Aalborg University

10th semester

Mathematics-Economics

master thesis

Detecting Financial Bubbles in a High Frequency Setup Using Neural Networks

June 3, 2022

AALBORG UNIVERSITET
STUDENTERRAPPORT

Fifth year w/
Department of Mathematical Sciences
Mathematics-Economics
Skjernvej 4A
9220 Aalborg Ø
http://www.studerende.math.aau.dk/

Title:

Detecting Financial Bubbles in a High Frequency Setup Using Neural Networks

Project:

Master Thesis

Project period:

1st of February 2022 - 3rd of June 2022

Participants:

Jonas Mikkelsen

Jonas Sveistrup Mikkelsen

Supervisor:

Orimar Sauri Arregui

Abstract:

This thesis aims to study financial bubbles in a continuous time setup. Our contribution to this field of research is to use artificial neural networks, and by use of them, develop a non parametric estimator of the functional form of the spot variance. The goal will be to determine whether neural networks can be used to detect financial bubbles, and at last, build a profitable trading strategy.
We will first present relevant theory regarding bubbles, and then powerful theorems that will be crucial in order to detect them. For our contribution to work, we will need a pointwise estimator of the spot variance process. Aftwards, we will use neural networks to learn the function that maps the price to the spot variance. The network will then be used to decide whether the price process is a bubble or not. Furthermore, both the spot variance estimator and the network method we developed will be tested in a simulation study, before being applied to real data in form of the S&P 500 ETF, SPY.
It is concluded that we could detect financial bubbles in a simulation study with our own proposed method. However, when applied to real data, our method could not detect any bubbles. On the other hand, investigating the 10 dates that, by a condition, are the closest to being bubbles, it is concluded that 9 of the days were during the housing bubble, whilst the last one was during the "Dot com" bubble. Lastly, we build a trading strategy based on shorting according to our methods, and showed that it outperformed the buy and hold strategy over the same time span.

Page count: 34

Preface

The group would like to give a special thanks to:
- *Orimar Sauri Arregui, supervisor, for help, advice and guidance throughout the duration of the project.*

The group has used the following programs in the writing of this report

- **Overleaf** - Writing.
- **Mendeley** - Sorting of literature.
- **Google drive** - Notes and file sharing.
- **R** - Statistical calculations and data analysis.
- **Github** - Version control.

Reader's Guide:

On page v, a table of contents is given. When viewing this report as a PDF, hyperlinks in the table of content will allow fast navigation to the desired section.

The bibliography on page 34 presents the literature used in this report. The references are given in the following format:

[Author][Title](Institution)(ISBN)[Year](URL)(Date Accessed)

Where fields in [square brackets] are mandatory, while regular parenthesis only are relevant for certain formats (e.g. books or web pages). The bibliography entries are sorted alphabetically.

Contents

Chapter 1

Introduction

The task of determining the fundamental price of an asset is a difficult one. Assuming the fundamental price can be determined, a plausible investment strategy would be to short sell overvalued ones, and buy the undervalued ones, i.e. short sell assets where the fundamental price exceeds the current market price and vice versa. Thus it is of great interest to determine fundamental prices in order to risk manage and create trading strategies.

The phenomenon of the price of an asset trading above its fundamental price is typically referred to as a bubble. The reason for this terminology comes from the idea that if an asset is currently trading above its fundamental price, the asset price should at some point decrease such that it once again trades at its fundamental value, i.e. the bubble bursts. Bubbles have been experienced several times throughout history, starting as early as in year 1636, where the tulip mania occurred [Ross, 2020]. In more recent time, a perfect example of a bubble is the "Dot com" bubble which took place in the late 1990s. This can be seen in Figure 1.1.

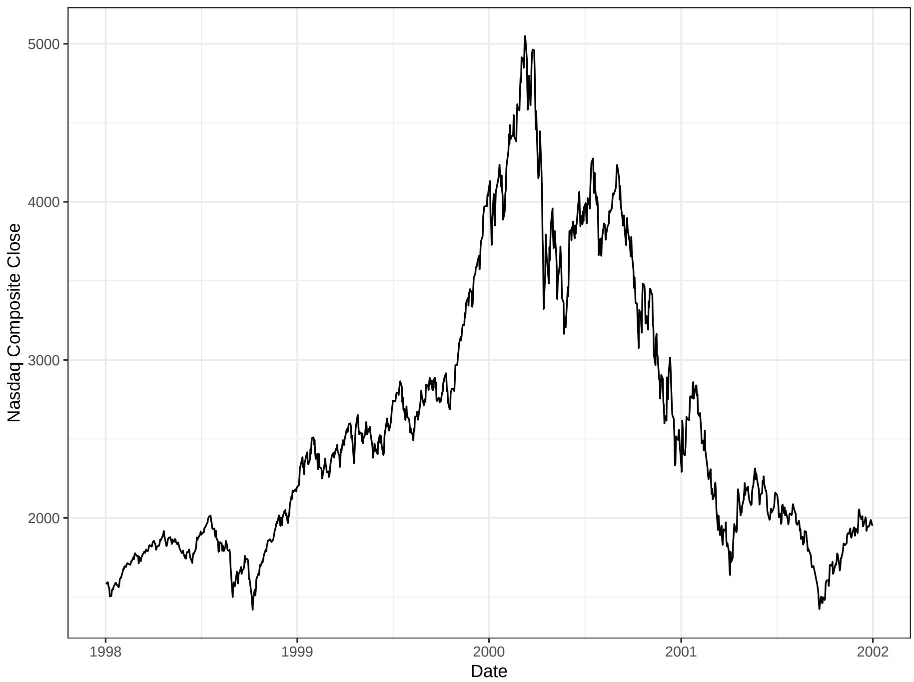

Figure 1.1. Daily close values of the Nasdaq Composite Index from 1st of January 1998 to 1st of January 2002. The prices are obtained from Yahoo finance.

From 1998 to 2002, the Nasdaq index experienced an enormous increase in value, tripling from early 1998 to 2000, before declining to levels similar to the ones in 1998. Due to the volatile behaviour of bubbles, starting with a rapid increase, followed by an enormous decrease, from a rational investors perspective, it would be of interest to determine whether an asset owned by an investor is currently in a bubble, because the investor would be able to either sell or short sell the asset, depending on risk willingness. Later, once the bubble has burst, the investor can take a long position at a cheaper price, or close a short position. In the example of the Nasdaq Composite Index, this would have been beneficial. Consequently, knowing that an asset is currently in a bubble is very helpful when managing risk, but how does one determine whether an asset is in a bubble, or an increase in price is justified?

This thesis aims to study bubbles with mathematical theory in a high frequency setup, and afterwards apply neural networks to see whether they can improve the traditional methods of detecting bubbles. This has already been studied in [Bashchenko and Marchal, 2020], but they used a sampling frequency of 2 minutes for their data, while simultaneously assuming no microstructure noise.

Our contribution to this problem is to increase the sampling frequency and take microstructure noise into account. We will follow an approach similar to that of [Bashchenko and Marchal, 2020], but we will not be training a neural network to classify bubbles, but instead use neural networks for regression purposes, and later use them in the key component of detecting financial bubbles.

For this setup to work, we will first need to specify some market dynamics, and in this framework, present powerful theorems, which will yield a concrete way to determine whether the price process of an asset is a bubble or not. In the setup we will present, it turns out that one needs to know the functional form of the spot variance in order to determine whether an asset is a bubble. However, in practice, this quantity is unknown, and thus has to be estimated. This is where we mainly differ from existing research such as [Jarrow et al., 2011]. They estimate the functional form of the spot variance, whereas we will proceed by first conducting point estimation, and when they are obtained, use neural networks to approximate the functional form of the function that maps the price level to the spot variance. Therefore, our method is a non parametric estimation of the spot variance. We will only investigate feedforward artificial neural networks throughout the thesis. These type of networks will be referred to as ”neural networks“, ”networks“ or similar in the report.

