Advantage of Filtering for Portfolio Optimization in Financial Markets with Partial Information

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

vorgelegt von

Leonie Maria Ruderer

Die Arbeit wurde angeleitet von Prof. Dr. Jörn Sass, TU Kaiserslautern, und zudem begutachtet von Prof. Dr. Ralf Wunderlich, BTU Cottbus-Senftenberg.

Datum der Disputation: 4.12.2015

Technische Universität Kaiserslautern

Fachbereich Mathematik

D 386

Abstract

In a financial market we consider three types of investors trading with a finite time horizon with access to a bank account as well as multliple stocks: the fully informed investor, the partially informed investor whose only source of information are the stock prices and an investor who does not use this information. The drift is modeled either as following linear Gaussian dynamics or as being a continuous time Markov chain with finite state space. The optimization problem is to maximize expected utility of terminal wealth.
The case of partial information is based on the use of filtering techniques. Conditions to ensure boundedness of the expected value of the filters are developed, in the Markov case also for positivity. For the Markov modulated drift, boundedness of the expected value of the filter relates strongly to portfolio optimization: effects are studied and quantified. The derivation of an equivalent, less dimensional market is presented next. It is a type of Mutual Fund Theorem that is shown here.
Gains and losses eminating from the use of filtering are then discussed in detail for different market parameters: For infrequent trading we find that both filters need to comply with the boundedness conditions to be an advantage for the investor. Losses are minimal in case the filters are advantageous. At an increasing number of stocks, again boundedness conditions need to be met. Losses in this case depend strongly on the added stocks. The relation of boundedness and portfolio optimization in the Markov model leads here to increasing losses for the investor if the boundedness condition is to hold for all numbers of stocks. In the Markov case, the losses for different numbers of states are negligible in case more states are assumed then were originally present. Assuming less states leads to high losses. Again for the Markov model, a simplification of the complex optimal trading strategy for power utility in the partial information setting is shown to cause only minor losses. If the market parameters are such that shortselling and borrowing constraints are in effect, these constraints may lead to big losses depending on how much effect the constraints have. They can though also be an advantage for the investor in case the expected value of the filters does not meet the conditions for boundedness.
All results are implemented and illustrated with the corresponding numerical findings.

Zusammenfassung

Wir betrachten drei Arten von Investoren, die in einem Markt mit stochastischer Drift, zum einen einer Markov Kette oder einer linear Gaußschen Dynamik, handeln. Der eine Investor ist voll informiert, einem partiell informierten sind nur die Aktienpreise bekannt und ein dritter nutzt diese Information nicht. Sie haben die Wahl zwischen mehreren Aktien und einer risikofreien Anlage und ihr Ziel ist es den erwarteten Nutzen des optimalen finalen Portfoliowerts zu maximieren.

Bei partieller Information kommt Filtern zum Einsatz. Dazu werden zunächst Bedingungen an den Erwartungswert der Filter hergeleitet, um zum Beispiel Beschränktheit des Diskretisierungsschemas zu gewährleisten. Der Markov-Filter muss zusätzlich auch positiv sein. Für diesen wird eine weitgehende Verbindung zur Portfoliooptimierung hergeleitet. Deren Effekte werden untersucht und quantifiziert. Diese Überlegungen führen zu einem äquivalenten Markt niedrigerer Dimension. Eine Art Mututal Fund Theorem wird dort bewiesen.

Im Anschluss werden Gewinne und Verluste durch Filtern in verschiedenen Marktsituationen behandelt. Handelt der Investor nur selten, so ist das Filtern von Vorteil, falls die Bedingungen an die Beschränktheit erfüllt sind. Dann verursacht seltenes Handeln aber nur geringe Verluste. Auch für steigende Anzahlen an Aktien im Portfolio muss die Beschränktheit beachtet werden. Der Zusammenhang zwischen Beschränktheit und Portfoliooptimierung im Markovmodell führt hier zu hohen Verlusten, falls die Beschränktheit für alle Aktienanzahlen gegeben sein soll. Im Markovmodell treten nur sehr geringe Verluste auf, wenn mehr Zustände für die Markovkette angenommen werden, als im tatsächlichen Modell vorhanden waren. Werden allerdings weniger Zustände angenommmen, so kommt es zu hohen Verlusten. Auch wird im Markovmodell die Auswirkung einer Vereinfachung der komplexen Gestalt der optimalen Handelsstrategie für partielle Information und Power-Nutzenfunktion untersucht: Es kommt nur zu vernachlässigbaren Verlusten. Gibt es Beschränkungen der Strategie, so kann es zu sehr hohen Verlusten kommen. Dies hängt davon ab, wie oft und stark die Beschränkungen zu Veränderungen der Strategie führen. Solche Beschränkungen können auch von Vorteil sein, falls zum Beispiel der Erwartungswert der Filter nicht beschränkt ist.

Alle Resultate sind implementiert und werden mit numerischen Ergebnissen illustriert.

Danksagung

Mein größter Dank gebührt meinem Betreuer Prof. Dr. Jörn Sass. Nicht nur hat er mir die Promotion aus der Ferne überhaupt erst ermöglicht durch seine Rücksichtnahme und Flexibilität, er hat jederzeit den passenden Input geliefert, sodass diese Arbeit inhaltlich Form annehmen konnte.

Auch bei Prof. Dr. Ralf Wunderlich möchte ich mich bedanken für die vielen konstruktiven Anmerkungen und die detaillierte Durchsicht.

Weiterhin hat mich die Zusammenarbeit mit Elisabeth Leoff immer sehr erfreut und positiv beeinflusst. Vielen Dank dafür.

Meine Familie hat mir in meinem Studium und später in meiner Promotionszeit ohne Pause kompromisslos beigestanden: Meinen Eltern und Großeltern möchte ich hier tiefen Dank aussprechen für den Einsatz und die Unterstützung.

Während dieser ganzen Zeit hatte ich meinen Mann an meiner Seite zur Aufmunterung, Realitätswahrung, Klärung des Kopfes und Begleitung. An ihn geht mein herzlichster Dank.

Contents

1 Introduction

In this work we consider an investor trading in the stock market with given initial capital. He has access to a bank account as well as a number of stocks and can trade up to some finite time. His goal is to maximize the expected utility of terminal wealth.

The drift is modeled either as following linear Gaussian dynamics or as being a continuous time Markov chain with finite state space. Apart from the fully informed investor we also consider one whose only source of information are the stock prices, which is a quite natural assumption, and the one who has no information at all. The investor with only access to stock prices is called partially informed. This lack of information is reflected in a smaller filtration and the necessary quantities are expressed by filters, especially those for the drift.
Throughout this work we address the issue of bankruptcy and the circumstances where it occurs.

The general topic of portfolio optimization under partial information has already been studied in detail for different models for the drift.
Lakner considered a stochastic drift and studied the optimal portfolio as well as the optimal trading strategy in his work [17] and especially for a Gaussian drift in [18]. The filter for the drift in this model is the well known Kalman filter.
The case of partial information with the drift modeled as a Markov chain leads to a Hidden Markov Model. Bäuerle and Rieder used a Markov modulated unobservable drift in [21] and solved the corresponding Hamilton-Jacobi- Bellman equation. In [24] the drift is also modeled with a Markov chain by Sass and Haussmann, but the martingale method is being applied and the optimal trading strategy computed.
The case of the Markov modulated drift uses the Wonham filter and filters introduced by Elliott in [10] which were then robustified following Clark's work [6] and discretized by James, Krishnamurthy and Le Gland in [14].

In the case of partial information parameters are unknown and can be estimated with an algorithm described by Elliott and Krishnamurthy in [11] for the Gaussian case and the algorithm proposed by [24] for the Markov drift which is based on a work by Dembo and Zeitouni [8].

These filters intuitively lead to a result for the partially informed investor somewhere inbetween the fully informed investor and the investor without information: if it were no better than the case without information it would be an unnecessary computational effort and the result with full information cannot be reached with less information. Nevertheless there are market situations in which filtering in fact does not contribute valuable information as well as those where the partially informed investor reaches a distinctively better result than in others when compared to the fully informed investor. Such situations will be discussed in detail, the different market settings include infrequent trading, a varying number of stocks and shortselling constraints. During this discussion the topic of filter performance plays a central role: the expected value of the filter should be bounded and in the HMM case also preserve positivity to be a useful approximation of the drift. We show that this is not always given, especially when considering infrequent trading. It also depends on the number of stocks and the nature of the stocks involved.

These results on filter performance are then put into relation with portfolio optimization. It becomes apparent that these two issues are closely related, mostly steered by the same matrix in the Hidden Markov Model.
This diagonalization of this matrix then leads us to a market reduction which is a formulation of a type of Mutual Fund Theorem in this Hidden Markov Model. It describes time independent funds in which the trader invests instead of the higher number of stocks that were in the original portfolio with the same expected utility of optimal terminal wealth. A more general Mutual Fund Theorem can be found in [25], but that fund is then time dependent.

The basic question in the final part of the work is how much money a not fully informed investor needs to invest to reach a terminal wealth comparable to the fully informed investor's. Since we usually regard the expected utility of the optimal terminal wealth, the difference in the initial capital is called the loss in utility. The idea of computing the loss in utility was brought up by Rogers in [22], although he did not consider the models that are studied here. Brendle developed this idea further for Gaussian drift in [5].
This not fully informed investor comes in various degrees: first we consider the partially informed investor from before. In this framework we then apply the definition of the loss in utility for the first time and present the compu-

tations and results.
Then this not fully informed investor is relaxed, i.e. he trades infrequently. In [22], the relaxed investor is treated as well and the idea was translated into the partial information setting by Bäuerle, Urban and Veraart in [3].
The next issue is the number of stocks available to the investor. We discuss among others the influence of the number of stocks on the variance in the portfolio and the behaviour of the loss. High dimensional portfolios were also the topic of Gandy and Veraart in [12] although they did not consider the problem under the aspect of loss and not in the two models that are used here.
Due to the complex nature of the expressions for the optimal trading strategy we then consider the loss that arises if the trading strategy is approximated by much simpler terms.
For the Markov modulated drift we also observe the effect of a varying number of states in the case of partial information.
Finally we discuss the consequences of shortselling and borrowing constraints. This requires a second method to compute the loss that includes the trading strategy. For the computation of the constrained trading strategy we apply known methods for models with only one stock and models with multiple stocks with logarithmic utility and extend the multidimensional case to power utility. A condition on the parameters used in computations such that the constraints come into effect is also included. The topic of trading infrequently under shortselling and borrowing constraints is being discussed as well. In [3] constraints on shortselling and borrowing were also included. These constraints were introduced by Cvitanic and Karatzas in [7] and applied to the Markov model using the martingale method by Sass in [23].

All these considerations are illustrated in detail by numerical results including parameter estimation and its difficulties in the different settings.

This work is structured as follows. In chapter two we start with the presentation of the market model, the models for the drift and the general guidelines we used for numerical evaluations. The optimization problem is being discussed in chapter three where we derive the explicit expressions for the optimal terminal wealth and the trading strategy. In that chapter the topic of bankruptcy is introduced and then comes up with regard to the various changes in the model in chapters four and five. The filter performance and its connection to portfolio optimization are the topic of chapter four which also contains the discussion of the filter in various market settings and the market reduction. The loss in utility is then defined in the fifth chapter and it is being computed for various changes in the market parameters as well. The first case is the one of partial information, followed by infrequent trading. With this relaxed investor we analyze difficulties in the boundedness and positivity of the filter. Next an increase in the number of stocks is considered, leading to a discussion of the variance in the portfolio and of the general shape of the loss. We continue with an approximation of the optimal fraction of wealth invested in the stocks. An increase in the number of states is being regarded, naturally only for the Markov modulated drift. The subject is to determine the loss if the number of states in the estimation for partial information is unknown. Finally shortselling and borrowing constraints are treated and the parameters determined where such restrictions come into effect. We also consider trading infrequently with constraints.

2 Market Model

In a complete probability space $(\Omega, \mathcal{A}, P)$ we have a filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in[0,T]}$ in $\mathcal{A}$ satisfying the usual conditions.
We then consider a financial market consisting of one riskless and n risky assets. The riskless asset is called bond and a risky asset is a stock.
We have a finite time horizon $T > 0$ and the bond S^0 and the stocks $(S^1, \cdots, S^n)^T$ follow the dynamics

$$\begin{aligned} dS_t^0 &= rS_t^0 dt,\ S_0^0 = 1 \\ dS_t &= \mu_t \ \operatorname{diag}(S_t)dt + \sigma \ \operatorname{diag}(S_t)dW_t,\ S_0 = s_0 > 0 \end{aligned}$$

where $r \geq 0$ is the interest rate of the riskless asset. σ is the constant and nonsingular $n \times n$-dimensional volatility matrix and $W = (W_t)_{t\in[0,T]}$ is an n-dimensional Brownian Motion with respect to $\mathcal{F}$. $(\mu_t)_{t\in[0,T]}$ is the progressively measurable n-dimensional drift vector.
$\beta_t = e^{-rt},\ t \in [0, T]$ denotes the corresponding discount factor and we define the return process $R = (R_t)_{t\in[0,T]}$ as

$$dR_t = (\operatorname{diag}(S_t))^{-1} dS_t = \mu_t dt + \sigma dW_t$$

and the corresponding excess return process $\tilde{R} = (\tilde{R}_t)_{t\in[0,T]}$ as

$$d\tilde{R}_t = dR_t - r\mathbf{1}_n dt.$$

Using the above, one can write the excess return as

$$\tilde{R}_t = \int_0^t (\mu_s - r\mathbf{1}_n)\, ds + \int_0^t \sigma\, dW_t.$$

The filtration reflects information, for example the knowledge of the investor. In the case where the filtration is the one already introduced, $\mathcal{F} = (\mathcal{F}_t)_{0\leq t\leq T}$, we will speak of full information since the Brownian motion as well as the drift are observable.
In a second scenario, the investor will only be able to observe the prices which we call the case of partial information. To reflect this realistic assumption of only knowing the stock prices, we introduce a new filtration which is only based on this knowledge:

$$\mathcal{F}^S = (\mathcal{F}_t^S)_{0\leq t\leq T},\ \mathcal{F}_t^S = \sigma(S_s : 0 \leq s \leq t) \quad .$$

2.1 Model for the drift

Two different models for the drift will be considered: a Gaussian drift and then the case where the drift is modeled as a Markov chain.

Markov modulated drift Let $Y = (Y_t)_{t \in [0,T]}$ be a stationary, continuous and irreducible Markov chain independent of the Brownian motion W. Its state space are the unit vectors $\{e_1, \ldots, e_d\}$ and we introduce a corresponding $n \times d$-dimensional state matrix B. In this case we will then assume that $\mu_t = BY_t$, i.e. the drift is a Markov chain with the state matrix B.

The Markov chain Y_t will be characterized by its rate matrix Q.

Gaussian drift We will assume the drift process is the solution to the following stochastic differential equation

$$d\mu_t = \alpha(\delta - \mu_t)\ dt + \beta\ dW_t^{(2)}, \quad \mu_0 \sim \mathcal{N}(m_0, \gamma_0)$$

where W_t^2 is a Brownian motion with respect to our filtration $\mathcal{F}$ and independent of W_t driving the prices. α and β are real $n \times n$-matrices with β invertible and δ is a real n-dimensional vector. m_0 and γ_0 are known.
Furthermore, the parameters are chosen such that in addition to being progressively measurable, the condition

$$\int_0^T \|\mu_s\|^2\ ds < \infty, \quad a.s., \tag{2.1}$$

where $\|\cdot\|$ is the Euclidean norm, holds.

2.2 Numerical Setup

The numerical setup presented here will be used as the default, changes to it will be marked and introduced at occurence.
Our investor's bank account grants interest rates of $r = 0$ for all numerical purposes. This is not a limitation of the model, it is simply a shift.
The investor can additionally choose between two risky assets in our computations.

The volatility matrix will be

$$\sigma = \begin{pmatrix} 0.2 & 0.12 \\ 0.1 & 0.18 \end{pmatrix}$$

and for future reference

$$\sigma\sigma^T = \begin{pmatrix} 0.0544 & 0.0416 \\ 0.0416 & 0.0424 \end{pmatrix}$$

with an approximate correlation coefficient of 0.87.

The Markov chain has three states reflecting a very good market situation, a mediocre one and a negative one.

Markov parameter sets:

$$B = \begin{pmatrix} 1.2 & 0.1 & -1 \\ 1 & 0.1 & -0.8 \end{pmatrix}$$

(MP1)

$$Q = \begin{pmatrix} -60 & 25 & 35 \\ 15 & -30 & 15 \\ 35 & 25 & -60 \end{pmatrix}$$

This parameter set will be referred to as (MP1). This parameter set contains extreme states with frequent changes of states which allows us to observe properties of the market distinctively. We have to tolerate a high standard deviation though.

In addition we will therefore introduce a second parameter set denoted by (MP2), which has more realistic states, does not show effects as clearly and has a much lower deviation of the results.

$$B = \begin{pmatrix} 0.2 & 0.05 & -0.1 \end{pmatrix}$$

(MP2)

$$Q = \begin{pmatrix} -25 & 15 & 10 \\ 5 & -10 & 5 \\ 10 & 15 & -25 \end{pmatrix}$$

Gaussian parameter set:

$$\alpha = \begin{pmatrix} 0.7 & 0.1 \\ 0.1 & 0.7 \end{pmatrix}$$

(GP1)

$$\beta = \begin{pmatrix} 0.3 & 0.1 \\ 0.1 & 0.4 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

This set will accordingly be referred to as (GP1).
We will usually list and display numerical results as the mean from one hundred runs of the algorithms.

3 Framework

3.1 Optimization Problem

In this section we will describe the general optimization problem that will be studied from different angles throughout this work. But first we need to introduce the basic instruments.

Trading Strategy

3.1 Definition (trading strategy)**.** *A trading strategy* $\pi = (\pi_t)_{t\in[0,T]}$ *is an n-dimensional* $\mathcal{F}^S$*-progressively measurable process which satisfies* $\int_0^T ||\pi_t||^2\, dt < \infty$ *a.s.*

A trading strategy is the wealth invested in an asset, more precisely π_t^i is the wealth invested in the i-th asset at time t.
The fraction of wealth will be denoted by $f_t = \frac{1}{X_t^\pi}\pi_t,\ t \in [0,T]$.

The investor's wealth is a process $(X_t^\pi)_t, t \in [0,T]$ following

$$dX_t^\pi \ = \pi_t^T(\mu_t\ dt + \sigma\ dW_t) + (X_t - \mathbf{1}_n^T \pi_t)r\ dt, \quad X_0^\pi = x_0$$

where π_t is a trading strategy, $\mathbf{1}_n$ is the n-dimensional vector of ones and $x_0 > 0$ is the investor's initial capital.

A trading strategy π is called admissible, if
$P(X_t^\pi \geq K$ for all $t \in [0,T]) = 1$ for some constant $K > -\infty$. The set of all admissible trading strategies will be denoted by $\mathcal{A} = \{\pi | \pi \text{ admissible}\}$.

Utility Function

3.2 Definition (utility function)**.** *A function* u *with* $u : [0,\infty) \to \mathbb{R}\cup\{-\infty\}$*, will be called utility function if it is strictly increasing, strictly concave, twice continuously differentiable on* $(0,\infty)$ *and its derivative* u' *satisfies* $\lim_{x\to\infty} u'(x) = 0$ *and* $\lim_{x\to 0^+} u'(x) = \infty$*. The inverse function of* u' *will be denoted by* $I : (0,\infty) \to (0,\infty)$*.*

We will consider logarithmic and power utility during this work:

$$u_1(x) = \log(x), \quad u_1'(x) = \frac{1}{x}, \quad I_1(y) = \frac{1}{y}$$
$$u_2(x) = \frac{x^\kappa}{\kappa}, \quad u_2'(x) = x^{\kappa-1}, \quad I_2(y) = x^{\frac{1}{\kappa-1}}$$

with $\kappa \neq 0$ and $\kappa < 1$. In our numerical considerations we will usually use three different values for κ: -10 ,-1 and 0.1, in Figure 3.1 the utility functions are illustrated.

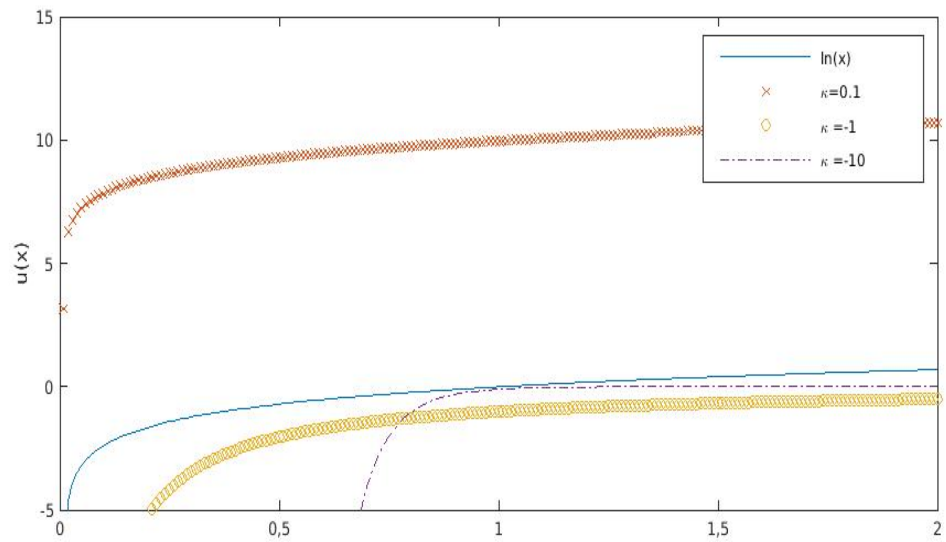

Figure 3.1: Utility functions

Risk Neutral Measure For the later optimization we will first change to the risk neutral measure $\tilde{\mathbb{P}}$.
The density process $Z = (Z_t)_{t\in[0,T]}$ is defined by

$$Z_t = \exp\left(-\int_0^t \sigma^{-1}(\mu_s - r\mathbf{1}_n)^T \, dW_s - \frac{1}{2}\int_0^t \|\sigma^{-1}(\mu_s - r\mathbf{1}_n)\|^2 \, ds\right) \quad .$$

Z is a martingale and the new measure $\tilde{\mathbb{P}}$ is then given by $\frac{d\tilde{\mathbb{P}}}{d\mathbb{P}} = Z_T$.
By Girsanov's Theorem a Brownian motion under the new measure can be defined as

$$\tilde{W}_t = W_t + \int_0^t \sigma^{-1}(\mu_s - r\mathbf{1}_n)\ ds \quad .$$

The process $Z^{-1} = (Z_t)^{-1}_{t\in[0,T]}$ is then a $\tilde{\mathbb{P}}$-martingale.
The expectation operator under the new measure $\tilde{\mathbb{P}}$ will be denoted by $\tilde{\mathbb{E}}$.

3.3 Remark. *The excess return process $d\tilde{R}_t = (\mu_t - r\mathbf{1}_n)dt + \sigma dW_t$ can be expressed in terms of the new measure $\tilde{\mathbb{P}}$ as $d\tilde{R}_t = \sigma d\tilde{W}_t$ and is also a Brownian motion under the new measure.*

We now can use the $\mathbb{P}$-martingale Z, but to do calculations in the case of partial information, we need to be able to operate in $\mathcal{F}^S$. Therefore we introduce the projection of Z

$$\zeta_t = \mathbb{E}[Z_t|\mathcal{F}_t^S] \quad ,$$

the conditional density process.

Optimization Problem With the previous definitions and results we can now state the optimization problem:

$$\sup_{\pi\in\mathcal{A}} \mathbb{E}[u(X_T^\pi)] \quad .$$

The trading strategy maximizing this expected value will then be the optimal trading strategy and denoted by π^* with the corresponding optimal fraction of wealth invested in stocks f^*. The terminal wealth affiliated with the optimal trading strategy is the optimal terminal wealth X_T^*.

[2] provide explicit expressions for the optimal trading strategy in the case of full information in the Markov case for logarithmic and power utility

$$\pi^*_{t,log} = (\sigma\sigma^T)^{-1}(\mu_t - r\mathbf{1}_n)\ X_t^*$$
$$\pi^*_{t,power} = \frac{1}{1-\kappa}(\sigma\sigma^T)^{-1}(\mu_t - r\mathbf{1}_n)\ X_t^* \quad .$$

Corresponding results for the Gaussian case are provided in [4].
Now we want to take a look at the case of partial information and we will

start with translating some of the quantities we defined for full information into the partial information setting.

In the case of partial information we cannot observe the drift directly, but only the process $\tilde{R}_t = \int_0^t (\mu_s - r\mathbf{1}_n)\, ds + \int_0^t \sigma\, dW_t$. We therefore are for the Markov model in the situation of a Hidden Markov Model where Y is the signal and $\tilde{R}$ the observation. For the Gaussian drift we are faced with a Kalman filtering problem.

3.2 Filtering in Hidden Markov Models

If one cannot observe the Markov chain itself, a simple approximation is its stationary distribution. It uses no informatin from observations and a definite improvement is a filter based on the observations that will be introduced next for Hidden Markov Models following [24]:

3.4 Definition. *We define the normalized filter* $\eta = (\eta_t)_{t\in[0,T]}$ *for* Y *as* $\eta_t = \mathbb{E}[Y_t|\mathcal{F}_t^S]$.
In addition to that, the unnormalized filter $\rho_t = (\rho_t)_{t\in[0,T]}$ *is defined by* $\rho_t = \tilde{\mathbb{E}}[Z_T^{-1}Y_t|\mathcal{F}_t^S]$.

The filter is the projection of the drift to the filtration with the information we have access to, namely $\mathcal{F}^S$. We will now derive an explicit representation of this filter following [10].

Explicit representation of the filter The rate matrix of a Markov chain satisfies

$$\frac{dp_t}{dt} = p_t Q$$

where p_t is the probability distribution $p_t = (p_t^1, \dots, p_t^d)^T$ with $p_t^i = \mathbb{P}(Y_t = e_i)$, $i = 1, \dots d$. The probability of jumping from state i to state j, $i, j = 1, \dots, d$, if the chain jumps, is given by $P_{i,j}$, the entry in position i, j of the transition matrix P. The transition matrix P belonging to this Markov chain can then be computed from the rate matrix:

$$P_{(t,s)} = \exp(Q(t-s)), \ s \leq t \quad .$$

We then have that

$$\frac{d}{dt}P_{(t,s)} = P_{(t,s)}Q \quad .$$

Y_t is a Markov chain and $\mathcal{Y}_t \subset \mathcal{F}_t$ where $\mathcal{Y}_t$ is the filtration generated by the Markov chain. Then for $s \leq t$

$$\mathbb{E}\big[Y_t|\mathcal{F}_s\big] = \mathbb{E}\big[Y_t|\mathcal{Y}_s\big] = P_{(t,s)}^T Y_s$$

which leads to the following martingale

$$M_t = Y_t - Y_0 - \int_0^t Q^T Y_u \, du$$

due to

$$\mathbb{E}\Big[M_t - M_s|\mathcal{F}_s\Big] = \mathbb{E}\Big[Y_t - Y_s - \int_s^t Q^T Y_r \, dr|\mathcal{Y}_s\Big]$$

$$= P^T_{(t,s)} Y_s - Y_s - \int_s^t Q^T P_{(r,s)} Y_s \, dr = 0 \quad .$$

Simply rearranging the results in the semimartingale representation of the Markov chain leads to

$$Y_t = Y_0 + \int_0^t Q^T Y_u \, du + M_t \quad .$$

For the complete representation we now need to perform the filtering into the filtration $\mathcal{F}^S$ and find out what M_t really is. This transition into the case of partial information will include the innovations process V_t.

3.5 Definition. *The innovation process is given by* $V = (V_t)_{t\in[0,T]}$, $dV_t = d\tilde{W}_t - \sigma^{-1}B\eta_t \, dt$.

3.6 Proposition. *The innovation process is an* $\mathcal{F}^S$*- Brownian motion under* $\mathbb{P}$.

Proof. This has been shown in a variety of books on the matter, see for example Propostion 2.30 in [1]. □

3.7 Remark. *The excess return process was discussed before as being* $d\tilde{R}_t = \sigma d\tilde{W}_t$. *In terms of the innovation process we now have* $d\tilde{R}_t = B\eta_t dt + \sigma dV_t$. *Under the filtration* $\mathcal{F}^S$ *this definition contains only observable quantities, thus with the help of the innovations process this definition relates to the full information setting. This conversion leads to valid results since the filtrations generated by the return and by the stock prices contain the same information and therefore results are optimal with regard to either.*

Let us now use the following theorem from [9] to learn more about the martingale.

3.8 Theorem. *Suppose* $\{M_t\}$, $M_0 = 0$ *is a square integrable* $\mathbb{P}$*-martingale with respect to the filtration* $\mathcal{F}_t$. *Then there is an* $\mathcal{F}_t$ *predictable process* γ_t *such that*

$$\int_0^t \mathbb{E}\left[|\gamma_t|^2\right] \, dt \;<\; \infty$$

and

$$M_t = \int_0^t \gamma_s \; dV_s \quad a.s.$$

Proof. This is Theorem 16.22 from [9]. □

We will denote processes with respect to the filtration $\mathcal{F}^S$ by ˆ and have now reached the following form of the filter:

$$\hat{Y}_t = \hat{Y}_0 + \int_0^t \; Q^T \hat{Y}_u \; du + \int_0^t \gamma_s \; dV_s \quad .$$

The next task is therefore to find the form of the process γ_t. This representation should ensure that representing it with respect to V_t in $\mathcal{F}^S$ and representing it with respect to W_t and then projecting it onto $\mathcal{F}^S$ leads to the same result. There is one process of which we already know that such a transition exists, namely the return:

$$dR_t = \mu_t dt + \sigma dW_t = \sigma d\tilde{W}_t = B\eta_t dt + \sigma dV_t \quad .$$

We will now use that knowledge to compute γ by computing the product with this process in both situations and identify γ by ensuring equality.

$$\begin{aligned} d(R_t Y_t) &= R_t Q^T Y_t dt + Y_t \mu_t dt + Y_t \sigma dW_t + R_t M_t + M_t \sigma dW_t \\ \widehat{d(R_t Y_t)} &= R_t Q^T \hat{Y}_t dt + \widehat{Y_t \mu_t} dt + \hat{Y}_t \sigma dW_t + \widehat{R_t M_t} + \hat{M}_t \sigma dW_t \end{aligned}$$

and the second procedure

$$d(\hat{R}_t \hat{Y}_t) = \hat{R}_t Q^T \hat{Y}_t dt + \hat{Y}_t B \eta_t dt + \hat{Y}_t \sigma dV_t + \hat{R}_t \gamma_t dV_t + \gamma_t \sigma dt \quad .$$

The martingale from the unfiltered expression translates into the integral with respect to V_t.
The term $\widehat{Y_t \mu_t} = \widehat{Y_t B Y_t}$ and since the Markov chain is twice in the same state, we can transform this into the sum of $Diag(g_i)\hat{Y}_t$, where g_i is the corresponding entry in the product of the inverse of the volatility matrix σ^{-1} and the state matrix $\tilde{B}$.
In one term the normalized filter η_t already appears, its definition was $\eta_t = \mathbb{E}\left[Y_t | \mathcal{F}^S\right]$, so this is the projection of the Markov chain into partial information, which is exactly what we denoted by $\hat{Y}_t$ in this derivation.
All these ingredients now accumulate to an expression for γ_t.

$$\gamma_t = Diag(g_i)\eta_t - (g_i^T \eta_t)\eta_t$$

and the normalized filter will be given by

$$\eta_t = \mathbb{E}[Y_0] + \int_0^t Q^T \eta_s ds + \sum_{i=1}^n \int_0^t \left(Diag(g_i)\eta_s - (g_i^T \eta_s)\eta_s \right) dV_s^i \quad .$$

Now we will turn towards the unnormalized filter and first establish its connection to the normalized filter by Bayes' law.

$$\rho_t = \tilde{\mathbb{E}}[Z_T^{-1} Y_t | \mathcal{F}_t^S] = \frac{\mathbb{E}\left[Z_T Z_T^{-1} Y_t | \mathcal{F}_t^S\right]}{\mathbb{E}\left[Z_T | \mathcal{F}_t^S\right]} = \frac{\mathbb{E}\left[Y_t | \mathcal{F}_t^S\right]}{\zeta_t} = \zeta_t^{-1} \eta_t$$

Computing this product on the right hand side provides us with an explicit characterization of ρ:

$$\rho_t = \mathbb{E}[Y_0] + \int_0^t Q^T \rho_s \; ds + \int_0^t \mathrm{diag}(\rho_s) \tilde{B}^T (\sigma\sigma^T)^{-1} \; d\tilde{R}_s \; ,$$

where $\tilde{B} = B - r\mathbf{1}_n$.
For all numerical computations we need a discretized version of this unnormalized filter and we wish for it to be robust.
If the filter is then Lipschitz continuous, it cannot react strongly since no jumps are possible. Thus we design filters that are locally Lipschitz continuous regarding the observation and call them robust. These robust versions were first introduced in [6] and then discretized in [14].
The first step in robustification is to find a process, that eliminates stochastic integrals in the filter if multiplied with it. For the unnormalized filter this would be

$$\begin{aligned} &\Phi_t = Diag(\phi_t) \quad \text{where} \\ &\phi_t^j = \exp\{\int_0^t \tilde{B^j}^T (\sigma\sigma^T)^{-1} \; d\tilde{R}_s - \frac{1}{2}\int_0^t \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B^j} \; ds\} \end{aligned}$$

with $j = 1, \ldots, d$ and $\tilde{B^j}$ being the j-th column of $\tilde{B}$.
Next we define

$$\bar{\rho}_t = \Phi_t^{-1} \rho_t \quad .$$

First we will take a closer look at this product with Ito and find out that Φ_t was chosen such that we get the simple relation

$$d(\bar{\rho}_t) = \Phi_t^{-1} Q^T \Phi_t \bar{\rho}_t dt \quad .$$

This version $\bar{\rho}_t$ is locally Lipschitz and thus of the desired form. It also leads to a robust version of the normalized filter which preserves the filter itself

$$\eta_t = \frac{\Phi_t \bar{\rho}_t}{\mathbf{1}_d^T \Phi_t \bar{\rho}_t}$$

since

$$\eta_t = \rho_t \zeta_t = \frac{\rho_t}{\mathbf{1}_d^T \rho_t} \quad .$$

This equation can now be discretized with the help of an Euler scheme

$$\bar{\rho}_{t_k} \approx \bar{\rho}_{t_{k-1}} + \Delta t \Phi_{t_{k-1}}^{-1} Q^T \Phi_{t_{k-1}} \bar{\rho}_{t_{k-1}}$$

where $k = 1, \dots, t/\Delta t$.
It appears in nicer form by introducing the process $\psi_{t_k} = \Phi_{t_k} \Phi_{t_{k-1}}^{-1}$ as

$$\rho_{t_k} = \psi_{t_k} \left(Id + \Delta t Q^T \right) \rho_{t_{k-1}} \quad . \tag{3.1}$$

3.3 Kalman Filtering

In the Gaussian model the drift is described as the solution of the stochastic differential equation

$$d\mu_t = \alpha(\delta - \mu_t)\ dt + \beta\ dW_t$$

with starting value μ_0 following an n-dimensional normal distribution with mean vector m_0 and covariance matrix γ_0 which are assumed to be known. The observation is the return process again and we are searching for the optimal filtering of these. This situation has been studied thoroughly already and was treated in several publications, among these are [1] and [20]. The aim in filtering is as always to find the normalized conditional distribution which can be gained by using the Kallianpur-Striebel formula on the unnormalized conditional distribution in form of the Zakai equation which then leads to the well known Kushner- Stratonovich equation for the normalized conditional distribution π_t as in $\pi_t\varphi = \mathbb{E}[\varphi|\mathcal{F}_t^S]$.

$$\begin{aligned}\pi_t(\varphi) =& \pi_0(\varphi) + \int_0^t \pi_s(A\varphi)ds \\ &+ \int_0^t \left(\pi_s(\varphi h^T) - \pi_s(h^T)\pi_s(\varphi)\right)(dR_s - \pi_s(h)ds)\end{aligned}$$

where A is the generator of the signal, $\mathcal{D}(A)$ its domain and $\varphi \in \mathcal{D}(A)$. The quantity h in this general equation is the factor of the drift in the observation process. Since the observation is the return here, this is just the identity. The signal on the other hand is μ and its generator for the one dimensional drift is

$$Af(x) = \alpha(\delta - x)f'(x) + \frac{\sigma^2}{2}f''(x) \quad .$$

The first valuable observation on the conditional distribution is then the following lemma.

