Technische Universität Kaiserslautern

Option Valuation, Optimization and Excursions of Commodity Indices

Melanie Hollstein

Vom Fachbereich Mathematik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr.rer.nat)
genehmigte Dissertation

1. Gutachter: Prof. Dr. Ralf Korn
2. Gutachter: Prof. Dr. Rudi Zagst

Datum der Disputation: 25. August 2010
D 386

To my son

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Prof. Dr. Ralf Korn, for his support and guidance throughout my PhD studies and industrial projects. I am truly grateful for the opportunity to work at the Fraunhofer Institute for Industrial Mathematics ITWM and the Technical University Kaiserslautern. For all this and his constant encouragement, I am very thankful.

I would also like to extend my gratitude to my colleagues for the wonderful time especially, Prof. Dr. Marlene Müller, Dr. Peter Ruckdeschel, Dr. Gerald Kroisandt, Dr. Kalina Natcheva-Acar and Dr. Johan de Kock, who have all influenced me very much. I have benefitted tremendously from the inspiring working atmosphere, for whom I would like to thank Dr. Jörg Wenzel, Dr. Ulrich Nögel, Dr. Stefan Lorenz, Dr. Christina Erlwein, Tilman Sayer, Sascha Desmettre, Dr. Frank Seifried and Nicole Tschauder.

I feel very fortunate to have been blessed with a loving family. My parents daily love, encouragement and support is beyond words. Last but not least, I would like to thank Önder Güngör for his inspiration, love and enrichment in my life. Finally, I owe everything to my son, Leo for his love and patience for which I dedicate this thesis to him.

Special thanks goes to the Raiffeisen Centrobank AG for their kind permission of inserting their picture on the title page.

Contents

Preface

This thesis deals with the solution of special problems arising in financial engineering or financial mathematics. The main focus lies on commodity indices. Commodity indices consist of futures or spot values of energy, livestock, grains, industry metals, precious metals and softs. A futures contract is an agreement to buy or sell an asset at a certain future time for a certain price. Today, commodity indices represent the easiest way of getting exposure to commodities. Commodity index options are tied to major commodity indices. The issuers of such notes are investment banks, financial institutions or individual commodity producers raising capital to extend their operations while providing investors with exposure to the commodity sector.

We first dedicate ourselves with the modeling of commodity indices. **Chapter 1** addresses the important issue for the financial engineering practice of developing well-suited models for certain assets (here: commodity indices). Descriptive analysis of the Dow Jones-UBS commodity index compared to the Standard & Poor 500 stock index provides us with first insights of some features of the corresponding distributions. Statistical tests of normality and mean reversion then helps us in setting up a model for commodity indices. Additionally, chapter 1 encompasses a thorough introduction to commodity investment, history of commodities trading and the most important derivatives, namely futures and European options on futures. The importance of commodity indices in investment today is outlined, too.

Chapter 2 proposes a model for commodity indices and derives fair prices for the most important derivatives in the commodity markets. It is a Heston model [Hes93] supplemented with a stochastic convenience yield. The Heston model belongs to the model class of stochastic volatility models and is currently widely used in stock markets. For the application in the commodity markets the stochastic convenience yield is included in the drift of the instantaneous spot return process. Motivated by the results of chapter 1 it seems reasonable to model the convenience yield by a mean reverting Ornstein-Uhlenbeck process. Since trading desks only apply and consider models with closed form solutions for options I derive such formulas for commodity futures by solving the corresponding partial differential equation. Additionally, semi-closed form formulas for European options on futures are determined. The Cauchy problem with respect to these options is more challenging than the first one. A solution can be provided by applying the same methodology as in Heston [Hes93] and Bakshi and Madan [BM00].

Besides the optimization of the rolling procedure for commodity futures we dedicate ourselves in **chapter 3** with the optimization of the weightings of the commodity futures that make up the index. To this end, I apply the Markowitz approach or mean-variance optimization. The mean-variance optimization penalizes up-side and down-side risk equally, whereas most investors do not mind up-side risk. To overcome this, I consider in the next step other risk measures, namely Value-at-Risk and Conditional Value-at-Risk. The main part of chapter 3 is devoted to the presentation of an approach of Rockafellar and Uryasev [RU00] and [RU02] to optimizing a portfolio so as to reduce the risk of high losses. The Conditional Value-at-Risk is generalized to *discontinuous* cumulative distribution functions of the loss. For continuous loss distributions, the Conditional Value-at-Risk at a given confidence level is defined as the expected loss exceeding the Value-at-Risk. Loss distributions associated with finite sampling or scenario modeling are, however, discontinuous. Various risk measures involving discontinuous loss distributions shall be introduced and compared. They depend on a decision variable x and the confidence level α. My contribution to this topic is to bridge a gap in the proof of the crucial theorem in the article of Rockafellar and Uryasev [RU02]. Furthermore, I present an alternative way of proving some parts of the main theorem. I then apply the theoretical results to the field of portfolio optimization with commodity indices.

Furthermore, I uncover graphically the behavior of these risk measures. For this purpose, I consider the risk measures as a function of the confidence level α. Based on a special discrete loss distribution, the graphs demonstrate the different properties of these risk measures. One recognizes graphically that the definition of the Conditional Value-at-Risk as given in Rockafellar and Uryasev [RU02] is the most reasonable generalization to distributions with possible discontinuities.

The goal of the first section of **chapter 4** is to apply the mathematical concept of excursions for the creation of optimal highly automated or algorithmic trading strategies. Algorithmic trading is widely used by pension funds, mutual funds, institutional traders and hedge funds. The idea is to consider the gain of the strategy and the excursion time it takes to realize the gain. In this section I calculate formulas for the Ornstein-Uhlenbeck process. I show that the corresponding formulas can be calculated quite fast since the only function appearing in the formulas is the so called imaginary error function. This function is already implemented in many programs, such as in Maple. My main contribution of this topic is the optimization of the trading strategy for Ornstein-Uhlenbeck processes via the Banach fixed-point theorem.

The second section of chapter 4 reviews a concept of statistical arbitrage as introduced in Hogan et al. [HJTW04], a long horizon trading opportunity that generates a riskless profit. The results of this section provide an investor with a tool to investigate empirically if some strategies (for example momentum strategies) constitute statistical arbitrage opportunities or not. I correct some proofs in the article [HJTW04] and furthermore, I supplement the main theorem. Especially, I prove that the necessary conditions for statistical arbitrage given in [HJTW04] are slso sufficient. This is important for testing statistical arbitrage.

Chapter 1

Commodity Indices

1.1. Introduction to Commodity Investment and Modeling

History of Commodity Trading

Commodities constitute the only spot markets which have existed nearly throughout the history of humankind. Over the centuries the scope of commodities available has grown from essential agricultural commodities to include metals and energy.

The nature of trading has evolved from barter organized on town marketplaces in the absence of any monetary vehicle, to forward contracting between producers and merchants, then to organized futures markets with clearing houses guaranteeing the creditworthiness of transactions. The specification of contracts has evolved from plain-vanilla forwards to exotic options and structured products allowing producers and market participants to hedge away risks.

In the 18th and 19th centuries, potato growers in the state of Maine started selling their crops at the time of planting in order to finance the production process. A need for standardization in terms of quantity, quality, delivery date emerged and led to the establishment of the New York Cotton Exchange in 1842 and the Chicago Board of Trade in 1848 (see Geman [Gem05]).

Futures markets were originally set up to meet the needs of hedgers, namely farmers who wanted to lock in advance a fixed price for their harvest. Commodity futures are still widely used by producers and users of commodities for hedging purposes. Suppose that the date of analysis is January and an airline knows that it will have to buy on September 25th of the same year one million tons of fuel. In order to hedge against the possible increase in fuel price between

January and the end of September, the airline company will buy *futures* contracts written on fuel, maturity September and in an amount corresponding to the necessary quantity of fuel. So the airline company has locked in at the beginning of the year the price it will pay in September and has done so with no cash flow payment at the beginning of the year. Another possible hedge would be to buy options written on the fuel as the underlying.

Most of the liquidity in the futures markets is generated by the combined activity of speculators and hedgers. Using the same example as before, a bank is betting that the fuel price will go up or will go down and is counting on the corresponding profits the bank will generate. Commodities are becoming increasingly attractive to investors and hedge fund managers who view them as an alternative asset class allowing one to reduce the overall risk of a financial portfolio and enhance the return as well.

Futures and Forwards

After having given a first motivation for futures as well as options I shall now give a formal introduction to futures and forwards. Then I explain why it is necessary to include the so called convenience yield.

I follow the description given in Hull [Hul06], Eydeland and Wolyniec [EW03] and Geman [Gem05]. A *forward contract* is a particularly simple derivative. It is an agreement to buy or sell an asset at a certain future time for a certain price. A forward contract is traded in the Over-the-Counter market – usually between two financial institutions or between a financial institution and one of its clients. One of the parties of a forward contract assumes a *long position* and agrees to buy the underlying asset on a certain specified future date for a certain specified price. The other party assumes a *short position* and agrees to sell the asset on the same date for the same price.

Like a forward contract, a *futures contract* is an agreement between two parties to buy or sell an asset at a certain time in the future for a certain price. It is traded on an exchange. If two people get in touch with each other directly and agree to trade an asset in the future for a certain price, there are obvious risks. It is possible that one of them may not have the financial resources to honor the agreement. One of the key roles of the exchange is to organize trading so that contract defaults are minimized. To this end, one has introduced so called *margin accounts*.

To illustrate how margins work, consider a trader who contacts a broker on Monday, June 3, to buy two December gold futures contracts on the New York Commodity Exchange. I suppose that the current futures price is \$ 400 per ounce. Because the contract size is 100 ounces, the trader has contracted to buy a total of 200 ounces at this price. The broker will require the trader to deposit funds in what is termed the margin account. The amount that must be deposited at the time the contract is entered into is known as the *initial margin*. We will suppose this is \$ 2,000 per contract, or \$ 4,000 in total. At the end of each trading day, the margin account is adjusted to reflect the trader's gain or loss. This is known as *marking to market* the account. Suppose, for example, that by the end of June 3, the futures price has dropped from \$ 400 to \$ 397. The trader has a loss of $200 * \$ 3 = \$ 600$. The balance in the margin account would therefore be reduced by \$ 600 to \$ 3,400. The trader's broker has to pay the exchange \$ 600 and the exchange passes the money on to the broker of a trader with a short position.

When interest rates are deterministic, as we assume in the thesis, futures prices are equal to forward prices (see for example Duffie and Stanton [DS92] on page 569) and so all our results for futures prices also apply for forward prices.

Rolling of Futures

The futures contract may be financially settled by design or the position closed prior to maturity by a symmetric position in futures with the same maturity. In both cases, the investor does not need to worry about physical delivery and related concerns. If the investment horizon of an investor is farther away than the most deferred liquidity maturity one has to *roll* the futures positions. Unlike equities, which typically entitle the holder to a continuing stake in a corporation, commodity futures contracts normally specify a certain date for the delivery of the underlying physical commodity. In order to avoid the delivery process and maintain a long futures position, nearby contracts must be sold and contracts that have not yet reached the delivery period must be purchased. This process is known as rolling a futures position.

Commodity Indices

Today, commodity indices represent the easiest way of getting exposure to commodities. Commodity-linked notes and commodity index options are tied to major commodity indices, such as the Dow Jones-UBS Commodity Index (formerly called Dow Jones-AIG Commodity Index) or Goldman Sachs Commodity Index. The issuers of such notes are investment banks, financial institutions or individual commodity producers raising capital to extend their operations while providing investors with exposure to the commodity sector.

The indices are designed to provide investors with a benchmark for investment performance in the commodity markets comparable to the stock index S&P 500. The indices show realizable returns attainable in the commodity markets.

Relationship between Futures and Spot Prices, Convenience Yield

We now consider the *relationship between the futures price* $F(0,T)$ *and spot price* $S(0)$ of an investment asset. The following relation must hold:

$$F(0,T) = S(0)e^{rT}$$

where r is the risk-free interest rate and T is the maturity of the futures contract. If $F(0,T) > S(0)e^{rT}$ arbitrageurs can adopt the following strategy:

1. Borrow $S(0)$ dollars at an interest rate of r.
2. Buy the spot price of the commodity $S(0)$.
3. Enter into the futures contract to sell the commodity for $F(0,T)$ dollars in T years.
4. At time T one then obtains the gain $F(0,T) - S(0)e^{rT}$.

If $F(0,T) < S(0)e^{rT}$ there would be the following simple arbitrage opportunity:

1. Short the spot commodity.
2. Invest the $S(0)$ proceeds at the risk-free rate.
3. Enter into a long futures contract to buy the commodity for $F(0,T)$ dollars in T years.
4. At time T one then obtains the strictly positive difference $S(0)e^{rT} - F(0,T)$.

Notice that the above mentioned pricing formula $F(0,T) = S(0)e^{rT}$ only holds if the commodity is an investment asset (for example, gold and silver). That means it is mainly hold for investment purposes and the investors are prepared to sell their holdings and go long futures contracts, if they look more attractive. For commodities that are not, to any significant extent, held for investment, this argument cannot be used. Individuals and companies who keep such a commodity in inventory do so because of its consumption value – not because of its value as an investment. They are reluctant to sell the commodity and buy futures contracts because futures contracts cannot be consumed. There is, therefore, nothing to stop $F(0,T) < S(0)e^{rT}$ from holding. This means that all we can assert for a consumption commodity is $F(0,T) \leq S(0)e^{rT}$. So users of the commodity must feel that there are benefits from ownership of the physical commodity that are not obtained by the holder of a futures contract. These benefits may include the ability to profit from temporary local shortages or the ability to keep a production process running. The benefits are sometimes referred to as the *convenience yield* provided by the product. The (constant) convenience yield δ is defined so that

$$F(0,T)e^{\delta T} = S(0)e^{rT} \Longleftrightarrow F(0,T) = S(0)e^{(r-\delta)T}$$

The greater the possibility that shortages of the commodity will occur during the life of the futures contract, the higher the convenience yield.

1.2. Descriptive Analysis of a Commodity Index

Analysis of the available data is the first step in understanding and quantifying the essential features of a particular market. Without thorough data analysis it is impossible to select the most appropriate model. My sample consists of daily closing values of the Dow Jones-UBS Commodity Index and adjusted closing values of the Standard & Poor 500 Index from 01/02/1991 to 01/31/2006. As references for comparison of the commodity index, I plot the distributions and calculate the test statistics of the Standard & Poor's 500 Index.

Commodities exhibit *seasonality.* But I will not eliminate the seasonality in the analysis because I am facing an index of several commodities. So the index shall not have such a pronounced seasonality compared with a single commodity such as for example natural gas.

I focus on financial pricing models. A great majority of them are based on an assumption of the normality or log-normality of underlying price/return distributions. Looking at the sample histograms of the commodity index makes us apprehensive of the normality assumption for these returns (see the figures on page 13) . So I shall conduct statistical tests to provide a formal foundation for acceptance and rejection of a particular choice of distributions built into our pricing models. More precisely I am interested in investigating the following items

- Can the prices be modeled as a geometric Brownian motion?
- Do the prices or the logarithm of the prices feature mean reversion, in particular is it recommendable to model the prices as a Vasicek process?

First I provide a description of the Dow Jones-UBS Commodity Index. The value of the Dow Jones-UBS Commodity Index (formerly Dow Jones-AIG Commodity Index until 7th May 2009) is computed on the basis of hypothetical investments in a basket of commodities that make up the index. It is composed of futures contracts on 19 physical commodities.

A long futures position is maintained by selling nearby contracts and purchasing contracts that have not yet reached the delivery period. In that sense, the Dow Jones-UBS Commodity Index is a rolling index. The roll period is the period of five business days, beginning with the sixth business day through and including the tenth business day of each month. In this time the value of the Dow Jones-UBS Commodity Index is gradually shifted from reliance on the basket of lead futures to the basket of next futures.

The index is composed of commodities traded on US exchanges, with the exception of aluminum, nickel and zinc which trade on the London Metal Exchange.

In stock indices the weightings of the constituent assets depend on market capitalization. For commodities, there is no direct counterpart to market capitalization. The problem is that commodities are held in a variety of ways – long futures positions, over-the-counter investments, long-term fixed-price purchasing contracts, physical inventory at the producer, etc. This makes a complete accounting of capital dedicated to holding commodities from the time they are produced to the time they are consumed infeasible. Different commodity indices have different ways to achieve a relatively close analog to market capitalization. The Dow Jones-UBS Commodity Index relies on five year averaging of both liquidity data and dollar-adjusted production data in order to determine the relative quantities of included commodities. Liquidity is an important indicator of the value placed on a commodity by financial market participants. Production data, on the other hand, are a useful measure of economic importance but may underestimate the economic significance of storable commodities (for example, gold) at the expense of relatively non-storable commodities (for example, live cattle). This is why the Dow Jones-UBS Commodity Index relies on production, a quantity exogenous to the futures markets, and liquidity, a quantity endogenous to these markets, to define the relative weightings (see table 1.1). Of course the actual percentages vary based on market price. In contrast to the Dow Jones-UBS Commodity Index, the Goldman Sachs Commodity Index is only world-production weighted; the quantity of each commodity in the index is determined by the average quantity of production in the last five years of available data.

To ensure that no single commodity or commodity sector dominates the index, the Dow Jones-UBS Commodity Index relies on several diversification rules. Among these rules are:

- no related group of commodities (for example, energy, or metals, or livestock and grains)

Table 1.1.: Dow Jones-UBS Commodity Index Target Weights 2005 (Rounded)

Commodity Group	***Commodity***	***Percentage***
Energy	Natural Gas	12%
	Crude Oil	13%
	Unleaded Gas	4%
	Heating Oil	4%
Livestock	Live Cattle	6%
	Lean Hogs	4%
Grains	Wheat	5%
	Corn	6%
	Soybeans	8%
	Soybean Oil	3%
Industry Metals	Aluminum	7%
	Copper	6%
	Zinc	3%
	Nickel	3%
Precious Metals	Gold	6%
	Silver	2%
Softs	Sugar	3%
	Cotton	3%
	Coffee	3%

may constitute more than 33% of the index;

- no single commodity may constitute less than 2% of the index.

Diversification rules will be applied each year, when the Dow Jones-UBS Commodity Index is reweighted and rebalanced on a price-percentage basis.

Quantitative Measures of Indices

First I estimate *mean*, *median*, *standard deviation*, *skewness* and *excess kurtosis*.

Definition 1.2.1. *Let X be a random variable with existing first four moments. With μ denoting the expectation, σ^2 the variance one can define the skewness $\mathbb{S}$ and excess kurtosis $\mathbb{K}$ as*

$$\mathbb{S}(X) = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3} \quad \text{and} \quad \mathbb{K}(X) = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4} - 3$$

Remark 1.2.1. *A high excess kurtosis distribution ($\mathbb{K} > 0$) has a sharper peak and longer, fatter tails, while a low excess kurtosis distribution ($\mathbb{K} < 0$) has a more rounded peak and shorter thinner tails. A distribution is called leptokurtic if $\mathbb{K} > 0$.*

Skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable. If a distribution has a negative skew ($\mathbb{S} < 0$) *the left tail is longer; the mass of the distribution is concentrated on the right of the distribution figure. In the case of a positive skew* ($\mathbb{S} > 0$) *is the right tail longer; the mass of the distribution is concentrated on the left of the figure.*

With n denoting the number of observed data and $x = (x_1, ..., x_n)$ the observed data the standard estimates for the mean and variance are here defined by

$$\widehat{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \widehat{x})^2 \tag{1.1}$$

and the estimates for skewness $\widehat{\mathbb{S}}$ and excess kurtosis $\widehat{\mathbb{K}}$ are calculated in the following way

$$\widehat{\mathbb{S}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x})^2\right)^{3/2}} \quad \text{and} \quad \widehat{\mathbb{K}} = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{x})^2\right)^{2}} - 3 \tag{1.2}$$

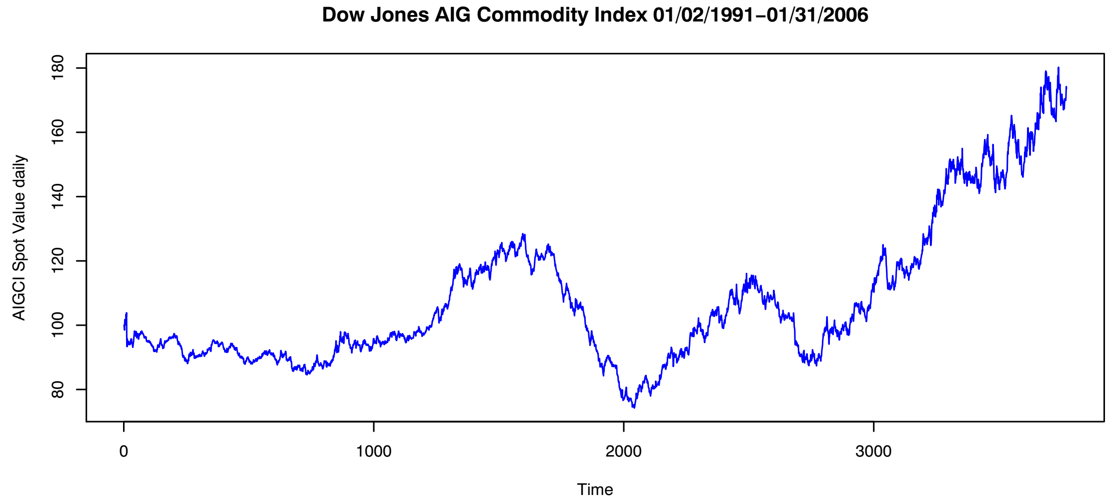
Note that the statistic software `R` also subtracts the kurtosis of the normal distribution (namely 3), so `R` calculates the excess kurtosis.

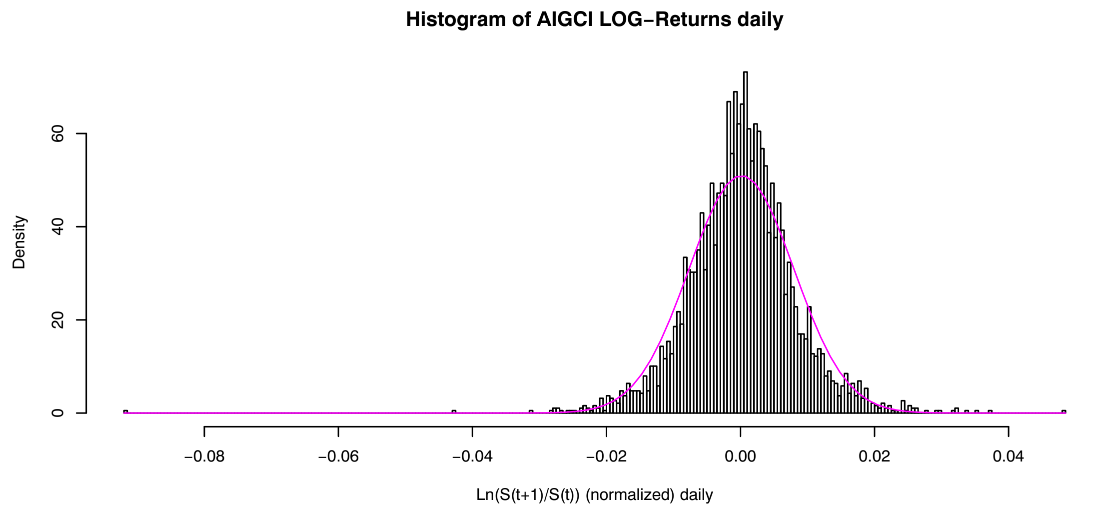
Table 1.2.: Quantitative Measures of Dow Jones-AIG Commodity Index

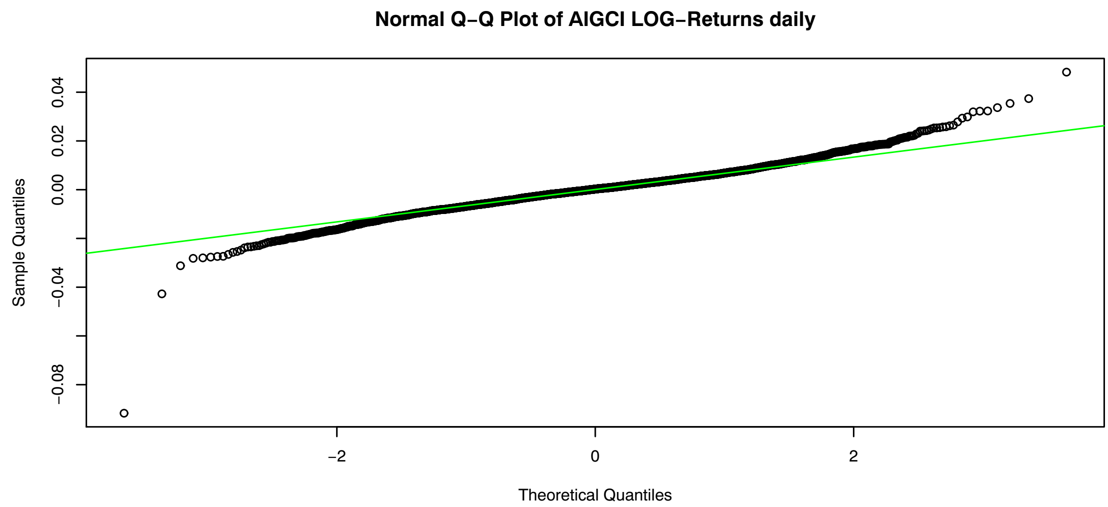
	$S_t - S_{t-1}$	$\ln S_t - \ln S_{t-1}$
Mean	0.020	0.000
Median	0.025	0.000
Standard Deviation	0.925	0.008
Skewness	0.137	−0.302
Excess Kurtosis	7.004	6.596
Amount of Data	3770	3770

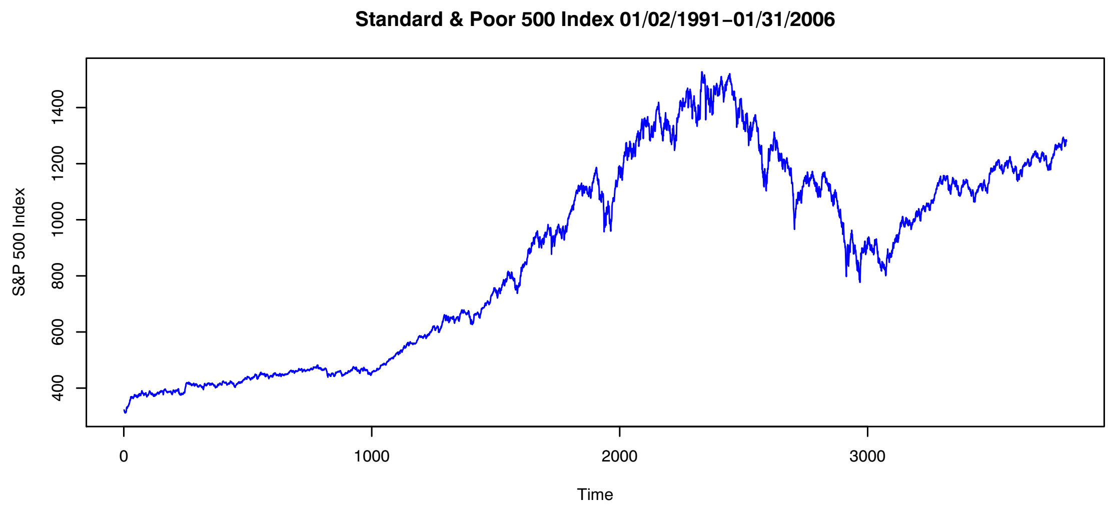
Table 1.3.: Quantitative Measures of Standard & Poor 500 Index

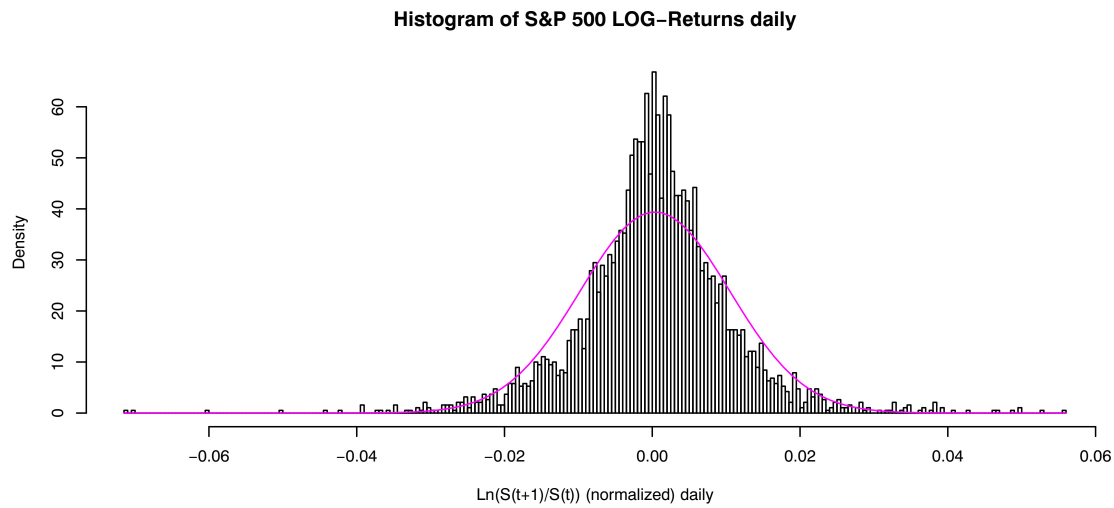
	$S_t - S_{t-1}$	$\ln S_t - \ln S_{t-1}$
Mean	0.252	0.000
Median	0.270	0.000
Standard Deviation	10.515	0.010
Skewness	−0.093	−0.101
Excess Kurtosis	5.749	3.998
Amount of Data	3801	3801

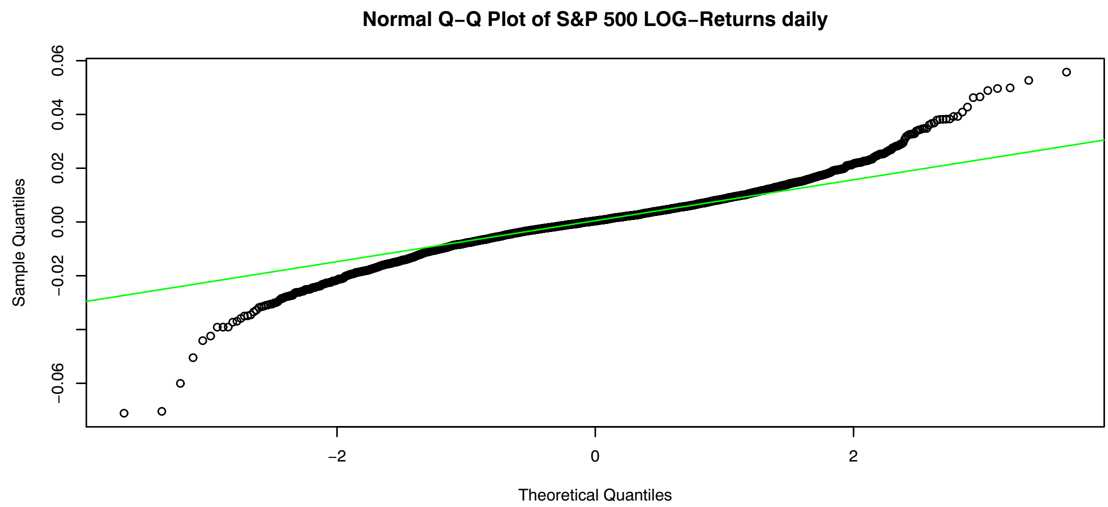

Dow Jones AIG Commodity Index 01/02/1991−01/31/2006
AIGCI Spot Value daily
80
100
120
140
160
180
0
1000
2000
3000
Time


Histogram of AIGCI LOG−Returns daily
Density
0
20
40
60
−0.08
−0.06
−0.04
−0.02
0.00
0.02
0.04
Ln(S(t+1)/S(t)) (normalized) daily


Normal Q−Q Plot of AIGCI LOG−Returns daily
Sample Quantiles
−0.08
−0.04
0.00
0.02
0.04
−2
0
2
Theoretical Quantiles


Standard & Poor 500 Index 01/02/1991−01/31/2006
S&P 500 Index
400
600
800
1000
1200
1400
0
1000
2000
3000
Time


Histogram of S&P 500 LOG−Returns daily
Density
0
10
20
30
40
50
60
−0.06
−0.04
−0.02
0.00
0.02
0.04
0.06
Ln(S(t+1)/S(t)) (normalized) daily


Normal Q−Q Plot of S&P 500 LOG−Returns daily
Sample Quantiles
−0.06
−0.02
0.00
0.02
0.04
0.06
−2
0
2
Theoretical Quantiles

Tests for Normality

To check for normality I use both the Jarque-Bera and Shapiro-Wilk test applied to the daily differences of the prices $(S_t - S_{t-1})$ and to the logarithmic returns $(\ln S_t - \ln S_{t-1})$.

The Jarque-Bera test compares skewness and kurtosis of the observed data with that of a normal distribution. Following Franke, Härdle and Hafner [FHH04] page 150 under the null hypothesis of normality the estimators of 1.2 are independent and asymptotically normal distributed with

$$\sqrt{n}\widehat{\mathbb{S}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,6) \quad \text{and} \quad \sqrt{n}\widehat{\mathbb{K}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,24)$$

The Jarque-Bera test uses the test statistic

$$JB = n\left[\frac{\widehat{\mathbb{S}}^2}{6} + \frac{\widehat{\mathbb{K}}^2}{24}\right]$$

For large sample sizes the Jarque-Bera statistic follows the chi-square distribution with 2 degrees of freedom. For normal distributions the skewness and excess kurtosis are both equal to zero, hence $JB = 0$. If the Jarque-Bera statistic is sufficiently large (greater than 6.0 at 5% significance level), the null hypothesis of normality is rejected. There exist many distributions which have the same skewness and kurtosis as the normal distribution, so in addition I apply the Shapiro-Wilk test which tests the complete sample for normality.

The Shapiro-Wilk test (see Shapiro and Wilk [SW65]) checks for linearity of the QQ-plot. A QQ-plot ("Q" stands for quantile) is a probability plot, a kind of graphical method for comparing two probability distributions, by plotting their quantiles against each other. In a probability plot, one can consider the regression of the ordered observations on the expected values of the order statistics from a standardized version of the hypothesized distribution – the plot tending to be linear if the hypotheses is true. Hence a possible method of testing the distributional assumption is by means of the comparison of the squared slope of the probability regression line with the sample sum of squares about the mean.

It tests the null hypothesis that a sample $x_1, \ldots, x_n$ came from a normally distributed population with unknown mean and variance. The SW test statistic for normality is defined by

$$SW = \frac{\left(\sum_{i=1}^{n} a_i X_i\right)^2}{\sum_{i=1}^{n} (x_i - \widehat{x})^2}$$

where

- X_i is the ith order statistic, that is, the ith-smallest number in the sample;

- $\widehat{x} = (x_1 + \cdots + x_n)/n$ is the sample mean;
- the constants a_i are given by

$$(a_1, \ldots, a_n) = \frac{m^\top V^{-1}}{(m^\top V^{-1} V^{-1} m)^{1/2}}$$

where $m = (m_1, \ldots, m_n)^\top$ and $m_1, \ldots, m_n$ are the expected values of the order statistics of independent and identically-distributed random variables sampled from the standard normal distribution, and V is the covariance matrix of those order statistics. The numerator of the test statistic $\sum_{i=1}^n a_i X_i$ is, up to a normalizing constant $C := (m^\top V^{-1} V^{-1} m)^{1/2}$, the best linear unbiased estimate of the slope of a linear regression of the ordered observations X_i on the expected values m_i of the standard normal order statistics. The constant C is so defined that the linear coefficients are normalized. It is clear that the null hypothesis is rejected if SW is too small.

Tests for Mean-Reversion

The mean-reversion tests are applied both to the price data S and to the logarithm of the price data. That means I test for the Vasicek and the exp(Vasicek) model. The Vasicek-process is given by

$$dS(t) = (\alpha - \kappa S(t))dt + \sigma dW(t) \tag{1.3}$$

with $S(0), \kappa, \alpha$ and σ being strictly positive constants and W denoting as usual a standard Brownian motion.

Proposition 1.2.1. *The stochastic differential equation 1.3 can be solved explicitly. $S(t)$ is normally distributed for all $t > 0$. For $\kappa > 0$ the process S is mean-reverting, for $\kappa < 0$ mean-exploding and for $\kappa = 0$ a Brownian motion with drift.*

Proof: The stochastic differential equation 1.3 can be solved explicitly. Indeed, if we consider $Y(t) = e^{\kappa t} S(t)$ and integrate by parts, it yields

$$\begin{aligned} dY(t) &= e^{\kappa t} dS(t) + S(t)\kappa e^{\kappa t} dt \\ &= \alpha e^{\kappa t} dt + \sigma e^{\kappa t} dW(t). \end{aligned}$$

Thus

$$S(t) = S(0)e^{-\kappa t} + \frac{\alpha}{\kappa}\left(1 - e^{-\kappa t}\right) + \int_0^t \sigma e^{-\kappa(t-s)} dW(s) \tag{1.4}$$

is normally distributed with

$$\mathbb{E}[S(t)] = S(0)e^{-\kappa t} + \frac{\alpha}{\kappa}\left(1 - e^{-\kappa t}\right) \tag{1.5}$$

$$\mathbb{V}[S(t)] = \sigma^2 \int_0^t e^{-2\kappa(t-s)} ds = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa t}\right) \tag{1.6}$$

For $\kappa > 0$ the process S is mean-reverting, since the expected value tends, for t going to infinity, to the value α/κ.

If $\kappa < 0$ (and $\alpha > 0$) the mean explodes for t going to infinity

$$\begin{aligned} \mathbb{E}[S(t)] &= S(0)e^{-\kappa t} + \frac{\alpha}{\kappa}\left(1 - e^{-\kappa t}\right) \\ &= \frac{\alpha}{\kappa} + \left(S(0) - \frac{\alpha}{\kappa}\right) e^{-\kappa t} \overset{t\to\infty}{\to} +\infty \end{aligned}$$

So for the model to be truly mean-reverting and not mean-exploding, we must have $\kappa > 0$.

If $\kappa = 0$ the process is a Brownian motion with drift and is a random walk with no mean-reversion. ■

The fact that α/κ can be regarded as a long term average value could also be inferred from the dynamics 1.3 itself. The drift of the process S is positive whenever the process is below α/κ and negative otherwise, so that S is pushed, at every time, to be closer on average to the level α/κ.

Discretizing stochastic differential equation 1.3 via the Euler scheme with equidistant time steps yields

$$S(t) = S(t-1) + (\alpha - \kappa S(t-1)) + \sigma\varepsilon_t$$

$$\Leftrightarrow S(t) = \alpha + (1-\kappa)S(t-1) + \sigma\varepsilon_t$$

where $(\varepsilon)_t$ are independent identically distributed normal random variables with mean 0 and variance equal to $\sqrt{t_i - t_{i-1}}$.

Define $\beta = 1 - \kappa$. I perform a statistical test to confirm whether the coefficient $\beta < 1$, that is, κ is positive, as required by the assumption of mean reversion.

For testing the mean reversion I apply the following tests:

- Dickey-Fuller test (DF) which assumes the discrete time version of the Vasicek model

$$S_t = \alpha + \beta S_{t-1} + \sigma\epsilon_t \tag{1.7}$$

and checks for

$$\text{DF}(<1): \; H_0 : \beta = 1 \text{ versus } H_1 : \beta < 1$$

$$\text{DF}(>1): \; H_0 : \beta = 1 \text{ versus } H_1 : \beta > 1$$

- Augmented Dickey-Fuller test (ADF) assumes the more general model

$$S_t = \alpha + \beta S_{t-1} + \sum_{j=2}^{k} \beta_j S_{t-j} + \sigma \epsilon_t \tag{1.8}$$

and checks for

$$\text{ADF}(<1): \; H_0 : \beta = 1 \text{ versus } H_1 : \beta < 1$$

$$\text{ADF}(>1): \; H_0 : \beta = 1 \text{ versus } H_1 : \beta > 1$$

- Kwiatkowski-Philips-Schmidt-Shin test (KPSS) which assumes model 1.7 and checks for

$$H_0 : \beta < 1 \text{ versus } H_1 : \beta \geqslant 1$$

The basis for the Dickey-Fuller test (see for example Franke, Härdle and Hafner [FHH04] section 10.6.1) is the regression

$$\Delta S(t) = \alpha + (\beta - 1) S(t-1) + \sigma \epsilon_t$$

which comes from model 1.7. The standard t-statistic then takes the form

$$T_n(x) = \frac{\widehat{\beta} - 1}{\sqrt{\widehat{\sigma}^2 (t_i - t_{i-1}) \sum_{i=2}^{n} x_{i-1}^2}} \tag{1.9}$$

where $\widehat{\beta}$ and $\widehat{\sigma}^2(t_i - t_{i-1})$ are the least squares estimates for β and the variance $\sigma^2(t_i - t_{i-1})$ of $\sigma\varepsilon_t$, respectively. For $n \to \infty$ the statistic does not converge to the standard normal distribution but to a distribution of a function of Brownian processes

$$T_n(x) \xrightarrow{\mathcal{L}} \frac{1 - W^2(1)}{2 \left(\int_0^1 W^2(u) du \right)^{1/2}}$$

where W is a standard Brownian motion.

In addition I also apply the Augmented Dickey-Fuller test which has the same test statistic 1.9, with critical values that can again be calculated with the help of `R`. The test is based on an augmented regression, namely 1.8, which encompasses in addition delayed differences. I set the number k in 1.8 equal to 2 (in `R` one has to type $k = 1$ due to different notations).

The KPSS test interchanges the hypotheses in comparison to the Dickey-Fuller test and the Augmented Dickey-Fuller test. It is described for example in Franke, Härdle and Hafner [FHH04] section 10.6.2.

I have implemented all the tests in `R` with the inclusion of package `tseries`. Tables 1.4 and 1.5 show the p-values for daily data of the Dow Jones-UBS Commodity Index and Standard & Poor 500 Index. Not formally speaking, the p-value is the probability of obtaining a result at least as extreme as the one that was actually observed, given that the null hypothesis is true.

Table 1.4.: P-Values for Daily Dow Jones-AIG Commodity Data

	$S_t - S_{t-1}$	$\ln S_t - \ln S_{t-1}$
Jarque-Bera	0.00000	0.00000
Shapiro-Wilk	0.00000	0.00000
AR(1) coefficient	0.99797	0.99824
ADF(<1)	0.96667	0.92113
ADF(>1)	0.03333	0.07887
DF<1	0.94211	0.88461
DF>1	0.05789	0.11539
KPSS	0.01000	0.01000
Amount of Data	3771	3771

Interpretation of Analyzed Indices

After having introduced and explained the tests we can now dedicate ourselves to interpret the analyzed commodity index: Both the Jarque-Bera and the Shapiro-Wilk test reject the hypotheses that the differences $S_t - S_{t-1}$ or the logarithmic returns $\ln S_t - \ln S_{t-1}$ are normally distributed (p-values equal to zero for daily data of both indices). The plotted histograms reveal the reason: the data is leptokurtic and fat-tailed. The high peaks at the origin come from the fact that the changes in the data are too small. But even when switching from daily to monthly data the disturbing leptokurticy does not vanish (p-value equal to zero for weekly data and small p-values for monthly data, not printed here). The fat tails can also be deduced from the

Table 1.5.: P-Values for Daily Standard & Poor 500 Data

	$S_t - S_{t-1}$	$\ln S_t - \ln S_{t-1}$
Jarque-Bera	0.00000	0.00000
Shapiro-Wilk	0.00000	0.00000
AR(1) coefficient	0.99906	0.99908
ADF(<1)	0.83870	0.90545
ADF(>1)	0.16130	0.09455
DF<1	0.76398	0.83844
DF>1	0.23602	0.16156
KPSS	0.01000	0.01000
Amount of Data	3802	3802

magnitude of the distribution excess kurtosis. In the case of the commodity index it is equal to 6.596 which is significantly higher than 0, the value of the excess kurtosis for the normal distribution. The excess kurtosis of the stock index is lower than the commodity index, namely 3.998. On the other hand, the skewness of the distributions are not far from the skewness of the normal distribution (which is equal to zero). Both indices are negatively skewed.

