an average trading period around 80%, α is chosen to be one.

3.4.2 Investment fractions

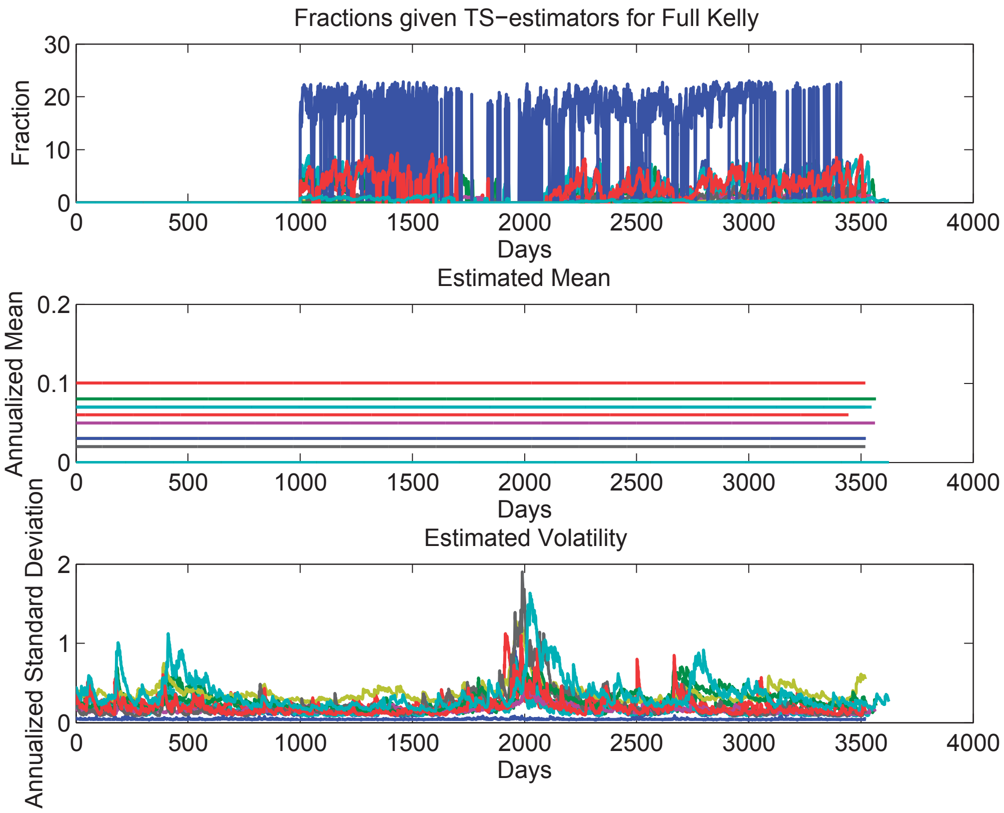


Figure 3.11: Fractions, means and variances over time, out-of-sample time-series/fixed

The main implications are that

- in periods of low uncertainty, I assume a long-term drift to be existent in the price series. In periods of high uncertainty, I will avoid any exposure in the invested market. The strategy is linked to the leverage effect for econometric time-series models (Nelson, 1991).
- An additional advantage is that for long times of low uncertainty, the unconditional volatility remains relatively low, implicating a direct cancellation of the position if the conditional volatility is increasing only slightly. In contrast, after a crises, unconditional volatility is relatively large, giving room for long investments profiting from a presumed, fast recovery.

The principle for the DAX investment can bee seen in figure 3.12. The first 1000 days from 2001-2005 are the loading period, in which no investments take place. The DAX is normalized to 100, as the wealth of the Kelly growth-investment at the beginning of 2005 in order to compare the series. Before the financial crisis levered long positions lead to an out-performance in the market. Within the financial crisis the conditional volatility forecasts from the GJR-model exceed the unconditional volatility from the ML-estimates, leading to squared positions in the DAX. Hence, draw-downs could be avoided.