3.9 Lemma. *The conditional distribution is also Gaussian and therefore determined by conditional mean vector and covariance matrix.*

The quantities to be determined are thus

$$\begin{aligned}&m_t = \mathbb{E}[\mu_t|\mathcal{F}_t^S] \quad \text{and} \\ &\gamma_t = \mathbb{E}[(\mu_t - m_t)(\mu_t - m_t)^T|\mathcal{F}_t^S] \quad .\end{aligned}$$

Following from the Kushner-Stratonovich equation and the usage of the generator, they are given by the Kalman-Bucy filter as in [18]:

$$dm_t = [-\alpha - \gamma_t(\sigma\sigma^T)^{-1}]m_t \ dt + \gamma_t(\sigma\sigma^T)^{-1} \ dR_t + \alpha\delta dt \ ,$$
$$d\gamma_t = -\gamma_t(\sigma\sigma^T)^{-1}\gamma_t - \alpha\gamma_t - \gamma_t\alpha^T + \beta\beta^T$$

with initial condition (m_0, γ_0) from above.

In [18] an additional useful equation is provided for the filter using the process ϕ which is the solution of

$$\phi'(t) = [-\alpha - \gamma(t)(\sigma\sigma^T)^{-1}]\phi(t)$$

with the initial condition that $\phi(0)$ is the identity matrix. This representation then reads

$$m_t = \phi(t)\Big[m_0 + \int_0^t \phi^{-1}(s)\gamma_s(\sigma\sigma^T)^{-1} \ dR_s + \int_0^t \phi^{-1}(s) \ ds \ \alpha\delta\Big] \quad .$$

Unfortunately we cannot use this representation in our computations though, the process ϕ then goes either to zero or diverges and thus the usage of this representation (or its inverse) leads to unstable results.

The deterministic equation for the conditional covariance is an example of the well known Ricatti equation. In the one dimensional model this equation has an explicit solution, but unfortunately not in higher dimensions. The equation in the one dimensional model then reads

$$d\gamma_t = -\frac{\gamma_t^2}{\sigma^2} - 2\alpha\gamma_t + \beta^2$$

and has the solution

$$\gamma_t = \sqrt{C}\frac{C_1 \exp\left(2\frac{\sqrt{C}}{\sigma}t\right) + C_2}{C_1 \exp\left(2\frac{\sqrt{C}}{\sigma}t\right) - C_2} - \alpha\sigma^2 \quad \text{where}$$
$$C = \alpha^2\sigma^2 + \beta^2$$
$$C_1 = \sqrt{C}\sigma + \gamma_0 + \alpha\sigma^2$$
$$C_2 = -\sqrt{C}\sigma + \gamma_0 + \alpha\sigma^2 \quad .$$

In this case the process ϕ also has an explicit solution

$$\phi_t = \exp\left(-\alpha t - \frac{1}{\sigma^2}\int_0^t \gamma_s\ ds\right)$$

and with this version the filter then reads

$$m_t = \phi_t\Big[m_0 + \frac{1}{\sigma^2}\int_0^t \frac{\gamma_s}{\phi_s}\ dR_s + \alpha\delta\int_0^t \frac{1}{\phi_s\ ds}\Big] \quad .$$

The filters for the drift in both models are applicable in the case of partial information provided we know the parameters: B and Q for the Markovian version and α, β and δ for the Gaussian. In the case of partial information, we will thus perform parameter estimations for these that will be described in Section 3.7.

3.4 Expected Utility of the Optimal Terminal Wealth

In [17] an explicit expression for the expected utility of the terminal wealth was derived that we will now introduce for both models.
With the above filters we can now state explicit representations for the conditional density process ζ_t from [24] and [18] respectively.
For the Markov chain drift we use a representation of $\zeta_t^{-1} = \tilde{\mathbb{E}}[Z_t^{-1}|\mathcal{F}_t^S]$ which is an $\mathcal{F}^S$-martingale with respect to $\tilde{\mathbb{P}}$. We know from Bayes' Theorem that $\rho_t = \zeta_t^{-1}\eta_t$ and by definition of the normalized filter $\mathbf{1}_n^T\eta_t = 1$. Therefore we also have $\mathbf{1}_d^T\rho_t = \zeta_t^{-1}$. Remembering that $\mathbf{1}_d^T Q^T = 0$, we get this explicit representation of the inverse of the density process straight from the filter equation of the unnormalized filter:

$$\zeta_t^{-1} = 1 + \int_0^t (\tilde{B}\rho_s)^T(\sigma\sigma^T)^{-1}\, d\tilde{R}_s \quad .$$

In the second case, the Gaussian drift, we have the representation

$$\zeta_t^{-1} = 1 + \int_0^t \zeta_s^{-1}(m_s - r\mathbf{1}_n)^T(\sigma\sigma^T)^{-1}\, d\tilde{R}_s \quad .$$

3.10 Remark. *These two definitions for the process ζ_t are of the same structure: We keep in mind that the normalized filter η_t in the HMM is the equivalent to the Kalman filter in the Gaussian model and that the normalized filter is given by $\rho_t\zeta_t$.*

3.11 Remark. *Under the measure $\tilde{\mathbb{P}}$, the expected value of ζ_T is one given L^2-integrability of the integrand since the return process $\tilde{R}$ is a Brownian motion, see Remark 3.3.*

With these preparations we are now able to compute the optimal terminal wealth both for the fully and partially informed investor with the help of the following theorem. Quantities relating to an optimal solution are denoted by *.

3.12 Theorem. *For every constant $x \in (0,\infty)$, suppose that $\tilde{\mathbb{E}}[\beta_T I(x\beta_T\zeta_T)] < \infty$. The optimal terminal wealth is then given by $X_T^* = I(y\beta_T\zeta_T)$ where y is a constant determined by $\tilde{\mathbb{E}}[\beta_T I(y\beta_T\zeta_T)] = x_0$.*
The process X^* and the optimal trading strategy π^* are determined by

$$\beta_t X_t^* = \tilde{\mathbb{E}}[\beta_T X_T^*|\mathcal{F}_t^S] = x_0 + \int_0^t \beta_s {\pi_s^*}^T\sigma\, d\tilde{W}_s \quad .$$

Proof. This is Lemma 6.5. and Theorem 6.6. from [17].
I is the inverse function of the derivative of the utility function and is therefore continuous and strictly decreasing on $(0, \infty)$. Then $\tilde{E}[\beta_T I(x\beta_T\zeta_T)]$ as a function of x is as well for every constant $x \in (0, \infty)$. We know from the definition that $\lim_{x\to\infty} u'(x) = 0$ and $\lim_{x\to 0^+} u'(x) = \infty$ and therefore $\lim_{x\to\infty} \tilde{E}[\beta_T I(x\beta_T\zeta_T)] = 0$ and $\lim_{x\to 0^+} \tilde{E}[\beta_T I(x\beta_T\zeta_T)] = \infty$. x_0 is positive and thus we have the existence of a y such that $\tilde{\mathbb{E}}[\beta_T I(y\beta_T\zeta_T)] = x_0$.

Let us consider an arbitrary random variable R that is $\mathcal{F}_t^S$-measurable and also fulfills $\tilde{\mathbb{E}}[R] = x_0$. The conditional expected value $\tilde{\mathbb{E}}[R|\mathcal{F}_t^S]$ is then a martingale and can be represented as $M_t = M_0 + \int_0^t \beta_s {\pi_s}^T \sigma \, d\tilde{W}_s$.
From the definition of the wealth process

$$dX_t^\pi = (X_t^\pi r + \pi_t(\mu_t - r))dt + \pi_t \sigma dW_t), \;\; X_0^\pi = x_0$$

we also know that

$$\begin{aligned} \beta_t X_t^\pi &= x_0 + \int_0^t \beta_s \pi_s^T \sigma \, dW_s + \int_0^t \beta_s \pi_s (\mu_t - r\mathbf{1}_n) \, ds \\ &= x_0 + \int_0^t \beta_s \pi_s^T \sigma \, d\tilde{W}_s \end{aligned}$$

and thus there exists a trading strategy such that the wealth process X^π represents $I(y\beta_T\zeta_T)$.
In addition, the integral in the martingale representation is a supermartingale and thus for every admissible portfolio process π we have $\tilde{E}[X_T^\pi] \leq x_0$. Thus we found a representation for the optimal terminal wealth, it remains to show is that π from the above equation is a tradings strategy and its optimality.

Let now another candidate X° be defined by $X_T^\circ = I(y^\circ \beta_T \zeta_T)$. We then have $\tilde{E}[\beta_T X_T^\circ] = x_0 \geq \tilde{E}[X^\pi(T)]$.
u is a utility function, in particular it is concave. Thus for a c such that $E[u^-(c)] < \infty$ with $u^- = -u$ in case $u < 0$ and $u^- = 0$ otherwise, we have that $u(I(x)) \geq u(c) + xI(x) - xc$. Setting $x = y^\circ \beta_T \zeta_T$, we get $u(X_T^\circ) \geq u(c) - y^\circ \beta_T \zeta_T c$ and thus $u^-(X^\circ) \leq u^-(c) - y^\circ \beta_T \zeta_T c$. From this we get that π° satisfies $E[u^-(X^{\circ \pi^\circ}(T))] < \infty$. Thus it is a candidate for a trading strategy.
Let π be arbitrary satisfying the previous equation. By the first inequality,

again with $x = y^\circ \beta_T \zeta_T$ and $c = X^{\circ \pi^\circ}(T)$, we get
$u(X_T^\circ) \geq u(X^\pi(T)) + y^\circ \beta_T \zeta_T X_T - y^\circ \beta_T \zeta_T X^\pi(T)$ and thus the desired result that π° is in fact optimal. □

3.13 Example. *In the case of power utility we have*

$$u_2(x) = \frac{x^\kappa}{\kappa}, \quad u_2'(x) = x^{\kappa-1}, \quad I_2(y) = x^{\frac{1}{\kappa-1}}$$

and thus

$$X^*_{T,power} = (\hat{y}\beta_T\zeta_T)^{\frac{1}{\kappa-1}}$$

with

$$\hat{y}^{\frac{1}{\kappa-1}} = \frac{x_0}{\mathbb{E}[(\beta_T\zeta_T)^{\frac{\kappa}{\kappa-1}}]} \quad .$$

For the optimal terminal wealth we then get

$$\mathbb{E}[u_2(X_T^*)] = \mathbb{E}[\frac{1}{\kappa}(\hat{y}\beta_T\zeta_T)^{\frac{\kappa}{\kappa-1}}] = \frac{x_0^\kappa}{\kappa}\ \mathbb{E}[(\beta_T\zeta_T)^{\frac{\kappa}{\kappa-1}}]^{1-\kappa} \quad .$$

3.5 Optimal Trading Strategy for Partial Information

Next we want to study the optimal trading strategy in the case of partial information. Its existence was shown in Theorem 3.12, now we aim to find an explicit expression.

Markov Model We will start with the Markov Model and will denote quantities in partial information by $\hat{}$.
The optimal trading strategy with logarithmic utility can be computed from Theorem 3.12 by comparing

$$\beta_t X_t^* = x_0 + \int_0^t \beta_s (\pi_s^*)^T \sigma \, d\tilde{W}_s$$

and

$$X_t^* = x_0 \zeta_t^{-1} = x_0 + x_0 \int_0^t (B\rho_s)^T (\sigma\sigma^T)^{-1} \, d\tilde{R}_s$$

resulting in

$$\hat{\pi}_{t,log}^* = x_0 (\sigma\sigma^T)^{-1} \tilde{B} \rho_t \beta_t^{-1} \quad .$$

The corresponding optimal fraction invested in the stocks is then given by

$$\hat{f}_{t,log}^* = (\sigma\sigma^T)^{-1} \tilde{B} \eta_t \ .$$

The case of logarithmic utility has once again an easy, straight forward solution which is not given for other utility functions, especially not for power utility. For these we will need some basics of the computation of Malliavin derivatives.

Malliavin Derivative In this paragraph we will introduce and present results on the Malliavin derivative as can be found in [15].

3.14 Definition. *Let $C_b^\infty(\mathbb{R}^m)$ be the set of C^∞-functions $f : \mathbb{R}^m \to \mathbb{R}$ that are bounded and have bounded derivatives of all orders. Additionally, let F be random variables of the form $F = f(W_{t_1}, \ldots, W_{t_k})$, where $(t_1, \ldots, t_n) \in [0,T]^n$ and $f(x^{11}, \ldots, x^{d1}, \ldots, x^{1n}, \ldots, x^{dn})$ belongs to $C_b^\infty(\mathbb{R}^{dn})$. These random variables build the class of smooth functionals S.*

The Malliavin derivative $DF = (D^1F, \dots, D^nF)^T$ *is then defined as a* $(L^2([0,T]))^d$*-valued random variable given by*

$$D^iF_t = \sum_{l=1}^{k} \frac{\partial}{\partial x_{li}} f(W_{t_1}, \dots, W_{t_k}) \mathbf{1}_{[0,t_l]}(t), \quad i = 1, \dots, n \quad .$$

3.15 Definition. *Denote by* $\mathbb{D}_{p,1}$ *the Banach space which is the closure of* S *under* $\|\cdot\|_{p,1}$ *with*

$$\|F\|_{p,1} = \|F\|_p + \| \ \|DF\|_{L^2} \ \|_p \quad .$$

3.16 Remark. [26] shows that DF is well defined on $\mathbb{D}_{p,1}$ by closure for any $p \geq 1$.

We can now introduce Clark's formula in this setting, again quoting from [15].

3.17 Theorem. *For every* $F \in \mathbb{D}_{1,1}$ *we have*

$$F = \mathbb{E}[F] + \int_0^T \mathbb{E}[(D_tF)^T | \mathcal{F}_t] dW_t.$$

Another result from [13] as it was presented in [24] is needed regarding Malliavin derivatives.

3.18 Theorem. *For* $d \in \mathbb{N}$ *we consider the* d*-dimensional SDE*

$$dX_t = f^\mu(t, X_t)\ dt + f^\sigma(t, X_t)\ dW_t, \quad t \in [0,T], \quad X_0 = x_0, \tag{3.2}$$

assuming that $x_0 \in \mathbb{R}^d$, f^μ *and* f^σ *are measurable* $\mathbb{R}$ *and* $\mathbb{R}^{d\times n}$*-valued functions which are continuously differentiable and satisfy*

$$\sup_{t\in[0,T], x\in\mathbb{R}^d} \left(|\frac{\partial}{\partial x_k} f_i^\mu(t,x)| + |\frac{\partial}{\partial x_k} f_{ij}^\sigma(t,x)| \right) \ < \ \infty,$$
$$\sup_{t\in[0,T]} \left(|f_i^\mu(t,0)| + |f_{ij}^\sigma(t,0)| \right) \ < \ \infty$$

for $i,k = 1, \dots, d,\ j = 1, \dots, n$. *Then (3.2) has a unique continuous solution* $(X_t)_{t\in[0,T]}$ *which satisfies* $X_s^k \in \mathbb{D},\ k = 1, \dots, d,$

$$\begin{aligned} D_tX_s = (f^\sigma(t,X_t))^T &+ \int_t^s D_tX_u (\partial_x f^\mu(u,X_u))^T\ du \\ &+ \int_t^s D_tX_u \sum_{j=1}^{n} (\partial_x f_{\cdot j}^\sigma(u,X_u))^T\ dW_u \end{aligned}$$

for $t \in [0,s]$ and $D_t X_s = 0$ for $t \in (s,T]$. Here ∂_x denotes the Jacobi matrix, i.e. $(\partial_x f^\mu)_{ij} = \frac{\partial}{\partial x_j} f_i^\mu$ and $f_{\cdot j}^\sigma$ is the j-th column of f^σ.

Application of this last theorem then leads to the Malliavin derivative of the unnormalized filter

$$D_t\rho_u = \sigma^{-1}\tilde{B}\,\mathrm{diag}(\rho_t) + \int_t^u (D_t\rho_s)Q\ ds + \int_t^u (D_t\rho_s)\,\mathrm{diag}(\tilde{B}^T(\sigma\sigma^T)^{-1}\ d\tilde{R}_s)$$

for all $u \in [0,T]$ and with $D_t\rho_u = 0$ for $t \in (u,T]$. More details on this can be found in [24].

This derivative can now be robustly discretized as we did with the unnormalized filter following again [24] leading to the recursion

$$(D_t\rho_u)_{t_k} = \psi_{t_{k-1}}\left(Id_d - \Delta t Q^T\right)(D_t\rho_u)_{t_{k-1}}$$

with $k = 1, \ldots, t/\Delta t$.

We will also need the Malliavin derivative of the density process ζ_t and its inverse.

We know already that $\zeta_T^{-1} = \mathbf{1}_d^T\rho_T$ and therefore $D_t\zeta_T^{-1} = (D_t\rho_T)\mathbf{1}_d$. The Malliavin derivative of the unnormalized filter was already established and keeping in mind that $Q\mathbf{1}_d = 0$, we get from [24]

$$D_t\zeta_T^{-1} = \sigma\tilde{B}\rho_t + \int_t^T (D_t\rho_s)\tilde{B}^T(\sigma\sigma^T)^{-1}\ d\tilde{R}_s \quad .$$

Using the chain rule, the result then is

$$D_t\zeta_T = -(\zeta_T)^2 D_t\zeta_T^{-1}$$

and

$$D_t\tilde{\zeta}_T = -\beta_T^{-1}(\tilde{\zeta}_T)^2 D_t\zeta_T^{-1} \quad .$$

With these preparations we can now prove Theorem 4.5 in [24] which will then allow us to compute the optimal trading strategy.

3.19 Theorem. *If $\hat{y}$ is the unique number which satisfies $\tilde{\mathbb{E}}[\beta_T I(\hat{y}\tilde{\zeta}_T)] = x_0$, if $\tilde{\mathbb{E}}[\beta_T I(x\tilde{\zeta}_T)] < \infty$ for all $x \in (0,\infty)$ and if $I(\hat{y}\tilde{\zeta}_T) \in L^q(\tilde{\mathbb{P}})$ for some $q > 1$, then $X_T^* = I(\hat{y}\tilde{\zeta}_T)$ and the optimal trading strategy $(\pi_t^*)_{t\in[0,T]}$ is given by*

$$\pi_t^* = \frac{\beta_t^{-1}}{\hat{y}}(\sigma\sigma^T)^{-1}\Big(\tilde{B}\rho_t\tilde{\mathbb{E}}[\psi(\hat{y}\tilde{\zeta}_T)|\mathcal{F}_t^S] \tag{3.3}$$

$$+\,\tilde{\mathbb{E}}[\psi(\hat{y}\tilde{\zeta}_T)\int_t^T(\sigma D_t\rho_s)\tilde{B}^T(\sigma\sigma^T)^{-1}d\tilde{R}_s|\mathcal{F}_t^S]\Big) \tag{3.4}$$

where $\psi(y) = -y^2 I'(y)$.

Proof. (Sketch)
Using Clark's formula from the paragraph on Malliavin derivatives leads to

$$\beta_T X_T^* = x_0 + \int_0^T \tilde{\mathbb{E}}[(D_t(\beta_T X_T^*))^T | \mathcal{F}_t^S] \, d\tilde{W}_t \quad .$$

We also know from Theorem 3.12 that $\beta_T X_T^* = \beta_T I(\hat{y}\tilde{\zeta}_T)$ and applying the Chain rule leads to

$$D_t(\beta_T X_T^*) = \hat{y}\beta_T I'(\hat{y}\tilde{\zeta}_T) D_t \tilde{\zeta}_T \quad .$$

Comparing the two representations of the optimal terminal wealth we get

$$\beta_t \sigma^T \pi_t^* = \tilde{\mathbb{E}}[D_t(\beta_T X_T^*) | \mathcal{F}_t^S] \quad .$$

This Malliavin derivative was already computed and with it we have

$$\beta_t \sigma^T \pi_t^* = \tilde{\mathbb{E}}\Big[\hat{y}(\tilde{\zeta}_T)^2 I'(\hat{y}\tilde{\zeta}_T) D_t(\zeta_T^{-1}) | \mathcal{F}_t^S\Big] \quad .$$

With the Malliavin derivative of the inverse of ζ_T and the definition of ψ this is the desired result. □

Application of this theorem then leads to

$$\begin{aligned} \hat{\pi}_{t,power}^* = & \frac{x_0 \beta_t^{-1}(\sigma\sigma^T)^{-1}}{(1-\kappa)\tilde{\mathbb{E}}[\beta_T \zeta_T^{\frac{1}{\kappa-1}}]} \Big\{ \tilde{B}\rho_t \tilde{\mathbb{E}}[\beta_T \zeta_T^{\frac{\kappa}{\kappa-1}} | \mathcal{F}_t^S] \\ & + \tilde{\mathbb{E}}[\beta_T \zeta_T^{\frac{\kappa}{\kappa-1}} \int_t^T (\sigma D_t \rho_s) \tilde{B}^T (\sigma\sigma^T)^{-1} \, d\tilde{R}_s | \mathcal{F}_t^S] \Big\} \quad . \end{aligned}$$

Gaussian Model In the Gaussian Model we find in [18] the following equations for the optimal trading strategy in partial information. For the investor with logarithmic utility we use

$$\hat{\pi}_{t,log}^* = x_0 \beta_t^{-1} (\sigma\sigma^T)^{-1} \frac{1}{\zeta_t}(m_t - r\mathbf{1}_n) \ .$$

It can be computed in much the same way by using 3.12. The optimal trading strategy for power utility presents itself to be much more complicated in this case as well

$$\hat{\pi}_{t,power}^* = \frac{x_0 \beta_t^{-1}(\sigma\sigma^T)^{-1}}{1-\kappa} \frac{1}{\zeta_t}(m_t - r\mathbf{1}_n) + G_t \quad , \tag{3.5}$$

where

$$G_t = y^{\frac{1}{\kappa-1}} \frac{1}{1-\kappa} \exp\left(rt + \frac{rT\kappa}{1-\kappa}\right) \frac{1}{\zeta_t} (\sigma\sigma^T)^{-1} \gamma(t)(\phi_t)^{-1}$$
$$\times \mathbb{E}\left[\zeta_T^{\frac{\kappa}{\kappa-1}} \int_t^T \phi^T(u)(\sigma^T)^{-1}\, d\tilde{W}_u | \mathcal{F}_t^S\right]$$

and ϕ is the solution of

$$\phi'(t) = [-\alpha - \gamma(t)(\sigma\sigma^T)^{-1}]\phi(t)$$

with the initial condition that $\phi(0)$ is the identity matrix.
The computation follows also [18] and the same path we presented for the Markov case. We compute the Malliavin derivative of the drift by first transforming it into a more suitable form

$$m_u = \phi_u \Big[m_0 + \alpha\delta \int_0^t \phi_s^{-1}\, ds + \int_0^u \phi_s^{-1}\gamma_s(\sigma^T)^{-1} d\tilde{w}_s$$
$$+ r\left(\int_0^u \phi_s^{-1}\gamma_s\, ds\right)(\sigma\sigma^T)^{-1}\mathbf{1}_n\Big]$$

and then determine its Malliavin derivative

$$D_t m_u = \sigma^{-1}\gamma_t(\phi_t^T)^{-1}\phi_u \mathbf{1}_{t\leq u} \quad .$$

Again we also need the Malliavin derivative of the density ζ which is given by

$$D_t\zeta_T = \zeta_T\Big[-\int_t^T (D_t m_u)(\sigma^T)^{-1} d\tilde{w}_u - \sigma^{-1}(m_t - r\mathbf{1}_n)$$
$$+ \int_t^T (D_t m_u)(\sigma\sigma^T)^{-1}(m_u - r\mathbf{1}_n)\, du\Big] \quad .$$

The last ingredient needed is

$$D_t I(x\zeta_T) = xI'(x\zeta_T)D_t\zeta_T$$

and we can apply Clark's formula again to the relation

$$\beta_t X_t^* = x_0 + \int_0^t \beta_s(\pi_s^*)^T\sigma\, d\tilde{W}_s$$

and use the Malliavin derivatives to achieve the above representation of the optimal trading strategy in partial information for power utility.

3.6 Bankruptcy

Whenever the investor's wealth diminishes to zero for the first time in an investment period, the investor is bankrupt.
In this case, the utility of this wealth is set by definition to $-\infty$ (except for power utility with $\kappa \in (0,1)$).
If this situation arises, the whole numerical experiment cannot be successful since we usually consider the empirical expected value of the utility which is then of course also $-\infty$ regardless of other possibly positive outcomes.

Occurrence of bankruptcy Using the characterization of the optimal wealth process from Theorem 3.12 and the bankruptcy requirement of it being less than zero, we get the following condition for logarithmic utility in both models

$$X_t^* = \frac{x_0}{\zeta_t \beta_t} \ > \ 0 \quad .$$

The process ζ_t is an exponential martingale and thus in theory it is always positive and bankruptcy cannot occur. In computations we use a different representation though which we introduced earlier that can very well become negative due to approximation and discretization.
For power utility the optimal wealth can be computed again by using Theorem 3.12 as

$$X_t^* = \frac{x_0^\kappa}{\kappa} \mathbb{E}\Big[(\beta_t \zeta_t)^{\frac{\kappa}{\kappa-1}}\Big]^{1-\kappa} \ > \ 0 \quad .$$

The initial capital x_0 is positive by definition and thus bankruptcy will again not be possible if ζ_t is positive as it should be by definition. If it were computed as a negative number, the terminal wealth would not even be defined though.

3.7 Parameter Estimation

In both models a version of the EM algorithm will be used. They are both based on the work by [8] where a general version of the EM algorithm for parameter estimation is introduced as follows. The first step is to establish an equivalent measure regarding the observation, i.e. the measure under which the observation was most probable. The Radon-Nikodym derivative for this change will then lead to the maximum likelihood estimator of the parameters by maximizing its logarithm.

HMM Parameter estimation The estimation in the Markov Model, i.e. estimation of the state and rate matrix, can be done using the EM algorithm in case $\sigma\sigma^T$ is known. We can compute an estimate for $\sigma\sigma^T$ based on the observations because of the relation $\sigma\sigma^T = \frac{1}{t}[\tilde{R}]_t$. Only knowing $\sigma\sigma^T$ and not σ itself suffices for the use of the EM algorithm.

With the EM algorithm we aim to estimate the parameters of the Markov chain. These will be described by the occupation time O_t^k in state e_k at time t, the number of jumps N_t^{kl} from state e_k to state e_l with $k \neq l$ and the level integrals G given by

$$\begin{aligned} O_t^k &= \int_0^t Y_s^k \, ds \\ N_t^{kl} &= \int_0^t Y_{s-}^k \, dY_s^l \\ G_t^k &= \int_0^t Y_s^k (\sigma\sigma^T)^{-1} \, d\tilde{R}_s \quad . \end{aligned}$$

The Likelihood functions L_t^B and L_t^Q for B and Q are expressed in terms of these quantities. Maximizing $\mathbb{E}\big[\log(L_t)|\mathcal{F}_t^S\big]$ then leads to updates Q' and

B' of the estimates where

$$
\begin{aligned}
L_t^Q L_t^B &= L_t = \frac{dP'}{dP}|_{\mathcal{F}_t} \\
L_t^Q &= \exp\left(\int_0^t \sum_{\substack{k,l=1\\k\neq l}}^d (Q_{kl} - Q'_{kl}) Y_s^k \; ds\right) \prod_{\substack{k,l=1\\k\neq l}}^d \left(\frac{Q'_{kl}}{Q_{kl}}^{N_t^{kl}}\right) \\
L_t^B &= \exp\left(\int_0^t ((\tilde{B}' - \tilde{B})Y_s)^T (\sigma\sigma^T)^{-1} (d\tilde{R}_s - \tilde{B}Y_s \; ds)\right. \\
&\qquad \left. -\frac{1}{2}\int_0^t ((\tilde{B}' - \tilde{B})Y_s)^T (\sigma\sigma^T)^{-1} ((\tilde{B}' - \tilde{B})Y_s) \; ds\right) \\
Q'_{kl} &= \frac{\tilde{\mathbb{E}}[Z_t^{-1} N_t^{kl} | \mathcal{F}_t^S]}{\tilde{\mathbb{E}}[Z_t^{-1} O_t^k | \mathcal{F}_t^S]} \\
B'_{\cdot k} &= (\sigma\sigma^T)^{-1} \frac{\tilde{\mathbb{E}}[Z_t^{-1} G_t^k | \mathcal{F}_t^S]}{\tilde{\mathbb{E}}[Z_t^{-1} O_t^k | \mathcal{F}_t^S]} \quad .
\end{aligned}
$$

In the expectation step the filters of the above quantities are computed in the same fashion as the unnormalized filter for the Markov chain itself.

For a more detailed description of the algorithms we refer to [24].

3.20 Example. *The following Figures 3.2, 3.3, 3.4 and matrices show the results of one hundred estimations from one hundred data sets with our usual data for estimation with $T = 10$ years. The trading frequency in the simulations is again once per trading day, $\Delta t = 0.004$. Each estimation consists of one application of the EM algorithm described above with the true values as starting values. The topic of different starting values will be discussed in the next example. We observe that all estimations lead to very satisfying results. The results for (MP1) were:*

$$
\widehat{(\sigma\sigma^T)} = \begin{pmatrix} 0.0571 & 0.0437 \\ 0.0437 & 0.0440 \end{pmatrix} \quad .
$$

$$
\hat{B} = \begin{pmatrix} 1.095 & 0.1016 & -0.8914 \\ 0.9165 & 0.1029 & -0.7082 \end{pmatrix}
$$

$$\hat{Q} = \begin{pmatrix} -60.8522 & 25.5257 & 35.3264 \\ 14.7691 & -29.5273 & 14.7582 \\ 35.3946 & 25.5440 & -60.9386 \end{pmatrix}$$

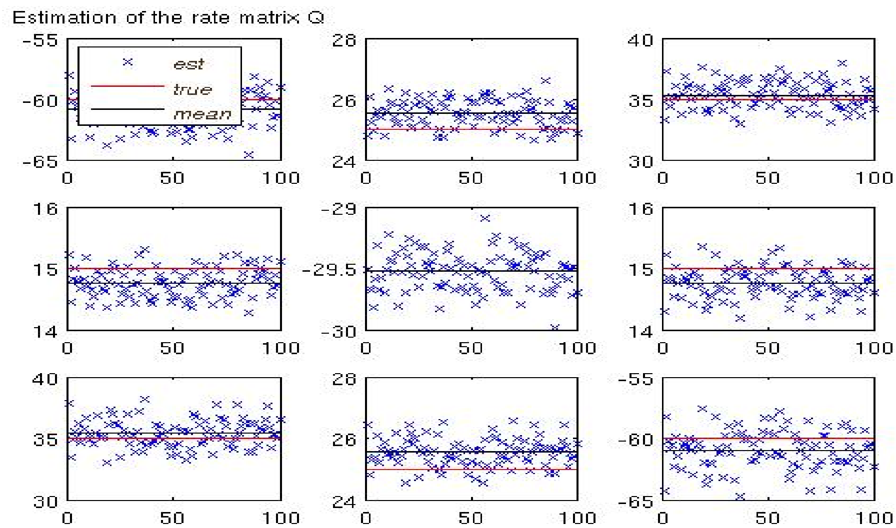

Figure 3.2: EM algorithm parameter estimation rate matrix Q, (MP1)

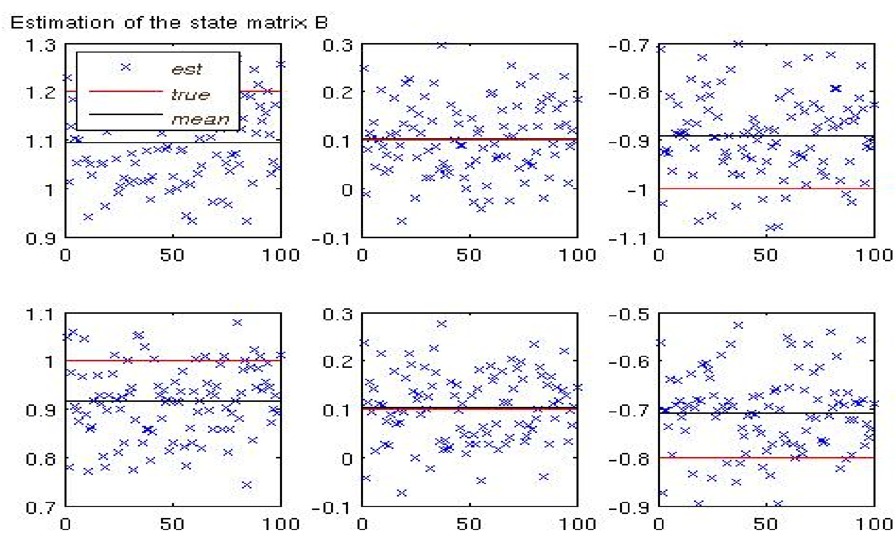

Figure 3.3: EM algorithm parameter estimation state matrix B, (MP1)

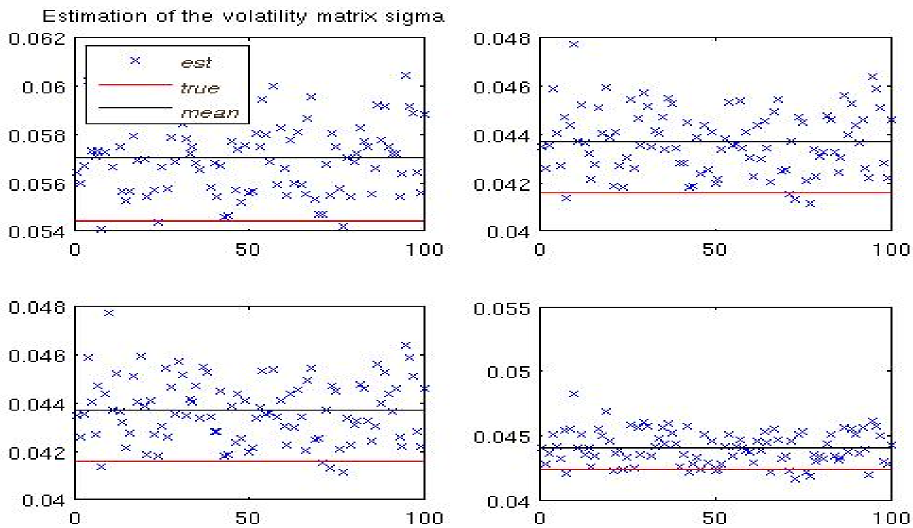

Figure 3.4: Parameter estimation volatility matrix σ

The above estimations were all done with the true parameters as starting values for the algorithm. That will be the case for all later applications due to the otherwise needed computational effort. If the starting values were not the true parameters, there is a risk of the algorithm terminating within a local extrema that is not a good approximation. To minimize that risk, one uses a grid of starting values and observes the estimations to choose the best fit. This results however in the mentioned noticeable computational effort. Nevertheless we want to present an example with guessed starting values next.

3.21 Example. *In this example the starting value for the EM algorithm will be a vector filled with ones. The true values are* $(15,\ -30,\ 15)$*, the middle row of the rate matrix, and are displayed in red in Figure 3.5. The starting values differ quite much from the true values to be estimated and the algorithm cannot cover such a distance in one run. Therefore one uses the result of one run as starting value for the next and thus the estimates change step by step although the data does not change. We did that 1000 times. This was an experiment not hitting a local extremum but converging to the true values. There is still some distance between the estimation and the true values, the displayed blue line is a mean of estimations and thus includes some deviations.*

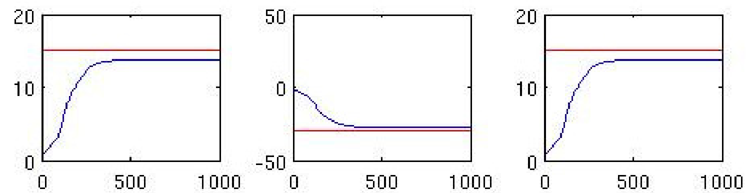

Figure 3.5: EM algorithm parameter estimation with arbitrary starting value for the second row of the rate matrix of (MP1)

Parameter estimation in linear Gaussian models We applied the algorithm from [11] for discrete time models, again based on the EM algorithm. To apply this algorithm, δ needs to be known and constant. We have set it to contain only zeros in (GP1) here. Other known and constant values would also be possible.