During a simulation study, it will be tested how the point estimator of the spot variance performs, and then afterwards, how well our neural network method performs on simulated prices which we know to be bubbles, and others price where we know they are not. We will then compare the performance of our neural network method to a simple case where we fit a linear model to the same point estimates.

Lastly, the method will also be applied to intraday SPY data, an S&P ETF, in order to see whether it also works in practice. Using the results obtained, we will try to develop a simple trading strategy that uses our method. This leads to the problem statement.

How can neural networks be applied to detect financial bubbles in a high frequency setup under the assumption of microstructure noise?
How can one build a trading strategy on SPY based on detecting financial bubbles with our proposed method?

Chapter 2

Prerequisites

This chapter is mainly based on [Jarrow et al., 2010].

In order to study bubbles, it is important to have a precise mathematical definition of this phenomena. The goal of this chapter is thus first to specify the dynamics of a market in order to define what a bubble mathematically is. Later, a test for bubbles in this framework will be presented. As will become clear during the chapter, the test relies on the functional form of the spot variance of the underlying price process. In practice, this form is unknown, so we will need to estimate it. However, we instead introduce a pointwise estimator of the spot variance and use neural networks to approximate the functional form from the point estimates.

2.1 The Setup

Initially, we begin with a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{0\leq t\leq T}, \mathbb{P})$ that satisfies the usual assumptions of right continuity and completeness. It is assumed that only a single risky asset and a money market account is traded in the economy. The money market account will also be a numéraire with a constant price of 1, i.e. the value of cash does not change throughout time. It should be noted that all prices and cash flows covered later is relative to the money market account, but since this is constant 1, it will not be written explicitly in the equations.

Let τ be a stopping time that represents the life of the risky asset, and let $D = (D_t)_{0\leq t<\tau}$ be an adapted process which is a càdlàg semimartingale, representing the cumulative dividend process of the risky asset. Furthermore, let $X_\tau \in \mathscr{F}_\tau$ be the liquidation value of the asset at time τ. It will be assumed that $X_\tau \geq 0$ and $D \geq 0$.

Continuing with the risky asset, the market price is given by the non-negative càdlàg semimartingale $S = (S_t)_{0\leq t\leq\tau}$. Related to the risky asset is the wealth process associated to it. This process is denoted by W and defined by

$$W_t = S_t + \int_0^{t\wedge\tau} 1 \, \mathrm{d}D_u + X_\tau \mathbf{1}_{\{\tau\leq t\}}, \tag{2.1}$$

that is, the wealth process is the market value of the risky asset plus the accumulated dividends. In the case of $t \geq \tau$, the liquidation value is added too. Because the risky asset does not exist in case $t > \tau$, all processes will be stopped at the stopping time τ, such that the interval $[0, \tau]$ represents the interval where trading the risky asset is possible. This also implies that $\mathscr{F} = \mathscr{F}_\tau$.

2.1.1 Market Dynamics

Before defining a bubble mathematically, some terms regarding the market has to be presented. First, a strategy is defined to be a pair of adapted processes (π, η) representing the amount of shares held in the risky asset and the money market account. The corresponding wealth process is then given by

$$V_t^{\pi,\eta} = \pi_t S_t + \eta_t.$$

A set of strategies we will need are the admissible strategies. They are defined in the following way.

Definition 1: Admissible Strategies

A strategy (π, η) is said to be admissible if

1. It is self financed: The change on the value of the strategy only depends on the change of the prices;

$$\mathrm{d}V_t^{\pi,\eta} = \pi_t \mathrm{d}S_t.$$

2. It has finite credit line; there exists a non random constant $C > 0$ such that

$$V_t^{\pi,\eta} > -C, \quad t \in [0, \tau],$$

almost surely. The set of all admissible strategies will be denoted $\mathcal{A}$.

The admissible strategies are thus strategies where no external injection of cash takes place due to it being self financed, as well as being bounded below by a non random constant, i.e. there is a limit to how much one can borrow. Admissible strategies are of interest because they are used when defining an arbitrage free market. It will be assumed throughout this report that the market satisfies No Free Lunch with Vanishing Risk, NFLVR. This means that there are no strategies in the market with all the following properties:

1. They are self financing
2. They have zero initial capital;

$$V_0^{\pi,\eta} = 0, \quad (\pi, \eta) \in \mathcal{A}.$$

3. They generate non-negative cash flows;

$$V_t^{\pi,\eta} \geq 0 \ a.s., \quad (\pi, \eta) \in \mathcal{A}, \quad t \in [0, \tau].$$

4. They generate strictly positive cash flows with positive probability;

$$\mathbb{P}(V_t^{\pi,\eta} > 0) > 0, \quad (\pi, \eta) \in \mathcal{A}, \quad t \in [0, \tau].$$

Moreover, there are no sequence of trading strategies that approaches any strategies with the properties above. More details regarding NFLVR can be found in [Jarrow et al., 2010]. The first fundamental theorem of asset pricing can then be used to determine whether a market satisfies NFLVR or not.

Theorem 2: First Fundamental Theorem of Asset Pricing

A market satisfies NFLVR if and only if there exists a local martingale measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that the discounted wealth process is a $\mathbb{Q}$-local martingale.

Proof. Omitted. □

In case there exists a unique local martingale measure, the market is called complete. However, if we were to make the assumption about a complete market, it will imply that a price process will be either a bubble or not for all $t \in [0, \tau]$. Since this is unrealistic, we will assume that the market is incomplete, i.e. a unique local martingale measure is not present.

2.2 The Fundamental Price and Bubbles

Because the market dynamics has been specified, defining the fundamental price of the risky asset is now possible.

We will continue in the same framework as in Section 2.1.1, such that all assumptions made in the span of that section will be assumed to hold here as well. Let $\mathcal{M}_{loc}$ denote the set of equivalent local martingale measures. By assumption of an incomplete market, $|\mathcal{M}_{loc}| \geq 2$, where $|.|$ denotes cardinality. It is of interest to use the assumption of an incomplete market in order to make it possible to change which risk neutral measure is used in the risk neutral pricing. This will be done in the following way.

First, let $(\kappa_i)_{i\geq 0}$ denote an increasing sequence of random stopping times, where $\kappa_0 = 0$. Next, let $(Y^i)_{i\geq 0}$ be a sequence of random variables independent of κ for each i. It will be assumed that both κ and Y are independent of the market filtration $\mathscr{F}$. The random times κ determines points in time where regime shifts take place in the economy, changing the local martingale measure chosen by the market to determine the arbitrage free prices. The random variables Y represents the relevant variables characterizing the state of the economy, e.g. unemployment rate.

Let $N_t = i,\ i = 1, 2, \ldots$ be a counting process representing the amount of regime shifts that has happened at time t, such that i regime shifts have happened at time t. Denote by $\mathbb{Q}^i \in \mathcal{M}_{loc}$ the i'th risk neutral measure chosen by the market at time t, depending on the state of the economy Y^i. Given this, we are able to define the fundamental price of the risky asset.

Definition 3: Fundamental Price

The fundamental price of the risky asset at time t, denoted S_t^*, is defined as

$$S_t^* = \sum_{i=1}^{\infty} \mathbb{E}_{\mathbb{Q}^i}\left[\int_t^\tau 1 \, \mathrm{d}Du + X_\tau \mathbf{1}_{\{\tau<\infty\}} | \mathscr{F}_t\right] \mathbf{1}_{\{t<\tau\}\cap\{t\in[\kappa_i,\kappa_{i+1})\}}, \quad t \in [0, \tau]. \tag{2.2}$$

Remark The terms appearing in (2.2) is the discounted cash flows, but since the money market account was assumed to be constantly equal to 1, the discounted and non-discounted value of the cash flows are the same. △

The fundamental price of the risky asset at time t is thus the expected value of the future cash flows, given the information at time t. Note that only one element of the sum will be nonzero for a given t because the indicator function is present. The reasoning for this is to only use the current risk neutral measure, the one the market is using, for the pricing and not the others. For instance, for any $t < \tau$, given that $\kappa_i \leq t < \kappa_{i+1}$, the fundamental price is given by

$$S_t^* = \mathbb{E}_{\mathbb{Q}^i}\left[\int_t^\tau 1 \ \mathrm{d}Du + X_\tau \mathbf{1}_{\{\tau<\infty\}} | \mathscr{F}_t\right].$$

Another thing to note is that the liquidation value of the risky asset at time infinity does not contribute to the fundamental price, since that value is not consumable by investors. Because the fundamental value is properly defined, it is now also possible to define the process of interest, namely the bubble process.