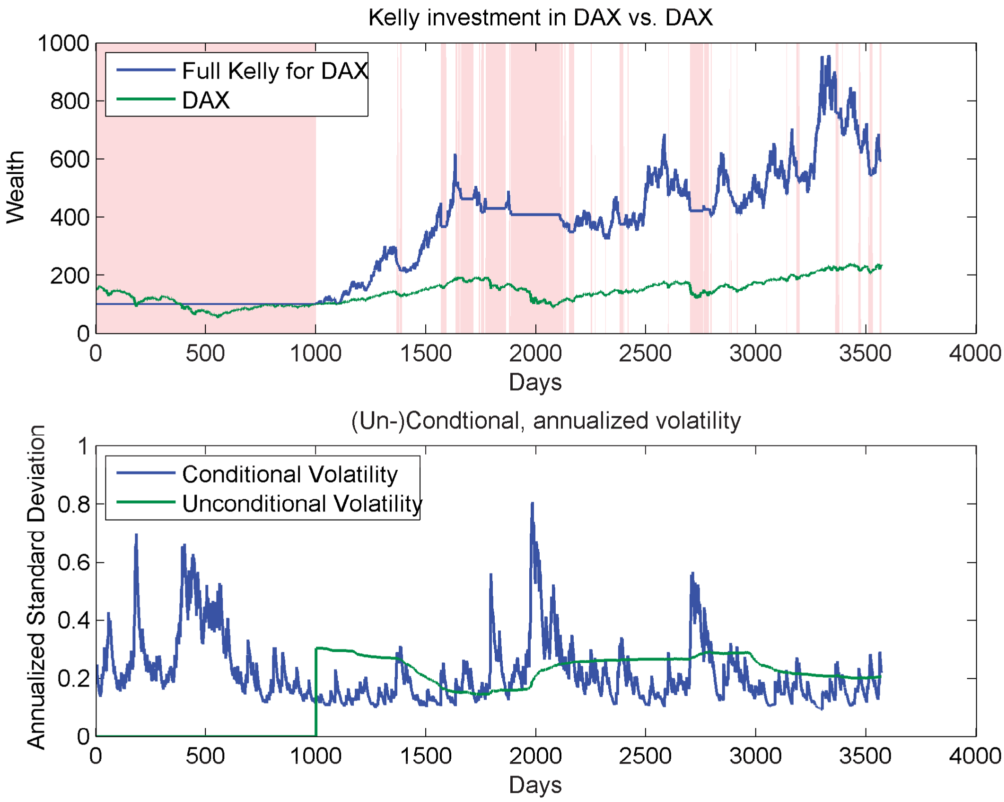


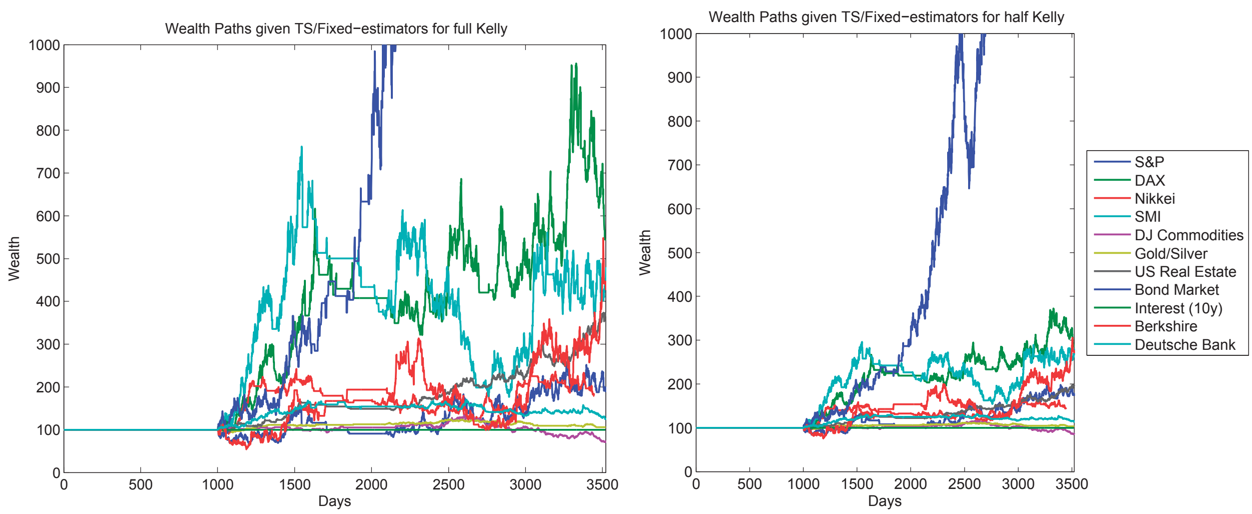
Figure 3.12: Two-regime example for the DAX-investment