Left to estimate are the matrices α, β and σ. The algorithm will estimate $\beta\beta^T$ and $\sigma\sigma^T$ though. This suffices for later computations, one can compute a matrix square root if necessary with suitable parameters. All estimates will again be maximum likelihood estimates. The log- likelihood function is here given by

$$\begin{aligned} L_t^\theta = & - T\log(|\beta|) - (T+1)\log(\sigma) \\ & - \frac{1}{2}\mathbb{E}\Big[\sum_{l=1}^{T}(\mu_l - \alpha\mu_{l-1})^T(\beta\beta^T)^{-1}(\mu_l - \alpha\mu_{l-1})|\mathcal{F}_T^S\Big] \\ & - \frac{1}{2}\mathbb{E}\Big[\sum_{l=0}^{T}(R_l - \mu_l)^T(\sigma\sigma^T)^{-1}(R_l - \mu_l)|\mathcal{F}_T^S\Big] + \mathbb{E}[\mathcal{R}_{\hat{\theta}}|\mathcal{F}_T^S] \end{aligned}$$

where θ is the vector of parameters to be estimated and the remaining term $\mathcal{R}_{\hat{\theta}}$ that does not depend on θ (these are terms that vanish when differentiating for the maximization step). In the maximization step the argmax of this function will be computed resulting in updates of estimates following

$$\alpha_{j+1} = \mathbb{E}\Big[\sum_{l=1}^{T} \mu_l \mu_{l-1}^T | \mathcal{F}_T^S\Big] \times \left(\mathbb{E}\Big[\sum_{l=1}^{T} \mu_l \mu_{l-1}^T | \mathcal{F}_T^S\Big]\right)^{-1}$$

$$\beta\beta_{j+1}^T = \frac{1}{T}\mathbb{E}\Big[\sum_{l=1}^{T} \mu_l - \alpha_{j+1}\mu_{l-1} \times (\mu_l - \alpha_{j+1}\mu_{l-1})^T | \mathcal{F}_T^S\Big]$$

$$\sigma\sigma_{j+1}^T = \frac{1}{T+1}\mathbb{E}\Big[\sum_{l=0}^{T} R_l - \mu_l \times (R_l - \mu_l)^T | \mathcal{F}_T^S\Big] \quad .$$

The parameter estimation for the Gaussian parameters does not work as nicely. The following matrices and Figures (3.6, 3.7, 3.8) show results from one hundred runs and with one hundred data sets from (GP1), results are much more volatile as the plots show:

$$\hat{\alpha} = \begin{pmatrix} 0.41 & 0.14 \\ 0.14 & 0.65 \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} 0.62 & -0.38 \\ -0.38 & 0.39 \end{pmatrix}$$

$$\hat{\sigma} = \begin{pmatrix} 0.66 & -0.22 \\ -0.22 & 0.34 \end{pmatrix} \quad .$$

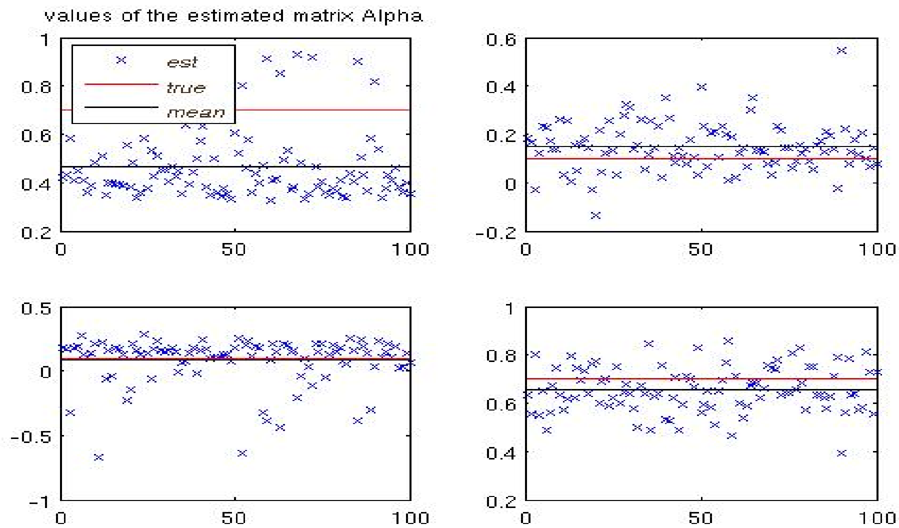

Figure 3.6: Parameter estimation matrix α, (GP1)

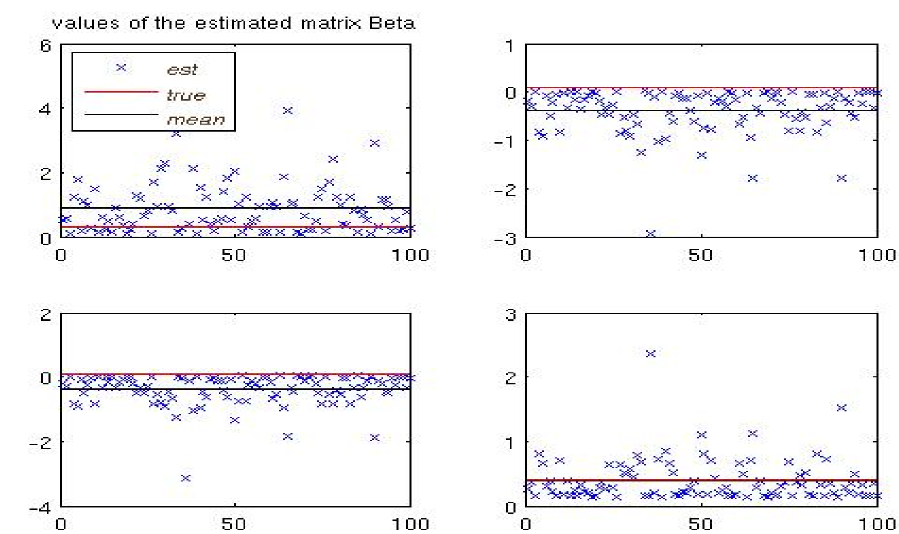

Figure 3.7: Parameter estimation matrix β, (GP1)

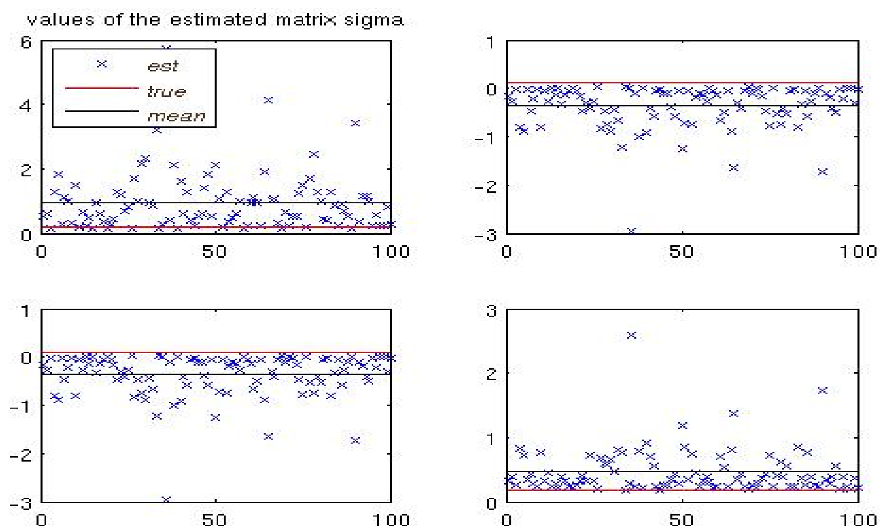

Figure 3.8: Parameter estimation volatility matrix

Utility of the Optimal Terminal Wealth with Parameter Estimation We will now first compare the optimal terminal wealth for the fully and the partially informed investor with and without parameter estimation. The comparison will consider four different cases: The first will be the fully informed investor, the second is about the investor who does not know the parameters but is apart from that fully informed, the third is for the partially informed investor who does know the parameters and the last deals with the partially informed investor without knowledge of the parameters.
We observe the expected difference in results: The fully informed investor is in a clear advantage compared to the partially informed. Also the parameter estimation leads to a slightly lesser terminal wealth.
The values for the expected utility of the optimal terminal wealth for (MP1) are displayed in Tables 3.1 and 3.2. They are once again the mean of one hundred estimations

utility function		full information true parameters	full information est. parameters
log	$\mathbb{E}[u(X_T^*)]$	6.69	4.90
	std	0.39	0.34
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	21.04	17.26
	std	0.85	0.60
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.35	-0.09
	std	0.03	0.05
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-777.66	-97878.73
	std	7757.06	96590.47

Table 3.1: Expected utility of the optimal terminal wealth of a fully informed investor (MP1)

The results for power utility with $\kappa = -10$ might seem unusual at first glance: These high standard deviations stem from the very steep increase of the utility functions on the interval $(0, 1)$. Minor deviations in the wealth then lead to huge differences in the utility.
It is also possible to compute an exact value for the expected utility of the optimal terminal wealth for the fully informed investor with parameter knowledge. With the help of the later introduced formulas from the proof of Proposition 4.14 numbered (4.1) and (4.3) for logarithmic utility we get an exact

value of 6.27 for parameter set (MP1). Let us now compare the partially informed investors.

utility function		partial info true parameters	partial info est. parameters
log	$\mathbb{E}[u(X_T^*)]$	0.69	0.51
	std	0.03	0.02
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	10.73	10.53
	std	0.035	0.027
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.51	-0.61
	std	0.014	0.013
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-0.0001	-0.0008
	std	0.00004	0.0001

Table 3.2: Expected utility of the optimal terminal wealth of a partially informed investor (MP1)

Less extreme values can be reached with a more moderate parameter set. Therefore we also include the results for parameter set (MP2).

utility function		full information true parameters	full information est. parameters
log	$\mathbb{E}[u(X_T^*)]$	0.15	0.23
	std	0.06	0.06
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	10.17	10.26
	std	0.06	0.06
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.93	-0.89
	std	0.06	0.05
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-0.11	-0.1
	std	0.07	0.08

Table 3.3: Expected utility of the optimal terminal wealth of a fully informed investor (MP2)

With this more moderate parameter set, we also do not detect these huge deviations for power utility with $\kappa = -10$ anymore. But they are still big in comparison.

utility function		partial info true parameters	partial info est. parameters
log	$\mathbb{E}[u(X_T^*)]$	0.028	-0.07
	std	0.022	0.05
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	10.03	9.94
	std	0.022	0.052
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.98	-1.1
	std	0.022	0.07
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-0.98	-1
	std	0.022	1.57

Table 3.4: Expected utility of the optimal terminal wealth of a partially informed investor (MP2)

For the Gaussian drift with (GP1) we observe generally the same structure regarding the success of the investors in Tables 3.5 and 3.6.

utility function		full information true parameters	full information est. parameters
log	$\mathbb{E}[u(X_T^*)]$	5.01	0.88
	std	0.59	0.17
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	28.3	11.11
	std	21	0.23
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.2	-0.69
	std	0.00	0.1
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-916.7	-1.02
	median	-0.0001	-0.03
	std	76348	6.77

Table 3.5: Expected utility of the optimal terminal wealth of a fully informed investor (GP1)

utility function		partial info true parameters	partial info est. parameters
log	$\mathbb{E}[u(X_T^*)]$	0.21	0.01
	std	0.01	0.001
power, $\kappa = 0.1$	$\mathbb{E}[u(X_T^*)]$	10.21	10.01
	std	0.01	0.001
power, $\kappa = -1$	$\mathbb{E}[u(X_T^*)]$	-0.81	-0.99
	std	0.01	0.001
power, $\kappa = -10$	$\mathbb{E}[u(X_T^*)]$	-0.01	-0.09
	std	0.001	0.001

Table 3.6: Expected utility of the optimal terminal wealth of a partially informed investor (GP1)

4 Advantage of Filtering

As briefly mentioned in the introduction to filtering, the filter is an improvement of the approximation of the drift that is just the starting value or the stationary distribution. The investor who filters is therefore intuitively in a better position than the one who assumes the starting value. This will be quantified in the following including a discussion of the prerequisites for such an advantage. The next chapter will also cover different views on this advantage in different market situations. In applications we usually consider the scheme for the unnormalized filter and we will now discuss when using it might be an advantage.

4.1 Filtering Advantage in the HMM

We will discuss for which market parameters the scheme for the unnormalized filter is bounded. If it were not bounded, then it is obviously not a good approximation and thus of no advantage.

The scheme for the unnormalized filter is given by

$$\rho_{t_k} = \psi_{t_k}(Id + \Delta t Q^T)\rho_{t_{k-1}} = (\prod_{i=1}^{k} \psi_{t_i}(Id + \Delta t Q^T))\rho_0$$

with $k = 1, \dots, t/\Delta t$ as it was introduced in Section 3.2, Formula (3.1). For simplicity, let $\rho_k := \rho_{t_k}$ be the scheme for the unnormalized filter at discretized time t_k. This is then a product of the process ψ, that depends on time, and two time independent factors: the starting value and $(Id + \Delta t Q^T)$ that appears in every step of the recursion. The starting value is the stationary distribution of the Markov chain, it is constant and does not influence the boundedness of the scheme of the unnormalized filter. The term boundedness will be properly introduced with the following two propositions, first for the risk neutral measure $\tilde{\mathbb{P}}$ and then for the real world measure $\mathbb{P}$.

4.1 Lemma. *The process ψ is componentwise of mean one under the risk neutral measure,*

$$\tilde{\mathbb{E}}[\psi_k^j] = 1 \quad .$$

Proof. The process ψ is defined by $\psi_k = \Phi_k \Phi_{k-1}^{-1}$ with

$$\Phi_k = Diag(\phi_k) \quad \text{where}$$
$$\phi_k^j = \exp\left\{ \int_0^k {\tilde{B}^j}^T (\sigma\sigma^T)^{-1} \, d\tilde{R}_s - \frac{1}{2} \int_0^k {\tilde{B}^j}^T (\sigma\sigma^T)^{-1} \tilde{B}^j \, ds \right\} \quad .$$

We can then use the mutliplicativity of the exponential function to write ψ componentwise as

$$\psi_k^j = \exp\left\{ \int_{k-1}^k {\tilde{B}^j}^T (\sigma\sigma^T)^{-1} \, d\tilde{R}_s - \frac{1}{2} \int_{k-1}^k {\tilde{B}^j}^T (\sigma\sigma^T)^{-1} \tilde{B}^j \, ds \right\} \quad .$$

The excess return can be written as $\sigma \tilde{W}$ and $\tilde{W}$ is a Brownian motion under the measure $\tilde{\mathbb{P}}$. Thus ψ can be represented as an exponential $\tilde{\mathbb{P}}$-martingale componentwise for $j = 1, \ldots, d$

$$\psi_k^j = 1 + \int_{k-1}^k (\tilde{B}^j)^T (\sigma\sigma^T)^{-1} d\tilde{R}_s \quad .$$

This process is of mean one under $\tilde{\mathbb{P}}$. □

Under the real world measure this is not true which is why we discuss the expected value of the scheme for the unnormalized filter in two separate propositions.

4.2 Proposition. *Under the risk neutral measure $\tilde{\mathbb{P}}$, the expected value of the scheme for the unnormalized filter is bounded by a finite bound $M < \infty$,*

$$\left\| \tilde{\mathbb{E}}[\rho_k] \right\| < M$$

with $k = 1, \ldots, t/\Delta t$, if

$$\left\| (\mathrm{Id} + \Delta t Q^T) \right\|^k < \frac{M}{\|\nu\|}$$

with a submultiplicative norm $\| \cdot \|$ and the stationary distribution of the Markov chain ν as starting value of the filter ρ_0.

Proof. The expected value of the scheme for the unnormalized filter under the risk neutral measure is given by

$$\tilde{\mathbb{E}}\big[\rho_k\big] = \tilde{\mathbb{E}}\Big[\prod_{i=1}^k \psi_i (\mathrm{Id} + \Delta t Q^T) \nu \Big] \quad .$$

The factors ψ_i are independent and of expected value equal to one under this measure. We will apply the norm now and use its submultiplicativity resulting in

$$\Big\|\mathbb{E}\big[\rho_k\big]\Big\| \leq \big\|(\mathrm{Id} + \Delta t Q^T)\big\|^k \cdot \|\nu\| \quad .$$

The norm of the initial distribution of the Markov chain is constant and introducing the bound $M < \infty$ then leads to the desired inequality

$$\big\|(\mathrm{Id} + \Delta t Q^T)\big\|^k < \frac{M}{\|\nu\|} \quad .$$

□

Let us first study another vital aspect to the advantage of filtering in the HMM, namely positivity. It will be needed for the result on boundedness under the real world measure. In the HMM model, the unnormalized filter is positive at all times. Unfortunately, this might not necessarily be true for the scheme for the filter.

4.3 Definition. *The robust discretized version of the unnormalized filter ρ_{t_k} is said to preserve positivity if it is positive for all times $t_k \in [0, T]$ and $k = 1, \ldots, T/\Delta t$.*

Again, this effect is closely related to the choice of the discretization step Δt and also the parameters. The following Lemma describes this relation.

4.4 Proposition. *The scheme preserves positivity if $(\mathrm{Id} + \Delta t\ Q^T)$ has only positive entries.*

Proof. This follows directly from the numerical representation. The process ψ is exponential and thus always positive as is the stationary distribution. □

4.5 Remark. *This is a very strict requirement and it might very well happen in a simulation, that the scheme preserves positivity even if it is not met. But to avoid negative values in one run at one time, which would destroy the whole simulation study, this requirement is very useful.*

Now we can state and prove the boundedness result under the real world measure.

4.6 Proposition. *Let* $(\mathrm{Id} + \Delta t\, Q^T)$ *have only positive entries. Under the real world measure and for a submultiplicative norm* $\|\cdot\|$*, the expected value of the scheme for the unnormalized filter is bounded by a finite bound* $M < \infty$*,*

$$\big\|\mathbb{E}[\rho_k]\big\| < M$$

with $k = 1, \ldots, t/\Delta t$*, if*

$$\max_j \big(\exp\big\{A_{jj} \Delta t\ k\big\}\big) \cdot \big\| \mathrm{Id} + \Delta t Q^T \big\|^k < \frac{M}{\|\nu\|} \quad ,$$

where $A = B^T(\sigma\sigma^T)^{-1}B$*.*

Proof. The expected value of the scheme for the unnormalized filter is

$$\mathbb{E}\big[\rho_k\big] = \mathbb{E}\Big[\prod_{i=1}^{k} \psi_i(\mathrm{Id} + \Delta t Q^T)\nu\Big] \quad .$$

The scheme is independent of the Markov chain, thus conditioning this expected value to the filtration induced by the Markov chain $\mathcal{F}_T^Y$ leaves us with

$$\mathbb{E}\big[\rho_k\big] = \mathbb{E}\Big[\mathbb{E}\Big[\rho_k \big| \mathcal{F}_T^Y\Big]\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\prod_{i=1}^{k} \psi_i(\mathrm{Id} + \Delta t Q^T)\nu \big| \mathcal{F}_T^Y\Big]\Big] \quad .$$

We know that the ψ_i, $i = 1, \ldots, k$ are independent given $\mathcal{F}_T^Y$. Let us thus have a closer look at the factors ψ_i and represent them again componentwise as in Lemma 4.1.

$$\begin{aligned}
\psi_k^j &= \exp\Big\{ \int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1}\ d\tilde{R}_s - \frac{1}{2}\int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B^j}\ ds \Big\} \\
&= \exp\Big\{ \int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B} Y_s\ ds \Big\} \\
&\quad \cdot \exp\Big\{ \int_{k-1}^{k} (\sigma^{-1}\tilde{B^j})^T\ dW_t - \frac{1}{2}\int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B^j}\ ds\} \Big\} \quad .
\end{aligned}$$

Given Y, the last factor is a martingale with mean one and thus by independence of W and Y

$$\mathbb{E}\big[\psi_k^j | \mathcal{F}_T^Y\big] = \exp\Big\{ \int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B} Y_s\ ds \Big\} \quad .$$

Let us then denote

$$c_k^j := \exp\left\{ \int_{k-1}^{k} \tilde{B^j}^T (\sigma\sigma^T)^{-1} \tilde{B} Y_s \; ds \right\}$$

and $c_k = \operatorname{diag}(c_k^j)$. Applying this to the expected value of the scheme leads by the tower property of conditional expectation to

$$\mathbb{E}\big[\rho_k\big] = \mathbb{E}\Big[\Big(\prod_{i=1}^{k} c_i (\mathrm{Id} + \Delta t Q^T)\Big)\nu\Big] \quad .$$

By positivity of $\mathrm{Id} + \Delta t Q^T$ we get

$$\mathbb{E}\big[\rho_k\big] \leq \max_j \big(\exp\big\{A_{jj} \Delta t \; k\big\}\big) \cdot \mathbb{E}\big[\big(\prod_{i=1}^{k} (\mathrm{Id} + \Delta t Q^T)\big)\nu\big]$$

since

$$\begin{aligned} c_k^j &\leq \max_{i,j} \big(\exp\big\{(B^j)^T (\sigma\sigma^T)^{-1} B^i \Delta t \; k\big\}\big) \\ &\leq \max_{j} \big(\exp\big\{(B^j)^T (\sigma\sigma^T)^{-1} B^j \Delta t \; k\big\}\big) \\ &= \max_{j} \big(\exp\big\{A_{jj} \Delta t \; k\big\}\big) \quad . \end{aligned}$$

For a submultiplicative norm $\|\cdot\|$ we then have

$$\|\mathbb{E}\big[\rho_k\big]\| = \max_j \big(\exp\big\{A_{jj} \Delta t \; k\big\}\big) \cdot \big\| \mathrm{Id} + \Delta t Q^T \big\|^k \cdot \|\nu\|$$

where $A = B^T (\sigma\sigma^T)^{-1} B$.
Introducing the finite bound $M < \infty$ and dividing by the constant norm of the stationary distribution $\|\nu\|$ then leads to the claim. □

4.7 Remark. *The scheme for the unnormalized filter*

$$\rho_k = \psi_k (Id + \Delta t Q^T) \rho_{k-1} = (\prod_{i=1}^{k} \psi_i (Id + \Delta t Q^T))\nu$$

will be bounded by one if $\|\nu\| \leq 1$ *and*
$\max_j \big(\exp\big\{A_{jj}\Delta t\big\}\big) \cdot \big\| \mathrm{Id} + \Delta t Q^T \big\| \leq 1.$

Boundedness of the term $\|(Id + \Delta t Q^T)^k\|$ depends on the trading frequency Δt. In case the rate matrix is diagonalizable, this diagonalization enables us to compute this Δt explicitly in a nice fashion.

4.8 Lemma. *For a diagonalizable rate matrix Q^T with $Q^T = T Q^T_{diag} T^{-1}$, the term $\|(Id + \Delta t Q^T)^k\|$ is bounded if*

$$\|Id + \Delta t Q^T_{diag}\| \leq 1$$

with a submultiplicative norm $\| \cdot \|$.

Proof. With the norm being submultiplicative we have

$$\begin{aligned} \|(Id + \Delta t Q^T)^k\| &= \|(Id + \Delta t T Q^T_{diag} T^{-1})^k\| = \|T(Id + \Delta t Q^T_{diag})^k T^{-1}\| \\ &\leq \|T\| \cdot \|(Id + \Delta t Q^T_{diag})^k\| \cdot \|T^{-1}\| \quad . \end{aligned}$$

This will be bounded if

$$\|(Id + \Delta t Q^T_{diag})\| \leq 1 \quad .$$

□

There is another factor influencing the boundedness, the process ψ. Its mean might be unbounded and in that case the scheme is unbounded as well. Figure 4.1 displays one simulation of the process ψ with increased mean. The figure is simulated with a parameter set with a diagonal volatility matrix to show this increased mean. The other parameters are taken from (MP1) and the usual simulation parameters were applied.

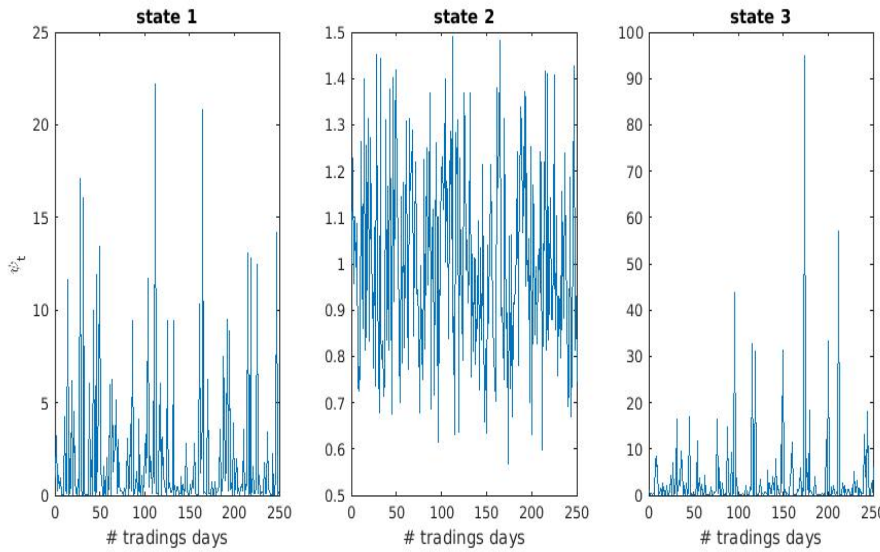

Figure 4.1: ψ at 25 stocks and diagonal volatility

In Proposition 4.6 this mean appears in the bound as the term $\| \max_j(\exp A_{jj}\Delta tk)\|$. Therefore controlling this term depends on the matrix $A = B^T(\sigma\sigma^T)^{-1}B$. This matrix will appear in various of the following paragraphs and its relation to filtering will be discussed in detail now.

Filtering and the matrix A We have now discussed all factors included in the scheme for the unnormalized filter and can thus ensure the boundedness and positivity of the filter by following the above rules.

Having the advantage of filtering is then only possible if all above criteria are met, which depends, apart from the choice of Δt, i.e. the trading frequency, mainly on the matrix $A = B^T(\sigma\sigma^T)^{-1}B$. To illustrate this further, we will now observe how changes in this matrix influence the filter. More influences of such a change besides on the filter will be discussed in Section 4.3.

To simulate such a change we first increased the values in the volatility matrix and simulated the scheme for the unnormalized filter 1000 times with the usual simulation parameters and (MP1) except for the volatility. In Figure 4.2 we can observe a decreasing mean of the sum of the entries of the unnormalized filter at terminal time at a decreasing biggest eigenvalue of the matrix A. With the parameters used, the maximal eigenvalue ist approximately the same value as the trace of the matrix: one large eigenvalue is again paired with two negligibly small ones.

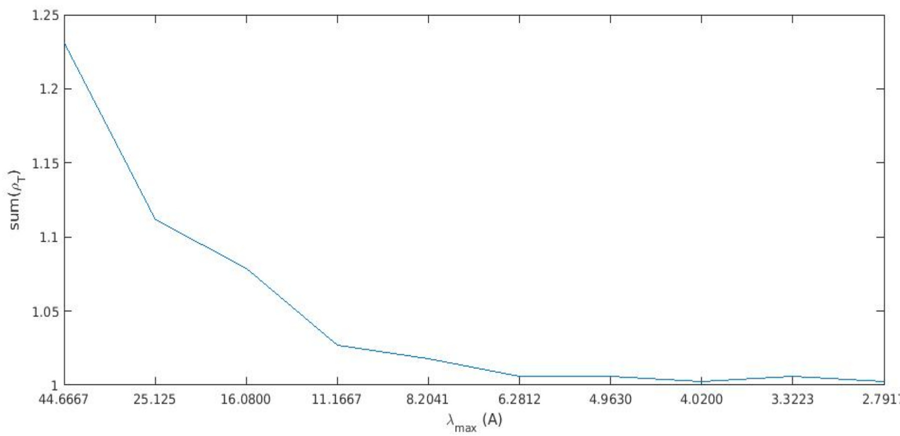

Figure 4.2: Sum of the entries of the unnormalized filter at terminal time with decreasing maximal eigenvalues of the matrix A caused by manipulation of σ

We can observe that with a decrease in the maximal eigenvalue of the matrix the sum of the entries of the unnormalized filter decreases as well, thus leading to more volatile results.
This effect occurs as well when manipulating the values of the state matrix instead of the volatility matrix.

4.2 Advantage in the Gaussian Model

In the Gaussian model positivity is of no importance, both the drift and hence the filter are very likely negative by definition at some point in time. Boundedness is still to be considered though and will be studied next.

Depending on the parameters we can observe that already the expected value of the drift itself might not be bounded.

4.9 Lemma. *The drift in this model is given by*

$$d\mu_t = \alpha(\delta - \mu_t)\ dt + \beta\ dW_t^2, \quad \mu_0 \sim \mathcal{N}(m_0, \gamma_0) \quad .$$

Then $h_t = \mathbb{E}[\mu_t]$ *satisfies*

$$dh_t = \alpha(\delta - h_t)dt, \quad h_0 \sim \mathcal{N}(m_0, \gamma_0)$$

with $t \in [0, T]$.

Proof. The last summand in μ_t is of mean zero after applying Fubini's Theorem. □

This process has to be bounded for the expected value of the drift to be bounded by a given bound at time t. We get the following condition on the boundedness of the process h_t.

4.10 Lemma. *The process* h_t *is bounded for all* $t \in [0, T],\ T \to \infty$ *if the matrix* α *is diagonalizable and positive semidefinite.*

Proof. The vector δ is constant thus the critical term to be controlled is the process H_t given by

$$dH_t = -\alpha H_t dt, \quad h_0 \sim \mathcal{N}(m_0, \gamma_0) \quad .$$

Its solution can be determined with help of the matrix exponential

$$H_t = \exp_m(-\alpha t)\, h_0 \quad .$$

This matrix exponential is bounded for diagonalizable, positive semidefinite matrices α. □

Kalman Filtering Now that the drift is bounded, we aim to find a similar condition for the filter. The filter is given by

$$m_t = \phi(t)\Big[m_0 + \int_0^t \phi^{-1}(s)\gamma_s(\sigma\sigma^T)^{-1}\ dR_s + \int_0^t \phi^{-1}(s)\ ds\ \alpha\delta\Big]$$

and

$$\phi'(t) = [-\alpha - \gamma(t)(\sigma\sigma^T)^{-1}]\phi(t)$$

with the initial condition that $\phi(0)$ is the identity matrix and

$$d\gamma_t = -\gamma_t(\sigma\sigma^T)^{-1}\gamma_t - \alpha\gamma_t - \gamma_t\alpha^T + \beta\beta^T$$

with initial condition (m_0, γ_0).
It is crucial for the boundedness of the filter that the process γ_t is bounded. Extensive work has been done in the study of this Ricatti equation and the convergence of its solutions: [16] proves when such a limit exists and shows of what form it is.

4.11 Lemma. *Let γ_t converge to a constant γ_∞. Then the filter is bounded if the matrix α is diagonalizable and positive semidefinite, as well as the matrix $\alpha + \gamma_t(\sigma\sigma^T)^{-1}$.*

Proof. Discussing boundedness of the filter is a discussion of the boundedness of the process ϕ. That will be bounded by the same argument as in Lemma 4.10 if $\alpha + \gamma_t(\sigma\sigma^T)^{-1}$ is positive semidefinite and γ_∞ exists. □

4.12 Remark. *The condition for the filter is not equivalent to the one for the drift. We can therefore encounter situations where the filter is bounded but the drift is not (example 4.13).*

4.13 Example. *Let α be 0.7 on the diagonal and -0.3 elsewhere for all numbers of stocks. If we consider three stocks, both the filter and the drift will be bounded, but with the fourth, the matrix α has eigenvalues -0.2 and 1 and thus the drift is not bounded. Using the volatility from (GP1), the eigenvalues of $-\alpha - \gamma_t(\sigma\sigma^T)^{-1}$ are -0.2 and -6.4214 which fulfills the required conditions and the filter is bounded. This changes then with the fifth stock: both filter and drift are not bounded.*

4.3 Portfolio Optimization with Filtering in the HMM

The matrix $A = B^T(\sigma\sigma^T)^{-1}B$ is present in considerations on the boundedness of the scheme for the unnormalized filter as we just saw and as well in computations on the expected utility of the terminal wealth:
In Proposition 4.6 we saw its close relation to the growth rate of the mean of the process ψ. It is thus part of the scheme for the unnormalized filter. It also appears in the optimal terminal wealth that was introduced in Theorem 3.12 as

$$\begin{aligned} X_t^* &= x_0 + \int_0^t \pi_s^{*T} \, d\tilde{R}_s \\ &= x_0 + x_0 \int_0^t (B\rho_s)^T(\sigma\sigma^T)^{-1}BY_s \, ds + x_0 \int_0^t (B\rho_s)^T(\sigma\sigma^T)^{-1}\sigma \, d\tilde{W}_s \\ &= x_0 + x_0 \int_0^t \rho_s^T A Y_s \, ds + x_0 \int_0^t \rho_s^T (A^{\frac{1}{2}})^T \, d\tilde{W}_s \end{aligned}$$

with $A = A^{\frac{1}{2}}(A^{\frac{1}{2}})^T, \; A^{\frac{1}{2}} = (\sigma^{-1}B)^T$ and logarithmic utility.
The entries of this matrix will thus have effects in both considerations: optimization and filtering. We saw that the scheme for the unnormalized filter is closer to unboundedness the bigger the eigenvalues of the matrix A are. To ensure an advantage due to filtering, the investor might therefore be interested in controlling these eigenvalues towards low values. The next section will show, that to maximize the expected value of the terminal wealth, his interest might just be the opposite.

To gain some first intuition, we repeat the simulations from Figure 4.2, but this time with only 100 runs and with respect to observing the optimal terminal wealth. Figure 4.2 was due to the higher number of runs smoother than the following figures. Theoretical results regarding the findings will be presented subsequently.
Figure 4.3 displays the expected utility of the terminal wealth under logarithmic utility for a decrease in the maximal eigenvalue of the matrix A caused by manipulation of the volatility. This utility decreases with the decrease in the eigenvalue which leads to the conclusion that this situation should be avoided in optimization of the expected utility. At the same time, Figure 4.4 shows that the distance of the normalized filter and the stationary distribution also decreases over the same decrease in the maximal eigenvalue leading

to more stable results. This distance is given by $\mathbb{E}[(\eta_t - \nu)^T(\eta_t - \nu)]$, we will denote this by mean squared distance (MSD).

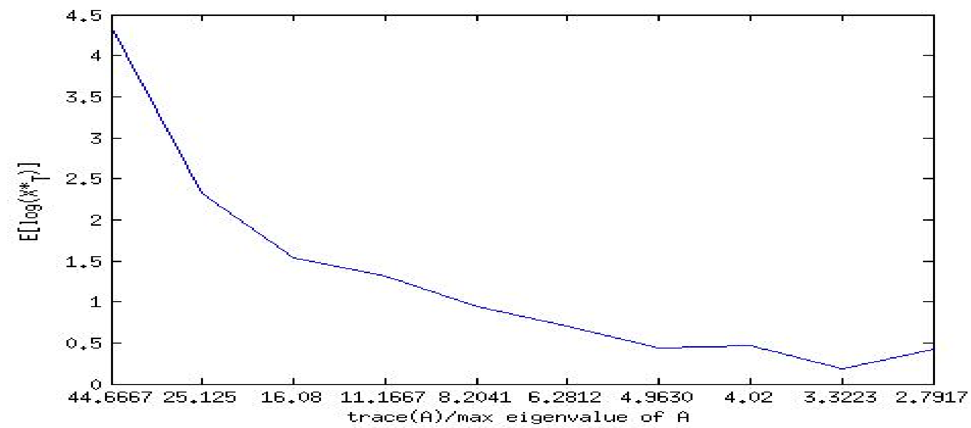

Figure 4.3: Expected utility at decreasing max eigenvalue of A caused by manipulating σ

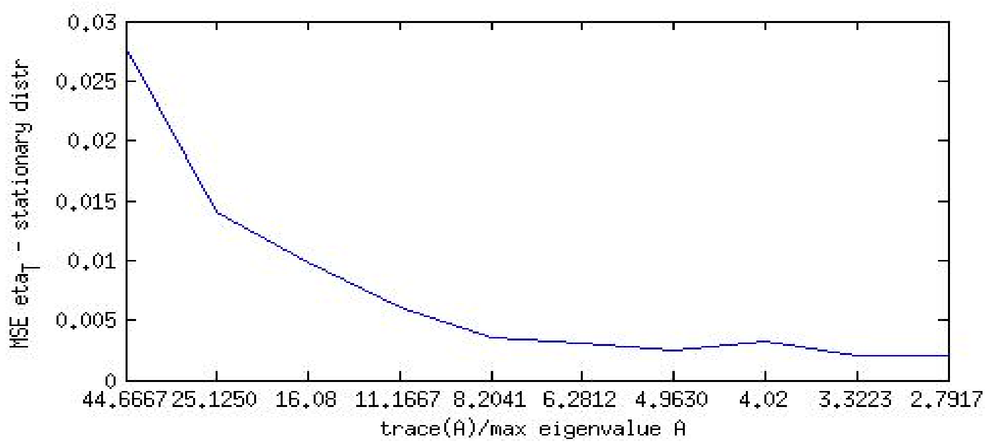

Figure 4.4: MSD of the normalized filter and the stationary distribution at decreasing max eigenvalue of A caused by manipulating σ

We also simulated the similar experiment where the matrix B is changed to decrease the maximal eigenvalue of A which again lead to similar results in

Figures 4.5 and 4.6.

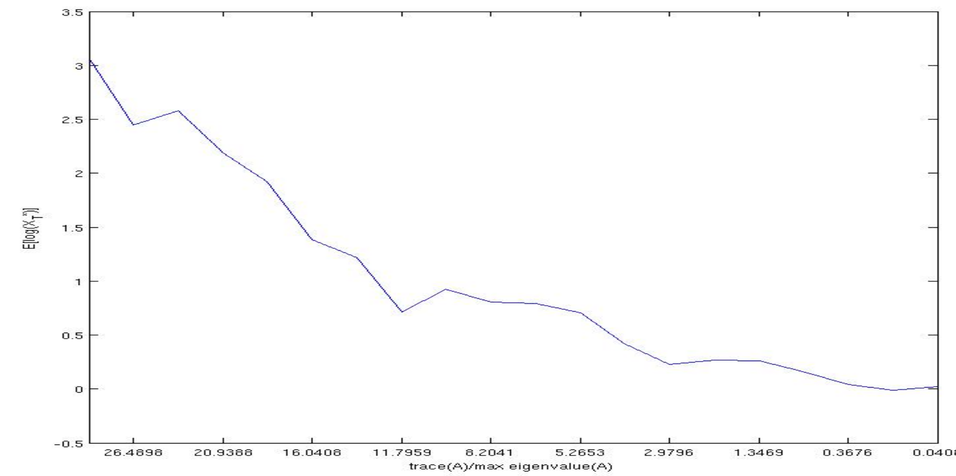

Figure 4.5: Expected utility at decreasing max eigenvalue of A caused by manipulating B

The above figures all indicate that the matrix A influences both filtering and portfolio optimization. The influence on filtering was already quantified previously, as for example in Lemma 4.6. Next we will also quantify its connection to portfolio optimization.