The estimated AR(1) coefficient (β in model 1.7) is in all cases near one; it descends slightly when changing from daily to weekly data and from weekly to monthly data (not printed here). All the applied mean reversion tests come to the same conclusion: One should not model the commodity price process nor the logarithm of the price process as a Vasicek model. The p-values for mean reversion tests of logarithmic returns are for both indices comparable: For example for the ADF(< 1) test the p-value of the commodity index is equal to 0.92 and 0.91 for the stock index. We can conclude that the mean-reversion is very slow, and we do not have enough data to form tests with enough power to detect it. In practical terms then, as long as we are not concerned about writing options with a 100 year maturity, we do not have to worry about the presence of mean reversion.

In the case of the differences $S_t - S_{t-1}$ the stock index seems to have a slightly more tendency to exhibit mean-reversion, for example the p-value of the DF(< 1)-test is 76% for the stock index in contrast to 94% for the commodity index. Since it is not usual to set up mean-reverting models in stock indices markets for the differences $S_t - S_{t-1}$ it is recommendable to discard mean reverting models for commodity indices, too.

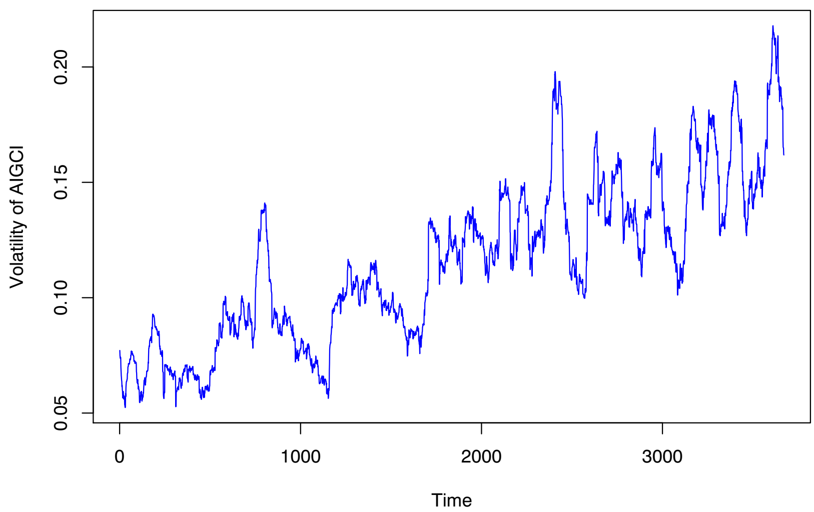
The first impression of the results in the table regarding the Jarque-Bera and Shapiro-Wilk test is that the distribution of the logarithmic returns is not normal. This inference may not be correct, however, and may stem from the misuse of a statistical methodology. I have assumed that each sample element x_i is a realization of the same random variable x. In other words, I have assumed some form of stationary of the evolution. But volatility plots on page 1.2 reveal that volatility is not constant over time for both indices.

Note that to create these graphs I used the volatility estimate with the 60-day moving window, which I shall explain: I assume that the increments $\ln S_i - \ln S_{i+1}$ are independent and normally distributed, and that the standard deviation of this distribution is proportional to the square root of time between observations (as it is the case in the Black-Scholes model); that is it can be represented as $\sigma\sqrt{t_i - t_{i-1}}$. The Black-Scholes volatility σ can be estimated using formula 1.1. Note that it is customary to represent time increments as year fractions. So with the moving window method, the volatility estimate at a given time t_k is given by the expression $\sigma(t_k) = \sqrt{\frac{1}{m-1}\sum_{i=k-m+1}^{k}\left(\frac{\ln S_i - \ln S_{i-1}}{\sqrt{t_i - t_{i-1}}} - \frac{1}{m}\sum_{i=k-m+1}^{k}\frac{\ln S_i - \ln S_{i-1}}{\sqrt{t_i - t_{i-1}}}\right)^2}$ with m being the specified number of observations preceding t_k used to estimate volatility.

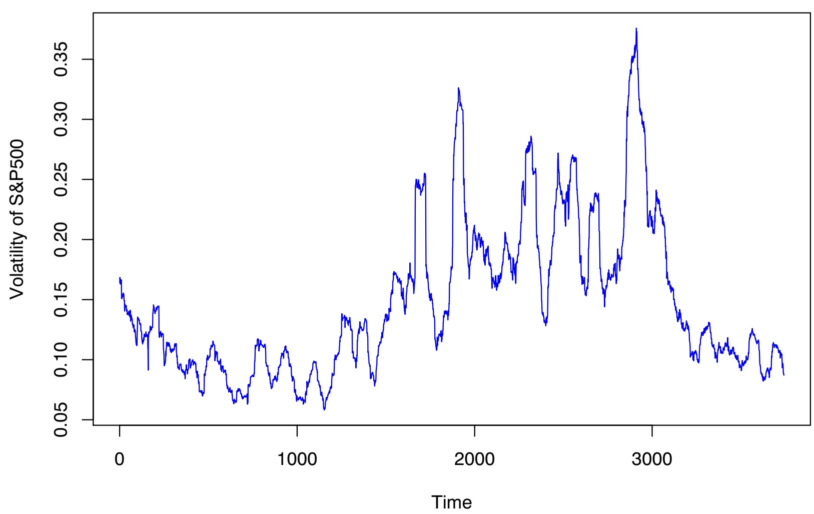
The volatility plots reveal that the commodity index as well as the stock index exhibit nonconstant volatility. The volatility of volatility of both indices are nearly the same. The average volatility of the commodity index is approximately 13% where the average volatility of the stock index seems higher, namely approximately 17%.

In the following chapter I shall introduce a model for commodity indices with stochastic volatility.

Volatility of Dow Jones AIG Commodity Index 01/02/1991−01/31/2006



Volatility of Standard & Poor 500 Index 01/02/1991−01/31/2006



Chapter 2

Heston Model for Commodity Indices

2.1. The Model

Descriptive analysis of commodity indices (as for example the Dow Jones-UBS Commodity Index which we considered in the last chapter) reveal that it is necessary to apply more sophisticated models than geometric Brownian motion for commodities. Several important characteristics of commodity return distributions was uncovered in the last chapter: First, commodity returns are leptokurtic. Second, the volatility of commodity returns changes randomly over time. Third, we do not have to worry about mean reversion.

I propose a stochastic volatility model, namely the Heston model [Hes93] supplemented with a stochastic convenience yield. For this model I calculate no-arbitrage prices for futures and European options on futures. Then I have implemented the solutions in Matlab and compared them with Monte Carlo prices.

The most basic and most frequently traded asset in commodity markets are futures. Using the technique developed by Heston [Hes93], Bakshi and Madan [BM00] and Yan [Yan02] I derive a closed form solution for futures in order to calibrate the parameters appearing in the Heston model. After the calibration of the parameters one is able to price non-traded options with the help of different kinds of simulations.

But an interesting result is that the futures prices do not depend on the volatility parameters. So when calculating the expectation of the index value $S(T)$ at time t the volatility parameters do not appear. So calibration is done by means of futures prices *and* European options on futures. To this end, I derive a semi-closed form solution for European options. Often jumps are also included in stochastic volatility models. But more complex models do not necessarily offer a better fit. While one can get a better fit with some complex functional specifications, the performance of those models out-of-sample is poor, and the parameters are unstable.

The Heston model [Hes93] is currently a frequently used model in the stock market. To apply it also in the commodity market I include a stochastic convenience yield. The model is specified by the following stochastic differential equations

$$dS(t)/S(t) = (r - \delta(t))\,dt + \sigma_S dW_1 + \sqrt{V(t)}dW_2,\ S(0) > 0 \tag{2.1}$$

$$d\delta(t) = (\theta_\delta - \kappa_\delta \delta(t))\,dt + \sigma_\delta dW_\delta,\ \delta(0) \in \mathbb{R} \tag{2.2}$$

$$dV(t) = (\theta_V - \kappa_V V(t))\,dt + \sigma_V \sqrt{V(t)}dW_V,\ V(0) > 0 \tag{2.3}$$

The processes S, δ and V are defined on a filtered probability space $((\Omega, \mathcal{F}, Q), \mathcal{F}(t))$. $(W_a)_{t\geq 0}$ with $a \in \{1, 2, \delta, V\}$ are all Q-Brownian motions with instantaneous correlations $\rho_1 \in [-1, 1]$ and $\rho_2 \in [-1, 1]$, that is, $d\left\langle W_1, W_\delta \right\rangle_t = \rho_1 dt$ and $d\left\langle W_2, W_V \right\rangle_t = \rho_2 dt$. Correlations between other Brownian motions are assumed to be zero. Note that I introduce two uncorrelated Brownian motions W_1 and W_2 in stochastic differential equation 2.1 modeling the instantaneous return. W_1 is correlated with the Brownian motion of the convenience yield W_δ and W_2 is correlated with the random walk of the volatility W_V. The reason for introducing two Brownian motions is to separate the influence of volatility and convenience yield in the stock returns. This makes the derivation and interpretation of the closed formula for futures and the semi-closed form solutions for European options on futures more tractable.

The involved parameters of the stock returns are:

- initial stock value $S(0) > 0$
- the instantaneous riskless short rate $r > 0$
- the stochastic volatility V given by a Cox-Ingersoll-Ross process 2.3 and
- the stochastic convenience yield $\delta(t)$ is specified by an Ornstein-Uhlenbeck process 2.2.

Remark 2.1.1. *The model is set up under an equivalent martingale measure Q. I note in passing that I am facing an incomplete market rather than a complete one. Due to the second fundamental theorem of asset prices there exist more than one equivalent martingale measure. As common in financial engineering the equivalent martingale measure is chosen from the mar-*

ket, that is, I calibrate the parameters with options. Since there exists an equivalent martingale measure the market is arbitrage-free (second fundamental theorem of asset prices). But I have to outline a peculiarity of commodities in comparison to equity or fixed-income securities: One cannot dynamically hedge. The action of borrowing and lending is hardly predictable in commodities. They are heavily grounded in their physical nature. One has been taught that securities are derived by arbitrage arguments that allow us to seamlessly borrow and lend, in order to move the asset and liability across time. In the arbitrage relationship, the future equals the spot times e^{rt}*. Now consider that you are trading in products that are not transferable into the future. Arbitrage becomes hardly possible – and, with it, the arbitrage argument. Storage can cause shrinkage in quantity (for example, electricity). Future oil may still be in the ground and might cost no carry to the producer, whereas the arbitrageur would have to bear onerous storage costs.*

So an important distinction with commodities as an asset class is storage, including transportation, insurance, warehousing and incidentals. Whereas equity and fixed-income securities can be easily stored as bits of paper or electronic entries after purchase, storage of physical commodities involves complications; for example, consider storing crude oil and wheat as part of one's investment.

So one has to be careful with any arbitrage involving physical delivery. Clearly, it is not possible to dynamically hedge a security that you cannot short, sometimes cannot easily own and that can be severely illiquid.

The return process is a Black-Scholes model with two modifications. First, the volatility is not constant anymore, because we have seen in chapter 1 that this is indeed not the case. Second, the convenience yield motivated in chapter 1 is included.

The so called Cox-Ingersoll-Ross diffusion for the *volatility* solve

$$dV(t) = (\theta_V - \kappa_V V(t))\,dt + \sigma_V \sqrt{|V(t)|}dW_V, \tag{2.4}$$

where the involved parameters are

- initial volatility $V(0) > 0$
- the mean reversion rate $\kappa_V > 0$
- the long-run mean $\theta_V \geq 0$ and
- the volatility of volatility $\sigma_V > 0$

and W_V being a standard Brownian motion. For every given value $V(0) \geq 0$, equation 2.4 admits a unique solution; this solution is strong, that is, adapted with respect to the natural filtration of W_V, and takes values in $[0, \infty[$ (see Göing-Jaeschke and Yor [GJY03] page 1). If $\theta_V = 0$ and $V(0) = 0$, the solution of 2.4 is $V(t) \equiv 0$, and from the comparison theorem for one-dimensional diffusion processes (see Revuz and Yor [RY91] Theorem 9.3.7), I deduce $V(t) \geq 0$ for $\theta_V \geq 0$ and $V(0) \geq 0$. Hence, in this case the absolute value in 2.4 can be omitted. If $\theta_V \geq 0.5\sigma_V^2$ it cannot reach zero almost surely (see Lamberton and Lapeyre [LL96] Proposition 6.2.4 on page 130). In the following I assume that it holds $\theta_V \geq 0.5\sigma_V^2$.

The volatility is chosen in this way by Heston [Hes93] because it is possible to derive semi-closed form solutions for a lot of derivatives. Closed form solutions are very important in trading derivatives because Monte Carlo methods take too much time.

An important feature of the volatility process for $\kappa_V > 0$ is the mean reversion, that is, it has a long-run mean. If $V(t)$ is greater than the long-run mean then the drift of the process is negative. When $V(t)$ is smaller than the long-run mean then the process is pushed up by the drift.

Proposition 2.1.1. *If $V(t)$ is a solution of 2.3 then with $s \geq t$ the conditional expectation and variance are given by*

$$\begin{aligned} \mathbb{E}[V(s)|V(t)] &= \frac{\theta_V}{\kappa_V} + e^{-\kappa_V(s-t)}\left(V(t) - \frac{\theta_V}{\kappa_V}\right) \\ \mathbb{V}[V(s)|V(t)] &= V(t)\left(\frac{\sigma_V^2}{\kappa_V}\right)\left(e^{-\kappa_V(s-t)} - e^{-2\kappa_V(s-t)}\right) + \frac{\theta_V}{\kappa_V}\left(\frac{\sigma_V^2}{2\kappa_V}\right)\left(1 - e^{-\kappa_V(s-t)}\right)^2 \end{aligned}$$

Proof: can be found in Ioffe [Iof10]. ■

The *convenience yield* is specified by an Ornstein-Uhlenbeck process because this process incorporates two crucial features, namely mean reversion (economic reason) and the fact that the process can become negative with positive probability. The last feature is implied by the expectation and variance of the Ornstein-Uhlenbeck process calculated in chapter 1, namely 1.5 and 1.6. In interest rate models the possibility of negative values is a drawback, but when modeling the convenience yield this in in fact desired. It enables us modeling strong contango: Remember the relationship between commodity spot values and futures prices in the case of a constant convenience yield δ

$$F(t,T) = S(t)e^{(r-\delta)(T-t)} \tag{2.5}$$

Knowledge of $S(t)$ and δ then leads to the whole futures curve (fix t, vary maturity T). One can observe that in the case where $r - \delta$ is negative, this futures curve is a decreasing function of maturity T. A decreasing futures curve is called a *backwardated* futures curve. This happens when $r < \delta$, that is, when interest rates are low and the benefit of holding the physical commodity high. Conversely, when the difference $r-\delta$ is positive, the forward curve is an increasing function of maturity and one obtains the situation of *contango*.

From section 1.2 we already know that the Ornstein-Uhlenbeck process is mean reverting if and only if parameter $\kappa_\delta > 0$. Furthermore, we have already derived the explicit solution 1.4 of stochastic differential equation 2.2. As in section 1.2 I assume that the parameters of the Ornstein-Uhlenbeck process, namely θ_δ, κ_δ and σ_δ, are all strictly positive.

In chapter 1 we have seen that the observed skewness and excess kurtosis of the log commodity index returns differ from the ones of a normal distribution. The correlation between the return process and the volatility process ρ_2 influences the skewness. The amount of excess kurtosis is managed by the volatility of volatility. If one choses the correlation ρ_2 to be positive then an increasing index value S implies an increase in volatility. So larger gains are possible. A down-move in the index price is associated with down-move in volatility and great losses are avoided. Thus the choice of a positive correlation implies a left skewed distribution of returns $dS(t)/S(t)$. A negative correlation $\rho_2 < 0$ implies a right skewed distribution: If the index value S rises then the volatility falls. Positive outliers occur less often than in comparison to $\rho_2 = 0$. Furthermore, a falling value S is associated with an increasing volatility value. The probability of observing negative outliers is higher.

The volatility of volatility σ_V influences the excess kurtosis of the return distribution of the spot: Setting σ_V equal to zero yields a normal distribution and thus the excess kurtosis is equal to 0. For increasing value for σ_V the excess kurtosis rises.

A positive correlation between instantaneous return and convenience yield ρ_1 yields the following: An up-move of the index value is connected with an up-move of the convenience yield which in turn yields a smaller drift of the return. A down-move of the index value forces the convenience yield to be smaller. The drift of the return process becomes larger. A negative correlation has the reverse effect.

It is clear that the above mentioned effects interfere with each other and cannot be separated in such a clear manner.

2.2. Closed form solution for Commodity Futures

Let $F(t, \tau, S(t), \delta(t), V(t))$ denote the futures price at time t with time to maturity τ. Since all traded assets are martingales and keeping the considerations of section 1.1 in mind the following holds

$$F(t, \tau, S(t), \delta(t), V(t)) = \mathbb{E}\left[S(t+\tau)|\mathcal{F}(t)\right] \tag{2.6}$$

Note that the model has a Markovian structure, so the value of the conditional expectation can be written as a function of the state variables. Define $L(t) := \ln S(t)$. The reason for considering $L(t)$ instead of $S(t)$ is the simplification of the valuation equations or partial differential equations: One has to solve one ordinary differential equation less in 2.10 and 2.29 because process S does not appear when switching to $\ln S$. With the help of Itô's lemma I obtain

$$dL = \left(r - \delta - 0.5(V + \sigma_S^2)\right) dt + \sigma_S dW_1 + \sqrt{V} dW_2$$

Since futures contracts cost nothing to enter, its expected return must be zero. Applying Itô's lemma to $F(t, \tau, L, \delta, V)$ and setting the drift equal to zero yield the following partial differential equation

$$\begin{aligned} -F_\tau + F_L(r - \delta - 0.5(V + \sigma_S^2)) + F_\delta(\theta_\delta - \kappa_\delta \delta) + F_V(\theta_V - \kappa_V V) \\ +0.5(\sigma_S^2 + V)F_{LL} + 0.5\sigma_\delta^2 F_{\delta\delta} + 0.5\sigma_V^2 V F_{VV} \\ +F_{L\delta}\sigma_S\sigma_\delta\rho_1 + F_{LV}\sigma_V V \rho_2 = 0 \end{aligned} \tag{2.7}$$

subject to the boundary condition

$$F(t, 0, L, \delta, V) = S(t) \tag{2.8}$$

where I suppress the dependency of the futures price (F instead of $F(t, \tau, L, \delta, V, r)$). Notice that I use $F_t = -F_\tau$: The reason for switching from time t to the remaining time τ is the fact that one can transform a terminal condition to an initial condition, namely 2.8. The problem of solving the partial differential equation 2.7 subject to 2.8 is called the Cauchy problem (see Friedman [Fri75] section 6.4, page 139). For solving the partial differential equation I try the ansatz

$$F(t, \tau, L, \delta, V) = \exp\left\{\ln(S(t)) + \beta_0(\tau) + \beta_\delta(\tau)\delta(t) + \beta_V(\tau)V(t)\right\} \tag{2.9}$$

The partial derivatives are $F_L = F$, $F_{LL} = F$, $F_\delta = \beta_\delta F$, $F_{\delta\delta} = \beta_\delta^2 F$, $F_{L\delta} = \beta_\delta F$, $F_V = \beta_V F$, $F_{VV} = \beta_V^2 F$, $F_{LV} = \beta_V F$ and $F_\tau = [\beta_0' + \beta_\delta' \delta + \beta_V' V] F$. Substituting the partial derivatives into the partial differential equation yields

$$-F\left(\beta_0' + \beta_\delta'\delta + \beta_V' V\right) + F(r - \delta - 0.5(V + \sigma_S^2)) + \beta_\delta F(\theta_\delta - \kappa_\delta \delta) + \beta_V F(\theta_V - \kappa_V V)$$
$$+0.5(\sigma_S^2 + V)F + 0.5\sigma_\delta^2\beta_\delta^2 F + 0.5\sigma_V^2 V \beta_V^2 F$$
$$+\beta_\delta F \sigma_S \sigma_\delta \rho_1 + \beta_V F \sigma_V V \rho_2 = 0.$$

I group by state variables δ and V

$$\begin{aligned} F\Big[\delta\left(-\beta_\delta' - 1 - \kappa_\delta\beta_\delta\right) \\ +V\left(-\beta_V' + \beta_V\left(\sigma_V\rho_2 - \kappa_V\right) + 0.5\sigma_V^2\beta_V^2\right) \\ +\left(-\beta_0' + 0.5\sigma_\delta^2\beta_\delta^2 + \left(\sigma_S\sigma_\delta\rho_1 + \theta_\delta\right)\beta_\delta + \beta_V\theta_V + r\right)\Big] = 0. \end{aligned} \tag{2.10}$$

So I obtain the following ordinary differential equations

$$\begin{aligned} \beta_\delta' &= -1 - \kappa_\delta\beta_\delta \\ \beta_0' &= 0.5\sigma_\delta^2\beta_\delta^2 + \left(\sigma_S\sigma_\delta\rho_1 + \theta_\delta\right)\beta_\delta + \beta_V\theta_V + r \\ \beta_V' &= \beta_V\left(\sigma_V\rho_2 - \kappa_V\right) + 0.5\sigma_V^2\beta_V^2 \end{aligned}$$

subject to the boundary conditions $\beta_\delta(0) = 0$, $\beta_V(0) = 0$ and $\beta_0(0) = 0$.

The first two equations are linear ordinary differential equations of order 1, which can be solved by standard methods. The general solution of the third ordinary differential equation is given in lemma 2.2.1.

Lemma 2.2.1. *The general solution of the ordinary differential equation*

$$y' = c_1 y + c_2 y^2, \quad c_1, c_2 \in \mathbb{R} \tag{2.11}$$

is given by

$$y(x) = \frac{c_1}{Cc_1 e^{-c_1 x} - c_2}, \quad C \in \mathbb{R}$$

Ordinary differential equation 2.11 has the singular solution $y \equiv 0$ for $C \to \infty$.

Proof: Substituting $u = y^{-1}$ I get $u' = -y^{-2}y' = -u^2 y'$ and hence $-u^{-2}u' = c_1 u^{-1} + c_2 u^{-2}$. This leads to the linear ordinary differential equation of first order $u' + c_1 u + c_2 = 0$ which has the general solution

$$u(x) = Ce^{-c_1 x} - \frac{c_2}{c_1}, \quad C \in \mathbb{R}$$

By resubstituting I obtain

$$y(x) = \frac{c_1}{Cc_1 e^{-c_1 x} - c_2}, \quad C \in \mathbb{R}$$

■

In view of $\beta_\delta(0) = 0$, $\beta_V(0) = 0$ and $\beta_0(0) = 0$ I obtain

$$\beta_\delta(\tau) = \frac{-\left(1 - e^{-\kappa_\delta \tau}\right)}{\kappa_\delta} \tag{2.12}$$

$$\begin{aligned} \beta_0(\tau) &= r\tau + \frac{\sigma_\delta^2 \tau}{2\kappa_\delta^2} - \frac{\sigma_S \sigma_\delta \rho_1 + \theta_\delta}{\kappa_\delta}\tau - \frac{\left(\sigma_S \sigma_\delta \rho_1 + \theta_\delta\right) e^{-\kappa_\delta \tau}}{\kappa_\delta^2} + \frac{4\sigma_\delta^2 e^{-\kappa_\delta \tau} - \sigma_\delta^2 e^{-2\kappa_\delta \tau}}{4\kappa_\delta^3} \\ &+ \frac{\sigma_S \sigma_\delta \rho_1 + \theta_\delta}{\kappa_\delta^2} - \frac{3\sigma_\delta^2}{4\kappa_\delta^3} \end{aligned} \tag{2.13}$$

$$\beta_V(\tau) \equiv 0$$

After providing a solution of the partial differential equation 2.7 I shall now give a rigorous proof that solution 2.9 coincides with the conditional expectation 2.6. Does 2.7 subject to boundary condition 2.8 admits a unique solution? The following theorem gives the answer:

Theorem 2.2.1. *If the spot index solves stochastic differential equations 2.1, 2.2 and 2.3 then futures prices at time t and maturity τ $F(t, \tau, \ln S(t), \delta(t))$ are given by formula*

$$F(t, \tau, \ln S(t), \delta(t)) = \exp\left\{\ln(S(t)) + \beta_0(\tau) + \beta_\delta(\tau)\delta(t)\right\} \tag{2.14}$$

where the functions $\beta_\delta(\tau)$ and $\beta_0(\tau)$ are given by 2.12 and 2.13, respectively.

Proof: The proof can be conducted by applying a uniqueness result for the Cauchy problem given in Friedman [Fri75], Corollary 6.4.4, page 141 and the Feynman-Kac theorem which can be found in Karatzas and Shreve [KS00], Theorem 5.7.6, page 366. Following the notations of Friedman [Fri75] I have to check that the coefficients matrix A is positive semi-definite

$$A = (a_{ij}(v))_{i,j=1,2,3} := \begin{pmatrix} \sigma_S^2 + v & \sigma_S \sigma_\delta \rho_1 & \sigma_V v \rho_2 \\ \sigma_S \sigma_\delta \rho_1 & \sigma_\delta^2 & 0 \\ \sigma_V v \rho_2 & 0 & \sigma_V^2 v \end{pmatrix}$$

By assumption σ_S, σ_δ, σ_V and v are strictly positive and $\rho_1, \rho_2 \in [-1, 1]$. Now I show that matrix A is positive definite if additionally $|\rho_1| \neq 1$ or $|\rho_2| \neq 1$. By Sylvester's criterion a symmetric matrix A is positive definite if and only if the three principle minors P_1, P_2 and P_3 are positive. Now

$$P_1 = \sigma_S^2 + v > 0$$

and

$$\begin{aligned} P_2 &= \begin{vmatrix} \sigma_S^2 + v & \sigma_S\sigma_\delta\rho_1 \\ \sigma_S\sigma_\delta\rho_1 & \sigma_\delta^2 \end{vmatrix} = \left(\sigma_S^2 + v\right)\sigma_\delta^2 - \sigma_S^2\sigma_\delta^2\rho_1^2 \\ &= \sigma_S^2\sigma_\delta^2 + v\sigma_\delta^2 - \sigma_S^2\sigma_\delta^2\rho_1^2 = \left(1-\rho_1^2\right)\sigma_S^2\sigma_\delta^2 + v\sigma_\delta^2 > 0 \end{aligned}$$

Furthermore,

$$\begin{aligned} P_3 &= \begin{vmatrix} \sigma_S^2 + v & \sigma_S\sigma_\delta\rho_1 & \sigma_V v\rho_2 \\ \sigma_S\sigma_\delta\rho_1 & \sigma_\delta^2 & 0 \\ \sigma_V v\rho_2 & 0 & \sigma_V^2 v \end{vmatrix} = \left(\sigma_S^2 + v\right)\sigma_\delta^2\sigma_V^2 v - \sigma_V^2 v^2\rho_2^2\sigma_\delta^2 - \sigma_V^2 v\sigma_S^2\sigma_\delta^2\rho_1^2 \\ &= \left(1-\rho_1^2\right)\sigma_V^2 v\sigma_S^2\sigma_\delta^2 + \left(1-\rho_2^2\right)v^2\sigma_\delta^2\sigma_V^2 > 0 \end{aligned}$$

It follows that A is positive definite, in particular positive semidefinite. I remark that Sylvester's criterion is not expandable to semidefinite matrices, so the additional assumption $|\rho_1| \neq 1$ or $|\rho_2| \neq 1$ is necessary.

Let $(b_i(v,\delta))_{i=1,2,3}$ denote the coefficients appearing in the partial differential equation 2.7 with respect to the partial derivatives $\frac{\partial f}{\partial l}$, $\frac{\partial f}{\partial \delta}$ and $\frac{\partial f}{\partial v}$, namely

$$(b_i(\delta,v))_{i=1,2,3} = \left(r - \delta - 0.5(v+\sigma_S^2), \theta_\delta - \kappa_\delta\delta, \theta_V - \kappa_V v\right)$$

The coefficients have to fulfill the polynomial/linear growth conditions for some constant C

$$\begin{aligned} |a_{ij}(v)| &\leq C(v^2+1) \text{ and} \\ |b_i(v,\delta)| &\leq C(\sqrt{v^2+\delta^2}+1). \end{aligned}$$

The coefficients can be estimated by

$$\begin{aligned} |a_{11}(v)| &= \left|\sigma_S^2 + v\right| = \sigma_S^2 + v \leq \left(\frac{\sigma_S^2}{2} + \frac{\sqrt{\sigma_S^4+1}}{2}\right)\left(v^2+1\right), \\ |a_{12}(v)| &= \sigma_S\sigma_\delta\,|\rho_1| \leq \sigma_S\sigma_\delta\left(v^2+1\right), \\ |a_{13}(v)| &= \sigma_V v\,|\rho_2| \leq \sigma_V v \leq \sigma_V\left(v^2+1\right), \\ |a_{22}(v)| &= \sigma_\delta^2 \leq \sigma_\delta^2\left(v^2+1\right) \text{ and} \\ |a_{33}(v)| &= \sigma_V^2 v \leq \sigma_V^2\left(v^2+1\right) \end{aligned}$$

By assumption θ_δ, κ_δ, θ_V and κ_V are positive, so the following estimates hold

$$\begin{aligned} |b_1(v,\delta)| &= \left|r - \delta - 0.5(v+\sigma_S^2)\right| \leq r + \sigma_S^2 + |\delta| + v \leq \left(\sqrt{2} + r + \sigma_S^2\right)\left(\sqrt{v^2+\delta^2}+1\right), \\ |b_2(v,\delta)| &= |\theta_\delta - \kappa_\delta\delta| \leq \theta_\delta + \kappa_\delta\,|\delta| \leq (\theta_\delta + \kappa_\delta)(\sqrt{v^2+\delta^2}+1) \text{ and} \\ |b_3(v,\delta)| &= |\theta_V - \kappa_V v| \leq \theta_V + \kappa_V v \leq (\theta_V + \kappa_V)(\sqrt{v^2+\delta^2}+1) \end{aligned}$$

where I applied the Cauchy-Schwarz inequality in the estimation of b_1. With

$$C := \max\left\{\frac{\sigma_S^2}{2} + \frac{\sqrt{\sigma_S^4+1}}{2}, \sigma_S\sigma_\delta, \sigma_V, \sigma_\delta^2, \sigma_V^2, \sqrt{2}+r+\sigma_S^2, \theta_\delta+\kappa_\delta, \theta_V+\kappa_V\right\}$$

the growth conditions are fulfilled. So I have checked the assumptions of Corollary 6.4.4 in Friedman [Fri75] and I conclude that there exists at most one solution of the Cauchy problem. But I have already derived a solution of the problem, so this solution is unique. Furthermore, the solution satisfies the growth condition (see Friedman [Fri75], Corollary 6.4.4)

$$|F(t,l,\delta,v)| \leq N\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q\right) \tag{2.15}$$

for some positive constants N, q.

The growth condition 2.15 implies the assumption of the Feynman-Kac theorem:

$$|F(t,l,\delta,v)| \leq M\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^{2\mu}\right) \tag{2.16}$$

for some positive constants M and $\mu \geq 1$.

Namely, if $q \geq 2$ then choose $M := N$ and $\mu := q/2$. Now let $0 < q < 2$. I set $K := \max\left\{(l,\delta,v) : \sqrt{l^2+\delta^2+v^2} \leq 1\right\}$. Since the function

$$(l,\delta,v) \rightarrow 1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q$$

is continuous on K and K is compact there exists an R such that

$$1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q \leq R \quad \text{for all } (l,\delta,v) \in K$$

Choose $M := NR$ and $\mu = 1$.

If $(l,\delta,v) \in K$ then the following estimation is valid

$$|F(t,l,\delta,v)| \leq N\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q\right) \leq NR = M \leq M\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^{2\mu}\right)$$

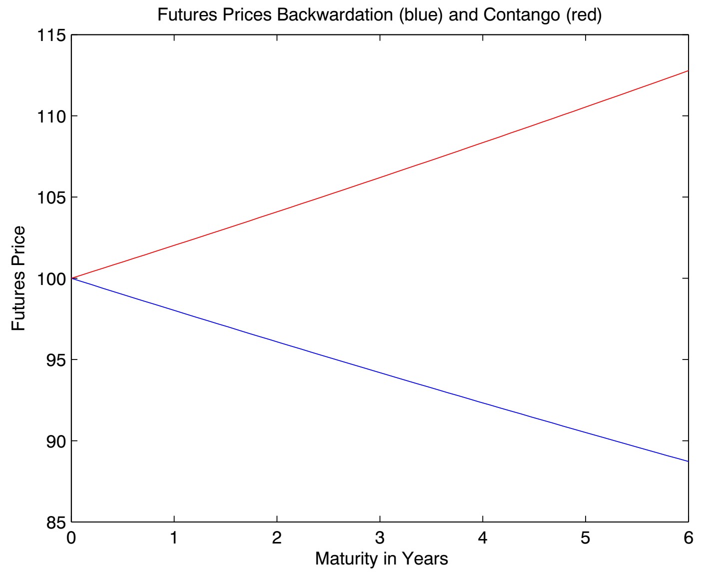
In the case of $(l,\delta,v) \notin K$, that is, $\sqrt{l^2+\delta^2+v^2} > 1$ then it holds

$$1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q \leq 1+\left(\sqrt{l^2+\delta^2+v^2}\right)^{2\mu}$$

So I obtain

$$|F(t,l,\delta,v)| \leq N\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^q\right) \leq M\left(1+\left(\sqrt{l^2+\delta^2+v^2}\right)^{2\mu}\right)$$

Figure 2.1.: Futures Prices: Backwardation (blue line) and Contango (red line) in dependence of $0.0 \leq \tau \leq 6.0$



Since growth condition 2.16 holds I can conclude via the Feynman-Kac theorem that the solution of the Cauchy problem $F(t, \tau, L(t), \delta(t))$ admits the stochastic representation

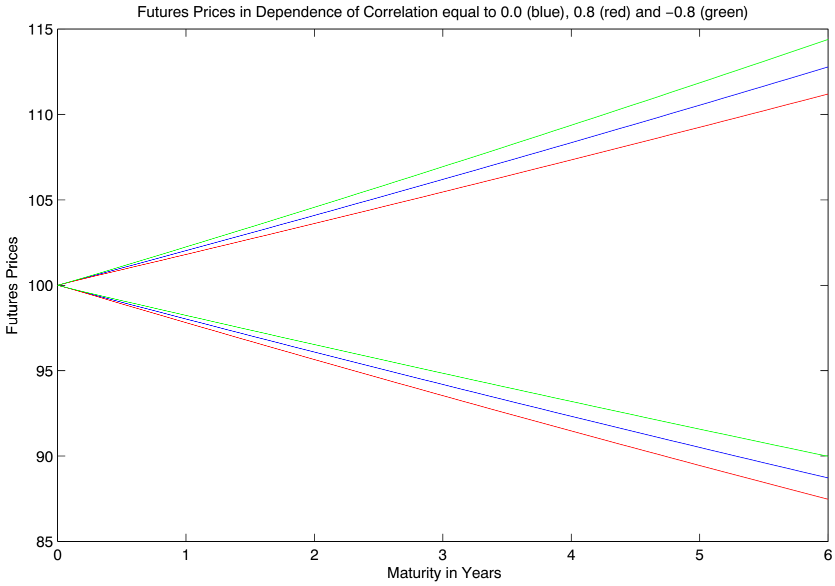
$$F(t, \tau, \ln S(t), \delta(t)) = \mathbb{E}\left[S(t+\tau)|\mathcal{F}(t)\right]$$

on $[t, \tau] \times \mathbb{R}^2$. ■

As already mentioned in the introduction the futures prices do not depend on the volatility parameters: In the formula for futures prices 2.14 the volatility parameters do not appear. So in our model setting with stochastic volatility and stochastic convenience yield the volatility has no influence on futures prices. But the convenience yield plays a crucial role. As mentioned above the values of futures prices are the expectation of the underlying spot value of the commodity. So more generally speaking the volatility does not change the first moment of the spot prices.

I have implemented futures formula 2.14 in MATLAB and shall discuss the behavior of futures prices in dependence on several parameters. First I am interested in displaying contango and backwardation. Figure 2.1 was generated with the following parameters: $S(0) = 100$, $\tau \in [0.0, 6.0]$, $r = 0.03$, $\delta(0) = 0.05$ in case of the blue line (backwardation) and $\delta(0) = 0.01$ in the case of contango, $\sigma_S = 0.3$, $\kappa_\delta = 10.0$, $\sigma_\delta = 0.1$, $\theta_\delta = 0.5$ for backwardation and $\theta_\delta = 0.1$ for

Figure 2.2.: Futures Prices in Dependence of Correlation: $\rho_1 = -0.8$ (green), $\rho_1 = 0.0$ (blue) and $\rho_1 = 0.8$ (red)



contango, $\rho_1 = 0.0$. To discover contango and backwardation I have chosen a very high value for the speed of mean reversion κ_δ so that the convenience yield behaves nearly nonrandom with value $\delta(t) \approx \delta(0) = 0.05$ or $\delta(t) \approx \delta(0) = 0.01$. Note that the mean value is given by $\frac{\theta_\delta}{\kappa_\delta}$. Additionally, the volatility of the convenience yield σ_δ was set to a relative low value, namely 0.1. Consult page 27 for an explanation why it is clear that this choice of parameters lead to the curves.

In the next step I am interested in analyzing the influence of the correlation between instantaneous return process and convenience yield ρ_1 on the term structure of futures. In figure 2.2 I have set the parameters equal to the parameters of figure 2.1. The only difference is the varying correlation ρ_1. A zero correlation is compared with a positive correlation $\rho_1 = 0.8$ as well with a negative correlation $\rho_1 = -0.8$. One can see that in backwardation as well as in contango a negative correlation leads to higher futures prices (the green lines lie above the blue lines) and a positive correlation leads to smaller futures prices as in the case of zero correlation (the red lines lie below the blue lines). This becomes immediately clear when considering the formula of futures prices 2.14:

$$F(0,\tau,L(0),\delta(0)) = \exp\left(\dots - \frac{\sigma_S\sigma_\delta\tau\rho_1}{\kappa_\delta} - \frac{\sigma_S\sigma_\delta e^{-\kappa_\delta\tau}\rho_1}{\kappa_\delta^2}\dots\right).$$

Note that all volatility parameters σ_S and σ_δ as well as κ_δ are greater than zero.

2.3. European Options on Futures

Let $C(t,\tau)$ denote the price of an European call option on a futures contract that matures in $\tilde{\tau} > \tau$, where τ is the time to maturity of the option contract and K is the strike price. Denote by $lag = \tilde{\tau} - \tau$ the fixed difference of the two maturities. The European call option is priced as the expected discounted payoff

$$C(t,\tau) = e^{-r\tau}\mathbb{E}\left[[F(t+\tau, lag, L, \delta) - K]^+\right] \tag{2.17}$$

where $F(t+\tau, lag, L, \delta)$ denotes the futures price. As derived in section 2.2 it does not depend on the volatility. As with the reasoning in subsection 2.2 the contingent claim $C(t,\tau)$ satisfies the pricing partial differential equation

$$\begin{aligned}-C_\tau + C_L(r - \delta - 0.5(V + \sigma_S^2)) + C_\delta(\theta_\delta - \kappa_\delta\delta) + C_V(\theta_V - \kappa_V V)\\ +0.5(\sigma_S^2 + V)C_{LL} + 0.5\sigma_\delta^2 C_{\delta\delta} + 0.5\sigma_V^2 V C_{VV}\\ +C_{L\delta}\sigma_S\sigma_\delta\rho_1 + C_{LV}\sigma_V V\rho_2 = rC\end{aligned} \tag{2.18}$$

subject to the boundary condition

$$C(t,0,\phi) = (F(t, lag, L, \delta) - K)^+ \tag{2.19}$$

In the following I abbreviate the futures price with $F(t+\tau, lag)$. I define the indicator function $1_{\{F(t+\tau,lag)>K\}}$ as being unity when $F(t+\tau, lag) > K$ and zero otherwise.

Proposition 2.3.1 taken from Bakshi and Madan [BM00] shows that knowing the prices of three basic options is equivalent to solving the expectation 2.17.

Proposition 2.3.1. *Model independent one can decompose the call price $C(t,\tau)$ to*

$$C(t,\tau) = G(t,\tau)\,\Pi_1(t,\tau) - e^{-r\tau}K\Pi_2(t,\tau) \tag{2.20}$$

where

$$G(t,\tau) := e^{-r\tau}\mathbb{E}\left[F(t+\tau, lag)\right]$$
$$\Pi_1(t,\tau) := e^{-r\tau}\frac{\mathbb{E}\left[F(t+\tau, lag)\,1_{\{F(t+\tau,lag)>K\}}\right]}{G(t,\tau)}$$
$$\Pi_2(t,\tau) := \mathbb{E}\left[1_{\{F(t+\tau,lag)>K\}}\right]$$

Proof: The assertions follow immediately from

$$\begin{aligned} C(t,\tau) &= e^{-r\tau}\mathbb{E}\left[F(t+\tau, lag)\,1_{\{F(t+\tau,lag)>K\}}\right] \\ &- e^{-r\tau}K\mathbb{E}\left[1_{\{F(t+\tau,lag)>K\}}\right] \\ &= G(t,\tau)\,e^{-r\tau}\frac{1}{G(t,\tau)}\mathbb{E}\left[F(t+\tau, lag)\,1_{\{F(t+\tau,lag)>K\}}\right] \\ &- e^{-r\tau}K\mathbb{E}\left[1_{\{F(t+\tau,lag)>K\}}\right] \\ &= G(t,\tau)\,\Pi_1(t,\tau) - e^{-r\tau}K\Pi_2(t,\tau) \end{aligned}$$

■

Note that Π_1 and Π_2 can be interpreted as options as well as probabilities since $\Pi_j \in [0,1]$, for $j = 1,2$.