Still, there are drawbacks, which have not been resolved yet,

- mainly as chosen mean estimates are chosen arbitrary from the past. Thus, I assume that the long-term trend will be existent in the future.
- The scaling factor α of the unconditional volatility is set to one in order to find a calibration for the trade-off between decreasing MSE and reduced trade frequency.
- Draw-downs can only be avoided if the prices decreases are sharp, inducing rapidly increasing volatility forecasts. As the same applies to large upswings in prices, prospectively, the measure for uncertainty needs to be varied to capture downside risks in particular.

3.4.3 Wealth dynamics

The wealth trajectories given the two-regime setup are given in figure 3.13 with according statistics in table 3.9 and 3.10. For the full Kelly bet ($\gamma = 0$), the least performing market was the Deutsche Bank investment, giving 83.985 (-1.7% p.a.) after then years. In contrast, the bond-market investment gave 57180 (88.72% p.a.). Averaging over the markets the average final wealth is 1591.8, whereas the median wealth trajectory of the S&P500 ends up with 187.53 (6.5% p.a.). For the risk-averse half Kelly bet ($\gamma = 0.5$), Deutsche Bank remains the least performing investment ending at 93.155 (-0.69% p.a.). The bond-market investment reduces to 11027 (60.1% p.a.). On average, the final is 315.77, whereas the median end wealth from the S&P500-investment is 172.88 (5.6% p.a.). Remembering the time series of the original assets over the horizon of 2005-2015, this is a significant improvement,



(a) (b)

Figure 3.13: Wealth Paths for full (left) and half (right) Kelly, out-of-sample Time-Series/Fixed, Gaussian

Except for the bond-market, in which it can be evaluated to drop the risky regime all-over, due to the robust parameters, the other markets strongly benefit by the usage of risk-averse betting. As tables 3.9 and 3.10 indicate, risk-averse betting reduces especially second and fourth moments of the according investments. The Berkshire-investment for example, exhibits for the full (half) Kelly bet an annual return of 15.59% (10.44%) with according standard deviation 46.36% (23.22%). Giving up one third of the return for reducing the standard deviation by one half seems to be a reasonable trade-off.

In contrast to past section 3.3, tables 3.9 and 3.10 demonstrate that

- the mean has decreased solely due to the bond-market investment, which is not dependent on the risky regime. The median strongly increases in contrast to the use of time-series estimators without the usage of the risky regime.
- The standard deviation of wealth returns decreased substantially
- skewness becomes less negative and
- kurtosis tends to significantly smaller values.

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	187.53	594.03	178.91	435.32	68.016	105.59	364.78	14896	100	426.06	129.19
Min Wealth	67.454	99.9	90.18	97.23	68.02	93.95	95.65	92.26	100	54.42	100
Max Wealth	258.3	955.99	252.32	761.88	133.59	124.17	373.57	16102	100	547.79	170.41
Mean Return (p.a.)	0.065	0.191	0.062	0.156	-0.037	0.005	0.138	0.648	0	0.156	0.025
Std Return (p.a.)	0.43	0.4	0.23	0.43	0.11	0.05	0.12	0.54	0	0.46	0.11
Skewness	-1.3148	-0.95	-0.53	-0.63	-0.78	-0.05	-0.19	0.24	NaN	0.16	0.02
Kurtosis	13.168	9.15	7.37	7.84	9.71	6.53	9.28	5.87	NaN	7.82	7.7
Sharpe Ratio	0.15	0.47	0.26	0.35	-0.34	0.1	1.16	1.19	NaN	0.33	0.23
Sortino Ratio	0.19	0.64	0.36	0.48	-0.44	0.15	1.71	1.8	NaN	0.49	0.33
Min Return	-0.31	-0.2	-0.11	-0.19	-0.05	-0.02	-0.05	-0.14	0	-0.19	-0.04
Max Return	0.12	0.12	0.06	0.17	0.04	0.01	0.04	0.19	0	0.16	0.04