4.14 Proposition. *A lower bound for the expected logarithmic utility of the optimal terminal wealth is given by*

$$\mathbb{E}[\log X_T^*] \geq \frac{T}{2}\left(\sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{max}(A)\right) + \frac{\lambda_{max}(A)}{2}\int_0^T \mathbb{E}[\eta_t^T \eta_t]\ dt \quad .$$

Proof. The optimal trading strategy for logarithmic utility in the partial information case is given by

$$\pi_t^* = (\sigma\sigma^T)^{-1} B\eta_t$$

and the wealth with initial wealth $x_0 = 1$ is then

$$X_T^* = \exp\left(\int_0^T \pi_t^* BY_t - \frac{1}{2}\pi_t^*(\sigma\sigma^T)\pi_t^*\ ds + \int_0^T \pi_t^* \sigma dW_t\right) \quad .$$

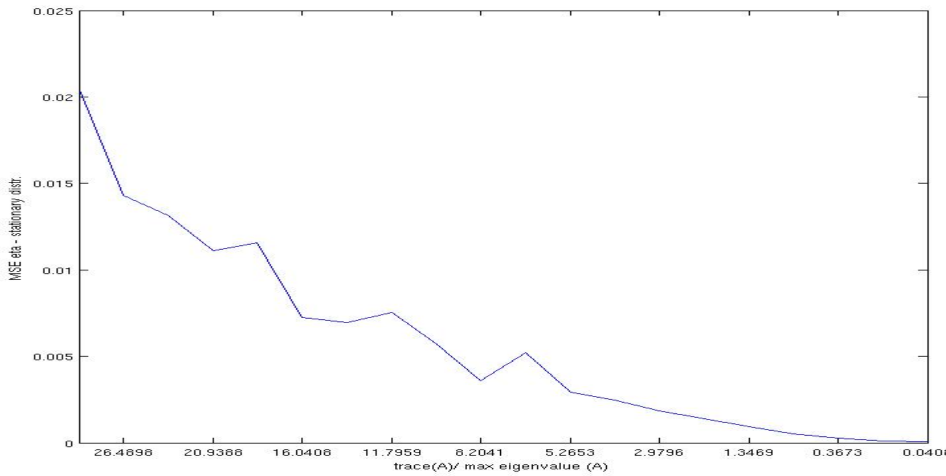

Figure 4.6: MSD of the normalized filter and the stationary distribution at decreasing max eigenvalue of A caused by manipulating B

The expected utility of the optimal terminal wealth will then be

$$\mathbb{E}\log(X_T^*) = \frac{1}{2}\mathbb{E}\int_0^T \eta_t^T B^T(\sigma\sigma^T)^{-1}B\eta_t \, dt \tag{4.1}$$

$$= \frac{1}{2}\int_0^T \mathbb{E}\Big[\eta_t^T A\eta_t\Big] \, dt \quad . \tag{4.2}$$

Of expected value $\mathbb{E}\Big[\eta_t^T A\eta_t\Big]$ we know

$$\begin{aligned} \mathbb{E}\left[(Y_t-\eta_t)^T A(Y_t-\eta_t)\right] &= \mathbb{E}\left[Y_t^T A Y_t\right] - \mathbb{E}\left[\eta_t^T A\eta_t\right] \\ \Leftrightarrow \mathbb{E}\left[\eta_t^T A\eta_t\right] &= \mathbb{E}\left[Y_t^T A Y_t\right] - \mathbb{E}\left[(Y_t-\eta_t)^T A(Y_t-\eta_t)\right] \quad . \end{aligned}$$

The Markov chain operates on the unit vectors and its expected value is ν. With that we get

$$\mathbb{E}\left[Y_t^T A Y_t\right] = \sum_{i=1}^d \nu_i A_{ii} \quad . \tag{4.3}$$

Now we can determine a lower bound for the optimal terminal wealth via the matrix A by first approximating

$$\mathbb{E}\left[(Y_t-\eta_t)^T A(Y_t-\eta_t)\right] \le \lambda_{max}(A)\mathbb{E}\left[(Y_t-\eta_t)^T(Y_t-\eta_t)\right] \quad ,$$

where $\lambda_{max}(A)$ is the biggest eigenvalue of A. It is also known that

$$\mathbb{E}\left[(Y_t - \eta_t)^T (Y_t - \eta_t)\right] = \mathbb{E}[Y_t^T Y_t] - \mathbb{E}[\eta_t^T \eta_t] = 1 - \mathbb{E}[\eta_t^T \eta_t]$$

and the desired bound is then given by

$$\begin{aligned}
\frac{1}{2}\int_0^T \mathbb{E}\Big[\eta_t^T A \eta_t\Big]\ dt &\geq \frac{1}{2}\int_0^T \sum_{i=1}^d \nu_i A_{ii} - \lambda_{max}(A)(1 - \mathbb{E}[\eta_t^T \eta_t])\ dt \\
&= \frac{T}{2}\left(\sum_{i=1}^d \nu_i A_{ii} - \lambda_{max}(A)\right) + \frac{\lambda_{max}(A)}{2}\int_0^T \mathbb{E}[\eta_t^T \eta_t]\ dt \quad .
\end{aligned}$$

□

4.15 Remark. *Equality is given for equal eigenvalues of A. This is not a very natural model though and will be discussed later.*

This bound does not necessarily lead to valuable information, it might very well be smaller than the value reached by the investor who has no information and does then not contribute any knowledge. It does provide us with information if its value at terminal time is bigger than what we compute without having any information at all, i.e. using the stationary distribution instead of the filter:

$$\frac{1}{2}\int_0^T \mathbb{E}\Big[\nu^T A \nu\Big]\ dt = \frac{1}{2} T \nu^T A \nu \quad .$$

4.16 Lemma. *The bound is bigger than the value reached by the investor without information if*

$$\frac{-\sum_{i=1}^d \nu_i A_{ii} - \nu^T A \nu}{\lambda_{max}(A)} + 1 \ < \ \mathbb{E}[\eta_T^T \eta_T] \quad .$$

Proof. For the bound to grow beyond the value that can be reached without information, it has to increase more rapidly with time than $\frac{T}{2}\nu^T A \nu$ since its starting value is zero. Therefore to determine whether there is a terminal time T for which the bound provides information, we study the derivative of the bound and determine for which parameters it is positive.
Let therefore

$$f(T) = \frac{T}{2}\left(\sum_{i=1}^d \nu_i A_{ii} - \lambda_{max}(A)\right) + \frac{\lambda_{max}(A)}{2}\int_0^T \mathbb{E}[\eta_t^T \eta_t]\ dt$$

be the function to be studied with

$$f'(T) = \frac{1}{2}\left(\sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{max}(A)(1 - \mathbb{E}[\eta_T^T \eta_T])\right) \quad .$$

This is bigger than $\frac{1}{2}\nu^T A \nu$ if

$$\begin{aligned} \frac{1}{2}\left(\sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{max}(A)(1 - \mathbb{E}[\eta_T^T \eta_T])\right) &> \frac{1}{2}\nu^T A \nu \\ \Leftrightarrow \quad \frac{-\sum_{i=1}^{d} \nu_i A_{ii} - \nu^T A \nu}{\lambda_{max}(A)} + 1 &< \mathbb{E}[\eta_T^T \eta_T] \quad . \end{aligned}$$

Hence if this is given, then the bound will contribute valuable information for some T from which onwards the bound is higher than the value computed without information. □

Next we will have a look at factors influencing the quality of the bound. The steeper it increases, the higher is its value and thus the higher is the quality. Therefore we aim to maximize $\frac{-\sum_{i=1}^{d} \nu_i A_{ii} - \nu^T A \nu}{\lambda_{max}(A)} + 1$ which is equivalent to minimizing $\frac{\sum_{i=1}^{d} \nu_i A_{ii} + \nu^T A \nu}{\lambda_{max}(A)}$. Hence, the bound gains quality for bigger maximal eigenvalues and also for smaller sums $\sum_{i=1}^{d} \nu_i A_{ii} + \nu^T A \nu$.

4.17 Remark. *All eigenvalues are nonnegative since the matrix A is positive semidefinite. In our usual model with two stocks and three states, at least one eigenvalue will always be zero since the state matrix has a nontrivial kernel in this case. If we now use the structure from previous examples where all stocks are represented by the same states, then there is only one positive eigenvalue. In general, stocks will of course not be represented by exactly the same states, but they will typically be similar which results in one large eigenvalue and several very small ones. This is due to the Implicit Function Theorem, from which follows that the coefficients of a polynomial are mapped continuously to its zeros as long as that zero is not an extremum.*
Eigenvalues of value zero can only occur if either we have less stocks than states, which is often used in theoretical discussions but not very realistic for multi-asset portfolios, or if there are stocks with linearly dependent paramters.

There exists another representation of the term we used in the proof of Proposition 4.14 with help of the diagonalization of the matrix $A = SDS^T$:

$$\mathbb{E}\left[(Y_t - \eta_t)^T A (Y_t - \eta_t)\right] = \mathbb{E}\left[\sum_{i=1}^{d} \lambda_i (S^T (Y_t - \eta_t))_i^2\right] \quad .$$

If we only regard the maximal eigenvalue and omit the terms belonging to others, then that is smaller than the above term, thus leading to another bound.

4.18 Proposition. *A lower bound for the expected log-utility of the optimal terminal wealth is given by*

$$\mathbb{E}[\log(X_T^*)] \geq \frac{1}{2} \int_0^T \sum_{i=1}^{d} \nu_i A_{ii} - \mathbb{E}[\lambda_{max} (S_{max}^T (Y_t - \eta_t))^2] \; dt$$

where λ_{max} *is the maximal eigenvalue of the matrix* A *and the diagonalization of* A *is given by* SDS^T *with* S_{max}^T *being the eigenvector corresponding to the maximal eigenvalue.*

Proof.

$$\begin{aligned}\mathbb{E}[\log(X_T^*)] &= \frac{1}{2} \int_0^T \sum_{i=0}^{d} \nu_i A_{ii} - \mathbb{E}[(Y_t - \eta_t)^T A (Y_t - \eta_t)] \; dt \\ &\geq \frac{1}{2} \int_0^T \sum_{i=0}^{d} \nu_i A_{ii} - \mathbb{E}[\lambda_{max} (S_{max}^T (Y_t - \eta_t))_i^2] \; dt \quad .\end{aligned}$$

□

In our usual models with all eigenvalues equal to zero or close to zero except for one large eigenvalue, these small eigenvalues will then be omitted in the computation leading to an approximation of a simpler nature.
In the case where there is only one positive eigenvalue, say in state k, we can compute the optimal terminal wealth using the diagonalization of the matrix A:

$$\begin{aligned}&\mathbb{E}\left[(Y_t - \eta_t)^T A (Y_t - \eta_t)\right] = \mathbb{E}\left[(Y_t - \eta_t)^T SDS^T (Y_t - \eta_t)\right] \\ &= \mathbb{E}\left[(S^T (Y_t - \eta_t))_k^2 \lambda_k(A)\right]\end{aligned}$$

and

$$\mathbb{E}[\log(X_T^*)] = \frac{1}{2}\int_0^T \sum_{i=1}^d \nu_i A_{ii} - \mathbb{E}\left[(S^T(Y_t - \eta_t))_k^2 \lambda_k(A)\right] \; dt \quad .$$

This computation includes still the possibly added error from omitted eigenvalues in case there are one large and otherwise very small eigenvalues. Then we do not have an equivalent representation but rather another approximation. In addition to that, we face a discretization error from the filter with this computation. The following examples illustrate the difference between the approximations.

In these examples we used a rate matrix with very small entries. This means that there occur few jumps which leads to a smaller discretization error in the expected value of the normalized filter. Otherwise this error may be very large. The rate matrix and its stationary distribution that will be used in the following examples are given by

$$Q = \begin{pmatrix} -5 & 3 & 2 \\ 1 & -2 & 1 \\ 4 & 1 & -5 \end{pmatrix} \quad , \quad \nu = \begin{pmatrix} 0.2727 & 0.5152 & 0.2121 \end{pmatrix} \quad .$$

In the figures we will use the notation bound 1 for the display of the bound from Proposition 4.14 and bound 2 corresponds to the bound in Proposition 4.18.

4.19 Example. *We will first examine a case with 2 stocks and 3 states. To model the existence of only one eigenvalue of the matrix A, the state and volatility matrices are given by*

$$B = \begin{pmatrix} 3 & 0.1 & -3 \\ 3 & 0.1 & -3 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{pmatrix} \quad .$$

The maximal eigenvalue of A is then 400.2222, both other eigenvalues are zero. The following Figure 4.7 shows in blue the expected utility of the optimal terminal wealth, in red the version for only one eigenvalue with the before mentioned second approximation. These values are supposed to be equal since there is only one positive eigenvalue, but due to simulation there appear small deviations. The bound which contains information in this case from Proposition 4.14 is in black and the wealth we reach without having any *information, i.e. using the stationary distribution instead of the filter, is displayed in green.*

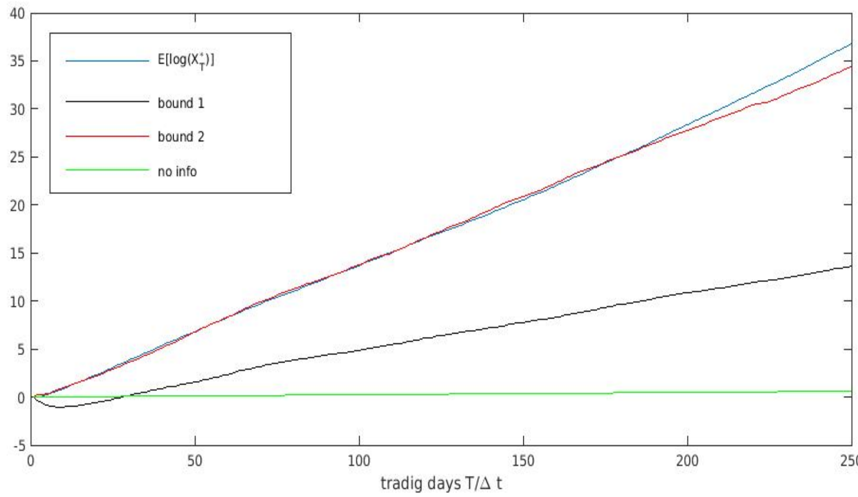

Figure 4.7: Expected utility of the optimal terminal wealth, bounds and terminal wealth without information

To improve the bound in black from Proposition 4.14 we will now have to choose parameters that lead to a larger maximal eigenvalue, i.e. larger entries in the state matrix and/or smaller ones in the volatility. We chose the latter with

$$\sigma = \begin{pmatrix} 0.15 & 0.08 \\ 0.08 & 0.15 \end{pmatrix}$$

leading to an eigenvalue 680.9074 *and the results in Figure 4.8 which are displayed in the same fashion as before.*

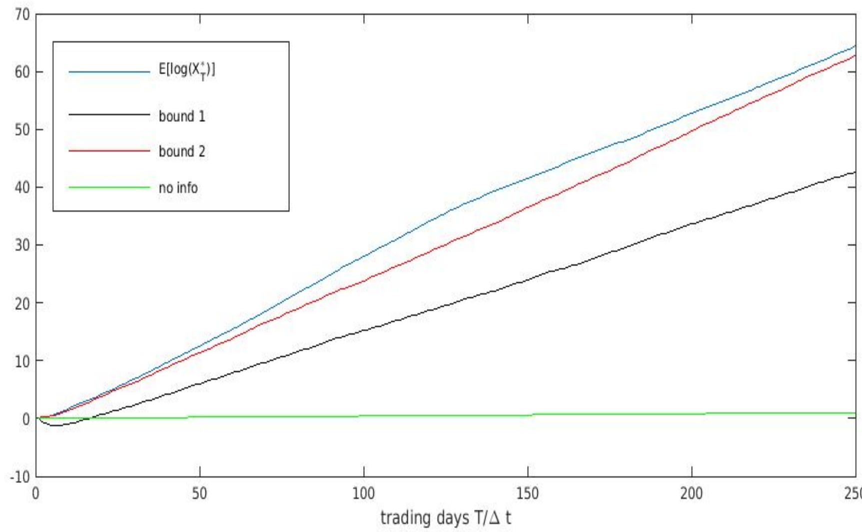

Figure 4.8: Optimal terminal wealth, bounds and expected utility of the terminal wealth without information

4.20 Example. *Next we will have a look at a model with three stocks and three states with the parameters*

$$B = \begin{pmatrix} 2.8 & -0.1 & -2.8 \\ 0.8 & 0.08 & -1.2 \\ 1.7 & -0.1 & -1 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} 0.2 & 0.12 & 0.1 \\ 0.11 & 0.18 & 0.09 \\ 0.11 & 0.09 & 0.21 \end{pmatrix} \quad .$$

With these parameters the matrix A has the eigenvalues 623.9941*,* 27.2268 *and* 0.148*, so we have one large eigenvalue paired with two comparably small ones.*
In Figure 4.9 we can then observe that the bound 1 is still containing information, but the red line for the model with only one eigenvalue has more distance to the expected utility of the optimal terminal wealth. In addition to the simulation deviation, we omitted the two albeit small eigenvalues that are not zero.

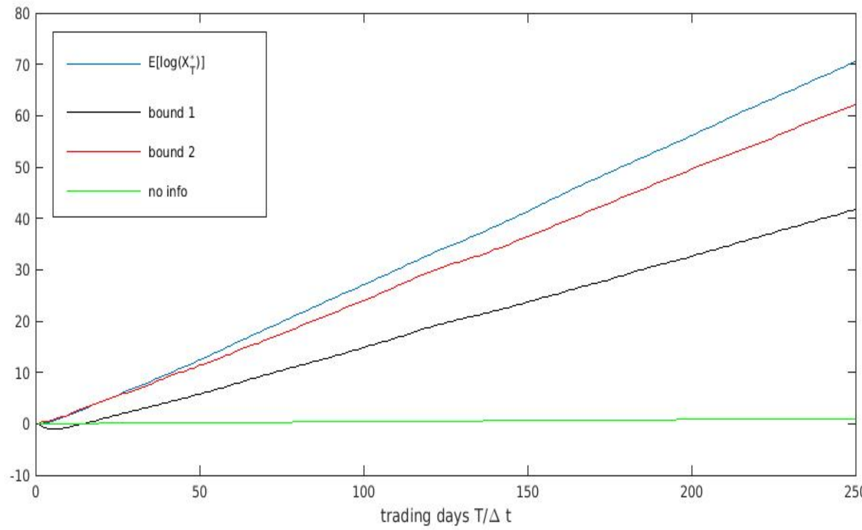

Figure 4.9: Expected utility of the optimal terminal wealth, bounds and expected utility of the terminal wealth without information

4.21 Example. *As mentioned before, if we have three equal eigenvalues, then we have equality for the bound 1 and it is equal to the expected utility of the optimal terminal wealth. We will now show an example for this with figure 4.10, although the parameters used are not realistic.*

$$B = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2.8 & 0 \\ 0 & 0 & 3.2 \end{pmatrix} \quad , \quad \sigma = \begin{pmatrix} 0.2 & 0 & 0 \\ 0 & 0.18 & 0 \\ 0 & 0 & 0.21 \end{pmatrix} \quad .$$

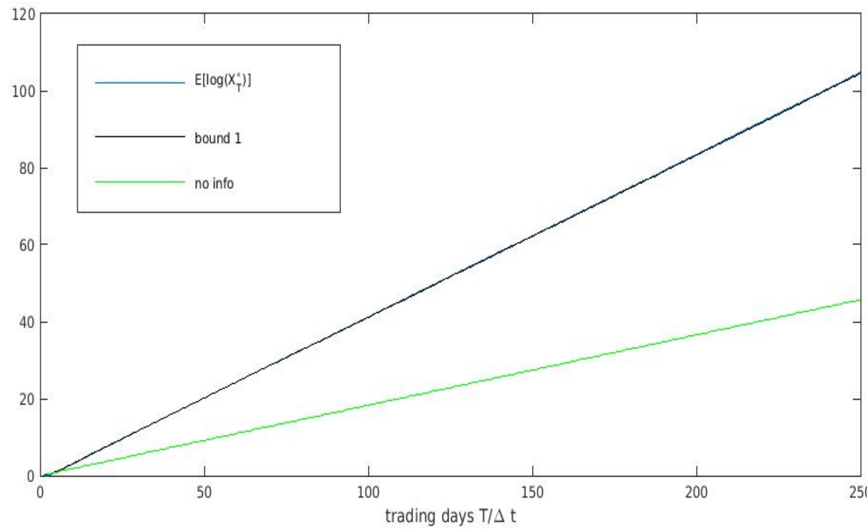

Figure 4.10: Expected utility of the optimal terminal wealth, bound and expected utility of the terminal wealth without information

The two approximations have shown a different behaviour depending on the size of the eigenvalues: If we have one large eigenvalue accompanied by several small ones, the second approximation from Proposition 4.18 leads to better results. The first version from Proposition 4.14 though is at its best if the eigenvalues are all equal.

4.22 Remark. *For an increase in the number of stocks, we discussed earlier, that the matrix A should have a decreasing maximal eigenvalue for the filter to remain stable. This goes hand in hand with a lower expected utility of the optimal terminal wealth and we can now add that the bound 1 is also of less quality the more stocks are involved if this decrease over an increase in the number of stocks is given.*

If we now performed the same computations as in Proposition 4.14, but with the minimal eigenvalue of A instead of the maximal, we would get an upper bound for the expected utility of the optimal terminal wealth accordingly.

4.23 Lemma. *An upper bound for the expected utility of the optimal terminal* *wealth is given by*

$$\mathbb{E}[\log X_T^*] \leq \frac{1}{2}\int_0^T \sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{min}(A)(1 - \mathbb{E}[\eta_t^T \eta_t])\ dt \quad .$$

Proof. This can be computed much the same way as was the bound in Proposition 4.14. □

In this case the two bounds often coincide since the minimal eigenvalue is often equal or close to zero and thus only the term $\nu_i A_{ii}$ remains to be integrated.

4.24 Example. *The following Figure 4.11 was computed with the same parameters we used in the Example 4.20 above, just with the minimal eigenvalue instead of the maximal. With these parameters, the smallest eigenvalue is not equal to zero, but close to, and the bounds are nearly indistinguishable.*

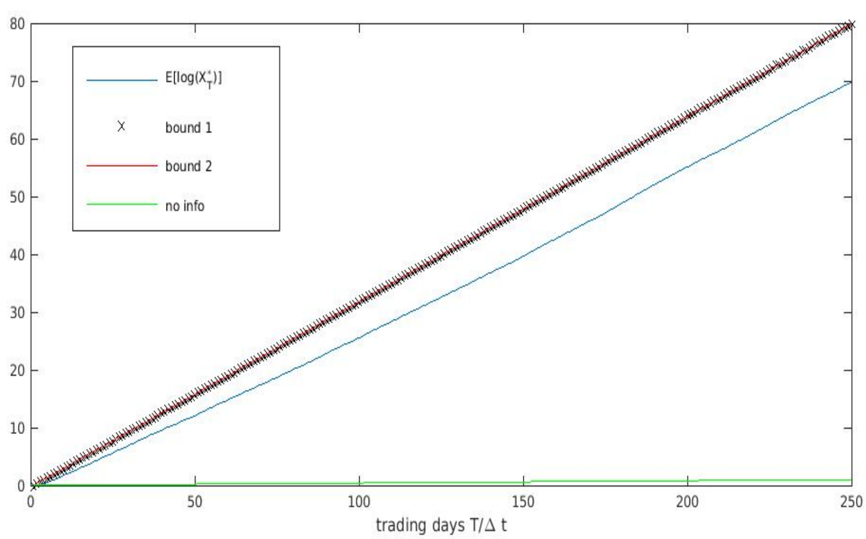

Figure 4.11: Expected utility of the optimal terminal wealth, bound and expected utility of the terminal wealth without information

4.25 Remark. *The lower bound for the optimal terminal wealth was very sensitive towards an increase in the number of stocks since the maximal eigenvalue is smaller with each additional stock. This is not the case for this upper bound, the smallest possible eigenvalue is zero since the matrix A is positive semidefinite and thus does not decrease with an increase in the number of stocks.*

Let us now assume, we have a decrease in the maximal eigenvalue in the matrix A in a model with two stocks. For optimization results, this was not desirable though. Therefore an investor might try to assume a bigger maximal eigenvalue than there actually is. He should expect more risk, but also a possibly higher gain. To model that, we increase the maximal eigenvalue in computations by changing the parameters in the model. With these constructed parameters we compute the matrix A and its eigenvalues as well as the trading strategy based on no information and the one using the filter. We will then study the trading strategy for the first stock and observe the outcome in Figure 4.12. The trading strategy with the increased eigenvalue leads to larger long positions and finally invests even less in the stocks than the investor without information. Not very surprisingly, this is not a solid method, it might lead to better results for small increases, though.

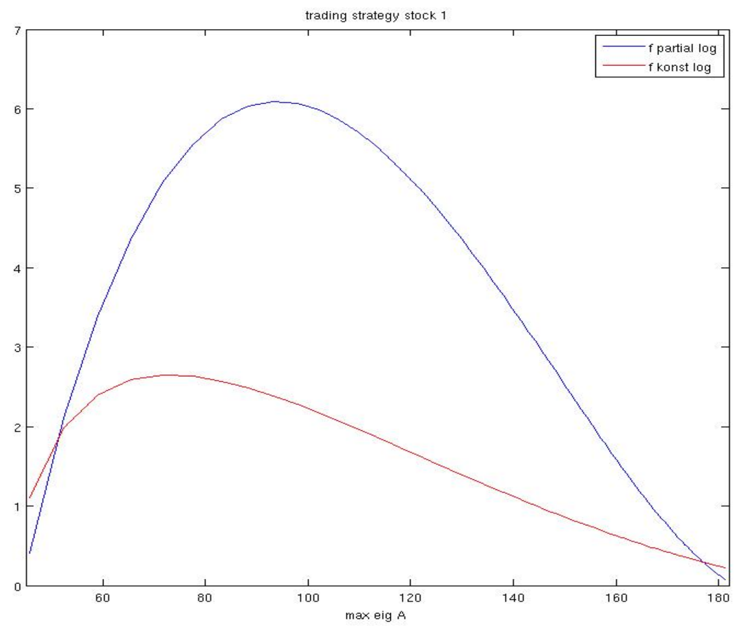

Figure 4.12: Trading strategy for stock 1 with no information and with the filter from the model with increased maximal eigenvalue

4.26 Remark. *The results from this and the previous section allow us to state a more general rule as to when it makes sense to perform portfolio optimization with one's filter at all. First we now know that the scheme for the unnormalized filter is required to be bounded and preserving positivity. Secondly, if the filter is too close to the stationary distribution, it is not worth the effort in comparison to the investor who has no information at all. This can be measured with the lower bound for the optimal terminal wealth from Proposition 4.14 which ensures*

$$\mathbb{E}\left[\log(X_T^*)\right] > \frac{1}{2}T\nu^T A\nu$$

if it contributes information.

If this inequality does not hold true, the investor does not need to apply filtering techniques. Unfortunately, he then faces a bad optimization result, we saw before that the investor without information reaches a much lower expected utility of the terminal wealth. In order to apply filtering to his advantage he might thus add or replace stocks in his portfolio.

4.27 Remark. *In conclusion of the previous thoughts the investor pursuits market parameters where the filter is close to being not bounded, but still bounded, to maximize his terminal wealth and stay as far away from the lower limit given by the investor without information as possible.*

4.4 Market Reduction in the HMM

Next we want to study another aspect of portfolio optimization with special respect to the matrix A. In order to do so let us start with recalling the expected log-utility of the optimal terminal wealth given as

$$\mathbb{E}[\log(X_T^*)] = \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T B^T(\sigma\sigma^T)^{-1}B\eta_t \; dt\Big] = \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T A\eta_t \; dt\Big] \quad .$$

Let us then again diagonalize the matrix A as $A = SDS^T$. The terminal wealth then reads

$$\mathbb{E}[\log(X_T^*)] = \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T SDS^T\eta_t \; dt\Big] \quad .$$

This new version can be interpreted as a lesser dimensional market with some rather nice properties which we will observe and study more closely in the following. Roughly speaking, the matrix S will be the translation for the states and the matrix D for the inverse of the volatility squared.
All quantities in this new market will be denoted by $\;^\circ$. Identifying the parameters by comparing these two versions of the expected utility of the optimal terminal wealth leads to the following representations in the diagonal market.

$$\begin{aligned}
f_t^\circ &= DS^T\eta_t \\
dW_t^\circ &= D^{-\frac{1}{2}}S^TB^T(\sigma^{-1})^T \; dW_t \\
dR_t^\circ &= S^TY_t\,dt + D^{-\frac{1}{2}}dW_t^\circ \\
d\rho_t^\circ &= Q^T\rho_t^\circ \; dt + \operatorname{diag}(\rho_t^\circ)SD \; dR_t^\circ
\end{aligned}$$

The matrix D is unfortunately not invertible in case A has eigenvalues of value zero. For now, we assume therefore that the matrix A is of full rank.

4.28 Lemma. *If* $\operatorname{rank}(A) = d$, W_t° *is a Brownian motion.*

Proof. Computing the quadratic covariance of the new quantity W_t° results in

$$d[W^\circ]_t = \mathrm{Id}_d\,dt$$

and thus we have by Lévy's characterization Theorem a Brownian motion in this new market. □

4.29 Remark. *As you may have noticed, quantities depending on the number of stocks are not of the same dimension as the original ones anymore: where we had n stocks and thus an n-dimensional Brownian motion, we now have a d-dimensional one. This leads to the before mentioned new properties. We are now able to represent an n-dimensional portfolio with only d funds. This is desirable in case there are more stocks than states. These d funds would of course have to act as exactly the combination of the n stocks that leads to the optimal terminal wealth and is described in the diagonalisation of the matrix A. In the portfolio we would now have funds as assets rather than single stocks. The composition of these funds, namely the linear combination of stocks in the fund, is independent of time, the amount invested in them is not.*

The introduction to this section covered logarithmic utility only and can be performed in much the same way for power utility as well: the filter does not need another translation, it does not depend on the utility. The strategy does though and will be discussed in Corollary 4.34.
We will show the equivalence of the original and the diagonal market for arbitrary utility next, but will need a preliminary result for the proof.

4.30 Lemma. *Let* $\operatorname{rank}(A) = d$. *Then the term* $B^T(\sigma\sigma^T)^{-1}\, dR_t$ *is equal to* $SD\, dR_t^\circ$.

Proof.

$$\begin{aligned}
& B^T(\sigma\sigma^T)^{-1}\, dR_t \\
&= B^T(\sigma\sigma^T)^{-1}BY_t\, dt + B^T(\sigma\sigma^T)^{-1}\sigma\, dW_t \\
&= AY_t\, dt + B^T(\sigma\sigma^T)^{-1}\sigma\, dW_t \\
&= SDS^TY_t\, dt + B^T(\sigma\sigma^T)^{-1}\sigma\, dW_t \\
&= SDS^TY_t\, dt + SDD^{-1}S^TB^T(\sigma\sigma^T)^{-1}\sigma\, dW_t \\
&= SD(S^TY_t\, dt + D^{-1}S^TB^T(\sigma\sigma^T)^{-1}\sigma\, dW_t) \\
&= SD(S^TY_t\, dt + D^{-\frac{1}{2}}\, dW_t^\circ) \\
&= SD\, dR_t^\circ \quad .
\end{aligned}$$

□

With this in mind we can now prove the following theorem.

4.31 Theorem. *Let* $\operatorname{rank}(A) = d$. *The original market and the newly introduced diagonal market lead to the same optimal terminal wealth and the filters coincide.*

Proof. To see that the filters are in fact the same, we apply Lemma 4.30 with the result

$$\begin{aligned} d\rho_t &= Q^T \rho_t \, dt + \operatorname{diag}(\rho_t) B^T (\sigma\sigma^T)^{-1} \, dR_t \\ &= Q^T \rho_t \, dt + \operatorname{diag}(\rho_t) SD \, dR_t^\circ \quad . \end{aligned}$$

The SDE for the scheme of the unnormalized filter in the new market reads

$$d\rho_t^\circ = Q^T \rho_t^\circ \, dt + \operatorname{diag}(\rho_t^\circ) SD \, dR_t^\circ \quad .$$

With the same initial value $\rho_0 = \rho_0^\circ = \nu$, the solution of the linear SDE is unique.
Showing the equality of the optimal terminal wealth can be done independently of the utility by using the definition from Theorem 3.12

$$X_T^* = I(y\zeta_T)$$

where y is a constant given by $\tilde{\mathbb{E}}[I(y\zeta_T)] = x_0$. The density ζ can be translated analogously to the filter using Lemma 4.30 leading to

$$\zeta_T^{\circ\,-1} = 1 + \int_0^T \rho_s^{\circ T} SD \, dR_s^\circ \quad .$$

The optimal terminal wealth translates then to

$$X_T^* = I(y\zeta_T) = I(y^\circ \zeta_T^\circ) = X_T^{\circ\,*} \quad .$$

□

Following Theorem 4.31 one may call the original market with return R and the new market with R° equivalent.

For our usual logarithmic and power utility functions we compute the expected utility of the optimal terminal wealth in the diagonal markets explicitly in the following corollary. But first we will translate the necessary Malliavin derivative using the product in which it is going to appear in the proof, namely $B^T {\sigma^T}^{-1} D_t \rho_u$.

