Fractional Brownian motion and its application in the Norwegian stock market

Ida Christine Gyldenløve

Supervisor: Jan Ubøe

Bergen, spring 2014

Master thesis in Economic Analysis

NORWEGIAN SCHOOL OF ECONOMICS

1 Abstract

In this thesis, I investigate the properties of fractional Brownian motion for use in the stock market, and I also look at what types of calculus that should be used when one works with it. I then use the calculus to see what would happen to the market if stock returns followed fractional Brownian motion.

The second part of the thesis consists of finding a method to estimate discretized fractional Brownian motion by using $ARFIMA$ models. I apply this theory to stocks in the Oslo Stock Exchange to look for long memory in the returns and volatility by estimating the Hurst coefficient. Would it be easy to make money by using this model?

I find that the fractional Brownian motion has several traits we appreciate when analysing stocks. However, the market would not be efficient if stock returns could be modelled by fractional Brownian motion, as it allows for arbitrage.

In the Norwegian stock market, I find that the main index shows some evidence of long memory in the returns. However, this is not much, as the Hurst coefficients estimated are quite close to $\frac{1}{2}$. This means that it is unlikely that one could make much money from trying to find arbitrage opportunities like this. The same is true for the analysed stocks, and the $ARFIMA$ model is not a perfect fit for any of them.

I do find evidence of long memory in the volatility of the stock returns, and this may be used to help understand and predict the risk of the stocks better.

2 Preface

This is a master thesis written as a part of a Master of Science in Economics and Business Administration at the Norwegian School of Economics (NHH). The election of fractional Brownian motion as a topic, reflects that my major is in Economic Analysis and my minor is Financial Economics.

I have chosen to write about something that lets me use a combination of mathematics and statistics as well as using this knowledge on the Norwegian stock market. Fractional Brownian motion is a topic that is not a part of the curriculum in any of the subjects at NHH, which makes it especially demanding to investigate. To work with a topic that is this advanced and rare, has probably been both the most challenging and the most exciting part of writing this thesis.

I would like to thank my supervisor, Jan Ubøe, for advice, feedback and motivation during the writing process. But also for getting me interested in regular Brownian motion from courses in statistics and mathematical statistics in the first place. I would also like to thank Jonas Andersson for great discussions about $ARFIMA$ models.

Contents

1 Abstract **3**

2 Preface **4**

3 Introduction **9**

4 Fractional Brownian motion **10**

4.1 History 10
4.2 Properties of fractional Brownian motion 12
4.3 Long-range dependence 13
4.4 Self-similarity 14
4.5 Continuity and path differentiability 15
4.6 Semimartingale 16

5 Calculus **17**

5.1 Wick Itô Skorohod 17
5.2 Pathwise forward integration 18
5.3 Wiener and divergence-type integrals 20
5.3.1 Wiener integrals for $H > \frac{1}{2}$ 20
5.3.2 Wiener integrals for $H < \frac{1}{2}$ 20
5.3.3 Divergence type integral for $H > \frac{1}{2}$ 21
5.3.4 Divergence type integral for $H < \frac{1}{2}$ 21
5.4 Fractional Wick Itô Skorohod for $H > \frac{1}{2}$ 21

6 Integration methods suited for finance **21**

6.1 Pathwise forward integrals 22
6.2 WIS integrals 22

7 Arbitrage with fBm **22**

7.1 What is arbitrage? 22
7.2 Why is arbitrage a problem? 23
7.3 Arbitrage for fBm when $\frac{1}{2} < H < 1$ 24

8 Simulation and estimation **26**
8.1 The Hoskin method 26
8.2 The Cholesky method 26
8.3 The Davies and Harte method 26
8.4 Approximated methods 27
8.5 Estimation of the Hurst coefficient 27
8.5.1 R/S analysis 27
8.5.2 Other methods 28

9 $ARFIMA$ models **28**
9.1 Discretizing the fBm 28
9.2 Fractional $ARIMA$ models 29
9.3 Different estimators 31
9.3.1 The Haslett and Raftery method 32
9.4 Long-term versus short-term dependence 33
9.5 Arbitrage in the $ARFIMA$ model 34
9.6 Identifying the correct model 36

10 Estimating models for stock price returns **38**
10.1 Data 38
10.2 Returns of the OSEBX 39
10.3 Squared returns of the OSEBX 44
10.4 Returns of other stocks 48
10.5 Squared returns of the stocks 55

11 Implications for investors **56**

12 Limitations and future work **57**

13 Concluding remarks **58**

14 Appendix **59**

List of Figures

1 Overview of the OSEBX 40
2 $ARFIMA(0, d, 0)$ for OSEBX 41
3 $ARFIMA(1, d, 0)$ for OSEBX 42
4 $ARFIMA(0, d, 1)$ for OSEBX 42
5 $ARFIMA(1, d, 1)$ for OSEBX 43
6 Diagnostics for OSEBX 43
7 Description of squared OSEBX 45
8 Diagnostics for squared OSEBX 47
9 Description of NAS 49
10 Autocorrelation of $ARFIMA(0, d, 1)$ for NAS 49
11 Diagnostics for $ARFIMA(0, d, 1)$ for NAS 50
12 Description of NAUR 51
13 Autocorrelation of $ARFIMA(1, d, 0)$ for NAUR 52
14 Diagnostics for $ARFIMA(1, d, 0)$ for NAUR 53
15 Autocorrelations for returns and squared returns for NAS and NAUR 55

List of Tables

1	The data used	39
2	Estimation of $ARFIMA$ models for OSEBX	42
3	Estimation of models for squared OSEBX	46
4	Estimation of models for NAS	50
5	Estimated models for NAUR	52
6	Estimated values of d for all chosen stocks	54
7	Estimated models for squared returns of NAUR	56
8	Estimated models for squared returns of NAS	56

3 Introduction

"It is possible to beat the market" , writes investor and editor Trygve Hegnar [21]. He is referring to an article about Warren Buffet's investment techniques [33]. The techniques are fairly simple and based on the relationships between easily accessible key numbers from companies, like the P/E ratio. Are there really such easy ways to make money in a market with almost an unliminted number of creative competitors?

While some today believe that one could beat the market by finding stocks with a favorable P/E ratio or a streak of high returns, it is not accepted in the scientific community that it is actually possible to do. The short explanation is that everyone would do it if it worked, and then the arbitrage opportunities would disappear. The same is the case with other so-called sure ways of beating the market.

Fractional Brownian motion has by some been thougth of as a revolutinary way of modelling stock returns, and hence finding that the returns have long memory. If this was the case with stock returns, one could use this to implement arbitrage strategies and beat the market. The technique would be different from the ones mentioned above, but the result would be the same if the strategies worked - someone would beat the market. It is however not commonly believed that there exists such ways of prediciting stock returns, and fractional Brownian motion is rarely discussed as a model for stock returns anymore.

I want to find an overview of the fractional Brownian motion, see if it has properties we find attractive for financial time series in the stock market, and try to apply the theory to stocks at the Oslo Stock Exchange. Is it possible to find stocks in the exchange with long memory, which could lead us to finding arbitrage opportunities?

I will start by introducing the history of fractional Brownian motion and look at its mathematical properties to understand the basic theory behind it. From this, I will explain why it is not possible to apply normal culculus to the fractional Brownian motion. Then I will investigate alternatives to regular calculus. I will also discuss which types of calculus that can be especially suited for finance.

The alternative calculus is needed when I discuss the reason why fractional Brownian motion is not used to analyse stock returns today, namely that the model implies that there exist arbitrage opportunities in the market.

After this I will shortly mention different ways fractional Brownian motion has been simulated and estimated traditionally to better understand how it is not an easy task to choose which methods to use.

In the last part of the paper, I will discuss the long memory model $ARFIMA$ and how it can be used to model fractional Brownian motion and the properties introduced in the first part of the thesis. The $ARFIMA$ model will be used in the application of the fractional Brownian motion in the Oslo Stock Exchange, where I will look for long memory in return and volatility for different stocks. I also shortly mention what implications the results could have for investors.

4 Fractional Brownian motion

4.1 History

The fractional Brownian motion, later also called fBm, has several fathers. One of them is Benoît Mandelbrot. He started asking questions about the basic theories of finance and how the market worked, as he could not see how the most commonly used financial theories could fit with the reality.

One of the characterisitcs of the market he questioned, was the large day to day changes. He mentioned as an example the summer of 1998 on Wall

Street. On august 4. the Dow Jones dropped by 3.5%. Three weeks later, it was down another 4.4%, and on august 31. the index dropped by 6.8% during the day. According to market theories which state that the daily changes are Gaussian, the chance of a drop that big is 1 to 20,000,000 [27]. Even if the markets traded daily for 100,000 years, a drop that big should not have happened [27]. Accordingly, the chance of having three big drops in august, was close to 1 in 500,000,000 [27]. Mandelbrot wished to find a model that could accomodate these big jumps.

There are several newer examples of this from the same market place as well. For example from the much known financial crisis during 2008. 29th of september started with a drop of 7.0% [36]. In addition to several smaller drops, there was one of 7.3% on 9.th of october, and one of 7.9% the 15th of october [36]. Then there also was a drop of 7.7% the 1st of december the same year [36]. There has been similar periods with jumps as well, but it only adds to the fact that these outcomes are extremely unlikely to happen under the Gaussian regime. So unlikely that the Gaussian assumptions should be wrong, and one needs to find another model that would allow these jumps to happen.

These jumps might have been caught by Lévy processes. An example of a well known Lévy process is regular Brownian motion combined with a Poisson process to account for the sudden jumps. However, another aspect that concerned Mandlebrot, was that the stock returns seemed to have patterns that repeated themselves in different scales, and that today's return could depend on returns from very long time ago [27]. There is no way to explain this sort of "memory" by using a Lévy process, as all the increments are independent.

But how did we end up with the Gaussian financial system in the first place? In the 1900, a french matematician called Louis Bachelier meant that price changes should be modelled as a random walk. He said that the prices had independent changes, and that these changes formed a bell shaped curve

[27]. This is some of the basic assumptions modern finance was built on. Later, we have the Efficient Market Hypothesis by Fama, the modern portfolio theory by Markovitz, the relationship between risk and reward by Sharpe and a formula for options pricing by Black, Scholes and Merton [27].

The point of finding a statistical model to describe the price changes in the financial market, is not to prove that the returns actually move randomly, but that they can be described as if they do, which means that such a model helps us [27]. Mandelbrot [27] meant that this way of looking at prices could give better ways of modelling them than we can get from both fundamental and technical analysis.

The fBm model was not originally devoted to finance when the first parts of the model took form. Harold Edwin Hurst tried to model the floods of the Nile when he discovered that the water movements did not seem to be independent. The sizes of the floods did not follow the Gaussian curve. Not only were the sizes of the individual floods different from the bell shapes curve, but the sequence they appeared in did not seem random [27].

There are two effects that are commonly mentioned in connection with the fBm. The first one is called the Noah effect. When Noah built the arc, it was because of the extreme amount of water that would fall down. This refers to the sizes of the jumps [27]. The other effect is called the Joseph effect. Joseph predicted that Egypt would have 7 fat years and 7 lean years [27]. This also had an extreme effect, but here it was the sequence of the lean and the fat years that made the outcome extreme. These two effects mix, and they can both have great impact on whatever phenomenon they are found in when they occur [27].

In today's fBm the Hurst coefficient is what tells us if the changes are positively or negatively correlated, and there has been made several attempts to estimate the Hurst coefficient for different stocks. Mandelbrot [27] was hoping that the coefficient would become a new yardstick in finance as much as the α or the β used in the CAPM. However, this has not been the case.

The patterns of fBm has been found several places in nature, art and types of science. Examples are the Sierpinski gasket, the Cantor dust, the Koch curve, clouds, branching of bronchia in human lungs, queing theory and so on [27]. But the model has not been the miracle Mandelbrot hoped it would become in finance. Is it at all possible to use the model in finance, or should we look for something different?

4.2 Properties of fractional Brownian motion

Definition 1 *H is called the Hurst constant and it belongs to the interval (0, 1). A fractional Brownian motion, which we denote by B_t^H, is a continuous and centered Gaussian process with covariance function*

$$\mathrm{E}[B_t^H B_s^H] = \tfrac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

And the increments $B_{t+h}^H - B_t^H$ and $B_{s+h}^H - B_s^H$ with $s + h \leq t$ and $t - s = nh$ have the covariance function

$$\rho_H(n) = \tfrac{1}{2}h^{2H}[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}].$$

When $H = \frac{1}{2}$, we see from the definition that the covariance becomes 0 in both cases. When that happens, the fractional Brownian motion behaves just like a standard Brownian motion with independent increments. Even when the increments are not independent, the fractional Brownian motion has some properties that are somewhat similiar to the properties of regular Brownian motion. Some of them are:

1. $B_0^H = 0$
2. $\mathrm{E}[B_t^H] = 0$ for all $t \geq 0$.
3. B^H has homogeneous increments, which means that $B_{(t+s)}^H - B_s^H$ has the same law of B_t^H for $s, t \geq 0$.

4. B^H is a Gaussian process and $\mathrm{E}[B_t^{2H}] = t^{2H}, t \geq 0, \forall H \in (0,1)$.

5. B^H has continuous trajectories.

From the first two bullets, we see that the fBm always starts at 0, and 0 is also our best prediction of where it will go, as the chance of it moving upwards is the same as the chance of it moving downwards. We have no information about the expected movement.

The information we have tells us something about the increments and their relative sizes. Two increments that have time intervals at the same length, are expected to be of the same size.

From the fourth point we see that the fBm is Gaussian, and that the expected value of the squared, i.e. the second moment, is not 0. This is an important point in both regular and fractional Brownian motion.

The last points tells us that the trajectories are continuous, but the fBm is nowhere differentiable [6].

When examining the fractional Brownian motion further, it is often both useful and necessary to distinguish between the cases where $H \in (0, \frac{1}{2})$ and $H \in (\frac{1}{2}, 1)$.

As shown earlier, the increments of the fBm are not independent when $H \neq \frac{1}{2}$. From definition 1 we see that the correlation is negative when $H \leq \frac{1}{2}$, and that it is positive when $H \geq \frac{1}{2}$. When it is positively correlated, a positive change is more likely to be followed by another positive change and vice versa. This makes the fBm with $H \geq \frac{1}{2}$ good at modelling series with memory and high persistence [6]. When a positive increment is more likely to be followed by a negative increment, the fBm can be used to model systems with antipersistence and intermittence [6].

4.3 Long-range dependence

Definition 2 *A stationary sequence $(X_n)_{n \in N}$ exhibits long-range dependence if the autocovariance functions $\rho(n) := cov(X_k, X_{k+n})$ satisfy*

$\lim_{n\to\infty} \frac{\rho(n)}{cn^{-\alpha}} = 1$ *for some constant* c *and* $\alpha \in (0,1)$. *In this case, the dependence between* X_k *and* X_{k+n} *decays slowly as* n *tends to infinity and*

$$\sum_{n=1}^{\infty} \rho(n) = \infty$$

There also exist several other definitions of long-range dependence which are used in different fields [6]. In short, the long-range dependence means that a series has a long term memory. The autocovariance function decays slowly because values far apart in the time series are highly correlated. For the fBm spesific we can show that for $H \geq \frac{1}{2}$. The covariance between the increments

$$\rho_H(n) = \frac{1}{2} h^{2H}[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}]$$

converges to $H(2H-1)n^{2H-2}$ as n approaches ∞ [6]. This could be used to our advantage, and if we use $H(2H-1)$ as the constant c, and similiarly use $2-2H$ as the constant α, we can see easily that

$$\lim_{n\to\infty} \frac{\rho_H(n)}{H(2H-1)n^{2H-2}} = 1.$$

Which means that for $H \geq \frac{1}{2}$ the fBm meets the definition of long-range dependency.