Table 3.9: Descriptives of the full Kelly Wealth Paths, out-of-sample time-series/fixed

Wealth Paths (Out-of-sample)	S&P	DAX	Nikkei	SMI	DJ Com.	PHLX	Real Estate	Bond Market	Interest (10y)	Berkshire	DBK
End Wealth	172.88	299.35	142.79	265.91	83.75	103.08	194.38	1766.9	100	270.05	115.34
Min Wealth	83.97	100	95.24	98.78	83.769	96.95	97.82	97.11	100	75.8	100
Max Wealth	197.62	372	167.66	295.71	116.34	111.63	196.64	1826.2	100	305.25	131.74
Mean Return (p.a.)	0.056	0.113	0.037	0.102	-0.0172	0.003	0.069	0.332	0	0.104	0.0138
Std Return (p.a.)	0.21	0.19	0.11	0.21	0.05	0.02	0.06	0.27	0	0.23	0.05
Skewness	-1.1	-0.81	-0.47	-0.51	-0.74	-0.04	-0.15	0.37	NaN	0.31	0.05
Kurtosis	11.34	8.55	7.14	7.58	9.55	6.51	9.21	6	NaN	7.9	7.7
Sharpe Ratio	0.26	0.57	0.32	0.46	-0.31	0.12	1.16	1.22	NaN	0.45	0.26
Sortino Ratio	0.34	0.77	0.44	0.64	-0.41	0.17	1.71	1.89	NaN	0.66	0.37
Min Return	-0.14	-0.09	-0.05	-0.09	-0.02	-0.01	-0.02	-0.07	0	-0.09	-0.02
Max Return	0.06	0.06	0.03	0.09	0.019	0.007	0.02	0.1	0	0.087	0.02

Table 3.10: Descriptives of the half Kelly Wealth Paths, out-of-sample time-series/fixed

Results

The Kelly growth-optimum criterion is no holy grail. As long as the true data generating process of the outcomes is known, the presented strategy outperforms any betting scheme in the long run, which is significantly different. Under different parametric assumptions, the maximum of the expected logarithm of wealth can be solved analytically as presented. If there is no suitable closed-form solution, especially for multivariate distributions beyond the Gaussian, the optimal fraction can also be estimated by stochastic optimization methods.

In chapter 2, the theorems of Breiman (1961) could be validated in a simulation study, as the parameters were assumed to be estimated correctly. Betting the Kelly wager leads in the long run to *the* strategy, outperforming the risk-averse and risk-seeking counterparts in terms of mean and time to reach specific wealth targets. In short- and medium-term the growth optimum investment is nevertheless risky and contains significantly draw-down risks. Betting fractional Kelly strategies, such as half Kelly is advised. For the assumed Gaussian returns it holds that the growth optimum portfolio (full Kelly) is optimal in outperforming any other significantly different strategy and requiring the minimal time to reach a certain goal. This result holds for short-, medium- and long-term simulations as long as the first two moments of the joint probability distribution, from which was sampled, are known. At each trial, the growth optimum portfolio maximizes the expected logarithm of wealth. Even when the distribution is sampled from its non-parametrical form instead of the Gaussian, the results hold. Those convincing asymptotics break down when the true first two moments from the simulated distribution are not known. Using maximum likelihood estimators with different starting data spans as foundation for the portfolio fractions, the risks in terms of draw-down probabilities are increased significantly. Additionally, the full Kelly strategy is not maximizing the expected logarithm of wealth anymore. On average the full Kelly strategy outperforms its fractional and over-betting competitors, but relatively few, extreme upward trajectories lead to that result. Still, a simple investment based upon the time varying maximum likelihood estimators for the fraction derived in e.g.