4.32 Lemma. *Let* $\operatorname{rank}(A) = d$ *and* $F_u^\circ := D^{-\frac{1}{2}} S^T B^T \sigma^{T^{-1}} D_t \rho_u$. *Then* F_u° *is a translation of the Malliavin derivative of the unnormalized filter into the diagonal market.*

Proof. The Malliavin derivative is given by

$$D_t\rho_u = \sigma^{-1} B \operatorname{diag}(\rho_t) + \int_t^u (D_t\rho_s) Q \; ds + \int_t^u (D_t \rho_s) \operatorname{diag}(B^T(\sigma\sigma^T)^{-1} \; dR_s)$$

and then $B^T\sigma^{T^{-1}} D_t\rho_u$ is with the use of Lemma 4.30

$$\begin{aligned} B^T\sigma^{T^{-1}} D_t\rho_u = & \; A \operatorname{diag}(\rho_t) + \int_t^u (B^T\sigma^{T^{-1}} D_t\rho_s) Q \; ds \\ & + \int_t^u (B^T\sigma^{T^{-1}} D_t\rho_s) \operatorname{diag}(SD \; dR_s^\circ) \quad . \end{aligned}$$

If we now multiply this also with $D^{-\frac{1}{2}} S^T$ we get

$$\begin{aligned} D^{-\frac{1}{2}} S^T B^T \sigma^{T^{-1}} D_t\rho_u = & D^{\frac{1}{2}} S^T \operatorname{diag}(\rho_t) + \int_t^u (D^{-\frac{1}{2}} S^T B^T \sigma^{T^{-1}} D_t\rho_s) Q \; ds \\ & + \int_t^u (D^{-\frac{1}{2}} S^T B^T \sigma^{T^{-1}} D_t\rho_s) \operatorname{diag}(SD \; dR_s^\circ) \quad . \end{aligned}$$

We denote $F_u^\circ := D^{-\frac{1}{2}} S^T B^T \sigma^{T^{-1}} D_t\rho_u$. The starting value of F_u° is now the direct translation of the original Malliavin derivative's starting value into the diagonal market. The dynamics of F_u° are the same as the dynamics of the Malliavin derivative of the unnormalized filter

$$\begin{aligned} dF_u^\circ &= F_u^\circ Q \; du + F_u^\circ \operatorname{diag}(SD \; dR_u^\circ) \\ &= F_u^\circ Q \; du + F_u^\circ \operatorname{diag}(B^T(\sigma\sigma^T)^{-1} \; dR_u) \quad . \end{aligned}$$

□

4.33 Remark. *This Lemma allows us to translate the product* $B^T\sigma^{T^{-1}} D_t\rho_u$ *into* $SD^{\frac{1}{2}} F_u^\circ$ *in the diagonal market.*

4.34 Corollary. *For logarithmic and power utility the expected utility of the optimal terminal wealth in the diagonal market is given by*

$$\mathbb{E}[\log(X_T^{\circ *})] = \mathbb{E}\Big[\frac{1}{2} \int_0^T \eta_t^T S D S^T \eta_t \; dt\Big] \quad ,$$

$$\mathbb{E}\Big[\frac{1}{\kappa}(X_T^{\circ *})^\kappa\Big] = \frac{x_0}{\kappa} + \mathbb{E}\Big[\int_0^T (X_s^*)^{\kappa-1} (\pi_s^{\circ * T} S^T \eta_s^\circ + \frac{\kappa - 1}{2} \pi_s^{\circ * T} D^{-1} \pi_s^{\circ *}) \; ds\Big] \quad .$$

Proof. The expected utility of the optimal terminal wealth for logarithmic utility can be computed straight forward as

$$\begin{aligned}\mathbb{E}[\log(X_T^*)] &= \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T B^T(\sigma\sigma^T)^{-1}B\eta_t\Big]\ dt \\ &= \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T A\eta_t\ dt\Big] = \mathbb{E}\Big[\frac{1}{2}\int_0^T \eta_t^T SDS^T\eta_t\ dt\Big] \\ &= \mathbb{E}[\log(X_T^{\circ *})] \quad .\end{aligned}$$

Let us now turn to power utility. The expected utility in the original market was given by

$$\mathbb{E}\Big[\frac{1}{\kappa}(X_T^*)^\kappa\Big] = \frac{x_0}{\kappa} + \mathbb{E}\Big[\int_0^T (X_s^*)^{\kappa-1}(\pi_s^{*T}B\eta_s + \frac{\kappa-1}{2}\pi_s^{*T}(\sigma\sigma^T)\pi_s^*)\ ds\Big]$$

with

$$\begin{aligned}\pi_t^* =& \frac{x_0(\sigma\sigma^T)^{-1}}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]}\Big\{B\rho_t\tilde{\mathbb{E}}[\zeta_T^{\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S] \\ &+ \tilde{\mathbb{E}}[\zeta_T^{\frac{\kappa}{\kappa-1}}\int_t^T(\sigma D_t\rho_s)B^T(\sigma\sigma^T)^{-1}\ dR_s|\mathcal{F}_t^S]\Big\} \quad .\end{aligned}$$

Inserting this leads with some rearranging to

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{\kappa}(X_T^*)^\kappa\Big] \\
&= \frac{x_0}{\kappa} + \mathbb{E}\Bigg[\int_0^T (X_s^*)^{\kappa-1}\left(\frac{x_0\tilde{\mathbb{E}}[\zeta_T^{\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]}\rho_s^T B^T(\sigma\sigma^T)^{-1}B\eta_s\right. \\
&+\frac{x_0}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]}\eta_s^T B^T(\sigma\sigma^T)^{-1}\tilde{\mathbb{E}}\Big[\zeta_T^{\frac{\kappa}{\kappa-1}}\int_t^T(\sigma D_t\rho_s)B^T(\sigma\sigma^T)^{-1}\; dR_s|\mathcal{F}_t^S\Big] \\
&+\frac{x_0^2(\kappa-1)\tilde{\mathbb{E}}[\zeta_T^{\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]^2}{2(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]^2}\rho_s^T B^T(\sigma\sigma^T)^{-1}B\rho_s \\
&+\frac{x_0^2(\kappa-1)\tilde{\mathbb{E}}[\zeta_T^{\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]}{(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]^2}\rho_s^T B^T(\sigma\sigma^T)^{-1}\tilde{\mathbb{E}}\Big[\zeta_T^{\frac{\kappa}{\kappa-1}}\int_t^T(\sigma D_t\rho_s)B^T(\sigma\sigma^T)^{-1}\; dR_s|\mathcal{F}_t^S\Big] \\
&+\frac{x_0^2(\kappa-1)(\sigma\sigma^T)^{-1}}{2(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\frac{1}{\kappa-1}}]^2}\tilde{\mathbb{E}}\Big[\zeta_T^{\frac{\kappa}{\kappa-1}}\int_t^T(\sigma D_t\rho_s)B^T(\sigma\sigma^T)^{-1}\; dR_s|\mathcal{F}_t^S\Big]^T \\
&\quad\left.\cdot\,\tilde{\mathbb{E}}\Big[\zeta_T^{\frac{\kappa}{\kappa-1}}\int_t^T\big(\sigma D_t\rho_s\big)B^T(\sigma\sigma^T)^{-1}\; dR_s|\mathcal{F}_t^S\Big]\right)\; ds\Bigg] \quad .
\end{aligned}$$

Applying Lemma 4.30 and the fact that the process ζ translates into the diagonal market with the same argument leave us with

$$
\begin{aligned}
&= \frac{x_0}{\kappa} + \mathbb{E}\Bigg[\int_0^T (X_s^*)^{\kappa-1} \Bigg(\frac{x_0 \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} | \mathcal{F}_t^S]}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\circ \frac{1}{\kappa-1}}]} \rho_s^{\circ T} S D S^T \eta_s^\circ \\
&+ \frac{x_0}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\circ \frac{1}{\kappa-1}}]} \eta_s^{\circ T} B^T (\sigma^T)^{-1} \tilde{\mathbb{E}}\Big[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} \int_t^T (D_t \rho_s) S D \; dR_s^\circ | \mathcal{F}_t^S \Big] \\
&+ \frac{x_0^2 (\kappa-1) \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} | \mathcal{F}_t^S]^2}{2(1-\kappa)^2 \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{1}{\kappa-1}}]^2} \rho_s^{\circ T} S D S^T \rho_s^\circ \\
&+ \frac{x_0^2 (\kappa-1) \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} | \mathcal{F}_t^S]}{(1-\kappa)^2 \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{1}{\kappa-1}}]^2} \rho_s^{\circ T} B^T (\sigma^T)^{-1} \tilde{\mathbb{E}}\Big[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} \int_t^T (D_t \rho_s) S D \; dR_s^\circ | \mathcal{F}_t^S \Big] \\
&+ \frac{x_0^2 (\kappa-1)}{2(1-\kappa)^2 \tilde{\mathbb{E}}[\zeta_T^{\circ \frac{1}{\kappa-1}}]^2} \tilde{\mathbb{E}}\Big[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} \int_t^T (D_t \rho_s) S D \; dR_s^\circ | \mathcal{F}_t^S \Big]^T \\
&\cdot \tilde{\mathbb{E}}\Big[\zeta_T^{\circ \frac{\kappa}{\kappa-1}} \int_t^T \big(D_t \rho_s\big) S D \; dR_s^\circ | \mathcal{F}_t^S \Big] \Bigg) \; ds \Bigg] \quad .
\end{aligned}
$$

The translation of the Malliavin derivative from Lemma 4.32 will be applied now and also a direct translation of the Malliavin derivative analogously to the filter leading to

$$
\begin{aligned}
&= \frac{x_0}{\kappa} + \mathbb{E}\Bigg[\int_0^T (X_s^*)^{\kappa-1}\Bigg(\frac{x_0\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{1}{\kappa-1}}]}\rho_s^{\circ T} SDS^T\eta_s^\circ \\
&+ \frac{x_0}{(1-\kappa)\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{1}{\kappa-1}}]}\eta_s^{\circ T} SD^{\frac{1}{2}}\tilde{\mathbb{E}}\Big[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}\int_t^T F_s^\circ SD\ dR_s^\circ|\mathcal{F}_t^S\Big] \\
&+ \frac{x_0^2(\kappa-1)\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]^2}{2(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{1}{\kappa-1}}]^2}\rho_s^{\circ T} SDS^T\rho_s^\circ \\
&+ \frac{x_0^2(\kappa-1)\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}|\mathcal{F}_t^S]}{(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{1}{\kappa-1}}]^2}\rho_s^{\circ T} SD^{\frac{1}{2}}\tilde{\mathbb{E}}\Big[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}\int_t^T F_s^\circ SD\ dR_s^\circ|\mathcal{F}_t^S\Big] \\
&+ \frac{x_0^2(\kappa-1)}{2(1-\kappa)^2\tilde{\mathbb{E}}[\zeta_T^{\circ\frac{1}{\kappa-1}}]^2}\tilde{\mathbb{E}}\Big[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}\int_t^T (D_t\rho_s^\circ)SD\ dR_s^\circ|\mathcal{F}_t^S\Big]^T \\
&\quad \cdot \tilde{\mathbb{E}}\Big[\zeta_T^{\circ\frac{\kappa}{\kappa-1}}\int_t^T \big(D_t\rho_s^\circ\big)SD\ dR_s^\circ|\mathcal{F}_t^S\Big]\Bigg)\ ds\Bigg] \\
&= \frac{x_0}{\kappa} + \mathbb{E}\Big[\int_0^T (X_s^*)^{\kappa-1}(\pi_s^{\circ *T}S^T\eta\circ_s + \frac{\kappa-1}{2}\pi_s^{\circ *T}D^{-1}\pi_s^{\circ *})\ ds\Big] \\
&= \mathbb{E}\Big[\frac{1}{\kappa}(X_T^{\circ *})^\kappa\Big] \quad .
\end{aligned}
$$

□

4.35 Remark. *This theorem describes how an investor trades only in the riskless asset and funds of risky assets, a linear combination of the risky assets in the market, and reaches the same expected utility of the optimal terminal wealth. This is a version of the Mutual Fund Theorem for this model. General applicability of the Mutual Fund Theorem was discussed in [25] for all utility functions under some completeness condition. Although in that work, the mutual fund depends on time and there is only one fund of risky assets next to the riskfree bond. Here we have as many funds as states, but their composition does not change with time, only the amount invested in each.*

In more detail, let us assume the Markov chain is in state i, $Y_t = e_i$. Then the investor will invest only in fund i with return $dR_t^\circ = S^T e_i\ dt + D^{-\frac{1}{2}}\ dW_t^\circ$ if the filter is equal to the corresponding eigenvector. If the filter were equal to a unit vector, the investor would invest in a linear combination of the funds:

$$f_t^{\circ *} = DS^T\eta_t = \begin{cases} \lambda_i e_i & \text{if } \eta_t = S^i \\ \sum_{i=1}^d \lambda_i S_k^i e_i & \text{if } \eta_t = e_k. \end{cases} \quad .$$

The case rank(A)<d Let us now turn to the discussion of the case where the matrix A is not of full rank, but has eigenvalues that are zero. That might either be the case if there are less stocks than states or if there are stocks with linearly dependent parameters. Both cases are not very important in application, but especially the one with less stocks than states often appears in the literature.

For that special case another transformation will be performed. Let the diagonalization be as before $A = SDS^T$, then the matrix D will have rows/columns consisting only of zeros. Eliminate these rows and columns and the corresponding eigenvectors in the matrix S. Then the matrix D shrinks then to dimension $p \times p$ where p is the number of eigenvalues different from zero and the rest of S has dimension $d \times p$. Quantities in this new lesser dimensional model will be denoted by ^{p}.

The trading strategy can be represented this way with p entries as in

$$f_t^{p*} = D^p S^{pT} \eta_t \quad .$$

For the Brownian motion we get

$$dW_t^p = (D^p)^{-\frac{1}{2}} S^{pT} B^T (\sigma\sigma^T)^{-1} \sigma \; dW_t$$

which has again the right quadratic variation. The return in p dimensions is then

$$dR_t^p = S^{pT} Y_t \; dt + (D^p)^{-\frac{1}{2}} \; dW_t^p \quad .$$

For the filter, unfortunately, the translation is not as straight forward. We attempt a reduce of the dimension to less than the number of the states, but the filter is of course exactly of that dimension and that will not change. If we were to translate naively without consideration of this, the result would be

$$d\rho_t^p = Q^T \rho_t^p \; dt + \text{diag}(\rho_t^p) S^p D^p \; dR_t^p \quad .$$

This is not equivalent to the filter of the original market though, translating the summand with the return as was done in Lemma 4.30 shows why. It includes the separation of the matrix S into the part belonging to the p positive eigenvalues S^p and the eigenvectors belonging to the eigenvalues that are zero S^{d-p}.

$$\begin{aligned}
&B^T(\sigma\sigma^T)^{-1}\ dR_t \\
=\ &AY_t\ dt + B^T(\sigma^T)^{-1}\ dW_t \\
=\ &AY_t\ dt + SS^TB^T(\sigma^T)^{-1}\ dW_t \\
=\ &S^pD^pS^{p^T}Y_t\ dt + (S^pS^{p^T} + S^{d-p}S^{d-p^T})B^T(\sigma^T)^{-1}\ dW_t \\
=\ &S^pD^pS^{p^T}Y_t\ dt + S^pD^p(D^p)^{-1}S^{p^T}B^T(\sigma^T)^{-1}\ dW_t \\
&+ S^{d-p}S^{d-p^T}B^T(\sigma^T)^{-1}\ dW_t \\
=\ &S^pD^p\ dR_t^p + S^{d-p}S^{d-p^T}B^T(\sigma^T)^{-1}\ dW_t
\end{aligned}$$

The two terms $s1 = S^pS^{p^T}B^T(\sigma^T)^{-1}\ dW_t$ and $s2 = S^{d-p}S^{d-p^T}B^T(\sigma^T)^{-1}\ dW_t$ are independent

$$Cov(s1, s2) = S^pS^{p^T}B^T(\sigma^T)^{-1}t\mathrm{Id}_n\sigma^{-1}BS^{d-p}S^{d-p^T} = 0 \quad .$$

When reducing the dimension to p we therefore have an additional noise term appearing. Without it, the drift and with it the terminal wealth are not optimal. This noise is necessary to be optimal, without it the filter does not move away from the invariant distribution as easily, it may even be reduced to be a mere oscillation around the invariant distribution. In this case, it does not contain information anymore.

4.5 Advantage in the Standard Market Model

The special knowledge available to the investor who filters can be an advantage. In this and the following sections we will discuss this possible advantage and we will distinguish the investor with full information who can observe the Markov chain directly, the investor with partial information who filters and the investor with no information who uses the stationary distribution of the Markov chain as its approximation.

Let us first observe the advantage in the standard market model numerically by comparing the expected utility of the terminal wealth that can be reached with full, partial or no information using (MP2) in Table 4.1. As expected, the investor with full information gains the most, whereas without information there is no gain and the filter improves that position without reaching the full information possibilities.

(MP2)	log utility	power $\kappa = 0.1$	$\kappa = -1$	$\kappa = -10$
full info	1.4	10.2	-0.91	-0.08
partial info	1.06	10.03	-0.99	-0.099
no info	1	10	-1	-0.1

Table 4.1: Filtering Advantage with (MP2)

The investor with no information uses the stationary distribution which has a sum of all entries of one by definition. The density ζ is given at any time-point as the inverse of the sum of the entries of the unnormalized filter and thus equal to one at all times for this investor. The computation of the expected utility of the optimal terminal wealth has only one time dependent parameter, namely ζ. Hence for the investor without information the expected utility of the optimal terminal wealth does not change with time and is given by the values displayed in Table 4.1 above for all times t.
For the Gaussian investor, this case of no information does also only depend on ζ and is identical to the Markovian case.
The difference between partially and fully informed investor is comparable to the Markovian case as well, for numerical results we refer to the Tables 3.5 and 3.6.

4.6 Filtering Advantage at Infrequent Trading

First we will discuss at which trading frequencies this advantage exists. Intuitively, very infrequent trading will obliterate the advantage, especially with the knowledge from the beginning of this chapter on the unboundedness of the scheme for the unnormalized filter for infrequent trading. In the following example the frequencies will be computed where the filter is not bounded and thus certainly no advantage in the HMM based on Lemma 4.8. The same consideration for the Gaussian Model will be presented subsequently.

4.36 Example. *Let us consider (MP1):*

$$\begin{aligned} Q^T &= \begin{pmatrix} -60 & 15 & 35 \\ 25 & -30 & 25 \\ 35 & 15 & -60 \end{pmatrix} \\ &= \begin{pmatrix} -1 & 1 & 3 \\ 0 & -2 & 5 \\ 1 & 1 & 3 \end{pmatrix} \begin{pmatrix} -95 & 0 & 0 \\ 0 & -55 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \\ \frac{5}{22} & -\frac{3}{11} & \frac{5}{22} \\ \frac{1}{11} & \frac{1}{11} & \frac{1}{11} \end{pmatrix} \end{aligned}$$

and therefore need to find the largest solution of the equations

$$|1 - \Delta t 95| < 1 \quad \textit{and} \quad |1 - \Delta t 55| < 1$$

leading to the result $\Delta t < \frac{2}{95} \approx 0.021$*. Figures 4.13 and 4.14 show the norm* $\left\|(\mathrm{Id} + \Delta t Q^T)\right\|^k$ *plotted against* k *at trading frequencies smaller and bigger than 0.021.*

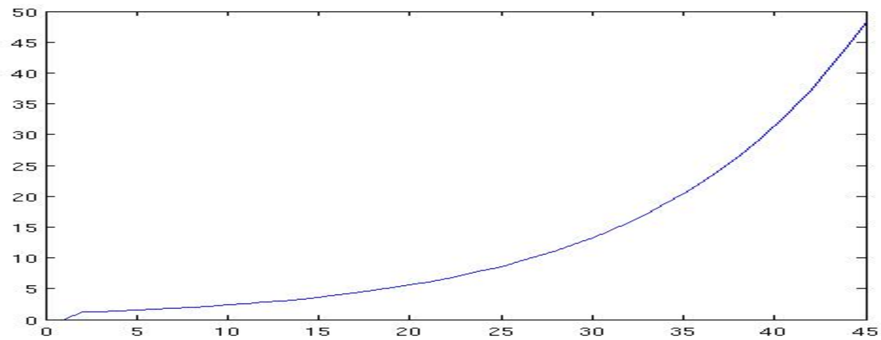

Figure 4.13: Behaviour of the norm $\left\|(\mathrm{Id} + \Delta t Q^T)\right\|^k$ for $\Delta t = 0.022$

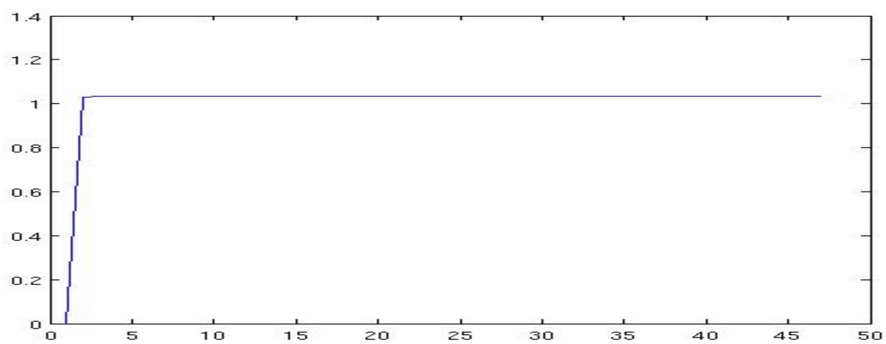

Figure 4.14: Behaviour of the norm $\left\|(\mathrm{Id} + \Delta t Q^T)\right\|^k$ for $\Delta t = 0.021$

The following Figures display the unnormalized filter, again at frequencies smaller and bigger than 0.021, Figures 4.15 and 4.16.

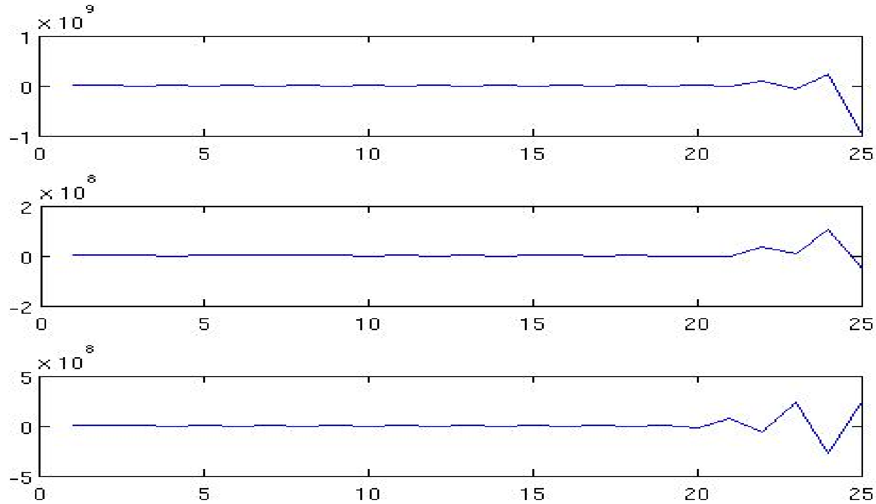

Figure 4.15: Unnormalized filter for 25 trades per year

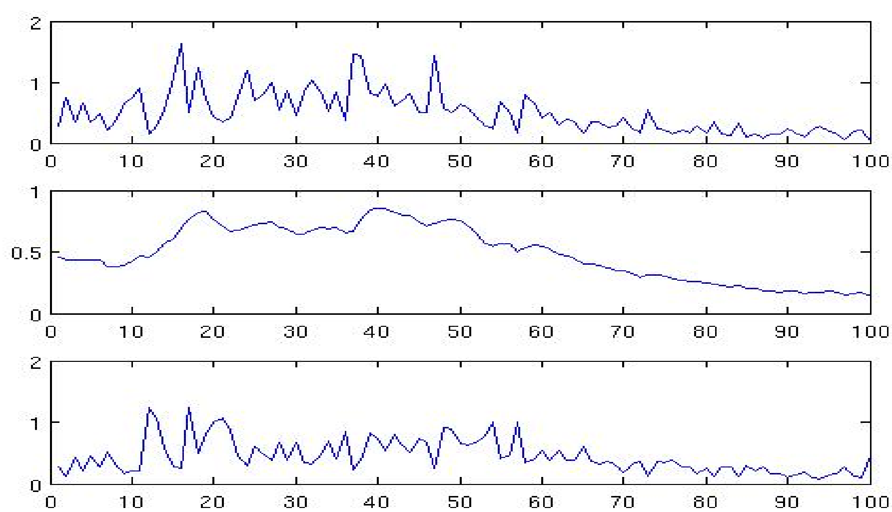

Figure 4.16: Unnormalized filter for 100 trades per year

Filter quality is also essential for the quality of parameter estimation if needed. The EM algorithm we use to estimate parameters depends among other similar filters on the one for the Markov chain. Therefore we expect the estimation to improve with a growing number of tradings. In Figures 4.17, 4.18 and 4.19 we can especially observe the significance of the earlier computed frequency from which onward the scheme for the unnormalized filter is bounded and preserving positivity for complete estimations at different trading frequencies.

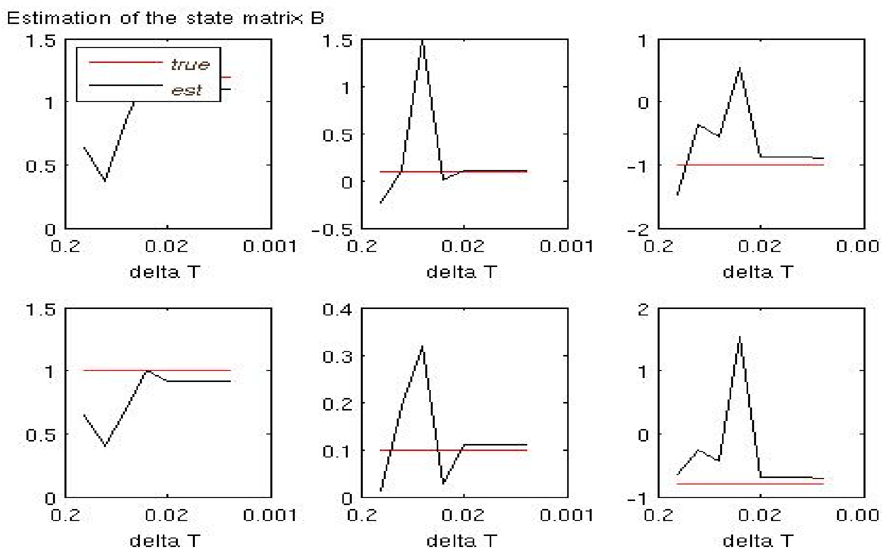

Figure 4.17: Estimation of the state matrix

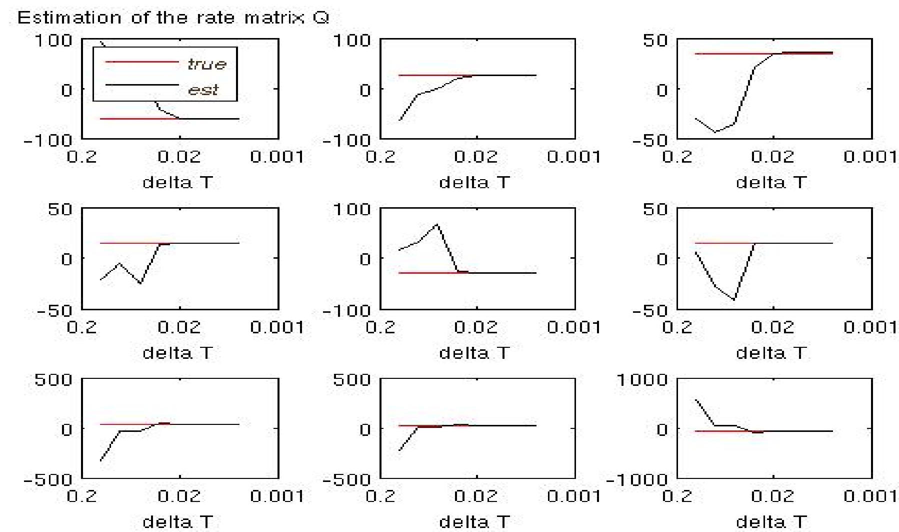

Figure 4.18: Estimation of the rate matrix

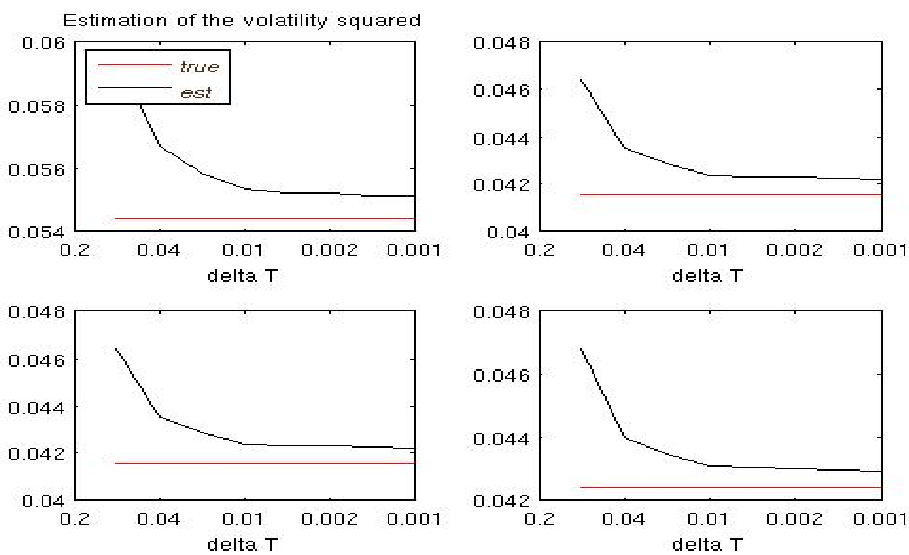

Figure 4.19: Estimation of the volatility matrix

In the Gaussian Model a similar behaviour of the filter occurs and again the Δt where the filter is bounded for all smaller stepsizes can be computed.

Figures 4.20 and 4.21 illustrate the filter at higher and lower frequencies. In this model the filter increases very rapidly for the bigger Δt, so much, that the simulation aborts after a short time.

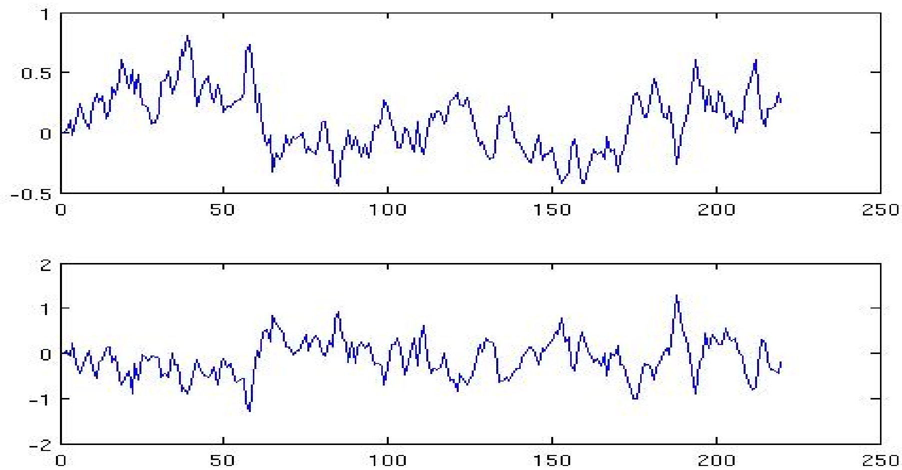

Figure 4.20: Filter for (GP1) at $\Delta t = 0.05$

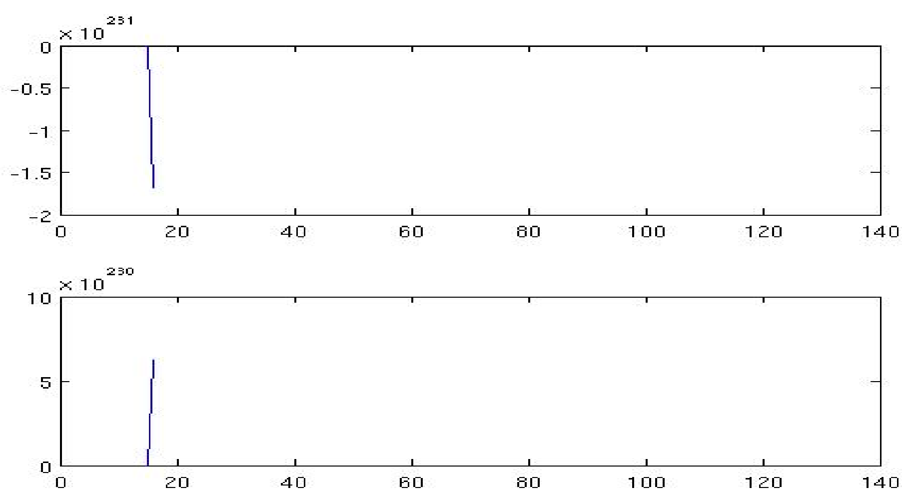

Figure 4.21: Filter for (GP1) at $\Delta t = 0.08$

4.37 Example. *For (GP1) we use*

$$\sigma = \begin{pmatrix} 0.2 & 0.12 \\ 0.1 & 0.18 \end{pmatrix} \quad .$$

and

$$\gamma_0 = \begin{pmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{pmatrix} \quad .$$

The condition

$$\|\Delta t(\sigma\sigma^T)^{-1}\gamma_0\| \leq 1$$

is then fulfilled if

$$\Delta t \;\leq\; \frac{1}{\|(\sigma\sigma^T)^{-1}\gamma_0\|} = 0.0637$$

Figures 4.22 and 4.23 illustrate this result.

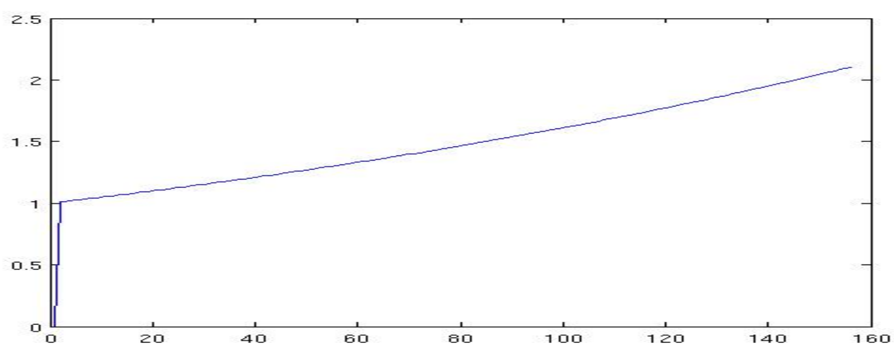

Figure 4.22: Behaviour of the norm $\|\Delta t(\sigma\sigma^T)^{-1}\gamma_0\|$ for $\Delta t = 0.064$

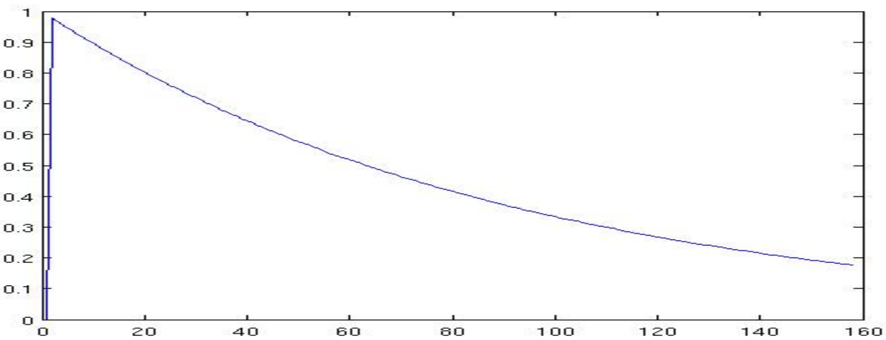

Figure 4.23: Behaviour of the norm $\|\Delta t(\sigma\sigma^T)^{-1}\gamma_0\|$ for $\Delta t = 0.063$

Parameter estimation in this model does also depends on the Kalman filter and will be studied next for infrequent trading. We find the same difficulties with parameter estimation when trading less frequent than the boundary for the filter to be bounded: Figure 4.24 illustrates that the algorithm does not produce estimates at all in the case of that rare trading until the before computed frequency. Displayed is the estimation of $\sigma\sigma^T$, but the estimation of the other two parameters shows the same behaviour.

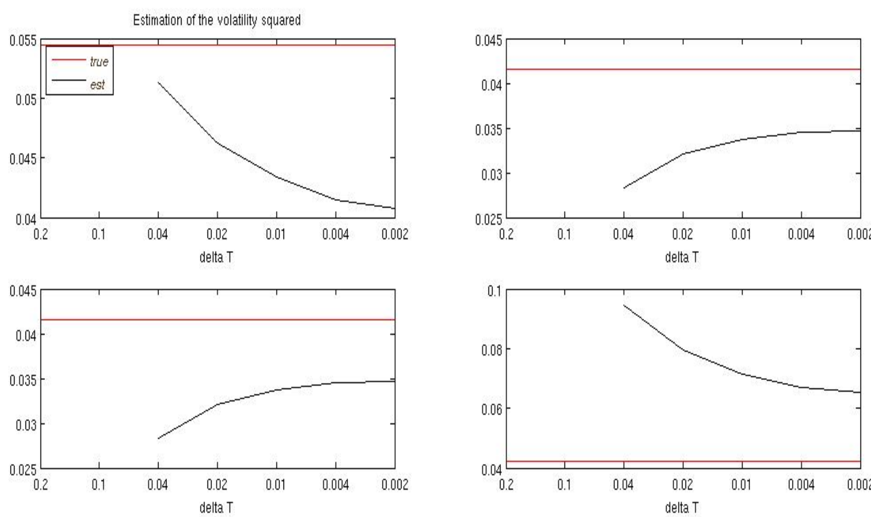

Figure 4.24: Estimation of $\sigma\sigma^T$

4.7 Advantage with an Increasing Number of Stocks

At an increase in the number of stocks, the dimension of the matrix A does not change, but its values do since both the states and the volatility grow in dimension. The driving Markov chain does not change, it depends only on the number of states.
The following example shows why the mean of the process ψ plays a particularly big role in the case of many stocks.

4.38 Example. *We consider 25 independent stocks with parameters*

$$\begin{aligned} B^1 &= \ 1 \cdot \mathbf{1}_{25} \\ B^2 &= \ 0.1 \cdot \mathbf{1}_{25} \\ B^3 &= -1 \cdot \mathbf{1}_{25} \\ \sigma_{ij} &= 0.2 \text{ for } i = j \text{ and } \ 0 \text{ for } i \neq j, \ i = 1, \dots, 25 \end{aligned}$$

and the rate matrix from (MP1).
Following Proposition 4.6 the growth rate of the mean of the process ψ is given by $\Delta t (B^j)^T (\sigma\sigma^T)^{-1} B\nu$ with $j = 1, \dots, d$.
This growth rate is with the above parameters for a general number of stocks n in the first state given by

$$\begin{aligned} & 0.004 \cdot \mathbf{1}_n \cdot 25 \cdot \mathbf{1}_n^T \cdot \left(\begin{pmatrix} 1 & 0.1 & -1 \end{pmatrix} \begin{pmatrix} 0.2727 \\ 0.4545 \\ 0.2727 \end{pmatrix} \right) \\ &= 0.004545 \cdot \mathbf{1}_n \cdot \mathbf{1}_n^T \\ &= n \cdot 0.004545 \quad . \end{aligned}$$

This value increases unboundedly with an increase in the number of stocks. The experiment will lead to an unbounded robust scheme for the unnormalized filter whenever the value is close to one which will happen eventually due to the increase. Figure 4.25 displays the unnormalized filter from the above situation.