I now want to reveal that the three securities appearing in 2.20 are all related to a characteristic function. This fact will be useful to derive a formula for $C(t,\tau)$. To this end, I define the discounted characteristic function of the logarithm of the futures price

$$f(t,\tau;\phi) \equiv e^{-r\tau}\mathbb{E}\left[e^{i\phi\ln F(t+\tau,lag)}\right] \tag{2.21}$$

which is implicitly the time-t price of a hypothetical claim that pays $e^{i\phi\ln F(t+\tau,lag)}$ (where i is the imaginary unit and ϕ is some parameter of the contract). So the characteristic function satisfies the pricing partial differential equation

$$\begin{aligned} -f_\tau + (r-\delta-0.5(V+\sigma_S^2))f_L + (\theta_\delta - \kappa_\delta\delta)f_\delta + (\theta_V - \kappa_V V)f_V \\ +0.5(\sigma_S^2+V)f_{LL} + 0.5\sigma_\delta^2 f_{\delta\delta} + 0.5\sigma_V^2 V f_{VV} \\ +f_{L\delta}\sigma_S\sigma_\delta\rho_1 + f_{LV}\sigma_V V\rho_2 = rf \end{aligned} \tag{2.22}$$

subject to the boundary condition

$$\begin{aligned} f(t,0,\phi) &= \exp\{i\phi\ln F(t,lag)\} \\ &= \exp\{i\phi\left[L(t) + \beta_0(lag) + \beta_\delta(lag)\delta(t)\right]\} \end{aligned} \tag{2.23}$$

where I applied the closed form solution for the futures price at maturity of the option derived in the previous subsection. The reason for introducing f becomes clear in the following theorem. The values of the three basic options of proposition 2.3.1 can be immediately obtained if one has calculated an explicit solution for the characteristic function f.

Theorem 2.3.1. *Let function $f(t,\tau;\phi)$ be the discounted characteristic function of the random variable $\ln F\,(t+\tau, lag)$ given by 2.21. If the random variable $\ln F\,(t+\tau, lag)$ exhibits a continuous cumulative distribution function then it holds that the three basic securities $G\,(t,\tau)$, $\Pi_1(t,\tau)$ and $\Pi_2(t,\tau)$ of proposition 2.3.1 are related to $f(t,\tau;\phi)$ in the following way*

$$G\,(t,\tau) = f\,(t,\tau;-i) \tag{2.24}$$

$$\Pi_j(t,\tau) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\ln K}\varphi_{\Pi_j}(t,\tau;\phi)}{i\phi}\right]d\phi \tag{2.25}$$

with

$$\varphi_{\Pi_1}(t,\tau;\phi) := \frac{f\,(t,\tau;\phi-i)}{G\,(t,\tau)} \tag{2.26}$$

$$\varphi_{\Pi_2}(t,\tau;\phi) := e^{r\tau}f(t,\tau;\phi) \tag{2.27}$$

In our model setting it is understood that $f(t,\tau;\phi)$ is available in closed form by solving the valuation equation 2.22 subject to the boundary condition 2.23.

Proof: For parsimony of presentation, let $F \equiv F\,(t+\tau, lag)$. Since

$$G\,(t,\tau) = e^{-r\tau}\mathbb{E}\,[F]$$

by definition and the discounted characteristic function equals

$$f(t,\tau;\phi) = e^{-r\tau}\mathbb{E}\left[e^{i\phi\ln F}\right]$$

it holds $G\,(t,\tau) = f(t,\tau;-i)$.

$\Pi_2(t,\tau)$ was given by

$$\begin{aligned}\Pi_2(t,\tau) &= \mathbb{E}\left[1_{\{F>K\}}\right] = \mathbb{E}\left[1_{\{\ln F>\ln K\}}\right] = \mathbb{P}\left(\{\ln F > \ln K\}\right)\\ &= 1 - \mathbb{F}_{\ln F}(\ln K)\end{aligned}$$

where $\mathbb{F}$ denotes the continuous cumulative distribution function of the logarithm of the futures price. With the help of a special form of the inversion theorem which provides a connection between the cumulative distribution function of the random variable $\ln F$ and the characteristic function $g(\phi) := \mathbb{E}\left[e^{i\phi \ln F}\right] = e^{r\tau} f(\phi)$ (see Chung [Chu01], page 168, inversion formula due to Gil-Palaez) it follows

$$\Pi_2(t,\tau) = 1 - \left(\frac{1}{2} + \int_0^\infty \frac{e^{i\phi \ln K} g(-\phi) - e^{-i\phi \ln K} g(\phi)}{2\pi i\phi} d\phi\right).$$

With $\overline{g(\phi)}$ being the complex conjugate of $g(\phi)$ I can easily calculate

$$\begin{aligned} g(-\phi) &= \mathbb{E}\left[e^{-i\phi \ln F}\right] = \mathbb{E}\left[\cos(-\phi \ln F)\right] + i\mathbb{E}\left[\sin(-\phi \ln F)\right] \\ &= \mathbb{E}\left[\cos(\phi \ln F)\right] - i\mathbb{E}\left[\sin(\phi \ln F)\right] = \overline{g(\phi)} \end{aligned}$$

So I continue and obtain

$$\begin{aligned} \Pi_2(t,\tau) &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty -\frac{1}{i\phi}\left[e^{i\phi \ln K}\overline{g(\phi)} - e^{-i\phi \ln K} g(\phi)\right] d\phi \\ &= \frac{1}{2} + \frac{1}{2\pi}\int_0^\infty \overline{\frac{1}{i\phi} e^{-i\phi \ln K} g(\phi)} + \frac{1}{i\phi} e^{-i\phi \ln K} g(\phi) d\phi \\ &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi \ln K} g(t,\tau;\phi)}{i\phi}\right] d\phi \\ &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi \ln K} e^{r\tau} f(t,\tau;\phi)}{i\phi}\right] d\phi \end{aligned}$$

$\Pi_1(t,\tau)$ is given by

$$\begin{aligned} \Pi_1(t,\tau) &= \frac{\mathbb{E}\left[F 1_{\{F>K\}}\right]}{\mathbb{E}\left[F\right]} \\ &= \mathbb{E}\left[1_{\{\ln F > \ln K\}} \frac{F}{\mathbb{E}\left[F\right]}\right] \end{aligned}$$

Since the futures price F is greater than zero $\frac{F}{\mathbb{E}[F]}$ defines a probability measure. So I can apply the same calculation as for $\Pi_2(t,\tau)$ but with respect to the changed probability measure:

$$\Pi_2(t,\tau) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{e^{-i\phi \ln K}\mathbb{E}\left[e^{i\phi \ln F}\frac{F}{\mathbb{E}[F]}\right]}{i\phi}\right] d\phi$$

But it holds

$$\frac{\mathbb{E}\left[e^{i\phi \ln F} F\right]}{\mathbb{E}\left[F\right]} = \frac{f(t,\tau;\phi - i)}{f(t,\tau;-i)}$$

which yields the assertion.

■

Remark 2.3.1. *If I could determine a solution of the characteristic function partial differential equation 2.22 subject to boundary condition 2.23 then I would have given a formula for European call options on futures $C(t,\tau)$ via proposition 2.3.1 and theorem 2.3.1. But both the valuation equation for European call prices on futures prices 2.18 and the partial differential equation for the characteristic function 2.22 are obviously indistinguishable, so what is the gain? The answer is the different boundary conditions 2.23 and 2.19. The one for the characteristic function, namely 2.23, is mathematically more tractable than the one for the European call option 2.19: Due to the characteristic function's exponential boundary condition, it is easier to provide a guess for the solution of the valuation partial differential equation subject to the characteristic function's boundary condition. The boundary condition for the characteristic function is smooth and infinitely differentiable, while the boundary condition for the call option fails to be differentiable.*

For solving the partial differential equation 2.22 subject to the boundary condition 2.23 I try the ansatz

$$f(t,\tau;\phi) = \exp\left\{i\phi\left[L(t) + \beta_0(\tau + lag) + \beta_\delta(\tau + lag)\delta(t)\right] + \vartheta_0(\tau) + \vartheta_\delta(\tau)\delta(t) + \vartheta_V(\tau)V(t)\right\}$$

The partial derivatives are

$$\begin{aligned}
&f_L = i\phi f;\ f_{LL} = i^2\phi^2 f = -\phi^2 f \\
&f_\delta = \left(\vartheta_\delta + i\phi\beta_\delta\right) f;\ f_{\delta\delta} = \left(\vartheta_\delta + i\phi\beta_\delta\right)^2 f \\
&f_V = f\vartheta_V;\ f_{VV} = f\vartheta_V^2 \\
&f_{L\delta} = i\phi\left(\vartheta_\delta + i\phi\beta_\delta\right) f \\
&f_{LV} = i\phi\vartheta_V f \\
&f_\tau = f\left(\vartheta_0' + \delta\vartheta_\delta' + V\vartheta_V' + i\phi\left[\beta_0' + \beta_\delta'\delta\right]\right)
\end{aligned}$$

Substituting the partial derivatives into the partial differential equation 2.22 I get

$$\begin{aligned}
-f\left(\vartheta_0' + \delta\vartheta_\delta' + V\vartheta_V' + i\phi\left[\beta_0' + \beta_\delta'\delta\right]\right) \\
+i\phi f(r - \delta - 0.5(V + \sigma_S^2)) + f\left(\vartheta_\delta + i\phi\beta_\delta\right)\left(\theta_\delta - \kappa_\delta\delta\right) + f\vartheta_V(\theta_V - \kappa_V V) \\
+0.5(\sigma_S^2 + V)(-\phi^2 f) + 0.5\sigma_\delta^2\left(\vartheta_\delta + i\phi\beta_\delta\right)^2 f + 0.5\sigma_V^2 V f\vartheta_V^2 \\
+i\phi f\left(\vartheta_\delta + i\phi\beta_\delta\right)\sigma_S\sigma_\delta\rho_1 + i\phi\vartheta_V f\sigma_V V\rho_2 - rf = 0
\end{aligned}$$

Sorting by state variables V and δ yields the partial differential equation

$$
\begin{aligned}
&f\left[\delta\left(-\vartheta_\delta' - \kappa_\delta \vartheta_\delta\right)\right. && (2.28)\\
&+V\left(-\vartheta_V' + 0.5\sigma_V^2\vartheta_V^2 + \left(\sigma_V\rho_2 i\phi - \kappa_V\right)\vartheta_V - 0.5i\phi - 0.5\phi^2\right)\\
&-\vartheta_0' + 0.5\sigma_\delta^2\vartheta_\delta^2 + \sigma_\delta^2 i\phi\beta_\delta\vartheta_\delta + \left(\sigma_S\sigma_\delta\rho_1 i\phi + \theta_\delta\right)\vartheta_\delta + \theta_V\vartheta_V - 0.5\sigma_S^2\phi^2\\
&\left.-0.5\sigma_S^2 i\phi - 0.5\sigma_\delta^2\phi^2\beta_\delta^2 + \left(-\phi^2\sigma_S\sigma_\delta\rho_1 + \theta_\delta i\phi\right)\beta_\delta - i\phi\beta_0' + (i\phi-1)r\right] = 0 && (2.29)
\end{aligned}
$$

where I applied in 2.28 the equality $\beta_\delta' = -1 - \kappa_\delta\beta_\delta$.

So I obtain the following ordinary differential equations

$$
\begin{aligned}
\vartheta_\delta' &= -\kappa_\delta\vartheta_\delta\\
\vartheta_V' &= 0.5\sigma_V^2\vartheta_V^2 + \left(\sigma_V\rho_2 i\phi - \kappa_V\right)\vartheta_V - 0.5i\phi - 0.5\phi^2\\
\vartheta_0' &= 0.5\sigma_\delta^2\vartheta_\delta^2 + \sigma_\delta^2 i\phi\beta_\delta\vartheta_\delta + \left(\sigma_S\sigma_\delta\rho_1 i\phi + \theta_\delta\right)\vartheta_\delta + \theta_V\vartheta_V - 0.5\sigma_S^2\phi^2\\
&\quad - 0.5\sigma_S^2 i\phi - 0.5\sigma_\delta^2\phi^2\beta_\delta^2 + \left(-\phi^2\sigma_S\sigma_\delta\rho_1 + \theta_\delta i\phi\right)\beta_\delta - i\phi\beta_0' + (i\phi-1)r
\end{aligned}
$$

subject to the boundary conditions

$$\vartheta_\delta(0) = 0,\ \vartheta_V(0) = 0,\ \vartheta_0(0) = 0$$

Define

$$\xi_V := \sqrt{\left(\kappa_V - \sigma_V\rho_2 i\phi\right)^2 - i\phi\left(i\phi - 1\right)\sigma_V^2}$$

The solutions of the first two ordinary differential equations are given by

$$
\begin{aligned}
\vartheta_\delta &\equiv 0\\
\vartheta_V &= \frac{i\phi\left(i\phi-1\right)\left(1-e^{-\xi_V\tau}\right)}{2\xi_V - \left[\xi_V - \kappa_V + \sigma_V\rho_2 i\phi\right]\left(1-e^{-\xi_V\tau}\right)}. \qquad (2.30)
\end{aligned}
$$

Thus the first three terms of the ordinary differential equation for ϑ_0 vanish completely, and to facilitate the presentation of the solution of the remaining ordinary differential equation, I define

$$
\begin{aligned}
A(\tau) &:= \theta_V\vartheta_V(\tau)\\
B(\tau) &:= -0.5\sigma_S^2\phi^2 - 0.5\sigma_S^2 i\phi + (i\phi-1)r\\
C(\tau) &:= -0.5\sigma_\delta^2\phi^2\beta_\delta^2(\tau + lag)\\
D(\tau) &:= \left(-\phi^2\sigma_S\sigma_\delta\rho_1 + \theta_\delta i\phi\right)\beta_\delta(\tau + lag)\\
E(\tau) &:= -i\phi\beta_0'(\tau + lag)
\end{aligned}
$$

To obtain a solution of the ordinary differential equation I have to integrate $A, B, ..., E$:

$$\begin{aligned}\int_0^\tau A(s)ds &= \frac{h_2\left(-\xi_V\tau\right)}{\xi_V\left(-2\xi_V+h_1\right)} - \frac{2h_2\ln\left(2\xi_V-h_1+h_1e^{-\xi_V\tau}\right)}{\left(-2\xi_V+h_1\right)h_1} \\ &+ \frac{2h_2\ln\left(2\xi_V\right)}{\left(-2\xi_V+h_1\right)h_1} \\ &=: \widetilde{A}(\tau)\end{aligned}$$

with

$$\begin{aligned}h_1 &:= \xi_V-\kappa_V+\sigma_V\rho_2 i\phi, \\ h_2 &:= \theta_V i\phi\left(i\phi-1\right)\end{aligned}$$

In proposition 2.4.1 I show that $\widetilde{A}$ can be simplified to

$$\widetilde{A}(\tau,\phi) = \frac{-\tau\theta_V\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}{\sigma_V^2} + \frac{2\theta_V}{\sigma_V^2}\left(\ln\left(2\xi_V\right)-\ln\left(2\xi_V-h_1+h_1e^{-\xi_V\tau}\right)\right)$$

Note that the above version of $\widetilde{A}$ has removed the singularity at point $\phi=-i$. I continue with the integration of the other expressions:

$$\begin{aligned}\int_0^\tau B(s)ds &= \left(-0.5\sigma_S^2\phi^2-0.5\sigma_S^2 i\phi+(i\phi-1)r\right)\tau \\ &=: \widetilde{B}(\tau) \\ \int_0^\tau C(s)ds &= \frac{\phi^2\sigma_\delta^2}{4\kappa_\delta^3}\left[4\left(e^{-\kappa_\delta lag}-e^{-\kappa_\delta(\tau+lag)}\right)+e^{-2\kappa_\delta(\tau+lag)}-e^{-2\kappa_\delta lag}-2\kappa_\delta\tau\right] \\ &=: \widetilde{C}(\tau) \\ \int_0^\tau D(s)ds &= \frac{\phi^2\rho_1\sigma_S\sigma_\delta-i\phi\theta_\delta}{\kappa_\delta^2}\left(e^{-\kappa_\delta(\tau+lag)}-e^{-\kappa_\delta lag}+\kappa_\delta\tau\right) \\ &=: \widetilde{D}(\tau) \\ \int_0^\tau E(s)ds &= -i\phi\beta_0(\tau+lag)+i\phi\beta_0(lag) \\ &=: \widetilde{E}(\tau)\end{aligned}$$

Thus

$$\vartheta_0(\tau) = \widetilde{A}(\tau)+\widetilde{B}(\tau)+\widetilde{C}(\tau)+\widetilde{D}(\tau)+\widetilde{E}(\tau) \tag{2.31}$$

solves the ordinary differential equation for ϑ_0 subject to the boundary condition $\vartheta_0(0)=0$.

After obtaining a closed-form solution for f, one can compute G, φ_{Π_1} and φ_{Π_2}. Π_1 and Π_2 are then recovered by Fourier inversion

$$\Pi_j(t,\tau) = \frac{1}{2}+\frac{1}{\pi}\int_0^\infty \text{Re}\left[\frac{e^{-i\phi\ln K}\varphi_{\Pi_j}(t,\tau;\phi)}{i\phi}\right]d\phi$$

where $j \in \{1,2\}$. So I have derived a semi-closed form solution for European call options on futures that only requires an additional numerical integration.

Theorem 2.3.2. *If the spot index solves stochastic differential equations 2.1, 2.2 and 2.3 then European call prices on futures $C(t,\tau)$ at time t and maturity τ and strike price K are given by formula*

$$\begin{aligned} C(t,\tau) &= f(t,\tau;-i)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\ln K}\varphi_{\Pi_1}(t,\tau;\phi)}{i\phi}\right]d\phi\right) \\ &- e^{-r\tau}K\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\ln K}\varphi_{\Pi_2}(t,\tau;\phi)}{i\phi}\right]d\phi\right) \end{aligned} \tag{2.32}$$

with φ_{Π_1} and φ_{Π_2} given by

$$\begin{aligned} \varphi_{\Pi_1}(t,\tau;\phi) &= \frac{f(t,\tau;\phi-i)}{f(t,\tau;-i)} \\ \varphi_{\Pi_2}(t,\tau;\phi) &= e^{r\tau}f(t,\tau;\phi) \\ f(t,\tau;\phi) &= \exp\left\{i\phi\left[L(t)+\beta_0(\tau+lag)+\beta_\delta(\tau+lag)\delta(t)\right]+\vartheta_0(\tau)+\vartheta_V(\tau)V(t)\right\} \end{aligned}$$

with $L(t)=\ln S(t)$. The functions β_0 and β_δ already appeared in the derivation of the futures formula 2.13 and 2.12, respectively. Functions ϑ_0 and ϑ_V are given by 2.31 and 2.30, respectively.

Proof: With the same reasoning as in theorem 2.2.1 the Cauchy problem for the characteristic function given by 2.22 subject to 2.23 has at most one solution. Since I have explicitly derived a solution it follows that this solution is unique. Partial differential equation 2.22 differs from the partial differential equation 2.18 by the additional term $-rf$. But since the interest rate r is a constant the assumptions of the uniqueness result in Friedman [Fri75], Corollary 6.4.4, page 141 are again valid. As in 2.2.1 the solution satisfies the growth condition 2.15 which is an assumption for the applicability of the Feynman-Kac theorem (see Karatzas and Shreve [KS00] theorem 5.7.6 on page 366).

I started this section by stating that the futures call price is given by

$$C(t,\tau)=e^{-r\tau}\mathbb{E}\left[\left[F(t+\tau,lag,L,\delta)-K\right]^+\right]$$

The expectation was then split in three basic securities 2.3.1 and the relationship of the securities to a characteristic function f was revealed in 2.3.1. But f is the unique solution of the respective Cauchy problem and with the help of the Feynman-Kac theorem it coincides with the expectation

$$f(t,\tau;\phi)\equiv e^{-r\tau}\mathbb{E}\left[e^{i\phi\ln F(t+\tau,lag)}\right]$$

■

Futures and European call options are (relatively) liquid traded. So the concerns mentioned on page 25 concerning arbitrage arguments in commodity markets are not relevant here. Shorting and lending of futures and options on futures are possible, so the Put-Call parity must hold in commodity markets, too:

Corollary 2.3.1. *In the model setting of theorem 2.3.2 the no-arbitrage prices for European put options $P(t,\tau)$ with strike price K and maturity $\tau > 0$ on futures with futures maturity $\tilde{\tau} > \tau$ are given by*

$$P(t,\tau) = C(t,\tau) + Ke^{-r\tau} - F(t,\tilde{\tau}-\tau) \tag{2.33}$$

where a semi-closed form solution of European call prices $C(t,\tau)$ is given in theorem 2.3.2 and a closed form solution of futures prices $F(t,\tilde{\tau}-\tau)$ in theorem 2.14.

Proof: see Appendix A. ■

In proposition 2.3.1 I decomposed the European call option in two options

$$C(t,\tau) = e^{-r\tau}\mathbb{E}\left[F\left(t+\tau, lag\right)1_{\{F(t+\tau,lag)>K\}}\right] =: AC(t,\tau) \tag{2.34}$$

$$- \quad e^{-r\tau}K\mathbb{E}\left[1_{\{F(t+\tau,lag)>K\}}\right] =: CC(t,\tau). \tag{2.35}$$

2.34 and 2.35 are the fair values of so called *binary options* or *digital options*. The first one, namely 2.34, is the value of the option which pays off nothing if the underlying futures price ends up below the strike price and pays an amount equal to the futures price itself if it ends up above the strike price. It is called *asset-or-nothing call* (Hull [Hul06]). The second one is the value of an option which pays off nothing if the futures price ends up below the strike price at time T and pays a fixed amount, namely the strike price K, if the futures price ends up above the strike price. It is called a *cash-or-nothing call*. So the European call option is equivalent to a long position in an asset-or-nothing call and a short position in a cash-or-nothing call where the cash payoff on the cash-or-nothing call equals the strike price.

Corollary 2.3.2. *As in the model setting of theorem 2.3.2 prices for the asset-or-nothing call $AC(t,\tau)$ and cash-or-nothing call $CC(t,\tau)$ are given by*

$$AC(t,\tau) = f\left(t,\tau;-i\right)\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty Re\left[\frac{e^{-i\phi\ln K}\varphi_{\Pi_1}(t,\tau;\phi)}{i\phi}\right]d\phi\right)$$

and

$$CC(t,\tau) = e^{-r\tau} K \left(\frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re \left[\frac{e^{-i\phi \ln K} \varphi_{\Pi_2}(t,\tau;\phi)}{i\phi} \right] d\phi \right),$$

respectively, where φ_{Π_1} *and* φ_{Π_2} *are related to function* f *as in Theorem 2.3.2.*

Proof: Follows immediately from theorem 2.3.2. ■

Remark 2.3.2. *In this section the Fourier transform for all* $\phi \in \mathbb{R}$

$$f(\phi, t) = \mathbb{E}\left[e^{i\phi X(t)} \right]$$

is a very useful vehicle to derive option prices. The state variable $X(t) = \ln F(t,\tau)$ *is equal to the logarithm of the futures price. The article of Duffie, Pan and Singleton [DPS98] provides a generalized Fourier transform for all* $u \in \mathbb{C}$*:*

$$f(u, t) = \mathbb{E}\left[e^{uX(t)} \right]$$

Deng [Den99] applies the methodology to the pricing of options with respect to three different electricity models. With the help of Itô's lemma for complex variables the valuation partial differential equations are stated. These partial differential equations with the corresponding boundary conditions for the different options are easier to solve than the original ones for the options. Similar to the methodology of this section basic options can be written as integrals of the general Fourier transform. These options then serve as building blocks in pricing more complex contingent claims.

The derived formulas for the different options can be used for a **calibration of the parameters of the model**. Calibration is the process of identifying the set of model parameters that are most likely given the observed data. I now remark how this can be conducted. The set $\mathcal{S}$ of parameters to be estimated is

$$\mathcal{S} = \{r, S_0, \sigma_S, \delta_0, \rho_1, \rho_2, \theta_\delta, \kappa_\delta, \sigma_\delta, V_0, \theta_V, \kappa_V, \sigma_V\}$$

It is clear that the more parameters have to be estimated the more unstable is the calibration. So this is a drawback of the model. Let $\mathcal{FU}$ be the set of observed futures prices with different maturities and let $\mathcal{C}$ be the set of market values for European call options. I calibrate by minimizing the squared relative differences between observed values and theoretical values

$$s^* = \text{argmin}_{\mathcal{S}} \left(\sum_{F^{\text{obs}} \in \mathcal{FU}} \left(\frac{F^{\text{theor}} - F^{\text{obs}}}{F^{\text{obs}}} \right)^2 + \sum_{C^{\text{obs}} \in \mathcal{C}} \left(\frac{C^{\text{theor}} - C^{\text{obs}}}{C^{\text{obs}}} \right)^2 \right)$$

As the calculation of C^{theor} involves numerical integration and F^{theor} has a closed form solution, I first precalibrate the parameters which influence the futures prices. These are then used as starting values for the minimization over the whole parameterset $\mathcal{S}$.

Hence, the first parameters to be estimated are

$$\mathcal{S}^f = \{r, S_0, \sigma_S, \delta_0, \rho_1, \theta_\delta, \kappa_\delta, \sigma_\delta\}$$

Minimizing the squared relative differences between observed values and theoretical values of the futures prices yields

$$s^f = \mathrm{a}rg \min_{\mathcal{S}^f} \left(\sum_{F^{\text{obs}} \in \mathcal{FU}} \left(\frac{F^{\text{theor}} - F^{\text{obs}}}{F^{\text{obs}}} \right)^2 \right)$$

Given s^f, all parameters in $\mathcal{S}$ can now be estimated and the resulting s^* can be used to price other derivatives, for example Asian Options.

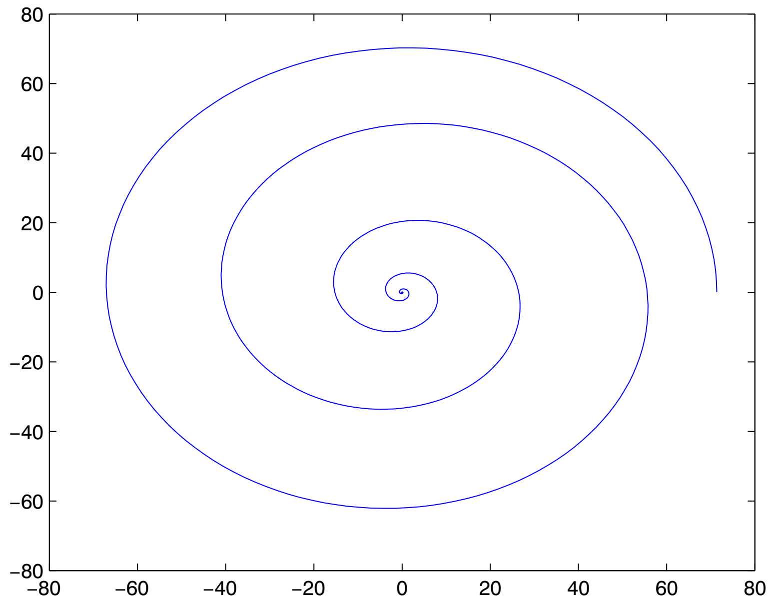
Another way of calibrating the parameters is via a time series of historical index data. The model setting in this chapter was under a risk neutral measure. So if one uses historical data the model should be formulated under the physical measure. The option data reflect the future expectation of the market whereas the times series is a historical view. In practice it is known that there exist problems with calibrating a model with both stochastic volatility and stochastic drift via time series data. The model cannot differentiate if for example an up-move in the index value comes from a changing drift or a changing volatility. But note that there are no problems if one applies option prices.

2.4. Implementation and Verification with Monte Carlo Simulation

I implemented the above formula in MATLAB and compared the calculated prices with Monte Carlo simulations. The reason for conducting a Monte Carlo simulation is to verify the derived semi-closed form solution for a European call option on futures and the closed form solution for futures prices.

In the implementation of closed form solution 2.32 one has to be aware of the fact, that the function $f(t, \tau; \phi)$ has a removable singularity at $\phi = -i$ and function f has to be evaluated at point $\phi = -i$ in the semi-closed form solution of European options. To illustrate the behavior of function $f(t, \tau; \phi)$ at point $\phi = -i$ I generated figure 2.4. On the x-axis I plotted the real

Figure 2.3.: $f(t, \tau, \phi - i)$ for $0.0001 \leq \phi \leq 50$: x-axis real part and y-axis imaginary part



part of $f(t, \tau; \phi - i)$ and on the y-axis the imaginary part with $\phi \in [0.0001, 50]$. One can see that for ϕ starting at 50 and approaching 0.0001 both the real and imaginary part of function $f(t, \tau; \phi - i)$ converge to 0. The plot was generated with the following parameters $S(0) = 100$, $r = 0.08$, $\tilde{\tau} = 0.8$, $\tau = 0.6$, $K = 100$, $\sigma_S = 0.3$, $\kappa_\delta = 0.6$, $\sigma_\delta = 0.2$, $\theta_\delta = 0.9$, $\kappa_V = 0.2$, $\sigma_V = 0.3$, $\theta_V = 0.3$, $\rho_1 = 0.8$, $\rho_2 = 0.5$, $\delta(0) = 0.15$ and $V(0) = 0.25$. Proposition 2.4.1 gives the proof that $f(t, \tau; \phi)$ has a removable singularity at point $\phi = -i$ for arbitrary parameter sets.

Proposition 2.4.1. *For fixed point t and maturity $\tau > 0$ let function $f(t, \tau; \phi)$ be given by*

$$f(t, \tau; \phi) = \exp\left\{i\phi\left[L(t) + \beta_0(\tau + lag) + \beta_\delta(\tau + lag)\delta(t)\right] + \vartheta_0(\tau) + \vartheta_\delta(\tau)\delta(t) + \vartheta_V(\tau)V(t)\right\} \tag{2.36}$$

where β_0 and β_δ are specified in 2.13 and 2.12, respectively. Functions ϑ_0 and ϑ_V are given by 2.31 and 2.30, respectively.
Then function $f(t, \tau; \phi)$ has a removable singularity at point $\phi = -i$.

Proof: A composition of holomorphic functions is again a holomorphic function in their joint domain of definition. Examining function f at $\phi = -i$ one observes that $\widetilde{A}$ appearing in function

ϑ_0 is causing a division by zero and the denominator is equal to zero as well:

$$\begin{aligned}\widetilde{A}(\tau,\phi=-i) &= \frac{h_1h_2\left(-\xi_V\tau\right)-\xi_V 2h_2\ln\left(2\xi_V-h_1+h_1e^{-\xi_V\tau}\right)}{\xi_V\left(-2\xi_V+h_1\right)h_1}\\ &+ \frac{2h_2\ln\left(2\xi_V\right)}{\left(-2\xi_V+h_1\right)h_1}\end{aligned}$$

with

$$\begin{aligned}\xi_V(-i) &= \kappa_V-\sigma_V\rho_2\\ h_1(-i) &= \xi_V(-i)-\kappa_V+\sigma_V\rho_2=0\\ h_2(-i) &= 0\end{aligned}$$

I now simplify $\widetilde{A}$ and show that $\widetilde{A}$ evaluated at point $\phi=-i$ is equal to 0

$$\begin{aligned}\widetilde{A}(\tau,\phi) &= \frac{-h_2\tau}{\left(-2\xi_V+h_1\right)}+\frac{-2h_2\ln\left(2\xi_V-h_1+h_1e^{-\xi_V\tau}\right)}{\left(-2\xi_V+h_1\right)h_1}+\frac{2h_2\ln\left(2\xi_V\right)}{\left(-2\xi_V+h_1\right)h_1}\\ &= \frac{-h_2\tau}{\left(-2\xi_V+h_1\right)}+\frac{2h_2}{\left(-2\xi_V+h_1\right)h_1}\left(\ln\left(2\xi_V\right)-\ln\left(2\xi_V-h_1+h_1e^{-\xi_V\tau}\right)\right)\end{aligned}$$

With

$$\begin{aligned}\frac{h_2\tau}{\left(-2\xi_V+h_1\right)} &= \frac{h_2\tau}{-2\xi_V+\xi_V-\kappa_V+\sigma_V\rho_2 i\phi}\\ &= \frac{h_2\tau\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}{\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)-\xi_V\right)\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}\\ &= \frac{\tau\theta_V i\phi\left(i\phi-1\right)\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}{\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-\xi_V^2}\\ &= \frac{\tau\theta_V i\phi\left(i\phi-1\right)\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}{\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-i\phi\left(i\phi-1\right)\sigma_V^2\right)}\\ &= \frac{\tau\theta_V\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}{\sigma_V^2}\end{aligned}$$

and

$$\begin{aligned}\frac{2h_2}{\left(-2\xi_V+h_1\right)h_1} &= \frac{2h_2}{\left(-2\xi_V+\xi_V-\kappa_V+\sigma_V\rho_2 i\phi\right)\left(\xi_V-\kappa_V+\sigma_V\rho_2 i\phi\right)}\\ &= \frac{2h_2}{\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)-\xi_V\right)\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)+\xi_V\right)}\\ &= \frac{2\theta_V i\phi\left(i\phi-1\right)}{\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-\xi_V^2}\\ &= \frac{2\theta_V i\phi\left(i\phi-1\right)}{\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-\left(\left(\sigma_V\rho_2 i\phi-\kappa_V\right)^2-i\phi\left(i\phi-1\right)\sigma_V^2\right)}\\ &= \frac{2\theta_V}{\sigma_V^2}\end{aligned}$$

I get

$$\widetilde{A}(\tau,\phi) = \frac{-\tau\theta_V\left((\sigma_V\rho_2 i\phi - \kappa_V) + \xi_V\right)}{\sigma_V^2} + \frac{2\theta_V}{\sigma_V^2}\left(\ln(2\xi_V) - \ln\left(2\xi_V - h_1 + h_1 e^{-\xi_V\tau}\right)\right)$$

Because of

$$\begin{aligned} \xi_V(-i) &= \kappa_V - \sigma_V\rho_2 \\ h_1(-i) &= \xi_V(-i) - \kappa_V + \sigma_V\rho_2 = 0 \end{aligned}$$

I obtain

$$\widetilde{A}(\tau,\phi=-i) = \frac{-\tau\theta_V\left((\sigma_V\rho_2 i\phi - \kappa_V) + \kappa_V - \sigma_V\rho_2\right)}{\sigma_V^2} + \frac{2\theta_V}{\sigma_V^2}\left(\ln(2\xi_V) - \ln(2\xi_V)\right) = 0$$

■

Due to the fact that the two integrands appearing in the futures call formula 2.32 can vary in its shape from almost simply exponentially decaying to highly oscillatory depending on the choice of parameters, most simple quadrature or numerical integration schemes are bound to fail. I used a more advanced scheme, namely the adaptive Lobatto quadrature to numerically evaluate the integrals in the formula. Plot 2.4 and 2.5 illustrate the different possible shapes of the integrands in formula 2.32. The first one is smooth in contrast to the oscillating second one. Note that the parameterset of the first plot was generated by the parameters as for table 2.5 but with $\tilde{\tau} = 0.8$ while for the second plot I increased the maturity of the call option and the volatilities (parameterset as on page 46 but with $\tilde{\tau} = 4.0$, $\tau = 0.2$ and $r = 0.02$).

In the Monte Carlo simulation I applied the Euler discretization to simulate the stochastic differential equations 2.1, 2.2 and 2.3. I use Monte Carlo simulation as a verification tool due to its almost sure convergence ensured by the strong law of large numbers but in this place I remark that although very flexible, the Monte Carlo simulation is not a very practical tool since the computational time it requires is often enormously long.
For each simulation $0 \leq i \leq N$ the end values of $S^i(\tau)$ and $\delta^i(\tau)$ at maturity of the option τ are then plugged in the futures formula 2.9. With N being the number of simulations and $F^i(\tau,\tilde{\tau},S(\tau),\delta(\tau))$ denoting the i-th simulated futures price the Monte Carlo estimator for the call option is then calculated via

$$MC = \frac{1}{N}\sum_{i=1}^{N} e^{-r\tau}\left(F^i(\tau,\tilde{\tau},S(\tau),\delta(\tau)) - K\right)^+ \tag{2.37}$$

Additionally, we used two different variance reduction approaches:

Figure 2.4.: Example of Smooth Integrands of Formula 2.32

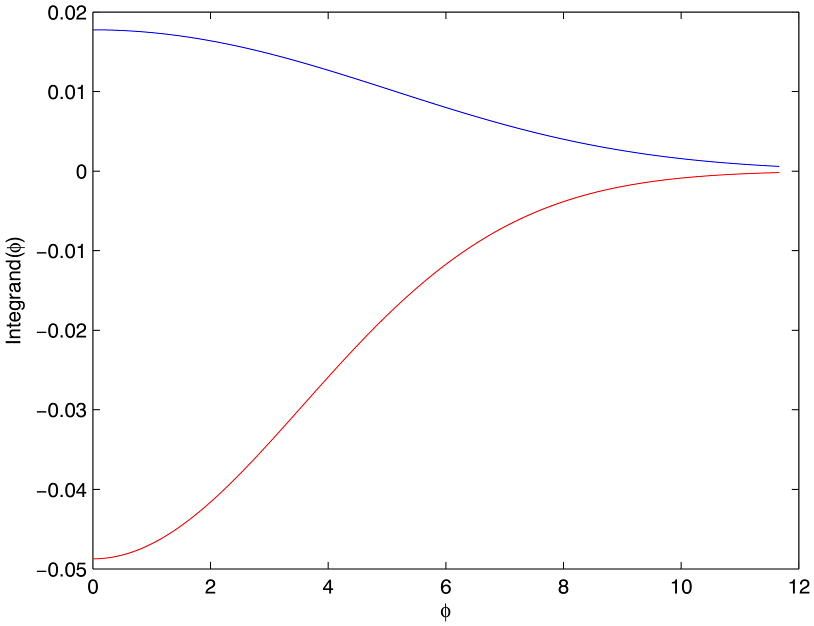
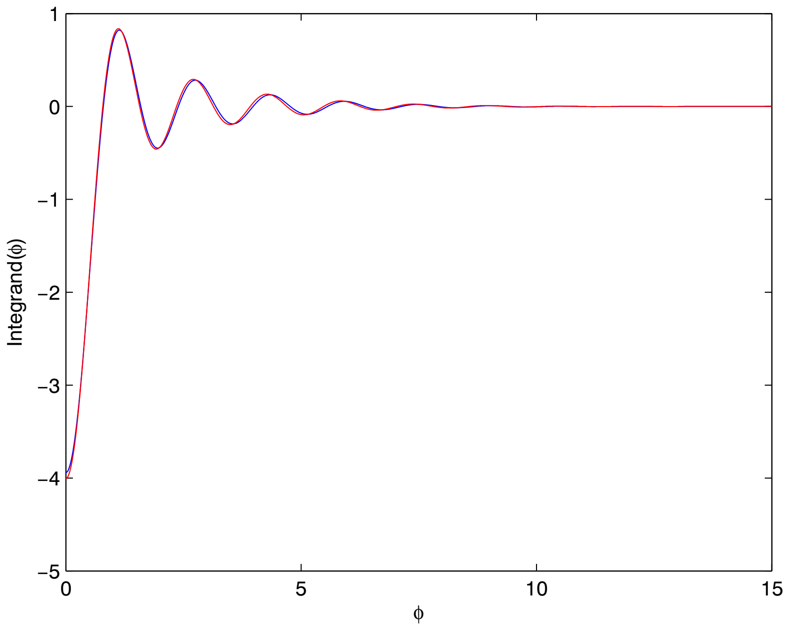


Figure 2.5.: Example of Oscillating Integrands of Formula 2.32



- an antithetic variate approach
- a control variate approach.

The efficiency of the two approaches is measured by the standard error defined in the following way (see for example Jäckel [Jäc02]) where I abbreviate $F^i\left(\tau, \tilde{\tau}, S(\tau), \delta(\tau)\right)$ with F^i

$$\begin{aligned}
\hat{\sigma}_N &= \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\left(e^{-r\tau}\left(F^i - K\right)^+\right)^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N} e^{-r\tau}\left(F^i - K\right)^+\right)^2} \\
\epsilon_N &= \frac{\hat{\sigma}_N}{\sqrt{N}}
\end{aligned}$$

So the standard error is defined by the realized standard deviation of the Monte Carlo simulation divided by the square root of the number of iterations. The smaller the standard error of the calculation, the more accurate will be the result in general.

To illustrate how I applied the **antithetic variate** approach I consider as an example the convenience yield δ. The antithetic variate approach simulates

$$\begin{aligned}
d\delta_1(t) &= \left(\theta_\delta - \kappa_\delta \delta_1(t)\right) dt + \sigma_\delta dW_\delta \quad \text{and} \\
d\delta_2(t) &= \left(\theta_\delta - \kappa_\delta \delta_2(t)\right) dt - \sigma_\delta dW_\delta
\end{aligned}$$

In the approach for each simulation $0 \leq i \leq N$ the end values of $S_1^i(\tau), S_2^i(\tau)$ and $\delta_1^i(\tau), \delta_2^i(\tau)$ at maturity of the option τ are then plugged in the futures formula and one obtains two values for call prices for each simulation, namely

$$\begin{aligned}
v_1(i) &:= e^{-r\tau}\left(F^i(\tau, \tilde{\tau}, S_1^i(\tau), \delta_1^i(\tau)) - K\right)^+ \quad \text{and} \\
v_2(i) &:= e^{-r\tau}\left(F^i(\tau, \tilde{\tau}, S_2^i(\tau), \delta_2^i(\tau)) - K\right)^+
\end{aligned}$$

In the antithetic variance approach I have only to count the pairwise average (Jäckel [Jäc02]) so that the Monte Carlo estimator is then given by

$$MC_a = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{2}v_1(i) + \frac{1}{2}v_2(i)\right)$$

Additionaly, I applied a **control variate**, namely the futures price given by formula 2.9 today with maturity equal to the maturity of the option $g^* := F(0, \tau, S(0), \delta(0))$. Let us remember that the futures price is equal to the expected spot value, that is, $F(0, \tau, S(0), \delta(0)) = \mathbb{E}[S(\tau)]$. The following holds for all $\beta \in \mathbb{R}$

$$\mathbb{E}[MC] = \mathbb{E}\left[MC + \beta\left(g^* - \frac{1}{N}\sum_{i=1}^{N} S^i(\tau)\right)\right]$$

The ordinary Monte Carlo estimator MC 2.37 is then replaced by

$$MC_{CV} = MC + \beta\left(g^* - \frac{1}{N}\sum_{i=1}^{N} S^i(\tau)\right) \tag{2.38}$$

and the Monte Carlo error then reads as

$$\hat{\sigma}_N^{cv} = \sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\left(e^{-r\tau}\left(F^i - K\right)^+ + g^* - S^i(\tau)\right)^2\right) - \left(\frac{1}{N}\sum_{i=1}^{N}\left(e^{-r\tau}\left(F^i - K\right)^+ + g^* - S^i(\tau)\right)\right)^2}$$

$$\epsilon_N^{cv} = \frac{\hat{\sigma}_N^{cv}}{\sqrt{N}}$$

The optimal choice of β is

$$\beta^* = \frac{\mathrm{Cov}\left[e^{-r\tau}\left(F\left(\tau, \tilde{\tau}, e^{-r\tau}S(\tau), \delta(\tau)\right) - K\right)^+, S(\tau)\right]}{\mathbb{V}[S(\tau)]}$$

which minimizes the variance of MC_{CV} (Jäckel [Jäc02]). An intuitive understanding of the control variate method is to observe that $e^{-r\tau}\left(F\left(\tau, \tilde{\tau}, S(\tau), \delta(\tau)\right) - K\right)^+$ and $S(\tau)$ are positively correlated. For any draw of the futures price that overestimates the result, the spot value is likely to overestimate g^*. As a result, the term multiplied by β in equation 2.38 is likely to correct the result by subtracting the aberration. The value of β^* is estimated from the same simulation that is used to calculate MC_{CV}: With $X_1^i := e^{-r\tau}\left(F^i\left(\tau, \tilde{\tau}, e^{-r\tau}S(\tau), \delta(\tau)\right) - K\right)^+$ denoting the ith simulated call price and $X_2^i := S^i(\tau)$ the ith simulated discounted spot index value β^* can be estimated via

$$\beta^* = \frac{\sum_{i=1}^{N}\left(X_1^i - \frac{1}{N}\sum_{i=1}^{N} X_1^i\right) * \left(X_2^i - \frac{1}{N}\sum_{i=1}^{N} X_2^i\right)}{\sum_{i=1}^{N}\left(X_2^i - \frac{1}{N}\sum_{i=1}^{N} X_2^i\right)^2}$$

After introducing and explaining the different variance reduction approaches for Monte Carlo simulations it is now interesting to quantify the amount of variance reduction in our case. To this end, I present table 2.2. One can see that for all three parameter sets the combination of the

Table 2.1.: Parameter Sets for Tables 2.3, 2.4 and 2.5

Parameters		Table 2.4	Table 2.5	Table 2.3
Initial Spot Value	$S(0)$	100	100	100
Interest Rate	r	0.03	0.03	0.03
Maturity Underlying Future	$\tilde{\tau}$	1.0	$\tilde{\tau} \in [0.7, 2.0]$	1.0
Maturity Option	τ	$\tau \in [0.3, 0.9]$	0.6	0.6
Strike	K	100	100	$K \in [40, 160]$
Volatility Spot	σ_S	0.1	0.1	0.1
Speed of Mean Reversion Conv. Y.	κ_δ	0.5	0.5	0.5
Volatility Conv. Y.	σ_δ	0.1	0.1	0.1
Mean Con. Y.$*\kappa_\delta$	θ_δ	0.025	0.025	0.025
Speed of Mean Reversion Volatility	κ_V	0.33	0.33	0.33
Volatility of Volatility	σ_V	0.1	0.1	0.1
Mean Volatility$*\kappa_V$	θ_V	0.03	0.03	0.03
Correlation Spot and Conv. Y.	ρ_1	0.0	0.0	0.0
Correlation Spot and Volatility	ρ_2	-0.3	-0.3	-0.3
Initial Value Conv. Y.	$\delta(0)$	0.05	0.05	0.05
Initial Value Volatility	$V(0)$	0.1	0.1	0.1

Table 2.2.: Comparison of Monte Carlo Errors w.r.t. different Variance Reduction Methods 100,000 Simulations

	Parameter 1	Parameter 2	Parameter 3
Without Variance Reduction	0.0512	0.0234	0.0239
Antithetic	0.0298	0.0135	0.0139
Control Variate	0.0246	0.0130	0.0130
Antithetic and Control Variate	0.0122	0.0071	0.0065

antithetic and control variate reduction yields the best Monte Carlo error, namely for example 0.0065 for parameter set 3 in contrast to an error of 0.0239 without any variance reduction. In all sets antithetic and control variate approach alone can halve the error without variance reduction. Via combination of antithetic and control variate this error is again divided by two. Parameter set 1 encompasses the parameters as for table 2.4 (see table 2.1) with $\tau = 0.6$. Parameter set 2 is the same as parameter set 1 but with another option maturity $\tau = 0.15$; parameter set 3 is the the same as parameter set 2 but with a modified spot-volatility-correlation $\rho_2 = 0.3$.