Definition 4: Bubble Process

The bubble process $\beta = (\beta_t)_{t\geq 0}$ is defined as

$$\beta = S - S^*, \tag{2.3}$$

where S and S^* are the market price and fundamental value of the risky asset respectively.

The bubble process is thus the difference between the market price and the fundamental price. In a well functioning market, the bubble process is zero. It is thus of interest to determine whether the market price is larger than the fundamental price, or equivalently, determine whether β is strictly positive. It turns out that only three scenarios exists for the bubble process.

Theorem 5

If there exists a nontrivial bubble $\beta \neq 0$, then three possibilities exist:

1. *If $\mathbb{P}(\tau = \infty) > 0$, then β is a $\mathbb{Q}$-local martingale, possibly a uniformly integrable martingale.*
2. *If τ is unbounded, but $\mathbb{P}(\tau < \infty) = 1$, then β is a $\mathbb{Q}$-local martingale, but not a uniformly integrable martingale.*
3. *If τ is a bounded stopping time, then β is a strict $\mathbb{Q}$-local martingale.*

Proof. A proof can be found in [Jarrow et al., 2006]. □

This theorem can be combined with the following, by first noting that (2.3) can be stated as

$$S = S^* + \beta, \tag{2.4}$$

which motivates the following theorem in order to decompose the bubble component in the market price.

Theorem 6: Bubble Decomposition

The price of the risky asset given by (2.4)*, when a bubble exists, admits a unique decomposition*

$$S = S^* + \beta = S^* + (\beta^1 + \beta^2 + \beta^3) \tag{2.5}$$

almost surely, where $\beta = \beta^1 + \beta^2 + \beta^3$*, and*

1. β^1 *is a càdlàg nonnegative, uniformly integrable martingale that satisfies* $\beta^1_t \to X_\infty$ *almost surely.*
2. β^2 *is a càdlàg nonnegative, nonuniformly integrable martingale satisfying* $\beta^2_t \to 0$ *almost surely.*
3. β^3 *is a càdlàg nonnegative strict local martingale, such that* $\mathbb{E}\left[\beta^3_t\right] \to 0$ *and* $\beta^3_t \to 0$ *almost surely.*

Proof. See [Jarrow et al., 2006] for a proof. □

Remark A strict local martingale is a local martingale that is not a martingale. △

Theorem 6 thus provides a decomposition into three terms, namely β^1, β^2 and β^3. The β^1 component will be referred to as a type 1 bubble, β^2 as a a type 2 bubble, etc. Type 1 bubbles occur when the lifetime of the risky asset is infinite and has a payoff at $\{\tau = \infty\}$. For type 2 bubbles, they occur when the lifetime of the risky asset is finite but unbounded, whereas type 3 bubbles takes place for a risky asset with bounded lifetimes.

The first thing to note about the bubbles are that type 1 bubbles converges to the payoff at infinity, while the others converge to zero. Consequently, type 1 bubbles only exists in infinite horizon markets, which are markets that will not be studied in this report. Furthermore, type 2 bubbles are also a phenomena in infinite horizon markets, which further excludes those types of bubbles, since only finite markets will be studied in this thesis. Therefore, type 3 bubbles will be the only bubble type of interest, and thus they will be denoted by β. Similarly, type 3 bubbles will be referred to as the “bubble component”, ”bubble process“, etc. and not as a type 3 bubble.

A thing to keep in mind is that the market was assumed to satisfy NFLVR, such that there are no strategies available that are an arbitrage. This does however not exclude bubbles. The most obvious strategy would be to take a short position in the risky asset during the bubble, hoping the bubble will burst and thus generate profit. Since arbitrage strategies are admissible by definition, the wealth process of an arbitrage strategy is bounded from below by some constant. However, with positive probability, the price of the risky asset can increase such that it violates the lower bound given by the admissibility condition. Thus shorting the risky asset during a bubble is not an arbitrage.

Because it is of interest to determine whether a bubble is present or not in the price process, it is possible to use Theorem 6 which states that the bubble component is a strict local martingale. Thus by (2.3), a bubble is present in the case that $S - S^*$ is a strict local martingale. Equivalently, one can also investigate the price process S and determine whether S is a strict local martingale or not. Thus, in the presented setup, even though the process of interest is the bubble process, we can still work with the original price process.

2.3 Test for Strict Local Martingales

Because the key component of bubbles are strict local martingales, it is of interest to develop a testing methodology for those, but before that is possible, it is necessary to make assumptions regarding the price process of the risky asset. It will be assumed that the price process is the solution to the stochastic differential equation

$$\mathrm{d}S_t = b(t, S_t, Y_t)\mathrm{d}t + \sigma(t, S_t)\mathrm{d}B_t, \quad S_0 = 1, \tag{2.6}$$

where the drift component b depends not only on time and the current asset price, but also on the underlying state of the economy, Y_t. The diffusion component σ also depends on time and the underlying asset price. B is a $\mathbb{P}-$Brownian motion. Such model is too general to be useful, so an additional assumption regarding the diffusion component will be made. More precisely, as in (2.2) where the risk neutral measure varies through time, we will allow the diffusion component to vary across time in the sense that it is time invariant locally, but not globally. Assume that we have a finite partition of the time interval $[0, \tau]$ such that $0 = t_0, \ldots, t_n = \tau$. It will then be assumed that the diffusion process σ has the form

$$\sigma(t, x) = \sum_{i=0}^{n-1} \sigma_i(x)\mathbf{1}_{t \in [t_i, t_{i+1})}. \tag{2.7}$$

The special form of σ is known as a regime shifting form. For instance, it could be assumed that t_0 is the time of the first observation on the first trading day, t_1 the first observation on the second trading day, etc. In this case, for each day, a different σ would have to be estimated. The diffusion component is thus globally time dependent, but not locally, since in each regime, the diffusion only depends on the asset price and not time. Because the market is incomplete by assumption, there exists an infinite amount of equivalent local martingale measures. Fix $\mathbb{Q} \in \mathcal{M}_{loc}$ and $t \in [t_i, t_{i+1})$, such that by use of Girsanov's Theorem, (2.6) becomes

$$\mathrm{d}S_t = \sigma(S_t)\mathrm{d}B_t^{\mathbb{Q}}, \quad S_0 = 1, \tag{2.8}$$

where $B_t^{\mathbb{Q}}$ denotes a $\mathbb{Q}-$Brownian motion. Hence the drift disappears, but the diffusion component stays the same after changing measure. Thus estimating σ on the original probability measure $\mathbb{P}$ or the risk neutral measure $\mathbb{Q}$ yields the same results. Since the price dynamics are now specified, it is possible to quantify whether the price process is a strict $\mathbb{Q}$-local martingale or not by the following theorem.