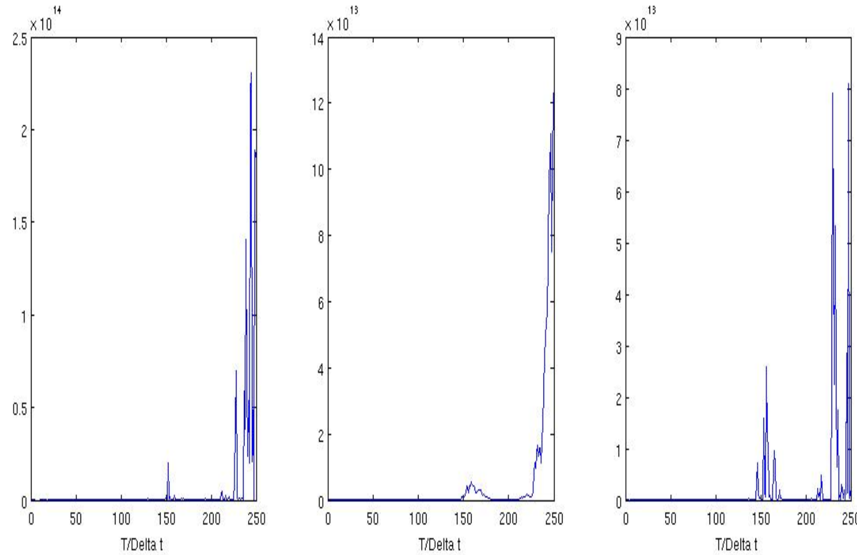

Figure 4.25: Unnormalized filter at 25 stocks with a diagonal volatility matrix

Parameter estimation is then either not possible at all or at least not leading to valuable results.

We will now adjust the state matrix such that the mean is a lot smaller and repeat the experiment.
The volatility matrix will not be changed, the states used will be 0.05, 0.01 and -0.05.
With these parameters, 20 stocks and one year simulation time, the scheme for the unnormalized filter was bounded and we got the following satisfying parameter estimation results for the rate matrix displayed in Figure 4.26. The estimates are again the mean of one hundred runs of the EM algorithm with one hundred data sets produced from the same parameters.

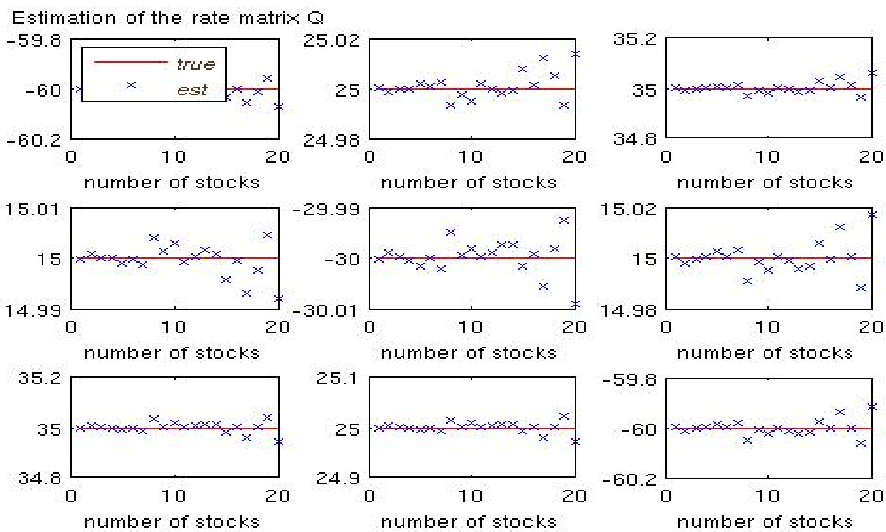

Figure 4.26: Estimation of the rate matrix with independent stocks

Unfortunately, we cannot observe the desired decrease of the standard deviation in Figure 4.27. These figures imply that although the scheme for the unnormalized filter is bounded and positive in this experiment and the parameter estimation works well enough, this will not be true for larger numbers of stocks.

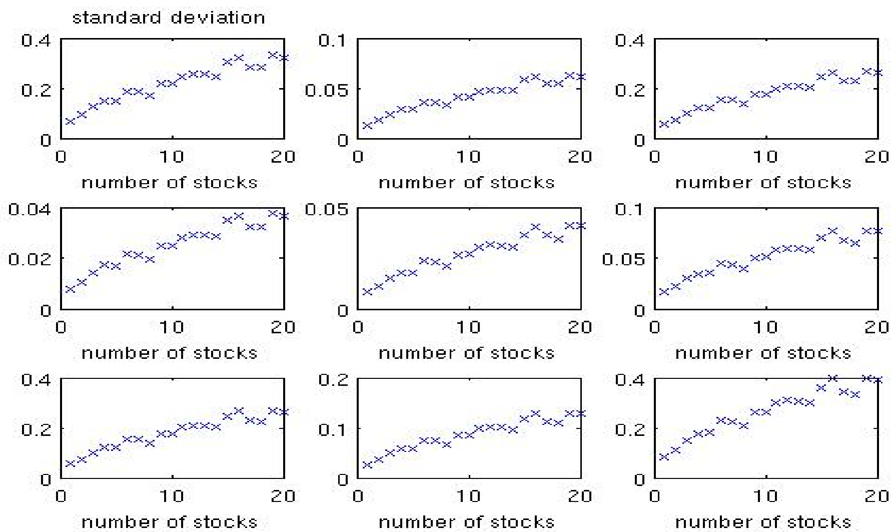

Figure 4.27: Standard deviation in the estimation of the rate matrix with independent stocks

The reason is again what we computed in the previous example: The general structure we computed in the example as $(n \cdot \text{number})$ is still valid and therefore the mean is still increasing with the number of stocks. Hence the mean will eventually pass the limit from which on the scheme for the unnormalized filter is not bounded anymore and the parameter estimation will not be satisfying.

4.39 Remark. *We saw before that with lowering the values of the states, the filter stays bounded and positive for some higher numbers of stocks, but not eventually. By increasing the trading frequency, i.e. using a smaller Δt, a similar effect would appear: The filter would be bounded and positive for higher numbers of stocks, but the increase in n would still have more weight.*

To break this pattern of unboundedness for higher numbers of stocks, we introduce another volatility matrix that has nonzero offdiagonal elements.

4.40 Example. *We used the following volatility matrix and did not change* *the state and rate matrices:*

$$\sigma_{i,j} = \begin{cases} 0.2, & \textit{if } i = j \\ 0.1, & \textit{if } i \neq j \end{cases}$$
$$\textit{for } i,j = 1,\ldots,n \quad .$$

We will now compute the growth rate of the mean of the process ψ with these parameters for a general number of stocks n for the first state.

$$\begin{aligned} &0.004 \cdot \mathbf{1}_n \cdot (\sigma\sigma^T)^{-1} \cdot \mathbf{1}_n^T \cdot \left(\begin{pmatrix} 1 & 0.1 & -1 \end{pmatrix} \begin{pmatrix} 0.2727 \\ 0.4545 \\ 0.2727 \end{pmatrix} \right) \\ &= 0.0001818 \cdot \mathbf{1}_n \cdot (\sigma\sigma^T)^{-1} \cdot \mathbf{1}_n^T \\ &= 0.0001818 \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} (\sigma\sigma^T)_{ij}^{-1} \end{aligned}$$

This sum is strictly decreasing since the sequence $\sum_{i=1}^{n}\sum_{j=1}^{n}(\sigma\sigma^T)_{ij}^{-1}$ is monotonously decreasing with limit zero for increasing n.
Parameter estimation is then as expected very well.

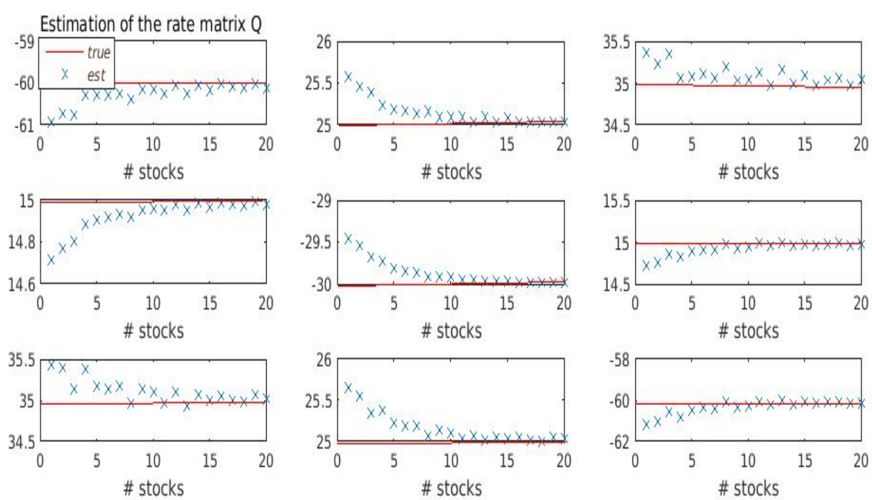

Figure 4.28: Estimation of the rate matrix

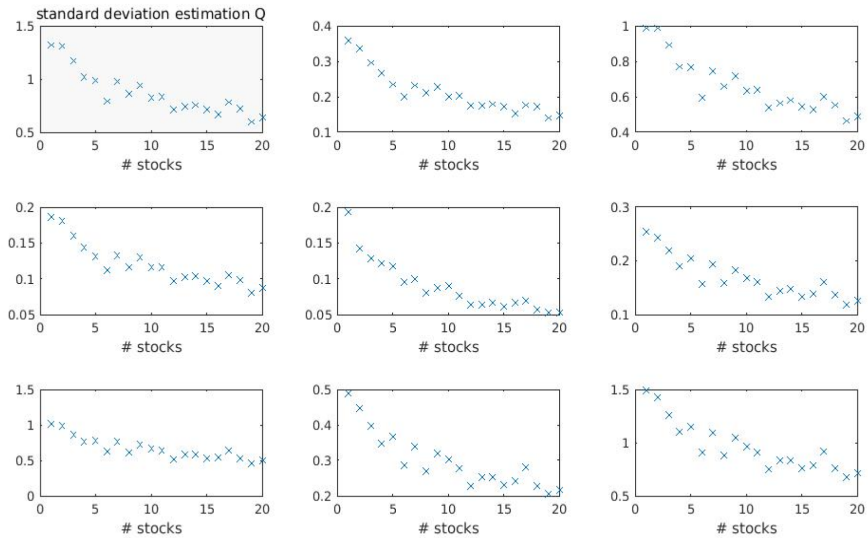

Figure 4.29: Standard deviation in the estimation of the rate matrix

4.41 Remark. *We have now seen two examples for the computation of the growth rate of the mean of the process* ψ *under the real world measure. Under the risk neutral measure, this expected value is equal to one (see Lemma 4.1), but under this measure the investor cannot gain anything. His objective is thus that the real world and the risk neutral measure differ. In this situation this would imply that the mean of the process* ψ *is as far from one as possible. But then we encounter problems with boundedness and positivity as was discussed earlier.*

4.42 Remark. *From the above considerations we can conclude that*

$$\Delta t(B^j)^T(\sigma\sigma^T)^{-1}B\nu$$

has to be positive, smaller than one and not increasing for the growth rate to be bounded with the increase in the number of stocks for all $j = 1, \ldots, d$ *.*

4.43 Example. *The parameter estimation algorithm for the HMM model depends heavily on filters including the filter we observed here. Now if this unnormalized filter is close to the stationary distribution, one could use the* *stationary distribution instead of the unnormalized filter and save some computation time. All other filters would still be computed according to their definitions, though. The results of such an experiment can be seen in Figure 4.30 where the estimation with the stationary distribution is displayed in red and blue shows the results for the original EM algorithm.*

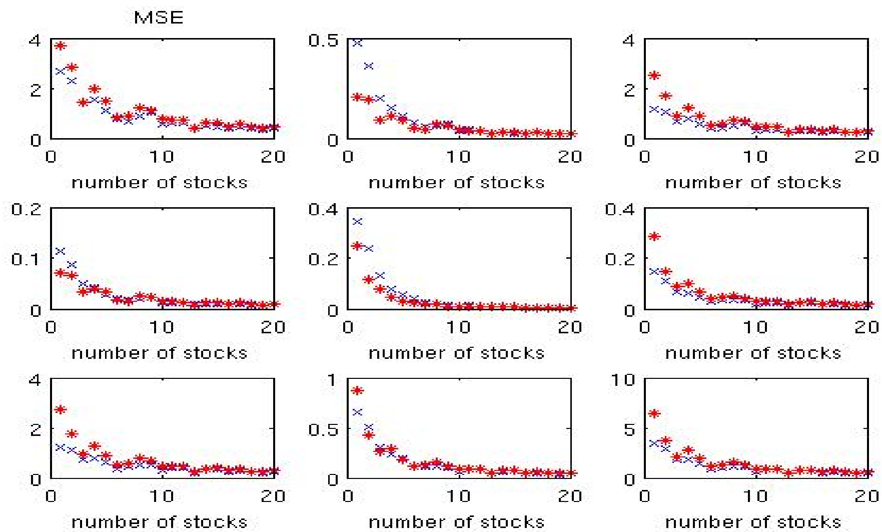

Figure 4.30: Estimation of the rate matrix with the stationary distribution

4.44 Remark. *If* $\Delta t(B^j)^T(\sigma\sigma^T)^{-1}B\nu$ *were negative and decreasing we would not encounter a problem with boundedness or positivity, but the filter would be close to approaching zero which is not desirable either since there is no information to be gained from it then. This same problem appears if the term is bigger than one and increasing. In Figure 4.31 we included an example where the scheme for the unnormalized filter is bounded, but still not of desirable shape since it is nearing zero.*

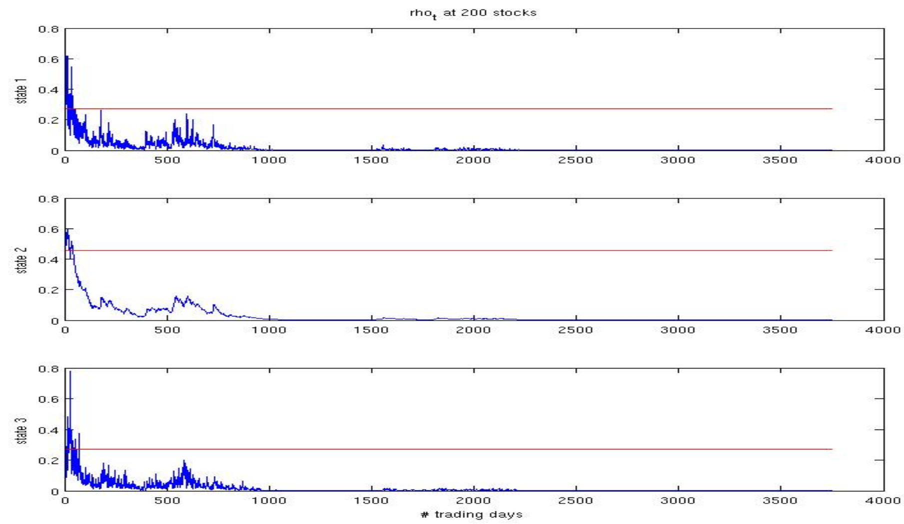

Figure 4.31: Unnormalized filter at 200 stocks, 3750 trading days

We now know all requirements for the scheme for the unnormalized filter to be bounded and preserving positivity at large numbers of stocks, but we can also observe that the normalized filter still changes over an increase in the number of stocks.
As a measure for the filter performance we will again use the MSD as

$$\mathbb{E}[(\nu - \eta_t)^T(\nu - \eta_t)] = \mathbb{E}[(\nu - \rho_t\zeta_t)^T(\nu - \rho_t\zeta_t)]$$

and observe it at terminal time T for the increasing number of stocks as well as for all $t \leq T$ for different numbers of stocks.

4.45 Remark. *Since both the stationary distribution and the normalized filter have only positive entries that sum up to one (at each time step in the case of the normalized filter), the MSD is contained in the unit interval* $[0, 1]$.

If we now compare the normalized filter with the stationary distribution of the Markov chain over all timesteps $t \leq T$, we will see that with an increase in the number of stocks, the size of the error lessens. Figure 4.32 displays the MSD for the normalized filter in all three states and for three different numbers of stocks: 1, 10 and 25. This relates to the earlier discussed result that the MSD decreases with a decrease in the matrix $A = B^T(\sigma\sigma^T)^{-1}B$

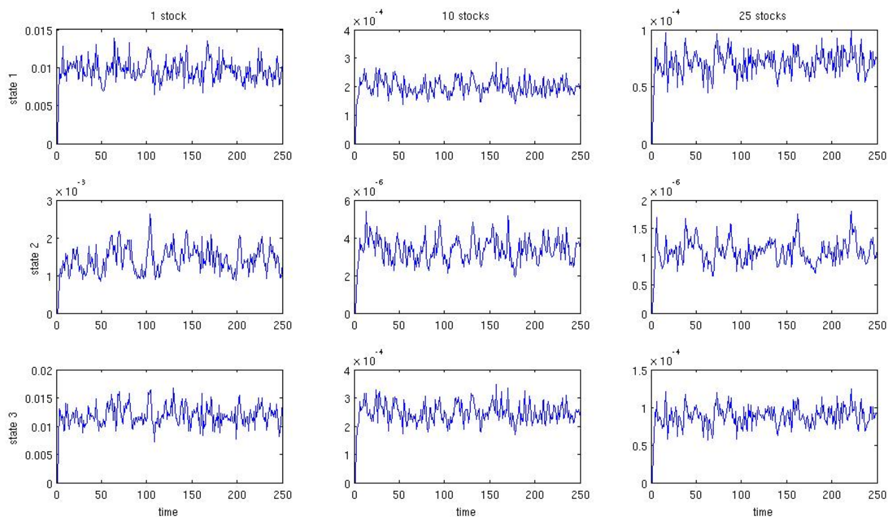

Figure 4.32: MSD of the normalized filter over time

which is naturally given in this model with the increase in the number of stocks.

4.46 Remark. *In Figure 4.32 we cannot only detect a decrease of the mean of the error, but also in the variance.*

The decrease in the MSD over an increase in the number of stocks will be our next subject of interest. Therefore we will not regard the increase over all timesteps any longer, but concentrate on the effect of the number of stocks by computing the MSD at terminal time T.
Figure 4.33 displays the results of the computation of the MSD of the filters and the stationary distribution at terminal time T over an increase in the number of stocks up to 25. As expected, the error lessens.

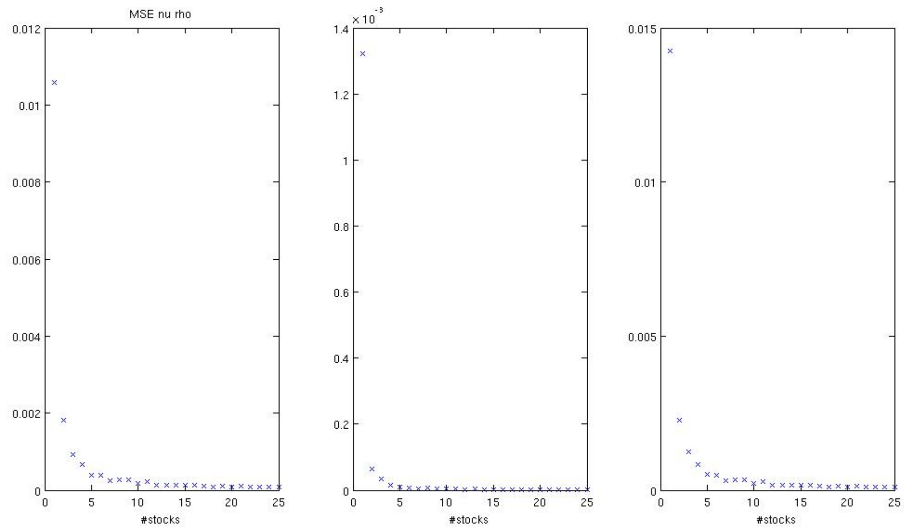

Figure 4.33: MSD of the normalized filter at terminal time

4.47 Remark. *Given an increasing number of stocks, the scheme for the unnnormalized filter* ρ_t *will eventually be near the stationary distribution of the Markov chain if the scheme is bounded for all numbers of stocks and is not nearing zero.*

This numerical observation relates to the result in [19] on the convergence of the unnormalized filter and the stationary distribution. To illustrate this result, we show a comparison of the unnormalized filter in the onedimensional model (Figure 4.34) and the one generated with 25 stocks (Figure 4.35) with the previous parameters. The initial distribution is in this case (0.2727 0.4545 0.2727) and the unnormalized filter is much closer to it with 25 stocks.

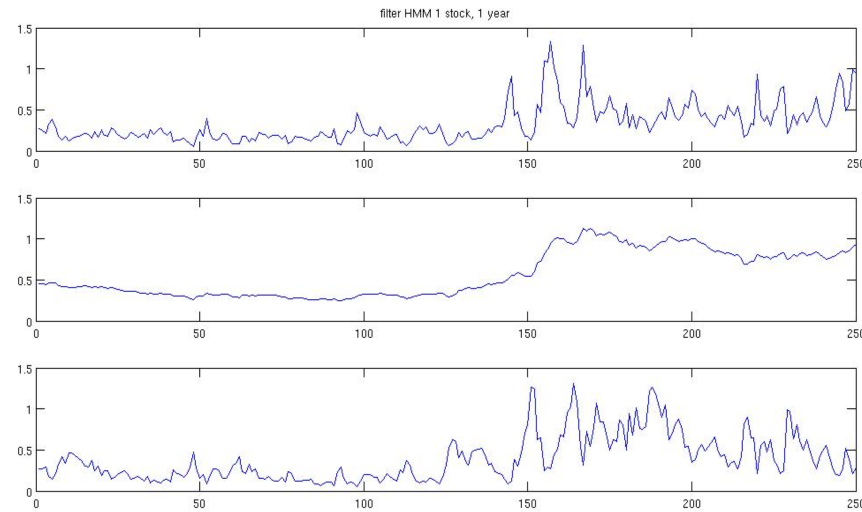

Figure 4.34: Unnormalized filter for one stock

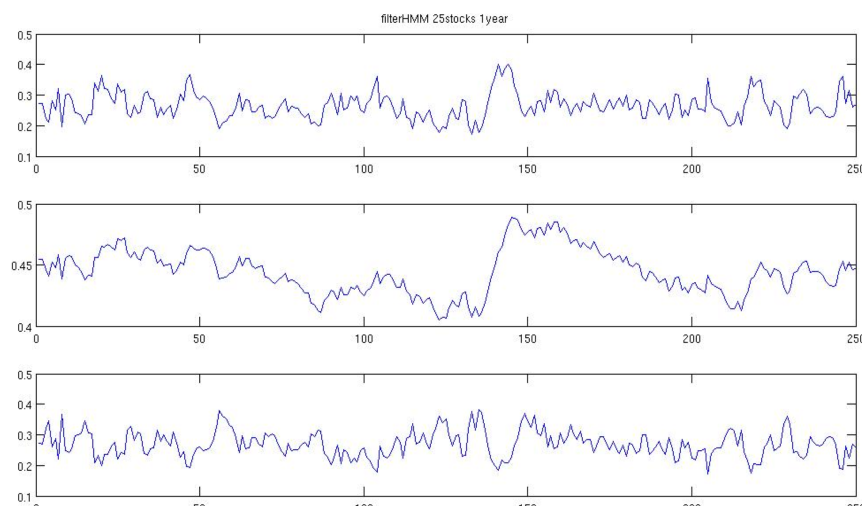

Figure 4.35: Unnormalized filter for 25 stocks

We turn to the study of the density ζ.

4.48 Remark. *The density ζ_t is given by*

$$\zeta_t^{-1} = \mathbf{1}_d^T \rho_t \quad .$$

Therefore it is close to one if the filter is close to the stationary distribution since the entries in the stationary distribution always sum up to one.

Figure 4.36 presents an overview over the behaviour of the density ζ over our usual 250 trading days in one year for different numbers of stocks. One can observe that the process stays closer to one the more stocks are involved.

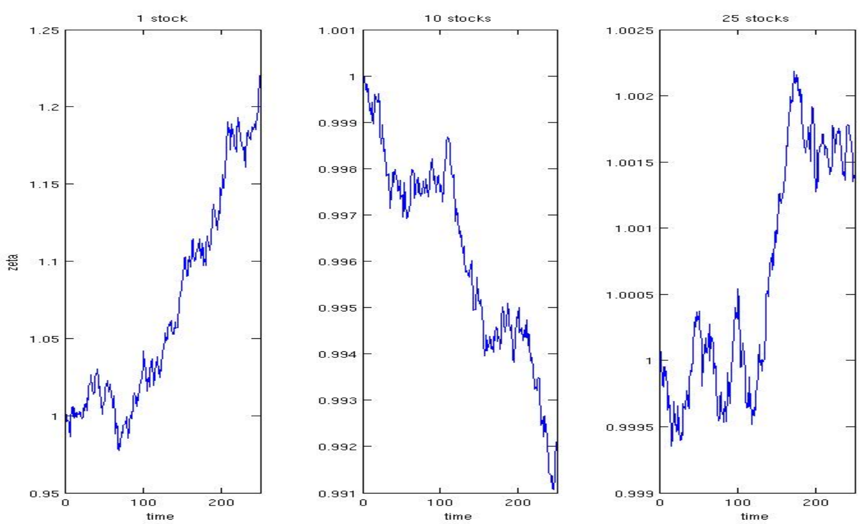

Figure 4.36: ζ_t for 1, 10 and 25 stocks

4.49 Remark. *We have now observed the process ζ_t in the situation of an increase in the number of stocks and found conditions under which it is positive while technically we were analyzing the MSD.*
It does have another positive side effect though: By ensuring ζ_t to be positive, bankruptcy is excluded.

With the parameters used in this chapter for a bounded and positive scheme for the unnormalized filter at all numbers of stocks, we get an appearance of the process ζ_T for 50 stocks as in Figure 4.37 which comes as usually from 100 computations.

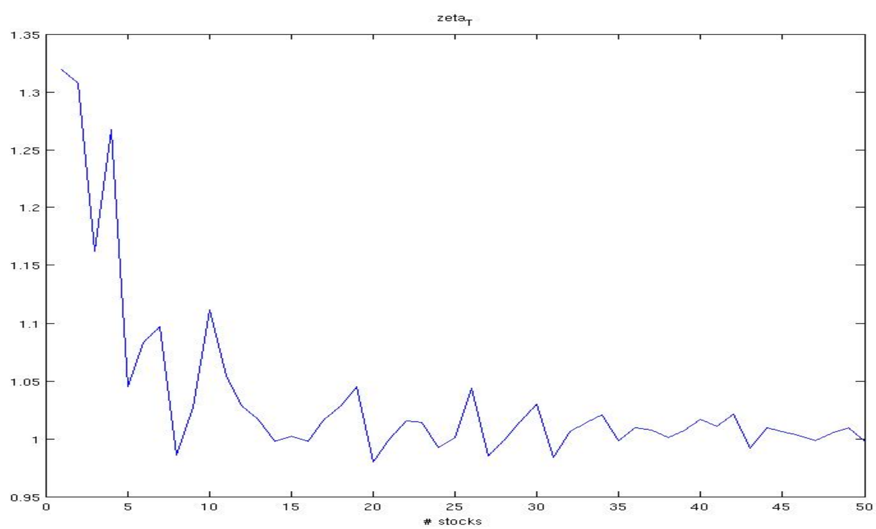

Figure 4.37: Process ζ at terminal time for up to 50 stocks

With the earlier stated changes that occur with an increasing number of stocks, we can also find that the variance in the portfolio reduces.

4.50 Lemma. *With the utility function $u(x) = \log(x)$, the variance of the utility of the optimal terminal wealth will reduce if $\tilde{B}_j^T(\sigma\sigma^T)^{-1}\tilde{B}\nu$ decreases towards zero for all $j = 1, \ldots, d$ at an increase in the number of stocks.*

Proof. (Sketch)

$$\begin{aligned}\mathrm{Var}\left(u(X_T^*)\right) =& \mathrm{Var}(\log(X_T^*)) = \mathrm{Var}(\log(x_0/\zeta_T)) = \mathrm{Var}(\log(\zeta_T))\\ =& \mathbb{E}\Big[\log\left(\zeta_T\right)^2\Big] - \mathbb{E}\Big[\log\left(\zeta_T\right)\Big]^2\end{aligned}$$

The function $\log^2(x)$ attains its minimum zero for $x = 1$. Since ζ_T is approaching one, $\log(\zeta_T)$ approaches zero and this leads to a decrease in the variance of the portfolio. □

If we consider power utility, the argument will not be as nice since we are a lot further away from actually computing the variance from it, but it is still valid.

4.51 Lemma. *With the utility function $u(x) = \frac{1}{\kappa}x^\kappa$, $\kappa < 0$, the variance of the utility of the optimal terminal wealth will reduce if $\tilde{B}_j^T(\sigma\sigma^T)^{-1}\tilde{B}\nu$ decreases for all $j = 1, \ldots, d$ over an increase in the number of stocks.*

Proof. With the same argument as before about the process ζ_T and a similar computation of the variance, the variance decreases. □

With these considerations we are now in a position to discuss the choice of the boundary M from Proposition 4.6.

The bound M restricts the expected value of the scheme for the unnormalized filter

$$\Big\|\mathbb{E}\big[\rho_t\big]\Big\| < M \quad .$$

Until now we have not made any comment on how to choose M other than that it should be greater than one and finite.
We know that $\zeta_t^{-1} = \sum_{j=1}^d \rho_t$. If we now use the L_1 norm in the inequality we get

$$\Big\|\mathbb{E}\big[\rho_t\big]\Big\| = \sum_{j=1}^d \mathbb{E}\big[\rho_t\big] = \mathbb{E}\big[\zeta_t^{-1}\big] < M$$

which then in turn gives us a bound on the expected utility of the optimal terminal wealth. We will continue with logarithmic utility and apply Jensen's inequality:

$$\mathbb{E}\big[\log(x_0\zeta_T^{-1})\big] \leq \log(x_0) + \log(\mathbb{E}\big[\zeta_T^{-1}\big]) < \log(x_0 M) \quad .$$

From this relation, the investor's interest would be to choose M as large as possible. On the other hand side the variance is a quantity usually desired to be as small as possible. Using Lemma 4.50 the variance is linked to the bound M in the following way

$$\begin{aligned}&\mathrm{Var}[\log(X_T^*)] = \mathrm{Var}[\log(\zeta_T^{-1})] = 2\mathbb{E}[\log(\zeta_T^{-1})] - \left(\mathbb{E}[\log(\zeta_T^{-1})]\right)^2 \\ &\leq\ 2\log\left(\mathbb{E}[\zeta_T^{-1}]\right) - \left(\log(\mathbb{E}[\zeta_T^{-1}])\right)^2 \ <\ 2\log(M) - \log(M)^2 \quad .\end{aligned}$$

With this relation the investor would be advised to choose M as small as possible to reduce the variance in the portfolio.

Gaussian Model In the Gaussian Model the requirements of Lemma 4.10 need to be met for the drift to be bounded.

4.52 Example. *We construct the parameters uniformly, in this case α is set to be* 0.7 *on the diagonal and* 0.1 *elsewhere. Then its eigenvalues are positive for all numbers of stocks. If we now replace all values* 0.1 *with* 1.1*, this will only be true for one stock and we get results as in Figure 4.38 where the drift does not comply with our required boundedness conditions.*

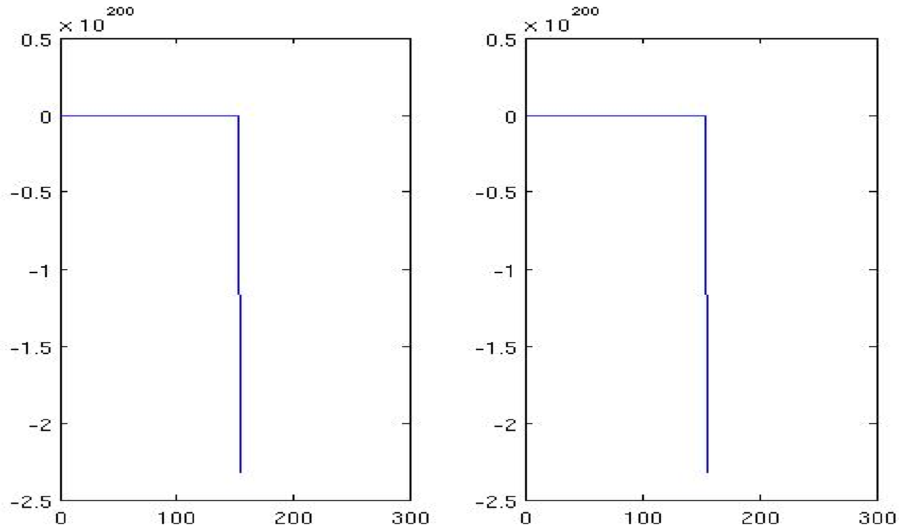

Figure 4.38: Drift at 25 stocks

With the results on the boundedness of the filter in mind, we will now once again observe the MSD of drift and filter where both processes are bounded. Figure 4.39 shows that the MSD does not decrease visibly over the number of stocks that increases to fifty in this figure.

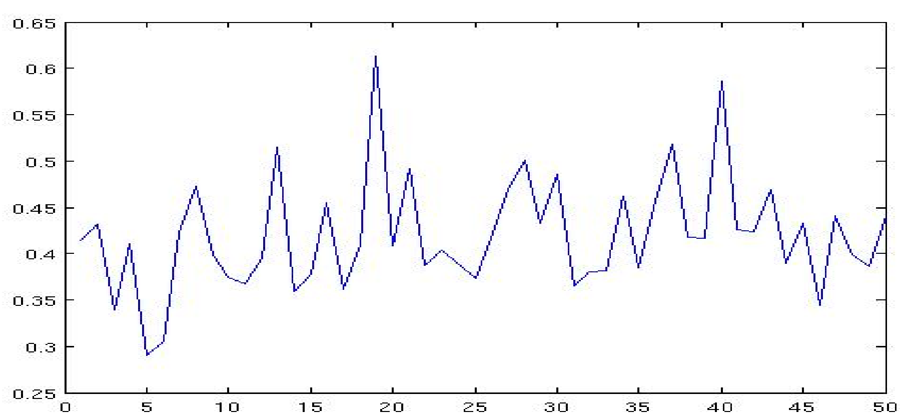

Figure 4.39: MSD of drift and filter

Let us now turn to the study of the process ζ in this model.

4.53 Remark. *The density ζ_t will be tending towards zero if it is bounded. Let us recall the definition of ζ_t as*

$$\zeta_t = \frac{1}{1 + \int_0^t \frac{1}{\zeta_u}(m_u - r\mathbf{1}_n)^T(\sigma\sigma^T)^{-1}\, d\tilde{R}_u} \quad .$$

It is obvious, that the convergence depends on the behaviour of the stochastic integral. If the integral is positive and increasing, the density will go to zero. If the integral is smaller than one, the density will take on negative values. In case the integral comes close to -1, we get problems with boundedness. Due to the factor $\frac{1}{\zeta_u}$, the stochastic integral only reacts to large changes in the value of the integral $\int_0^t \frac{1}{\zeta_u}(m_u - r\mathbf{1}_n)^T(\sigma\sigma^T)^{-1}\mu_u\, du$. If this integral is about the same for an interval, the stochastic integral will be close to zero. This deterministic integral in turn follows the process $(m_u - r\mathbf{1}_n)^T(\sigma\sigma^T)^{-1}\mu_u$. Thus the process ζ_t will not be close to zero if $(m_u - r\mathbf{1}_n)^T(\sigma\sigma^T)^{-1}\mu_u$ alters only with very small deviation in which case we do not have boundedness.

If the process $(m_u - r\mathbf{1}_n)^T\mu_u$ is not volatile, the stochastic integral will take on values around zero which leaves the process ζ_t with the possibility of being negative, taking on too large values or even being unbounded.