In the following, I shall illustrate and visualize the dependence of the semi-closed form solution for the call price on futures with respect to different parameters. Furthermore, I compare these prices to the simulated prices. The Monte Carlo simulation is conducted with the inclusion of two variance reduction methods, namely the above-mentioned combination of the antithetic approach and control variate. For each figure I used $100,000$ simulations for each of the 400 $(20 * 20)$ simulated prices.

In plots 2.8 to 2.11 I do not present the call prices but the implied Black volatilities. To explain how one maps the futures call price obtained by the semi-closed form solution 2.32 to the corresponding implied Black volatility I first have to recall the famous Black's formula for futures call prices:

Proposition 2.4.2. *Let $F(t,\tilde{\tau})$ denote the the futures price at time t with maturity $\tilde{\tau}$. Assume that $F(t,\tilde{\tau})$ follows the stochastic differential equation*

$$dF(t,\tilde{\tau}) = F(t,\tilde{\tau})\sigma dW(t)$$

where $\sigma > 0$ denotes the volatility of the futures contract and W is a standard Brownian motion. Under this condition, the values of the futures call $C(t,\tau)$ and put $P(t,\tau)$ with maturity $\tau \leq \tilde{\tau}$ are given by formulas

$$\begin{aligned} C^{Bl}(t,\tau) &= e^{-r\tau}\left(F(t,\tilde{\tau})\Phi(d_1) - K\Phi(d_2)\right) \\ P^{Bl}(t,\tau) &= e^{-r\tau}\left(K\Phi(d_1) - F(t,\tilde{\tau})\Phi(d_2)\right) \end{aligned}$$

where $r > 0$ is the risk-free interest rate and

$$\begin{aligned} d_1 &= \frac{\ln(F(t,\tilde{\tau})/K) + (\sigma^2/2)\tau}{\sigma\sqrt{\tau}} \\ d_2 &= \frac{\ln(F(t,\tilde{\tau})/K) - (\sigma^2/2)\tau}{\sigma\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau} \end{aligned}$$

Φ marks the cumulative distribution function of the standard normal distribution.

Proof: see Black [Bla76] . ■

The *implied volatility* for a given futures call price is defined as the volatility σ_0 in the Black formula such that the Black call price C^{Bl} and the Heston call price C^H coincide, that is,

$$C^H \stackrel{!}{=} C^{Bl}(\sigma_0)$$

In the plots below I set $F(0,\tau) = S(0) = 100$.

Table 2.3.: Dependence on Strike K

K	Closed Formula	Monte Carlo	price difference
40	57.0942	57.0993	0.0051
50	47.3221	47.3130	-0.0091
60	37.6990	37.7117	0.0128
70	28.6591	28.6618	0.0027
80	20.6629	20.6616	-0.0013
90	14.0819	14.0860	0.0041
100	9.1031	9.1383	0.0352
110	5.6367	5.6569	0.0202
120	3.3730	3.3580	-0.0150
130	1.9589	1.9343	-0.0247
140	1.1005	1.0829	-0.0176
150	0.5920	0.5989	0.0069
160	0.3023	0.3119	0.0096

Table 2.4.: Dependence on Option Maturity τ

τ	Closed Formula	Monte Carlo	price difference
0.30	6.2365	6.3315	0.0950
0.35	6.8059	6.8777	0.0718
0.40	7.3303	7.3901	0.0598
0.45	7.8172	7.8672	0.0501
0.50	8.2720	8.3122	0.0401
0.55	8.6995	8.7384	0.0389
0.60	9.1031	9.1460	0.0430
0.65	9.4858	9.5265	0.0406
0.70	9.8501	9.8657	0.0156
0.75	10.1981	10.2086	0.0106
0.80	10.5313	10.5547	0.0233
0.85	10.8514	10.8503	-0.0011
0.90	11.1595	11.1678	0.0083

Table 2.5.: Dependence on Underlying Futures Maturity $\tilde{\tau}$

$\tilde{\tau}$	Closed Formula	Monte Carlo	price difference
0.70	9.2776	9.2991	0.0216
0.80	9.2127	9.2459	0.0332
0.90	9.1548	9.1870	0.0322
1.00	9.1031	9.1296	0.0266
1.10	9.0569	9.0553	-0.0016
1.20	9.0156	9.0415	0.0258
1.30	8.9788	9.0073	0.0285
1.40	8.9459	8.9859	0.0401
1.50	8.9165	8.9209	0.0044
1.60	8.8901	8.9089	0.0188
1.70	8.8666	8.8800	0.0134
1.80	8.8455	8.8720	0.0265
1.90	8.8266	8.8888	0.0621
2.00	8.8097	8.8375	0.0278

In the first figure 2.6 I varied strike price K and the option maturity τ, in the second one 2.8 the interest rate r and the initial value of the convenience yield δ_0 and the third one 2.10 is concerned with the option and futures maturities τ and $\tilde{\tau}$.

The surfaces represent the prices/ implied volatilities obtained via the formula and the black dots mark the prices/ implied volatilities obtained by the Monte Carlo simulation.
The histograms illustrate the empirical distribution of the differences of the Black implied volatilities with respect to the Monte Carlo prices and the closed form prices. The red line shows a normal distribution fitted to the differences having their empirical mean and variance.

The chosen parameters for the figures can be looked up in column *Table 2.4* of table 2.1 with option maturity $\tau = 0.8$. Note that the mean value for the convenience yield is given by $m_\delta := \frac{\theta_\delta}{\kappa_\delta} = 0.05$ and the one for the mean volatility by $m_V := \frac{\theta_V}{\kappa_V} = 0.1$. Since the interest rate r is equal to 0.03 we have a setting in backwardation: $r - m_\delta < 0$. The volatility of the instantaneous spot index return $dS(t)/S(t)$ is given by $\sqrt{\sigma_S^2 + V(t)} \approx \sqrt{\sigma_S^2 + m_V} \approx 0.33$.

After the illustration of the simultaneous dependence of the call price on some pairs of parameters, I then varied only one parameter at a time. Tables 2.4, 2.5 and 2.3 display the dependence of European options on futures with respect to the maturity of the option τ, maturity of the underlying future $\tilde{\tau}$ and strike price K. The parameter sets for tables 2.4, 2.5 and 2.3 are given in table 2.1.

All the figures and tables reveal the following futures call price features with respect to the chosen parameters:

1. With increasing strike price K decreases the value of the price (see table 2.3 and figure 2.6).
2. With increasing option maturity τ increases the value of the price/ Black implied volatility (see table 2.4, figure 2.6 and figure 2.10).
3. With increasing futures maturity $\tilde{\tau}$ decreases the value of the price/ Black implied volatility (see table 2.5 and figure 2.10).
4. The prices are the highest in the case of high interest rates combined with very low (negative) initial values for the convenience yield δ_0 (see figure 2.8).

Figure 2.6.: Dependence of European Call Price on K and τ

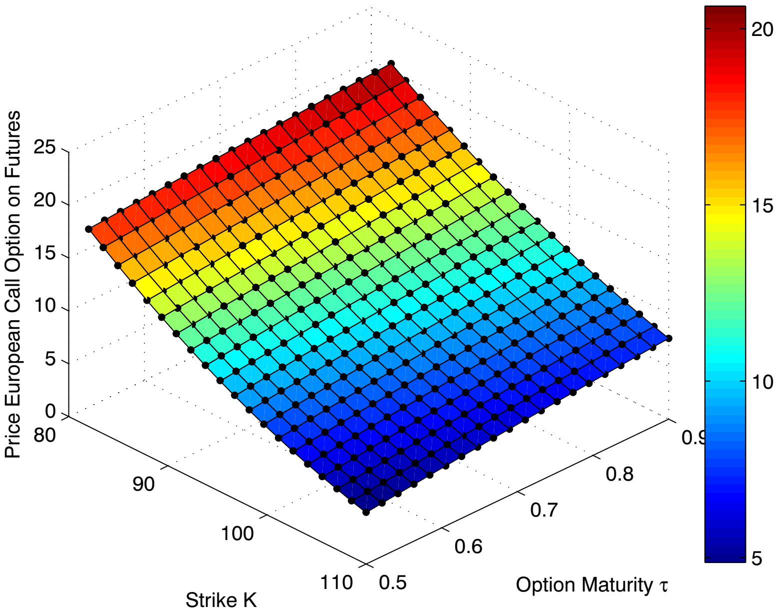


Figure 2.7.: Error between Semi-Closed Form Solution and MC

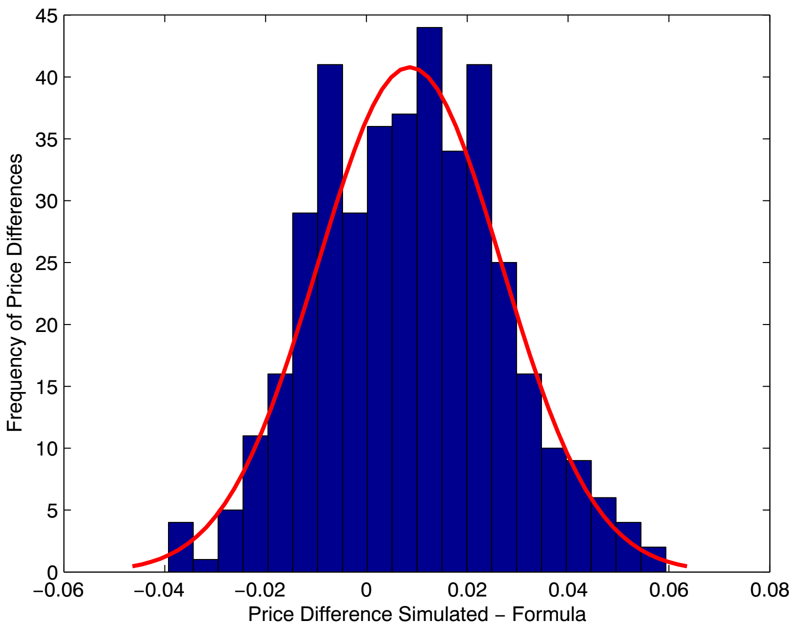


Figure 2.8.: Dependence of Implied Volatility of European Options on Futures on r and δ_0

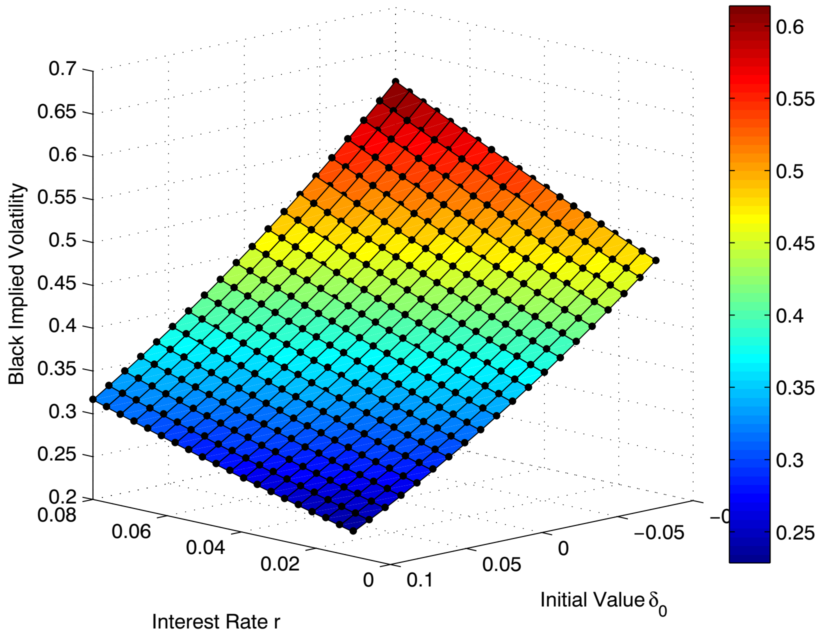


Figure 2.9.: Error between Semi-Closed Form Solution and MC

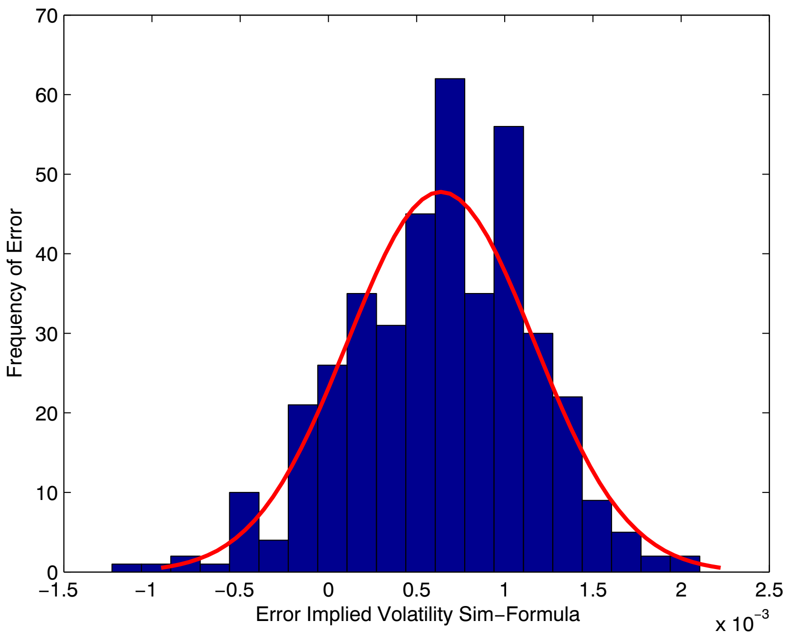


Figure 2.10.: Dependence of Implied Volatility of European Options on Futures on $\tilde{\tau}$ and τ

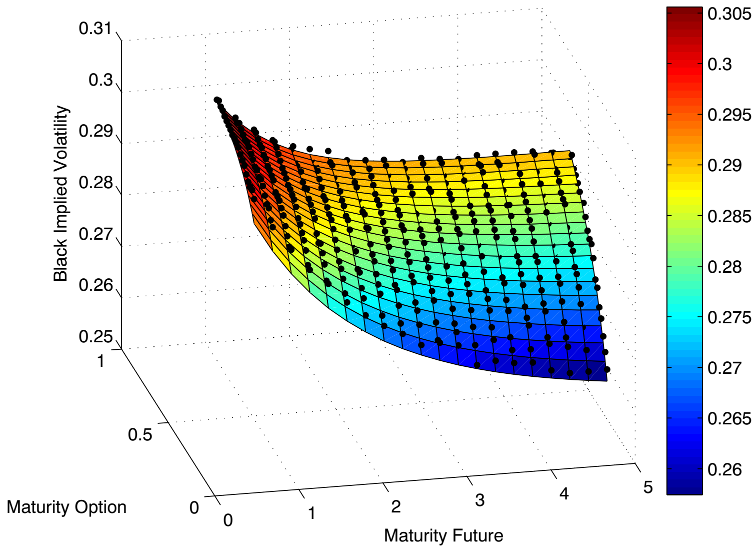
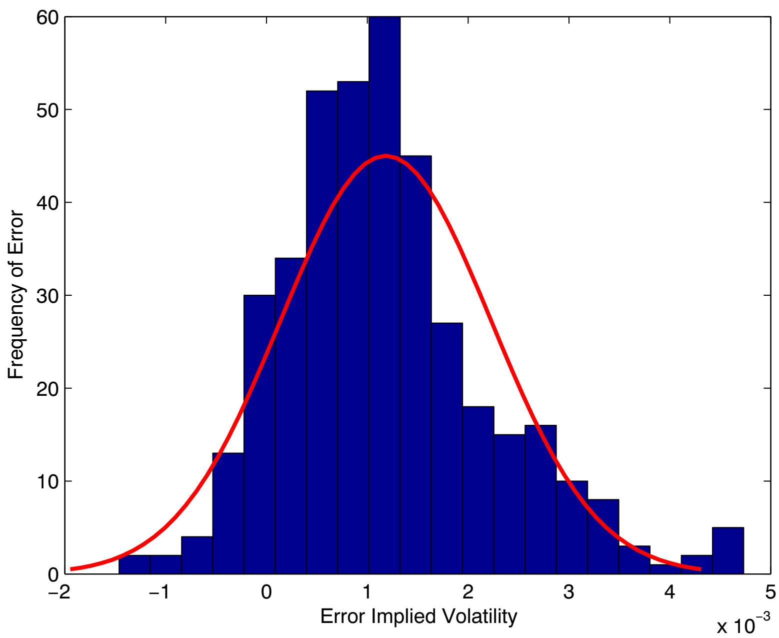


Figure 2.11.: Error between Semi-Closed Form Solution and MC in dependence of maturities $\tilde{\tau}$ and τ



Chapter 3

Construction of Optimal Commodity Indices and Portfolios

3.1. Optimal Rolling of Commodity Futures

Let us recall the rolling procedure of the Dow Jones-UBS Commodity Index as mentioned in section 1.2. Rolling was necessary to maintain a long futures position. To this end, nearby contracts (lead futures) are sold and futures that have not yet reached the delivery period (next futures) are bought. The roll period for all commodities of the index is the period of five business days, beginning with the sixth business day through and including the tenth business day of each month. In this time the value of the Dow Jones-UBS Commodity Index is gradually shifted from reliance on the basket of lead futures to the basket of next futures. But is this rolling period optimal? Optimal rolling days are those days where the spread between the next futures prices and the lead futures prices is as small as possible because then the loss associated with the rolling is minimized.

We now show that we have an indifference of the rolling with respect to different rolling dates *under the assumption that futures prices are martingales*. But then we will explore that in practice there is indeed a difference.

At time t_0 we buy a futures with maturity $t_1 > t_0$, at time t_1 we sell it and invest the money in a futures with maturity $t_2 > t_1$. Figure 3.1 illustrates the investment.

We calculate the expected value of the investment using only that the commodity futures are

Figure 3.1.: Two Periods Rolling Strategy

t_0	t_1	t_2
$-F(t_0, t_1)$	$+F(t_1, t_1)$	
	$-F(t_1, t_2)$	$F(t_2, t_2)$

martingales

$$\begin{aligned}\mathbb{E}\left[-F(t_0,t_1)+F(t_1,t_1)-F(t_1,t_2)\right] &= -F(t_0,t_1)+F(t_0,t_1)-F(t_0,t_2)\\ &= -F(t_0,t_2)\end{aligned}$$

Which shows that the above mentioned rolling is equal in the mean to buying at time t_0 directly a futures with maturity t_2. In reality, due to asymmetry of the supply and demand shortly before the maturity of the futures, their prices may strongly rise and thus violate the assumptions we have made for their distributions.

Müller [Mül05] reveals that for some commodities, for example, Brent and Gasoil, in the analyzed time period from 01/01/1995 to 08/20/2004 optimal rolling days can be determined and those days differ with that of the Dow Jones-UBS Commodity Index. The energy products analyzed in Müller [Mül05] are Brent and Gasoil traded on the International Petroleum Exchange based in London. The contract size of Brent is $1,000$ barrel and of Gasoil 100 metric tons. The contract price is in US dollars and cents per barrel. The International Petroleum Exchange is one of the world's largest energy futures and options exchanges. Its flagship commodity, Brent Crude is a world benchmark for oil prices, but the exchange also handles futures contracts and options on gas oil, natural gas, electricity, coal contracts and carbon emission allowances with the European Climate Exchange.

The industry metals are Primary Aluminum and Grade A Copper both traded on the London Metal Exchange. Both metals have the same contract size, namely 25 metric tons. The contract price is in US dollars and cents per tonne. The London Metal Exchange is the major international market for the main industrially-used non-ferrous metals, namely aluminum, aluminum alloy, copper, lead, nickel, tin, and zinc. Aluminum has the highest volume of spot and futures trade on the exchange, followed closely by copper. The two metals are also amongst the most important metals in an industrial sense.

The ingredients for the optimization of rolling days are historical futures prices with two different maturities at every time point. The lead futures expires in the midst of the current month and the other one expires in the midst of the next month. If the lead futures is terminated it is substituted by the next futures price and the next futures price is the one with expiration one month later. In table 3.1 we have extracted the data for Brent to give an illustration: We see that on January, the 17th 1995 the contract day is equal to one. This means that the lead futures has expired on January, the 16th and has to be substituted by the next future. The best rolling day was February, the 13th, because on this day the difference between 1st and 2nd is minimized. Remember that the lead futures (2nd) are sold and the next futures (1st) are bought. So the loss is equal to the spread.

Let us denote the lead futures prices with F^{2nd} and the next futures F^{1st}. In Müller [Mül05] the spreads $\Upsilon = F^{1st} - F^{2nd}$ for two times series of futures are considered. Given the observations f_{it}^{1st} and f_{it}^{2nd} $(i = 1, \ldots, n_t)$ for each contract day $t = 1, \ldots, 20$, one can calculate the spreads for each contract day t by

$$s_{it} = f_{it}^{1st} - f_{it}^{2nd}$$

The optimal time points for switching are to be determined by finding the contract days t with smallest spreads Υ_t. We are therefore interested to compare spreads Υ_{t_1} and Υ_{t_2} for any pair of different contract days t_1 and t_2. This leads to the test problem

$$H_0 : \Upsilon_{t_1} \succeq \Upsilon_{t_2} \quad \text{versus} \quad H_1 : \Upsilon_{t_1} \prec \Upsilon_{t_2}$$

where $\Upsilon_{t_1} \prec \Upsilon_{t_2}$ is used to indicate that the realizations of Υ_{t_2} are typically (in a stochastic sense) larger than those of Υ_{t_1}. If we denote the cumulative distribution functions of Υ_{t_1} and Υ_{t_2} by H_{t_1} and H_{t_2}, respectively, this gives

$$\Upsilon_{t_1} \prec \Upsilon_{t_2} \quad \Longleftrightarrow \quad H_{t_1}(x) > H_{t_2}(x) \quad \text{for all } x,$$

that is, H_{t_1} stochastically dominates H_{t_2}.

In Müller [Mül05] two statistical tests are applied, the t-test and the Wilcoxon Signed Rank test.

Table 3.1.: Extraction of Analyzed Time Series

Date	IPE Brent Contract Day	Brent 2nd	Brent 1st	Spread 1st – 2nd
16-Jan-95	21	16,37	16,43	0,06
17-Jan-95	1	16,82	16,64	-0,18
18-Jan-95	2	16,93	16,75	-0,18
19-Jan-95	3	16,8	16,61	-0,19
20-Jan-95	4	16,87	16,66	-0,21
23-Jan-95	5	16,64	16,47	-0,17
24-Jan-95	6	16,86	16,72	-0,14
25-Jan-95	7	16,79	16,63	-0,16
26-Jan-95	8	16,67	16,57	-0,1
27-Jan-95	9	16,39	16,34	-0,05
30-Jan-95	10	16,57	16,44	-0,13
31-Jan-95	11	16,8	16,64	-0,16
01-Feb-95	12	16,88	16,67	-0,21
02-Feb-95	13	16,89	16,67	-0,22
03-Feb-95	14	17,2	16,78	-0,42
06-Feb-95	15	17,22	16,72	-0,5
07-Feb-95	16	17,07	16,56	-0,51
08-Feb-95	17	17,07	16,48	-0,59
09-Feb-95	18	17,11	16,59	-0,52
10-Feb-95	19	17,33	16,7	-0,63
13-Feb-95	20	17,42	16,73	-0,69
14-Feb-95	1	16,75	16,62	-0,13
15-Feb-95	2	16,76	16,61	-0,15

The t-test assumes the distributions of Υ_{t_1} and Υ_{t_2} to be normal. We moreover assume that we have paired observations such that we consider

$$d_i = s_{it_1} - s_{it_2}, \quad i = 1, \ldots, n$$

as observations of $D = \Upsilon_{t_1} - \Upsilon_{t_2} \sim N(\Delta, \sigma_D^2)$. Thus, the test problem reduces to

$$H_0 : \Delta \geq 0 \quad \text{versus} \quad H_1 : \Delta < 0$$

and the test statistic is given by

$$t = \sqrt{n} \cdot \frac{\overline{D}}{s_D} \sim t_{n-1}, \quad \text{where} \quad \overline{D} = \sum_{i=1}^{n} d_i, \quad s_D = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (d_i - \overline{D})^2}$$

The hypothesis is rejected if

$$t < -t_{n-1,1-\alpha}$$

where α denotes the significance level and $t_{n-1,1-\alpha}$ the $(1-\alpha)$-quantile of the t_{n-1} distribution. As n gets large, the $t_{n-1,1-\alpha}$ can be approximated by the $(1-\alpha)$-quantile of the standard normal.

Furthermore a distribution-free test, namely the Wilcoxon signed rank test is applied. It is based on paired observations where the test problem reads as

$$H_0 : H_{t_1}(x) \leq H_{t_2}(x) \quad \text{for all } x \quad \text{versus} \quad H_1 : H_{t_1}(x) > H_{t_2}(x) \quad \text{for all } x$$

The test statistic is calculated from the ranks

$$r_i = \text{rank}(d_i) = \text{rank}(s_{it_1} - s_{it_2}), \quad i = 1, \ldots, n$$

where the rank denotes the position of the observation d_i in the ordered sample. Define

$$W_+ = \sum_{d_i > 0} r_i \,, \quad W_- = \sum_{d_i < 0} r_i \,,$$

and the test statistic by

$$W = \min\{W_+, W_-\}$$

There are tables for the exact distribution of W in case that $n \leq 30$. For larger values of n, the distribution of W is normal (under H_0) with expectation $\frac{n(n+1)}{2}$ and variance $\frac{n(2n+1)(n+1)}{24}$.

In Müller [Mül05] the following results for the optimal contract days are:

- Brent:

  Particular small p-values are found for both tests starting from contract days $i = 3$ to $i = 7$ and $j > 8$. This means that smaller spreads can be found at the contract days around 3 to 7 (corresponds approximately to the 18th to 25th calendar day of a month).

- Gasoil:

  For both tests small p-values can be found for contract days $i = 1, \ldots, 4$ and $j > 3$. This means that the smallest spreads can be found at the very first days (corresponds approximately to the 15th to 19th calendar day of a month).

- Aluminum and Copper:

  For these two commodities we hardly find p-values smaller than 5% or 10%. For this reason we conclude that all contract days give nearly similar spreads.

3.2. Markowitz Optimization

Many commodities, like those in the energy and industrial metals sectors, have liquid futures contracts that expire every month. Therefore, these commodities can be rolled forward every month. In the last chapter we discovered the best rolling days for each of the four commodities. I now fix these optimal days:

- Brent: 3rd, 4th and 5th contract day
- Gasoil, Copper, Aluminum: 2nd, 3rd and 4th contract day

In addition of the two energy products and the two industry metals I also consider the precious metal gold. It is traded on the London Bullion Market Association with a contract size of one ounce. In contrast to the other four commodities Gold has a liquid spot market. So for Gold no futures prices are analyzed but the London PM fix. The London PM fix is normally considered the main reference price for the day and is the price most often used in contracts. The price of gold is quoted in US dollars and cents per troy ounce.

After having fixed the rolling days I now also fix for each rolling day the proportions of lead futures being substituted by next futures. With $V_i(0) > 0$ invested in the lead futures of the $i-$th commodity the rolling strategy $V_i(t), t > 0$ then takes the following form: The amount $V_i(t)$ is fully invested in the lead futures of the i-th commodity until the first rolling day is reached. On the first rolling day 33% of the money invested in the lead futures are sold. The money is then immediately reinvested in the next futures. The roll process continues the following two rolling days. The basket of lead futures is gradually shifted to a basket of next futures in the same fashion as on the first day: 50% on the second day, 100% on the third and last day. At the end of the rolling period one has sold all the lead futures and is fully invested in the next futures which are then called the lead futures. A second nearby basket of futures is formed (with futures maturities further in the future) for use in the nest month's roll. So the strategy is a self-financing one since after the initial endowment no further money flows in or out.

After having determined the optimal weightings of commodity futures (*optimal* in the sense as shall be described below) the whole initial money V is allocated to the 5 commodities $V_1, \ldots, \mathrm{V}_5$ corresponding to the weightings. But the fluctuating futures prices and the rolling procedure causes that the actual weightings of the respective commodities in the index change and are not equal to the one-period Markowitz weightings after a while. To maintain the Markowitz weightings I apply in addition a rebalancing mechanism. Every month on the third day of month (or the next trading day where the exchange is open) the actual weightings of the commodities are compared to the Markowitz ones and adjusted. That means that some futures are sold and the money is invested in the ones where the weighting was too small.

I also take into account the trading costs for selling and buying of futures in the roll period and for the rebalancing of the portfolio. The transaction costs are set to the following values

- International Petroleum Exchange (IPE) and London Metal Exchange (LME): 0.3 percent per barrel/ per tonne which is sold or bought
- London Bullion Market Association (LBMA): 5 percent per ounce which is sold or bought

Note that the transaction costs are tracked in separate time series. That means that all the results and plots are without friction losses.

The goal of this section is the construction of an optimal index in the sense that I want to find the optimal weightings of the five commodities in the index via the mean-variance approach (or often called Markowitz optimization [Mar52]). This approach is described for example in section 1.2 in Korn [Kor97]. It is a one period model. The crucial observation is that a pure maximization of expected return would lead to putting all of the money in the asset with the highest expected return. As such a strategy represents a highly risky position Markowitz suggested to also quantify the risk by the variance of the position. Furthermore, he recommended to consider in addition to the single return and single risk of an asset, respectively, the covariance risk of all assets together. The optimization consists of maximizing the index return R^π over all weightings $\pi = (\pi_1, \pi_2, \ldots, \pi_5)$ with $\pi_i \in [0,1]$ for all $i = 1, 2, \ldots, 5$ and $\sum_{i=1}^5 \pi_i = 1$ which fulfill that the index variance is not greater than a given boundary B:

$$\max_\pi \mathbb{E}\left[R^\pi\right] \text{ with constraint } \mathbb{V}\left[R^\pi\right] \leq B$$

The index return is the sum of the 5 weighted rolling strategies returns $R_i(T) = \frac{V_i(T)}{V_i(0)}$

$$R^\pi = \sum_{i=1}^{5} \pi_i R_i(T)$$

Thus

$$\mathbb{E}\left[R^\pi\right] = \sum_{i=1}^{5} \pi_i \mathbb{E}\left[R_i(T)\right] \tag{3.1}$$

The variance of the index reads as

$$\mathbb{V}\left[R^\pi\right] = \mathbb{V}\left[\sum_{i=1}^{5} \pi_i R_i(T)\right] = \sum_{i,j=1}^{5} \pi_i \pi_j \operatorname{Cov}\left[R_i(T), R_j(T)\right] \tag{3.2}$$

I now dedicate myself to the estimation of the expectations 3.1 and covariances 3.2. The applied model for the four rolling strategies and the Gold PM fixing prices V_i is the Black-Scholes model

$$\frac{dV_i(t)}{V_i(0)} = \mu_i dt + \sigma_i dW_i(t), \; V_i(0) > 0 \tag{3.3}$$

where $W_i(t)$ with $i = 1, \ldots, 5$ are correlated Brownian motions with correlations $\rho_{ij} \in [-1, 1]$.

For the calculation and estimation of the expectations and covariances appearing in the Markowitz optimization the following proposition will be useful.

Proposition 3.2.1. *Let M be a positive random variable where the logarithm $\ln M$ is normally distributed with mean α_1 and standard deviation ν_1. Then it holds for all positive integer values n*

$$\mathbb{E}\left[M^n\right] = \exp\left(n\alpha_1 + 0.5n^2\nu_1^2\right)$$

Especially, the mean and variance are given by

$$\begin{aligned}
\mathbb{E}\left[M\right] &= \exp\left(\alpha_1 + 0.5\nu_1^2\right) \\
\mathbb{V}\left[M\right] &= \exp\left(2\alpha_1 + \nu_1^2\right)\left(e^{\nu_1^2} - 1\right) = \mathbb{E}\left[M\right]^2\left(e^{\nu_1^2} - 1\right)
\end{aligned}$$

With N being an additional lognormal random variable with mean α_2 and standard variation ν_2 then the product $M * N$ is lognormally distributed, too. The mean of $\ln(M * N)$ is equal to $\alpha_1 + \alpha_2$ and the variance is $\mathbb{V}[M] + \mathbb{V}[N] + 2\,\mathrm{Cov}[M, N]$. So it follows that

$$\mathbb{E}[NM] = \exp(\alpha_3 + 0.5\nu_3) \tag{3.4}$$

with $\alpha_3 = \alpha_1 + \alpha_2$ and $\nu_3 = \nu_1^2 + \nu_2^2 + 2\,\mathrm{Cov}[M, N]$.

Proof: Is given in Aitchison and Brown [AB57], page 8 and theorem 2.4 on page 12. ■

Corollary 3.2.1. *With $V_i(t)$ $t \geq 0$ being specified by 3.3 for every $i = 1, \ldots, 5$ the expectations, variances and covariances of the returns $R_i(T) = \frac{V_i(T)}{V_i(0)}$ appearing in the Markowitz optimization are given by*

$$\begin{aligned}
\mathbb{E}[R_i(T)] &= \exp(\mu_i T) \\
\mathbb{V}[R_i(T)] &= \mathbb{E}[R_i(T)]^2 \left(\exp\left(\sigma_i^2 T\right) - 1\right) \\
\mathrm{Cov}[R_i(T), R_j(T)] &= \mathbb{E}[R_i(T)]\,\mathbb{E}[R_j(T)]\,(\exp(\rho_{ij}\sigma_i\sigma_j T) - 1)
\end{aligned}$$

where the investment horizon in years is denoted by T.

Proof: The model 3.3 can be written as $\frac{V_i(t)}{V_i(0)} = \exp\left\{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i dW_i(t)\right\}$. This is equivalent to $\ln\frac{V_i(t)}{V_i(0)} = \left\{\left(\mu_i - \frac{1}{2}\sigma_i^2\right)t + \sigma_i dW_i(t)\right\}$. Thus the log-returns $\ln\frac{V_i(t+\Delta t)}{V_i(t)}$ are normally distributed with mean $\left(\mu_i - \frac{1}{2}\sigma_i^2\right)\Delta t$ and volatility $\sigma_i^2\Delta t$. Since the Brownian increments are independent also the log-returns are independent. So for all rolling strategies V_i the returns $R_i(T) = V_i(T)/V_i(0)$ are lognormally distributed. Via proposition 3.2.1 I can calculate

$$\mathbb{E}[R_i(T)] = \exp\left(\mu_i T - 0.5\sigma_i^2 T + 0.5\sigma_i^2 T\right) = \exp(\mu_i T)$$

and

$$\mathbb{V}[R_i(T)] = \mathbb{E}[R_i(T)]^2\left(\exp\left(\sigma_i^2 T\right) - 1\right)$$

With

$$\begin{aligned}
\alpha &= \mu_i T - 0.5\sigma_i^2 T + \mu_j T - 0.5\sigma_j^2 T \\
\nu &= \sigma_i^2 T + \sigma_j^2 T + 2\rho_{ij}\sigma_i\sigma_j T
\end{aligned}$$

the covariances can be written as

$$\begin{aligned}
\mathrm{Cov}\left[R_i(T), R_j(T)\right] &= \mathbb{E}\left[R_i(T)R_j(T)\right] - \mathbb{E}\left[R_i(T)\right]\mathbb{E}\left[R_j(T)\right] \\
&= \exp\left(\alpha + 0.5\nu\right) - \exp\left(\mu_i T + \mu_j T\right) \\
&= \exp\left(\mu_i T + \mu_j T\right)\left(\exp\left(\rho_{ij}\sigma_i\sigma_j T\right) - 1\right)
\end{aligned}$$

■

So I am left with the estimation of the Black-Scholes parameters μ_i, σ_i and ρ_{ij}. To this end, I consider the k-th daily log-return of the i-th rolled commodity $r_{ik} = \ln(V_i(k)/V_i(k-1))$. As above T denotes the investment horizon in years:

$$\begin{aligned}
\overline{m}_i &:= \frac{1}{n}\sum_{k=1}^{n} r_{ik} = \frac{1}{n}\ln\frac{V_i(T)}{V_i(0)} \\
\overline{\mu}_i &:= \frac{n}{T}\overline{m}_i \\
\overline{\sigma}_i^2 &:= \frac{n}{T}\frac{1}{n-1}\sum_{k=1}^{n}\left(r_{ik} - \overline{m}_i\right)^2 \\
\overline{\sigma}_{ij} &:= \frac{n}{T}\frac{1}{n-1}\sum_{k=1}^{n}\left(r_{ik} - \overline{m}_i\right)\left(r_{jk} - \overline{m}_j\right) \\
\overline{\rho}_{ij} &:= \frac{\overline{\sigma}_{ij}}{\overline{\sigma}_i\overline{\sigma}_j}
\end{aligned}$$

Note that the estimated expectation of the return is simply the observed end value of V_i divided by the initial endowment

$$\mathbb{E}\left[R_i(T)\right] = \exp\left(\mu_i T\right) = \exp\left(\ln\frac{V_i(T)}{V_i(0)}\right) = R_i(T)$$

These are printed in the first row of table 3.2. The annualized yield is calculated via $(R_i(T))^{1/T} - 1$ (printed in the second row of table 3.2). Consult proposition 3.2.2 about different types of annualization which yield different values for the annualized returns.

Since there exist different alternatives of annualizations we have implemented two of them and observed that one of them always yields smaller annualized returns than the second one. Proposition 3.2.2 introduces the two types of annualizations and gives the reasoning for the differences.

Proposition 3.2.2. *Let* $y = (y_1, y_2, ..., y_n)$ *be a sequence of real-valued values greater than*

Table 3.2.: Expected Returns of Commodities as Input for Markowitz Optimization

	Brent	Gasoil	Aluminum	Copper	Gold
01/01/1995 – 08/20/2004	6.05	3.95	0.51	0.94	1.08
Annualized	21%	15%	-7%	-1%	1%

-100%, *interpreted as returns, and* T *the investment horizon in years. Then for the two annualizations* $a_1(y)$ *and* $a_2(y)$

$$a_1(y) := \left(1 + \frac{1}{n}\sum_{i=1}^{n} y_i\right)^{n/T} - 1 \quad \textit{and} \quad a_2(y) := \sqrt[T]{\prod_{i=1}^{n}(1+y_i)} - 1 \tag{3.5}$$

holds

$$a_1(y) \geq a_2(y) \tag{3.6}$$

Proof: We have to compare $\left(1 + \frac{1}{n}\sum_{i=1}^{n} y_i\right)^n$ with $\prod_{i=1}^{n}(1+y_i)$. The first expression can be written as

$$\begin{aligned}\left(1 + \frac{1}{n}\sum_{i=1}^{n} y_i\right)^n &= \exp\left(n \ln\left(1 + \frac{1}{n}\sum_{i=1}^{n} y_i\right)\right) \\ &= \exp\left(n \ln\left(\frac{1}{n}\sum_{i=1}^{n}(1+y_i)\right)\right)\end{aligned}$$

and the second one as

$$\prod_{i=1}^{n}(1+y_i) = \exp\left(\sum_{i=1}^{n}\ln(1+y_i)\right)$$

The logarithm is a concave function so the negative logarithm is convex. So the inequality of Jensen can be applied and yields

$$-n \ln\left(\frac{1}{n}\sum_{i=1}^{n}(1+y_i)\right) \leq n\frac{1}{n}\sum_{i=1}^{n}(-\ln(1+y_i))$$

which is equivalent to

$$a_1(y) \geq a_2(y)$$

■

As in the preceding section the time period of the observed futures data is 01/01/1995 – 08/20/2004 (= 9.63 years). I assume in the following that the commodities are divisible in

Table 3.3.: Covariance Matrix of Commodities as Input for Markowitz Optimization

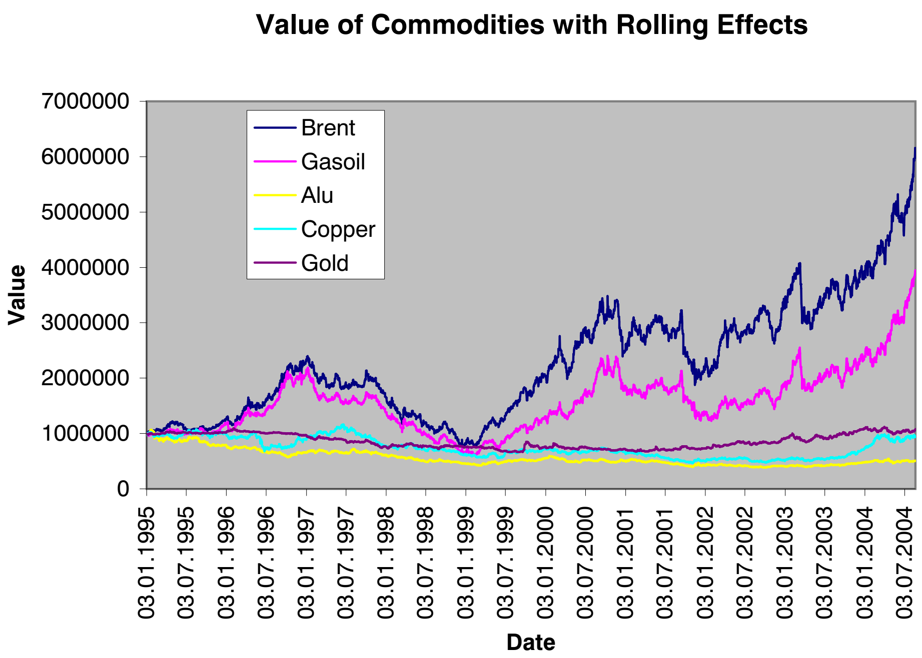
	Brent	Gasoil	Aluminum	Copper	Gold
Brent	63.28	20.52	0.05	0.24	0.11
Gasoil	20.52	22.73	0.04	0.17	0.13
Alu	0.05	0.04	0.08	0.10	0.01
Copper	0.24	0.17	0.10	0.46	0.03
Gold	0.11	0.13	0.01	0.03	0.20

the sense that I assume that one can buy for example half a tonne of Gasoil. Of course this is in practice not possible, since contract sizes have to be complied. The prescription of the data is given on page 62. I have programmed the estimation of the parameters of the Black-Scholes models and input values of the optimization in VBA/EXCEL. For the Markowitz optimization I use EXCEL-SOLVER. It applies the gradient descent algorithm to determine a local maximum of the expected return. I randomized the starting values to get the global maximum.