Long-range dependence is often found in internet traffic modelling, linguistics, hydrology, and as we will be discussing, may also be found in financial markets [38]. A trait that is often found in series that have long range dependence is called self-similarity [38].

4.4 Self-similarity

Self-similarity is a property that is found in both man-made and natural phenomena. Known for this are for example snowflakes, coastlines, traffic processes and trees [38][27]. A self-similar process is visually seen as the same pattern repeating both seen up close and seen from afar. In other words, there are small versions of the larger pattern repeated inside larger patterns.

Definition 3 *We say that an $\mathbb{R}^d$-valued random process $X = (X_t)_{t\geq 0}$ is self-similar or satisfies the property of self-similarity if for every $a > 0$ there exists $b > 0$ such that*

$$Law(X_{at}, t \geq 0) = Law(bX_t, t \geq 0).$$

Definition 4 *If $b = a^{-H}$ in the definition above, then we say that $X = (X_t)_{t\geq 0}$ is a self-similar process with Hurst index H or that it satifies the property of (statistical) self similarity with Hurst index H. The quantity $D = \frac{1}{H}$ is called the statistical fractal dimension of X.*

From the covariance function, we see that it is homogeneous of order $2H$ [6]. This is because when the argument n is multiplied by a factor, then the result is multiplied by $\frac{1}{2}h$ in the power $2H$.

The fBm is actually the only Gaussian process that is self-similar [37].

Is it reasonable that stock returns could be self-similar? One way of intuitively explain that they could behave that way is to look at how investors react to different kinds of news. It is not unreasonable to believe that they could react the same way each time they recieve news of the same type. When this happens, it is possible that the return would repeat the same kind of pattern.

4.5 Continuity and path differentiability

Theorem 1 *Let $H \in (0, 1)$. The fBm B^H admits a version whose sample paths are almost surely Hölder continuous of order strictly less than H.*

Theorem 2 *A function $f : \mathbb{R} \to \mathbb{R}$ is Hölder continuous of order $\alpha, 0 < \alpha \leq 1$, and we write $f \in C^\alpha(\mathbb{R})$, if there exists $M > 0$ such that*

$$|f(t) - f(s)| \leq M|t - s|^{\alpha},$$

for every $s, t \in \mathbb{R}$.

By using the Kolmogorov criterion, Biagini, Hu, Øksendal and Zhang [6] shows that the sample paths of B^H are almost everywhere Hölder continuous of order strictly less than H, and for no order greater than H.

The Hölder continuity is setting a limit for how big a change can be during the specific parts of the process. A Hölder condition can be used to show existence and uniqueness of solutions to stochastic differential equations [14].

That the fBm is Hölder continuous does not need to mean that we are able to differentiate the paths it creates. In fact the sample paths of B^H are not differentiable [6]. This is also found with the regular Brownian motion.

Proposition 1 *Let* $H \in (0, 1)$. *The fBm sample path* $B^H_{\cdot}$ *is not differentiable. In fact, for every* $t_0 \in [0, \infty)$

$$\lim_{t \to \infty} sup|\frac{B^H_t - B_{t_0}}{t - t_0}| = \infty \textit{ with propability 1.}$$

This is proved by [6] by utilizing the fact that the fBm is self-similar, as we have shown earlier.

Processes that are not differentiable are of course harder to work with, but it is a property we appreciate. Later, we will see how the fBm could be used to analyze price processes in financial markets, and if we were able to differentiate it, we would see what the price would become next, which is not realistic.

One of the difficulties we get from not being able to differentiate the fBm, is that we need to use something other than the regular calculus used to integrate and differentiate the fBm. For regular Brownian motion, this is done by using Itô calculus, but as we shall see later, this is not possible for fBm.

4.6 Semimartingale

A regular Brownian motion is known for being a martingale.

Definition 5 *A process $M_t, t \geq 0$ is a martingale if it has the properties*

- $\mathrm{E}[|M_t|] < \infty$
- M_t *is* $\mathcal{F}_t$*-measurable*
- $\mathrm{E}[M_t - M_s|\mathcal{F}_s] = 0, s < t$

A martingale could be explained as being a fair game. When one considers all the past information about the process, the best predicition for the next value, will be today's value. The martingale property of the Brownian motion, is especially put to good use when one needs to price derivatives. The technique where one use different equivalent martingale measures is for example what makes us able to find the Black-Scholes-Merton formula without having to deal with the added difficulty of covariance.

Definition 6 *A semimartingale is a process $Z_t = M_t + A_t$ where M_t is a local martingale and A_t is an adapted càdlàg process with a bounded variation.*

Definition 7 *A càdlàg is a function that is right continuous and has limits on the left side.*

We have already shown that the sample path of fBm is not differentiable. Since this is the case, we would like to be able to use the Itô calculus to find the derivatives, which we need when we wish to price a claim in a market that includes fBm. However, to be able to do this, the fBm would have to be a semimartingale [6].

This proof is valid for $H \neq \frac{1}{2}$

$$Y_{n,p} = n^{pH-1} \sum_{j=1}^{n} |B_{j/n} - B_{(j-1)/n}|^p$$

From the self-similar property of fBm, we have that the sequence $\{Y_{n,p}, n \geq 1\}$ has the same distribution as $\{\tilde{Y}_{n,p}, n \geq 1\}$ [30] where

$$\tilde{Y}_{n,p} = n^{-1} \sum_{j=1}^{n} |B_j - B_{j-1}|^p$$

Because the stationary sequence $\{B_j - B_{j-1}, j \geq 1\}$ is mixing, we can use a version of the Ergodic Theorem [30]. This tells us that $\tilde{Y}_{n,p}$ converges almost surely and in L^1 to $\mathrm{E}(|B_1^p|)$ as n tends to infinity [30]. And because of this, $Y_{n,p}$ converges in probability as n approaches infinity to $\mathrm{E}(|B_1^p|)$. Because of this

$$V_{n,p} = n^{-1} \sum_{j=1}^{n} |B_j - B_{j-1}|^p$$

converges in probability [30], i.e. other outcomes become less and less likely, to 0 as n approaches infinity if $pH > 1$, and to infinity if $pH < 1$.

Therefore, the fBm is not a semimartingale as long as $H \neq \frac{1}{2}$, and we can not use the regular Itô-calculus. This means that we do not have a way to differentiate or integrate expressions containing fBm, and we need to look at different methods that can be used to do this.

5 Calculus

5.1 Wick Itô Skorohod

Wick Itô Skorohod integration (WIS integral) is a technique used to handle integration of fBm, and it is often used in mathematical finance. The stochastic integral can be defined for all $H \in (0, 1)$. It is built by using white noise theory and Mallavian calculus [6]. The first is theory about how to utilize the fact that some movements or signals are just random disturbance or noise. The second is calculus used on the regular Brownian motion. The parts of the integral are known as the Wick product and the Skorohod integral.

The integration technique uses something called the M operator to relate the fBm to the regular Brownian motion [6].

Definition 8 *Let $0 < H < 1$. The operator $M = M_H$ is defined on functions $f \in \mathcal{S}(\mathbb{R})$ by*

$$\widehat{Mf}(y) = |y]^{\frac{1}{2}-H}\hat{f}(y), y \in \mathbb{R}$$

where

$$\hat{g}(y) := \int_{\mathbb{R}} \mathrm{e}^{-ixy} g(x)\mathrm{d}x$$

denotes the Fourier transform.

We then end up finding the relationship

$$\int_{\mathbb{R}} f(t)\mathrm{d}B^H(t) = \int_{\mathbb{R}} Mf(t)\mathrm{d}B(t), f \in L^2_H\mathbb{R}. \tag{1}$$

[6]. Now we see that the M operator works as some sort of bridge to translate the regular Brownian motion to fBm.

Definition 9 *Let $Y : \mathbb{R} \to (\mathcal{S})^*$ be such that $Y(t) \diamond W_t^H$ is dt integrable in $(\mathcal{S})^*$ Then we say that Y is WIS integrable and we define the WIS integral of $Y(t) = Y_t(\omega)$ with respect to B_t^H by*

$$\int_{\mathbb{R}} Y_t(\omega)\mathrm{d}B_t^H := \int_{\mathbb{R}} Y_t \diamond W_t^H \mathrm{d}t,$$

where $\diamond$ is the Wick product and W_t^H is the fractional white noise.

This definition can be used to find a fractional alternative to the classical Itô formula.

Theorem 3 *Let $H \in (0,1)$. Assume that $f(s,x) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ belongs to $C^{1,2}(\mathbb{R} \times \mathbb{R})$, and assume that the random variables*

$$f(t, B_t^H), \int_0^t \frac{\partial f}{\partial s}(s, B_s^H)\mathrm{d}s \text{ and } \int_0^t \frac{\partial^2 f}{\partial x^2}(s, B_s^H)s^{2H-1}\mathrm{d}s$$

all belong to $L^2(\mathbb{P})$. Then

$$f(t,B_t^H) = f(0,0) + \int_0^t \frac{\partial f}{\partial s}(s,B_s^H)\mathrm{d}s + \int_0^t \frac{\partial f}{\partial x}(s,B_s^H)\mathrm{d}B_s^H + H\int_0^t \frac{\partial^2 f}{\partial x^2}(s,B_s^H)s^{2H-1}\mathrm{d}s \tag{2}$$

This gives us an alternative that could be used insted of the Itô calculus. There is also found a fractional version of the classical Girsanov theorem, which gives us a way to change the measure, but we will not show this here.

5.2 Pathwise forward integration

There are not just one way to find integration techniques for fBm, and Pathwise forward integration is another. Pathwise integration builds on Riemann sums [6], which are the integrals most people know. The method divides the area under the function into squares. There are several ways to do this. For example, the squares can be under the line, over, or both. This will never give the true value of the integral, but as the squares are divided into smaller and smaller parts, the sum of the areas approaches the real value of the integral.

Definition 10 *Let $H \in (0,1)$. Let $(u_t)_{t\in(0,1)}$ be a process with integrable trajectories. The symmetric integral of u with respect to B^H is defined as*

$$\lim_{\epsilon\to 0} \tfrac{1}{2\epsilon} \int_0^T u(s)[B^H_{(s+\epsilon)} - B^H_{(s-\epsilon)}]\mathrm{d}s,$$

provided that the limit exists in probability, and is denoted by $\int_0^T u(s)\mathrm{d} \circ B^H_s$.

Definition 11 *Let $H \in (0,1)$. Let $(u_t)_{t\in(0,1)}$ be a process with integrable trajectories. The forward integral of u with respect to B^H is defined as*

$$\lim_{\epsilon\to 0} \tfrac{1}{\epsilon} \int_0^T u(s)\tfrac{B^H_{(s+\epsilon)} - B^H_s}{\epsilon}\mathrm{d}s,$$

provided that the limit exists in probability, and is denoted by $\int_0^T u(s)\mathrm{d}^- B^H_s$. The backward integral is defined as

$$\lim_{\epsilon\to 0} \tfrac{1}{\epsilon} \int_0^T u(s)\tfrac{B^H_{(s-\epsilon)} - B^H_s}{\epsilon}\mathrm{d}s,$$

provided that the limit exists in probability, and is denoted by $\int_0^T u(s)\mathrm{d}^+ B^H_s$.

In these definitions, we see that our regular integration approach with the Riemann sums is extended to the case with fBm. The alternative ways are easy to explain from their definitions. The last one integrates from where you are standing and back, while the middle one integrates by moving one step forward. The first integral goes equally long in both directions, and is thus called symmetric.

As with the other ways of integrating over fBm, we get an alternative version of the Itô formula. It is extended from the Itô formula for forward integrals used on regular Brownian motion [6].

Lemma 1 *Let $G \in (\mathcal{S})^*$ and suppose that ψ is forward integrable. Then*

$$G(\omega) \int_0^T \psi_t \mathrm{d}^- B_t^H = \int_0^T G(\omega)\psi_t \mathrm{d}^- B_t^H.$$

Together with

Definition 12 *Let ψ be a forward integrable process an let $\alpha(s)$ be a measurable process such that $\int_0^t |\alpha(s)| \mathrm{d}s < \infty$ almost surely for all $t \geq 0$. Then the process*

$$X(t) := x + \int_0^t \alpha(s)\mathrm{d}s + \int_0^t \psi(s)\mathrm{d}^- B_s^H, t \geq 0$$

is called a fractional forward process.

Theorem 4 *Let*

$$\mathrm{d}^- X_t = \alpha_t \mathrm{d}t + \psi_t \mathrm{d}B_t^H, X(0) = x$$

be a fractional forward process. Suppose $f \in C^2(\mathbb{R})$ and put $Y_t = ft(X_t)$. Then if $\frac{1}{2} < H < 1$, we have

$$\mathrm{d}^- Y_t = \tfrac{\partial f_t}{\partial t}(X_t)\mathrm{d}t + \tfrac{\partial f_t}{\partial x}(X_t)\mathrm{d}^- X_t.$$

When we have the lemma above, the proof is fairly simple. We apply the lemma along with Taylor expansion to $\mathrm{d}Y_t$ [6]. $\mathrm{d}Y_t$ is of course just another way of saying $Y_t - Y_0$, and this can be expressed by the sum of all the increments. At least, we take into account that the quadratic variation of the fBm is zero when $\frac{1}{2} < H < 1$ and end up with the expression from the theorem above [6].

5.3 Wiener and divergence-type integrals

Wiener integrals are integrals of deterministic functions with respect to a Gaussian process. When $H = \frac{1}{2}$, the integral is the same as the Itô formula.

Definition 13 *For any $H \in (0,1)$, the (abstract) Wiener integral with resepct to the fBm is defined as the linear extension from $\mathcal{H}$ in $L^2(\mathbb{P}^H)$ of the isometric map $\mathcal{I}^H$:*

$$\mathcal{I}^H : \mathcal{H} \to L^2(\mathbb{P}^H)$$
$$R^H_{(t,\cdot)} \to B^H_t$$

This is as it sounds like, a kind of map that tells us how to go from the integral definition to the actual fBm.

The divergence type integral for fBm is stochastic and uses something called the derivative operator [6].