Theorem 7

The price process S given by (2.8) *is a strict $\mathbb{Q}$-local martingale if and only if for all $\varepsilon > 0$ it holds that*

$$\int_{\varepsilon}^{\infty} \frac{s}{\sigma(s)^2} ds < \infty. \tag{2.9}$$

Proof. See [Mijatović and Urusov, 2010] for a proof. □

Thus to determine whether the price process is a strict $\mathbb{Q}$-local martingale or not, one has to evaluate (2.9) and determine whether it is finite or not. If the integral is finite, a bubble is present by Theorem 6. Note that the integral is finite if the diffusion process grows at a faster rate than the underlying price process, i.e. bubbles are associated with large return variances at high price levels. Additionally, as one can see, this procedure requires the functional form of the diffusion process up until time infinity. Hence, one must first estimate the functional form of the diffusion component and later extrapolate in order to evaluate the integral. Even though the functional form of the diffusion component is required, we will instead focus on a point estimation of it, because this thesis will use the point estimates to estimate the functional form with a neural network.

2.4 Estimating the Diffusion Process

Due to the results of Theorem 7, it is of strong interest to estimate the diffusion process. As we already have mentioned, we will proceed by doing point estimates first, and therefore, an estimator of those will be presented. Since the condition actually needs the squared diffusion process, namely the spot variance, we will proceed by estimating that process instead of the spot volatility.

There are different ways of estimating the spot variance, but the method that will be used throughout this project will be the following based on [Figueroa-López and Wu, 2020]. Only a minor part of the details will be mentioned here, as the main purpose of this project is not to investigate different estimators of the spot variance. The rest of the details can be found in [Figueroa-López and Wu, 2020].

2.4.1 Setup

First, we assume that we are in a high frequency setup, such that intraday price data is available, and the log price process X_t is a continuous semimartingale satisfying

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t, \quad t \in [0,1],$$

with B being a Brownian motion, μ and σ being predictable and having continuous sample paths, and σ being non-negative. All processes are assumed to be defined on a filtered probability space. Furthermore, it is assumed that the spot variance process σ^2 is an Itô semimartingale with the following dynamics

$$\mathrm{d}\sigma_t^2 = \widetilde{\mu}_t \mathrm{d}t + \widetilde{\sigma}_t \mathrm{d}\widetilde{B}_t, \quad t \in [0,1],$$

with $\widetilde{B}$ being a standard Brownian Motion as well, with $\mathrm{d}\langle B, \widetilde{B}\rangle_t = \rho_t \mathrm{d}t$. Moreover, $\widetilde{\mu}$ and ρ are both adapted and locally bounded, $\widetilde{\sigma}$ is adapted and càdlàg, and ρ is also càdlàg.

An important matter to take into account when doing high frequency analysis is microstructure noise. If not, estimates of the volatility will be contaminated due to the noise being present in the observations. To account for this, it will further be assumed that X will not be observed, but $Y_t = X_t + \varepsilon_t$ is observed in its place, at a discrete time grid, where $t_i = \frac{i}{n}$ for $i = 0, 1, \ldots, n$, while n denotes the amount of observations present in the interval $[0,1]$. Furthermore, the noise process $\{\varepsilon_{t_i}\}_{i=1}^n$ is i.i.d. with mean 0 and variance ω^2, and also independent of the true log price process X.

2.4.2 The Estimator

The estimator in [Figueroa-López and Wu, 2020] is based on the pre-average estimator of the integrated volatility and kernel estimation. Thus to define it, some things must be presented first. More in depth, the following will be needed:

1. A sequence of positive integers k_n representing the pre-averaging window, defined as
$$k_n = \left\lfloor \frac{1}{\theta\sqrt{1/n}} \right\rfloor, \quad \theta > 0.$$
2. A real valued function g defined on $[0,1]$ being continuous, piecewise C^1 with piecewise Lipschitz derivative g' such that
$$g(0) = g(1) = 0, \qquad \int_0^1 g(s)^2 \mathrm{d}s = 1.$$
The function g plays the role of a weight function. Throughout this project, the function $g(x) = \min\{x, 1-x\}$ will be used.

Before presenting the estimator itself, some notation will be introduced for convenience. For a process U, we define

$$\overline{U}_i^n = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)(U_{(i+j-1)/n} - U_{(i+j-2)/n}),$$
$$\widehat{U}_i^n = \sum_{j=1}^{k_n} \left(g\left(\frac{j}{k_n}\right) - g\left(\frac{j-1}{k_n}\right)\right)^2 \left(U_{(i+j-1)/n} - U_{(i+j-2)/n}\right)^2.$$

Furthermore, for the weight function g we define

$$\phi_{k_n}(g) = \sum_{i=1}^{k_n} g\left(\frac{i}{k_n}\right)^2.$$

With these definitions, we can finally define the estimator.

Definition 8: Spot Variance Estimator

The pre averaging spot variance estimator is defined as

$$\widehat{\sigma}_t^2(\theta) = \frac{1}{\phi_{k_n}(g)} \sum_{j=1}^{n-k_n+1} \frac{1}{m_n/n} K\left(\frac{t_{j-1}-t}{m_n/n}\right)\left((\overline{Y}_j^n)^2 - \frac{1}{2}\widehat{Y}_j^n\right), \quad t \in [0,1], \tag{2.10}$$

where K is a kernel function, and m_n is a sequence of positive integers such that $m_n \to \infty$, $\frac{m_n}{n} \to 0$, $\frac{m_n}{\sqrt{n}} \to \infty$ and $\frac{m_n}{n^{3/4}} \to \beta$ for some $\beta \in (0,\infty)$.

Throughout this project, we will use the exponential kernel defined by

$$K(x) = \frac{1}{2}\exp(-|x|).$$

Moreover, we will let $m_n = n^{3/4}$. Choosing this specific kernel and m_n ensures consistency of the estimator which is always favorable. The details regarding this can be found in [Figueroa-López and Wu, 2020].

One should keep in mind that even though an estimator of the spot variance is now available, we cannot evaluate (2.9) yet. The first reason being that the estimator yields the value of $\sigma(s)^2$ at different point in time, not the functional form of σ^2. Furthermore, it also suffers from the finite amount of points the price process has been observed at. As a result, once the spot variance has been estimated, it is required to also estimate the function $S \mapsto \sigma(S)^2$ from the point estimates in order to evaluate the integral.

Chapter 3

Simulation Studies

The purpose of this chapter is to first study the estimator of the spot variance in finite samples in order to pick the parameter value that produces the best results in terms of some error measures. Next, we will also investigate whether we can detect known bubbles by the methodology described in Chapter 2. This is done by choosing a model such that under certain parameter values, the price process is known to be a strict local martingale, and under others, a martingale.

3.1 The Spot Variance Estimator in Finite Sample

To do this, a model to generate artificial prices first need to be specified. We will use the model (2.8) with a parametric form of the diffusion component, since that model will also be used in the later part of the simulation study that investigates whether we are able to detect bubbles. Namely, we will use the following model for prices:

$$\mathrm{d}S_t = \sigma(S_t)\mathrm{d}B_t = \gamma_0 S_t^{\gamma_1}\mathrm{d}B_t, \quad S_0 = 100. \tag{3.1}$$

We are intentionally leaving out a drift term, since we can always transform the model of the prices with a Girsanov type of transformation, which effectively removes the drift.