First I examine the input data of the Markowitz optimization, namely the expectations and covariances which can be found in tables 3.2 and 3.3, respectively. The first row of table 3.2 shows the return as the devision of end value and initial value. The energy products have the highest return/ annualized yield (21%, 15%) followed by Gold (1%). But the variances of Brent and Gasoil in the time period of approximately 10 years is extreme high. The annual yield of the industry metals in the observed time period was negative where the variances are moderate (0.46 for Copper and 0.08 for Aluminum). Since covariances depend on the Black-Scholes parameters given in table 3.4 the parameters should be examined. The correlation of the Brownian motions of the two energy products is equal to 65%. This high correlation is also reflected in the first plot of figure 3.2. The two industry metals are positive correlated, too. Their correlation amounts to 57%. In the observed time period the energy and industry metals are nearly uncorrelated. Gold is nearly uncorrelated to both the energy and industry metals (the correlation is for all commodities smaller than 12%).

The negative expected return of the industry metals combined with the (approximately) zero correlation with the other commodities make it clear why Aluminum and Copper do not appear in the optimal portfolios presented in table 3.5. In this table the variance of the index was set to different values (starting from the lowest value of 1.00 to the highest value 64.00). The results are not surprising: A small value of variance leads to a high weight of the sound gold. If more

Figure 3.2.: Results of Markowitz Optimization



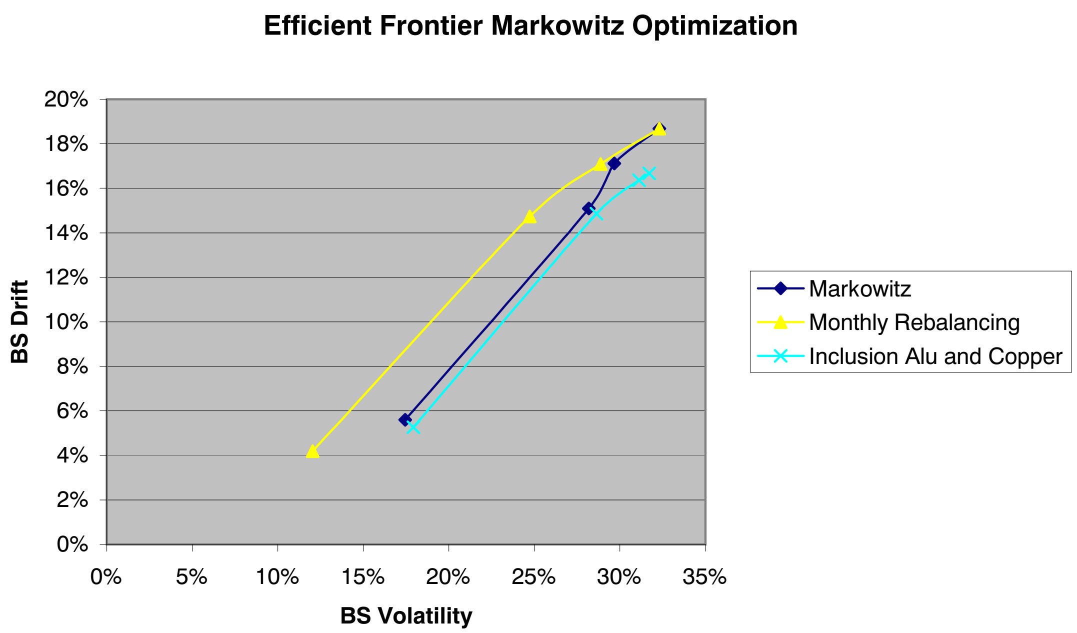


Table 3.4.: Estimated Black-Scholes Parameters: Correlation of Brownian motions ρ, volatilities σ and drift μ

	Brent	Gasoil	Aluminum	Copper	Gold
Volatility	32%	31%	17%	21%	13%
Drift	19%	14%	-7%	-0.6%	0.8%
Correlation	Brent	Gasoil	Aluminum	Copper	Gold
Brent	100%	65%	3%	6%	4%
Gasoil	65%	100%	4%	7%	8%
Alu	3%	4%	100%	57%	11%
Copper	6%	7%	57%	100%	12%
Gold	4%	8%	11%	12%	100%

variance of the index is allowed Gasoil and Brent dominate the index. With the index return and variance I can calculate the Black-Scholes drift and volatility of the index via corollary 3.2.1. I plot these values in figure 3.2 because they are more vivid.

In the second plot of figure 3.2 I compare three different efficient frontiers. The dark blue one results from an investment where the start values of the commodity weightings are equal to the Markowitz ones (see table 3.5). Two other efficient frontiers are also included in figure 3.2, namely the frontier where the index values are rebalanced monthly (yellow line) and another one where I forced that the industry metals are included by at least 10% (light blue one). It is clear that a forced inclusion of Aluminum and Copper yields the worst results since I have included more constraints in the optimization. We see that monthly rebalancing leads to the best results in our case. Due to the fluctuating futures prices and the rolling procedure (backwardation/ contango) the weightings differ with the Markowitz weightings after a while: If $s_{ik} = V_i(k)/V_i(k-1)$ denotes the $k-$th daily return of the $i-$th commodity then the return R is given by

- First Day: $R_1 = \pi_1 s_{11} + \pi_2 s_{21} + ... + \pi_5 s_{51}$
- Second Day: $R_2 = R_1 \left(\frac{\pi_1 s_{11}}{R_1} s_{12} + \frac{\pi_2 s_{21}}{R_1} s_{22} + ... + \frac{\pi_5 s_{51}}{R_1} s_{52} \right) = \pi_1 s_{11} s_{12} + ... + \pi_5 s_{51} s_{52}$
- ...
- Last Day: $R_N = \sum_{i=1}^{5} \pi_i V_i(N)/V_i(0)$

In the case where for example $s_{11} \neq 1$ the weighting of the first commodity on the second day is not equal to the start weighting ($\pi_1 \neq \pi_1 s_{11}$). If one would rebalance on the second day one

Table 3.5.: Results of Markowitz Optimization: Optimal Weightings

Markowitz				
Brent	7%	35%	59%	100%
Gasoil	10%	51%	41%	0%
Alu	0%	0%	0%	0%
Copper	0%	0%	0%	0%
Gold	83%	14%	0%	0%
Index Return	1.71	4.28	5.19	6.05
Index Variance	1.00	21.00	36.00	63.28
Index BS Drift	6%	15%	17%	19%
Index BS Volatility	17%	28%	30%	32%

Table 3.6.: Results of Markowitz Optimization: Optimal Weightings with Forced Inclusion of Aluminum and Copper

Brent	7%	42%	73%	80%
Gasoil	11%	38%	7%	0%
Alu	10%	10%	10%	10%
Copper	10%	10%	10%	10%
Gold	62%	0%	0%	0%
Index Return	1.66	4.18	4.84	4.98
Index Variance	1.00	21.00	36.00	40.55
Index BS Drift	5%	15%	16%	17%
Index BS Volatility	18%	29%	31%	32%

would buy/sell futures contracts to force the weightings to be equal to $(\pi_i)_{i=1,\dots,5}$. In our special case the yellow rebalancing frontier outperforms enormously the static blue Markowitz frontier.

A last note on the transaction costs: At first site it seems quite surprising that the trading costs are lower when I rebalance in comparison of no rebalancing. Rebalancing costs are additional costs to the costs of rolling. But observe that the trading costs of the index depend highly on the amount of Brent in the index. The transaction costs on the International Petroleum Exchange is set to 0.3% US dollars per ounce/ per tonne. For example, on January, the 3rd 1995 the Brent lead futures has a value of $\$\,16$ per barrel and Gasoil $\$\,145$ per tonne. If one invests $\$\,10,000$ in Brent (Gasoil) one has to pay $\approx \$\,2$ ($\$\,0.2$). So the transaction costs of Brent are ten times higher than those of Gasoil. This is the reason why transaction costs depend only on the amount of Brent in the index. In the case of monthly rebalancing the value of Brent is thwarted, so are

Table 3.7.: Monthly Rebalancing

Markowitz-Weightings as Start Values				
Brent	7%	35%	59%	100%
Gasoil	10%	51%	41%	0%
Alu	0%	0%	0%	0%
Copper	0%	0%	0%	0%
Gold	83%	14%	0%	0%
Monthly Rebalancing Index Results				
Index Return	1.50	4.13	5.19	6.05
Index Trading Costs	3,040	21,667	38,601	68,338
Index BS Drift	4%	15%	17%	19%
Index BS Volatility	12%	25%	29%	32%

the transaction costs compared to the ones of no rebalancing.

Mean variance optimization penalizes up-side and down-side risk equally, whereas most investors do not mind up-side risk. To overcome this, I consider in the next section other risk measures: Value-at-Risk and Conditional Value-at-Risk.

3.3. Minimization of Conditional Value-at-Risk for General Loss Distributions

This section presents an approach of Rockafellar and Uryasev [RU00] and [RU02] to optimizing a portfolio so as to reduce the risk of high losses. My contribution of this topic is to bridge a gap in the proof of the crucial theorem in the article of Rockafellar and Uryasev [RU02]. Furthermore, I present an alternative way of proving some parts of the main theorem. I then apply the theoretical results in the field of portfolio optimization with commodity indices.

The Conditional Value-at-Risk CVaR is generalized to discontinuous cumulative distribution functions of the loss. For continuous loss distributions, the CVaR at a given confidence level is defined as the expected loss exceeding the Value-at-Risk VaR. Loss distributions associated with finite sampling or scenario modeling are, however, discontinuous. Various risk measures involving discontinuous loss distributions shall be introduced and compared. They depend on a decision variable x and the confidence level α.

Furthermore, I uncover graphically the behavior of these risk measures. Especially, I consider the risk measures as a function of α. Based on a special discrete loss distribution, the graphs demonstrate the different properties of these risk measures. One recognizes graphically that the definition of the Conditional Value-at-Risk as given in Rockafellar and Uryasev [RU02] is the most reasonable generalization to distributions with possible discontinuities. It seems rather surprising that three loss points are sufficient to reveal graphically the different behaviors of all four risk measures.

Loss can be envisioned as a function $z = f(x, y)$ of a decision vector $x \in X \subseteq \mathbb{R}^n$ representing what we may generally call a portfolio or index, with X expressing decision constraints, and a vector $y \in Y \subseteq \mathbb{R}^m$ representing the future values of a number of variables like in our case commodity rolling strategies returns, bond returns and structured product returns. When y is taken to be random with known probability distribution, z comes out as a random variable having its distribution dependent on the choice of x. As an example consider the case of three returns of three assets $y = (y_1, y_2, y_3)$ and $x = (x_1, x_2, x_3)$ representing the weightings of the respective assets in the portfolio. The weightings should be in $X := \left\{ x \mid x_i \in [0,1] \quad \text{for all} \quad i = 1,2,3 \quad \text{and} \quad \sum_{i=1}^{3} x_i = 1 \right\}$. Then a possible way of defining the loss function z is the negative return $z = f(x, y) = -x_1 y_1 - x_2 y_2 - x_3 y_3$.

Any optimization problem of maximizing the expected return of a portfolio in terms of the choice of x should then also take into account the riskiness of the portfolio which depends on x. In section 3.2 the risk was measured by the variance of the portfolio. The higher the variance the riskier was the portfolio. Mean-variance optimization penalizes up-side and down-side risk equally, whereas most investors do not mind up-side risk. To overcome this, we consider in this section various other risk measures with main focus on the last one: Value-at-Risk and Conditional Value-at-Risk.

In everything that follows, we suppose the random vector y (the "return vector") is governed by a probability measure $\mathbb{P}$ on Y (a Borel-measure) that is independent of the index weightings x. For each x, we denote by $\Psi(x, \cdot)$ on $\mathbb{R}$ the resulting distribution function for the loss $z = f(x, y)$:

$$\Psi(x, \zeta) = \mathbb{P}\left[\{y \mid f(x, y) \leq \zeta\}\right],$$

making the technical assumptions that $f(x, y)$ is continuous in x and measurable in y, and that $\mathbb{E}\left[|f(x, y)|\right] < \infty$.

We consider a confidence level $\alpha \in]0, 1[$, which in applications would be something like $\alpha = 0.95$ or $\alpha = 0.99$. At this confidence level, there is a corresponding VaR and CVaR:

Definition 3.3.1. *The $\alpha-$ VaR of the loss associated with a decision x is the value*

$$\zeta_\alpha(x) = \min\{\zeta \mid \Psi(x, \zeta) \geq \alpha\} \tag{3.7}$$

The minimum in 3.7 is attained because $\Psi(x, \zeta)$ is nondecreasing and right-continuous in ζ.

Definition 3.3.2. *In the case of $\Psi(x, \zeta_\alpha(x)) < 1$ (so there is a chance of a loss greater than $\zeta_\alpha(x)$) the $\alpha-$CVaR is defined by*

$$\phi_\alpha(x) = \frac{\Psi(x, \zeta_\alpha(x)) - \alpha}{1 - \alpha}\, \zeta_\alpha(x) + \frac{1 - \Psi(x, \zeta_\alpha(x))}{1 - \alpha}\, \mathbb{E}\left[f(x, y) \mid f(x, y) > \zeta_\alpha(x)\right] \tag{3.8}$$

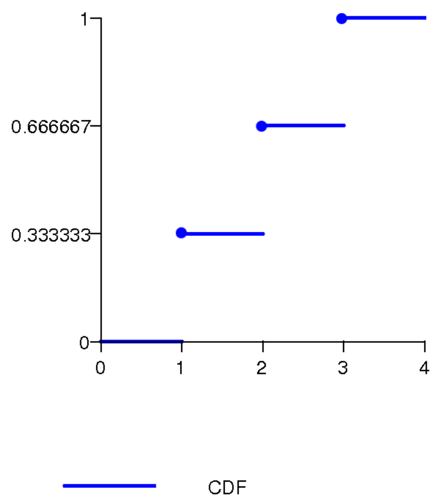
If $\Psi(x, \zeta_\alpha(x)) = 1$ (so $\zeta_\alpha(x)$ is the highest loss that can occur and thus the conditional expectation in 3.8 is ill-defined) then

$$\phi_\alpha(x) = \zeta_\alpha(x)$$

The crucial feature in definition 3.3.2 of $\alpha-$CVaR is the splitting of probability atoms (if present): In the circumstances of figure 3.3 where for $\alpha = 1/2$ (thus $\zeta_\alpha(x) = 2$) holds that

$$1/3 = \lim_{\zeta \nearrow \zeta_\alpha(x)} \Psi(x, \zeta) < \alpha < \lim_{\zeta \searrow \zeta_\alpha(x)} \Psi(x, \zeta) = \Psi(x, \zeta_\alpha(x)) = 2/3,$$

Figure 3.3.: Cumulative Distribution Function $\Psi(x, \cdot)$ of Loss (x-axis) of Example 3.3.3



an atom at $\zeta_\alpha(x)$ having total probability $\Psi(x, \zeta_\alpha(x)) - \lim_{\zeta \nearrow \zeta_\alpha(x)} \Psi(x, \zeta) = 1/3 > 0$ is effectively split into two pieces with probabilities $\Psi(x, \zeta_\alpha(x)) - \alpha = 2/3 - 1/2 = 1/6$ and $\alpha - \lim_{\zeta \nearrow \zeta_\alpha(x)} \Psi(x, \zeta) = 1/6$, respectively. Only the first of these pieces is adjoined to the interval $]\zeta_\alpha(x), \infty[$, which itself has probability $1 - \Psi(x, \zeta_\alpha(x)) = 1/3$. So one achieves a probability of $1 - \alpha = 1/2$, whereas, if the atom could not be split, we would have to choose between the intervals $]\zeta_\alpha(x), \infty[$ and $[\zeta_\alpha(x), \infty[$, neither of which actually has probability $1 - \alpha = 1/2$.

Definition 3.3.3. *The upper α-CVaR of the loss associated with a decision x is the value*

$$\phi_\alpha^+(x) := \mathbb{E}\left[f(x,y) \mid f(x,y) > \zeta_\alpha(x)\right], \tag{3.9}$$

whereas the lower α-CVaR of the loss is the value

$$\phi_\alpha^-(x) := \mathbb{E}\left[f(x,y) \mid f(x,y) \geq \zeta_\alpha(x)\right] \tag{3.10}$$

The conditional expectation in 3.10 is well defined because $\mathbb{P}\left[\{y \mid f(x,y) \geq \zeta_\alpha(x)\}\right] \geq 1 - \Psi(x, \zeta_\alpha(x)) = 1 - \alpha > 0$, since we have assumed that $\alpha \in \,]0,1[$. But the one in 3.9 only makes sense as long as $\mathbb{P}\left[\{y \mid f(x,y) > \zeta_\alpha(x)\}\right] > 0$, that is, $\Psi(x, \zeta_\alpha(x)) < 1$, which is not assured through our assumption that $\alpha \in \,]0,1[$, since there might be a probability atom at $\zeta_\alpha(x)$ large enough to cover the interval $1 - \lim_{\zeta \nearrow \zeta_\alpha(x)} \Psi(x, \zeta)$.

The reasons for defining the α-CVaR for general loss distributions as a weighted average of α-VaR and upper α-CVaR are the following:

- The *fundamental minimization formula* (3.3.1, upcoming) which states that the calculation of α-CVaR can be conducted by a convenient optimization of convex type in one dimension does hold for the α-CVaR as defined in 3.3.2. But it does not hold for the upper and lower α-CVaR as will be demonstrated in corollary 3.3.1.
- α-CVaR as defined in 3.3.2 is *continuous in the confidence level* α. Note that for the risk measures α-VaR, upper α-CVaR and lower α-CVaR a jump in the respective risk measures is sure to occur if a slightly higher confidence level is demanded. Of course in practice this is an undesirable feature. Example 3.3.3 highlights these characteristics of the four risk measures.
- α-CVaR as defined in 3.3.2 is a *coherent risk measure* in the sense of Artzner et al. [ADEH99] where α-VaR, upper and lower α-CVaR fail to be coherent (see on page 1458 in Rockafellar and Uryasev [RU02]).

I now calculate the α-VaR and α-CVaR in three special cases. In the first example the distribution function of the loss $\Psi(x,\cdot)$ is continuous. Example 3.3.1 shows that the definitions for the α-VaR and α-CVaR comprise the respective "standard" definitions given in the article of Rockafellar and Uryasev [RU00]. The second example picks up a well-known situation in practice: The probability measure $\mathbb{P}$ of the return vector y is the result of a Monte-Carlo simulation (for example $10,000$ simulations). So $\mathbb{P}$ is concentrated in finitely many points y_k in Y. The last example shows the continuity of the α-CVaR with respect to the confidence level α in contrast to α-VaR, upper and lower α-CVaR.

Example 3.3.1. *If the distribution function of the loss $\Psi(x,\cdot)$ is continuous then for the α-VaR holds*

$$\zeta_\alpha(x) = \min\{\zeta \mid \Psi(x,\zeta) \geq \alpha\} = \min\{\zeta \mid \Psi(x,\zeta) = \alpha\}$$

If $\Psi(x,\cdot)$ is in addition strictly increasing, $\zeta_\alpha(x)$ is simply the unique ζ satisfying $\Psi(x,\zeta) = \alpha$. If $\Psi(x,\cdot)$ is not strictly increasing, that means the distribution function has "flat" spots, the equation $\Psi(x,\zeta) = \alpha$ can have a whole range of solutions.

With $\alpha \in]0,1[$ the upper α-CVaR is well defined since

$$\mathbb{P}\left[\{y|f(x,y) > \zeta_\alpha(x)\}\right] = \mathbb{P}\left[\{y|f(x,y) \geq \zeta_\alpha(x)\}\right] > 0$$

The α-CVaR and the lower and upper α-CVaR coincide

$$\phi_\alpha^-(x) = \phi_\alpha(x) = \phi_\alpha^+(x)$$

since

$$\begin{aligned}\mathbb{E}\left[f(x,y) \mid f(x,y) \geq \zeta_\alpha(x)\right] &= 0\,\zeta_\alpha(x) + 1\,\mathbb{E}\left[f(x,y) \mid f(x,y) > \zeta_\alpha(x)\right] \\ &= \mathbb{E}\left[f(x,y) \mid f(x,y) > \zeta_\alpha(x)\right]\end{aligned}$$

Example 3.3.2. *Suppose the probability measure $\mathbb{P}$ is concentrated in finitely many points y_k in Y, so that for each $x \in X$ the distribution of the loss $z = f(x,y)$ is likewise concentrated in finitely many points, and $\Psi(x,\cdot)$ is a step function with jumps at those points. Fixing x, let those corresponding loss points be ordered as $z_1 < z_2 < \cdots < z_N$, with the probability of z_k being $p_k > 0$. Let k_α be the unique index such that*

$$\sum_{k=1}^{k_\alpha} p_k \geq \alpha > \sum_{k=1}^{k_\alpha - 1} p_k \tag{3.11}$$

The α-VaR of the loss is then given by

$$\zeta_\alpha(x) = z_{k_\alpha} \tag{3.12}$$

The conditional expectation appearing in the definition of the $\alpha-$CVaR can be calculated in our example as (see Billingsley [Bil95], Example 34.1, page 446)

$$\mathbb{E}\left[f(x,y) \mid f(x,y) > \zeta_\alpha(x)\right] = \frac{1}{1 - \sum_{k=1}^{k_\alpha} p_k} \sum_{k=k_\alpha+1}^{N} p_k z_k \tag{3.13}$$

where I have used

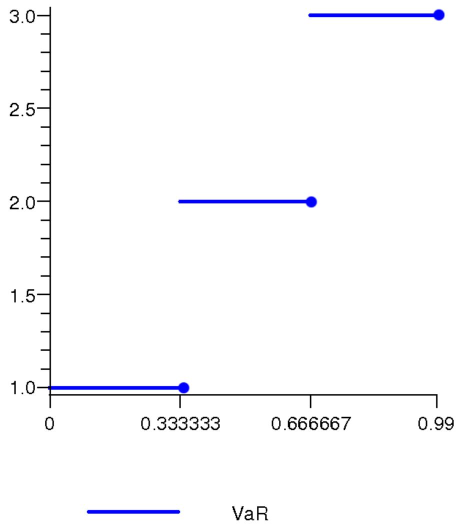
$$\Psi(x, \zeta_\alpha(x)) = \sum_{k=1}^{k_\alpha} p_k$$

When $\Psi(x,\zeta_\alpha(x)) < 1$ then the α-CVaR can then be written as

$$\phi_\alpha(x) = \frac{1}{1-\alpha}\left[\left(\sum_{k=1}^{k_\alpha} p_k - \alpha\right) z_{k_\alpha} + \sum_{k=k_\alpha+1}^{N} p_k z_k\right] \tag{3.14}$$

Otherwise $\phi_\alpha(x) = \zeta_\alpha(x)$. Note that I have suppressed the dependence on x for the loss $z_i = f(x, y_i)$ and the index value $k_\alpha(x)$.

Figure 3.4.: $\alpha-$VaR in Dependence of Confidence Level α (X-axis) in Example 3.3.3



Remark 3.3.1. *In the setting of Example 3.3.2 imagine we are facing the problem of minimizing the $\alpha-$CVaR in dependence of the decision variable $x \in X$ ("portfolio weightings"). For each $x \in X$ we would have to run the algorithm for the calculation of the $\alpha-$CVaR: First sorting the losses, determination of the $\alpha-$ VaR and applying the $\alpha-$CVaR formula 3.14. To conduct the algorithm one has to discretize the set X. Furthermore, it is not clear if formula 3.14 is convex. The following propositions and theorems shall be concerned with providing some insight how we can solve the problem of efficiently calculate the $\alpha-$CVaR.*

Example 3.3.3. *I now calculate and visualize in MAPLE the different risk measures in a concrete version of example 3.3.2. For fixed $x \in X$ the distribution of loss $z = f(x, y)$ is concentrated in three points $y_1, y_2, y_3 \in Y$, whereby the values of the corresponding loss points z_1, z_2, z_3 with probability p_i are given by table 3.8. Then the distribution function of the loss $\Psi(x, \cdot)$ and the α-VaR in dependence of the confidence level α take the form as in figures 3.3 and 3.4. The diagram for distribution function $\Psi(x, \cdot)$ shows that for example for $\alpha \in]\,1/3, 2/3\,[$ the equation $\Psi(x, \zeta) = \alpha$ has no solution whereas the equation $\Psi(x, \zeta) = 1/3$ has many solutions in ζ. The diagram for the $\alpha-$ VaR $\zeta_\alpha(x)$ exposes the instability of the $\alpha-$ VaR with respect to α: A slightly higher confidence level can cause a jump.*

For calculating the $\alpha-$ VaR and upper $\alpha-$CVaR I apply formulas 3.12 and 3.13. The formula

Table 3.8.: Losses and Probabilities in Example 3.3.3

	y_1	y_2	y_3
z_i	1	2	3
p_i	1/3	1/3	1/3

for the lower $\alpha-$CVaR takes the form

$$\phi_\alpha^- = \frac{1}{1-\sum_{k=1}^{k_\alpha-1} p_k} \sum_{k=k_\alpha}^{N} p_k z_k$$

Table 3.9 shows the values of the three risk measures in dependence of α. This simple example shows that the risk measures lower and upper $\alpha-$CVaR are generally instable with respect to the confidence level α. In addition we see that the upper $\alpha-$CVaR is always greater or equal than the lower $\alpha-$CVaR and the last one is always greater or equal than the $\alpha-$VaR (for a formal justification of these features one can consult Proposition 5, page 1450 in Rockafellar and Uryasev [RU02]). Plot 3.5 illustrates the values of the four risk measures.

Table 3.9.: Values for $\alpha-$VaR, lower and upper $\alpha-$CVaR

	$\alpha-$VaR	lower $\alpha-$CVaR	upper $\alpha-$CVaR	k_α
$\alpha \in]0, 1/3]$	1	2	2.5	1
$\alpha \in]1/3, 2/3]$	2	2.5	3	2
$\alpha \in]2/3, 1[$	3	3	not defined	3

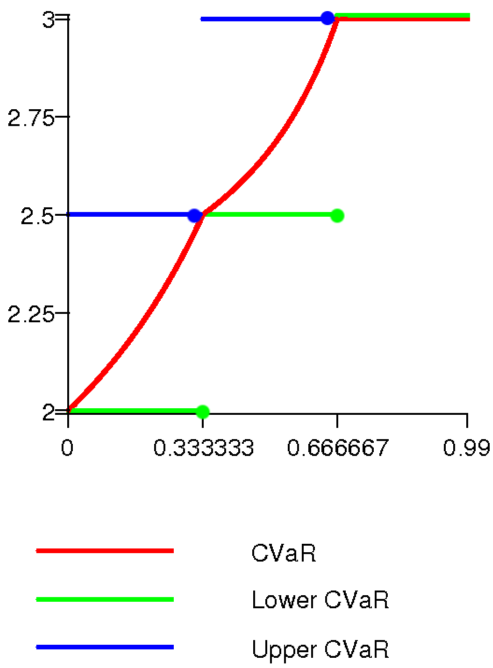
If one defines the $\alpha-$CVaR as the weighted average of the $\alpha-$VaR and the upper $\alpha-$CVaR (where the weighting depends on α as in 3.3.2) then the risk measure is continuous with respect to the confidence level α (see proposition 13, page 1458 in Rockafellar and Uryasev [RU02]). In our example the $\alpha-$CVaR can be calculated via formula 3.14):

$$\alpha \in]0, 1/3] \quad : \quad \phi_\alpha(x) = \frac{1}{1-\alpha}\left[(1/3-\alpha)\,1 + \left(\frac{1}{3}2 + \frac{1}{3}3\right)\right] = \frac{2-\alpha}{1-\alpha}$$

$$\alpha \in]1/3, 2/3] \quad : \quad \phi_\alpha(x) = \frac{1}{1-\alpha}\left[(2/3-\alpha)\,2 + \left(\frac{1}{3}2 + \frac{1}{3}3\right)\right] = \frac{7/3-2\alpha}{1-\alpha}$$

$$\alpha \in]2/3, 1[\quad : \quad \phi_\alpha(x) = \frac{1}{1-\alpha}\left[(1-\alpha)\,3\right] = 3$$

Figure 3.5.: Risk Measures in Dependence of Confidence Level α in Example 3.3.3



So the α-CVaR takes the form

$$\phi_\alpha(x) = \begin{cases} \frac{2-\alpha}{1-\alpha} & \textit{for} \quad \alpha \in \,]\,0, 1/3\,] \\ \frac{7/3-2\alpha}{1-\alpha} & \textit{for} \quad \alpha \in \,]\,1/3, 2/3\,] \\ 3 & \textit{for} \quad \alpha \in \,]\,2/3, 1\,] \end{cases}$$

So our simple example shows that among the risk measures $\alpha-$ VaR, $\alpha-$CVaR, lower and upper $\alpha-$CVaR only $\alpha-$CVaR is stable with respect to the confidence level α.

In the article of Rockafellar and Uryasev [RU00] is shown that the $\alpha-$CVaR and $\alpha-$VaR of the loss z can be calculated simultaneously by solving an elementary optimization problem of convex type in one dimension (if $f(x, y)$ is convex in x). The optimization of a convex function is a very convenient problem since a local minimum is a global minimum. In [RU00] the crucial theorems have the unpractical assumption that for each $x \in X$ the distribution function for the loss $\Psi(x, \zeta)$ is everywhere continuous with respect to ζ. In their follow-up article Rockafellar and Uryasev [RU02] dropped this assumption and they proved that the theorems also hold in the general case if one defines the $\alpha-$CVaR via 3.3.2. Let us note that no such formulas hold for the lower and upper $\alpha-$CVaR as shown in corollary 3.3.1.

In preparation for the fundamental minimization theorem I now provide some definitions and propositions.

We denote by $\Psi(x,\zeta^-)$ the left limit of $\Psi(x,\cdot)$ at ζ; thus

$$\Psi(x,\zeta^-) = \mathbb{P}\left[\{y \mid f(x,y) < \zeta\}\right] \tag{3.15}$$

For $\zeta \in \mathbb{R}$ let $B(\zeta)$ be defined by

$$B(\zeta) := \mathbb{E}\left[[f(x,y)-\zeta]^+\right]$$

Since $B(\zeta)$ is finite and convex the one-sided derivatives of $B(\zeta)$ exist as will be shown in proposition 3.3.1. There I determine the right and left derivatives of $B(\zeta)$. The assertions can be found implicitly in the proof of the fundamental minimization theorem [RU02]. But I provide an alternative way of calculating the derivatives: Since the existence of the one-sided derivatives is assured, it suffices to determine the limit of the difference quotient of $B(\zeta)$ for a special sequence $\zeta + 1/n$ and $\zeta - 1/n$, respectively. This simplifies the proof of Rockafellar and Uryasev [RU02].

Proposition 3.3.1. *Let the decision vector $x \in X$ be fixed. Then for all $\zeta \in \mathbb{R}$ holds*

1. *The right derivative of $B(\zeta)$ exists with $B'_r(\zeta) = \Psi(x,\zeta) - 1$*
2. *The left derivative of $B(\zeta)$ exists with $B'_l(\zeta) = \Psi(x,\zeta^-) - 1$.*

Proof: For each return vector $y \in Y$ the function $b(\zeta,y) = [f(x,y)-\zeta]^+$ is a convex function with respect to ζ since b is a composition of the convex function $\zeta \to f(x,y)-\zeta$ and the nondecreasing convex function $t \to [t]^+$ (see Rockafellar [Roc97], Theorem 5.1). It follows that $B(\zeta) = \mathbb{E}\left[[f(x,y)-\zeta]^+\right]$ is also convex and finite since by assumption $\mathbb{E}\left[|f(x,y)|\right] < \infty$. As a finite convex function, $B(\zeta)$ has finite right and left derivatives at any ζ (see Rockafellar [Roc97], Theorem 23.1 and Theorem 24.1).

1. Let $\zeta_0 \in \mathbb{R}$ and $s_n = 1/n$. The right difference quotient of b in ζ_0 reads as $g_n(y) = \dfrac{b(\zeta_0+s_n,y)-b(\zeta_0,y)}{s_n}$. The sequence $g_n(y)$ is monotone decreasing since $b(\zeta,y)$ is convex. Now, if $f(x,y) > \zeta_0$ then there exists $N \in \mathbb{N}$ with $\zeta_0 + s_N < f(x,y)$ and it follows that
$$\frac{[f(x,y)-(\zeta_0+s_n)]^+ - [f(x,y)-\zeta_0]^+}{s_n} = \frac{f(x,y)-(\zeta_0+s_n)-(f(x,y)-\zeta_0)}{s_n} = -1$$

for all $n \geq N$.

If $f(x, y) \leq \zeta_0$ then

$$\frac{[f(x,y) - (\zeta_0 + s_n)]^+ - [f(x,y) - \zeta_0]^+}{s_n} = 0 \qquad \text{for all} \qquad n \in \mathbb{N}$$

It follows that

$$\lim_{n\to\infty} g_n(y) = \frac{b(\zeta_0 + s_n, y) - b(\zeta_0, y)}{s_n} = \begin{cases} -1, & \text{if} \quad f(x,y) > \zeta_0 \\ 0, & \text{if} \quad f(x,y) \leq \zeta_0 \end{cases}$$

$\lim_{n\to\infty} \mathbb{E}[g_n(y)] = \mathbb{E}[\lim_{n\to\infty} g_n(y)]$ by the monotone convergence theorem. So I obtain

$$\begin{aligned} \lim_{n\to\infty} s_n^{-1} \left(\mathbb{E}[b(\zeta_0 + s_n, y)] - \mathbb{E}[b(\zeta_0, y)]\right) &= \lim_{n\to\infty} \frac{\mathbb{E}[b(\zeta_0 + s_n, y) - b(\zeta_0, y)]}{s_n} \\ &= \lim_{n\to\infty} \mathbb{E}[g_n(y)] = \mathbb{E}\left[\lim_{n\to\infty} g_n(y)\right] \\ &= -1\, \mathbb{P}[f(x,y) > \zeta_0] = -1\ (1 - \Psi(x, \zeta_0)) \\ &= \Psi(x, \zeta_0) - 1 \end{aligned}$$

Hence $B_r'(\zeta) = \Psi(x, \zeta_0) - 1$ is proved.

2. Let $\zeta_0 \in \mathbb{R}$ and $t_n = -1/n$. Setting $h_n(y) = \dfrac{b(\zeta_0 + t_n, y) - b(\zeta_0, y)}{t_n}$ the sequence $(h_n(y))$ is increasing for each y (since $b(\zeta, y)$ is convex with respect to ζ) and one obtains analogously to 1

$$\lim_{n\to\infty} h_n(y) = \frac{b(\zeta_0 + t_n, y) - b(\zeta_0, y)}{t_n} = \begin{cases} -1, & \text{if} \quad f(x,y) \geq \zeta_0 \\ 0, & \text{if} \quad f(x,y) < \zeta_0 \end{cases}$$

and

$$\begin{aligned} \lim_{n\to\infty} t_n^{-1} \left(\mathbb{E}[b(\zeta_0 + t_n, y)] - \mathbb{E}[b(\zeta_0, y)]\right) &= \lim_{n\to\infty} \frac{\mathbb{E}[b(\zeta_0 + t_n, y) - b(\zeta_0, y)]}{t_n} \\ &= \lim_{n\to\infty} \mathbb{E}[h_n(y)] = \mathbb{E}\left[\lim_{n\to\infty} h_n(y)\right] \\ &= -1\, \mathbb{P}[f(x,y) \geq \zeta_0] \\ &= -1\ (1 - \mathbb{P}[f(x,y) < \zeta_0]) \\ &= \Psi(x, \zeta_0^-) - 1 \end{aligned}$$

where the last step applies definition 3.15. So I have shown that $B_l'(\zeta) = \Psi(x, \zeta_0^-) - 1$.

■

Next, we shall turn to the fundamental minimization theorem as stated in Rockafellar and Uryasev [RU00] and [RU02]. This important theorem states in particular that

$$\begin{aligned} \phi_\alpha(x) &= \min_{\zeta\in\mathbb{R}} F_\alpha(x,\zeta) \qquad \text{with} \\ F_\alpha(x,\zeta) &= \zeta + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y)-\zeta]^+\right] \end{aligned} \tag{3.16}$$

Remark 3.3.2. *Rockafellar and Uryasev proved in [RU02] that the $\alpha-$VaR $\zeta_\alpha(x)$ is in the set $\operatorname{argmin}_\zeta F_\alpha(x,\zeta)$ and conclude from 3.16 the identity $\phi_\alpha(x) = F_\alpha(x,\zeta_\alpha(x))$. But in their proof one cannot find any hint for the correctness of 3.16. I close this gap by choosing the opposite way: I first prove directly the identity $\phi_\alpha(x) = F_\alpha(x,\zeta_\alpha(x))$ and from $\zeta_\alpha(x) \in \operatorname{argmin}_\zeta F_\alpha(x,\zeta)$ I can conclude the correctness of $\phi_\alpha(x) = \min_\zeta F_\alpha(x,\zeta)$.*

Proposition 3.3.2. *For each $x \in X$ one has*

$$\phi_\alpha(x) = F_\alpha(x,\zeta_\alpha(x))$$

Proof: The auxiliary function $F_\alpha(x,\zeta)$ evaluated at the $\alpha-$VaR reads as

$$F_\alpha(x,\zeta_\alpha(x)) = \zeta_\alpha(x) + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y)-\zeta_\alpha(x)]^+\right] \tag{3.17}$$

If $\Psi(x,\zeta_\alpha(x)) < 1$ the expectation appearing in 3.17 can be written as

$$\begin{aligned} \mathbb{E}\left[[f(x,y)-\zeta_\alpha(x)]^+\right] &= \mathbb{E}\left[f(x,y)1_{\{f(x,y)>\zeta_\alpha(x)\}}\right] - \zeta_\alpha(x)\mathbb{E}\left[1_{\{f(x,y)>\zeta_\alpha(x)\}}\right] \\ &= \mathbb{P}\left[f(x,y)>\zeta_\alpha(x)\right]\mathbb{E}\left[f(x,y)\mid f(x,y)>\zeta_\alpha(x)\right] \\ &- \zeta_\alpha(x)\left(1-\mathbb{P}\left[f(x,y)\le\zeta_\alpha(x)\right]\right) \\ &= (1-\Psi(x,\zeta_\alpha(x)))\left[\mathbb{E}\left[f(x,y)\mid f(x,y)>\zeta_\alpha(x)\right]-\zeta_\alpha(x)\right] \end{aligned} \tag{3.18}$$

Note that I applied in 3.18 the formula

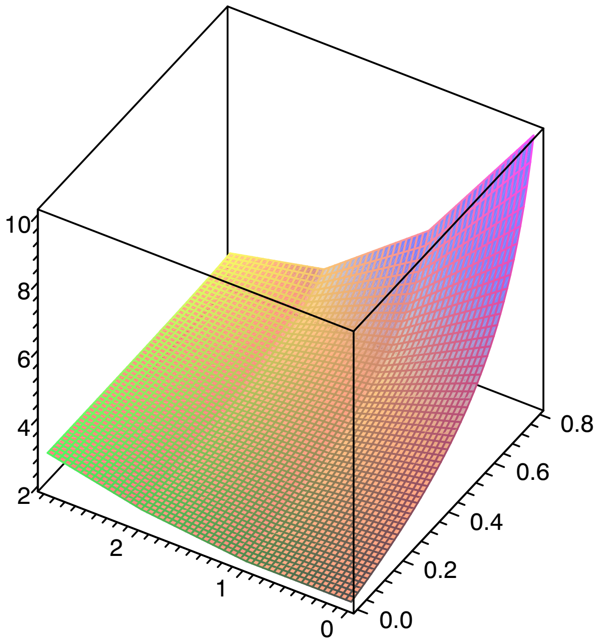
$$\mathbb{E}\left[f(x,y)\mid f(x,y)>\zeta_\alpha(x)\right] = \left(\mathbb{P}\left[f(x,y)>\zeta_\alpha(x)\right]\right)^{-1}\mathbb{E}\left[f(x,y)1_{\{f(x,y)>\zeta_\alpha(x)\}}\right]$$

which can be found for example in Billingsley [Bil95], Example 34.1, page 446.