5.3.1 Wiener integrals for $H > \frac{1}{2}$

There are two different ways of defining the Wiener integrals which are based on two different isometries, but here we use the second type from [6] and we obtain:

Theorem 5 *Let π_n be an increasing sequence of partitions of $[0,T]$ such that the mesh size $|\pi_n|$ of π_n tends to 0 as n goes to infinity. The sequence of processes $(W^n)_{n\in N}$ defined by*

$$W^n_t = \sum_{t_i^{(n)}\in\pi_n} \frac{1}{t_{i+1}^{(n)} - t_i^{(n)}} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} K^H_t(s)\mathrm{d}s[B_{t_{i+1}^{(n)}} - B_{t_i^{(n)}}]$$

converges to B^H in $L^2(\mathbb{P}\otimes \mathrm{d}s)$, where here $\mathbb{P}$ denotes the probability measure induced by the standard Brownian motion.

And we have

Definition 14 *Consider $H > \frac{1}{2}$ Let u be a stochastic process $u.(\omega) : [0,T] \to \mathcal{H}$ such that $K^*_H u$ is Skorohod integrable with respect to the standard Brownian motion B_t. Then we define the extended Wiener integral of u with respect to the fBm B^H as:*

$$B_u^H := \int_o^T (K_H^* u)_s \delta B_s$$

5.3.2 Wiener integrals for $H < \frac{1}{2}$

Proposition 2 *For $H < \frac{1}{2}$ the Wiener type integral B_ϕ^H with respect to fBm can be defined for $\phi \in \mathcal{H} = I_T^{\frac{1}{2}-H}(L^2([0,T]))$ and the following holds:*

$$B_\phi^H = \int_0^T (K_H^* \phi)(t) \mathrm{d}B_t.$$

5.3.3 Divergence type integral for $H > \frac{1}{2}$

Here, the divergence operator coincides with the generalized Wiener integral [6].

Proposition 3 *Let $u \in dom\delta_H$. Then $\delta_H(u)$ coincides with the extended Wiener integral of u*

$$\delta_H(u) = B_u^H.$$

5.3.4 Divergence type integral for $H < \frac{1}{2}$

One problem here is that the paths of the fBm becomes too irregular (which is a trait of the fBm when H is low) when $H < \frac{1}{4}$, so that the standard divergence integral ceases to exist [6].

An extended divergence operator is then defined and we get:

Definition 15 *Let $u_t, t \in [0,T]$ be a measurable process such that $\mathrm{E}[\int_0^T u_t^2 \mathrm{d}t] < \infty$. We say that $u \in dom^*\delta_H$ if there exists a random variable $\delta_H(u) \in L^2(\mathbb{P}^H)$ such that for all $F \in \mathrm{S}_\mathcal{K}$ we have*

$$\mathrm{E}[u_t K_H^{*,a} K_H^* D_t^H F]\mathrm{d}t = \mathrm{E}[\delta_H(u)F].$$

5.4 Fractional Wick Itô Skorohod for $H > \frac{1}{2}$

For this type of stochastic integral, fractional white noise is used along with the ϕ derivative [6].

Definition 16 *Suppose* $Y : \mathbb{R} \to (\mathcal{S})^*_H$ *is a given function such that* $Y_t \diamond W^H_t$ *is dt integrable in* $(\mathcal{S}^H)$. *Then we define its fWIS integral,* $\int_{\mathbb{R}} Y_t \mathrm{d}B^H_t$, *by*

$$\int_{\mathbb{R}} Y_t \mathrm{d}B^H_t := \int_{\mathbb{R}} Y_t \diamond W^H_t \mathrm{d}t$$

Where $\diamond$ *is the Wick product and* W^H_t *is fractional white noise.*

6 Integration methods suited for finance

We need to chose an integration method that is suited for the financial market. Since we assume that stocks have $H > \frac{1}{2}$, we can for example use pathwise forward integration or WIS integration. However, both methods are problematic [6].

6.1 Pathwise forward integrals

Pathwise forward integrals are natural to use in finance because $H > \frac{1}{2}$ [6]. If we choose this form of integration, this gives us the Itô formula shown earlier. When we replace the classical Brownian motion with fBm in the Black Scholes market, we can apply this to get an expression for S_t. We can prove that this makes it possible to find strong arbitrage opportunities in the market [6]. However, this disappears when one accounts for transaction costs in the market [6].

6.2 WIS integrals

The WIS integration technique could also be used in finance. Here, we get an alternative Itô formula for the integration as well, but the expression it gives for S_t is not the actual value of the firm, but rather a value that we are forced to use if we need one [6]. There is no strong arbitrage opportunities in this case, but we can prove that there is weak [6].

It is much more difficult to show that there exists weak arbitrage in the Black Scholes market with WIS integrals than there is to prove the existence of arbitrage with Pathwise forward integrals. We need to apply theory from quantum mechanics and then show that we have a portfolio that fits the critera of weak arbitrage [6].

But we cannot justify to use this model. This is shown by Björk and Hult [10]. They explain that both the definition of a self-financing strategy and the definition of the value of the claim are problematic. They do not have an economical interpretation in the real world when this form of integration is used [10]. An example of this is taken from a portfolio used of Øksendal and Hu in 2003. It is shown that when the portfolio holds a positive number of shares with positive prices, there is a positive probability that the value of the portfolio is negative [10].

7 Arbitrage with fBm

7.1 What is arbitrage?

Arbitrage is often called a free lunch. This is because we are essentially making money without investing or taking any risks. It is done by exploiting the fact that some assets are mispriced by making a portfolio that take a short position in the overpriced assets and a long position in the underpriced ones.

Definition 17 *A self-financing portfolio is one that has*

$$X_t^{\pi} = X_0^{\pi} + \int_0^t \gamma_u \mathrm{d}S_u(\mu)$$

This means that the value of the portfolio always consists of the starting value in addition to any losses or gains one may have during the investment period. There are no additional in- or outflows of cash during this time. This property is important to us because it makes pricing easier when we use a portfolio to replicate a claim.

Definition 18 *Let π be a self-financing strategy and V be wealth. π is called an arbitrage strategy if $P(V_{\pi}(0) = 0) = 1$ and the terminal wealth satisfies the conditions [7]:*

$$P(V_\pi(T) \geq 0) = 1 \ a.s. \ and \ P(V_\pi(T) > 0) > 0$$

This means that we would start off without any cashflow at the beginning of the investment period, but in the end of the period, we will have a positive cashflow. If we were able to to this, we would have a money machine.

Most financial theories are built on the basis of an arbitrage-free market. This is a market where there does not exist any arbitrage strategies.

Theorem 6 *A market is arbitrage-free if there exists at least one equivalent martingale measure.This is known as the fundamental theorem of asset pricing.*

Definition 19 *If there exists one and only one equivalent martingale measure, the market is said to be complete*

A complete market is one where we are able to replicate all the different claims, and hence price them [7].

7.2 Why is arbitrage a problem?

If there exists arbitrage in a market, the market is not in equlibrium. One claim could have several different prices. One theory about complete markets says that an arbitrage opportunity will disappear because investors would buy the relatively underpriced claims until their prices had risen, and sell the overpriced ones until the prices had sunk. This would happen until the market ended up in equlibrium and all the claims had only one correct price.

Definition 20 *A price is C_T is rational if*

$$C_T = inf\{x \geq 0 : \exists \pi \ with \ X_0^\pi = x, X_T^\pi \geq f_T\}$$

Here $\pi = (\beta, \gamma)$ is a portfolio, and the corresponding capital is

$$X_t^\pi = \beta_t B_t(r) + \gamma_t S_t(\mu), t \leq T.$$

In short, this definition means that there exists a portfolio we can use to replicate the claim. We also assume that π is self-financing.

Well known pricing formulas such as the Black-Scholes-Merton model are based on the possibility to replicate claims using the prices of others with something called the martingale approach.

7.3 Arbitrage for fBm when $\frac{1}{2} < H < 1$

We want to show that there exists arbitrage opportunities in the fractal version of the Black-Scholes-Merton market model. The BSM-market model is commonly used because it does not allow the prices to become negative [32]. This is a trait we appreciate, as it makes our market model more realistic. The following proof is made by Shiryaev [32].

The model consists of a risk-free asset $B_t(r)$ and the risky asset $S_t(r)$

$$\begin{aligned} B_t(r) &= \mathrm{e}^{rt}, \\ S_t(r) &= \mathrm{e}^{rt+B_t^H}. \end{aligned}$$

We start by differentiating the two assets. This is later used as input when using the Itô-formula for fBm on the chosen portfolio.

$$\begin{aligned} \mathrm{d}B_t(r) &= rB_t(r)\mathrm{d}t \\ \mathrm{d}S_t(r) &= S_t(r)(r\mathrm{d}t + \mathrm{d}B_t^H). \end{aligned}$$

We would now like to show that there exists a portfolio $\pi = (\beta, \gamma)$ that gives us arbitrage in this market model.

$$\begin{aligned} \beta_t &= 1 - \mathrm{e}^{2B_t^H}, \\ \gamma_t &= 2(\mathrm{e}^{B_t^H} - 1). \end{aligned}$$

By applying this portfolio to our market model, we obtain

$$X_t^\pi = \beta_t B_t(r) + \gamma_t S_t(r) = \mathrm{e}^{rt}(\mathrm{e}^{B_t^H} - 1)^2.$$

Then we need to prove that this stragey is a self-financing arbitrage strategy. Now we use an alternative Itô formula to differentiate X_t^π to see that the only change in value in the portfolio comes from the summarized changes in S_t and B_t. Remember how alternatives to the Itô formula was found by using integration techniques for fBm.

The Itô formula used is

$$F(B_t^H) - F(B_0^H) = \int_0^t f(B_u^H)\mathrm{d}B_u^H.$$

We apply this and use the derivatives of the risky and the risk-free asset. Then we get

$$\mathrm{d}X_t^\pi = r\mathrm{e}^{rt}(\mathrm{e}^{B_t^H} - 1)^2\mathrm{d}t + 2\mathrm{e}^{rt+B_t^H}(\mathrm{e}^{B_t^H} - 1)\mathrm{d}B_t^H.$$

When rearranging the items and remembering the derivatives of the risky and risk-free asset along with the portfolio, one can see that the expression becomes

$$\mathrm{d}X_t^\pi = \beta_t\mathrm{d}B_t(r) + \gamma_t\mathrm{d}S_t(r)$$

When we integrate over the time period 0 to T, we find that

$$X_T^\pi = \beta_0 B_0 + \gamma_0 S_0 + \int_0^T \beta_u B_u(r)\mathrm{d}B_u + \int_0^T \gamma_u S_u(r)\mathrm{d}S_u$$

If we remember the definition of a self-financing model, we see that the last equation fits the criteria. Also, we need to prove that the portfolio meets the additional requirements for arbitrage. When using the fact that B_0^H=0, we can easily see that the first two terms of the expression, namely S_0^π, is 0. We can also see that for $t > 0$, the rest of the terms, which are X_t^π, will be greater than 0 since our model does not allow for negative values. This is enough to prove that this market model admits arbitrage.

Similiar proofs can also be shown for other market models including fractional Brownian motion. For example the Bachelier model [32]. But a somewhat interesting fact is discussed by among others Bender, Sottinen and

Valkeila [4]. They show that one can restrict the available trading strategies in a way that does not admit arbitrage in the market. There are also examples when fBm is used together with classic Brownian motion to create a market without arbitrage.

8 Simulation and estimation

There are several methods used to simulate fBm. Some of them are exact, while others are approximations. They use different properties of the fBm, and they have their own advantages and disadvantages. The simulation is done in discrete time.

I do not use simulation futher in this thesis, but when I refer to other papers, simulation is often used to evaluate different methods used in estimation of fBm. This is why I find it relevant to mention here.

8.1 The Hoskin method

This algorithm generates X_{n+1} given $X_n, ..., X_0$ recursively [15]. The sample is obtained by taking cumulative sums of fractional Gaussian noise [15]. It can be used on all stationary Gaussian processes, and the simulation is exact [15].

The complexity of the algorithm is of order N^2 when one need N observations [15]. The advantages of this method is that it is easy, and that one does not need to know the size of the sample one creates in advance [15].

8.2 The Cholesky method

This method uses something called the Cholesky decomposistion on the covariance matrix to simulate X_{n+1} given $X_n, ..., X_0$ recursively [15]. It can be used on both stationary and non-stationary Gaussian processes, and the method produces exact values of the fBm [15].

As with the Hoskin method, one does not need to know the number of observations one needs in advance of the simulation [15]. The Cholesky method is however slower, and uses a lot of storage space when used [15]. When the number of observations needed are N, the complexity is of order N^3 [15].

8.3 The Davies and Harte method

This algorithm uses somewhat of the same tehcnique as the Cholesky method by trying to find a square root of the covariance matrix [15]. The method produces exact values of the fBm [15].

From one simulation, we get two samples of size N, but these can not be combined to create a larger sample since the covariances between them are not correct [15]. The other sample of N observations can not be used alone either, as it is not independent from the first sample [15]. To use this would give skewed results.

The main advantage with the Davies and Harte method is the speed. When one need N observations, the complexity is only $N \log(N)$ [15].

8.4 Approximated methods

There are some approximated methods that are not used anymore, but are still interesting because of their historical place in fBm science [15]. This is for example the stochastic representation method defined by Mandelbrot and van Ness [15].

The advantages of using approximated methods instead of exact, are that they are faster and easier to compute [15]. Some of the methods are modified and easier versions of the exact methods, such as the Random Midpoint Displacement and spectral simulation. Some of the modifications are made to have easier and fewer calculations by changing formulas, while other ways of simplifying is to not use all the past simulated values when finding new ones.

When one need large sample sizes, the speed and memory usage of the methods are often important, and the while the fastest exact method had a complexity of $N \log(N)$, approximated methods such as the random midpoint displacement have a complexity of order N [15].

8.5 Estimation of the Hurst coefficient

8.5.1 R/S analysis

The most popular method used to estimate H has traditionally been the method developed by Mandelbrot and is called R/S analysis. Mandlebrot explained it as a way to distinguish the Noah effect from the Joseph effect [27]. The method is used to find evidence of long memory in different time series.

- Divide the time series into n subseries $Z_{i,m}$ of length n.
- Find the mean $\mathrm{E}_{i,m}$ and the standard deviation $\mathrm{S}_{i,m}$ of each of the subseries.
- Normalize the subseries by subtracting the mean $X_{i,m} = Z_{i,m} - \mathrm{E}_m$.
- Create the cumulative series $Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ for $i = 1, ..., n$.
- Find the range $R_m = max\{Y_{1,m}, ..., Y_{n,m}\} - min\{Y_{1,m}, ...mY_{n,m}\}$.
- Rescale the range (R_m/S_m).
- Find the mean value $(R/S)_n$ of the rescaled range for all the subseries with length n.
- The R/S statistics asymptotically follows $(R/S)_n cn^H$. Find H by running a linear regression over $\log(R/S)_n = \log c + H \log n$.

This method has been tested empirically by running Monte Carlo simulation to create simulations of random walks. When the series are not large enough, there has been found evidence of long memory in more of the series than what should be acceptable [18]. This could be a problem, as this method has often been used by researchers who have found evidence of long memory in stock markets.

8.5.2 Other methods

Because of the problems one finds when testing the R/S method empirically, one should also consider other methods. Granero, Segovia and Perez has suggested several improvements in the R/S analysis that did not have the same problems [18].

Other methods are detrended fluctuation analysis and periodogram regression. The first tries to distinguish trends from long range fluctuations by eliminate trends of different orders [24]. The second method tries to estimate spectral density of signals [35]. We will come back to other methods later.