However, since the spot variance estimator only works when using log prices, we will transform the model for prices by using Itô's formula on the process $X_t = \log(S_t)$. This yields the model for log prices:

$$\mathrm{d}X_t = \widetilde{\sigma}(S_t)\mathrm{d}B_t + \widetilde{\mu}\mathrm{d}t = \frac{1}{S_t}\gamma_0 S_t^{\gamma_1}\mathrm{d}B_t - \frac{1}{2}\left(\frac{\gamma_0 S_t^{\gamma_1}}{S_t}\right)^2 \mathrm{d}t. \tag{3.2}$$

Notice that the diffusion process for the log prices is the diffusion process for the original prices, divided by the current asset price, i.e. $\widetilde{\sigma}(S_t) = \frac{1}{S_t}\sigma(S_t)$. Therefore, when we apply our estimator of the spot variance to the log prices, we will end up estimating $\widetilde{\sigma}(S_t)^2$, i.e. we will actually estimate

$$\widetilde{\sigma}(S_t)^2 = \left(\frac{\gamma_0 S_t^{\gamma_1}}{S_t}\right)^2 = \frac{\sigma(S_t)^2}{S_t^2}.$$

Thus to obtain estimates of the spot variance of the original prices, we will have to multiply the estimates obtained from the log prices by S_t^2. On top of the log prices, we will also add a noise term, such that the log prices observed will have the form

$$Y_t = X_t + \varepsilon_t.$$

The term ε is added to represent the microstructure noise that is present in high frequency financial data. We will assume that $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \omega^2)$ for each i. Moreover, it will be assumed that ω^2 can take 4 different values, namely $0.01, 0.001, 0.0001$ and 0.00001. The reason we allow for this flexibility is to investigate how the estimator performs in different scenarios of noise. For simulating, the Euler Scheme will be used to simulate (3.1), and it will be assumed that $23,400$ equally spaced observations are available each trading day, corresponding to an observation every second, i.e. $t_i = \frac{i}{n}$, where $n = 23,400$.

Because this thesis focuses on bubbles, and we already have allowed the diffusion component to change through time as in (2.7), we also want to incorporate this in the simulation. For simplicity, we will assume that there is two regimes, one where the price process is a bubble and one where it is not. Due to the parametric form of the diffusion process, the price process will be a strict local martingale in case $\gamma_1 > 1$, and a martingale in case $\frac{1}{2} < \gamma_1 \leq 1$. Moreover, we will settle with a fixed scaling parameter of $\gamma_0 = 1.2$, while simulating with both $\gamma_1 = 0.9$ and $\gamma_1 = 1.1$ to test how the estimator performs in two different regimes. The reason we allow for a relatively large value of γ_0 is to make the prices vary more, such that the price process will potentially visit larger price levels, and thereby, we would be able to obtain estimates of the spot variance at those levels.

For each scenario of noise and regime, we will simulate 1000 paths, and for the j'th path, $j = 1, \ldots, 1000$, calculate the average squared error, ASE, defined by

$$ASE_j = \frac{1}{n-2l} \sum_{i=l}^{n-l} (\widehat{\sigma}_{i,j} - \sigma_{i,j})^2,$$

where $l = \lfloor \frac{n}{10} \rfloor$ is used to avoid problems of estimating the boundary. Keep in mind that the estimator presented estimates the spot variance. However, we will report the errors as the difference in the spot volatility, such that after estimating the spot variance, we will calculate the square root of the estimates, obtaining estimates of the spot volatility instead. Next, we estimate the root mean squared error by

$$\widehat{RMSE} = \sqrt{\frac{1}{m} \sum_{j=1}^{m} ASE_j},$$

where $m = 1000$ represent the amount of simulated paths. In the following tables, the RMSE estimate is reported for each combination of the θ parameter of the estimator and level of noise variance.

θ	$\omega^2 = 0.00001$	$\omega^2 = 0.0001$	$\omega^2 = 0.001$	$\omega^2 = 0.01$
0.25	12.550	12.536	12.478	13.457
0.5	9.130	9.302	9.139	10.093
0.75	7.670	7.347	7.648	7.986
1	7.002	6.588	7.116	7.338
1.25	6.051	5.932	6.135	7.008
1.5	5.496	5.840	5.568	6.870
1.75	5.139	5.151	5.477	6.851
2	5.070	5.005	5.165	7.038
2.25	4.598	4.860	4.916	7.775
2.5	4.637	4.519	4.778	7.909
2.75	4.302	4.263	4.685	8.176
3	4.090	4.271	4.622	9.181
3.25	4.199	4.012	4.326	9.865
3.5	3.953	4.039	4.193	10.281
3.75	3.916	3.935	4.304	12.140
4	3.801	3.980	4.373	11.651
4.25	3.747	3.768	4.565	13.480
4.5	3.704	3.623	4.402	13.843
4.75	3.685	3.723	4.260	15.469
5	3.642	3.620	4.463	17.784
5.25	3.484	3.521	4.381	17.071
5.5	3.575	3.650	4.501	20.028
5.75	3.526	3.519	4.436	19.496
6	3.396	3.569	4.398	20.166

Table 3.1. RMSE of the estimator in different scenarios with $\gamma_1 = 0.9$.

By Table 3.1, one can deduct that in a scenario where the price process is a martingale and the variance of the noise is low, the best estimates in term of the estimated RMSE is obtained with larger values of θ. Intuitively, this makes sense since larger values of θ implies a smaller pre averaging window, i.e. one uses less of the future weighted returns to make the effect of the noise diminish. As the variance of the noise term increases, the optimal value of θ decreases. This can easily be seen in the last column in Table 3.1, as the optimal value of θ is 1.75 based on the RMSE estimates. Compared to the case with the lowest variance, the RMSE estimate is not fluctuating as much, and the best value seems to 6, but it could possibly be larger.

We will further analyze the estimator in case where the price process is a strict local martingale, i.e. a bubble process. The results are reported in the following table.

θ	$\omega^2 = 0.00001$	$\omega^2 = 0.0001$	$\omega^2 = 0.001$	$\omega^2 = 0.01$
0.25	83.654	174.267	219.769	173.571
0.5	163.373	99.045	106.175	93.140
0.75	92.313	54.615	74.228	57.928
1	45.141	59.643	52.022	67.321
1.25	61.224	54.626	39.102	158.093
1.5	70.854	286.099	138.646	38.821
1.75	58.205	51.823	127.107	66.760
2	48.556	134.380	50.616	42.804
2.25	70.396	74.400	57.429	302.405
2.5	96.015	82.169	122.815	429.408
2.75	47.891	63.495	87.868	64.194
3	109.199	90.133	36.124	53.901
3.25	30.790	53.125	100.535	176.632
3.5	51.603	186.960	35.715	179.498
3.75	175.728	30.168	50.401	56.280
4	136.610	119.228	39.076	45.576
4.25	49.539	423.192	36.736	285.726
4.5	38.036	111.537	65.376	42.773
4.75	76.490	111.460	71.280	54.544
5	42.059	58.283	101.620	607.103
5.25	107.219	82.962	48.059	78.934
5.5	98.768	55.012	143.287	45.782
5.75	50.499	89.099	73.815	39.022
6	65.085	58.613	37.942	61.756

Table 3.2. RMSE of the estimator in different scenarios with $\gamma_1 = 1.1$.

When estimating the RMSE in case where the price process is a strict local martingale, one can deduct from Table 3.2 that the errors are substantially larger compared to those reported in Table 3.1. In fact, the RMSE estimates vary so much that it is hard to even come up with optimal values of θ in each of the 4 scenarios. This can be seen in all the cases, where e.g. the RMSE estimates in the case $\omega^2 = 0.01$ are very close for $\theta = 1.5$ and $\theta = 5.75$. We expected the RMSE estimates to be more volatile, because bubbles are volatile by nature, but certainly not to this extent. Therefore, when we later will use the estimator, we will not use the results from Table 3.2 to choose an optimal θ, but only the results in Table 3.1 because those estimates looks more trustworthy.

3.2 Detecting Strict Local Martingales

In this section, we will simulate processes that are known to be strict local martingales, and therefore bubbles, to test whether we can detect them. To do this, we will stick with the same model as in Section 3.1, meaning we can control whether the price process is a bubble or not. We also choose the same model parameters. After simulating a price path, we will apply our method to quantify whether the simulated price process is a strict local martingale or a martingale. We will proceed by presenting our method before applying it.