4.54 Example. *In Figure 4.40 we can observe the situation discussed in the previous lemma. The matrix α was 0.7 on the diagonal and 0.1 elsewhere, β was 0.3 on the diagonal and -0.1 elsewhere and the volatility was set to be 0.2 on the diagonal and 0.1 elsewhere.*

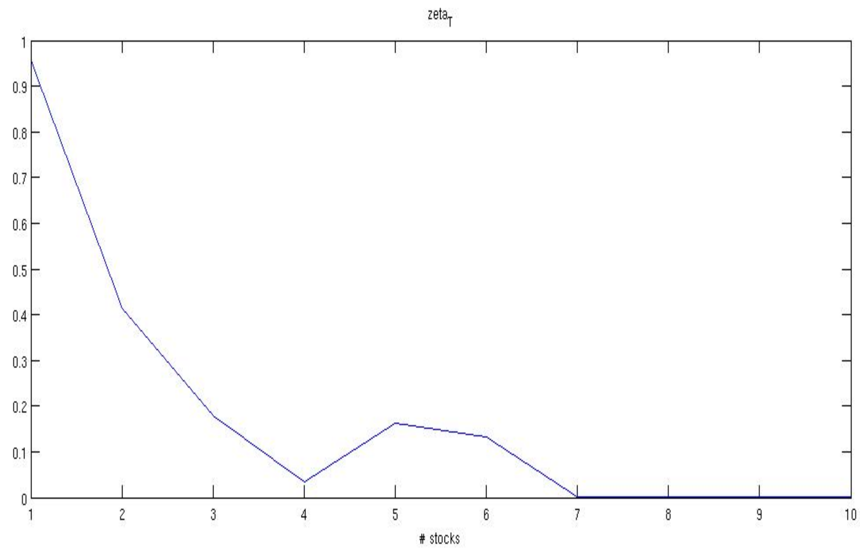

Figure 4.40: Process ζ at terminal time

4.55 Remark. *The process* $(m_u - r\mathbf{1}_n)^T\mu_u$ *will be more volatile with an increase in the number of stocks if the parameters* α *and* β *have an increasing sum of absolute values of column or row elements respectively.*
The term accounting for deviation in the drift is $\beta\, dW_t$ *and thus the deviation will increase if the sum of row elements does.*
The drift is also present in the filter through the return process. In addition, the deterministic part will produce higher values in absolute value if the matrix α *increases its sum of column elements in absolute value.*

The possibility that ζ_t might be negative came up in this paragraph. This means that the investor goes bankrupt and depending on the parameters, this might be a frequent event in this scenario. But we also discovered that this happens less the more stocks are involved, i.e. the more stocks the investor has in his portfolio, the less frequent he will go bankrupt.

The algorithm for parameter estimation does not deal well with an increase in the number of stocks. The algorithm is based on an iterative recursive method to approximate the finite dimensional filters for linear dynamic systems as given here and was introduced by [11]. The Kalman filter for the drift is a special case of these filters, but as in the HMM, more quantities need to be filtered for parameter estimation. The parameter estimates are then given as fractions of such recursive finite dimensional filters. These filters are called finite dimensional because they can be expressed by a finite number of sufficient statistics and these statistics can then be computed recursively. In numerical implementations though, these recursions may attain very large values during computation which would in theory cancel out, but lead to a termination of computations without results. Figure 4.41 illustrates this by means of one of the statistics (a_0) and the comparison between one and four stocks.

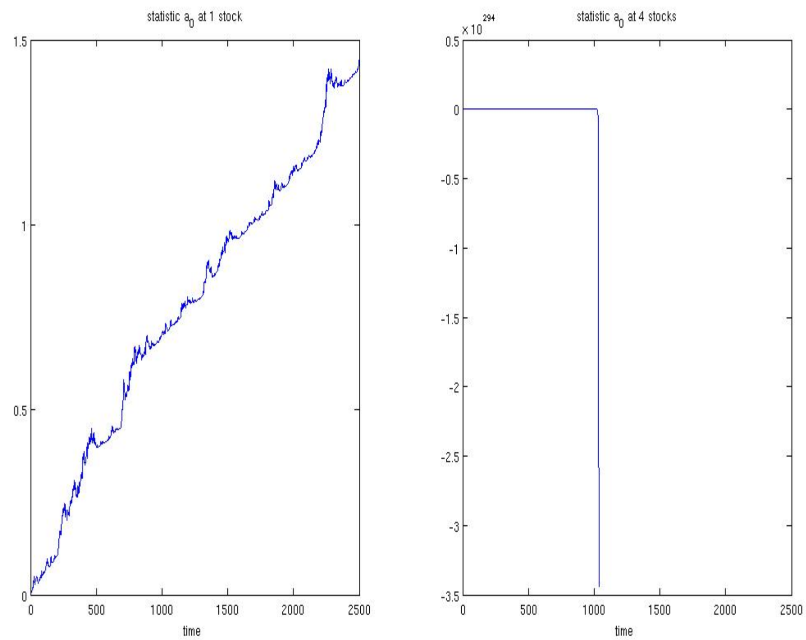

Figure 4.41: Statistic in the model with one stock (left) and four stocks (right)

5 Loss in Utility at Different Market Parameters

When the partially and fully informed investors start with the same initial capital, we have already seen that the partially informed investor is at a disadvantage. The question we want to answer in this section is now how much more initial capital the partially informed investor needs to reach a comparable terminal wealth at different market parameters. If not otherwise stated, we only consider market situations where filtering is an advantage according to the previous chapter, for example the trading frequency and number of stocks will be chosen accordingly.

5.1 A Measure for the Loss in Utility

With this approach we follow an idea from [5] who introduces the difference in the initial capital as a measure for the loss in utility. The measure will be based on the earlier introduced method to compute the optimal terminal wealth:
For $u(x) = \frac{x^\kappa}{\kappa}$ we know from Theorem 3.12 how to compute the optimal terminal wealth in the case of full information:

$$X_T^* = (\hat{y}\beta_T Z_T)^{\frac{1}{\kappa-1}}$$

with

$$\hat{y}^{\frac{1}{\kappa-1}} = \frac{x_0}{\mathbb{E}[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}]} \quad .$$

To measure the loss, we need to find $\hat{x}_0$ so that the partial investor has the same expected utility of the terminal wealth as the fully informed investor.

$$\mathbb{E}[u(\hat{X}_T^*)] \overset{!}{=} \mathbb{E}[u(X_T^*)]$$

From this we then get the desired expression for the initial capital

$$\hat{x}_0 = \frac{x_0\ \mathbb{E}[(\beta_T \zeta_T)^{\frac{\kappa}{\kappa-1}}]^{\frac{\kappa-1}{\kappa}}}{\mathbb{E}[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}]^{\frac{\kappa-1}{\kappa}}} \quad .$$

For logarithmic utility we follow the same principle to get

$$\hat{x}_0 = x_0\ \exp\left(\mathbb{E}\left[\log\left(\frac{\zeta_T}{Z_T}\right)\right]\right) \quad .$$

The loss is then as described above the difference between the initial capitals

$$\text{loss} = \hat{x}_0 - x_0 \quad .$$

Numerical results We use this measure for the loss to observe the loss for the fully informed investor who knows the parameters and the one who estimates parameters (marked as fi -,fi PE in Table 5.1).
The same quantities for the partially informed investors with and without knowledge of the parameters are denoted by under pi -,pi PE.
Additionally the differences fi -, pi - (full information without parameter estimation and partial information without parameter estimation) and fi -, pi PE (full information without parameter estimation and partial information with parameter estimation) are listed.

As the tables on the terminal wealth already indicated, the loss that occurs due to partial information is much bigger than the loss caused by parameter estimation.

Markov Model The results for the first parameter set in the Markov Model are displayed in Table 5.1. Once again, they stem from one hundred samples.

parameter set (MP1)	fi -,fi PE	pi -,pi PE	fi -,pi -	fi - ,pi PE
mean log	6.13	0.20	434.50	524.04
std log	4.84	0.59	181.35	209.32
mean power $\kappa = 0.1$	7.67	0.21	905.25	1094.41
std power $\kappa = 0.1$	6.37	0.02	373.40	453.61
mean power $\kappa = -1$	3.01	0.19	21.88	26.31
std power $\kappa = -1$	4.87	0.02	16.56	19.85
mean power $\kappa = -10$	3.63	0.18	3.90	4.81
std power $\kappa = -10$	10.72	0.02	5.16	6.12

Table 5.1: loss in utility for (MP1)

Next we will assume worse estimation results to see how much influence bad parameter estimation can have. We display next the loss in initial capital for partial information with true parameters and a set of state matrices with different values. The values in these state matrices centre around the true values and deviate to smaller and higher values to the right and left respectively. Figure 5.1 serves mainly to show the possible extreme losses, more details can be observed in Figure 5.2.

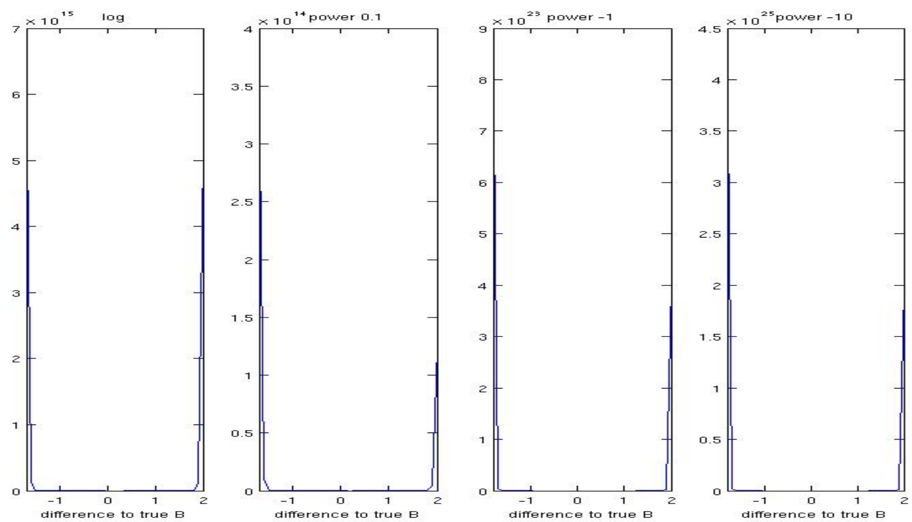

Figure 5.1: Loss due to bad parameter estimation

We can see here, that with a bad parameter estimation result, we may face a major loss. If we now zoom into the figure and take a closer look at a more reasonable range for the error in the estimation, we can see in Figure 5.2 that the loss due to partial information is more severe than the loss due to parameter estimation if we assume reasonable quality in the parameter estimation.

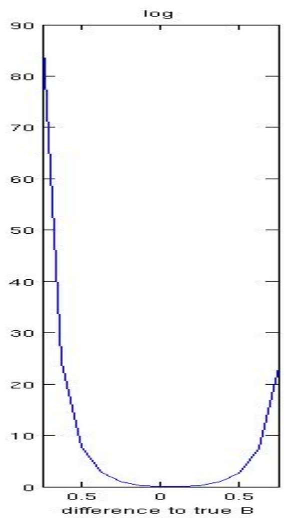

Figure 5.2: Loss due to bad parameter estimation

Gaussian Drift The method we introduced in the previous paragraph was based on a theorem that is valid for both models. Therefore we will follow the same approach with the Gaussian drift and state only the numerical results in Table 5.2.

parameter set (GP1)	fi -,fi PE	pi -,pi PE	fi -,pi -	fi - ,pi PE
mean log	75.55	0.22	146.07	179.01
median	55.67	0.22	117.49	143.68
std log	57.31	0.01	108.63	133.06
mean power $\kappa = 0.1$	$3.6 \cdot 10^{10}$	0.223	$9.6 \cdot 10^{10}$	$1.12 \cdot 10^{11}$
median	156.62	0.222	366.51	451.07
std	$3.6 \cdot 10^{11}$	0.01	$9.5 \cdot 10^{11}$	$1.12 \cdot 10^{12}$
mean power $\kappa = -1$	n.a.	0.221	n.a.	n.a.
median	2.5	0.221	3.4	4.4
std	n.a.	0.01	n.a.	n.a.
mean power $\kappa = -10$	n.a.	0.22	n.a.	n.a.
median	0.7	0.22	0.6	0.9
std	n.a.	0.01	n.a.	n.a.

Table 5.2: loss in utility for (GP1)

Due to the high volatility in the model, we deal with extreme values in this case. These may be too extreme for numerical evaluation in some cases, therefore the median was computed as well. The numerical evaluation fails due to the shape of the utility functions in this case: very slight changes in the wealth cause a huge difference in utility whenever the wealth is smaller than one. Add to that the possible larger changes due to the volatility in the model and numerical evaluation fails. Nevertheless, we can observe in this model as well, that the loss due to parameter estimation is much smaller than the one due to partial information.

5.2 Infrequent Trading

Usually our investor trades once a day on two hundred and fifty days per year. Now we want to determine what effect less frequent trading has on the loss in expected utility due to partial information.
All computations will be based on the same theoretical basis we introduced before. For the data to be consistent, we generate all data for the smallest Δt, i.e. the largest number of tradings, but only use the data at the time points we wish to examine.
But before we compare the optimal terminal wealth, please recall that the trading frequency cannot be chosen freely. Lemma 4.8 showed that the filter is not bounded for too infrequent trading. In fact, if we now observe the loss due to parameter estimation in partial information (Figure 5.3) at different trading frequencies, we will see that it does not lead to big losses in case the filter complies with the boundedness and positivity assumptions. It does not make sense to observe at frequencies where the filter is neither bounded nor preserving positivity, but after that the loss stays at about the same level we already computed before. That is consistent with our previous results on parameter estimation at different trading frequencies: The estimation at infrequent trading is rather poor, but from the before computed timepoint onward it is very good and and it then causes only minor losses. One may expect smoother outcomes in Figure 5.3, but the lack of smoothness is mainly due to the small range on the y-axis. In fact, the difference between data points is very small.

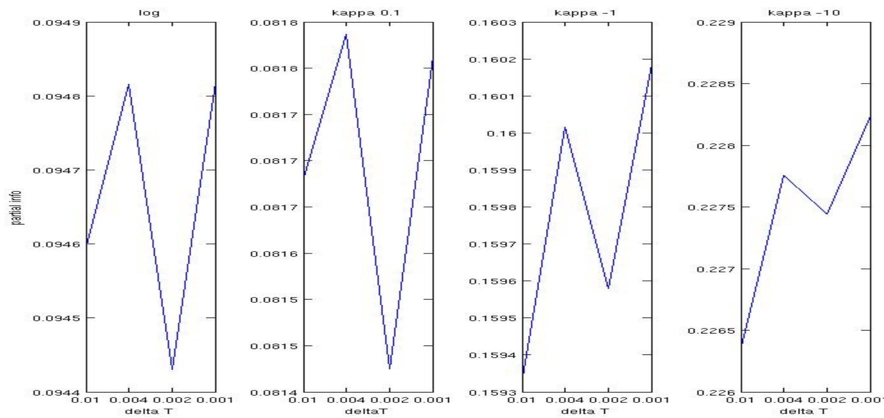

Figure 5.3: Loss in utility due to parameter estimation at infrequent trading, partial information, (MP2)

Figure 5.4 displays the loss due to partial information in comparison to full information without parameter estimation. As expected, the loss due to partial information increases with the number of tradings.

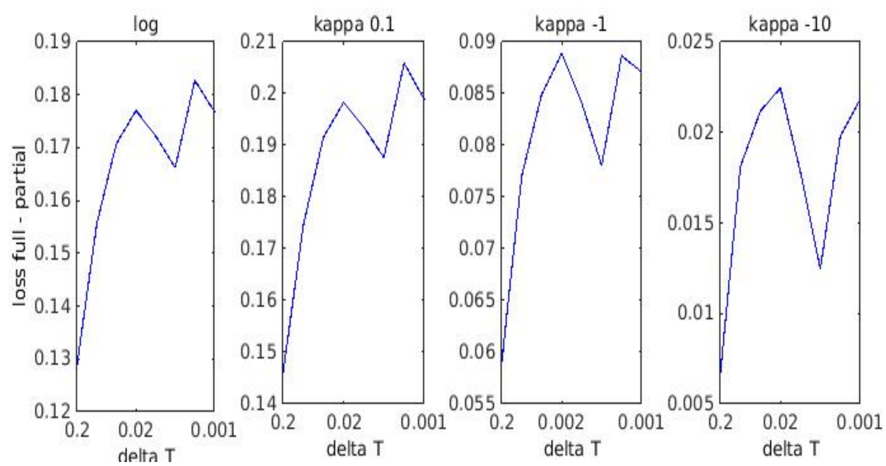

Figure 5.4: Loss in utility due to partial information at infrequent trading, (MP2)

An observation of the loss at different market parameters was also accomplished by [22], although for a different model. There was a constant drift in consideration leading to the reverse result of higher losses due to parameter uncertainty than due to infrequent portfolio rebalancing in the classical Merton model. There infrequent trading does not have large effects if the proportion invested is close to the Merton proportion, whereas straying from this proportion due to parameter uncertainty leads to higher losses.

5.1 Remark. *If the filter is not bounded, then the process ζ_t will not be either since it depends on the filter. In this case bankruptcy cannot be ruled out. In fact it will occur frequently.*

5.3 Increasing Number of Stocks

We will now take a closer look at the loss in utility for an increasing number of stocks. This model was already discussed in the previous chapter and we will now apply some of the results to discuss the loss.
In both the Markov and the Gaussian Model, the process ζ was analyzed closely for increasing number of stocks in Section 4.7. For a filter that is stable for all numbers of stocks, this density decreases towards one as in Figure 4.37 and zero in Figure 4.40 respectively.
To determine the loss, we now compare the investor who only trades in one stock with the investor with many stocks in his portfolio, starting with two and increasing by one in each iteration. Let us assume that all stocks depend on the same parameters for this first experiment. Since the loss then only depends on the density for one stock and the one for more stocks which decreases in each iteration, the loss itself increases for each additional stock. It does not increase unboundedly, since in both models the densities decrease boundedly.

If we now do not require similarity of the stocks, the properties of a single stock gain influence. As illustration we will now consider ten different stocks with the state matrix

$$B = \begin{pmatrix} 1.2 & 0.1 & -1.0 \\ 1.0 & 0.1 & -0.8 \\ 0.9 & -0.1 & -1.1 \\ 1.1 & 0.2 & -1.2 \\ 1.2 & 0.05 & -1.1 \\ 1 & 0.1 & -1.5 \\ 1.5 & 0.3 & -0.7 \\ 1 & 0.1 & -1.1 \\ 1.2 & -0.1 & -1 \\ 1.2 & 0.2 & -0.9 \end{pmatrix} .$$

The estimation of this matrix in the mean is very well:

$$\tilde{B} = \begin{pmatrix} 1.18 & 0.1 & -0.99 \\ 1.02 & 0.13 & -0.77 \\ 0.91 & -0.09 & -1.06 \\ 1.08 & 0.16 & -1.19 \\ 1.18 & 0.01 & -1.12 \\ 1.00 & 0.08 & -1.38 \\ 1.47 & 0.29 & -0.73 \\ 1.03 & 0.12 & -0.97 \\ 1.16 & -0.1 & -0.99 \\ 1.15 & 0.13 & -0.92 \end{pmatrix}$$

The estimation of the volatility matrix works as nicely and for the rate matrix we can see, that the estimation converges as for the standard stocks we had before in Figure 5.5.

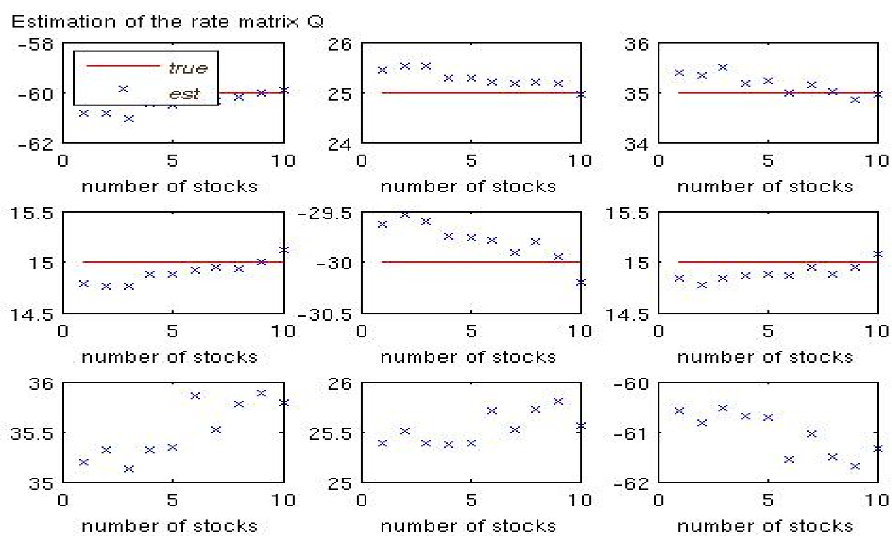

Figure 5.5: Estimation of the rate matrix

The loss that the investor who only invests in one stock has in relation to the one who has two to ten stocks is displayed in Figure 5.6. The influence of the very good stock number seven can be clearly seen, the loss does jump at that point.

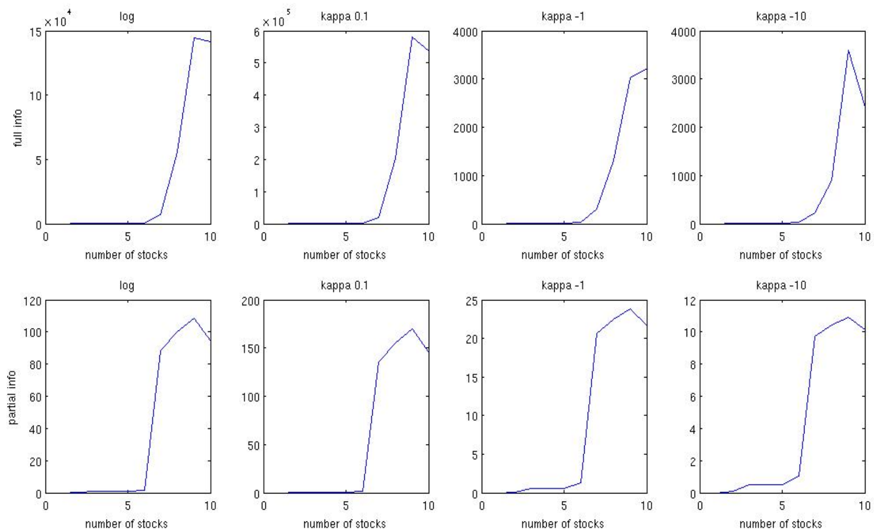

Figure 5.6: Loss at an increasing number of stocks

5.4 Increasing Number of States

This section will naturally only contain results regarding the Markov model. We will examine the loss in utility for a different number of states in the estimation of the Markov chain. To do that, we will consider the case of full information exactly as we did before, with three states according to (MP1). For the parameter estimation and the case of partial information though, we will consider an increasing number of states, i.e. we will apply the EM algorithm assuming the data was generated with different numbers of states. The loss in utility originating from the comparison of the optimal utility of the terminal wealth in full information and the case of full information with estimated parameters for a number of states increasing from one to ten is displayed in Figure 5.7. The same considerations for partial information can be observed in Figure 5.8.

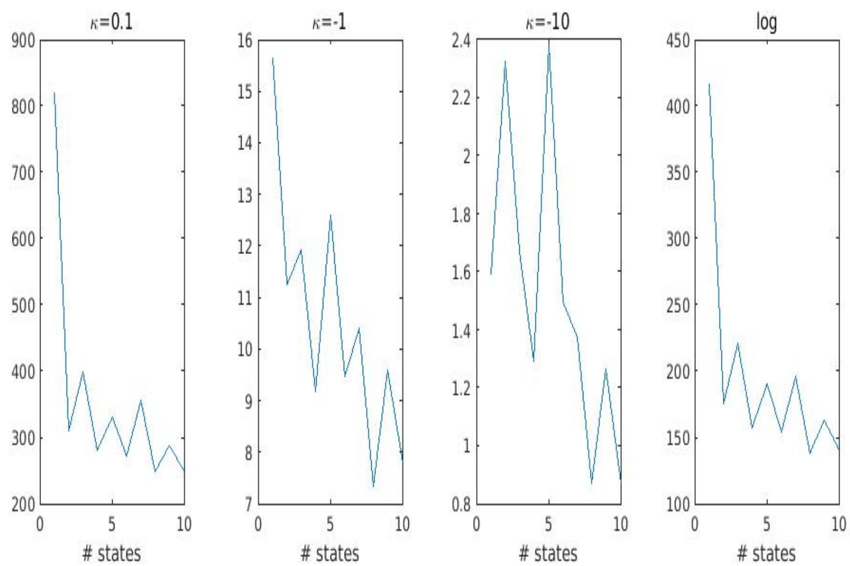

Figure 5.7: Loss due to a different number of states in parameter estimation at full information

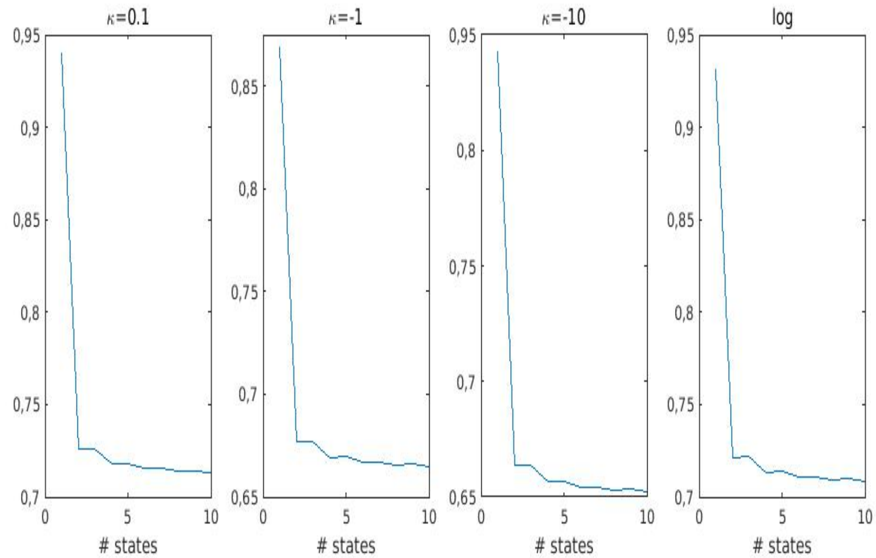

Figure 5.8: Loss due to a different number of states in parameter estimation at partial information

We can conclude that assuming less states than were originally used in the model is punished gravely whereas assuming more will cost some computation time, but not lead to higher losses.

If we were now to increase the number of states used in the original model, we can represent the market in more detail, but also face more computational effort. Especially due to the rising size of the matrices involved. For example, let us consider 10 states (which is already a very large number)

$$B = \begin{pmatrix} 1.2 & 0.1 & -1 & 0.5 & -0.8 & -0.1 & -0.9 & 0.8 \\ 1 & 0.1 & -0.8 & -0.6 & 0.7 & -0.1 & 1 & -0.9 \end{pmatrix} \quad .$$

The first consideration will be the parameter estimation, but the estimator for the state matrix B is absolutely satisfying:

$$\hat{B} = \begin{pmatrix} 1.07 & 0.07 & -0.9 & 0.23 & -0.51 & -0.08 & -0.62 & 0.43 \\ 0.93 & 0.11 & -0.7 & -0.5 & 0.53 & -0.05 & 0.74 & -0.76 \end{pmatrix} \quad .$$

The rate matrix also changes in dimension and we chose the entries so that the product of rate matrix and stationary distribution of the Markov chain

stays roughly the same.
We observe that the EM algorithm also deals well with this estimation:

$$Q = \begin{pmatrix} -80 & 30 & 10 & 10 & 10 & 10 & 5 & 5 \\ 20 & -60 & 10 & 5 & 5 & 5 & 5 & 10 \\ 20 & 5 & -50 & 5 & 5 & 5 & 5 & 5 \\ 10 & 5 & 5 & -40 & 5 & 5 & 5 & 5 \\ 10 & 5 & 5 & 5 & -40 & 5 & 5 & 5 \\ 10 & 5 & 10 & 5 & 10 & -50 & 5 & 5 \\ 20 & 10 & 5 & 10 & 5 & 5 & -60 & 5 \\ 30 & 10 & 5 & 10 & 5 & 10 & 10 & -80 \end{pmatrix}$$

And the corresponding estimator from one estimation was

$$\hat{Q} =$$

$$\begin{pmatrix} -75.84 & 31.08 & 10.56 & 8.24 & 7.76 & 10.42 & 3.82 & 3.95 \\ 19.53 & -53.56 & 10.02 & 3.96 & 3.74 & 5 & 3.68 & 7.63 \\ 20.08 & 5.07 & -45.54 & 3.99 & 3.8 & 5.04 & 3.74 & 3.83 \\ 12.78 & 6.67 & 6.81 & -48.03 & 5.25 & 6.7 & 5.17 & 4.64 \\ 15.56 & 7.88 & 7.80 & 6.71 & -57.2 & 7.84 & 4.86 & 6.55 \\ 9.84 & 5.01 & 9.98 & 3.97 & 7.47 & -43.77 & 3.68 & 3.82 \\ 32.23 & 16.38 & 8.15 & 13.9 & 5.19 & 8.17 & -90.82 & 6.79 \\ 41.91 & 14.68 & 7.6 & 10.84 & 5.85 & 14.8 & 11.52 & -107.21 \end{pmatrix}$$

The estimation of the volatility does not require any input that depends on the number of states and is therefore equally good for all numbers of states.

5.5 Loss due to the Approximation of the Trading Strategy

In both models, the optimal trading strategy in the case of partial information and power utility is a very complicated expression containing Malliavin derivatives, they are given as Formulas 3.3 and 3.5 respectively. In the case of logarithmic utility, it was much simpler: Exchanging the drift for the filter did the trick. We want to examine now what losses occur if the investor assumed this simple method also worked for power utility.

Method The previously applied method to determine losses in expected utility did not use the trading strategy and thus we will now introduce another way to determine the difference in the initial capital. It will be based on another representation of the terminal wealth that was introduced in Theorem 3.12

$$\beta_t X_t = x_0 + \int_0^t \beta_s \pi_s^T \sigma \, d\tilde{W}_s \quad .$$

We are again interested in comparing the expected utility of the optimal terminal wealth, i.e.:

$$\mathbb{E}[\log(X_T^*)] \quad \text{and} \quad \mathbb{E}[\log(\tilde{X}_T^*)]$$

as well as

$$\mathbb{E}\left[\frac{1}{\kappa}(X_T^*)^\kappa\right] \quad \text{and} \quad \mathbb{E}\left[\frac{1}{\kappa}(\tilde{X}_T^*)^\kappa\right] \quad .$$

In the case of logarithmic utility, we could simply compute the logarithm of the optimal terminal wealth and then compare both sides.

$$\tilde{x}_0 = x_0 \, \exp\Bigg(\mathbb{E}\Bigg[\int_0^T (f_s^T - \tilde{f}_s^T)\beta_s\sigma \, dW_s \\ + \int_0^T (f_s^T - \tilde{f}_s^T)\beta_s\mu_s - \frac{1}{2}(f_s^T - \tilde{f}_s^T)^T\beta_s(\sigma\sigma)^T(f_s^T - \tilde{f}_s^T)\beta_s \, ds\Bigg]\Bigg) \quad .$$

But we are for now interested in the case of power utility where a little more work is necessary since it is not as easy to extract x_0 after applying Ito's

Lemma:

$$\begin{aligned}
&\mathbb{E}\Big[\frac{1}{\kappa}(X_T^*)^\kappa\Big] \\
&= \frac{x_0^\kappa}{\kappa} + \mathbb{E}\Bigg[\int_0^T (X_s^*)^\kappa f_s^T \beta_s \sigma \; dW_s + \int_0^T (X_s^*)^\kappa f_s^T \beta_s \mu_s \; ds \\
&\qquad + \frac{1}{2}(\kappa-1)\int_0^T (X_s^*)^\kappa f_s^T \beta_s (\sigma\sigma^T) f_s \beta_s \; ds\Bigg] \quad .
\end{aligned}$$

x_0 is in this case still present in the terms $(X_s^*)^\kappa$. To enable extraction, we use the representation of $(X_s^*)^\kappa$ from Theorem 3.12 leading to

$$\begin{aligned}
&\frac{x_0^\kappa}{\kappa} + \mathbb{E}\Bigg[\int_0^T \left(\frac{x_0(\beta_T Z_T)^{\frac{1}{\kappa-1}}}{\mathbb{E}\Big[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}\Big]}\right)^\kappa f_s^T \beta_s \sigma \; dW_s \\
&\quad + \int_0^T \left(\frac{x_0(\beta_T Z_T)^{\frac{1}{\kappa-1}}}{\mathbb{E}\Big[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}\Big]}\right)^\kappa f_s^T \beta_s \mu_s \; ds \\
&\quad + \frac{1}{2}(\kappa-1)\int_0^T \left(\frac{x_0(\beta_T Z_T)^{\frac{1}{\kappa-1}}}{\mathbb{E}\Big[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}\Big]}\right)^\kappa f_s^T \beta_s (\sigma\sigma^T) f_s \beta_s \; ds\Bigg] \quad .
\end{aligned}$$

In this term we can now isolate x_0 and perform the desired comparison.

$$\begin{aligned}
&\tilde{x}_0^\kappa = x_0^\kappa \cdot \\
&\mathbb{E}\Bigg[\frac{1}{\kappa} + \left(\frac{(\beta_T Z_T)^{\frac{1}{\kappa-1}}}{\mathbb{E}\Big[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}\Big]}\right)^\kappa \cdot \\
&\quad \left(\int_0^T f_s^T \beta_s \sigma \; dW_s + \int_0^T f_s^T \beta_s \mu_s \; ds + \frac{1}{2}(\kappa-1)\int_0^T f_s^T \beta_s (\sigma\sigma^T) f_s \beta_s \; ds\right)\Bigg] \cdot \\
&\mathbb{E}\Bigg[\frac{1}{\kappa} + \left(\frac{(\beta_T Z_T)^{\frac{1}{\kappa-1}}}{\mathbb{E}\Big[(\beta_T Z_T)^{\frac{\kappa}{\kappa-1}}\Big]}\right)^\kappa \cdot \\
&\quad \left(\int_0^T \tilde{f}_s^T \beta_s \sigma \; dW_s + \int_0^T \tilde{f}_s^T \beta_s \mu_s \; ds + \frac{1}{2}(\kappa-1)\int_0^T \tilde{f}_s^T \beta_s (\sigma\sigma^T) \tilde{f}_s \beta_s \; ds\right)\Bigg]^{-1}
\end{aligned}$$

The loss The loss due to the approximation of the trading strategy is very small, not even distinguishable from its standard deviation as can be seen in Figure 5.9 and Table 5.3 which were simulated using (MP1) and one hundred runs.

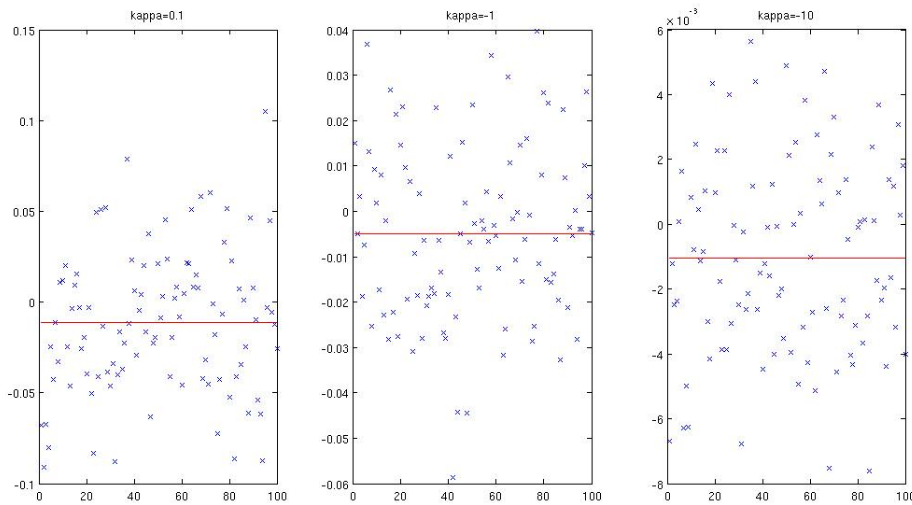

Figure 5.9: Loss due to the approximation of the trading strategy

loss	$\kappa = 0.1$	$\kappa = -1$	$\kappa = -10$
mean	-0.0114	-0.0049	-0.001
std	0.0394	0.0190	0.0030

Table 5.3: Loss due to the approximation of the trading strategy with (MP1)

In the Gaussian case the loss due to the approximation of the trading strategy is very small, but we detect a much higher standard deviation in this model: Figure 5.10 and Table 5.4. The outlier is the result of run 49 and increases the loss to size 10^6, the scales belonging to the other two utility functions are much smaller.

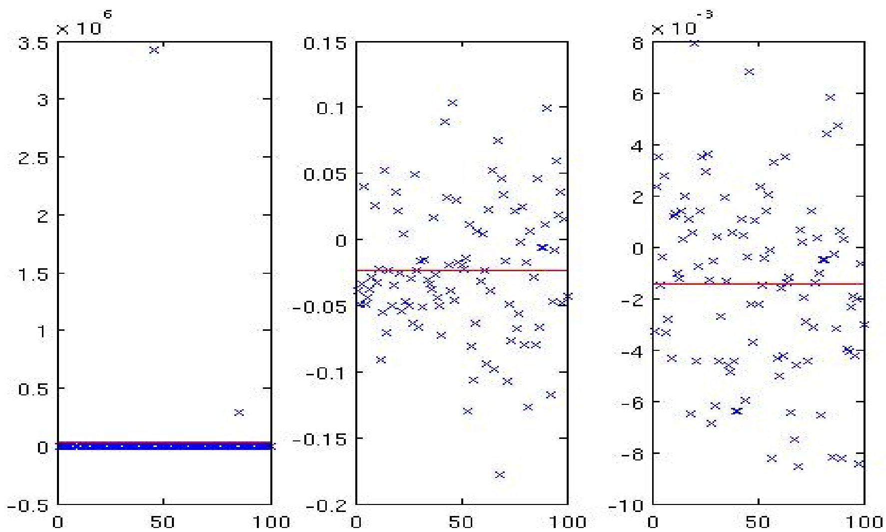

Figure 5.10: Loss due to the approximation of the trading strategy

loss	$\kappa = 0.1$	$\kappa = -1$	$\kappa = -10$
mean	3.73 10^6	n.a.	n.a.
median	0.55	-0.03	-0.001
std	3.44 10^5	n.a.	n.a.