9 $ARFIMA$ models

9.1 Discretizing the fBm

When we try to estimate a fractional Brownian motion model for stock price returns, our observations are discrete, as all real life observations are. The discrete version of the regular Brownian motion is known as the random walk, and it can be modelled by $ARIMA(0,1,0)$ [23], which stands for Autoregressive, Integrated, Moving-Average model.

Definition 21 *The derivative of a regular Brownian motion is defined as*

$$\delta x_t = (1-L)x_t = a_t$$

and is called a random walk. The first difference of the random walk is the discrete white-noise process $\{a_t\}$

We need to find a similar way to discretize the fractional Brownian motion in a way that captures the properties of the continouos, fractional Brownian motion.

9.2 Fractional $ARIMA$ models

From definition 2 of long-range dependence in fBm, we remember that $\sum_{n=1}^{\infty} \rho(n) = \infty$. This tells us something about the properties we are looking for in a long memory process. When the memory is long, the summation of the autocorrelations will go to infinity, as the memory does not disappear. The AR and MA terms from the regular $ARIMA$ model decay geometrically, while those of the d in a fractional $ARIMA$ model decay hyperbolically. This makes the combination of the two ideal to model both long and short memory. The correlations should decay slowly, or more spesifically it should be hyperbolic if it has long memory [5].

Definition 22 *Let X_t be a stationary process for which the following holds: There exists a real number $\alpha \in (0,1)$ and a constant $c_\rho > 0$ such that*

$$\lim_{k\to\infty} \frac{\rho(k)}{c_\rho k^{-\alpha}} = 1$$

Then X_t is called a stationary process with long memory or long-range dependence or strong dependence, or a stationary process with slowly decaying or long-range correlations.

It does however exist several different definitions of what a time series with long term memory is.

There is a direct relationship between the parameter α and our H from earlier

$$H = 1 - \frac{1}{\alpha}.$$

The definition only tells us something about the correlations as k goes to infinity [5]. It does not tell us anything about the absolute sizes of the lags, so a time series could be slowly decaying even though they are initially small. This means we still need a method to find the sizes for specific lags.

As shortly mentioned above, the Autoregressive moving-average model ($ARMA$) finds by itself only the short-term memory. An $AR(p)$ model actually has infinite memory, but the effect of earlier observations follows a geometric decay, and the effect on today's observation is therefore quickly diminishing [2]. It is easier to see that a pure $MA(q)$ model only has a short memory, as the effect on today's value disappears after q lags [2].

We start by looking at a regular $ARMA(p,q)$ model.

- $\mu = \mathrm{E}[X_t] = 0$, or else we need to remove μ from X_t.
- $\phi(x) = 1 - \sum_{j=1}^{p} \phi_j x^j$.
- $\psi(x) = 1 + \sum_{j=1}^{q} \psi_j x^j$.
- We use the backshift operator L to express the differences i.e. $X_t + X_{t-1} = (1+L)X_t$.
- P and q are integers.
- The error terms $\epsilon_t (t = 1, 2, 3, ...)$ are independent and identically distributed (iid) with expectd value 0 and variance σ_ϵ^2.
- All solutions of $\phi(x_0) = 0$ and $\psi(x_0) = 0$ lies outside the unit circle.

The $ARMA(p,q)$ model is the stationary solution of $\phi(L)X_t = \psi(L)\epsilon_t$ [17]. The model is then extended to an $ARIMA(p,d,q)$. This model is an $ARMA(p,q)$ model which is integrated by order d if the solution holds for the dth difference of the orginal time series.

$$\phi(L)(1-L)^d X_t = \psi(L)\epsilon$$

We can use an extension from the regular $ARIMA$ model to make it fit our needs [5]. Where the $d = 0$ in a usual $ARMA$ process. When we go from an $ARMA$ process to an $ARIMA$ process, we violate the last of the bullet points. Now the solutions to the equations can lie inside the unit circle [17].

In a regular $ARIMA$ model, the d is only allowed to be an integer, but the expression is also defined when d takes on any real number. This is when an $ARIMA$ model becomes fractional, and this is often called an ARFIMA model. This can be defined as

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k$$

Where

$$\binom{d}{k} = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

Here Γ is the gamma function. It is defined for all real numbers, and this is why we can extend the model to accommodate d other than integers [5]. This is an infinite series. When this is estimated in practice, we have to choose how many terms we wish to include.

For our solution to be stationary and invertible [23], $-\frac{1}{2} < d < \frac{1}{2}$. If d lies outside this interval, our function would be nonintegrable, and will not fit with our earlier theory with $0 < H < 1$ [5]. This is because $H = \frac{1}{2} + d$.

Definition 23 *Let X_t be a stationary process such that*

$$\phi(L)(1-L)^d X_t = \psi(L)\epsilon_t$$

This is the same expression as when we used a regular $ARIMA$ model. Note that the possible values of d are the only thing that is changed.

Definition 24 *The covariance function between X_t and X_{t-k} is*

$$cov(X_t, X_{t-k}) = \gamma_k = \frac{(-1)^k \Gamma(1-2d)}{\Gamma(1-k-d)\Gamma(1+k-d)}$$

Definition 25 *The correlation function is*

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{(-1)^k \Gamma(1-d)^2}{\Gamma(1-k-d)\Gamma(1+k-d)}$$

It is this function that will show hyperbolic decay when $0 < d < \frac{1}{2}$ and the time series have long-term memory. Observations that are far apart will have highly correlated if d is high. Hosking [23] defines the correlation for as $ARFIMA(0,d,0)$ as

$$\rho_k = \frac{(-d)!(k+d-1)!}{(d-1)!(k-d)!}$$

and show that as $k \to \infty$

$$\rho_k \sim \frac{(-d)!}{(d-1)!} k^{2d-1}.$$

This implies that $\{x_t\}$ is asymptotically self-similar [23]. Our way of modelling a discrete form of the fBm has then been shown to have several of the properties we went through in first part of the paper. The $ARFIMA$ model seems to be a good choice as it will try to separate the long and the short memory.

9.3 Different estimators

There are mainly two different ways of estimating an $ARFIMA$ model, namely parametric and semi-parametric methods [1].

In the semi-parametric methods, the short memory components are not important, and one can estimate the long-term behavior without knowledge of the short-term behavior. These approaches can be enough if one wants to show that there is long memory in the data, but if one needs more information, like values of the parameters or the short-term behavior, more refined methods should be used [5].

The parametric methods can model the short-term behavior as well, but we do need to specify the $ARMA$ components [2]. To find the correct components migth be difficult, but this is not our only problem. The model needs to separate the short memory from the long, and where does the asymptotic behavior actually start? [5]. But this is also the main reason to use these kind of methods, as we do not want to mistakenly find that a time series has long-term memory when what we are actually seeing is short-term behavior from a regular $ARMA$ model.

In an earlier section, we went through some of the methods that traditionally has been used to estimate the Hurst coefficient, but we will not use any of those here. According to Beran [5], maximum likelihood estimators (MLE) are the most efficient one. There are several methods within this group. Some of them are exact Gaussian MLE, Whittle's approximate MLE and the approximated method of Haslett and Raftery. As the names indicate, only the first one is exact, but the downside is that it demands extreme computational time if one has a large data set [5].

Later, we are using the package fracdiff in R, and this package uses the approximated method of Haslett and Raftery.

9.3.1 The Haslett and Raftery method

Let $Xt = (X_1, ..., X_t)$ and$X_i^t = .(X_{i1}, ..., X_{it})$. We then know that the expectation of one observation conditional on all the earlier ones has a multivariate normal distribution [20]. We have that $\mathrm{E}[X_{it}|X^{t1}] = \mathrm{E}[X_{it}|X_i^{t-1}]$, $var[X_{it}|X^{t-1}] = var[X_{it}|X_i^{t1}]$ and $corr[X_{it}, X_{jt}|X^{t-1}] = \alpha \mathrm{e}^{(-\beta d_{ij})}$ [20].

The method also assumes that $0 < d < \frac{1}{2}$ [20], which is what we are looking for in the stock returns

Then we would like to maximize the likelihood, and this is possible to do exactly, while using a numerical approach [20]. However, it would take a lot of time and computer capacity [20].

Haslett and Raftery [20] uses several approximations in their estimations. Conditional mean and variance are found using the partial autocorrelation function (PACF) of the series $ARFIMA(0, d, 0)$ instead of the full $ARFIMA(p, d, q)$ [20]. Then they analitically find approximations for μ and σ^2. The final approximation is of the partial linear regression coefficients of the $ARFIMA(0, d, 0)$ process [20]. The likelihood function is then dependent on $\alpha, \beta, d, \phi(L), \theta(L)$ [20]. As with all similar approaches, the point of the method is to estimate the differetn parameters by maximizing the likelihood of them being correct, given what observations our time series give us.

I do not go any further into the detailed calculations of the approximations here, but this is an easy way of explaining how the package I am applying to R works. While this method was originally used by Haslett and Raftery to estimate a fractional $ARIMA$ model for Ireland's wind power resources, it could still be used in other areas, as the method itself is not dependent on this.

9.4 Long-term versus short-term dependence

Could we only have used a model consisting of μ, σ and H, which would be equivalent to a $ARFIMA(0, d, 0)$ model? According to Hosking [23] it has been claimed that such a model would lack flexibility as it would not be able to give accurate short-term predictions, and that it performs poorly when compared with other models by using the Akaike information criterion.

The best reason why the $ARFIMA$ model is such a good chioce for separating the long-term and the short-term behavior is that the effect of the parameter d decays hyperbolically, while the effect from the AR and MA terms diminish exponentially [23]. This means that the effect of the last two will die out faster, and eventually just leave the effect of the d parameter.

Other problems could be that we find long term dependence where no such thing actually exists because we model short term behavior as long term. To adequatly separate these two behaviors is not easy when we use the $ARFIMA$ model, but it gives us a chance to try. The main problem is to decide where the asymptotic behavior start. In other words, what constitutes long memory, and what is just short?

How does one separate the two different dynamics in practice? One have to try to determine if the correlation follow a Hyperbolic curve k^{2H-2} or an exponential curve c^k. It is not enough to determine if the correlation function has large or small values, as the absolute values of the lags can be small, but still diminish slowly. There is no exact way of doing this, and two different scientists could estimate two different models by using the same data set.

There does not appear to yet exist an easy solution to this problem of determining what kind of an $ARFRIMA$ model to use. And as we will show later, different choices of lags to use in the AR and MA part will give sligthly different estimates of d. According to Hosking [23] the behavior of an $ARFIMA(0, d, 0)$ will be almost the same as that of an $ARFIMA(p, d, q)$ model on very distant observations.

The AR or MA part of a solution will as we have explained account for the dynamics seen in the first PACFs and ACFs, but it will not have much to say for the long term dynamics. But could it be different if we just chosse to use enough lags in our model? To some extent, yes, but after the chosen number of lags, the effect from the MA and AR terms still will diminish much faster than that of the fractional differenced part, so it would not do a good job for very distant observations. This would also affect the parsimony of the model. A trait we appreciate and strive to fulfill when we fit a model to time series data.

9.5 Arbitrage in the $ARFIMA$ model

Both properties of the long- and short-term modelling could give us arbitrage opportunities. Let us first look at the alternative values of d and how these determine the behavior of the model.

When d=0 in the discrete model, this is the same as $H = \frac{1}{2}$ in the continuous fBm. The long term dynamics of the time series will not be the source of any arbitrage opportunities, as we in part 7 showed that arbitrage will not occur when $H = \frac{1}{2}$. We also note that the correlation is 0 in this case, as it is supposed to be since this actually models a regular Brownian motion.

When $0 < d < \frac{1}{2}$, the times series show long term persistence, and it is also stationary as we showed earlier. It is analogue to $\frac{1}{2} < H < 1$ in the continuous case, and we showed in part 7 that this will give us arbitrage opportunities if a stock price has these dynamics. This is what we will be looking for in financial time series later.

When $-\frac{1}{2} < d < 0$, the time series shows antipersistent, and has a short memory [23]. This is not behavior we expect to find in financial time series, and in part 7 of the paper, we did not show any proof of how this could lead to arbitrage, but it has been shown by for example Cheridito [12].

When we estimate a fractional $ARFIMA$ model for the changes in stock prices minus μ, we would expect to find a model consisting of only white noise if economic theory is correct. If we find any form of either long or short memory, the changes depend on each other. This means that d should be 0, as this means $H = \frac{1}{2}$, and there should not be any MA or AR lags.This is because we do not want it to be possible to predict future returns from neither short- or long-term dynamics.

However, I will not spend much time analysing the consequences of short memory in the data, as this is not the objective of this thesis. The short memory is mainly included here so the model will not estimate wrong values of d because it confuses short memory with long.

The higher estimates we find of d, the more long memory will be present in the model, and there is a greater chance of finding arbitrage opportunities that pays off when they are implemented in the real world with transaction costs and limits to how often we can trade.

Such findings would have great implications for economic theory. The most prominent are that there will no longer be market efficiency and that martingale models for stocks are no longer possible to use. This means that our known approaches for pricing for example derivatives will not work, as the martingale approach is used today.

The main problem that concerns us are long memory in stock returns, this is also what I spend most time on here. However, long memory in the squared returns will also have consequences. From intuition, we know that there could exist arbitrage opportunities in some cases if we were able to predict the volatility of a stock. For example, if we knew for sure that the predicted volatility was not large enough for the stock return to be negative in the next period, the price should already have risen.

9.6 Identifying the correct model

Now we know how the chosen method estimates the parameters of an $ARFIMA(p, d, q)$ model, but they can only be used if we know the true model that lies behind the data. When working with real time series, we usually do not know this in advance. Then we have to use different techniques to identify the true underlying model.

We are used to depend on the autocorrelations (ACF) and the partial autocorrelations (PACF) from identifying regular $ARIMA$ models, and there is a pretty clear step by step process to do this. Even though the same data set could be assigned different models by two different researchers. The same will be true when estimating the $ARFIMA$ model.

The fracdiff package in R does some of the work for us, by estimating the value of d as well as the values of the AR or MA lags. We still need to tell the program that the correct model to use includes a fractional d. We also need to find the number of lags to include. This is not straigth forward.

According to Hosking [23], the PACF is complicated to read and of little help when one tries to identify the correct model when he tried to analyse a $ARFIMA(1, d, 0)$ process. The ACF of this model seem to have a sharper drop from lag 1 to lag 2 than that for the $ARFIMA(0, d, 1)$ [23].

Hosking [23] suggests a way of identifying the correct $ARFIMA$ model. First we define $u_t = \bigtriangledown^d y_t$ such that $\{u_t\}$ is an $ARFIMA(0, d, 0)$ process and $x_t = (\theta(L))^{-1}\phi(L)y_t$ such that $\{x_t\}$ is an $ARFIMA(0, d, 0)$ process.

1. Estimate d.
2. Define $u_t = \bigtriangledown^d y_t$.
3. Find and estimate ϕ and θ in the $ARFIMA(p, 0, q)$ model.
4. Define $x_t = (\theta(L))^{-1}\phi(L)y_t$
5. Estimate d in $\bigtriangledown^d x_t = a_t$.
6. If the parameters d, ϕ and θ are not converging, go back to step 2.