Merton (1992) is not advised, especially when daily re-balancing is, due to transactions costs, not realistic. In the context of the simulation it would be sufficient to use larger loading periods, as the data generating process is assumed to be constant over time.

This holds for the empirical back-test for various financial markets in chapter 3. Even when the first two moments of the unknown future distribution are known, for the period covering the financial crisis of 2008, constant portfolio returns are not realistic. As the Kelly growth-criterion proposed levered positions, the according portfolios of the eleven chosen assets underlie rapid changes in wealth. Loosing the assumption of known moments by estimating the moments by maximum likelihood estimators from the past, intensifies the draw-downs significantly and leads to severe losses in the financial crises. The only exception is the bond-market investment, which underlying parameters are robust over time, favoring the use of the proposed Kelly bet. In order to overcome the heavy losses during the financial crises, the average portfolio wealth has increased by using time-series estimators, especially due to the conditional volatility estimates from the APARCH-model family. For the case of high conditional volatility the investment fraction is decreased substantially. For lower times of risk, the effect is vice versa. In essence, this comes at a price of increasing standard deviation of wealth returns, further decrease skewness and highly increasing kurtosis. After the financial crisis, the mean-estimators are turning negative, leading to, in combination with low volatilities, crucial miss-specification of the presumed data-generating process. A starting point for a real implementation is a proposed two-regime approach: At first, mean estimators are fixed according to long-term estimates and secondly, all positions are covered if the market is in the risky state, implying that time-series volatility forecast exceeds the volatility forecast from past MLE. For all tested strategies risk-averse strategies are calculated accordingly, indicating that a major reduction in variance or kurtosis can be realized by sacrificing, relative to the risk, a small amount of expected return.

There is not *one* ad-hoc method, which is worth implementing over various markets. A robust edge needs to be estimated for each market separately in order to use the strengths of the Kelly Criterion. The main statistical focus remains the improvement of the mean estimate in terms of errors, although I have shown that the cancellation of positions in times of high uncertainty decreases draw-downs substantially. Further research will be a two-fold: On the one hand the optimization for the Kelly bet needs to be restricted to measures of security in order to reduce especially tail risks. On the other hand, parameter estimation techniques, especially for the mean estimator, need to be improved in order to estimate a robust edge for a given investment opportunity.

Erklärung der Urheberschaft

Hiermit erkläre ich, Niels Wesselhöfft, dass ich die vorliegende Arbeit allein und nur unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Die Prüfungsordnung ist mir bekannt. Ich habe in meinem Studienfach bisher keine Masterarbeit eingereicht bzw. diese nicht endgültig nicht bestanden.

Niels Wesselhöfft
Berlin, February 28, 2016

Bibliography

Algeot, P., and T. Cover (1988): "Asymptotic optimality and asymptotic equipartition properties of log-optimum investment," *Annals of probability*, 16, 876 – 898.

Barron, A. R., and T. Cover (1988): "A bound on the financial value of information," *IEEE Transactions of Information Theorys*, 34(5), 1097 – 1100.

Bicksler, J., and E. Thorp (1973): "The capital growth model: an empirical investigation," *Journa of Financial and Quantitative Analysis*, 8(2), 273 – 287.

Bollerslev, T. (1990): "Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498 – 505.

Breiman, L. (1961): "Optimal gambling system for favorable games," *in Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, 1, 63 – 68.

Browne, S. (1997): "Survival and growth with a liability: Optimal portfolios in continuous time," *Math of Operations Research*, 22, 468 – 493.

——— (2000): "Risk-constrained dynamic active portfolio management," *Management Science*, 46(9), 1188 – 1199.

Chopra, V., and W. T. Ziemba (1993): "The effect of errors in the mean, variance, and covariance estimates on optimal portfolio choice," *Journal of Portfolio Management*, winter, 6 – 11.