Table 5.4: Loss due to the approximation of the trading strategy with (GP1)

5.6 Shortselling and Borrowing Constraints in the one-dimensional Model

We will now discuss the effect of shortselling and borrowing constraints on the loss. For a onedimensional model it suffices to limit the optimal fractions of wealth invested in a stock to the boundaries $[0,1]$ to exclude shortselling and borrowing. E.g. in the Markov model for logarithmic utility $f_t^* = \frac{\mu_t - r}{\sigma^2}$ has to be limited to $[0,1]$.

5.2 Remark. *Depending on the matrix A, shortselling constraints might not be influential. If its values are small enough, then no shortselling can occur and constraints are unnecessary. Borrowing cannot be prevented by manipulating A though, the matrix is positive definite and does not influence the sign of the trading strategy.*

Numerical results We will be using (MP1), but only the first stock, and compare the constrained investor to the unconstrained in Table 5.5.

		$\mathbb{E}[u(X_T^*)]$	$\mathbb{E}[u(X_T^{*,nss})]$	$\mathbb{E}[u(\hat{X}_T^*)]$	$\mathbb{E}[u(\hat{X}_T^{*,nss})]$
log	mean	5.02	0.35	0.54	0.1
	std	0.41	0.02	0.11	0.02
power $\kappa = 0.1$	mean	16.58	10.36	10.6	10.08
	std	0	0.024	1.03	0.16
power $\kappa = -1$	mean	-0.07	-0.72	-1	-1
	std	0.02	0.014	0.032	0.001
power $\kappa = -10$	mean	-0.001	-0.01	-0.1	-0.1
	std	0.001	0.002	0.0007	0.0007

Table 5.5: Optimal utility for the constrained and unconstrained investor with (MP1)

The results for partial information with power utility and $\kappa = -1$ or $\kappa = -10$ indicate that there is nearly no investment in stocks, up to the standard deviation. We may expect high losses thus, losses are displayed in Table 5.6.

		loss full info	loss partial info
log	mean	210.21	0.38
power $\kappa = 0.1$	mean	840.43	2.5
power $\kappa = -1$	mean	0.26	1.17
power $\kappa = -10$	mean	-0.004	0.0000

Table 5.6: Loss due to no shortselling/borrowing with (MP1)

Short selling is more attractive depending on the parameters: with (MP2), the constrained investor is much more successful in comparison as can be seen in Table 5.7.

		$\mathbb{E}[u(X_T^*)]$	$\mathbb{E}[u(X_T^{*,nss})]$	$\mathbb{E}[u(\hat{X}_T^*)]$	$\mathbb{E}[u(\hat{X}_T^{*,nss})]$
log	mean	0.09	0.05	0.02	0.02
	std	0.04	0.21	0.021	0.02
power $\kappa = 0.1$	mean	10.1	10.05	10	9.99
	std	0.05	0.2	0.28	0.22
power $\kappa = -1$	mean	-0.96	-0.97	-1	-1
	std	0.02	0.013	0.001	0.001
power $\kappa = -10$	mean	-0.09	-0.1	-0.1	-0.1
	std	0.004	0.003	0.0002	0.0002

Table 5.7: Optimal utility for the constrained and unconstrained investor with (MP2)

We also include table 5.8 displaying the loss. As expected, these values are very small especially in comparison to the ones in Table 5.6.

		loss full info	loss partial info
log	mean	0.075	0.0000
	std	0.03	0.0000
power $\kappa = 0.1$	mean	0.06	0.1
	std	0.06	0.39
power $\kappa = -1$	mean	0.009	-0.006
	std	0.02	0.15
power $\kappa = -10$	mean	0.002	0.002
	std	0.005	0.02

Table 5.8: Loss due to no shortselling/borrowing with (MP2)

Based on the same method, we also provide the numbers for (GP1) in Table 5.9.

		$\mathbb{E}[u(X_T^*)]$	$\mathbb{E}[u(X_T^{*,nss})]$	$\mathbb{E}[u(\hat{X}_T^*)]$	$\mathbb{E}[u(\hat{X}_T^{*,nss})]$
log	mean	4.76	0.3	0.6	0.09
	std	1.83	0.05	0.13	0.02
power $\kappa = 0.1$	mean	NaN	10.32	10.2	10.08
	std	NaN	0.05	2.62	0.19
power $\kappa = -1$	mean	-0.4	-0.8	0.0003	-0.95
	std	0.05	0.03	0.003	0.20
power $\kappa = -10$	mean	-0.03	-0.07	$-8.99 \cdot 10^{12}$	-0.12
	std	0.06	0.006	$-8.99 \cdot 10^{13}$	0.10

Table 5.9: Optimal utility for the constrained and unconstrained investor with (GP1)

These results are again accompanied by the results for the loss in Table 5.10.

		loss full info	loss partial info
log	mean	252.01	0.78
	std	222.48	0.22
power $\kappa = 0.1$	mean	$2.31 \cdot 10^{10}$	12.69
	std	$2.28 \cdot 10^{11}$	59.09
power $\kappa = -1$	mean	-1.13	-0.02
	std	5.69	0.14
power $\kappa = -10$	mean	-0.05	0.0000
	std	0.11	0.0000

Table 5.10: Loss due to no shortselling/borrowing with (MP2)

5.3 Remark. *With the parameter sets (MP1) and (GP1) big long and short positions occur. If we now compute the terminal wealth using the trading strategy instead of following Theorem 3.12, the results are thus much more volatile. Therefore the unconstrained investor goes bankrupt much more often and in the case of power utility with* $\kappa = 0.1$ *and (GP1) even in every experiment. Not allowing for shortselling and borrowing excludes these positions and leads to stable results. Thus even if the value of the portfolio is much lower with these strategies, bankruptcy does not occur and these strategies are worth a consideration.*

Choice of parameters With (MP1), the constraints on the optimal trading strategy in partial information came to effect in every run, changing half of all the values in these runs. Now with the smaller values in the parameter set (MP2), the trading strategy lies inside the boundaries much more often, only two thirds of the runs contained changes at all. In these runs, only one fifth of the values were limited. This leads to similar results for the constrained and unconstrained investor, but effected by the randomness in the equation for the terminal wealth. Thus the constrained investor might sometimes even outperform the unrestricted one numerically.
The question arises which choice of parameters results in limitations for the constrained investor and which does not. The following Lemma describes such a choice for the full information case in the Markov model, similar considerations lead to results for the Gaussian model and partial information in

both models.

5.4 Lemma. *For $n = 1$ and in the case of full information, no limitations occur if the entries in the state matrix are not smaller than the interest rate and no greater than σ^2 plus interest rate for logarithmic utility or $(1-\kappa)\sigma^2$ plus interest rate for power utility respectively.*

Proof. The optimal fractions are given by

$$f^*_{t,\text{power}} = \frac{BY_t - r}{(1-\kappa)\sigma^2}$$
$$f^*_{t,\log} = \frac{BY_t - r}{\sigma^2} \quad .$$

No limitations occur if these are inside the interval $[0, 1]$. The Markov chain Y_t acts on the unit vectors and these inequalitys follow

$$0 \ \le\ \frac{BY_t - r}{\sigma^2} \ \le\ 1$$
$$\Leftrightarrow 0 \ \le\ BY_t - r \ \le\ \sigma^2$$

and

$$0 \ \le\ \frac{BY_t - r}{(1-\kappa)\sigma^2} \ \le\ 1$$
$$\Leftrightarrow 0 \ \le\ BY_t - r \ \le\ (1-\kappa)\sigma^2 \quad .$$

Adding the interest rate in both inequalitys leads to the claim. □

These bounds present a strong constraint on the choice of parameters, so we will usually see an effect of the no shortselling or borrowing rule.

5.5 Example. *Let us consider (MP2) with*

$$B = \begin{pmatrix} 0.2 & 0.05 & -0.1 \end{pmatrix},$$

$r = 0$ *and* $\sigma = 0.25$. *The condition is then fulfilled for the second state only leading to changes due to the constraints for the other two states for logarithmic utility. For power utility, no effects occur in the first states if* $\kappa < -2.2$ *while the third state still violates the no borrowing constraint.*

Trading infrequently with constraints Trading in the Markov model leads to large long and short positions with parameter set (MP1) and is therefore risky when trading infrequently. Therefore the constrained strategy can perform better in this situation, but it always will be outperformed by the unconstrained investor when increasing the number of tradings. In Figure 5.11, we can observe this increase in the loss with an increase in the frequency for full information. For partial information only higher frequencies are displayed since the filter is not bounded and preserving positivity otherwise. That does only influence the data for the unconstrained case though, the constrained strategy still leads to results, but a comparison via the loss is not possible.

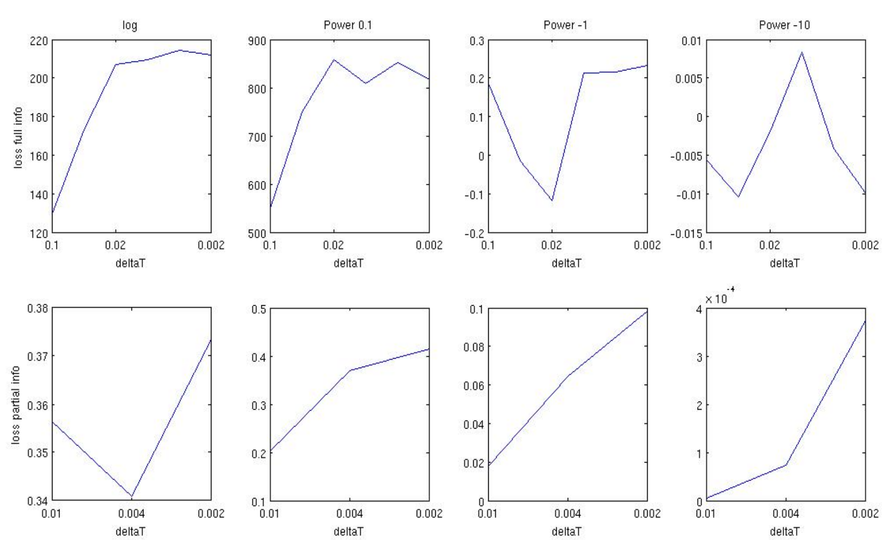

Figure 5.11: Loss due to constraints, (MP1)

Again the computation was also done in the Gaussian model leading to a similar picture 5.12 but with the usual higher deviation. Due to this not

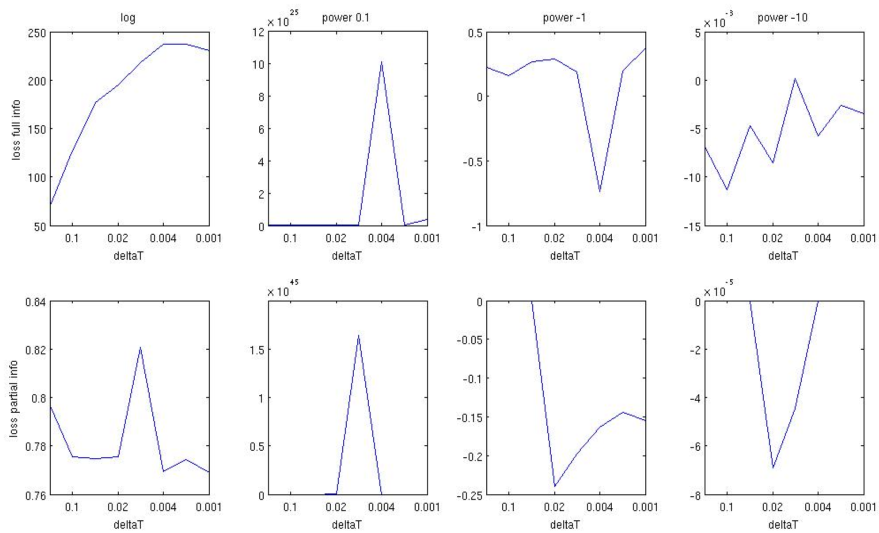

Figure 5.12: Loss due to constraints, (GP1)

possible comparison via the loss, we include tables containing the utility of the expected terminal wealth for low frequencies. Tables 5.11, 5.12, 5.13 and 5.14 show the results for (MP1), Tables 5.15, 5.16, 5.17 and 5.18 for (GP1).

log	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	3.2513	4.6711	5.7242	5.6331
	std	0.1377	0.2239	0.2844	0.3234
$\mathbb{E}[X_{T_{nss}}^*]$	mean	0.2527	0.2980	0.3298	0.3403
	std	0.0240	0.0220	0.0201	0.0196
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	0.1626	1.1447	3.2152	0.1301
	std	0.0751	0.0940	0.1861	0.0206

Table 5.11: Trading infrequently with constraints, log utility,(MP1)

power $\kappa = 0.1$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	14.2931	16.8751	19.3762	18.9980
	std	0.2391	0.3742	0.6392	0.8032
$\mathbb{E}[X_{T_{nss}}^*]$	mean	10.2599	10.3066	10.3396	10.3447
	std	0.0250	0.0232	0.0249	0.0241
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	10.1947	10.2375	NaN	10.1301
	std	0.2205	0.7639	NaN	0.0271

Table 5.12: Trading infrequently with constraints, power utility 0.1,(MP1)

power $\kappa = -1$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	-0.4091	-0.1092	-0.1026	-0.0850
	std	2.0517	0.2069	0.3039	0.1325
$\mathbb{E}[X_{T_{nss}}^*]$	mean	-0.8018	-0.7640640	-0.7392	-0.7295
	std	0.0148	0.0167	0.0150	0.0133
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	-1.5370	-0.3890	-1.4039	-0.8984
	std	7.6730	2.8229	9.8097	0.0228

Table 5.13: Trading infrequently with constraints, power -1,(MP1)

power $\kappa = -10$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	-0.0091	-0.0035	-0.0011	-0.0009
	std	0.0047	0.0043	0.0011	0.0008
$\mathbb{E}[X_{T_{nss}}^*]$	mean	-0.0360	-0.0214	-0.0137	-0.0114
	std	0.0042	0.0036	0.0028	0.0020
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	$-1.6355 \cdot 10^{16}$	$-1.1085 \cdot 10^{13}$	NaN	-0.0798
	std	$1.6352 \cdot 10^{17}$	$7.4362 \cdot 10^{13}$	NaN	0.0190

Table 5.14: Trading infrequently with constraints, power -10,(MP1)

log	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	2.9470	3.9724	4.4278	4.0522
	std	0.2901	0.4600	0.7098	0.7025
$\mathbb{E}[X_{T_{nss}}^*]$	mean	0.2291	0.2705	0.2918	0.3033
	std	0.0377	0.0431	0.0470	0.0467
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	0.0799	0.0833	0.0878	0.0859
	std	0.0213	0.0203	0.0192	0.0192

Table 5.15: Trading infrequently with constraints, log utility,(GP1)

power $\kappa = 0.1$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	14.4101	17.8518	24.1558	31.2979
	std	0.4855	1.3215	5.2302	44.4137
$\mathbb{E}[X_{T_{nss}}^*]$	mean	10.2358	10.2795	10.3110	10.3202
	std	0.0392	0.0446	0.0517	0.0507
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	10.0575	10.0543	10.0517	10.0523
	std	0.0205	0.0214	0.0240	0.0216

Table 5.16: Trading infrequently with constraints, power utility 0.1,(GP1)

power $\kappa = -1$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	-0.4523	-0.5229	-0.4081	-1.1473
	std	0.1639	1.0928	0.0577	7.5285
$\mathbb{E}[X_{T_{nss}}^*]$	mean	-0.8406	-0.8227	-0.8105	-0.8081
	std	0.0253	0.0271	0.0282	0.0291
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	-0.9613	-0.9643	-0.9687	-0.9729
	std	0.0158	0.0178	0.0190	0.0195

Table 5.17: Trading infrequently with constraints, power -1,(GP1)

power $\kappa = -10$	Δt	0.2	0.1	0.04	0.02
$\mathbb{E}[X_T^*]$	mean	-0.0331	-0.0307	-0.0296	-0.0304
	std	0.0067	0.0060	0.0061	0.0064
$\mathbb{E}[X_{T_{nss}}^*]$	mean	-0.0661	-0.0655	-0.0648	-0.0647
	std	0.0057	0.0074	0.0056	0.0058
$\mathbb{E}[\hat{X}_{T_{nss}}^*]$	mean	-0.1716	-0.2003	-0.2653	-0.2025
	std	0.1988	0.1859	0.3609	0.1277

Table 5.18: Trading infrequently with constraints, power -10,(GP1)

5.6 Remark. *These shortselling and borrowing constraints naturally hinder bankruptcy.*

5.7 Constraints in the Multidimensional Model

We are basically going to follow the onedimensional case, except for the computation of the optimal trading strategy in the constrained case. Here just limiting to the boundaries does not suffice anymore: The fraction of wealth is now a vector that has to consist of entries in $[0,1]$, but they also have to add up to a result smaller or equal to one. For logarithmic utility, [3] suggested the following method:

Logarithmic utility We have for logarithmic utility

$$J_f(x) = \mathbb{E}[log(X_T^f)] = log(x) + \mathbb{E}[\int_0^T (r + f_s^T(\hat{\mu}_s - r1) - \frac{1}{2} f_s^T \Sigma f_s ds)]$$

and

$$f_t^* = \Sigma^{-1}(\hat{\mu}_t - r1) \quad .$$

So to maximize the first equation without shortselling, we have to solve a quadratic optimization problem:

$$\begin{aligned} \max_a \quad & (a^T \hat{\mu}_t - \frac{1}{2} a^T \Sigma a) \\ \text{s.t.} \quad & a \in [0,1]^d \\ & a^T \mathbf{1} \leq 1 \end{aligned}$$

Transfer to power utility The basic problem is the same as before except for the nonvanishing stochastic integral:

$$\mathbb{E}\Big[\frac{(X_T^f)^\kappa}{\kappa}\Big] = \frac{x^\kappa}{\kappa}\mathbb{E}\Big[\frac{1}{\kappa}\exp\{\kappa \int_0^T r + f_s^T(\hat{\mu}_s - r1) - \frac{1}{2} f_s^T \Sigma f_s ds + \kappa \int_0^T f_s^T \sigma dV_s\}\Big]$$

where V_s is the innovations process with

$$dV_s = \sigma^{-1}(dZ_s - \hat{\mu}_s ds)$$

and $Z_s = \mu_s + \sigma W_s$. Remember that the innovations process (V_s) is an $\mathcal{F}_s$-Brownian motion under $\mathbb{P}$.

5.7 Proposition. *In the case of power utility with a constant drift we have to solve the optimization problem*

$$\begin{aligned} \max_{a} \quad & (\kappa a^T \hat{\mu}_t - \frac{1}{2}\kappa(1-\kappa)a^T \Sigma a) \\ s.t. \quad & a \in [0,1]^d \\ & a^T \mathbf{1} \leq 1 \quad . \end{aligned}$$

Proof. We need again to locate f^* that maximizes the basic problem. Another change of measure will take care of the nonvanishing stochastic integral.

$$\begin{aligned} & \frac{x^\kappa}{\kappa} \quad \mathbb{E}\Big[\frac{1}{\kappa}\exp\{\kappa \int_0^T r + f_s^T(\hat{\mu}_s - r1) - \frac{1}{2} f_s^T \Sigma f_s ds + \kappa \int_0^T f_s^T \sigma dV_s\}\Big] \\ = & \frac{x^\kappa}{\kappa} \quad \mathbb{E}\Big[\frac{1}{\kappa}\exp\{\kappa \int_0^T f_s^T \sigma dV_s - \frac{1}{2}\kappa^2 \int_0^T f_s^T \Sigma f_s ds \\ & + \kappa \int_0^T r + f_s^T(\hat{\mu}_s - r1) - \frac{1}{2} f_s^T \Sigma f_s ds + \frac{1}{2}\kappa^2 \int_0^T f_s^T \Sigma f_s ds\}\Big] \\ = & \frac{x^\kappa}{\kappa} \quad \mathbb{E}\Big[\frac{1}{\kappa}\exp\{\kappa \int_0^T f_s^T \sigma dV_s - \frac{1}{2}\kappa^2 \int_0^T f_s^T \Sigma f_s ds\} \\ & \exp\{\kappa \int_0^T r + f_s^T(\hat{\mu}_s - r1) - \frac{1}{2} f_s^T \Sigma f_s ds + \frac{1}{2}\kappa^2 \int_0^T f_s^T \Sigma f_s ds)\}\Big] \end{aligned}$$

Now we first have to show, that the first exponential term is a martingale to proceed with the change of measure:
Let $X_t := \kappa \int_0^t f_s^T \sigma dV_s$, the Doléans exponential of X_t is then given by

$$\varepsilon(X)_t = \exp\{\kappa \int_0^t f_s^T \sigma dV_s - \frac{1}{2}\kappa^2 \int_0^T f_s^T \Sigma f_s ds\}$$

which is the process we are examining. For that to be a martingale we have to verify Novikov's condition, i.e.

$$\mathbb{E}\Big[\exp\{\frac{1}{2}\kappa^2 \int_0^t f_s^T \Sigma f_s ds\}\Big] < \infty \quad .$$

But we know already by definition that

$$\int_0^t \|f_s\|^2 ds < \infty \quad .$$

The expected value of this exponential is then equal to one since it is a supermartingale by Fatou's Lemma.
Now we can perform the change of measure following Girsanov's Theorem to a new measure $\mathbb{P}_\kappa$ with the martingale $\varepsilon(X)_T$ and the Radon-Nikodym derivative

$$\frac{d\mathbb{P}_\kappa}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \varepsilon(X)_T$$

In this new measure, we now have the optimization problem of maximizing

$$\mathbb{E}\Big[\frac{(X_T^f)^\kappa}{\kappa}\Big] = \frac{x^\kappa}{\kappa}\mathbb{E}_\kappa\Big[\frac{1}{\kappa}\exp\{\int_0^T \kappa((r + f_s^T(\hat{\mu}_s - r1)) - \frac{1}{2}\kappa(1-\kappa)f_s^T\Sigma f_s ds\}\Big]$$

which we can then reduce as we also did in the logarithmic case:

$$\begin{aligned} \max_a \quad & (\kappa a^T\hat{\mu}_t - \frac{1}{2}\kappa(1-\kappa)a^T\Sigma a) \\ \text{s.t.} \quad & a \in [0,1]^d \\ & a^T\mathbf{1} \leq 1 \end{aligned}$$

□

If we now return to our stochastic drift and use the above result despite the condition of a constant drift, optimality is not given anymore.

Numerical results We will now display the results of the computations corresponding to the two methods we introduced in the previous paragraphs, first for (MP1) in Table 5.19.

		$\mathbb{E}[u(X_T^*)]$	$\mathbb{E}[u(X_T^{*,nss})]$	$\mathbb{E}[u(\hat{X}_T^*)]$	$\mathbb{E}[u(\hat{X}_T^{*,nss})]$
log	mean	5.75	0.36	0.77	0.11
	std	0.33	0.02	0.08	0.02
power $\kappa = 0.1$	mean	20.74	10.36	10.56	10.11
	std	9.49	0	0.06	0.02
power $\kappa = -1$	mean	-0.05	-1.33	-1	-1
	std	0.96	0.02	1.2	0
power $\kappa = -10$	mean	-0.0004	-3.64	-0.07	-0.1
	std	0.96	0.75	0.01	0

Table 5.19: Optimal terminal wealth with (MP1) for the constrained and unconstrained investor

It can clearly be seen, that the investor who is not allowed to short sell and borrow has a disadvantage compared to the one who does in the case of full information. This is in line with the results in [3] and we also have a gap between the infrequent trader who is not allowed to short sell and borrow and the continuous investor who is.
This is also reflected in the loss as displayed in 5.20.

		loss full info	loss partial info
log	mean	537.97	0.56
	std	184.45	0.14
power $\kappa = 0.1$	mean	4185.8	-0.25
	std	1921.7	0.28
power $\kappa = -1$	mean	0.18	0.001
	std	0.7	0.06
power $\kappa = -10$	mean	0.06	0.0000
	std	0.07	0.0000

Table 5.20: Loss due to no shortselling/borrowing (MP1)

Next we will turn to our second model and compare the results for (GP1) in Table 5.21. We will observe very high standard deviations due to the use of the trading strategy which has even more effect in this model.

		$\mathbb{E}[u(X_T^*)]$	$\mathbb{E}[u(X_T^{*,nss})]$	$\mathbb{E}[u(\hat{X}_T^*)]$	$\mathbb{E}[u(\hat{X}_T^{*,nss})]$
log	mean	37.5	0.5	30.07	0.36
	std	6.43	0.05	3.41	0.03
power $\kappa = 0.1$	mean	NaN	10.53	NaN	9.76
	std	NaN	0.05	NaN	0.06
power $\kappa = -1$	mean	-0.04	-2.01	-0.001	-1
	std	0.018	0.14	0.027	0
power $\kappa = -10$	mean	-0.0019	$-6.64 \cdot 10^8$	$-2.55 \cdot 10^8$	-0.1
	std	0.002	$5.23 \cdot 10^9$	$2.55 \cdot 10^9$	0

Table 5.21: Optimal terminal wealth with (GP1) for the constrained and unconstrained investor

We also computed the loss with the results in Table 5.22.

		loss full info	loss partial info
log	mean	$3.14 \cdot 10^{28}$	$4.31 \cdot 10^{15}$
	std	$3.14 \cdot 10^{29}$	$2.07 \cdot 10^{16}$
power $\kappa = 0.1$	mean	$1.07 \cdot 10^{21}$	NaN
	std	$6.01 \cdot 10^{21}$	NaN
power $\kappa = -1$	mean	-0.66	NaN
	std	2.71	NaN
power $\kappa = -10$	mean	0.001	NaN
	std	0.008	NaN

Table 5.22: Loss due to no shortselling/borrowing (GP1)

6 Conclusions

We will now shortly recapitulate the central topics of this work.

When do we filter? If the expected value of the filter is unbounded, there will be no information to be gained from it, filtering is then of no avail. For the HMM filter, this means

$$\left\|\tilde{\mathbb{E}}[\rho_k]\right\| < M$$

is only given if

$$\left\| \max_j \left(\exp\left\{ A_{jj} \Delta t \ k \right\} \right) \right\| \cdot \left\| \operatorname{Id} + \Delta t Q^T \right\|^k < \frac{M}{\|\nu\|} \quad .$$

For a diagonalizable rate matrix Q the frequency Δt for which $\left\|(\operatorname{Id} + \Delta t Q^T)^k\right\|$ is bounded can be computed nicely as

$$\|Id + \Delta t Q^T_{diag}\| \leq 1 \quad .$$

In this model positivity is another criterion that has to be met since the Markov chain operates on the unit vectors and is positive at all times.

$$(\operatorname{Id} + \Delta t \ Q^T)$$

needs to be positive for the scheme to preserve positivity.
A similar bound was given for the Kalman filter for the drift following linear Gaussian dynamics: the expected value of the filter is bounded if the matrix α is diagonalizable and positive semidefinite, as well as the matrix $\alpha + \gamma(\sigma\sigma^T)^{-1}$. In addition, γ_t needs to approach a constant γ_∞.

Why do we filter? The partially informed investor generally achieves a lower expected utility of the optimal terminal wealth in comparison to the fully informed investor and a higher one in comparison to the investor without information. The following bounds allow a closer observation of the advantage in comparison to the investor without information.
A lower bound for the expected utility of the optimal terminal wealth for logarithmic utility was computed to be

$$\mathbb{E}[\log X_T^*] \geq \frac{T}{2} \left(\sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{max}(A) \right) + \frac{\lambda_{max}(A)}{2} \int_0^T \mathbb{E}[\eta_t^T \eta_t] \ dt \quad .$$

This bound can though only contribute information if

$$\frac{-\sum_{i=1}^{d} \nu_i A_{ii} - \nu^T A \nu}{\lambda_{max}(A)} + 1 \;<\; \mathbb{E}[\eta_T^T \eta_T] \quad .$$

The corresponding upper bound is then given by

$$\mathbb{E}[\log X_T^*] \leq \frac{1}{2} \int_0^T \sum_{i=1}^{d} \nu_i A_{ii} - \lambda_{min}(A)(1 - \mathbb{E}[\eta_t^T \eta_t]) \; dt \quad .$$

This bound works best for equally large eigenvalues, another bound was introduced that is more precise the smaller the eigenvalues are except for one large one:

$$\mathbb{E}[\log(X_T^*)] \geq \frac{1}{2} \int_0^T \sum_{i=0}^{d} \nu_i A_{ii} - \mathbb{E}[\lambda_{max}(S_{max}^T (Y_t - \eta_t))^2] \; dt$$

where the diagonalization of A is given by SDS^T, λ_{max} is the maximal eigenvalue of the matrix A and S_{max}^T is the corresponding eigenvector.

Where do we use the filter? The filter and the corresponding expected utility optimal terminal wealth were introduced in a model with n stocks and d states. In case there are more stocks than states, it is more efficient to filter and invest in a model with d Mutual Funds in the HMM. It was shown that this market with d Mutual Funds leads to the same expected utility of the optimal terminal wealth and the filters in this market and the original one coincide.

What is the quality of filtering at different market parameters? At infrequent trading, the above mentioned criteria for boundedness and positivity have to be met. The trading frequency for which this is given, can be computed explicitly and for more frequent trading, filtering is of good quality.
For an increasing number of stocks, the mean of the process ψ changes and has to be observed with regard to boundedness. If the mean is bounded, then filtering is again of good quality and the market can be reduced to the Mutual Fund version.
The quality of the filter does also depend on the investor's choices: The

filter contains no information if it is close to the stationary distribution and leads to a bad optimization result. If the filter is unbounded, he faces bad optimization results again, but near the border to unboundedness, the best result can be reached, although with high risk. The dilemma is thus to maximize the expected utility of optimal terminal wealth while ensuring good filter quality.

What loss is to be expected at these different market parameters?
For the above mentioned scenarios infrequent trading and increasing number of stocks, the loss is negligible if the filter quality is good. For an increasing number of stocks, it naturally depends on the quality of the added stocks.
In the HMM, we also observed the loss for different numbers of states with the result that assuming less states than were orginially present in the model is punished with big losses, whereas assuming more is not.
In the case of partial information with power utility, the optimal trading strategy is of complex appearance. Simplifying it appropriately leads to nearly the optimal result, only very small losses occur.
Under shortselling and borrowing constraints, the losses depend first and foremost on the parameters. If the parameters are such that big long and short positions occur, then constraints lead to big losses. Constraining can on the other hand lead to an advantage if without constraints strategies would be so extreme that with a high probability they lead to bankruptcy.

There exists an essential connection between filtering and portfolio optimization and the partially informed investor gains with advantageous filters in all observed market situations compared to the investor not using information.

References

[1] A. Bain and D. Crisan. *Fundamentals of Stochastic Filtering*, volume 3. Springer, 2009.

[2] N. Bäuerle and U. Rieder. Portfolio optimization with Markov-modulated stock prices and interest rates. *Automatic Control, IEEE Transactions on*, 49(3):442–447, 2004.

[3] N. Bäuerle, S.P. Urban, and L.A.M. Veraart. The relaxed investor with partial information. *SIAM Journal on Financial Mathematics*, 3(1):304–327, 2012.

[4] S. Brendle. Portfolio selection under partial observation and constant relative risk aversion. *Available at SSRN 581921*, 2004.

[5] S. Brendle. Portfolio selection under incomplete information. *Stochastic processes and their Applications*, 116(5):701–723, 2006.

[6] J.M.C. Clark. The design of robust approximations to the stochastic differential equations of nonlinear filtering. *Communication systems and random process theory*, 25:721–734, 1978.

[7] J. Cvitanić and I. Karatzas. Convex duality in constrained portfolio optimization. *The Annals of Applied Probability*, pages 767–818, 1992.

[8] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous time stochastic processes via the EM algorithm. *Stochastic Processes and their Applications*, 23(1):91–113, 1986.

[9] R.J. Elliott. *Stochastic Calculus and Applications*, volume 18. Springer New York, 1982.

[10] R.J. Elliott. New finite-dimensional filters and smoothers for noisily observed Markov chains. *Information Theory, IEEE Transactions on*, 39(1):265–271, 1993.

[11] R.J. Elliott and V. Krishnamurthy. New finite-dimensional filters for parameter estimation of discrete-time linear Gaussian models. *Automatic Control, IEEE Transactions on*, 44(5):938–951, 1999.

[12] A. Gandy and L.A.M. Veraart. The effect of estimation in high-dimensional portfolios. *Mathematical Finance*, 2012.

[13] U.G. Haussmann. On the integral representation of functionals of Itô processes. *Stochastics*, 3(1-4):17–27, 1980.

[14] M.R. James, V. Krishnamurthy, and F. Le Gland. Time discretization of continuous-time filters and smoothers for HMM parameter estimation. *Information Theory, IEEE Transactions on*, 42(2):593–605, 1996.

[15] I. Karatzas, D. L. Ocone, and J. Li. An extension of Clark's formula. *Stochastics: An International Journal of Probability and Stochastic Processes*, 37(3):127–131, 1991.

[16] V. Kučera. A review of the matrix Riccati equation. *Kybernetika*, 9(1):42–61, 1973.

[17] P. Lakner. Utility maximization with partial information. *Stochastic processes and their applications*, 56(2):247–273, 1995.

[18] P. Lakner. Optimal trading strategy for an investor: the case of partial information. *Stochastic Processes and their Applications*, 76(1):77–97, 1998.

[19] E. Leoff, L. Ruderer, and J. Sass. Convergence of filters and mutual funds for markov modulated drift. *Work in Progress*, 2015.

[20] R.S. Lipster, A.B. Aries, and A.N. Shiryayev. *Statistics of Random Processes II: Applications*. Springer-Verlag, 1974.

[21] U. Rieder and N. Bäuerle. Portfolio optimization with unobservable Markov-modulated drift process. *Journal of Applied Probability*, 42(2):362–378, 2005.

[22] L.C.G. Rogers. The relaxed investor and parameter uncertainty. *Finance and Stochastics*, 5(2):131–154, 2001.

[23] J. Sass. Utility maximization with convex constraints and partial information. *Acta Applicandae Mathematicae*, 97(1-3):221–238, 2007.

[24] J. Sass and U.G. Haussmann. Optimizing the terminal wealth under partial information: The drift process as a continuous time Markov chain. *Finance and Stochastics*, 8(4):553–577, 2004.

[25] W. Schachermayer, M. Sîrbu, and E. Taflin. In which financial markets do mutual fund theorems hold true? *Finance and Stochastics*, 13(1):49–77, 2009.

[26] I. Shigekawa. Derivatives of Wiener functionals and absolute continuity of induced measures. *Kyoto Journal of Mathematics*, 20(2):263–289, 1980.

Leonie Ruderer

(geb. Karbach)

Akademischer Werdegang

✉ leonie.ruderer@gmail.com

Abschlüsse

06/2010 **Master of Science in Engineering (Engineering Mathematics)**, *Lunds Tekniska Högskola (LTH)*, Schweden.

12/2009 **Diplom in Mathematik mit Nebenfach Wirtschaftswissenschaften**, *Technische Universität*, Kaiserslautern.

03/2005 **Allgemeine Hochschulreife**, *Kant-Gymansium*, Boppard.

Ausbildung

10/2010-heute **Promotion in Mathematik im Bereich Finanzmathematik**, *TU Kaiserslautern*.

08/2007-06/2008 **Studium im Doppelabschlussprogramm**, *Lunds Tekniska Högskola (LTH)*, Schweden.

03/2005-12/2009 **Studium der Mathematik mit Nebenfach Wirtschaftswissenschaften**, *Technische Universität*, Kaiserslautern.

09/1996-03/2005 **Gymnasiast**, *Kant-Gymnasium*, Boppard.

Arbeiten

2015 **Convergence of filters and mutual funds for Markov modulated drift**, *Work in Progress*, E. Leoff, L. Ruderer and J. Sass.

Leonie Ruderer

(formerly Karbach)

Academic History

✉ leonie.ruderer@gmail.com

Education

10/2010-today **Graduate Studies in financial mathematics**, *TU Kaiserslautern*.

08/2007-06/2008 **Participant in Double Degree Program**, *Lunds Tekniska Högskola (LTH)*, Sweden.

03/2005-12/2009 **Undergraduate Studies: Mathematics and Economy**, *Technische Universität*, Kaiserslautern.

Degrees

06/2010 **Master of Science in Engineering (Engineering Mathematics)**, *Lunds Tekniska Högskola (LTH)*, Schweden.

12/2009 **Diploma in Mathematics and Business Sciences**, *Technische Universität*, Kaiserslautern.

03/2005 **Allgemeine Hochschulreife**, *Kant-Gymansium*, Boppard.

Publications

2015 **Convergence of filters and mutual funds for Markov modulated drift**, *Work in Progress*, E. Leoff, L. Ruderer and J. Sass.