A simulation study was done by Abraham, Lopes and Reisen [1] to test this procedure, and they found that there was usually only need for one iteration to get good enough estimates, as the values for ϕ, θ and d from doing the first three steps was very close to the converged values. They also found that the estimation of the AR lags had an impact on the estimation of d.

We start by finding the ACF just to look at it. We have to determine if the time series seems to have long memory. The decay of the ACF is telling us if there is long memory or not. It is important to remember that the absolute sizes of the lags do not have anything to say. Other visual tools that could be used are varaince plots, R/S plots or variograms. But the use visual tools is not an exact science.

If the decay of the ACF plot implies that the time series have long memory, then we estimate d using the package fracdiff. Then we would like to look at the residuals after the $ARFIMA(0, d, 0)$ model. This is not straightforward to find and there is no implemented method for this in the fracdiff package. In short, this consits of choosing the number of lags to use from

$$(1 - L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k L^k$$

and then calculate each of the residuals manually after that.

From these residuals, we try to identify the correct number of AR and MA lags. From identification of regular $ARMA(p, q)$ models, we know that an ACF with geomatrically decaying pattern and a PACF that is 0 after p lags, tells us that we should have p AR lags [9]. In the same manner, a geometrically decaying PACF along with an ACF that is 0 after q lags, tells us that we should have q MA lags. There is no exact pattern to look for if the model is mixed. We would like to use as few lags as possible to keep our model parsimonous.

After we have identified the correct number of lags, we go back and estimate the entire model from the start by giving the fracdiff package the correct inputs. It will then give us estimates for θ, ϕ and d.

The residuals from this model should not have any pattern, in other words, there should not be any autocorrelation left. Other thing we can check for in the residuals are normality, homoscedasticity and autoregressice conditional heteroscedasticity (ARCH) effects.

The autocorrelation is found by estimating ACF again. The normality could both be tested for and be checked visually by checking that the residuals form a bell shaped curve. To check if the residuals are homoscedastic, we plot the residuals against the independent variables. If they are, they should be scattered randomly around 0. ARCH in the residuals is found if there is autocorelation in the squared residuals.

As well as trying to find the correct model by using this method, I would like to find the Akaike information criterion (AIC) for a set of relatively simple models and check which one give us the best fit. The AIC compares the goodness of fit of a model with the complexity. It does not tell us if a model is completely inapropriate or not, but it compares different models. The model that has the lowest score from AIC has the best relationship between goodness of fit and complexity.

10 Estimating models for stock price returns

I want to look at the daily changes in stock prices to see if I can find any evidence of long-term memory. I will both be looking at some single stocks and an index. The datapoints are downloaded from Yahoo Finance to the program R where I use the package fracdiff to work with the estimation of the models.

The data first needs to be inverted, as they are downloaded in the opposite order, then we remove the days that are weekends and holidays if those are not already removed by Yahoo. This is done by removing the days where no trading has occured. We want to look at the adjusted closing price for each day so we do not catch any price jumps from dividends or splits. We then take the difference of the logarithm of each day's prices.

We will both be looking at the regular returns of the stock prices and the squared absolute returns. The first is clearly to look for long memory in the price dynamics. The second can be used as a proxy for the variance of the prices, and long memory here will suggest that there are some predictability in the stock's risk.

When working with the series, we have to remember to remove the mean so the expected returns are 0.

10.1 Data

Here is a selection of different types of stocks from the Oslo Stock Exchange. Most of the stocks have data for about ten years, but some of them only have for a couple of years. It will be interesting to see if we are able to estimate a

Ticker	Name	From	To
AUSS	Austevoll Seafood	11.10.06	13.06.14
CECON	Cecon	16.08.10	13.06.14
ECHEM	Eitzen Chemicals	28.06.04	13.06.14
FOE	F. Olsen Energy	03.01.00	13.06.14
GOGL	Golden Ocean	15.12.04	13.06.14
HAVI	Havila Shipping	24.05.05	13.06.14
NAUR	Northland Resouces	23.10.06	13.06.14
NAS	Norwegian Air Shuttle	18.12.03	13.06.14
PLCS	Polarcus	09.04.10	13.06.14
PROS	Prospector offshore Drilling	25.05.11	13.06.14
SN I	Stolt-Nielsen	08.03.01	13.06.14
TIDE	Tide	03.01.00	13.06.14
VIZ	Vizrt	12.05.05	13.06.14
WWASA	Wilh. Wilhelmsen	24.06.10	13.06.14
OSEBX	OSE Bench IDX	03.01.83	13.06.14

Table 1: The data used

long memory model with this amount of information. There is an overview of all the stocks I have used in table 1.

The length of each of the time series used here differs, but all observations are of the adjusted closing prices from each trading day. None of the data sets are extremely long, but if they were, it would most likely have been necessary to divide them into smaller series. This is because structural changes over the years could have made the H coefficient change, and our analysis could suffer if we made the assumption that it was the same over all the observed years. This is for example done by Bayrakter, Poor and Sircar [3] when analysing the S&P 500 Index.

10.2 Returns of the OSEBX

We start by looking at the index at the Oslo Stock Exchange, named OS-EBX. There is no registered trading volume of the OSEBX, so it will not be possible to remove holidays that have not already been removed by Yahoo. This will hopefully effect the model minimally.

As with most stocks and indexes, it has rosen sharply the last years. We can see this in plot a in figure 1. We have data from January 1983 and until June 2014. This means almost 8000 observations, and it could have been possible to divide it into smaller parts.

The returns have clear tendencies to cluster, and there are three clear points where the returns have large fluctuations, as we see from plot b. The ACF in plot c do not have any large spikes. It is hard to determine how slowly it decays, but there seem to be several spikes with almost the same size as the ones in the beginning. This means there could be a reason to fit an $ARFIMA$ model to the data. The PACF in figure d do not tell us much. It oscilliates with no clear decay.

We estimate d and find the residuals for the $ARFIMA(0, d, 0)$ model. The estimate for d is 0.0545129 which indicates some long memory. We would like to use the ACF and PACF to find out how many lags to use for the MA and AR part. There are however no clear pattern here from figure 2, so we try several models.

In figure 3, 4 and 5 I have estimated $ARFIMA(1, d, 0)$, $ARFIMA(0, d, 1)$ and $ARFIMA(1, d, 1)$ and plotted the ACFs and PACFs of the residuals. We see that the ACF and PCF plots are almost identical for the $ARFIMA(1, d, 0)$ in figure 3 and the $ARFIMA(0, d, 1)$ in figure 4. None of them seem to be making an especially good fit, and the estimates of d are lower than for the $ARFIMA(0, d, 0)$ model. The choice of short run dynamics clearly effects the estimation of the long memory, in other words the number of lags for MA or AR will change the estimate of d. It can seem as if d in $ARFIMA(0, d, 0)$ has tried to account for some of the short memory.

However, the opposite is true for the $ARFIMA(1, d, 1)$ model. The esti-

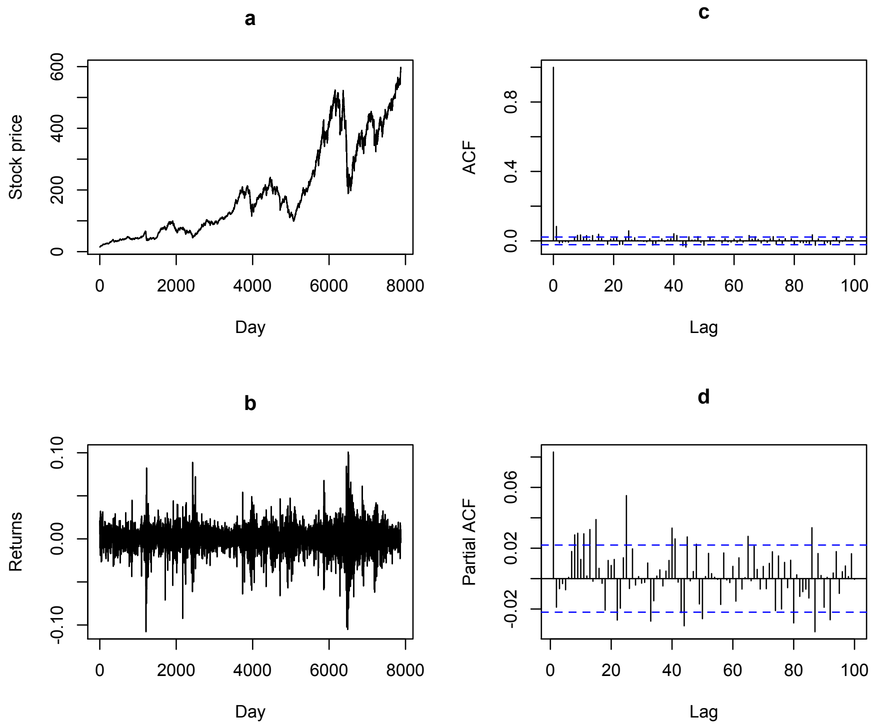

Figure 1: Overview of the OSEBX

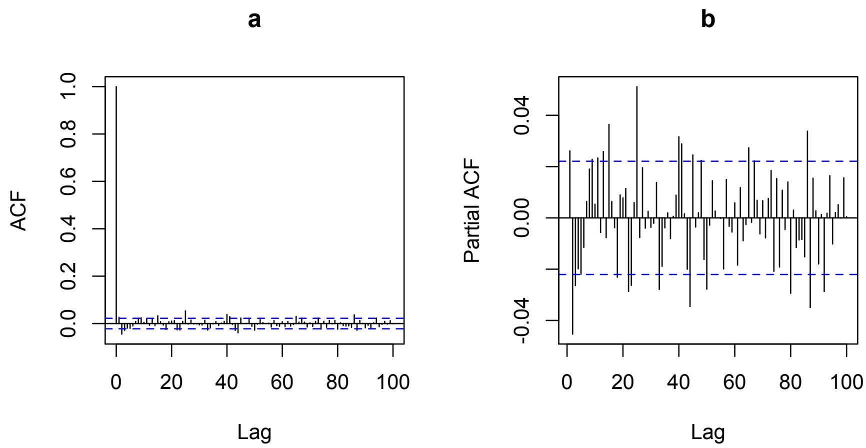

Figure 2: $ARFIMA(0, d, 0)$ for OSEBX

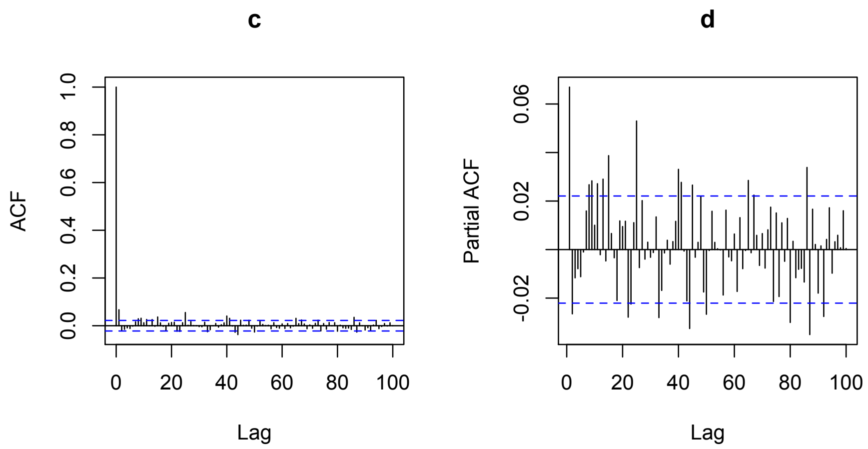

Figure 3: $ARFIMA(1, d, 0)$ for OSEBX

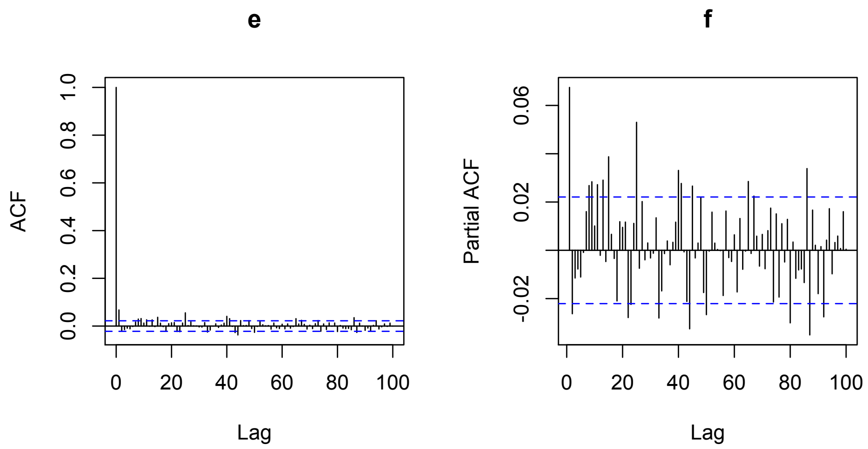

Figure 4: $ARFIMA(0, d, 1)$ for OSEBX

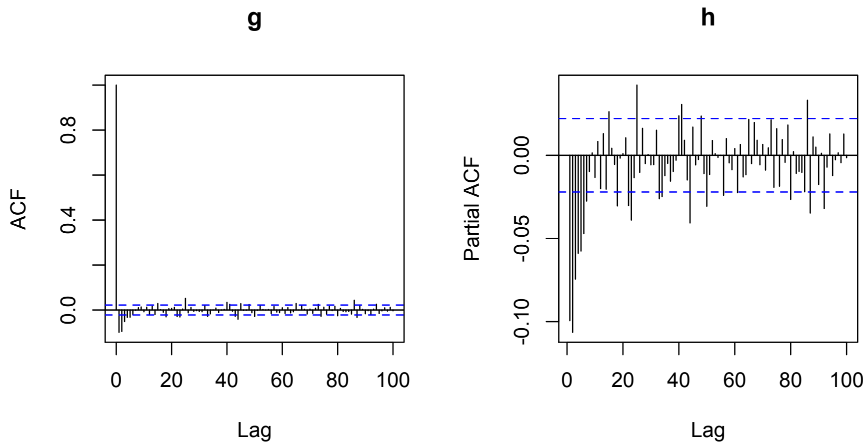

Figure 5: $ARFIMA(1, d, 1)$ for OSEBX

OSEBX	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-45667.94	-45679.75	-45681.49	-45684.01
p	0	0.07032	0	0.64621
d	0.0545129	0.01259	0.01210	0.20914
q	0	0	-0.07392	0.78831

Table 2: Estimation of $ARFIMA$ models for OSEBX

mate of d is much higher than for the three other models. But both the ACF and PACF plots shows that there are clearer autocorrelation in the residuals in this model than for the others, as seen in figure 5. From table 2, we can also see that the estimates of AR and MA lags are quite low for the first three models, but somewhat higher in the $ARFIMA(1, d, 1)$ model, which indicate that we can make some predictions of the returns based on the most recent returns.

The AIC implies that the better model is the $ARFIMA(1, d, 1)$ since this has the lowest value. Hovever, I would like to look at some more diagnostics for this model.