So I obtain

$$\begin{aligned} F_\alpha(x,\zeta_\alpha(x)) &= \zeta_\alpha(x) + (1-\alpha)^{-1}(1-\Psi(x,\zeta_\alpha(x)))\left[\mathbb{E}\left[f(x,y)\mid f(x,y)>\zeta_\alpha(x)\right]-\zeta_\alpha(x)\right] \\ &= \frac{\Psi(x,\zeta_\alpha(x))-\alpha}{1-\alpha}\,\zeta_\alpha(x) + \frac{1-\Psi(x,\zeta_\alpha(x))}{1-\alpha}\,\mathbb{E}\left[f(x,y)\mid f(x,y)>\zeta_\alpha(x)\right] \\ &= \phi_\alpha(x) \end{aligned}$$

Figure 3.6.: Function $F_{\cdot}(x, \cdot)$ in Dependence of Confidence Level α and Losses ζ in Example 3.3.3



In the case of $\Psi(x, \zeta_\alpha(x)) = 1$ the $\alpha-$CVaR is by definition equal to the $\alpha-$VaR and it holds in that case for the auxiliary function

$$F_\alpha(x, \zeta_\alpha(x)) = \zeta_\alpha(x) + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y) - \zeta_\alpha(x)]^+\right] = \zeta_\alpha(x)$$

■

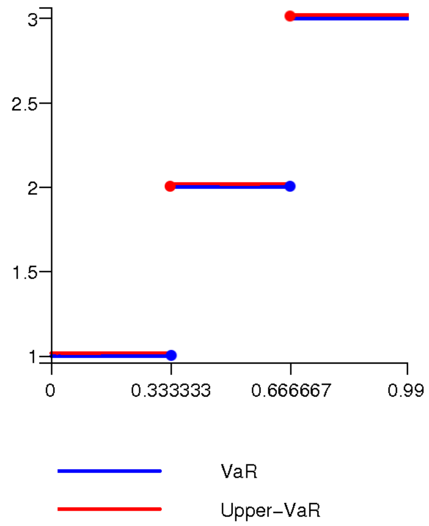
Remark 3.3.3. *At first glance the definition of the $\alpha-$CVaR as a weighted average of $\alpha-$VaR and upper $\alpha-$CVaR seemed artificial. But proposition 3.3.2 in combination with the upcoming theorem 3.3.1 explain why we defined it in that way.*

Definition 3.3.4. *(Upper $\alpha-$VaR) The upper $\alpha-$VaR of the loss with respect to the decision vector x is defined by*

$$\zeta_\alpha^+(x) = \inf\{\zeta \mid \Psi(x,\zeta) > \alpha\}$$

Clearly, $\zeta_\alpha^+(x) \geq \zeta_\alpha(x)$. These values are equal except when $\Psi(x, \zeta)$ is constant at level α over

Figure 3.7.: Differences of VaR and upper-VaR in Dependence of Confidence Level α in Example 3.3.3



a $\zeta-$interval. This is illustrated in diagrams 3.7 where I apply the example 3.3.3 to the risk measures upper $\alpha-$VaR and $\alpha-$VaR.

Theorem 3.3.1. *(Fundamental Minimization Formula) Function*

$$F_\alpha(x,\zeta) = \zeta + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y)-\zeta]^+\right]$$

is finite and convex (hence continuous) in $\zeta \in \mathbb{R}$ with

1. $\phi_\alpha(x) = \min_\zeta F_\alpha(x,\zeta)$
2. $[\zeta_\alpha(x), \zeta_\alpha^+(x)] = \operatorname{argmin}_\zeta F_\alpha(x,\zeta)$

Proof: Function $F_\alpha(x,\zeta) = \zeta + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y)-\zeta]^+\right]$ is finite for each $x \in X$ since by assumption $\mathbb{E}\left[|f(x,y)|\right] < \infty$. Furthermore, $F_\alpha(x,\zeta)$ is convex as the composition of convex functions.

Applying Proposition 3.3.1 one has

$$\begin{aligned}
\frac{\partial^+ F_\alpha(x,\zeta)}{\partial\zeta} &= 1 + \frac{\Psi(x,\zeta)-1}{1-\alpha} = \frac{\Psi(x,\zeta)-\alpha}{1-\alpha} \\
\frac{\partial^- F_\alpha(x,\zeta)}{\partial\zeta} &= 1 + \frac{\Psi(x,\zeta^-)-1}{1-\alpha} = \frac{\Psi(x,\zeta^-)-\alpha}{1-\alpha}
\end{aligned}$$

These two one-sided derivatives are nondecreasing in ζ since the distribution function is nondecreasing. Furthermore, I have

$$\lim_{\zeta\to\infty}\frac{\partial^+ F_\alpha(x,\zeta)}{\partial\zeta} = \lim_{\zeta\to\infty}\frac{\partial^- F_\alpha(x,\zeta)}{\partial\zeta} = 1 \qquad \text{and}$$
$$\lim_{\zeta\to-\infty}\frac{\partial^- F_\alpha(x,\zeta)}{\partial\zeta} = \lim_{\zeta\to-\infty}\frac{\partial^- F_\alpha(x,\zeta)}{\partial\zeta} = -\frac{\alpha}{1-\alpha}$$

It follows that the level sets $\{\zeta \mid F_\alpha(x,\zeta)\le c\}$ are bounded for each $c\in\mathbb{R}$. Because of continuity the minimum $\min_\zeta F_\alpha(x,\zeta)$ is attained by the theorem of Weierstrass. Hence, in view of the convexity $\operatorname{argmin}_\zeta F_\alpha(x,\zeta)$ is a nonempty bounded interval and closed since $\operatorname{argmin}_\zeta F_\alpha(x,\zeta) = F_\alpha^{-1}(x,\{c\})$, $c=\min_\zeta F_\alpha(x,\zeta)$, is the closed inverse image of the continuous function $F_\alpha(x,\zeta)$.

It follows

$$\begin{aligned}\operatorname{argmin}_\zeta F_\alpha(x,\zeta) &= \left\{\zeta \mid \frac{\partial^- F_\alpha(x,\zeta)}{\partial\zeta}\le 0\le\frac{\partial^+ F_\alpha(x,\zeta)}{\partial\zeta}\right\}\\ &= \left\{\zeta \mid \frac{\Psi(x,\zeta^-)-\alpha}{1-\alpha}\le 0\le\frac{\Psi(x,\zeta)-\alpha}{1-\alpha}\right\}\\ &= \left\{\zeta \mid \Psi(x,\zeta^-)\le\alpha\le\Psi(x,\zeta)\right\}\end{aligned}$$

In view of definitions $\zeta_\alpha(x) = \min\{\zeta \mid \Psi(x,\zeta)\ge\alpha\}$ and $\zeta_\alpha^+(x) = \inf\{\zeta \mid \Psi(x,\zeta)>\alpha\}$ I can conclude that $[\zeta_\alpha(x),\zeta_\alpha^+(x)] = \operatorname{argmin}_\zeta F_\alpha(x,\zeta)$ which proves assertion 2.

Combining the identity of proposition 3.3.2 $\phi_\alpha(x) = F_\alpha(x,\zeta_\alpha(x))$ and assertion 2 which yields $\min_\zeta F_\alpha(x,\zeta) = F_\alpha(x,\zeta_\alpha(x))$ I obtain assertion 1: $\phi_\alpha(x) = \min_\zeta F_\alpha(x,\zeta)$. ■

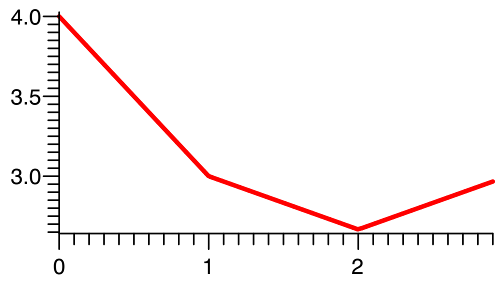
Example 3.3.4. *I now calculate and illustrate the behavior of function $F_\alpha(x,\zeta)$ in the setting of example 3.3.3. Remember that in this basic example the loss can take three values with equal probability, namely 1, 2 and 3.*

In this case $F_\alpha(x,\zeta) = \zeta + (1-\alpha)^{-1}\mathbb{E}\left[[f(x,y)-\zeta]^+\right]$ takes the form

$$\begin{aligned}\zeta\in\,]-\infty,1[\quad &: \quad F_\alpha(x,\zeta) = \zeta+(1-\alpha)^{-1}\left((1-\zeta)1/3+(2-\zeta)1/3+(3-\zeta)1/3\right) = \frac{2-\alpha\zeta}{1-\alpha}\\ \zeta\in[1,2[\quad &: \quad F_\alpha(x,\zeta) = \zeta+(1-\alpha)^{-1}\left((2-\zeta)1/3+(3-\zeta)1/3\right) = \frac{(1/3-\alpha)\zeta+5/3}{1-\alpha}\\ \zeta\in[2,3[\quad &: \quad F_\alpha(x,\zeta) = \zeta+(1-\alpha)^{-1}\left((2-\zeta)1/3+(3-\zeta)1/3\right) = \frac{(2/3-\alpha)\zeta+1}{1-\alpha}\\ \zeta\in[3,\infty[\quad &: \quad F_\alpha(x,\zeta) = \zeta\end{aligned}$$

The graph of this function is given in figure 3.6. The graph of $F_\cdot(x,\cdot)$ shows that $F_\alpha(x,\cdot)$ is continuous, not generally differentiable and convex which holds in general as proved in theorem 3.3.1

Figure 3.8.: Function $F_{0.5}(x, \cdot)$ in Example 3.3.3



and proposition 3.3.3.

In figures 3.9 and 3.8 I plot function $F_\alpha(x,\zeta)$ *for fixed values of the confidence level* $\alpha = 0.5$ *and* $\alpha = 1/3$*. These reveal that the extremal sets* $\operatorname{argmin}_{\zeta\in\mathbb{R}} F_\alpha(x,\zeta)$ *can vary from being just a single point (as in the case for* $\alpha = 0.5$*) to being a whole interval (as it is the case for* $\alpha = 1/3$*). The* $0.5-$ *VaR is equal to* 2 *and* $\phi_{0.5}(x) = F_{0.5}(x,2) = 2.\overline{6}$*. For* $\alpha = 1/3$ *I have* $\zeta_{1/3}(x) = 1$*,* $\zeta^+_{1/3}(x) = 2$ *and* $\phi_{1/3}(x) = 2.5$*.*

Corollary 3.3.1. *For the risk measures upper and lower* $\alpha-$*CVaR the fundamental minimization theorem 3.3.1 does not hold in general.*

Proof: Is already given in example 3.3.3. The lower $0.5-$CVaR is in this case equal to 2.5 and the upper $0.5-$CVaR is 3 (see table 3.9). But it holds $\min_\zeta F_{0.5}(x,\zeta) = F_{0.5}(x,2) = 2.\overline{6}$. The graph of $F_{0.5}(x,\cdot)$ can be found in figure 3.8.

■

Theorem 3.3.2. *(Optimization Shortcut) It holds*

1.

$$\min_{x\in X} \phi_\alpha(x) = \min_{(x,\zeta)\in X\times\mathbb{R}} F_\alpha(x,\zeta)$$

2.

$$F_\alpha(x^*,\zeta^*) = \mathit{minimal} \Longleftrightarrow \begin{cases} \phi_\alpha(x^*) = \mathit{minimal} \\ \zeta^* \in \operatorname{argmin}_{\zeta\in\mathbb{R}} F_\alpha(x^*,\zeta) \end{cases}$$

Proof:

1. One has

$$\min_{x\in X}\phi_\alpha(x) \geq \min_{(x,\zeta)\in X\times\mathbb{R}} F_\alpha(x,\zeta)$$

In fact, if

$$\min_{x\in X}\phi_\alpha(x) < \min_{(x,\zeta)\in X\times\mathbb{R}} F_\alpha(x,\zeta)$$

then there exists $x_0 \in X$ with

$$\phi_\alpha(x_0) \leq F_\alpha(x,\zeta)$$

for all $(x,\zeta) \in X \times \mathbb{R}$, in particular

$$\phi_\alpha(x_0) < F_\alpha(x_0,\zeta_\alpha(x_0))$$

which contradicts the assertion of the proposition 3.3.2.

On the other hand it holds

$$\min_{x\in X}\phi_\alpha(x) \leq \min_{(x,\zeta)\in X\times\mathbb{R}} F_\alpha(x,\zeta)$$

Suppose that

$$\min_{x\in X}\phi_\alpha(x) > \min_{(x,\zeta)\in X\times\mathbb{R}} F_\alpha(x,\zeta)$$

Then there exists $(x_0,\zeta_0) \in X \times \mathbb{R}$ with

$$\phi_\alpha(x) > F_\alpha(x_0,\zeta_0)$$

for all $x \in X$. By the minimization theorem 3.3.1 one has however

$$F_\alpha(x_0,\zeta_0) \geq F_\alpha(x_0,\zeta_\alpha(x_0)) = \phi_\alpha(x_0)$$

hence I have a contradiction.

2. Let $F_\alpha(x^*,\zeta^*) =$ minimal. Clearly, one has $\zeta^* \in \operatorname{argmin}_{\zeta\in\mathbb{R}} F_\alpha(x^*,\zeta)$. Assume that there exists $x_0 \in X$ with $\phi_\alpha(x_0) < \phi_\alpha(x^*)$. In view of the minimization theorem one has

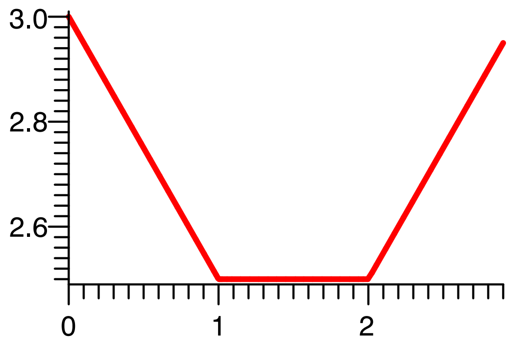
$$\phi_\alpha(x_0) = F_\alpha(x_0,\zeta_\alpha(x_0)) \geq F_\alpha(x^*,\zeta^*) = \min_{\zeta\in\mathbb{R}} F_\alpha(x^*,\zeta) = \phi_\alpha(x^*)$$

which is a contradiction.

Conversely, let $\phi_\alpha(x^*) =$ minimal and $\zeta^* \in \operatorname{argmin}_{\zeta\in\mathbb{R}} F_\alpha(x^*,\zeta)$. Assume that there exists $(x_0,\zeta_0) \in X \times \mathbb{R}$ with $F_\alpha(x^*,\zeta^*) > F_\alpha(x_0,\zeta_0)$. Applying the minimization theorem again one obtains

$$\begin{aligned} F_\alpha(x_0,\zeta_0) &\geq F_\alpha(x_0,\zeta_\alpha(x_0)) = \phi_\alpha(x_0) \geq \phi_\alpha(x^*) \\ &= \min_{\zeta} F_\alpha(x^*,\zeta) = F_\alpha(x^*,\zeta^*) \end{aligned}$$

Figure 3.9.: Function $F_{1/3}(x, \cdot)$ in Example 3.3.3



whence I have a contradiction.

■

In the following proposition 3.3.3 I simplify a proof of Rockafellar and Uryasev [RU02] by applying a result which I have proven in proposition 3.3.2.

Proposition 3.3.3. *If the loss $f(x, y)$ is convex with respect to x, then*

1. *$F_\alpha(x, \zeta)$ is jointly convex in (x, ζ).*
2. *$\phi_\alpha(x)$ is convex with respect to x.*

Proof:

1. For each return vector $y \in Y$ the function $[f(x, y) - \zeta]^+$ is convex with respect to (x, ζ) since this function is the composition of the convex function $\zeta \to f(x, y) - \zeta$ and the nondecreasing convex function $t \to [t]^+$ (see theorem 5.1 in Rockafellar [Roc97]). It follows that $F_\alpha(x, \zeta) = \zeta + (1 - \alpha)^{-1}\mathbb{E}\left[[f(x, y) - \zeta]^+\right]$ is also convex since the integrand is convex.
2. Follows directly from assertion 1 and the fact that by proposition 3.3.2 it holds $\phi_\alpha(x) = F_\alpha(x, \zeta_\alpha(x))$.

■

I am now interested in reconsidering the optimization problem of section 3.2. In this section I optimize the expected return of the index value with the constraint that the variance of the index is smaller than given bounds. By varying the bound value of the constraint one obtains the so called *efficient frontier* which is the graphical representation of risk versus return. Since the risk measure $\alpha-$CVaR allows a convenient optimization as illustrated in theorem 3.3.1 and proposition 3.3.3 (in the case of a convex loss function with respect to x) I now modify the above mentioned optimization problem by substituting the risk measure variance by the $\alpha-$CVaR.

To this end, the following theorem is a very useful vehicle.

Theorem 3.3.3. *For a given objective function $g : X \to \mathbb{R}$ and a given bound for the risk u let the optimization problem 1 be given by*

$$\textit{minimize} \quad g(x) \quad \textit{over} \quad x \in X \quad \textit{satisfying} \quad \phi_\alpha(x) \leq u$$

and optimization problem 2 be given by

$$\textit{minimize} \quad g(x) \quad \textit{over} \quad (x, \zeta) \in X \times \mathbb{R} \quad \textit{satisfying} \quad F_\alpha(x, \zeta) \leq u$$

Then the two problems are equivalent in the sense that

$$(x^*, \zeta^*) \quad \textit{solves the second problem} \Longleftrightarrow \begin{cases} x^* \quad \textit{solves the first problem} \\ F_\alpha(x^*, \zeta^*) \leq u \end{cases}$$

Proof: Suppose that x^* solves the first problem, that is, $g(x^*)$=minimal satisfying $\phi_\alpha(x^*) \leq u$. Assume that there exist $b^* \in X$ and $\zeta^* \in \mathbb{R}$ with $g(b^*) < g(x^*)$ and $F_\alpha(b^*, \zeta^*) \leq u$. Because of $\phi_\alpha(b^*) = \min_{\zeta \in \mathbb{R}} F_\alpha(b^*, \zeta)$ one has $\phi_\alpha(b^*) \leq F_\alpha(b^*, \zeta^*) \leq u$ which contradicts the fact that x^* is a solution of the first problem.

Conversely, if $g(x^*)$=minimal with $F_\alpha(b^*, \zeta^*) \leq u$ then x^* solves the first problem. In fact, assume that there exists $b^* \in X$ with $g(b^*) < g(x^*)$ and $\phi_\alpha(b^*) \leq u$. By proposition 3.3.2 one has $\phi_\alpha(b^*) = F_\alpha(b^*, \zeta_\alpha(b^*)) \leq u$ which contradicts that x^* is a solution of the second problem.

■

Example 3.3.5. *Remember example 3.3.2. Y is in this case a discrete probability space with elements y_k, $k = 1, 2, \ldots, N$, having probability $p_k = 1/N$. Then function $F_\alpha(x, \zeta)$ reads as*

$$F_\alpha(x, \zeta) = \zeta + ((1 - \alpha)N)^{-1} \sum_{k=1}^{N} [f(x, y_k) - \zeta]^+$$

The return of an index is the sum of the returns on the individual instruments in the index, scaled by the proportions x. Since I want to maximize the expected end value of the index the objective *function g appearing in theorem 3.3.3 is set to the negative expected return of the commodity index, namely $g(x) := -\frac{1}{N}\sum_{k=1}^{N}[x_1 y_{1k} + x_2 y_{2k} + x_3 y_{3k}]$. The loss $f(x, y)$ is the negative return of the index, namely $f(x, y_k) = -[x_1 y_{1k} + x_2 y_{2k} + x_3 y_{3k}]$. Note that both the objective function g and the loss function f are linear (convex) with respect to x. With the help of theorem 3.3.3 we are now facing the following optimization problem*

$$\begin{array}{ll} \textit{minimize} & g(x) = -\dfrac{1}{N}\displaystyle\sum_{k=1}^{N}[x_1 y_{1k} + x_2 y_{2k} + x_3 y_{3k}] \\ \textit{over} & (x, \zeta) \in X \times \mathbb{R} \textit{ satisfying } F_\alpha(x, \zeta) = \zeta + ((1-\alpha)N)^{-1}\displaystyle\sum_{k=1}^{N}[f(x, y_k) - \zeta]^+ \leq u \end{array}$$

The objective function is linear and even the constraint can be equivalently formulated as a linear one: We introduce additional variables η_k for $k = 1, 2, \ldots, N$ subject to the conditions

$$\eta_k \geq 0, \quad f(x, y_k) - \zeta - \eta_k \leq 0,$$

and requiring that

$$\zeta + ((1-\alpha)N)^{-1}\sum_{k=1}^{N}\eta_k \leq u$$

Thus we are in the very convenient situation where the optimization problem can be solved by applying linear programming techniques. Since the number of constraints is greater than the number of scenarios of the Monte Carlo-Simulation (for example $N = 100{,}000$) note that some tools like for example EXCEL-SOLVER or MATLAB cannot handle this large number of constraints or is too slow. I found out that CPLEX is very fast and can handle the amount of data, QSOPT has exactly the same performance as CPLEX and LP_SOLVE has nearly the same performance as CPLEX.

Let us summarize:

$$\begin{array}{ll} \textit{minimize} & g(x) = -\frac{1}{N}\sum_{k=1}^{N}\left[x_1 y_{1k} + x_2 y_{2k} + x_3 y_{3k}\right] \\ \textit{over} & x = (x_1, x_2, x_3), \zeta, \eta = (\eta_1, \eta_2, \ldots, \eta_N) \end{array}$$

subject to:

1. $0 \leq x_k \leq 1, \quad \sum_{k=1}^{3} x_k = 1 \quad \textit{for} \quad k = 1,2,3$
2. $\eta_k \geq 0 \quad \textit{for} \quad k = 1,2,\ldots,N$
3. $f(x, y_k) - \zeta - \eta_k \leq 0 \quad \textit{for} \quad k = 1,2,\ldots,N$
4. $\zeta + ((1-\alpha)N)^{-1}\sum_{k=1}^{N}\eta_k \leq u$

I now apply example 3.3.5 in a special portfolio optimization problem. An institutional investor allocates his wealth between *bonds*, a *commodity index* and a *structured product* on the commodity index. The investment in the commodity index could for example take place in the form of a commodity index of the type as described in section 3.2: One invests in futures of Aluminum, Copper, Brent, Gasoil and Gold where the weightings of the respective commodities are equal to the Markowitz-weightings.

It is now our task to optimize the weightings of the three assets in the sense that the expected mean of the portfolio is maximized with the constraint that the 0.95-CVaR is smaller than fixed bound values. The result shall be an efficient frontier.

I assume that the structured product is offered by a bank especially for the institutional investor and is not traded on the market. Thus no market values for the custom made product is available. The derivative consists of a guaranteed sum and a participation in the upside potential of the commodity index. The participation is realized by buying so called hindsight options which have the following payoffs at the end of the investment horizon

$$\max\left(\xi - S_0, 0\right)$$

where ξ is defined as the maximum of the commodity index with respect to yearly observation points and S_0 denotes the commodity value at beginning (At-the-Money Option). The participation rate k is derived as follows: Calculate the difference between the guaranteed sum and discounted guaranteed sum. The difference is then divided by the simulated option price of the hindsight option. That implies that the bank sells the hindsight option for the "fair" price of the option. It is clear that in practice the participation rate k is reduced so that the bank can get a fee (and can reduce the losses which can occur due to model misspecification). But in our example k is not reduced. The specific payoff structure is taken from Martellini, Simsek and Goltz [MSG05] who consider stock options.

I give an example for the payoff-structure of the derivative: Imagine we offer a guaranteed sum of \$ 100. So at initial date an investor has to pay \$ 100. The investment horizon is set to 10 years. Furthermore, I assume that a zero-coupon bond with maturity 10 years has the price 0.6 and a "fair" price of the hindsight option with the same maturity is \$ 60. Then the calculation of the participation rate reads as follows:

$$k = \frac{100 - 100 * 0.6}{60} = \frac{2}{3}.$$

That means at initial date \$ 60 is spent in bonds and \$ 40 in the hindsight option which corresponds to buying a 2/3 hindsight option. After ten years the investor gets the \$ 100 from the investment in bonds and in addition the payoff of the hindsight option.

To start with the optimization I first have to describe the models for the assets. The short rate r is assumed to be constant over the whole investment horizon.

Under the physical measure the commodity index S is modeled as a Heston model which we already came across in chapter 2. For simplicity the stochastic convenience yield is dropped

$$\begin{aligned} dS(t)/S(t) &= \mu^S(t)dt + \sqrt{V(t)}dW^S(t) \\ dV(t) &= \kappa\left(\theta - V(t)\right)dt + \sigma_V\sqrt{V(T)}dW^V(t) \end{aligned}$$

The drift μ^S is expressed in terms of the market prices of risk of the commodity index λ_S:

$$\mu^S(t) = r + \sqrt{V(t)}\lambda_S$$

The Brownian motions $W^S(t)$ and $W^V(t)$ are correlated with correlation ρ_{SV}.

To derive a fair value of the hindsight option I further have to model the commodity index under a risk-neutral measure. Facing an incomplete market I assume that we can switch from the true measure to a risk-neutral measure by substituting the drift of the index μ_S by the short rate r.

After simulating I first have to calculate the participation rate k, which is done in the following steps:

Table 3.10.: Model Parameters for the Portfolio Optimization Problem

Parameter	Stochastic Volatility	Constant Volatility
Investment Horizon	10 years	10 years
Guaranteed Sum	100	100
Initial Value Commodity Index $S(0)$	100	100
Correlation Index-Variance ρ_{SV}	0.7	0.7
Market Price of Risk λ_S	0.2	0.2
Initial Value Volatility $V(0)$	0.16	0.16
Speed of Mean Reversion κ	40%	500%
Long-Run Mean θ	0.16	0.16
Volatility of Volatility σ_V	50%	50%
Short Rate r	4%	4%

1. Derive the average of the payoff of the hindsight option at maturity with respect to the risk-adjusted commodity index process and multiply it by the discounting factor.
2. The difference between guaranteed sum and the guaranteed sum multiplied by the discounting factor is divided by the average of the hindsight option.

The returns of the commodity index and the structured product are given by the general formula

$$\text{Return} := \frac{\text{end value}}{\text{initial value}} - 1$$

where the initial value is known at beginning and the end value depends on the simulation, that is, it is random. For the *commodity index return* the commodity index is observed under the physical measure. For the nonrandom *bond return* one has

$$\text{Bondreturn} = \frac{1}{\exp(-rT)} - 1 = \exp(rT) - 1$$

The end value of the *structured product* consists of the guaranteed sum which results from the investment in bonds and the payoff of the hindsight option under the physical measure

$$\text{StrReturn} = \frac{\text{Guaranteed Sum + Participation rate * Payoff Hindsight}}{\text{Guaranteed Sum}} - 1.$$

I do not annualize the returns. The mean of the returns of the assets can be found in tables 3.11 in the case of the stochastic volatility model and in table 3.12.

The simulation is conducted via discretizing the stochastic differential equations for the commodity index, the risk-adjusted commodity index and the variance process for the commodity

Figure 3.10.: Results of an Implementation of Example 3.3.5

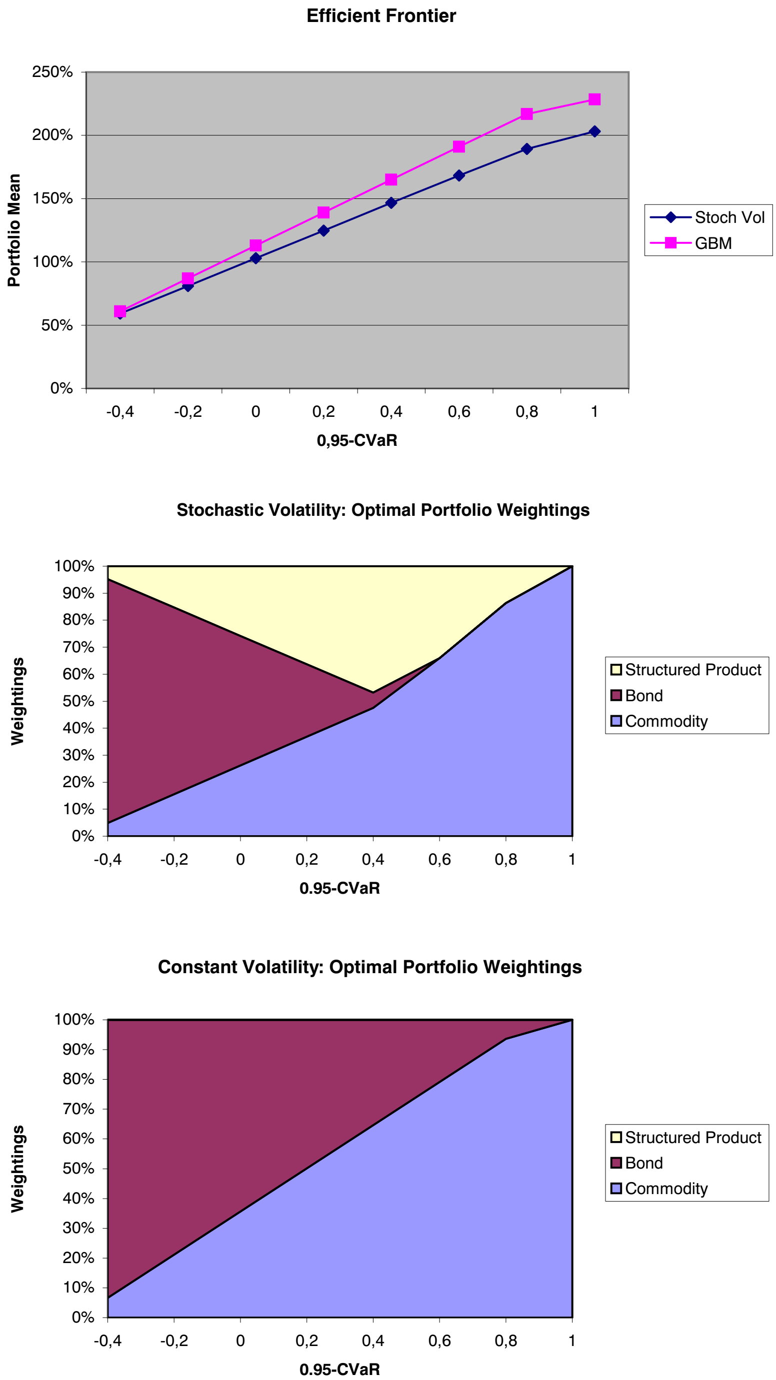


Table 3.11.: Stochastic Volatility: 10 Years Mean Returns of Assets

Asset	Mean Return
Commodity Index	203%
Bond	49%
Structured Product	101%

Table 3.12.: Constant Volatility: 10 Years Mean Returns of Assets

Asset	Mean Return
Commodity Index	228%
Bond	49%
Structured Product	103%

index. I use the Euler scheme. So for example the discretized variance process reads as

$$V(t+\Delta t) = V(t) + \kappa\left(\theta - V(t)\right)\Delta t + \sigma_V \sqrt{V(t)}\sqrt{\Delta t}\mathcal{N}(0,1)$$

To obtain stable results (that is, portfolio weightings which do not depend on the random seed) I set the number of simulations to $25,000$. For the simulation of the standard independent normal random variables I have used the algorithms `ran1` and `gasdev` which can be found in Press et al. [PTVF02]. The Brownian motions of the commodity index and volatility are correlated. To realize the correlations I have applied a *Cholesky decomposition*. Since it a very useful formula I here provide the decomposition for three processes, also I only need the two-dimensional decomposition.

Proposition 3.3.4. *Let X_1, X_2, X_3 be three independent normal random variables and ρ_{12}, ρ_{13}, ρ_{23} be in $[-1,1]$. Then Y_1, Y_2 and Y_3 defined by*

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= \rho_{12}X_1 + \sqrt{1-\rho_{12}^2}X_2 \\ Y_3 &= \rho_{13}X_1 + \frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}}X_2 + \frac{\sqrt{1+2\rho_{12}\rho_{13}\rho_{23}-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2}}{\sqrt{1-\rho_{12}^2}}X_3 \end{aligned}$$

are correlated with $Corr(Y_i, Y_j) = \rho_{ij}$ (for $1 \le i < j \le 3$).

I have implemented the simulation in C++. Furthermore, I have to solve a linear programming problem with thousands of variables as described in example 3.3.5. To handle this, I included the

Table 3.13.: Stochastic Volatility: Optimal Weightings

Assets			Portfolio		
Commodity Index	Bonds	Struct. Product	10 Years Mean	0.95−VaR	0.95−CVaR
0%	100%	0%	49%	-0.49	-0.49
5%	90%	5%	59%	-0.40	-0.40
16%	69%	15%	81%	-0.21	-0.20
26%	48%	26%	103%	-0.02	0.00
37%	27%	36%	125%	0.17	0.20
48%	6%	47%	147%	0.36	0.40
66%	0%	34%	168%	0.55	0.60
86%	0%	14%	189%	0.74	0.80
100%	0%	0%	203%	0.94	1.00

LP_SOLVE C++-File. To use this tool I have to assign the linear programming problem in a text file (saved with file extension .lp instead of .txt) where the given syntax of the LP_SOLVE has to be kept. To this end, I have programmed a C++ file which samples the simulated returns and the other relevant input parameters and exports these values in the above mentioned text file. To have a convenient user interface the input parameters and results are handled in EXCEL/VBA.

Table 3.10 records the parameters of the model. The correlation of the Brownian motions of the commodity index and the volatility is set to a positive value, namely $\rho_{SV} = 0.7$. A positive correlation is a typical feature in commodity markets. Typically a negative correlation is observed in stock markets $\rho_{SV} \approx -0.7$ to -0.9. The initial value of volatility $V(0)$ and the long run mean of the volatility θ are set to 0.16. This corresponds to a volatility of the instantaneous commodity index returns of $\sqrt{0.16} = 40\%$ per year. The volatility of volatility amounts to $20\% = 50\% * \sqrt{0.16}$ per year. So we can conclude that the volatility value of the commodity return is expected to fluctuate between $36\% - 44\%$.

The optimal portfolio weightings are presented in tables 3.14 and 3.13. In addition I show the portfolio mean, the 0.95−VaR and 0.95−CVaR. We see that the 0.95−VaR is always smaller than the 0.95−CVaR. This was already clear from the beginning due to the definitions. Remember that the loss of the portfolio was defined as the negative expected portfolio return. So the values of the 0.95−VaR in tables 3.14 and 3.13 are equal to the 5%−worst portfolio return and the values of the 0.95−CVaR represent the average of the 5%−worst portfolio returns. In the case of a full investment of the nonrandom bond it is clear that 0.95−VaR and 0.95−CVaR coincide and

Table 3.14.: Constant Volatility: Optimal Weightings

Assets			Portfolio		
Commodity Index	Bonds	Struct. Product	10 Years Mean	0.95−VaR	0.95−CVaR
0%	100%	0%	49%	-0.49	-0.49
7%	93%	0%	61%	-0.40	-0.40
21%	79%	0%	87%	-0.21	-0.20
36%	64%	0%	113%	-0.02	0.00
50%	50%	0%	139%	0.17	0.20
65%	35%	0%	165%	0.36	0.40
79%	21%	0%	191%	0.55	0.60
94%	6%	0%	217%	0.74	0.80
100%	0%	0%	228%	0.92	1.00

are equal to the negative portfolio mean $-0*203\%-1*49\%-0*101\% = -49\% = -\exp(0.04*10)$.

A very restrictive 0.95−CVaR (for example equal to −0.49) corresponds to a very risk-averse investor. This value can only be preserved with a full investment in riskless bonds. A risk-seeking investor (for example 0.95−CVaR= 0.4) should invest all his money in commodities. An investor who wishes to have a balance between risk and return (e.g. 0.95−CVaR=0.20) is recommended to put 37% in the commodity index, 27% in bonds and 36% in the structured product in the stochastic volatility model. In the constant volatility model exactely half of the money should be put in bonds and the other half in the commodity index.

The crucial and quite interesting observation is that in the stochastic volatility model an investment in the structured product is recommendable for all nonextreme 0.95−CVaR values where in the constant volatility model the product does not appear in any optimal portfolio. The structured product is compounded of the riskless bonds and the hindsight option which pays off more if the path of the commodity option exhibits at least one high peak at one observation point. The index is expected to fluctuate more in the stochastic volatility model than in the constant one. Even if the index return at the end of the investment horizon has performed not overwhelming, the hindsight option nonetheless has in the stochastic volatility case a higher probability to deliver a very good performance. So the hindsight option presents definitely a complementary investment possibility which is not the case in the constant volatility model. This holds also when one changes the correlation of the Brownian motions to −0.7. A negative correlation is typical in stock markets. The optimization results are substantially the same.

Chapter 4

Excursions and Statistical Arbitrage

4.1. Excursions

The goal of this section is to apply the mathematical concept of excursions for the creation of optimal highly automated or algorithmic trading strategies. The idea is to consider the gain of the strategy and the excursion time it takes to realize the gain.

In this section I present formulas for the Ornstein-Uhlenbeck process. I show that the corresponding formulas can be calculated quite fast since the only function appearing in the formulas is the so called imaginary error function. This function is already implemented in many programs, such as in Maple. Algorithmic trading is widely used by pension funds, mutual funds, institutional traders and hedge funds. In this high frequency trading computers make the decision to initiate orders based on information that is received electronically.

My main contribution to this topic is the optimization of the trading strategy via the Banach fixed-point theorem.

Example 4.1.1. *A famous and widely used algorithmic trading strategy is the so called momentum trading (see for example Risk [Mad08]). The construction of momentum strategies and the mentioning of the efficiency of these strategies in stock markets first appeared in Jegadeesh and Titman [JT93]. The (basic) idea behind these kind of strategies is the observation that high performing stocks have an upward trend for a specific time horizon in the future. In addition assets with a low performance exhibit lower performance in the future. So the strategy consists of a short-selling of loser assets and buying of winner assets. I now give a precise description* *which is extracted from Rouwenhorst [Rou98].*

At the end of each month, all stocks are ranked into deciles based on their past J-month return (J equals 3, 6, 9, or 12) and assigned to one of ten relative strength portfolios (1 equals lowest past performance, or loser, 10 equals highest past performance or winner). These portfolios are equally weighted at formation, and held for K subsequent months (K equals 3, 6, 9, or 12 months) during which time they are not rebalanced. The holding period exceeds the interval over which return information is available (monthly), which creates an overlap in the holding period returns. The paper of Rouwenhorst [Rou98] follows the article of Jegadeesh and Titman [JT93] who report the monthly average return of K strategies, each starting one month apart. This is equivalent to a composite portfolio in which each month $1/K$ of the holdings are revised. For example, toward the end of month t the $J = 6$, $K = 3$ portfolio of winners consists of three parts: a position carried over from an investment at the end of month $t-3$ in the 10 percent of firms with highest prior six-month performance as of $t-3$, and two similar positions resulting from an investment in the top-performing firms at the end of months $t-2$ and $t-1$. At the end of month t, the first of these holdings will be liquidated and replaced with an investment in the stocks with highest six-month performance as of time t. The amount of cash available at the beginning of each holding period defines the size of investment in the winner stocks. At the same time, the loser stocks are sold short for exactly the same amount, leading to a zero-cost net investment.

In the following I denote the value of the long-short portfolio as $(X(t))_{t\geq 0}$. I enter the trade when the value equals $X(t) = a$ and exit the trade at $X(t) = b > a$, and waiting until the process returns to $X(t) = a$, to complete the trading cycle. Since $X(t)$ is a stochastic process, the time taken to complete the trade cycle are a sequence of random variables henceforth called *sequence of excursions of value a via value b*. So the values a and b are trading signals.

I am interested in considering the expected gain of a strategy *per unit time*. Since process X models the value of a long-short portfolio the gain of each trade cycle amounts to $b-a-z$ where $z \in \mathbb{R}_0^+$ denotes the transaction cost per trade cycle. Note that the values of the trading cycle length are independent and identically distributed random variables if process X is a stationary strong Markov process. The expected value of the gain for the strategy per unit time is then given by

$$\mathbb{E}\left[R(a,b,z)\right] = \mathbb{E}\left[\frac{b-a-z}{\mathcal{T}}\right] = (b-a-z)\,\mathbb{E}\left[\frac{1}{\mathcal{T}}\right]$$

By applying the inequality of Jensen for convex functions one can write

$$\mathbb{E}\left[1/\mathcal{T}\right] \geq 1/\mathbb{E}\left[\mathcal{T}\right]$$

So $(b-a-z)/\mathbb{E}\left[\mathcal{T}\right]$ is a lower bound for the expected gain per unit time.

Definition 4.1.1. *Given a continuous process $(X(t))_{t\geq 0}$ with state space I the first passage time that it hits b after a is defined by*

$$\begin{aligned}\mathcal{T}_a &= \inf\{t \geq 0 : X(t) = a\} \\ \mathcal{T}_{a;b} &= \inf\{t \geq \mathcal{T}_a : X(t) = b\}\end{aligned}$$

for all $a, b \in I$. If a or b is never reached, we set $\inf\{\emptyset\} = \infty$.

When calculating moments of random variables the Laplace transform is a convenient tool. The following examples for not necessarily stationary processes illustrate that the expected trading cycle length can become infinite. I consider three basic processes: the Brownian motion, Brownian motion with drift and geometric Brownian motion.

Definition 4.1.2. *(see Chung [Chu01] page 197) Let $\mathcal{T}$ be a positive (≥ 0) random variable having the density function F so that F has support in $[0,\infty[$, namely $F(0-) = 0$. The Laplace transform of $\mathcal{T}$ or F is the function $\mathcal{L}$ on $\mathbb{R}^+ = [0,\infty[$ given by*

$$\mathcal{L} : \alpha \to \mathbb{E}\left[e^{-\alpha\mathcal{T}}\right] = \int_0^\infty e^{-\alpha x} dF(x).$$

It is obvious that

$$\mathcal{L}(0) = \lim_{\alpha\searrow 0} \mathcal{L}(\alpha) = 1, \quad \text{and} \quad \mathcal{L}(\infty) = \lim_{\alpha\to\infty} \mathcal{L}(\alpha) = F(0).$$

Proposition 4.1.1. *(Laplace transform as moment generating function) For $\mathcal{T}$ being a positive (≥ 0) random variable the first moment can be calculated with the help of the Laplace transform*

$$\mathbb{E}\left[\mathcal{T}\right] = -\lim_{\alpha\to 0+} \frac{d\mathcal{L}(\alpha)}{d\alpha} \tag{4.1}$$

Proof: Applying the dominated convergence theorem one can interchange limes and integral. ■

In the following $\mathbb{P}^a$ indicates the probability measure corresponding to a Markov process with initial position $a \in \mathbb{R}$. $\mathbb{E}^a$ means taking the expectation with respect to the probability measure $\mathbb{P}^a$. A rigorous introduction can be looked up in Karatzas and Shreve [KS00] on page 74 and 75.