As noted during Chapter 2, determining whether a price path is a bubble, we must determine whether (2.9) holds or not, i.e. determine whether the price process is a strict $\mathbb{Q}-$local martingale or not. Our method aims to use all the theory presented throughout Chapter 2. It will be assumed that the spot volatility process is not dependent on time throughout a trading day, such that t_i and t_{i+1} in (2.7) represents the first observation on day i and $i+1$ respectively. In essence, the method is based on the following procedure. For each day, we do the following;

1. For each intraday data point, estimate the spot variance by (2.10).
    - This yields point estimates of the function $S \mapsto \sigma(S)^2$ for points the stochastic process generating the prices has visited.
2. Estimate the function $S \mapsto \sigma(S)^2$ by using the point estimates obtained from the first point.
    - We use two methods to see which works best. The first one is a simple linear model, while the other one is a more complicated neural network. The goal of both methods is to learn the map $S \mapsto \sigma(S)^2$. We will use the *keras* package in *R* to train the networks. The architecture of the network and the linear model will be elaborated later.
3. Determine whether the integral of the function $\frac{s}{\sigma(s)^2}$ over the positive real line is finite by using the function estimate obtained from the second point.
    - The reason this is of interest stems from the fact that it decides whether (2.9) holds. In case the integral is finite, we conclude a bubble is present.
    - To determine whether the integral is finite or not, we will first use the function estimate obtained from the previous point, such that we are deciding whether the integral of $\frac{s}{\widehat{\sigma}(s)^2}$ is finite or not. To check whether the integral is finite or not, we will investigate the limit

      $$\lim_{s\to\infty} \frac{s}{\widehat{\sigma}(s)^2}.$$

      If the limit is 0, we will conclude that the integral is finite, and thus a bubble is present. Keep in mind that technically, this condition is a necessary but not sufficient condition for the integral in (2.9) to be finite, i.e. even though we conclude that the limit is 0, we cannot actually state that the integral is finite, and thus conclude that we have a bubble. Nevertheless, we will still use this approach, but we will discuss this matter further in the discussion part of the thesis.

One must keep in mind that it is not the behaviour of $\sigma(S)^2$ as $t \to \infty$ that is of interest, that is, the asymptotic behaviour of the spot variance as a function of time. What is important is determining the structure of the spot variance when the prices approaches infinity, i.e. how the spot variance behaves at large prices. If the volatility increases rapidly, we classify a stock as a bubble, which also makes sense intuitively, since bubbles are often linked with high volatility and large price deviations. We will proceed by applying each step to a simulated path such that one understands how the method works.

3.2.1 Estimating the Spot Variance Process

The first part of our method involves estimating the spot variance process. First, we will simulate a path with $\gamma_0 = 1.2, \gamma_1 = 1.1$ and $\omega^2 = 0.00001$. Recall that we mentioned that the RMSE estimates in Table 3.2 was not particularly good. Therefore, even though we are simulating a path with $\gamma_1 = 1.1$, we will still use Table 3.1 in order to pick the value of θ to use in the estimator. Thus we deduce that the optimal value of θ for the spot variance estimator is 6. We then apply the estimator to the simulated path. The simulated price along with the true spot variance and estimated spot variance process can be seen in the following figure.

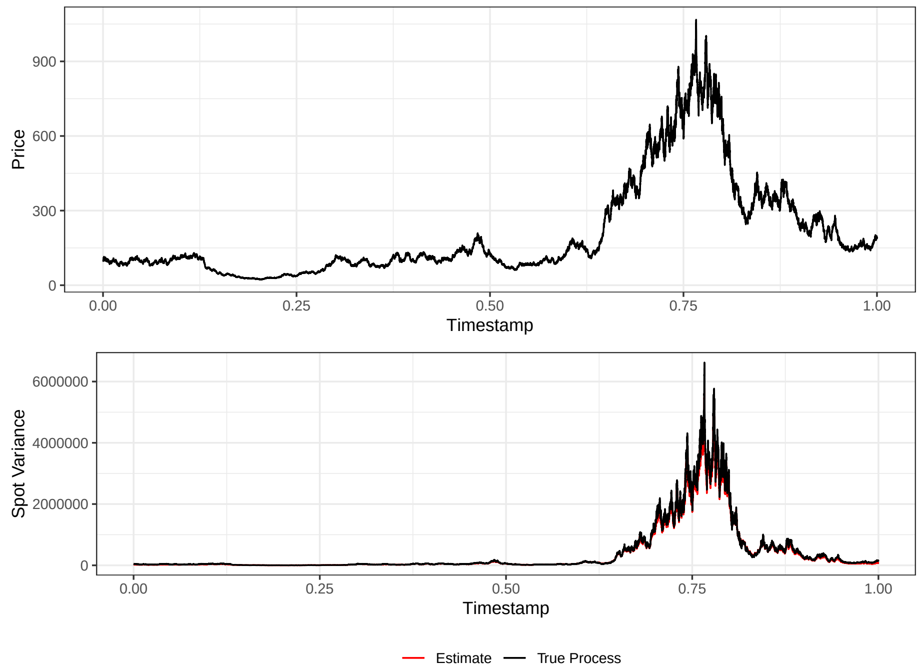

Figure 3.1. Simulated price data and corresponding spot variance and estimated spot variance process.

It can be seen by Figure 3.1 that the estimator has problems with capturing sudden spikes in volatility, but in general it estimates the behaviour of the true process quite well. For this particular simulation, the price level starts at 100 and then increases all the way up to 1000, before dropping massively, replicating what one would think is bubble behaviour. The estimator has some problems estimating the boundary, but due to the large values of the spot variance in this particular simulation, one cannot see it. The boundary problem of the estimator was previously alleviated by simply discarding the first and last 10% of the estimated spot variance points. We will do the same here, which discards effectively 20% of the data points.

Furthermore, Figure 3.1 reveals a potential cause as to why the RMSE estimates in Table 3.2 are so volatile. Because the estimator does not react to sudden spikes in the spot variance, and if one price path behaves similarly to the path seen in Figure 3.1, this single simulation would massively effect the average squared error, and thereby the RMSE estimate. We will further discuss this in the discussion part of the thesis.

Note that by assumption, at each timestamp $t_i \in [0, 1]$, we have effectively obtained an estimate of $\sigma(S_{t_i})^2$. Because it is exactly the function $S \mapsto \sigma(S)^2$ that is of interest, we want to use the spot variance estimates to estimate this function.

3.2.2 Estimating the Function From the Spot Variance Estimates

Because the spot variance estimates are now available at each timestamp, we can start estimating $S \mapsto \sigma(S)^2$ with a linear model and a neural network. Before diving further into this task, we would like to visualize the point estimates beforehand, such that one has a grasp of what function we are actually trying to estimate. That is, we plot the estimated spot variance as a function of the price. However, this gets really messy, so we round the prices to 1 digits and then calculate the average estimated spot variance for each price level, resulting in 5222 observations. Another advantage to this is that for a given price level, if we have several estimates of the spot variance, we can use the multiple estimates to obtain a more accurate estimate by calculating the average. To visualize, we obtain the following:

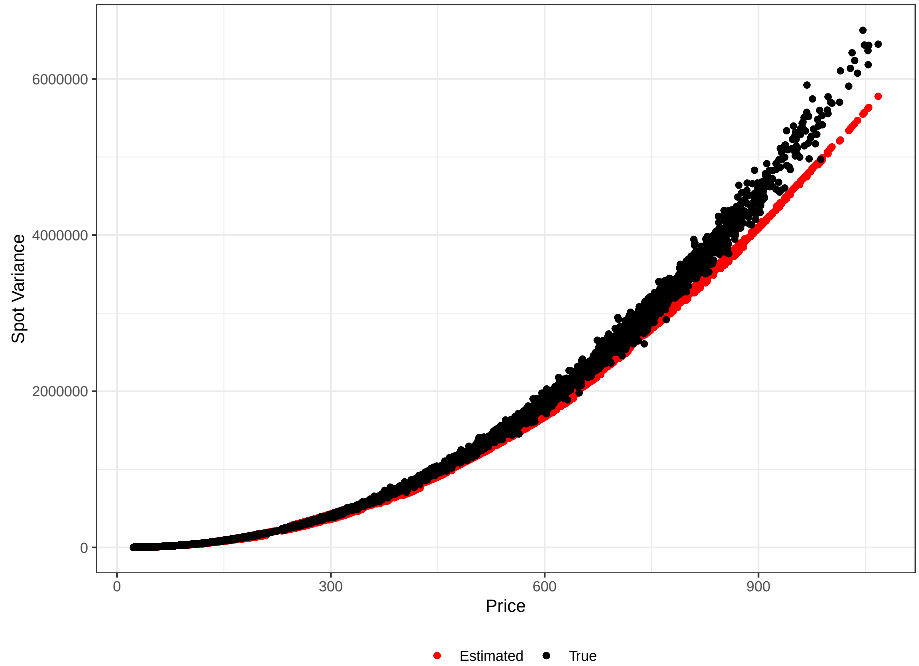

Figure 3.2. The spot variance and estimated spot variance as a function of the simulated prices.