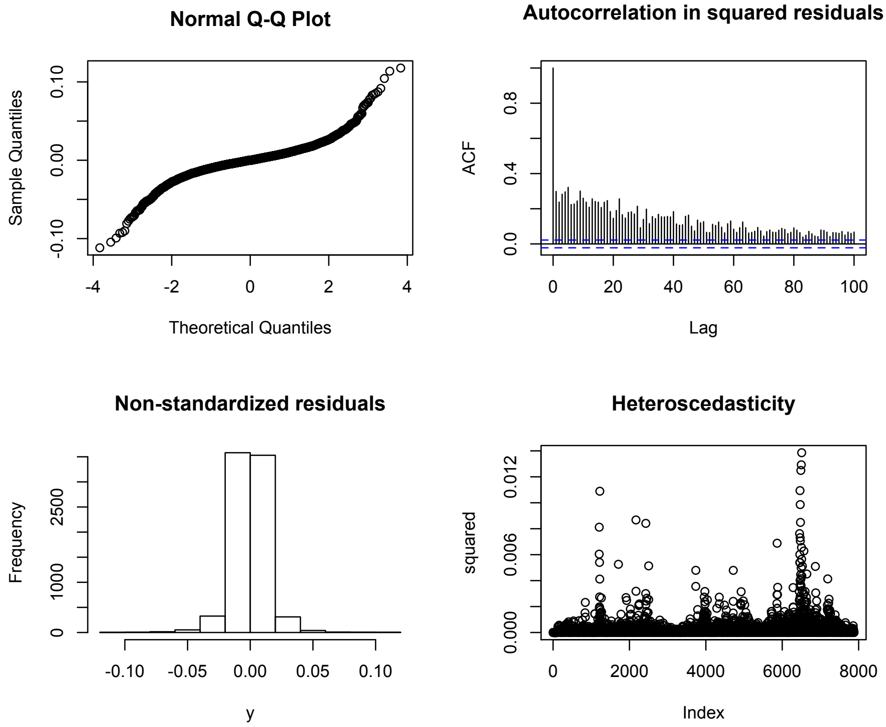

Figure 6: Diagnostics for OSEBX

In figure 6, I have only plotted graphs to do visual diagnostics of the model. One usually needs to use statistical tests to formally conclude, but the violations here are so obvious that this will not be necessary. Let us first look at the normal Q-Q plot and the histogram of the residuals, all from figure 6. The first should form a straight line if the residuals were normally distributed, and the histogram should have formed a bell like shape. The histogram shows clearly high kurtosis, in other words are there a lot more residuals close to 0 than what should have been the case if they were normally distributed.

There are clearly autocorrelation in the squared residuals. This is also known as Garch effect. This might tell us that we should try a variant of a $ARFIMA - GARCH$ model.

The squared residuals plotted against the independent variable do not look as they should either, which tells us that they are heteroscedastic. The residuals should have been scattered randomly, but there are several places where the residuals lie far away from the x-axis in clusters. We remember from the plot of the returns that there were several places where the volatility seemed higher than for the returns overall. The peaks for the residuals seem to be at the same places as these peaks in volatility for the returns. Our model has clearly not been able to account for these movements.

There are some evidence of long memory in the different models. But the estimates of d in the three first models are not high. In fact, for $ARFIMA(1, d, 0)$ and $ARFIMA(0, d, 1)$ the value of d is not even significantly larger than 0. As the models do not give a very good fit with the data either, I would be very careful about trying to interperet this as any kind of evidence that there are long memory in the OSEBX without further investigations. This means that it would be highly unlikely that we could find arbitrage opportunities large enough for earning money in a market with transaction costs.

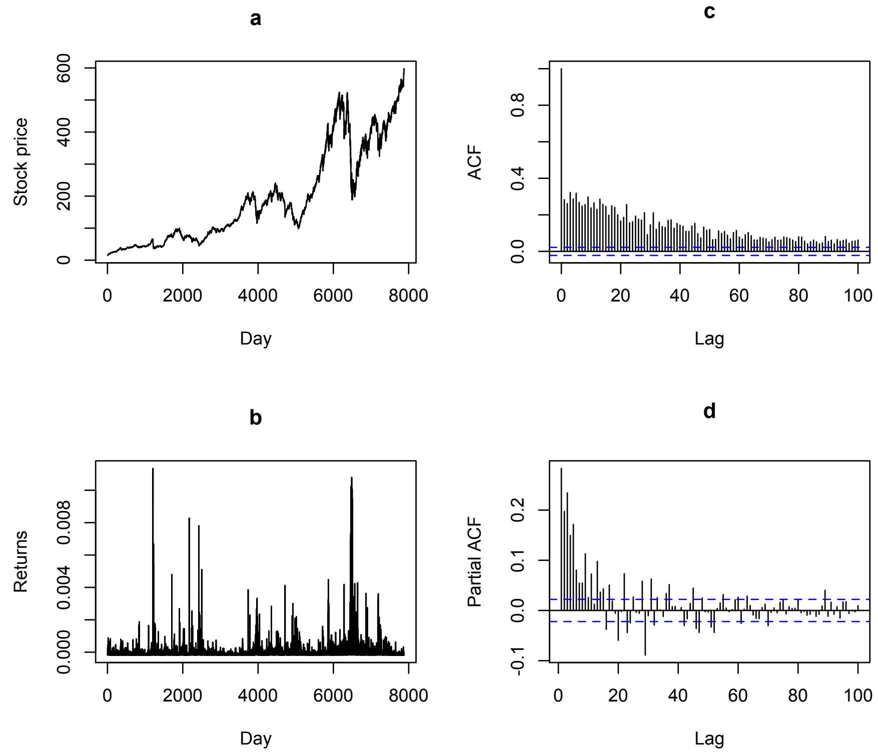

Figure 7: Description of squared OSEBX

10.3 Squared returns of the OSEBX

We do the same analysis for the squared returns for the index at the Oslo Stock Exchange. The initial squared returns plotted in figure 7 b shows the same peaks as the regular return did. The ACF in c looks more like the plots we have described for fractional time series than the ACF of the regular return did. The ACF plot decays very slowly over the 100 lags contained in this figure, and they keep decaying slowly after this as well. The PACF has a sharper drop.

By testing the same models as we did for the regular return, we find that the $ARFIMA(0, d, 0)$ seems to be the best fit, even though the differences are not big. This model has the same problems as the model we chose for

OSEBX^2	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-97400.12	-96954.7	-96953.64	-96952.69
p	0	-0.0773365	0	-0.0798912
d	0.2209035	0.2538745	0.2546283	0.2537898
q	0	0	0.0755728	0.0025993

Table 3: Estimation of models for squared OSEBX

the regular returns. Which means that the residuals are not normal or homoscedastic, and that there are still some autocorrelation left. The residuals also seems to have GARCH effects left, as the ACF of the squared residuals are quite high.

One explanation for the GARCH effect in the residuals, could be that the model does not account for the fact that investors may react different to good and bad news. It is known that volatility seems to rise when investors receive bad news, while it falls when they get good news which is found by for example Ding, Granger and Engle [16].

The biggest difference between the models estimated for the squared returns and the models estimated for the regular returns, is that the estimates of d for the squared returns are higher. They are always significantly different from 0 on a 95% confidence level or more, while they were not for the regular returns. This was not unexpected, and has been found by for example Grau-Carles [19] in several of the world's largest indexes.

Even though the $ARFIMA$ model is no where near an exact fit, it seems to be more long memory in the squared returns. The values estimated of d here would give an H that was more than 0.7. This could be used to predict future volatiltiy, but how well it could be done, can not be stated surely as we have not done any tests of how good predicitons would have been here. However, we see that it looks like we can be able to model and understand some of the volatility in the index when we assume that it has long memory.

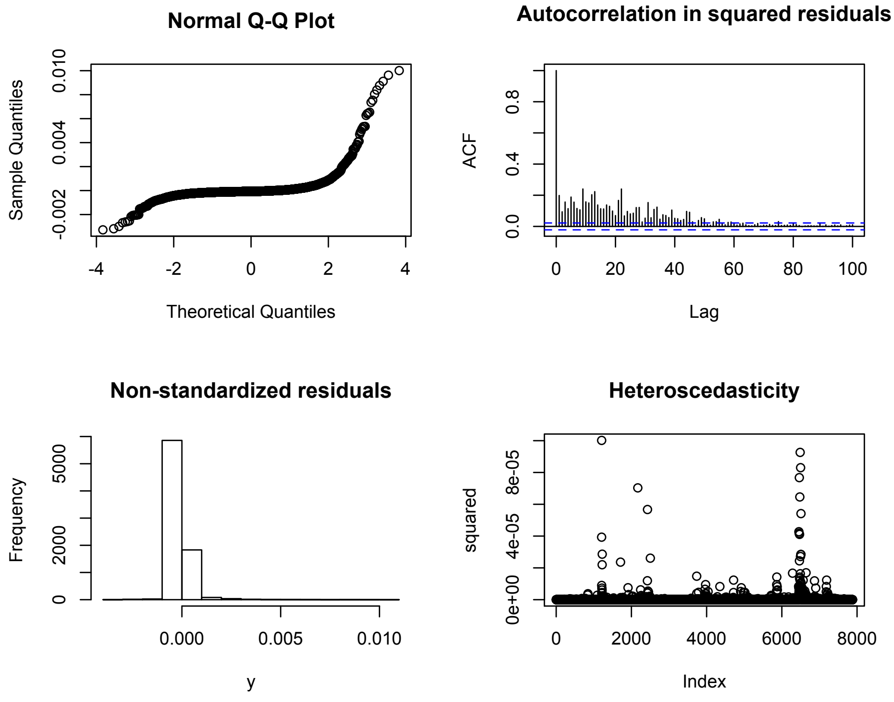

Figure 8: Diagnostics for squared OSEBX

We also note that the estimates of the AR and MA lags are relatively close to 0, and they are not always significant.

10.4 Returns of other stocks

I have then proceeded to look at ACF plots and tried to fit $ARFIMA$ models to several stocks in the Oslo Stock Exchange to find good examples of stocks both with and without long memory. Most stocks exhibit the same behavior. The autocorrelation is generally low, but as far as I can see, it may sometimes be described as slowly decaying. However, the estimates of d are generally close to 0 for a lot of the models, but there are still some variations within the same stock when different $ARFIMA$ models are fitted to the data. Most models are also clearly ill fitted, as the residuals have both autocorrelation, heteroscedasticity, GARCH effects and are non-normal.

An example of a stock that shows signs of long memory is the Norwegian Air Shutte (NAS).

The autocorrelation is low, as seen in figure 9, but at the same time, it does not seem to decay quickly. After trying severalt $ARFIMA$ models, the $ARFIMA(0, d, 1)$ seem to be the better fit from the ACF plots. It is however not the model with the lowest AIC. This model gives an estimate of d that is 0.03. This is not particularly high, but it is statistically significant according to the fracdiff package in R, which means it is not 0. This is the same as an fBm with $H = 0.53$.

The diagnostics for the chosen model is slightly better than for other stocks. There are still some autocorrelation in the squared residuals, and there are signs of heteroscedasticity. The normal plots are not too far off, even though they show some kurtosis. This is the case for all the stocks I have analysed, but in this case, it is less prominent than for other stocks.

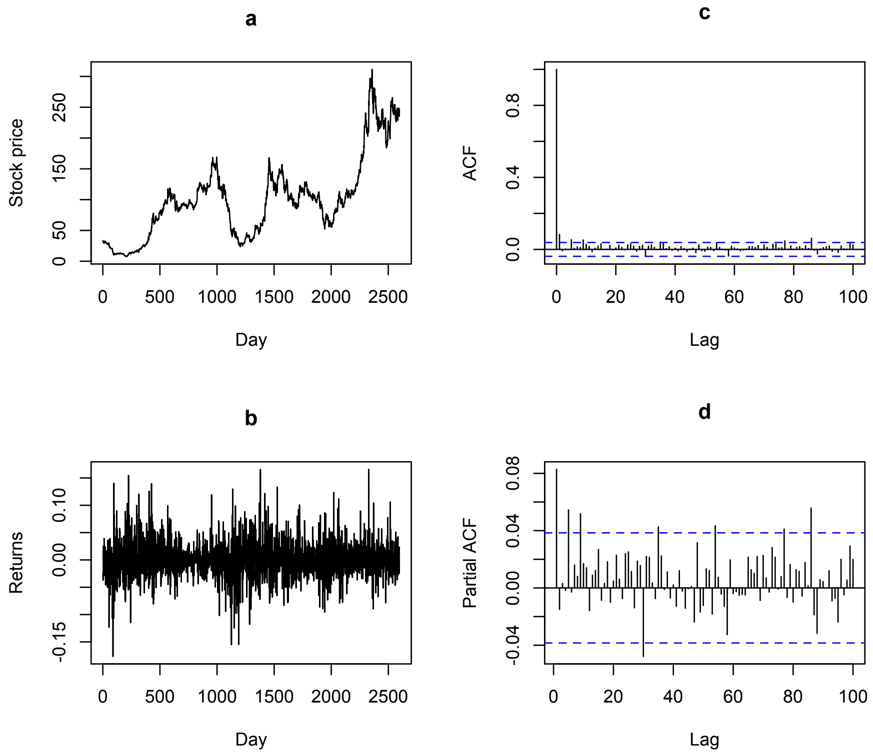

Figure 9: Description of NAS

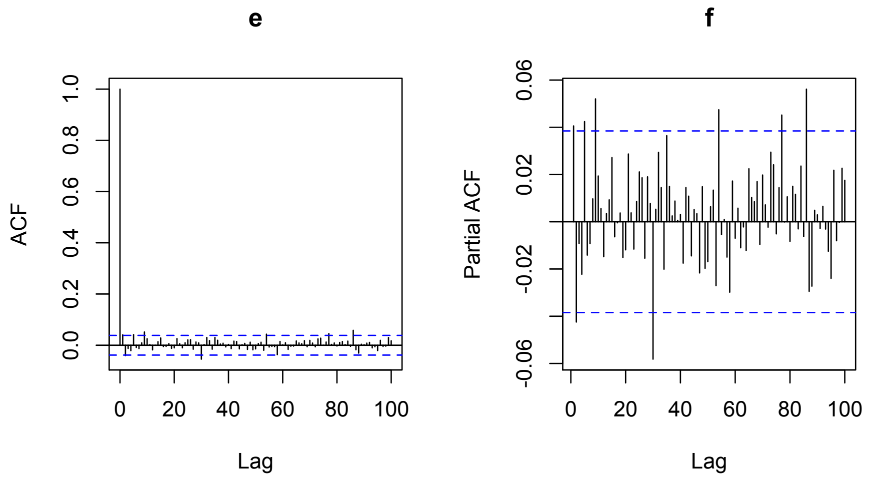

Figure 10: Autocorrelation of $ARFIMA(0, d, 1)$ for NAS

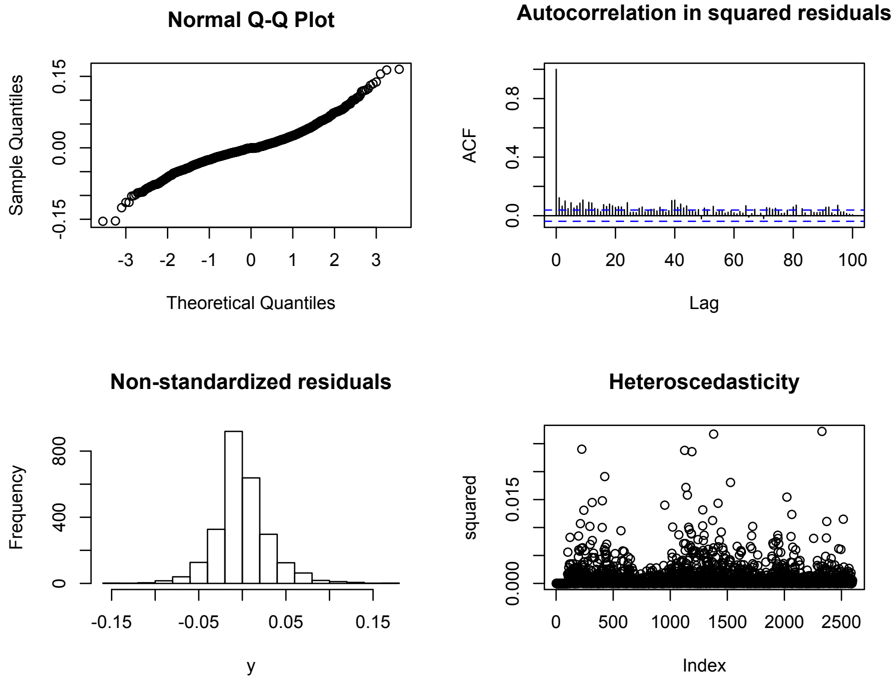

Figure 11: Diagnostics for $ARFIMA(0, d, 1)$ for NAS

Even though this stock seem to fit the chosen model pretty well, it could just be a coincidence. When one tries several stocks, one is bound to find something that fits the criteria just by pure chance. This is in no way a good enough reason to conclude that there is long memory on the Oslo Stock Exchange or in the stock NAS, but it could be.