Ding, Z., C. Granger, and R. Engle (1993): "long memory property of stock market returns and a new model," *Journal of Empirical Finance*, (1), 83 – 106.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50(4), 987 – 1007.

Finkelstein, M., and R. Whitley (1981): "Optimal strategies for repeated games," *Advanced Applied Probability*, 13, 415 – 428.

Glosten, L., R. Jagannathan, and D. Runkle (1993): "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779 – 1801.

Hakansson, N. H. (1970): "Optimal investment and consumption strategies under risk for a class of utility functions," *Econometrica*, 38, 587 – 607.

——— (1971): "On optimal myopic portfolio policies, with and without serial correlation of yields," *Journal of Business*, 44, 324 – 334.

Hakansson, N. H., and W. T. Ziemba (1995): "Capital Growth Theory," *Handbooks in Operations Research and Management Science*, 9.

Kelly, J. (1956): "A new interpretation of information rate," *Bell System Technology Journal*, 35, 917 – 926.

Latané, H. (1959): "Criteria for choice among risky ventures," *Journal of Political Economy*, 67, 144 – 155.

Lintner, J. (1965): "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, 47(1), 13 – 37.

Lv, Y., and B. K. Meister (2009): "Appliation of the Kelly criterion to Ornstein-Uhlenbeck processes," *Lecture notes of the institute for Computer Sciences*, 4, 1051 – 1062.

MacLean, L., E. Thorp, Y. Zhao, and W. Ziemba (2010): "Medium term Simulations of Kelly and fractional Kelly and proportional betting Strategies," *Working paper*.

MacLean, L. C., E. O. Thorp, and W. T. Ziemba. (2011): *The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific Press, Singapore.

MacLean, L. C., W. T. Ziemba, and G. Blazenko (1992): "Growth versus Security in dynamic Investment Analysis," *Management Science*, 38(11), 1562 – 1585.

Markowitz, H. M. (1952): "Portfolio Selection," *Journal of Finance*, 7, 77 – 91.

Merton, R. C. (1969): "Lifetime portfolio selection under uncertainty: The continuous time case," *Review of Economics and Statistics*, 51, 247 – 259.

——— (1992): *Continuous Time Finance*. MA Blackwell Publishers Inc., Malden.

NELSON, D. (1991): "Conditional heteroscedasticity in asset returns: A new approach," *Econometrica*, 59, 347 – 370.

OSORIO, R. (2008): "A Prospect-Theory Approach to the Kelly Criterion for Fat-Tail Portfolios: The Case of the Student T-Distribution," *Working papers series*.

POON, S. (2010): "Intertemporal Portfolio Selection. Lecture note in Continuous Time Finance," University of Manchester.

ROLL, R. (1973): "Evidence on the growth optimum model," *The journal of Finance*, 28(3), 551 – 566.

SHARPE, W. F. (1964): "Capital asset prices: A theory of market equilibrium under conditions of risk," *Journal of Finance*, 19(3), 425 – 442.

SIEGEL, J. (1998): *Stocks in the long run*. Mc Graw Hill, New York.

THORP, E. O. (1966): *Beat the dealer*. Knopf Doubleday Publishing Group, New York.

——— (1971): "Portfolio choice and the Kelly criterion," *Proceedings of the Business and Economics Section of the American Statistical Association*, pp. 215 – 224.

——— (1984): *The Mathematics of Gambling*. John Wiley and Sons, New York.

——— (2006): "The Kelly criterion in blackjack, sports betting and the stock market," *In S. A. Zenios and W. T. Ziemba (Eds.), Handbook of Asset and Liability Management*, 1, 387 – 428.

TOBIN, J. (1958): "Liquidity preference as behavior towards risk," *The Review of Economic Studies*, 25(67), 65 – 86.

TSAY, R. S. (2002): *Analysis of Financial Time Series*. Wiley, New York.

ZIEMBA, W., AND D. HAUSCH (1986): "Betting at the racetrack," *Dr. Z Investments Inc.*