From the looks of Figure 3.2, one can see that the spot variance increases with larger price levels. One can also see that at the largest price levels visited by the price process, we have that the estimator underestimates the true spot variance, further supporting the fact that it is not good at reacting to spikes in prices, but mostly constructs a more smooth version of the true spot variance process.

Since there has been a lot of price movements for this simulation, namely from 23 to 1067, we have obtained estimates at relatively high price levels. However, there is no way of really knowing the behavior of the spot variance as the price approaches infinity, simply cause we do not have the data available, and thus we are forced to extrapolate. We are doing this, as already mentioned, with a simple linear model and a neural network. For the architecture of the network, we will settle with the following;

1. Input layer with 1 neuron such that the input of the network is a given price.
2. Hidden layer with 50 neurons, equipped with activation function softplus.
3. Output layer with 1 neuron. This yields the spot variance estimate of the inputted price.

We will train the network with 100 epochs. For training, we will also use scaling of the input variable by the following method

$$\widetilde{S} = \frac{S - \min S}{\max S - \min S},$$

which forces the prices into the interval $[0, 1]$. We will do the same for the estimated spot variance. This eases the training procedure. Then, once we want to predict, we simply first apply the same scaling to the price level we want to predict, then use the trained network, and at last apply the reverse transformation to recover the original scale of the spot variance.

For the linear model, we will settle with something very simple. We will estimate the model

$$\sigma_i^2 = \beta_1 S_i + \beta_2 S_i^2 + \beta_3 S_i^3,$$

that is, explain the spot variance with the original, squared and cubed prices. Moreover, we will try to estimate the linear model and train a network on both the true spot variance process and the estimated one to see how estimating the spot variance impacts the results.

Results Using The True Process

After estimating the linear model and training a network, we plotted the resulting estimates along with the raw data to visualize the results. They can be seen in the following figure.

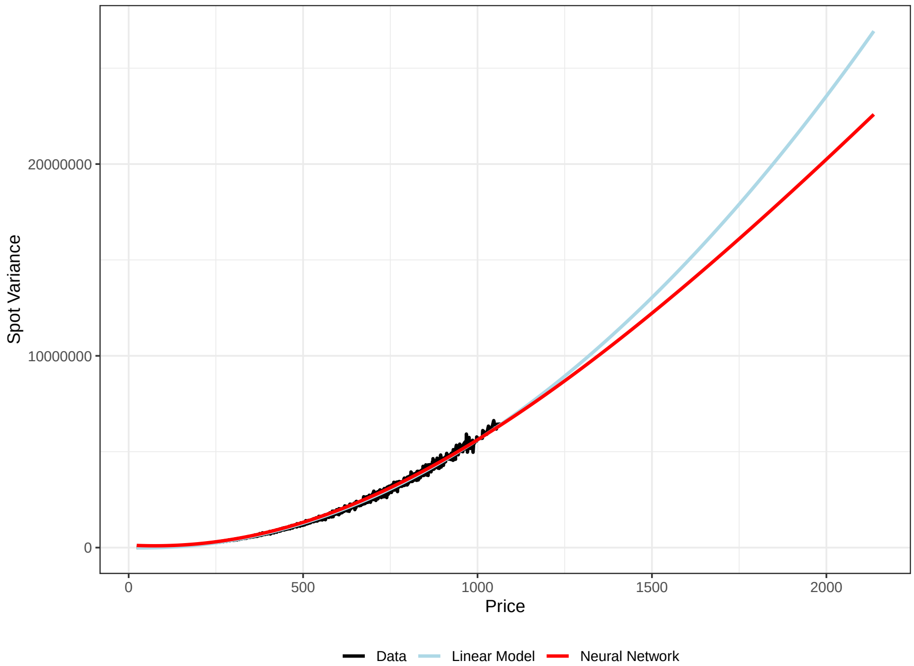

Figure 3.3. The linear model and the neural network fitted to exact spot variance and prices.

It is worth noticing that by Figure 3.3, for the price levels displayed in the plot, the linear model predicts a more steep increase in the spot variance as the price level increases compared to the trained network. It is of interest to see how the two models behave in case where we use the estimated spot variance process because that will be relevant once we start applying it to real data.

Results Using The Estimated Process

Applying the same steps as before on the estimated spot variance process yields the following plot.

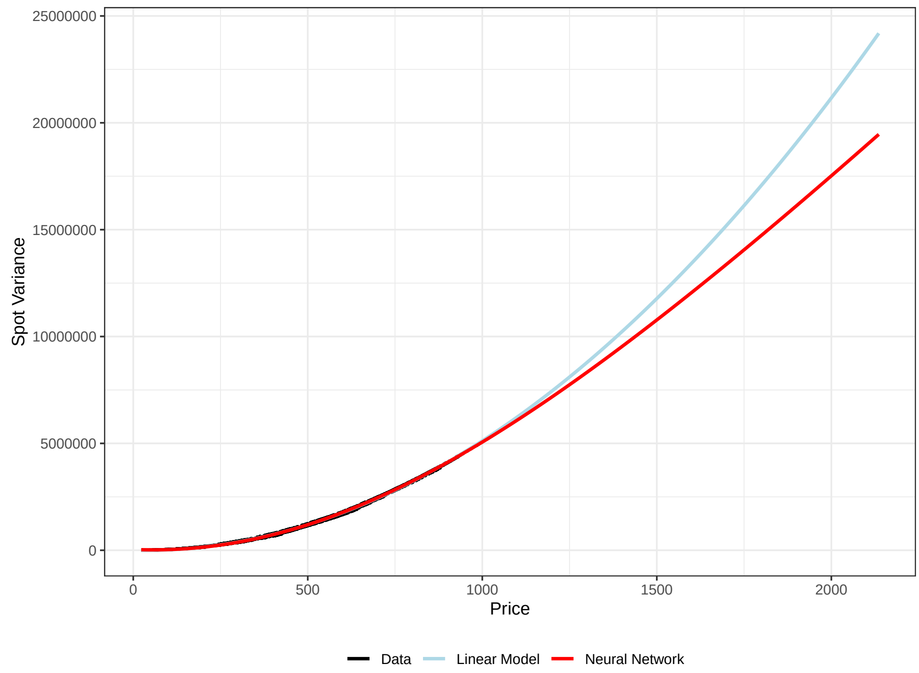

Figure 3.4. The linear model and the neural network fitted to estimated spot variance and prices.

The results are very similar to the case where we used the true spot variance process as training. The neural network does not predict as steep an increase in spot variance at larger price levels compared to the linear model. However, this is only for the displayed data. We must determine the asymptotic behavior and check how fast the spot variance grows when prices increase.

3.2.3 Determining the Limit

The last point is to evaluate the limit

$$\lim_{s\to\infty} \frac{s}{\widehat{\sigma}(s)^2},$$

which in turn will help us determine whether (2.9) is true or not. We will once again go through the two cases, the first one we use the actual spot variance process, and the other one where we use the estimated spot variance process.