From table 4, we can again see that the estimations of the AR and MA lags are low in the first three models, but higher in $ARFIMA(1, d, 1)$.

An example of a stock that do not fit the model well, and that do not exhibit any long memory tendencies is Northland Resources (NAUR). On the next few pages, we can see the plots used to determine which model to use. The autocorrelation in figure 12c looks like it is decaying a little faster

NAS	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-10545.61	-10544.85	-10545.14 [	-10548.7
p	0	0.04412	0	0.61842
d	6.254e-02	0.03707	0.03499	0.20171
q	0	0	-0.04895	0.75474

Table 4: Estimation of models for NAS

NAUR	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-4595.927	-4607.491	-4606.158	-4605.53
p	0	0.14690	0	0.17748
d	1.201e-01	0.02299	0.046250	0.01968
q	0	0	-0.117353	0.02761

Table 5: Estimated models for NAUR

than the example with NAS. From using both the AIC and the ACFs and PACFs of different models, I choose to use the $ARFIMA(1, d, 0)$ model.

In the chosen model, the estimate of d is significantly different from 0 on a 95% confidence level, but it only gives an H of 0.52, which means it does not have much long memory. The fit of the model is not good either. The autocorrelation of the regular residuals actually looks worse after differencing the data. The residuals are clearly not normally distributed. There are som autocorrelation in the squared residuals, which indicate GARCH effects. However, the heteroscedasticity plot is not the worst I have seen. But all in all, it does not look like any of the models tried on these data are a good fit.

In table 6 there is a collection of estimates of d for different $ARFIMA$ models for several stocks. The stocks are all from the Oslo Stock Exchange. For many of the stocks, d is estimated so close to 0 that I have set the value as 0 in the table. If we compare with the data table, we see that the stocks where we have estimated a value of d that is 0, we only have data from a cou-

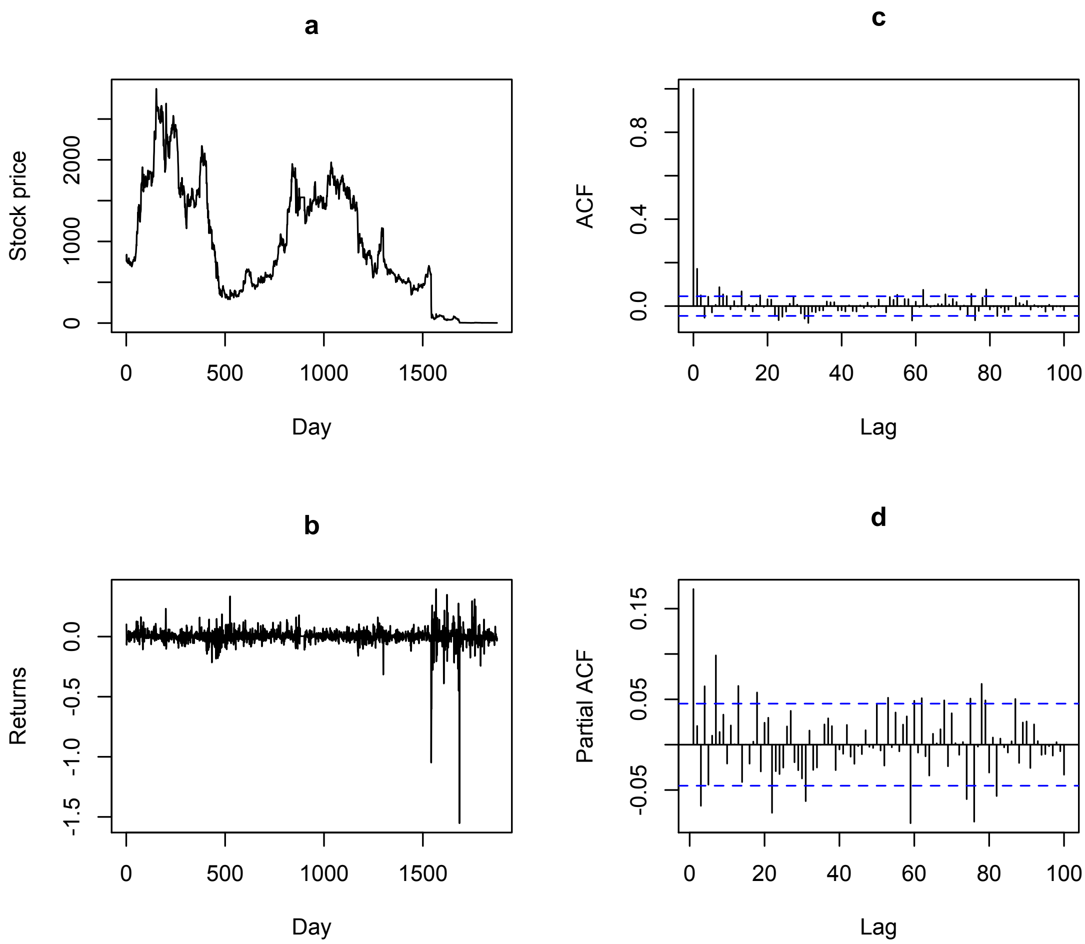

Figure 12: Description of NAUR

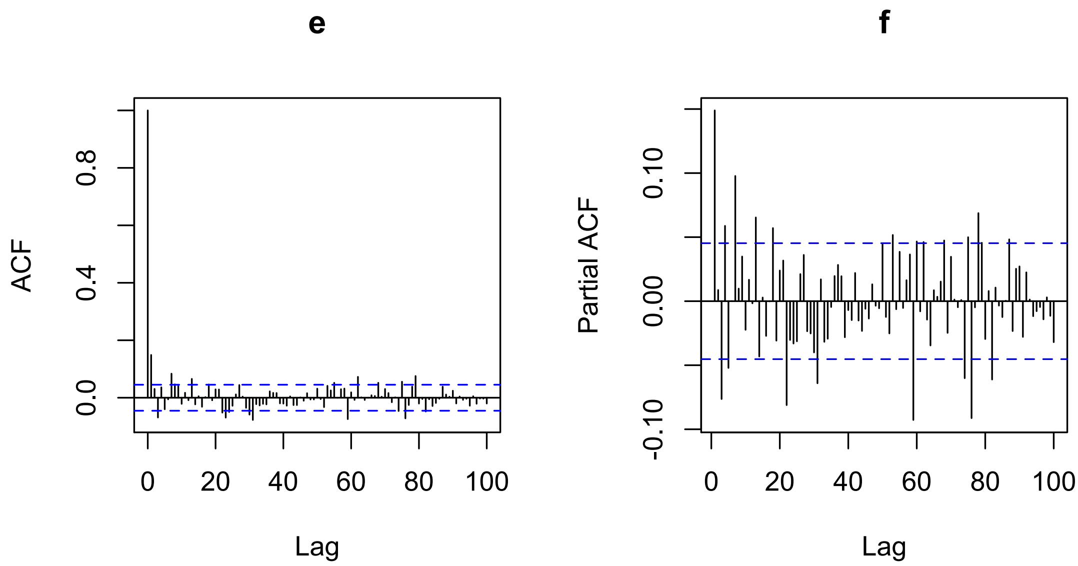

Figure 13: Autocorrelation of $ARFIMA(1, d, 0)$ for NAUR

Ticker	Best AIC	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AUSS	(0, d, 0)	0.04436	0.05510	0.05429	0.11958
CECON	(0, d, 1)	0	0	0	0
ECHEM	(1, d, 1)	0	0	0	0
FOE	(0, d, 1)	Unable	0.03348	0.04255	0.05158
GOGL	(1, d, 1)	0	0	0	0.22981
HAVI	(0, d, 1)	0	0	0.03956	0.04494
NAUR	(0, d, 1)	0.1201	0.02299	0.046250	0.01968
NAS	(1, d, 0)	0.06245	0.03707	0.03499	0.20171
PLCS	(0, d, 0)	0.0217881	0	0	0.19106
PROS	(0, d, 0)	0	0	0	0
SN I	(0, d, 0)	0.01434	0.02085	0.02117	Unable
TIDE	(1, d, 0)	0	0	0	0
VIZ	(1, d, 0)	0	0.01810	0.02034	0.01423
WWASA	(0, d, 0)	0	0	0	0

Table 6: Estimated values of d for all chosen stocks

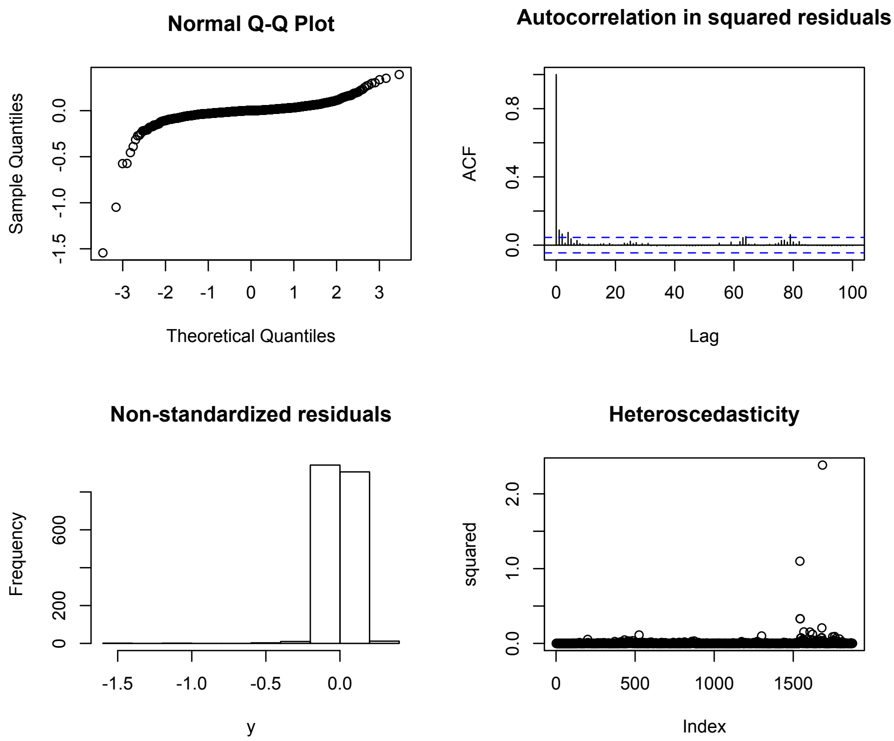

Figure 14: Diagnostics for $ARFIMA(1, d, 0)$ for NAUR

ple of years back, and this could mean that these stocks would have higher estimates of d later when more data is provided.

In two cases, market "unable" in the table, fracdiff could not provide an estimate for the given model inputs. When testing models with more lags from AR and MA this happend frequently.

There are a few cases, especially for the $ARFIMA(1, d, 1)$ where the estimates of d are pretty high. For example for GOGL, NAS and PLCS. Only for GOGL had this model the lowest AIC value. As none of the estimates of d in the other models for these stock come close to the value of d in $ARFIMA(1, d, 1)$, it might get us to believe that these estimates do not provide a good description of the stock returns.

Most of the estimates of d are between 0 and 0.055. With such low estimates of d, it is difficult to see how one could use this in a trading strategy to make money when one accounts for the cost of buying and selling. But how an arbitrage strategy in a market with transaction costs would be implemented, is not discussed here. The estimates are close to what is found for stocks in other exchanges [26] [19] [22].

This is what we expected to find, as there would not be market efficiency at the exchange if there were long memory in the stock returns. With this in mind, we should remember the article mentioned by Hegnar [21] where he wrote that it is possible to beat the market. When we find estimates of H that are close to $\frac{1}{2}$, the market behaves efficiently, and there should not exist easy ways to beat the market without taking on extra risk.

10.5 Squared returns of the stocks

By using the same method as before. I have analysed the squared returns of NAS and NAUR and chosen models for them. From the plotted autocorrela-

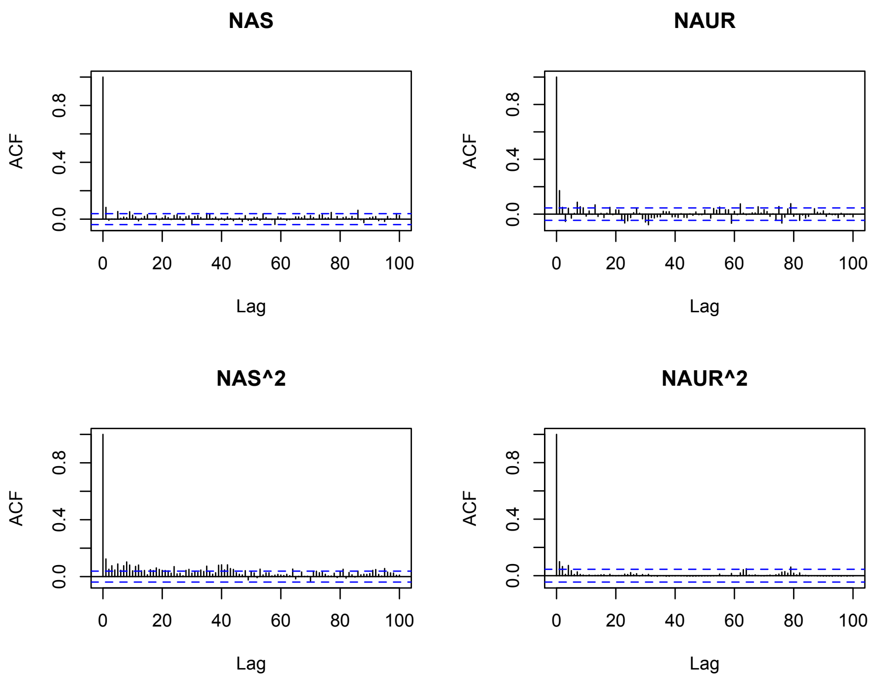

Figure 15: Autocorrelations for returns and squared returns for NAS and NAUR

tions in figure 15, it is fairly easy to see that the squared returns have a more prominent form of the slowly decaying pattern described in autocorrelations for long memory time series. As was also the case with OSEBX.