Example 4.1.2. *In the case of a Brownian motion* $(W(t))_{t\geq 0}$ *starting at* $W(0)=0$ *the Laplace transform of* $\mathcal{T}_b$ *for* $b>0$ *takes the form*

$$\mathcal{L}(\alpha)=\mathbb{E}^0\left[e^{-\alpha\mathcal{T}_b}\right]=e^{-b\sqrt{2\alpha}}$$

(see Karatzas and Shreve [KS00], Remark 2.8.3). Differentiating the Laplace transform yields

$$-\frac{d\mathcal{L}(\alpha)}{d\alpha}=-e^{-b\sqrt{2\alpha}}(-b)\frac{1}{\sqrt{2\alpha}}\overset{\alpha\to 0+}{\longrightarrow}\infty.$$

Furthermore, it holds $\mathbb{P}^0\left[\mathcal{T}_b<\infty\right]=1$ *(see again Karatzas and Shreve [KS00], Remark 2.8.3). So the probability of a hit of the barrier b is unity. But very long excursions away from the barrier can occur. For illustrative purposes the expectation can be approximated by the sum of the products of the hitting times and their probabilities. It diverges. The reasoning for this is that the probabilities do not fall sufficiently fast for longer hitting times. So we shall not model the long-short portfolio as a Brownian motion.*

For arbitrary a,b *we obtain*

$$\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b}\right]=e^{-|a-b|\sqrt{2\alpha}}$$

(see Borodin and Salminen [BS02], page 198). The expected time length of the first passage time $\mathcal{T}_b$ *for arbitrary* $a\neq b$ *is infinite, too.*

Example 4.1.3. *The Laplace transform of the first passage time for the Brownian motion with drift, namely* $W^\mu(t)=\mu t+W(t)$ $(\mu\in\mathbb{R})$ *can be found in Borodin and Salminen [BS02] on page 295*

$$\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b}\right]=\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b};\ \mathcal{T}_b<\infty\right]=e^{\mu(b-a)-|b-a|\sqrt{2\alpha+\mu^2}}.$$

The first equality follows from the fact that if $\mathcal{T}_b=\infty$ *it follows* $e^{-\alpha\mathcal{T}_b}=0$ *and therefore* $\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b}\right]=\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b};\ \mathcal{T}_b<\infty\right]$*, the qualification under the expectation signifying that the contributions to the expectation occur only for the sample paths over which* $\mathcal{T}_b$ *is finite.*

Differentiating the Laplace transform for $\mu\neq 0$ *with respect to* α *gives*

$$\begin{aligned}-\frac{d\mathbb{E}^a\left[e^{-\alpha\mathcal{T}_b}\right]}{d\alpha} &= -e^{\mu(b-a)-|b-a|\sqrt{2\alpha+\mu^2}}(-|b-a|)\frac{1}{\sqrt{2\alpha+\mu^2}}\\ &\overset{\alpha\to 0+}{\longrightarrow} e^{\mu(b-a)-|\mu||b-a|}\frac{|b-a|}{|\mu|}\end{aligned}$$

So for $b>a$ *we obtain*

$$\mathbb{E}^a\left[\mathcal{T}_b\right]=\begin{cases}\frac{b-a}{\mu} & : \ \mu>0\\ \infty & : \ \mu\leq 0\end{cases}$$

and for $b < a$

$$\mathbb{E}^a[\mathcal{T}_b] = \begin{cases} \frac{a-b}{|\mu|} & : \quad \mu < 0 \\ \infty & : \quad \mu \geq 0 \end{cases}$$

Note that for special values of μ, a and b (for example $\mu = 1$, $b = 0.2$ and $a = 0.3$) the probability that the Brownian motion with drift starting in a never reaches level b is strictly positive (see Borodin and Salminen [BS02] page 295):

$$\mathbb{P}^a[\mathcal{T}_b = \infty] = 1 - e^{\mu(b-a)-|\mu||b-a|}$$

So the Brownian motion with drift does not seem to be an interesting model for calculating the mean length of the trading cycle.

Example 4.1.4. *The problem appearing in the previous two examples also occurs in the case of the geometric Brownian motion $dX(t)/X(t) = \mu dt + \sigma dW(t)$ ($\sigma > 0$): The following result is taken from Wilmott, Dewynne and Howison [Wil98] on page 371: For $b > a$*

$$\mathbb{E}^a[\mathcal{T}_b] = \begin{cases} \frac{1}{\mu - 0.5\sigma^2}\ln(b/a) & : \quad \mu > 0.5\sigma^2 \\ \infty & : \quad \mu \leq 0.5\sigma^2 \end{cases}$$

and for $b < a$

$$\mathbb{E}^a[\mathcal{T}_b] = \begin{cases} \frac{1}{0.5\sigma^2 - \mu}\ln(a/b) & : \quad \mu < 0.5\sigma^2 \\ \infty & : \quad \mu \geq 0.5\sigma^2 \end{cases}$$

Remark 4.1.1. *I will now show that the Ornstein-Uhlenbeck mean-reveartìng process exhibits the desired feature $\mathbb{E}^a[\mathcal{T}_{b;a}] < \infty$. The drawback of the Brownian motion with drift and the geometric Brownian motion is the fact that for a (for example) positive drift the mean length to reach level $b > a > 0$ is finite, but the mean length of the recurrence to the smaller value a is infinite.*

I now apply the Ornstein-Uhlenbeck process for the long-short portfolio $(X(t))_{t \geq 0}$. We already came across the process in lemma 1.2.1:

$$dX(t) = -\theta X(t)dt + \sigma dW(t). \tag{4.2}$$

with $\sigma > 0$ (to ovoid divisions by zero in the following calculations). I assume that $\theta > 0$. Then the process is stationary (see Borodin and Salminen [BS02] pages 136-137). If $\theta < 0$ the process is not stationary but transient. The Ornstein-Uhlenbeck process is the oldest example of a stochastic differential equation (Uhlenbeck and Ornstein [UO30]).

The process has the entire real line $]-\infty,\infty[$ as its state space (Karlin and Taylor [KT81] page 170 C). This feature is well-suited for our purpose of modeling a long-short portfolio taking on both negative and positive values. As mentioned in proposition 1.2.1 the infinitesimal drift parameter reflects a restoring force directed towards the origin and of a magnitude proportional to the distance.

As calculated in proposition 1.2.1 the solution of 4.2 is

$$X(t) = X_0 e^{-\theta t} + \sigma \int_0^t e^{-\theta(t-s)} dW(s); \quad 0 \le t < \infty$$

If $\mathbb{E}\left[X_0^2\right] < \infty$, the expectation and variance are given by

$$m(t) := \mathbb{E}\left[X(t)\right] = m(0)e^{-\theta t} \tag{4.3}$$

$$V(t) := \mathbb{V}\left[X(t)\right] = \frac{\sigma^2}{2\theta} + \left(V(0) - \frac{\sigma^2}{2\theta}\right) e^{-2\theta t} \tag{4.4}$$

If the initial random variable $X(0)$ has a normal distribution with mean zero $m(0) = 0$ and variance $V(0) = \sigma^2/(2\theta)$, then X is a stationary process, that is, the distribution of $(X(t + t_1), X(t + t_2), ..., X(t + t_k))$ for $0 \le t_1 < t_2 < ... < t_k < \infty$ does not depend on t (Karatzas and Shreve [KS00] Example 5.6.8). In fact, if $m(0) = 0$ and variance $V(0) = \sigma^2/(2\theta)$ then the time dependence of the mean and the variance in equations 4.3 and 4.4 cancels out.

Proposition 4.1.2. *The Ornstein-Uhlenbeck process specified by 4.2 is a stationary strong Markov process.*

Proof: Follows directly from Friedman [Fri75], Theorem 3.4 on page 112 and Karatzas and Shreve [KS00] Example 5.6.8. ■

Definition 4.1.3. *An $\mathbb{R}-$valued stochastic process $(X(t))_{t\ge 0}$ is called Gaussian if, for any integer $k \ge 1$ and real numbers $0 \le t_1 < t_2 < ... < t_k < \infty$, the random vector $(X_{t_1}, X_{t_2}, ..., X_{t_k})$ has a joint normal distribution.*

Proposition 4.1.3. *Let X be given by the linear stochastic differential equation*

$$dX(t) = (A(t)X(t) + a(t))\, dt + \sigma(t)dW(t), \quad X(0) = \xi,$$

where W is a Brownian motion independent of the initial vector ξ, and $A(t)$, $a(t)$ and $\sigma(t)$ are nonrandom, measurable, and locally bounded.

If $X(0)$ has a normal distribution, then X is a Gaussian process.

Proof: Karatzas and Shreve [KS00] page 355. ■

Corollary 4.1.1. *The Ornstein-Uhlenbeck process is a Gaussian process.*

Proof: follows immediately from proposition 4.1.3. ■

In the following calculation of the mean time of the trading cycle the *speed measure* and the *scale function* are very valuable functions. Let X be a solution of the one-dimensional stochastic differential equation

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t) \tag{4.5}$$

with Borel-measurable coefficients $b : I \to \mathbb{R}$ and $\sigma : I \to \mathbb{R}$ with $I =]l, r[$ and $-\infty \leq l < r \leq \infty$. I require that

- for all $x \in \mathbb{R}$ it holds that $\sigma^2(x) > 0$ and
- for all $x \in \mathbb{R}$ exists an $\varepsilon > 0$ such that

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(y)|dy}{\sigma^2(y)} < \infty$$

We fix an arbitrary number $c \in I$ and define for $l < x < r$ the *scale function* S by

$$\begin{aligned} s(\eta) &:= \exp\left\{ -\int_c^\eta \frac{2b(\phi)}{\sigma^2(\phi)} d\phi \right\} \\ S(x) &:= \int_c^x s(\eta) d\eta \end{aligned}$$

and the speed density m for $x \in I$ by

$$m(x) = 2/\left(\sigma^2(x)s(x)\right)$$

Remark 4.1.2. *The finiteness (or nonfiniteness) of* $\lim_{x\to+\infty} S(x)$ *or* $\lim_{x\to-\infty} S(x)$ *do not depend on the choice of c (Karatzas and Shreve [KS00] page 339 5.5.12). In the following applications of the scale function only differences of the scale function do appear so that the we can choose an arbitrary c.*

Proposition 4.1.4. *Let X be a solution to*

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t) \tag{4.6}$$

with state space $I =]l, r[$ *where* $-\infty \leq l < r \leq \infty$ *and with Borel-measurable coefficients* $b : I \to \mathbb{R}$ *and* $\sigma : I \to \mathbb{R}$ *with* $I =]l, r[;\ -\infty \leq l < r \leq \infty$ *and let the coefficients satisfy* $\sigma(x) > 0$ *for all* $x \in I$ *and it holds for all* $x \in I$ *there exists an* $\epsilon > 0$ *such that*

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(y)|}{\sigma^2(y)} dy < \infty.$$

Furthermore, let b and σ *be bounded on compact subintervals of* I *and let the scale function* $S(x)$ *and the speed measure* $m(x)$ *satisfy*

$$p(l+) = -\infty \quad and \quad p(r-) = \infty \quad and \quad m(I) < \infty \tag{4.7}$$

where $m(I) := \int_l^r m(x)dx$. *In addition we assume that for all* $x, z \in I$ *and* $t > 0$

$$\mathbb{P}^x [X(t) = z] = 0$$

Then the solution X *to 4.5 starting at* $x \in I$ *never exits* I. *The process is positive recurrent with* $l < a < x < u < b < r$:

$$\begin{aligned}
\mathbb{E}^x [\mathcal{T}_a] &= -\int_a^x (S(x) - S(y))\, m(dy) + (S(x) - S(a))\, m\,(]a, r[) < \infty, \\
\mathbb{E}^x [\mathcal{T}_b] &= -\int_x^b (S(y) - S(x))\, m(dy) + (S(b) - S(x))\, m\,(]l, b[) < \infty, \\
\mathbb{E}^a [\mathcal{T}_{b;a}] &= \mathbb{E}^a [\mathcal{T}_b] + \mathbb{E}^b [\mathcal{T}_a] = (S(u) - S(x))m(I) < \infty.
\end{aligned}$$

Proof: see Pollak and Siegmund [PS85] or Karatzas and Shreve [KS00] pages 352-353 ∎

Remark 4.1.3. *It is clear that a positively recurrent process is recurrent in the sense that for all* $x, y \in I$ *it holds*

$$\mathbb{P}^x [\mathcal{T}_y < \infty] = 1$$

since otherwise the expectations appearing in proposition 4.1.4 would not be finite. A positively recurrent process is a regular process, that is, for all interior points x *and* y *of* I *it holds*

$$\mathbb{P}^x [\mathcal{T}_y < \infty] > 0.$$

I now recall the definition of the imaginary error function.

Definition 4.1.4. *For all* $x \in \mathbb{R}$ *the function* $\mathrm{Erfi}(x)$ *is defined by*

$$\mathrm{Erfi}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{t^2} dt = \frac{2}{\sqrt{\pi}} \sum_{k=0}^{\infty} \frac{x^{2k+1}}{k!(2k+1)}$$

Corollary 4.1.2. *Let the Ornstein-Uhlenbeck process X be given by the stochastic differential equation 4.2 with $\sigma > 0$ and $\theta > 0$. The state space is $I =]-\infty, \infty[$. Let $a, b \in I$ and $a < b$. Then the expected trading length amounts to*

$$\mathbb{E}^a[\mathcal{T}_{a;b}] = \frac{\pi}{\theta}\left(\text{Erfi}\left(b\sqrt{\theta}/\sigma\right) - \text{Erfi}\left(a\sqrt{\theta}/\sigma\right)\right)$$

Proof: I first check the assumptions of proposition 4.1.4. The coefficients are bounded on compact subintervals of I. Furthermore,

$$\begin{aligned}\int_{x-\varepsilon}^{x+\varepsilon} \frac{1+|b(y)|}{\sigma^2(y)}dy &= \int_{x-\varepsilon}^{x+\varepsilon} \frac{1+|-\theta y|}{\sigma^2}dy \\ &= \int_{x-\varepsilon}^{x+\varepsilon} \frac{1+\theta|y|}{\sigma^2}dy = \frac{2\varepsilon}{\sigma^2} + \frac{\theta}{\sigma^2}\int_{x-\varepsilon}^{x+\varepsilon} |y|dy.\end{aligned}$$

- For $x > 0$ I define $\varepsilon = |x|$. Then it follows that the integral is bounded:

$$\int_{x-|x|}^{x+|x|} ydy = \int_0^{2x} ydy = \frac{1}{2}(2x)^2 = 2x^2 < \infty$$

- For $x < 0$ I define again $\varepsilon = |x|$. Then I obtain:

$$\int_{2x}^{0} -ydy = \int_0^{2x} ydy = 2x^2 < \infty$$

- For $x = 0$ I choose $\varepsilon = 1$. Then

$$\int_{-1}^{1} |y|dy = 1$$

For an arbitrary $c \in I$ the integrand of the scale function takes the form

$$\begin{aligned}s(\eta) &= \exp\left\{-\int_c^\eta \frac{2\mu(\phi)}{\sigma^2(\phi)}d\phi\right\} = \exp\left\{\int_c^\eta \frac{2\theta\phi}{\sigma^2}d\phi\right\} \\ &= \exp\left\{\frac{2\theta}{\sigma^2}\int_c^\eta \phi d\phi\right\} = \exp\left\{\frac{2\theta}{\sigma^2}\frac{1}{2}\eta^2\right\} = \exp\left\{\frac{\theta}{\sigma^2}\eta^2\right\}.\end{aligned}$$

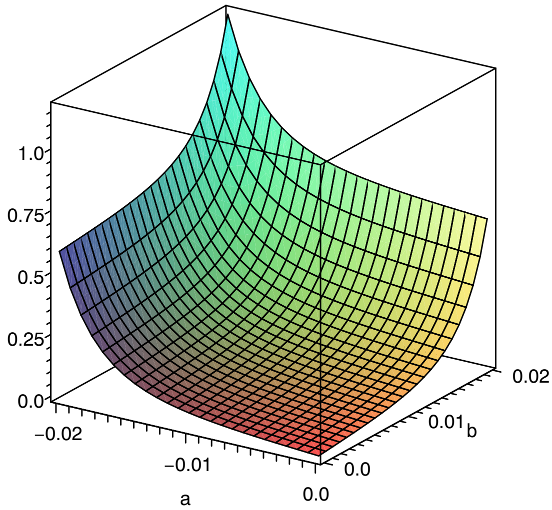
So for $-\infty < x < \infty$ the scale function reads as

$$\begin{aligned}S(x) &= \int_c^x s(\eta)d\eta = \int_c^x \exp\left\{\frac{\theta}{\sigma^2}\eta^2\right\}d\eta \\ &= \frac{\sigma}{\sqrt{\theta}}\int_c^{\frac{\sqrt{\theta}}{\sigma}x} \exp\left\{z^2\right\}dz = \frac{\sigma\sqrt{\pi}}{2\sqrt{\theta}}\text{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}x\right)\end{aligned}$$

Taking the limit to the left and right endpoints, namely $r = +\infty$ and $l = -\infty$, yields that the first two assumptions of 4.7 are fulfilled:

$$S(x) = \frac{\sigma\sqrt{\pi}}{2\sqrt{\theta}}\text{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}x\right) = d * \int_0^{x\frac{\sqrt{\theta}}{\sigma}} e^{\nu^2}d\nu \overset{x\to+\infty}{\to} +\infty$$

Figure 4.1.: Expected Length of Excursion $\mathbb{E}^a[\mathcal{T}_{b;a}]$ of the Ornstein-Uhlenbeck process for $\theta = 100$ and $\sigma = 0.1$



where $d := \frac{\sigma\sqrt{\pi}}{2\sqrt{\theta}} > 0$. Furthermore,

$$\lim_{x\to-\infty} d * \int_0^{x\frac{\sqrt{\theta}}{\sigma}} e^{\nu^2} d\nu = \lim_{x\to\infty} -d * \int_0^{x\frac{\sqrt{\theta}}{\sigma}} e^{\nu^2} d\nu = -\infty.$$

The difference of the scale function evaluated at the trading signal points $S(b) - S(a)$ is given by

$$S(b) - S(a) = \frac{\sigma\sqrt{\pi}}{2\sqrt{\theta}} \left(\text{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \text{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right) \tag{4.8}$$

I am now left with the calculation of the speed measure $m(I)$

$$\begin{aligned} m(x) &= \frac{2}{\sigma^2(x)s(x)} = \frac{2}{\sigma^2 \exp\left\{\frac{\theta}{\sigma^2}x^2\right\}} \\ m(I) &= \int_{-\infty}^{\infty} m(x)dx = \frac{2}{\sigma^2}\int_{-\infty}^{\infty} \exp\left\{-\frac{\theta}{\sigma^2}x^2\right\} dx = \frac{4}{\sigma^2}\int_0^{\infty} \exp\left\{-\frac{\theta}{\sigma^2}x^2\right\} dx \\ &= \frac{4}{\sigma^2}\frac{\sqrt{\pi}\sigma}{2\sqrt{\theta}} = \frac{2\sqrt{\pi}}{\sqrt{\theta}\sigma} < \infty \end{aligned}$$

Since the Ornstein-Uhlenbeck process is a Gaussian process 4.1.1 it follows that $X(t)$ at time $t > 0$ is normally distributed which entails that the set $\{X(t) = z\}$ has probability zero:

$$\mathbb{P}^x[X(t) = z] = 0 \quad \text{for all} \quad x, z \in I \quad \text{and} \quad t > 0.$$

Figure 4.2.: Expected Length of Excursion $\mathbb{E}^a\left[\mathcal{T}_{b;a}\right]$ of the Ornstein-Uhlenbeck process for $\theta = 100$ and $\sigma = 0.3$

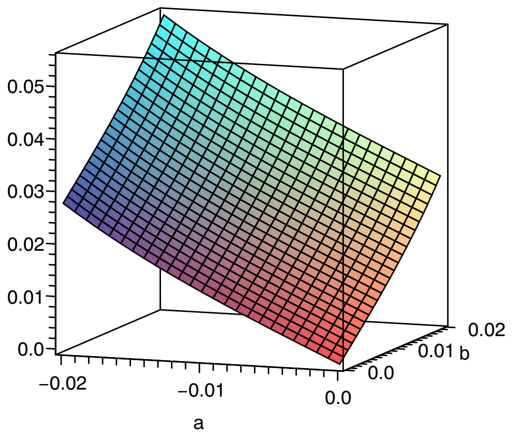
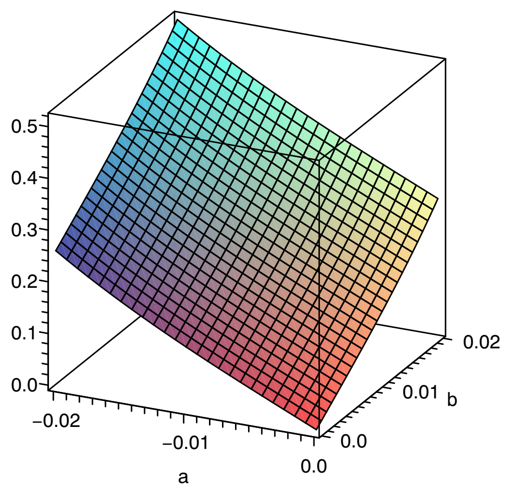


Figure 4.3.: Expected Length of Excursion $\mathbb{E}^a\left[\mathcal{T}_{b;a}\right]$ of the Ornstein-Uhlenbeck process for $\theta = 10$ and $\sigma = 0.1$



So I obtain for the mean trading cycle length

$$\begin{aligned}\mathbb{E}^a\left[\mathcal{T}_{b;a}\right] &= \frac{\sigma\sqrt{\pi}}{2\sqrt{\theta}}\frac{2\sqrt{\pi}}{\sqrt{\theta}\sigma}\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)-\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right)\\ &= \frac{\pi}{\theta}\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)-\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right)\end{aligned}$$

■

The expected excursion length of corollary 4.1.2 is plotted in figures 4.1, 4.2 and 4.3. The mean excursion length seems to decrease with a higher volatility and lower speed of mean reversion.

Definition 4.1.5. *For $a, b \in \mathbb{R}$ with $a < b$ and $\theta > 0$, $\sigma > 0$ and $z \geq 0$ we define*

$$F(a,b) := \frac{\theta\,(b-a-z)}{\pi\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)-\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right)}.$$

Remark 4.1.4. *Next I determine the global maximum of a lower bound of the expected return per unit time with respect to the trading signals a and b. Function $F(a,b)$ is not concave in general as figure 4.4 shows. This fact complicates the determination of the global maximum. Then I calculated the Hesse matrix but this seems too cumbersome (see Appendix B). I solve this problem by applying the Banach fixed-point theorem. The equation delivering the local extreme points is transformed into an equivalent fixed-point equation. The fixed-point is equal to the unique local maximum.*

Lemma 4.1.1. *For $f(x) = 1 - x\sqrt{\pi}\operatorname{Erfi}(x)e^{-x^2}$ it holds*

$$|f(x)| < 1 \quad \textit{for all} \quad x > 0.$$

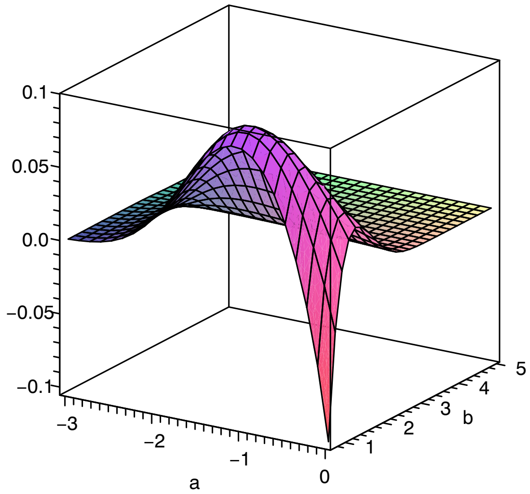
Proof: Let $x > 0$. I can write

$$x\sqrt{\pi}\operatorname{Erfi}(x)e^{-x^2} = x\sqrt{\pi}\frac{2}{\sqrt{\pi}}\int_0^x e^{t^2}dte^{-x^2} = 2x\int_0^x e^{t^2}dte^{-x^2}$$

Since I do not know the antiderivative of function $t \to e^{t^2}$ but of function $t \to te^{t^2}$ I tried that this function comes into play. By the mean value theorem for integration there exists a number $t_0 \in\,]0, x[$ such that

$$\int_0^x te^{t^2}dt = t_0\int_0^x e^{t^2}dt. \tag{4.9}$$

Figure 4.4.: Function F as defined in 4.1.5 for $\theta = \sigma = z = 1$ is not concave



It follows

$$\int_0^x te^{t^2}dt - t_0\int_0^x e^{t^2}dt = \int_0^x (t-t_0)\,e^{t^2}dt = \int_0^{t_0}(t-t_0)\,e^{t^2}dt + \int_{t_0}^x (t-t_0)\,e^{t^2}dt = 0$$

and hence

$$\int_0^{t_0}(t-t_0)\,e^{t^2}dt = -\int_{t_0}^x (t-t_0)\,e^{t^2}dt.$$

Since function $t \to te^{t^2}$ is strictly increasing it holds $t_0 > \frac{x}{2}$. It follows by 4.9

$$\int_0^x te^{t^2}dt > \frac{x}{2}\int_0^x e^{t^2}dt$$

and I obtain

$$\int_0^x e^{t^2}dt < \frac{2}{x}\int_0^x te^{t^2}dt = \frac{1}{x}\left(e^{x^2}-1\right) < \frac{1}{x}e^{x^2}.$$

From this it follows

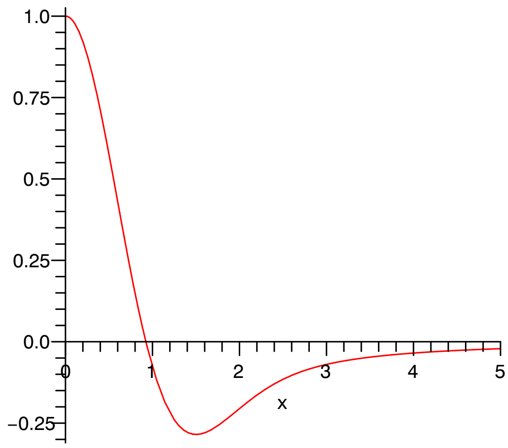
$$0 < x\sqrt{\pi}\,\mathrm{Erfi}(x)e^{-x^2} = 2x\int_0^x e^{t^2}dte^{-x^2} < 2x\frac{1}{x}e^{x^2}e^{x^{-2}} = 2$$

and hence

$$|f(x)| = |1 - x\sqrt{\pi}\,\mathrm{Erfi}(x)e^{-x^2}| < 1.$$

■

Figure 4.5.: Function f of Lemma 4.1.1



Proposition 4.1.5. *Function $F(a,b)$ as defined in definition 4.1.5 has a global maximum in $\left(-\frac{\sigma}{\sqrt{\theta}}u^*, \frac{\sigma}{\sqrt{\theta}}u^*\right)$ where u^* is given by the limit of the iteration sequence*

$$u_{n+1} = \frac{\sqrt{\pi}}{2}\,\mathrm{Erfi}\,(u_n)\,e^{-u_n^2} + \frac{z}{2}\frac{\sqrt{\theta}}{\sigma}, \quad n = 0, 1, 2, 3, \ldots$$

with arbitrary starting value $u_0 > 0$.

Proof: I first determine the set $M = \left\{(a,b) \in \mathbb{R}^2 : \ \mathrm{grad}\, F(a,b) = 0\right\}$. To this end, I define

$$\begin{aligned} f(a,b) &:= \theta\,(b - a - z) \\ g(a,b) &:= \pi\left(\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right). \end{aligned}$$

So the partial derivatives take the following form

$$\begin{aligned} \frac{\partial F(a,b)}{\partial a} &= \frac{-\theta g(a,b) - f(a,b)(-\pi)\frac{2}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}e^{\left(\frac{\sqrt{\theta}}{\sigma}a\right)^2}}{g^2(a,b)} \\ \frac{\partial F(a,b)}{\partial b} &= \frac{\theta g(a,b) - f(a,b)\pi\frac{2}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2}}{g^2(a,b)} \end{aligned}$$

Thus

$$\begin{aligned} -\theta g(a,b) + f(a,b)\pi \frac{2}{\sqrt{\pi}} \frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}a\right)^2} &\stackrel{!}{=} 0 \\ \theta g(a,b) - f(a,b)\pi \frac{2}{\sqrt{\pi}} \frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} &\stackrel{!}{=} 0 \end{aligned}$$

Summing up the above equations and setting to zero it is clear that $M \subseteq \left\{(a,b) \in \mathbb{R}^2 : \ a^2 = b^2\right\}$. Since our strategy only makes sense when $a < b$ I obtain $b > 0$ and $a = -b$.

If $F(a,b)$ has in $(-b^*, b^*)$ a maximum then the intersection curve

$$h(b) := F(-b,b) = \frac{\theta\,(2b - z)}{\pi\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}(-b)\right)\right)} = \frac{\theta\,(2b-z)}{2\pi \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)}$$

possesses a maximum in b^*. Setting

$$h'(b) = \frac{4\theta\pi \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \theta\,(2b-z)\,\frac{4\pi}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2}}{\left(2\pi \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)\right)^2} \stackrel{!}{=} 0.$$

I have to solve

$$\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - (2b - z)\,\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} = 0. \tag{4.10}$$

It seems to be that this equality is not analytically solvable in b, hence I have to determine a solution numerically.

From 4.10 I get

$$\begin{aligned} \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \frac{2b}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} + \frac{z}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} &= 0 \\ \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) + \frac{z}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} &= \frac{2b}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma} e^{\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} \\ \frac{\sqrt{\pi}}{2}\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) e^{-\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} + \frac{z}{2}\frac{\sqrt{\theta}}{\sigma} &= \frac{\sqrt{\theta}}{\sigma}b. \end{aligned}$$

Setting $u := \frac{\sqrt{\theta}}{\sigma}b$ I am facing the equality

$$u = \frac{\sqrt{\pi}}{2}\operatorname{Erfi}(u)\,e^{-u^2} + \frac{z}{2}\frac{\sqrt{\theta}}{\sigma} =: p(u).$$

A fixpoint u^* of $p(u)$ gives a solution $b^* = \frac{\sigma}{\sqrt{\theta}}u^*$ of the equality $h'(b) = 0$. The derivative of $p(u)$ is given by

$$p'(u) = \frac{\sqrt{\pi}}{2}\left(\frac{2}{\sqrt{\pi}}e^{u^2}e^{-u^2} - 2u\,\mathrm{Erfi}(u)e^{-u^2}\right) = 1 - u\sqrt{\pi}\,\mathrm{Erfi}(u)e^{-u^2}.$$

Since $|p'(u)| \leq L < 1$ in each interval $[u_0, \infty[$ with $u_0 > 0$ by lemma 4.1.1, $p(u)$ is a contraction in $[u_0, \infty[$. Note that $b > 0$ and thus $u > 0$.

By the Banach fixed-point theorem a fixpoint u^* of $p(u)$ exists and is unique, hence there exists a unique solution $b^* = \frac{\sigma}{\sqrt{\theta}}u^*$ of the equality $h'(b) = 0$.

If $u_0 > 0$ is any number, then the fixed-point iteration

$$u_{n+1} = \frac{\sqrt{\pi}}{2}\,\mathrm{Erfi}\left(u_n\right)e^{-u_n^2} + \frac{z}{2}\frac{\sqrt{\theta}}{\sigma}, \quad n = 0, 1, 2, 3, ...$$

converges to u^*.

It remains to show that $h(b)$ possesses a global maximum in b^*. The second derivative is given by

$$\begin{aligned}
h''(b) &= \frac{\left(\theta\frac{\sqrt{\theta}}{\sigma}\frac{2}{\sqrt{\pi}}e^{\frac{\theta}{\sigma^2}b^2} - \theta\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}\left(2e^{\frac{\theta}{\sigma^2}b^2} + (2b-z)\frac{2\theta b}{\sigma^2}e^{\frac{\theta}{\sigma^2}b^2}\right)\right)\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2\pi}{\pi^2\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)^4} \\
&- \frac{\left(\theta\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \theta(2b-z)\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)2\frac{\sqrt{\theta}}{\sigma}\frac{2}{\sqrt{\pi}}e^{\frac{\theta}{\sigma^2}b^2}\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)\pi}{\pi^2\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)^4} \\
h''(b) &= \frac{\theta\frac{\sqrt{\theta}}{\sigma}\frac{2}{\sqrt{\pi}}e^{\frac{\theta}{\sigma^2}b^2} - \theta\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}\left(2e^{\frac{\theta}{\sigma^2}b^2} + (2b-z)\frac{2\theta b}{\sigma^2}e^{\frac{\theta}{\sigma^2}b^2}\right)}{\pi\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)^2} \\
&- \frac{\left(\theta\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \theta(2b-z)\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)2\frac{\sqrt{\theta}}{\sigma}\frac{2}{\sqrt{\pi}}e^{\frac{\theta}{\sigma^2}b^2}}{\pi\,\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right)^3}
\end{aligned}$$

Since b^* is a solution of equality 4.10 the second term vanishes for $b = b^*$ and one obtains under the assumption $2b > z$

$$
\begin{aligned}
h''(b^*) &= \frac{\theta\frac{\sqrt{\theta}}{\sigma}\frac{2}{\sqrt{\pi}}e^{\frac{\theta}{\sigma^2}b^{*2}} - \theta\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}\left(2e^{\frac{\theta}{\sigma^2}b^{*2}} + (2b^* - z)\frac{2\theta b^*}{\sigma^2}e^{\frac{\theta}{\sigma^2}b^{*2}}\right)}{\pi \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b^*\right)^2} \\
&= -\frac{\theta\frac{1}{\sqrt{\pi}}\frac{\sqrt{\theta}}{\sigma}\left((2b^* - z)\frac{2\theta b^*}{\sigma^2}e^{\frac{\theta}{\sigma^2}b^{*2}}\right)}{\pi \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b^*\right)^2} < 0
\end{aligned}
$$

If $2b = z$ function $F(-b, b) = h(b)$ is equal to zero since the numerator is equal to zero. In the case $2b < z$ function $F(-b, b) = h(b)$ is negative: The numerator of function F is negative and the denominator (the expected trading cycle length) is greater than zero. It is obvious that function F is strictly positive for $2b > z$.

So I can conclude that $h(b) = F(-b, b)$ has a local maximum in b^* which is a global maximum since b^* is the unique solution of $h'(b) = 0$. ■

Example 4.1.5. *For the choice of $\theta = 1$, $\sigma = 0.2$, $z = 0.1$ and start value $u_0 = 1$ the first values of the iteration sequence*

$$
u_{n+1} = \frac{\sqrt{\pi}}{2}\operatorname{Erfi}(u_n)e^{-u_n^2} + \frac{z}{2}\frac{\sqrt{\theta}}{\sigma} = \frac{\sqrt{\pi}}{2}\operatorname{Erfi}(u_n)e^{-u_n^2} + \frac{1}{4}
$$

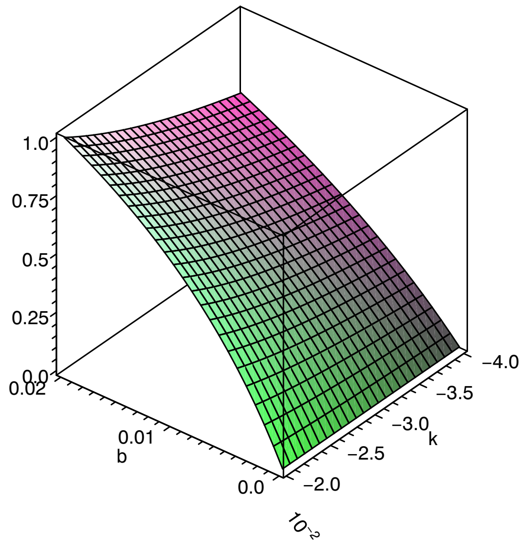
are given by

$$
\begin{aligned}
u_1 &= 0.78810 \\
u_2 &= 0.78025 \\
u_3 &= 0.77895 \\
u_4 &= 0.77870 \\
u_5 &= 0.77865 \\
u_6 &= 0.77860 \\
u_7 &= 0.77860
\end{aligned}
$$

These values are calculated with the aid of the tool Maple. Function $\operatorname{Erfi}(x)$ *is predefined in Maple. In view of proposition 4.1.5 the expected return per unit time*

$$
F(a, b) = \frac{b - a - 1}{\pi\left(\operatorname{Erfi}(b) - \operatorname{Erfi}(a)\right)}
$$

Figure 4.6.: Probability 4.13 for $\theta = 0.5$ and $\sigma = 0.3$ with $a = -b$: The Probability of reaching Lower Value k before Upper Value $b > 0$ with Start Value $-b$



has a global maximum in

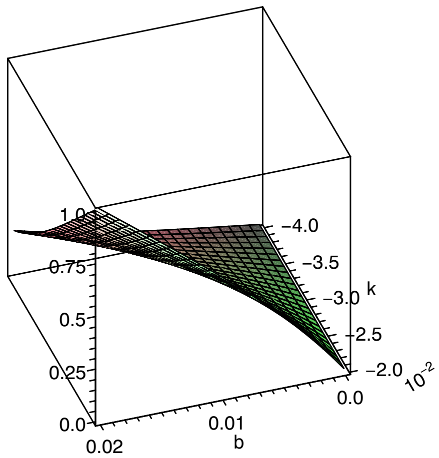
$$\left(-\frac{\sigma}{\sqrt{\theta}}u_6, \frac{\sigma}{\sqrt{\theta}}u_6\right) = (-0.15572, 0.15572)$$

I remark that by using the five-digit floating form the last value u_7 is identical with the global maximum since the input u_6 of the fixed-point iteration sequence coincides with the output u_7.

Remark 4.1.5. *Since the Ornstein-Uhlenbeck process has a state space equal to $I =]-\infty, \infty[$ the process is not bounded from below. So losses can become arbitrarily high. It is clear that traders or investors are forced to quit a non-performing portfolio. So one can modify the strategy: If the portfolio value reaches a level $k < a$ then the portfolio is liquidated and we wait until the portfolio value recurres to level k. Then we start again investing in the long-short portfolio. For liquidation or restart one can assume that one has to pay the constant z for transaction costs. In the following we shall calculate the probabilities of the following events (with $b > 0$ and $k < -b$):*

1. *Starting in $-b$ the process first reaches value b and then value k*
2. *Starting in $-b$ the process first reaches value k and then value b*

Figure 4.7.: Probability 4.13 for $\theta = 200$ and $\sigma = 0.3$ with $a = -b$: The Probability of reaching Lower Value k before Upper Value $b > 0$ with Start Value $-b$



Proposition 4.1.6. *Let X be a solution to*

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t) \tag{4.11}$$

with state space $I =]l, r[$ where $-\infty \leq l < r \leq \infty$ and assume that the coefficients $b : I \to \mathbb{R}$ and $\sigma : I \to \mathbb{R}$ satisfy $\sigma(x) > 0$ for all $x \in I$ and it holds for all $x \in I$ there exists an $\epsilon > 0$ such that

$$\int_{x-\varepsilon}^{x+\varepsilon} \frac{1 + |b(y)|}{\sigma^2(y)} dy < \infty.$$

Then for $k < a < b$ holds

$$\mathbb{P}^a \left[\sup_{0 \leq s \leq T_k} X(s) \geq b \right] = \frac{S(a) - S(k)}{S(b) - S(k)}, \quad \mathbb{P}^a \left[\inf_{0 \leq s \leq T_b} X(s) \leq k \right] = \frac{S(b) - S(a)}{S(b) - S(k)}$$

Proof: see Karatzas and Shreve [KS00] pages 342-344 formula 5.61. ∎

Corollary 4.1.3. *Let X be an Ornstein-Uhlenbeck process specified by stochastic differential* *equation 4.2. Then it holds for $k < a < b$*

$$\mathbb{P}^a\left[\sup_{0\leq s\leq T_k} X(s) \geq b\right] = \frac{\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right) - \mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}k\right)}{\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}k\right)} \quad \textit{and} \tag{4.12}$$

$$\mathbb{P}^a\left[\inf_{0\leq s\leq T_b} X(s) \leq k\right] = \frac{\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)}{\mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \mathrm{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}k\right)} \tag{4.13}$$

Proof: The assumptions of lemma 4.1.6 are checked in corollary 4.1.2. Since I have already calculated the scale function S and the difference $S(x) - S(y)$ for $x > y$ in equation 4.8 I can deduce the assertion. ■

I close this section by summarizing how relevant the provided formulas are in practice. After a calibration of the Ornstein-Uhlenbeck process the optimal trading signals can be determined. Since the portfolio is not bounded from below one is recommended to specify an emergency exit level k. Alternatively, a trader starting in value $-b$ and waiting extremely longer than the expected trading cycle length to reach level b should also exit the trade.

4.2. Statistical Arbitrage

This section reviews a concept of statistical arbitrage as introduced in Hogan, Jarrow, Teo and Warachka [HJTW04], a long horizon trading opportunity that generates a riskless profit and is designed to exploit persistent anomalies.

My contribution is the correction of parts of some proofs in the article [HJTW04]. Furthermore, I supplement the main theorem.

The following results shall provide an investor with a tool to investigate empirically if some strategies (for example momentum strategies) constitute statistical arbitrage opportunities.

Traded in the economy are several assets $\left(S^j(t)\right)_{j=1,2,\ldots,n}$ $(t \geq 0)$ and a money market account $B(t)$ initialized at one Euro ($B(0) = 1$). Let the stochastic processes $(x(t), y(t))_{t\geq 0}$ represent a zero initial cost, self-financing trading strategy involving $x(t) = \left(x^j(t)\right)_{j=1,\ldots,n}$ units of the assets and $y(t)$ units of a money market account at time t. These trading strategies, by definition, must have zero initial cost, $x(0)S^T(0) + y(0) = 0$. In addition, by definition, the trading strategy is self-financing which is described for example in Karatzas and Shreve [KS00] section 5.8 A.

We denote the value process of the portfolio by $(X(t))_{t\geq 0} = \left(x(t)S^T(t) + y(t)B(t)\right)_{t\geq 0}$. In the following we work with the discounted value of the value process $(\nu(t) = X(t)/B(t))_{t\geq 0}$.

For a given trading strategy, let $\Delta\nu(i) = \nu(t_i) - \nu(t_{i-1})$ denote increments of the discounted cumulative trading profit measured at equidistant time points $t_i - t_{i-1} = \Delta$ with $t_i = i\Delta$. The quantity Δ denoting the time between equidistant increments may be set to one without loss of generality (with t_i understood as being i).

Assumption 4.2.1. *(Model Assumption) As in Hogan et al. [HJTW04] on page 536 formula 13 I assume that the discounted cumulative trading profits are given by*

$$\nu(n) = \sum_{i=1}^{n} \Delta\nu(i) \overset{d}{\sim} \mathcal{N}\left(\mu \sum_{i=1}^{n} i^{\theta}, \sigma^2 \sum_{i=1}^{n} i^{2\lambda}\right)$$

and $\Delta\nu(i) = \mu i^{\theta} + \sigma i^{\lambda} z(i)$ *where* $(z(i))_{i=0,1,2,\ldots}$ *are i.i.d.* $\mathcal{N}(0,1)$ *−distributed random variables. Note that* $\mathcal{N}$ *denotes the normal distribution.*

Observe that the process for incremental trading profits is nonstationary when θ or λ is nonzero.

Definition 4.2.1. *(Statistical Arbitrage) As in Hogan et al. [HJTW04] on page 531 Definition 1 I define a statistical arbitrage as a zero-initial cost, self-financing trading strategy* $(x(t))_{t\geq 0}$ *with cumulative discounted value* $(\nu(t))_{t\geq 0}$ *and incremental trading profit* $\Delta\nu$ *such that*

1. $\nu(0) = 0$,
2. $\lim_{t\to\infty} \mathbb{E}[\nu(t)] > 0$,
3. $\lim_{t\to\infty} \mathbb{P}[\nu(t) < 0] = 0$, *and*
4. $\lim_{i\to\infty} \mathbb{V}[\Delta\nu(i)] = 0$ *for* $\Delta\nu(i) < 0$

Note that the fourth condition in 4.2.1 is not taken from the article of Hogan et al. [HJTW04], but from the follow-up article of Jarrow, Teo, Tse and Warachka [JTTW05]. In this article they modified/ improved the original definition of statistical arbitrage.

Proposition 4.2.1. *Under assumption 4.2.1 the following assertions are equivalent*

1. $\lim_{n\to\infty} \mathbb{E}[\nu(n)] > 0$
2. $\mu > 0$.