Both the squared returns of NAUR and NAS seem to fit best with the $ARFIMA(0, d, 0)$. This came both from looking at the autocorrelations of the residuals and from the AIC values. The estimations of d do not vary greatly between the different models shown here, and they are all higher than in the models for the regular returns. Again does this point out that there might be better chances of finding long memory in the stocks' volatility than in the returns.

NAUR^2	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-5050.318	-5046.276	-5046.276	-5044.628
p	0	0.002313	0	0.70633
d	0.08987	0.088556	0.088636	0.05217
q	0	0	-0.002184	0.67003

Table 7: Estimated models for squared returns of NAUR

NAS^2	(0, d, 0)	(1, d, 0)	(0, d, 1)	(1, d, 1)
AIC	-24324.72	-24242.91	-24242.92	-24244.7
p	0	0.027872	0	-0.756092
d	0.1183877	0.102905	0.103147	0.106662
q	0	0	-0.027735	-0.785776

Table 8: Estimated models for squared returns of NAS

11 Implications for investors

I found that it is very litte, if any, long memory in the returns in the Norwegian stock market, and investors have to decide if it is enough to use in an arbitrage strategy. They have to consider the transaction costs, limits to which stocks they can trade and how often. They also need to consider if the return from this is high enough to cover the implicit cost of using the time of an analyst to find the best portfolios to use.

In chapter 7, I showed an example of a portfolio from Shiryaev [32]. This only consisted of two assets. In a real market, there are several stocks, and investors need to consider how to put this mini-portfolio together with their other assets.

In the introduction, I mentioned an article by Hegnar [21]. He wrote about investors that beat the market by using simple strategies based on key ratios. In the last decades, several traders have started to trade more often, and this should eliminate most of the simple strategies one could use to find arbitrage. I believe that it would be harder for an investor to use trading strategies based on long memory, as the theory behind is difficult to understand. This might mean that it could be easier for the traders who use it to actually benefit from the little long memory that has been found in these stocks, but it could also mean that the profit from using this is smaller because it takes time to fully understand and implement the strategies.

Long memory in the volatility, as I have found in the Norwegian stock market, could be particularly important to understand for investors managing large hedge funds. Hedge funds have a business model that heavily relies on being able to predict risk correctly, in order to assemble portfolios that yield high returns with very low portfolio risk. Long memory in the squared returns would help an investor understand the risk associated with the stocks he buys if the model gives good out-of-sample predictions for volatility. It could help prevent large losses, and for a hedge fund, only a minor improvement in the understanding of risk can yield large returns.

12 Limitations and future work

It is important to know which weaknesses and limitations the fBm gives us. The first part of this paper discussed the properties of fBm, and then we looked at how this would fit with known financial theory. Our main problem was that there was possible to create arbitrage portfolios in different markets based on the fBm movements. Even though it was mentioned that these opportunities would disappear with limitations on the frequency of trades and by the transaction costs, this still creates a problem as it violates the basis of today's finacial theory.

We also need to see if we find evidence of this kind of behavior in reality. I have tried to give some examples of its practical applications here, but the results were in no way only in favour of the fBm.

The stocks tested are not selected at random, and I have not tested all the stocks in the Oslo Stock Exchange either, which means that we can not conclude that all stocks would behave in the same way.

From the code written in the program R, there might be misspellings that could have affected the results, even though I have done my best to assure that this did not happen.

We need to assess possible problems with the methods used as well. It is not a given that I have chosen the best fitting $ARFIMA$ models for each case, as the selection method is subjective. I have also kept the testing to parsimonous models. Other methods used to estimate d for the $ARFIMA$ models could also have given slightly different results.

It is also not a given that the $ARFIMA$ is the best model to use for mdoelling the returns and squared returns in the stock market. Other models with varieties of the GARCH model is an example we could have tried. This model is especially relevant because we saw tendencies of GARCH effects in the residuals.

Some of the series used seemet to be so short that we had troube finding any signs of long memory in them at all. However, especially for the OSEBX, it would be possible to split the data into groups and estimate different models for all of them separately.

For future work, I would test all of the stocks at the exchange if possible, or choose a random set. The estimations could also be done with several methods, but the most important change would be to explore the possibility of using $ARFIMA-GARCH$ models in the Oslo Stock Exchange.

Another interesting topic would be to investigate if there would be any possibilities of making money from estimating these models and implementing arbitrage strategies in a market with transaction costs even though none of the stocks seem to have much long memory. I would also like to see if we could use the long memory in the squared returns as a tool in risk management. If this is the case, it could help investors reduce losses in their portfolios.

13 Concluding remarks

From mathematical theory about fractional Brownian motion, we saw that if stock returns behaved as if they were fBm, this would violate the efficient market hypothesis. Or in other words, it would be possible to make money in the market through arbitrage strategies.

I showed that an $ARFIMA$ model could be used as a discrete way of modelling both short and long memory in stock returns. A high estimate of d would mean that the model had much long memory, and the more long memory a stock return had, the greater was the possibility to earn money from arbitrage strategies.

Then I estimated $ARFIMA$ models for several stock returns in the Oslo Stock Exchange. The estimated models were not a pefect fit for the stock returns in any of the cases. Most of the estimates of d were relatively low, and they mostly gave us an H between 0.5 and 0.55. This indicates some long memory in the stock returns, but not much. It would probably not be possible to beat the market by using this model to find arbitrage strategies when accounting for costs of trading, but this is not investigated further in this thesis.

However, we did find long memory in the volatility of the stocks, which could help us to understand and predict risk better.

14 Appendix

These are the main parts of code used to estimate models.

```
# READ DATA
yar=read.table("http://ichart.yahoo.com/table.csv?s=ECHEM.OL",sep=",",header=T)


#Remove days with no trading
 yar <- yar[yar$Volume!=0,]

#Use only column with adjusted close data
p=yar[,"Adj.Close"]

#Reverse order of data
p=p[length(p):1]
na.omit(p)



#Choose alternative 1 or 2:
#ALT 1: Log returns
ret=diff(log(p))

#ALT 2: Squared log returns
ret=(diff(log(p))^2)

#Remove mean
ret=ret-mean(ret)



par(mfcol=c(2,2))
#Show time series plot
ts.plot(p, ylab="Stock price", xlab="Day", main="a")
ts.plot(ret, ylab="Returns", xlab="Day", main="b")
#ACF and PACF
acf(ret, main="c", lag.max=100)
pacf(ret, main="d", lag.max=100)


#FIT ARFIMA MODEL
arf1=fracdiff(ret, nar=0, nma=0)

summary(arf1)

#Find residuals of arfima model
n=length(ret)
L=100
d=arf1$d
```

```
fdc=d
fdc[1]=fdc
for (k in 2:L) fdc[k] = fdc[k-1]*(d+1-k)/k
y= rep(0,L)
for (i in (L+1):n){
        csm=ret[i]
        for (j in 1:L) csm= csm + ((-1)^j)*fdc[j]*ret[i - j]
        y[i]= csm
}
par(mfcol=c(1,2))
acf(y, main="e", lag.max=100); pacf(y, main="f", lag.max=100)

#Diagnostics
#Normality plots of residuals
par(mfcol=c(2,2))
qqnorm(y); hist(y, main="Non-standardized residuals")
#ACF of squared residuals
squared=y^2
acf(squared, main="Autocorrelation in squared residuals", lag.max=100)
#Homoskedasticity
plot(squared, main="Heteroscedasticity")
```

References

[1] Abraham, B., Lopes, S., Reisen, V. (2001) Estimation of parameters in ARFIMA processes: a simulation study, Communications in Statistics, vol. 30, no. 4, pp. 787-803.

[2] Baum, C. F (2013) ARIMA and ARFIMA models, Boston College: EC823.

[3] Bayraktar, E., Poor, H. V., Sircar, K. R. (2004) Estimating the Fractal Dimension of the S &P Index using Wavelet Analysis, International Journal of Theoretical and Applied Finance, vol. 7, no. 5, p 615.

[4] Bender, C., Sottinen, T., Valkeila, E. (2008) Pricing by hedging and no-arbitrage beyond semimartingales, Finance and Stochastics, vol. 12, no. 4, pp. 441-468.

[5] Beran, J. (1994) Statistics for long-memory processes, New York: Chapman & Hall.

[6] Biagini, F., Hu, Y., Øksendal, B., Zhang, T., (2007) Stochastic calculus for fractional Brownian motion and applications, Berlin: Springer.

[7] Bingham, N. H., Kiesel, R. (2010) Risk neutral valuation, 2nd edition, London: Springer.

[8] Contreras-Reyes, J. E., Palma, W (2013) Statistical Analysis of Autoregressive Fractionally Integrated Moving Average Models, Computational Statistics, vol. 28, no. 5 , pp 2309-2331.

[9] Cowpertwait, P. S. P., Metcalfe, A. V. (2009) Introductory time series with R, Dordrecht: Springer.

[10] Bjørk, Tomas; Hult, Henrik (2005) A note on Wick products and the fractional Black-Scholes model, SSE/EFI Working Paper Series in Economics and Finance, no. 596.

[11] Cheridito, P. (2001) Mixed fractional Brownian motion, Bernoulli, vol. 5, no. 6, pp 913-934.

[12] Cheridito, P. (2003) Arbitrage in Brownian motion models, Finance and Stochastics, vol. 7, no. 4, pp 533-553.

[13] Decreusefond, L., Ustunel, A. S. (1999) Stochastic Analysis of the Fractional Brownian Motion, Potential Analysis, vol. 10, no. 2 , pp 177-214.

[14] Deelstra, G., Delbaen, F. (1998) Existence of Solutions of Stochastic Differential Equations related to the Bessel process, [Online], `Available:http://www.math.ethz.ch/~delbaen/ftp/preprints/existentie_rev.pdf`, [14 Apr 2014].

[15] Dieker, T. (2004) Simulation of fractional Brownian motion, [Online] Available: `http://www.wstein.org/home/wstein/www/home/simuw/simuw08/refs/fractal/dieker-2004-thesis-simulation_of_fractional_Brownian_motion.pdf`, [14 Apr 2014].

[16] Ding, Z., Granger, C. W. J., Engle, R. F. (1993) A long memory property of stock market returns and a new model*, Journal of Empirical Finance, vol. 1, pp. 83-106.

[17] Enders, W. (2010) Applied econometric time series, Hoboken: John Wiley and sons.

[18] Granero, S., Segoivia, T., Perez, G. (2008) Some comments on Hurst exponent and the long memory processes on capital markets, [Online], Available: http://www.ual.es/~jgarcia/index_archivos/HURST.pdf, [14 Apr 2014]

[19] Grau-Carles, P. (2000) Empirical evidence of long-range correlations in stock returns, Physica A, vol. 287, pp. 396 404.

[20] Haslett, J., Raftery, A. E. (1989) Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power, Journal of the Royal Statistical Society. Series C (Applied Statistics), vol. 38, no. 1, pp. 1-50.

[21] Hegnar, T. (2014) Best med Buffet, Kapital, vol. 10, p. 12.

[22] Henry, O. T (2002) Long memory in stock returns: some international evidence, Applied Financial Economics, vol 12, pp 725-729.

[23] Hosking, J. R. M. (1981) Fractional differencing, Biometrika, vol. 2, no. 1, pp 165-176.

[24] Kantelhardt , J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., Bunde, A (2001) Detecting Long-range Correlations with Detrended Fluctuation Analysis, [Online], Available: http://arxiv.org/pdf/cond-mat/0102214.pdf?origin=publication_detail, [14 Apr 2014].

[25] Lalley, S. (2013) Gaussian processes; Kolmogorov-Chentsov theorem, [Online], Available: http://galton.uchicago.edu/~lalley/Courses/385/GaussianProcesses.pdf [14 Apr 2014].

[26] Lopez Herrera, F., Ortiz, E., de Jesus, R. (2012) Long Memory Behavior in the Returns of the Mexiacan Stock Market: Arfima Models and Value at Risk Estimation, vol. 2, no. 10.

[27] Mandelbrot, B., Hudson, R. L. () The misbehavior of markets: A fractal view of financial turbulence, New York: Basic Books.

[28] Mandelbrot, B., Van Ness, J. W. (1968) Fractional Brownian motion, fractional noises and applications, SIAM Review, no. 10, pp 422-437.

[29] Norros, I., Valkeila, E., Virtamo, J. (1999) An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions, Bernoulli, vol. 5, no. 4, pp. 571-760.

[30] Nualart, D. (2010) Stochastic integration with respect to fractional Brownian motion and applications, [Online], Available: `http://www.mat.ub.edu/~nualart/fbm.pdf`, [14 Apr 2014] .

[31] Pollock, D. S. G, (1993) Arima identification, lecture 8

[32] Shiryaev, A. N. (1998) On arbitrage and replication for fractal models, In A. Shiryaev and A. Sulem (editors): Workshop on Mathematical Finance, INRIA, Paris.

[33] Solem, R. (2014) Bli best med Buffet, Kapital, vol. 10, pp. 130134.

[34] Taqqu, M. S. (2013) Benoit Mandelbrot and Fractional Brownian Motion, Statistical Science, vol. 28, no. 1, pp. 131134.

[35] Velasco, C. (1998) Non-stationary log-periodogram regression, Journal of Econometrics, vol. 91, no. 2, pp. 325-371.

[36] Wall Street Jojurnal (2014) Dow Jones Industrial Average All-Time Largest One Day Gains and Losses, [Online], Available: `http://online.wsj.com/mdc/public/page/2_3047-djia_alltime.html`, [14 Apr 2014].

[37] Wikipedia (2014) Fractional Brownian motion, [Online], Available: http://en.wikipedia.org/wiki/Fractional_Brownian_motion, [14 Apr 2014].

[38] Yang, B. (2008) LRD of Fractional Brownian Motion and application in Data Network, [Online], Available: http://math.arizona.edu/~flaschka/Topmatter/527files/termpapers/bole.pdf [14 Apr 2014].

[39] Yin, Z. M (1996) New Methods for Simulation of Fractional Brownian Motion, Journal of computational physics, no. 127, pp. 6672.

[40] Øksendal, B. (1997) An introduction to Malliavin calculus with applications to economics, Dept. of Mathematics, University of Oslo.