Proof: $\lim_{n\to\infty} \mathbb{E}[\nu(n)] = \lim_{n\to\infty} \mu \sum_{i=1}^{n} i^{\theta} > 0$ if and only if $\mu > 0$. ■

Remark 4.2.1. *Hogan et al. [HJTW04] proved in theorem 1 on page 537 that the third condition of statistical arbitrage, namely* $\lim_{t\to\infty}\mathbb{P}\left[\nu(t)\right]=0$ *holds if* $\theta > \max\left\{\lambda-\frac{1}{2},-1\right\}$. *For the proof they applied the inequality 25 on page 538*

$$\sum_{i=1}^{n} i^{2\lambda} \leq \int_1^n (s+1)^{2\lambda}\, ds$$

which, however, is generally false for positive and negative λ. *Indeed, setting* $n=2$ *and* $\lambda=-1$ *one obtains*

$$\sum_{i=1}^{2} i^{-2} = \frac{1}{1^2}+\frac{1}{2^2}=1.25$$

whereas

$$\int_1^2 (s+1)^{-2}\, ds = \left[-\frac{1}{s+1}\right]_1^2 = -\frac{1}{3}+\frac{1}{2}=\frac{1}{6}.$$

In addition, I provide a counterexample for positive λ: *Choose* $n=2$ *and* $\lambda=0.25$

$$\sum_{i=1}^{2} i^{0.5} = 1+\sqrt{2}\approx 2.414 \quad \textit{and} \quad \int_1^2 (s+1)^{0.5} ds = \left[\frac{2}{3}(s+1)^{\frac{3}{2}}\right]_1^2 \approx 1.578.$$

In the proof of the following proposition I shall correct this gap. For this reason I prove the following lemma.

Lemma 4.2.1. *It holds:*

1. $$\sum_{i=1}^{n} i^{\alpha} \geq \int_1^n s^{\alpha} ds \quad \textit{for all} \quad \alpha\in\mathbb{R}$$

2. $$\begin{cases} \sum_{i=1}^n i^{\alpha} \leq \int_1^n s^{\alpha} ds + n^{\alpha} & \textit{if} \quad \alpha \geq 0 \\ \sum_{i=1}^n i^{\alpha} \leq \int_1^n s^{\alpha} ds + 1 & \textit{if} \quad \alpha < 0 \end{cases}$$

3. $$\sum_{i=1}^{n} i^{\alpha} \leq \int_1^n s^{\alpha} ds + 1 + n^{\alpha} \quad \textit{for all} \quad \alpha\in\mathbb{R}$$

Proof: If $\alpha \geq 0$ then i^{α} is monotone increasing and it holds

$$\int_1^n s^{\alpha} ds = \sum_{i=1}^{n-1}\int_i^{i+1} s^{\alpha} ds \leq \sum_{i=1}^{n-1}(i+1)^{\alpha} = \sum_{i=2}^{n} i^{\alpha} \leq \sum_{i=1}^{n} i^{\alpha}$$

and

$$\int_1^n s^{\alpha} ds = \sum_{i=1}^{n-1}\int_i^{i+1} s^{\alpha} ds \geq \sum_{i=1}^{n-1} i^{\alpha},$$

hence

$$\int_1^n s^\alpha ds + n^\alpha \geq \sum_{i=1}^n i^\alpha.$$

If $\alpha < 0$ then i^α is monotone decreasing and one has

$$\int_1^n s^\alpha ds = \sum_{i=1}^{n-1} \int_i^{i+1} s^\alpha ds \leq \sum_{i=1}^{n-1} i^\alpha \leq \sum_{i=1}^{n} i^\alpha$$

and

$$\int_1^n s^\alpha ds = \sum_{i=1}^{n-1} \int_i^{i+1} s^\alpha ds \geq \sum_{i=1}^{n-1} (i+1)^\alpha = \sum_{i=2}^{n} i^\alpha,$$

hence

$$\int_1^n s^\alpha ds + 1 \geq \sum_{i=1}^n i^\alpha.$$

This proves assertion 1 and 2. Assertion 3 follows directly from 2. ■

Remark 4.2.2. *As mentioned above Hogan et al. [HJTW04] proved that*

$$\lim_{t\to\infty} \mathbb{P}\left[\nu(t) < 0\right] = 0$$

holds if $\theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$*, but they do not prove that the condition* $\theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$ *is sufficient. This however is import for testing statistical arbitrage. My contribution is to prove the converse in the following proposition. Note that the authors of article [HJTW04] skipped a boundary point (with respect to statistical arbitrage/no statistical arbitrage), namely the point* $\theta = -1$ *and* $\lambda = -\frac{1}{2}$*. If the parameter values are equal to this boundary point we are also facing a statistical arbitrage opportunity.*

Proposition 4.2.2. *The following assertions are equivalent*

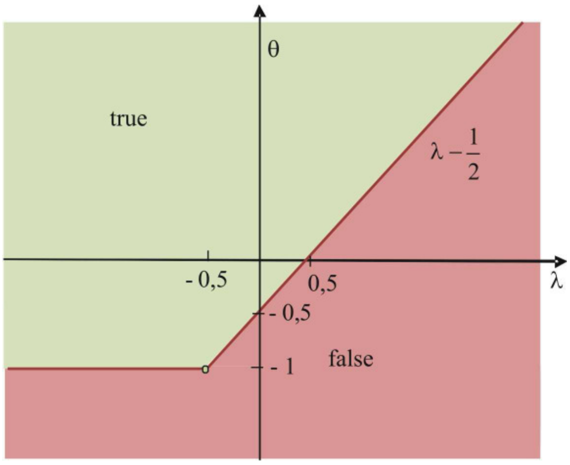
1.
$$\theta > \max\left\{\lambda - \frac{1}{2}, -1\right\} \quad \textit{or} \quad \left\{\theta = -1 \quad \textit{and} \quad \lambda = -\frac{1}{2}\right\} \quad \textit{and} \quad \mu > 0$$

2.
$$\lim_{t\to\infty} \mathbb{P}\left[\nu(t) < 0\right] = 0$$

Proof: By assumption 4.2.1 we have that $\nu(n)$ is distributed $\mathcal{N}\left(\mu \sum_{i=1}^n i^\theta, \sigma^2 \sum_{i=1}^n i^{2\lambda}\right)$, hence

$$\mathbb{P}\left[\nu(n) < 0\right] = \frac{1}{\sqrt{2\pi\sigma^2 \sum_{i=1}^n i^{2\lambda}}} \int_{-\infty}^0 \exp\left\{-\frac{\left(x - \mu \sum_{i=1}^n i^\theta\right)^2}{2\sigma^2 \sum_{i=1}^n i^{2\lambda}}\right\} dx = \Phi\left(-\frac{\mu \sum_{i=1}^n i^\theta}{\sigma \sqrt{\sum_{i=1}^n i^{2\lambda}}}\right)$$

Figure 4.8.: The Set of all (θ, λ) fulfilling the Third Condition of Statistical Arbitrage: $\lim_{t\to\infty} \mathbb{P}[\nu(t) < 0] = 0$



where Φ denotes the standard normal distribution function.

Thus the following assertions are equivalent

a.

$$\lim_{t\to\infty} \mathbb{P}[\nu(t) < 0] = 0$$

b.

$$\lim_{n\to\infty} \frac{\mu \sum_{i=1}^{n} i^{\theta}}{\sigma \sqrt{\sum_{i=1}^{n} i^{2\lambda}}} = \infty$$

c.

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} = \infty \text{ and } \mu > 0$$

$(1) \implies (2)$

Let $\theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$ and $\mu > 0$. We show that $\lim_{n\to\infty} \frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} = \infty$.

(i) Let $\lambda \neq -\frac{1}{2}$. By lemma 4.2.1 one has

$$\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} \geq \frac{\int_1^n s^{\theta} ds}{\sqrt{\int_1^n s^{2\lambda} ds + 1 + n^{2\lambda}}} = \frac{\frac{n^{1+\theta}-1}{1+\theta}}{\sqrt{\frac{n^{2\lambda+1}-1+(2\lambda+1)\left(1+n^{2\lambda}\right)}{2\lambda+1}}}$$

Multiplying both the numerator and denominator by $n^{-(1+\theta)}$ it follows

$$\begin{aligned}
&\frac{\frac{n^{1+\theta}-1}{1+\theta}}{\sqrt{\frac{n^{2\lambda+1}-1+(2\lambda+1)\left(1+n^{2\lambda}\right)}{2\lambda+1}}} \\
&= \frac{\frac{1-n^{-(1+\theta)}}{1+\theta}}{\sqrt{\frac{n^{(2\lambda+1)-2(1+\theta)}-n^{-2(1+\theta)}+(2\lambda+1)\left(n^{-2(1+\theta)}+n^{2\lambda-2(1+\theta)}\right)}{2\lambda+1}}} \\
&\stackrel{n\to\infty}{\to} \infty
\end{aligned}$$

since by assumption $1+\theta > 0$ and $2(1+\theta) > 2\lambda+1$.

(ii) For $\lambda = -\frac{1}{2}$ we obtain

$$\begin{aligned}
\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} &= \frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{-1}}} \geq \frac{\int_1^n s^{\theta} ds}{\sqrt{\int_1^n s^{-1} ds + 1}} = \frac{n^{1+\theta}-1}{\sqrt{\ln n + 1}} \frac{1}{1+\theta} \\
&= \frac{1-n^{-(1+\theta)}}{\sqrt{(\ln n)\, n^{-2(1+\theta)} + n^{-2(1+\theta)}}} \frac{1}{1+\theta} \stackrel{n\to\infty}{\to} \infty
\end{aligned}$$

since by L'Hôpital

$$\lim_{x\to\infty} \frac{\ln x}{x^{2(1+\theta)}} = \lim_{x\to\infty} \frac{1}{2\,(1+\theta)\, x^{2(1+\theta)}} = 0.$$

For the special case $\theta = -1$ and $\lambda = -\frac{1}{2}$ one has

$$\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} = \frac{\sum_{i=1}^{n} i^{-1}}{\sqrt{\sum_{i=1}^{n} i^{-1}}} = \sqrt{\sum_{i=1}^{n} i^{-1}} \stackrel{n\to\infty}{\to} \infty$$

since the harmonic series $\sum_{i=1}^{\infty} i^{-1}$ is divergent.

<u>(2) $\Longrightarrow$ (1)</u>

Assume that $\mu > 0$ and $\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} \stackrel{n\to\infty}{\to} \infty$. It follows that $\theta \geq -1$. We remark that $\sum_{i=1}^{\infty} i^{\theta}$ is divergent if $\theta \geq -1$ and converges if $\theta < -1$.

(i) Suppose that $-1 < \theta \leq \lambda - \frac{1}{2}$. Then by lemma 4.2.1

$$\begin{aligned}\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} &\leq \frac{\int_{1}^{n} s^{\theta} ds + 1 + n^{\theta}}{\sqrt{\int_{1}^{n} s^{2\lambda} ds}} = \frac{\frac{n^{1+\theta}-1+(1+\theta)+(1+\theta)n^{\theta}}{1+\theta}}{\sqrt{\frac{n^{2\lambda+1}-1}{2\lambda+1}}} \\ &= \frac{\frac{1+\theta n^{-(1+\theta)}+(1+\theta)n^{-1}}{1+\theta}}{\sqrt{\frac{n^{(2\lambda+1)-2(1+\theta)}-n^{-2(1+\theta)}}{2\lambda+1}}}\end{aligned}$$

where I again multiply numerator and denominator by $n^{-(1+\theta)}$.

If $-1 < \theta$ and $\theta < \lambda - \frac{1}{2} \Leftrightarrow 0 < 2\lambda - 1 - 2\theta$ then the sequence on the right hand side converges to 0 for $n \to \infty$. The sequence converges to $\frac{\sqrt{2\lambda+1}}{1+\theta}$ if $\theta = \lambda - \frac{1}{2} \Leftrightarrow 2\lambda + 1 = 2(1+\theta)$ and $\theta > -1$. This contradicts

$$\frac{\sum_{i=1}^{n} i^{\theta}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} \overset{n\to\infty}{\to} \infty.$$

(ii) Let $\theta = -1$ with $\theta < \lambda - \frac{1}{2}$. Thus $2\lambda + 1 > 0$. It holds by lemma 4.2.1

$$\frac{\sum_{i=1}^{n} i^{-1}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} \leq \frac{\int_{1}^{n} s^{-1} ds + 1}{\sqrt{\int_{1}^{n} s^{2\lambda} ds}} = \frac{\ln n + 1}{\sqrt{n^{2\lambda+1} - 1}} \sqrt{2\lambda + 1} \overset{n\to\infty}{\to} 0,$$

since by L'Hôpital

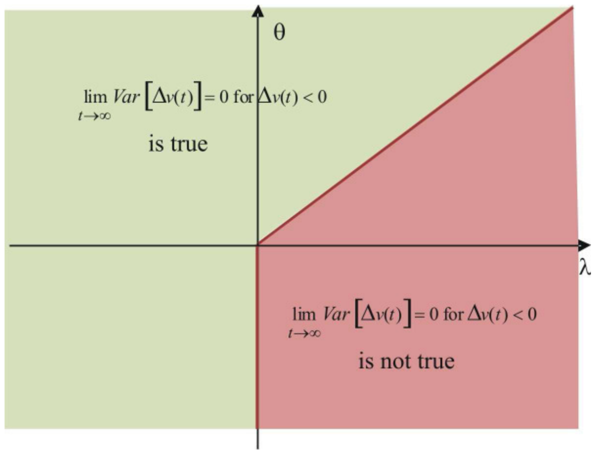
$$\begin{aligned}\lim_{x\to\infty} \frac{\ln x + 1}{\sqrt{x^{2\lambda+1} - 1}} &= \lim_{x\to\infty} \frac{2\sqrt{x^{2\lambda+1} - 1}}{(2\lambda+1)\, x^{2\lambda+1}} \leq \lim_{x\to\infty} \frac{2\sqrt{x^{2\lambda+1}}}{(2\lambda+1)\, x^{2\lambda+1}} \\ &= \lim_{x\to\infty} \frac{2}{2\lambda+1} \frac{1}{\sqrt{x^{2\lambda+1}}} = 0.\end{aligned}$$

This again contradicts $\frac{\sum_{i=1}^{n} i^{-1}}{\sqrt{\sum_{i=1}^{n} i^{2\lambda}}} \to \infty$ for $n \to \infty$. ∎

Remark 4.2.3. *Proposition 4.2.1 and 4.2.2 show that under assumption 4.2.1 condition 3 of statistical arbitrage, namely* $\lim_{t\to\infty} \mathbb{P}[\nu(t) < 0] = 0$ *implies condition 2 of statistical arbitrage, namely* $\lim_{t\to\infty} \mathbb{E}[\nu(t)] > 0$.

R. Jarrow et al. [JTTW05] proved that the fourth condition of statistical arbitrage, namely $\lim_{t\to\infty} \mathbb{V}[\Delta\nu(t)] = 0$ for $\Delta\nu(t) < 0$, holds if $\lambda < 0$ or $\theta > \lambda$. In the next proposition I show that this condition is also sufficient which is important for testing statistical arbitrage.

Figure 4.9.: The Set of all (θ, λ) fulfilling the Fourth Condition of Statistical Arbitrage



Proposition 4.2.3. *Let $\mu > 0$. Then the following assertions are equivalent:*

1. $\lambda < 0$ *or* $\theta > \lambda$
2. $\lim_{t\to\infty} \mathbb{V}\left[\Delta\nu(t)\right] = 0$ *for* $\Delta\nu(t) < 0$.

Proof: The distribution of the incremental trading profits is the normal one $\mathcal{N}\left(\mu t^{\theta}, \sigma^2 t^{2\lambda}\right)$, hence the variance of the random variable $\Delta\nu(t)$ on the set $\{\Delta\nu(t) < 0\}$ is given by

$$\frac{1}{\sqrt{2\pi\sigma^2 t^{2\lambda}}} \int_{-\infty}^{0} \left(x - \mu t^{\theta}\right)^2 e^{-\frac{\left(x-\mu t^{\theta}\right)^2}{2\sigma^2 t^{2\lambda}}} dx$$

Substituting $y = \frac{x-\mu t^{\theta}}{\sigma t^{\lambda}}$ one obtains

$$\frac{1}{\sqrt{2\pi\sigma^2 t^{2\lambda}}} \int_{-\infty}^{0} \left(x - \mu t^{\theta}\right)^2 e^{-\frac{\left(x-\mu t^{\theta}\right)^2}{2\sigma^2 t^{2\lambda}}} dx = \frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}} dy.$$

(1) $\Longrightarrow$ (2)

Let $\lambda < 0$. One has

$$\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}} dy \leq \frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2}} dy = \sigma^2 t^{2\lambda} \overset{t\to\infty}{\to} 0$$

Now, let $\theta > \lambda$. Applying partial integration we obtain

$$\int u^2 e^{-\frac{u^2}{2}} du = -\int u(-u) e^{-\frac{u^2}{2}} du = -u e^{-\frac{u^2}{2}} + \int e^{-\frac{u^2}{2}} du,$$

hence

$$\begin{aligned}\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy &= \frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\left(-\left(-\frac{\mu t^{\theta-\lambda}}{\sigma}\right)e^{-\frac{\left(-\frac{\mu t^{\theta-\lambda}}{\sigma}\right)^2}{2}} + 0 + \Phi\left(\frac{-\mu t^{\theta-\lambda}}{\sigma}\right)\right)\\ &= \frac{\sigma\mu t^{\theta+\lambda}}{\sqrt{2\pi}e^{\frac{\mu^2 t^{2(\theta-\lambda)}}{2\sigma^2}}} + \frac{\sigma^2}{\sqrt{2\pi}}t^{2\lambda}\Phi\left(\frac{-\mu t^{\theta-\lambda}}{\sigma}\right)\end{aligned}$$

where Φ denotes the cumulative normal distribution. It follows

$$\frac{\sigma\mu t^{\theta+\lambda}}{\sqrt{2\pi}e^{\frac{\mu^2 t^{2(\theta-\lambda)}}{2\sigma^2}}} + \frac{\sigma^2}{\sqrt{2\pi}}t^{2\lambda}\Phi\left(\frac{-\mu t^{\theta-\lambda}}{\sigma}\right) \overset{t\to\infty}{\to} 0$$

since the exponential function converges faster to infinity for $\theta - \lambda > 0$ than any power function.

<u>(2) $\Longrightarrow$ (1)</u>

Let $\lim_{t\to\infty} \mathbb{V}[\Delta\nu(t)] = 0$ for $\Delta\nu(t) < 0$. Assume that $\lambda \geq 0$ and $\theta \leq \lambda$.

If $\lambda = 0$ and $\theta \neq 0$ then

$$\lim_{t\to\infty}\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy = \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{0} y^2 e^{-\frac{y^2}{2}}\,dy = \frac{\sigma^2}{\sqrt{2\pi}}\frac{1}{2} > 0$$

and for $\lambda = \theta = 0$

$$\begin{aligned}\lim_{t\to\infty}\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy &= \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy\\ &= \frac{\sigma^2}{\sqrt{2\pi}}\left(\frac{\mu}{\sigma}e^{-\frac{\mu^2}{2\sigma^2}} + \Phi\left(-\frac{\mu}{\sigma}\right)\right) > 0\end{aligned}$$

For $\lambda > 0$ I obtain

$$\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy \overset{t\to\infty}{\to} \infty$$

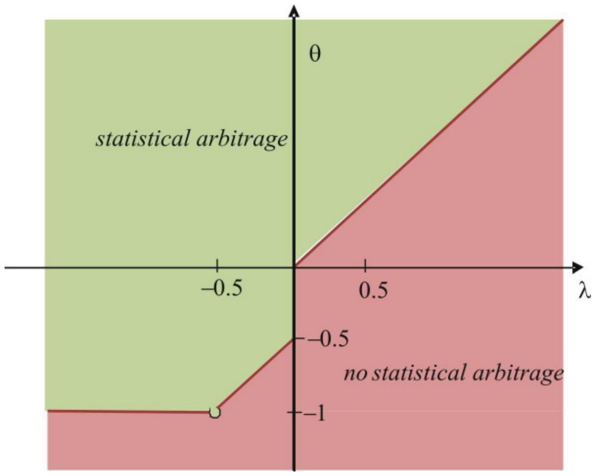
This contradicts

$$\frac{\sigma^2 t^{2\lambda}}{\sqrt{2\pi}}\int_{-\infty}^{-\frac{\mu t^{\theta-\lambda}}{\sigma}} y^2 e^{-\frac{y^2}{2}}\,dy \overset{t\to\infty}{\to} 0$$

■

Conditions for statistical arbitrage are proven by Hogen et al. [HJTW04], theorem 1, page 537. I show that the *equivalence* holds if one enlarges condition $H3$ by the additional condition $(\theta, \lambda) = (-1, -0.5)$.

Figure 4.10.: The Set of all (θ, λ) fulfilling all Conditions of Statistical Arbitrage



Theorem 4.2.1. *A trading strategy generates statistical arbitrage if and only if incremental trading profits satisfy the following conditions:*

1. $A_1 : \mu > 0$ *and*
2. $A_2 : \lambda < 0$ *or* $\theta > \lambda$, *and*
3. $A_3 : \theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$ *or* $(\theta, \lambda) = (-1, -0.5)$

Proof: This follows directly from proposition 4.2.1, 4.2.2 and 4.2.3. ■

Remark 4.2.4. *(Standard Arbitrage Opportunity versus Statistical Arbitrage) A standard arbitrage has* $X(0) = 0$ *where there exists a finite time* $T > 0$ *such that* $\mathbb{P}[X(T) > 0] > 0$ *and* $\mathbb{P}[X(T) \geq 0] = 1$. *To transform the standard arbitrage opportunity into an infinite horizon self-financing trading strategy, we just invest the proceeds at time* T *into the money market account, i.e.* $X(s) = X(T)\frac{B(s)}{B(T)}$ *for* $s \geq T$. *Note that* $\nu(s) = X(T)\frac{B(s)}{B(T)}\frac{1}{B(s)} = \nu(T)$ *for* $s \geq T$. *Then* $\lim_{s\to\infty} \mathbb{E}[\nu(s)] = \mathbb{E}[\nu(T)] > 0$ *which satisfies condition 2 and* $\lim_{s\to\infty} \mathbb{P}[\nu(s) < 0] = \mathbb{P}[\nu(T) < 0] = 0$ *which satisfies condition 3. Condition 4 is trivially fulfilled.*

A statistical arbitrage is not necessarily a standard arbitrage opportunity: Choose $\mu = 0.1$, $\theta = -1$, $\sigma = 0.2$ *and* $\lambda = -0.5$. *With the help of theorem 4.2.1 we know that we are facing a statistical arbitrage opportunity. Now suppose that this statistical arbitrage opportunity is also a standard arbitrage opportunity: Then there exists a* $N > 0$ *so that* $\mathbb{P}\left[\sum_{i=1}^{N} \Delta\nu(i) \geq 0\right] = 1$. *It* *follows*

$$\mathbb{P}\left[0.1\sum_{i=1}^{N}\frac{1}{i}+0.2\sum_{i=1}^{N}\frac{1}{\sqrt{i}}z(i)\geq 0\right]=1.$$

But this is contradiction since $z(i)$ is normal distributed with a strictly positive variance.

In the next step we dedicate ourselves with the estimation of the model parameters. For this purpose, we first calculate the log-likelihood function for the increments in assumption 4.2.1. Since the increments are independent we can write for the joint density

$$\begin{aligned}\ln L\left(\mu,\sigma^2,\lambda,\theta \mid \Delta v\right) &= \ln\prod_{i=1}^{n}\frac{1}{\sigma i^{\lambda}\sqrt{2\pi}}\exp\left(-\frac{(\Delta\nu_i-\mu i^{\theta})^2}{2\sigma^2 i^{2\lambda}}\right)\\ &= -n\ln\sqrt{2\pi}-\frac{1}{2}\sum_{i=1}^{n}\ln\left(\sigma^2 i^{2\lambda}\right)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\left(\Delta\nu_i-\mu i^{\theta}\right)^2}{i^{2\lambda}}\end{aligned}$$

Taking partial derivatives with respect to μ, σ, λ and θ of the log-likelihood function and setting the derivatives equal to zero yield the following:

1.

$$\begin{aligned}\frac{\partial\ln L}{\partial\mu}\left(\mu,\sigma^2,\lambda,\theta \mid \Delta v\right) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{2\left(\Delta\nu(i)-\mu i^{\theta}\right)\left(-i^{\theta}\right)}{i^{2\lambda}}\overset{!}{=}0\Leftrightarrow\\ \sum_{i=1}^{n}\frac{i^{\theta}}{i^{2\lambda}}\Delta\nu(i)-\sum_{i=1}^{n}\mu i^{2(\theta-\lambda)} &\overset{!}{=} 0\Leftrightarrow\\ \mu &= \frac{\sum_{i=1}^{n}\frac{i^{\theta}}{i^{2\lambda}}\Delta\nu(i)}{\sum_{i=1}^{n}i^{2(\theta-\lambda)}}\end{aligned}$$

2.

$$\begin{aligned}\frac{\partial\ln L}{\partial\sigma^2}\left(\mu,\sigma^2,\lambda,\theta \mid \Delta v\right) &= -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}\frac{\left(\Delta\nu(i)-\mu i^{\theta}\right)^2}{i^{2\lambda}}\overset{!}{=}0\Leftrightarrow\\ \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\Delta\nu(i)-\mu i^{\theta}\right)^2}{i^{2\lambda}}\end{aligned}$$

3.

$$\begin{aligned}\frac{\partial\ln L}{\partial\lambda}\left(\mu,\sigma^2,\lambda,\theta \mid \Delta v\right) &= -\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma^2 i^{2\lambda}}\sigma^2 i^{2\lambda}2\ln i-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\left(\Delta\nu(i)-\mu i^{\theta}\right)^2}{i^{2\lambda}}(-2)\ln i\\ &= -\sum_{i=1}^{n}\ln i+\frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{\left(\Delta\nu(i)-\mu i^{\theta}\right)^2}{i^{2\lambda}}\ln i\overset{!}{=}0\Leftrightarrow\\ \sigma^2\sum_{i=1}^{n}\ln i &= \sum_{i=1}^{n}\frac{\ln i}{i^{2\lambda}}\left(\Delta\nu(i)-\mu i^{\theta}\right)^2\end{aligned}$$

4.

$$\begin{aligned}
\frac{\partial \ln L}{\partial \theta}\left(\mu, \sigma^2, \lambda, \theta \mid \Delta v\right) &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{1}{i^{2\lambda}}2\left(\Delta\nu(i) - \mu i^{\theta}\right)(-\mu) i^{\theta}\ln i \stackrel{!}{=} 0 \Leftrightarrow \\
\sum_{i=1}^{n}\Delta\nu(i)\ln(i) i^{\theta-2\lambda} &= \mu\sum_{i=1}^{n}\ln(i) i^{2(\theta-\lambda)}
\end{aligned}$$

So we obtain

$$\begin{aligned}
\hat{\mu} &= \frac{\sum_{i=1}^{n}\Delta\nu(i) i^{\hat{\theta}-2\hat{\lambda}}}{\sum_{i=1}^{n} i^{2\left(\hat{\theta}-\hat{\lambda}\right)}} \\
\hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}\frac{\left(\Delta\nu(i) - \hat{\mu} i^{\hat{\theta}}\right)^2}{i^{2\hat{\lambda}}} \\
\hat{\sigma}^2\sum_{i=1}^{n}\ln(i) &= \sum_{i=1}^{n}\frac{\ln(i)}{i^{2\hat{\lambda}}}\left(\Delta\nu(i) - \hat{\mu} i^{\hat{\theta}}\right)^2 \\
\sum_{i=1}^{n}\Delta\nu(i)\ln(i) i^{\hat{\theta}-2\hat{\lambda}} &= \hat{\mu}\sum_{i=1}^{n}\ln(i) i^{2\left(\hat{\theta}-\hat{\lambda}\right)}
\end{aligned}$$

One can find solutions for the four parameters numerically with the Excel Solver. Since we have to check if these values are the global maximum of the log-likelihood function we then have to visualize the log-likelihood function.

To obtain valuable start values for the mean μ and θ we consider the incremental trading profits without random noise $\Delta\nu_i = \mu i^{\theta}$. If $\Delta\nu_i > 0$ then we can write $\ln\Delta\nu_i = \ln\mu + \theta\ln i$. Note that the incremental trading profits which are negative are skipped with an index i also skipping the negative ones. Now we can apply the least squares method to estimate the start values for μ and θ denoted by $\tilde{\mu}, \tilde{\theta}$. To fit the volatility σ and λ we consider $\ln\left|\Delta\nu_i - \tilde{\mu} i^{\tilde{\theta}}\right| = \ln\sigma + \lambda\ln i$.

Now we provide a statistical methodology to test for statistical arbitrage via *bootstrapping*. The existence of statistical arbitrage consists of joint restrictions on the parameters underlying the evolution of trading profits. Following theorem 4.2.1 the following restrictions have to be satisfied simultaneously for a statistical arbitrage opportunity to exist:

1. $A_1 : \mu > 0$, <u>and</u>
2. $A_2 : \lambda < 0$ or $\theta > \lambda$, <u>and</u>
3. $A_3 : \theta > \max\left\{\lambda - \frac{1}{2}, -1\right\}$ or $(\theta, \lambda) = (-1, -0.5)$

Thus, no statistical arbitrage means

1. A_1^c : $\mu \leq 0$, or
2. A_2^c : $\lambda \geq 0$ and $\theta \leq \lambda$, or
3. A_3^c : $\theta \leq \max\left\{\lambda - \frac{1}{2}, -1\right\}$ and $(\theta, \lambda) \neq (-1, -0.5)$

We remark that A_3 and A_3^c are modified and rectified conditions of the sub-hypotheses H_3/H_3^c of [HJTW04] on page 537.

Because we are interested in testing a given trading strategy for statistical arbitrage, the null hypothesis is no statistical arbitrage.

Let $\left(\hat{\mu}, \hat{\lambda}, \hat{\theta}, \hat{\sigma}\right)$ denote the maximum likelihood estimator with respect to an observed incremental trading profit time series $\Delta\nu_i$ with $i = 1, 2, ..., n$. Plugging the estimator in model 4.2.1 one obtains

$$\Delta\hat{v}(i) = \hat{\mu} i^{\hat{\theta}} + \hat{\sigma} i^{\hat{\lambda}} z(i). \tag{4.14}$$

One should simulate the process $\Delta\hat{v}$ with $i = 1, 2, ..., n$ $10,000$ times and obtains $10,000$ time series $\Delta\nu(i)^*$ with $i = 1, 2, ..., n$. For each time series we calculate the maximum likelihood estimator $(\mu^*, \lambda^*, \theta^*, \sigma^*)$ and check if one has statistical arbitrage or not. The number of the occurrence of no statistical arbitrage is divided by $10,000$ and can be interpreted as the p-value. Before testing one chooses a level of significance α (for example $\alpha = 5\%$). If the p-value is smaller or equal α one rejects the null hypothesis of no statistical arbitrage. So one has found a statistical arbitrage opportunity.

If one has detected a statistical arbitrage opportunity one can hope that the strategy is profitable in the future, too. Since a statistical arbitrage opportunity as defined in Hogan et al. [HJTW04] can take a long time to unfold, we cannot be sure that the estimated parameters do not change significantly over time. It is a long-horizon strategy. Actually, for a fixed time horizon it can happen that we have to bear severe losses.

Appendix A

Put-Call Parity for European Options on Futures

The Put-Call parity for put and call options underlying futures contracts can be derived with the same arbitrage arguments as in stock markets.

With $C(t,T)$ and $P(t,T)$ denoting the price of a European call and put option at time t both with strike K and time to expiration T on a futures contract with maturity $\tilde{T} \geq T$ named by $F(t,\tilde{T})$ we can form two portfolios:

Portfolio A: a European call futures option plus an amount of cash equal to Ke^{-rT}

Portfolio B: a European put futures option plus a long futures contract plus an amount of cash equal to $F(t,\tilde{T})e^{-rT}$.

In portfolio A the cash can be invested at the risk-free rate, r, and will grow to K at time T. Let $F(T,\tilde{T})$ be the futures price at maturity of the option. If $F(T,\tilde{T}) > K$, the call option in portfolio A is exercised and portfolio A is worth $F(T,\tilde{T})$. If $F(T,\tilde{T}) \leq K$, the call is not exercised and portfolio A is worth K. The value of portfolio A at time T is, therefore, $\max\Big(F(T,\tilde{T}), K\Big)$.

In portfolio B the cash can be invested at the risk-free rate to grow to $F(T,\tilde{T})$ at time T. The put option provides a payoff of $\max\Big(K - F(T,\tilde{T}), 0\Big)$. The futures contract provides a payoff of $F(T,\tilde{T}) - F(t,\tilde{T})$. The value of portfolio B at time T is, therefore,

$$F(t,\tilde{T}) + \Big(F(T,\tilde{T}) - F(t,\tilde{T})\Big) + \max\Big(K - F(T,\tilde{T}), 0\Big) = \max\Big(F(T,\tilde{T}), K\Big)$$

Because the two portfolios have the same value at time T and there are no early exercise opportunities, it follows that they are worth the same at time t. The value of portfolio A today is $C(t,T) + Ke^{-rT}$. The futures contract in portfolio B is worth zero at time t. Therefore, portfolio B is worth $P(t,T) + F(t,\tilde{T})$.

$$C(t,T) + Ke^{-rT} = P(t,T) + F(t,\tilde{T})$$

Appendix B

Hesse Matrix

For $a, b \in \mathbb{R}$ with $a < b$ and $\theta > 0$, $\sigma > 0$ and $c \geq 0$ we have already defined (see definition 4.1.5)

$$F(a,b) = \frac{\theta\,(b-a-c)}{\pi\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right)}$$

We now calculate the partial derivatives of function F.

The numerator and denominator of function F are abbreviated with

$$\begin{aligned} f(a,b) &= \theta\,(b-a-c) \\ g(a,b) &= \pi\left(\operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}b\right) - \operatorname{Erfi}\left(\frac{\sqrt{\theta}}{\sigma}a\right)\right) \end{aligned}$$

$$\frac{\partial^2 F(a,b)}{\partial^2 a} = \frac{\partial}{\partial a}\left(\frac{-\theta g(a,b) + f(a,b) 2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}}{g^2(a,b)}\right)$$

$$= \frac{\left(\theta\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2} - \theta 2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2} + f(a,b)2\sqrt{\pi}\frac{2\sqrt{\theta}\theta a}{\sigma^2\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)g^2(a,b)}{g^4(a,b)}$$

$$- \frac{\left(-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)2g(a,b)\left(-\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)}{g^4(a,b)}$$

$$= \frac{\frac{\sqrt{\pi}\theta\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\left(4f(a,b)\frac{a}{\sigma^2}g(a,b) + 4\left(-g(a,b) + 2\sqrt{\pi}(b-a-c)\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)\right)}{g^3(a,b)}$$

$$\frac{\partial^2 F(a,b)}{\partial^2 b} = \frac{\partial}{\partial b}\left(\frac{-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}}{g^2(a,b)}\right)$$

$$= \frac{\left(\theta\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2} - \theta 2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2} + f(a,b)2\sqrt{\pi}\frac{2\sqrt{\theta}\theta b}{\sigma^2\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)g^2(a,b)}{g^4(a,b)}$$

$$- \frac{\left(-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)2g(a,b)\left(-\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)}{g^4(a,b)}$$

$$= \frac{\frac{\sqrt{\pi}\theta\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\left(4f(a,b)\frac{b}{\sigma^2}g(a,b) + 4\left(-g(a,b) + 2\sqrt{\pi}(b-a-c)\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)\right)}{g^3(a,b)}$$

$$\frac{\partial^2 F(a,b)}{\partial a\partial b} = \frac{\partial}{\partial b}\left(\frac{-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}}{g^2(a,b)}\right)$$

$$= \frac{\left(-\theta\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2} + \theta 2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)g^2(a,b) - \left(-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)2g(a,b)2\sqrt{\pi}\frac{\theta}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}}{g^4(a,b)}$$

The determinant of the Hesse matrix evaluated at point $(-b,b)$ $(b>0)$ then takes the following form

$$\frac{\partial^2 F(-b,b)}{\partial^2 a}\frac{\partial^2 F(-b,b)}{\partial^2 b} - \left(\frac{\partial^2 F(-b,b)}{\partial a\partial b}\right)^2$$

$$= \frac{\frac{\sqrt{\pi}\theta\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\left(4f(a,b)\frac{a}{\sigma^2}g(a,b) + 4\left(-g(a,b) + 2\sqrt{\pi}(b-a-c)\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)\right)}{g^3(a,b)}$$

$$* \frac{\frac{\sqrt{\pi}\theta\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\left(4f(a,b)\frac{b}{\sigma^2}g(a,b) + 4\left(-g(a,b) + 2\sqrt{\pi}(b-a-c)\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}\right)\right)}{g^3(a,b)}$$

$$- \left(\frac{\left(-\theta\frac{2\sqrt{\pi}\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}b^2} + \theta 2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)g^2(a,b) - \left(-\theta g(a,b) + f(a,b)2\sqrt{\pi}\frac{\sqrt{\theta}}{\sigma}e^{\frac{\theta}{\sigma^2}a^2}\right)2g(a,b)2\sqrt{\pi}\frac{\theta}{\sigma}e^{\frac{\theta}{\sigma^2}b^2}}{g^4(a,b)}\right)^2$$

Bibliography

[AB57] J. Aitchison and J. Brown. *The Lognormal Distribution.* Cambridge University Press, 1957.

[ADEH99] P. Artzner, F. Delbaen, J. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–228, 1999.

[Bil95] P. Billingsley. *Probability and Measure.* Wiley Series in Probability and Mathematical Statistics, third edition, 1995.

[Bla76] F. Black. The pricing of commodity contracts. *Journal of Financial Economics*, 3:167–179, 1976.

[BM00] G. Bakshi and D. Madan. Spanning and derivative-security valuation. *Journal of Financial Economics*, 55:205–238, 2000.

[BS02] A. Borodin and P. Salminen. *Handbook of Brownian Motion-Facts and Formulae.* Birkheuser Verlag, second edition, 2002.

[Chu01] K. Chung. *A Course in Probability Theory.* Academic Press, third edition, 2001.

[Den99] S. Deng. Stochastic models of energy commodity prices and their applications: Mean reversion with jumps and spikes. Working Paper, Georgia Institute of Technology, 1999.

[DPS98] D. Duffie, J. Pan, and K. Singleton. Transform analysis and option pricing for affine jump-diffusion. Working Paper, Graduate School of Business, Stanford University, 1998.

[DS92] D. Duffie and R. Stanton. Pricing continuously resettled contingent claims. *Journal* *of Economic Dynamics and Control*, 16:561–573, 1992.

[EW03] A. Eydeland and K. Wolyniec. *Energy and Power Risk Management.* Wiley Finance, 2003.

[FHH04] J. Franke, W. Härdle, and Ch. Hafner. *Einführung in die Statistik der Finanzmärkte.* Springer, 2004.

[Fri75] A. Friedman. *Stochastic Differential Equations and Applications.* Academic Press, New York, 1975.

[Gem05] H. Geman. *Commodities and commoditiy derivatives.* Wiley Finance, 2005.

[GJY03] A. Göing-Jaeschke and M. Yor. A survey and some generalizations of bessel processes. *Bernoulli*, 9:313–349, 2003.

[Hes93] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6:327–343, 1993.

[HJTW04] S. Hogan, R. Jarrow, M. Teo, and M. Warachka. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics.*, 73/3:525–565, 2004.

[Hul06] J.C. Hull. *Options, Futures & Other Derivatives.* Prentice Hall, sixth edition, 2006.

[Iof10] M. Ioffe. Probability distribution of cox-ingersoll-ross process. Working Paper, Egar technology, New York, 2010.

[Jäc02] P. Jäckel. *Monte Carlo Methods in Finance.* Wiley, 2002.

[JT93] N. Jegadeesh and S. Titman. Returns to buying winners and selling losers implications for stock market efficiency. *Journal of Finance*, 48:65–91, 1993.

[JTTW05] R. Jarrow, M. Teo, Y. Tse, and M. Warachka. Statistical Arbitrage and Market Efficiency, Enhanced Theory, Robust Tests and Further Applications. *SSRN eLibrary*, February 2005.

[Kor97] R. Korn. *Optimal Portfolios.* World Scientific, 1997.

[KS00] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, 2000.

[KT81] S. Karlin and H. Taylor. *A Second Course in Stochastic Processes.* Academic Press, New York, 1981.

[LL96] D. Lamberton and B. Lapeyre. *Introduction to Stochastic Calculus Applied to Finance.* Chapman & Hall, 1996.

[Mad08] P. Madigan. Don't follow the herd. *Risk*, pages 88–90, September 2008.

[Mar52] H. Markowitz. Portfolio selection. *Journal of Finance*, 7/1:77–91, 1952.

[MSG05] L. Martellini, K. Simsek, and F. Goltz. Structured forms of investment strategies in institutional investors' portfolios. Working Paper, Edhec Risk and Asset Management Research Centre, 2005.

[Mül05] M. Müller. Tests for optimal trading days. Working Paper, Fraunhofer ITWM, Kaiserslautern, 2005.

[PS85] M. Pollak and D. Siegmund. A diffusion process and its applications to detecting a change in the drift of brownian motion. *Biometrika*, 72:267–280, 1985.

[PTVF02] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing.* Cambridge University Press, second edition, 2002.

[Roc97] R. Rockafellar. *Convex Analysis.* Princton Landmarksin Mathematics and Physics, 1997.

[Rou98] K. Rouwenhorst. International momentum strategies. *Journal of Finance*, 53/1:267–284, 1998.

[RU00] R. T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–41, 2000.

[RU02] R. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, 26/7:1443–1471, 2002.

[RY91] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag Berlin Heidelberg, 1991.

[SW65] S. Shapiro and M. Wilk. An analysis of variance test for normality (complete samples). *Biometrika*, 52:591–611, 1965.

[UO30] G. Uhlenbeck and L. Ornstein. On the theory of brownian motion. *Physical Review*, 36:823–841, 1930.

[Wil98] P. Wilmott. *Derivatives. The Theory and Practice of Financial Engineering.* John Wiley & Sons, 1998.

[Yan02] X. Yan. Valuation of commodity derivatives in a new multi-factor model. *Review of Derivatives Research*, 5:251–271, 2002.

Scientific Career

June 1998	*Abitur at Maria Ward-Gymnasium Mainz*
October 1998 to May 2004	*Study of Mathematics with Minor Subject Computer Sciences at the Ludwig-Maximilians-University Munich, Graduation: Diplom-Mathematikerin*
September 2004 to March 2010	*Research Associate Fraunhofer ITWM Institute for Industrial Mathematics Kaiserslautern, Department of Financial Mathematics*
since April 2010	*Research Associate Technical University Kaiserslautern*
February 2007 to April 2008	*Maternity Leave*

Wissenschaftlicher Werdegang

Juni 1998	*Abitur am Maria Ward-Gymnasium Mainz*
Oktober 1998 bis Mai 2004	*Studium der Mathematik mit Nebenfach Informatik an der Ludwig-Maximilians-Universität München, Abschluss: Diplom-Mathematikerin*
September 2004 bis März 2010	*Wissenschaftliche Mitarbeiterin am Fraunhofer ITWM Institut für Techno- und Wirtschaftmathematik Kaiserslautern, Abteilung Finanzmathematik*
seit April 2010	*Wissenschaftliche Mitarbeiterin an der Technischen Universität Kaiserslautern*
Februar 2007 bis April 2008	*Elternzeit*