We simply calculate the ratio $\frac{s}{\widehat{\sigma}(s)^2}$ and then determine whether the ratio is converging to 0 as $s \to \infty$. If the fraction does converge to 0, we have a bubble by (2.9), and if not, we do not have a bubble. This is seen in the following figure.

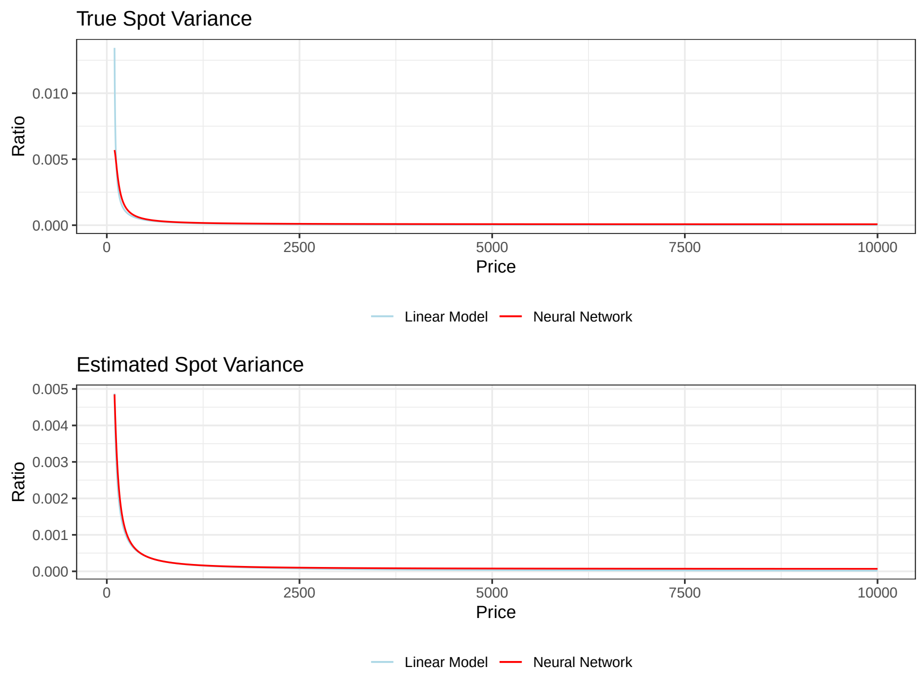

Figure 3.5. Plot of the ratio $\frac{s}{\widehat{\sigma}(s)^2}$. The top plot is when the true spot variance process is used, and the bottom one is with the estimated spot variance process.

The top part of Figure 3.5 is the ratio plotted for both models in the case where the true spot variance process is used. Similarly, the bottom one is the results of using the estimated process. It can be seen that as the price level increases, the ratio decreases in both cases. Heuristically, this means that as the underlying price level increases, the spot variance increases at a faster pace, which is exactly what one would expect from a bubble process; increasingly high volatility at larger price levels.

For the classification itself, a threshold will be proposed for the ratio $\frac{s}{\widehat{\sigma}(s)^2}$, such that if the ratio is lower than the given threshold, the price path will be classified as a bubble. We will settle with a threshold of 0.01. Thus, if

$$\frac{s}{\widehat{\sigma}(s)^2} < 0.01,$$

the price path is classified as a bubble. If the inequality does not hold, we instead classify the price path as not being a bubble. Moreover, we cannot evaluate the true limit, but we will instead calculate the value of the ratio using a very large s. We will use $s = 1,000,000$. The conclusion is the same for both cases, namely that both the linear model and the network classifies this particular price path as a bubble. Thus for this particular simulation, both methods succeeded in classifying correctly.

3.2.4 Simulation Results

Next, we simulate 500 price paths which we know are martingales, each with 23,400 observations. Similarly, we simulate another 500 paths which are strict local martingales. Afterwards, the method just presented will be applied to see how well it classifies these paths. We will use both the linear model and the network method, but only on the estimated spot variance process, because during the application part of the thesis, the true spot variance process will not be available. The results are summarized in the following misclassification tables. The first table is the performance of our network method.

	Actual Martingale	Actual Strict Local Martingale
Predicted Martingale	193	8
Predicted Strict Local Martingale	307	492

Table 3.3. Misclassification table for the network method.

One can see that the network is great at identifying strict local martingales, but it is not as outstanding in detecting martingales, at least under the chosen model and parameters. For the price processes that were bubble processes, it had a success rate of 98.4%. On the other hand, when simulating processes which are not bubbles, it ended up classifying 38.6% of the paths correctly, which is quite disappointing. The results for the linear model are reported in the next table.

	Actual Martingale	Actual Strict Local Martingale
Predicted Martingale	284	334
Predicted Strict Local Martingale	216	166

Table 3.4. Misclassification table for the linear model.

In the case where bubbles are simulated, the linear model does a poor job in identifying them. In fact, it only accurately classifies 23.2% of the simulated paths correctly. On the other hand, it does a much better job in identifying the paths that are not bubbles, achieving an accuracy of 56.8%. This is also much better than the network method.

The results obtained during this chapter clearly shows that our network method is capable of detecting bubbles in a controlled environment as a simulation study. Therefore, the next thing is to test the method on real data.

Chapter 4

Data Analysis

The goal of this chapter is to use the theory from Chapter 2 and the results obtained in Chapter 3 on real data. The application will be focused on applying our method presented in Section 3.2 and see how often we find bubbles. Some minor adjustments to the method will be needed, and they will be covered shortly.

4.1 Pre Processing

First, we will need to assume a model for our prices. During the simulation study, we used a general model, but introduced a parametric form of the spot volatility process. In this chapter, a paramatric form of the spot volatility will not be imposed, because our method is non parametric. It is worth emphasizing that, as already mentioned in Chapter 3, one can eliminate the drift term by a Girsanov transformation, so we will use a model for prices of the form:

$$\mathrm{d}S_t = \sigma(S_t)\mathrm{d}B_t.$$

Then by Itô's formula, the model for the log prices $X_t = \log(S_t)$ is then given by

$$\mathrm{d}X_t = \frac{\sigma(S_t)}{S_t}\mathrm{d}B_t - \frac{1}{2}\left(\frac{\sigma(S_t)}{S_t}\right)^2 \mathrm{d}t.$$

This reveals that in order to obtain the estimates of the spot variance process for the original price process, we need to multiply the estimates obtained under the log prices by the squared price level, just as in Chapter 3.

For the application, we will use intraday S&P 500 index data from 1st of January 1999 to 31 of December 2009 in order to hopefully detect the housing bubble. Technically, the data is from SPY, an S&P 500 ETF, but it will still be referred to as "S&P 500 data". The data was obtained from the NYSE Trade and Quote database. Before starting, the data must be cleaned. This was done in the same way as in [Jensen and Mikkelsen, 2021].

Furthermore, when we applied our method during Section 3.2, we knew the variance of the noise process because the prices were simulated. When handling real data, this is obviously an unknown quantity, i.e. it must be estimated. The reason the variance of the noise process is of interest is due to the results obtained in Tables 3.1 and 3.2. Here, the parameter of the estimator, θ, is shown to have a great effect on the performance of the estimator. Moreover, the variance of the noise plays a large role in the value of θ, which is why the noise level is worth estimating before starting the entire procedure. The variance of the noise can be estimated by appropriately scaling the realized variance estimator of the integrated variance. More formally, the variance of the noise can be estimated consistently under the assumption of *i.i.d.* noise by

$$\widehat{\omega}^2 = \frac{1}{2n}\sum_{i=1}^{n}(X_{t_i} - X_{t_{i-1}})^2, \tag{4.1}$$

where n is the number of observations in the discrete time grid $[0,1]$, and X_{t_i} is the i'th observed log price.

Because our method works best if the price process has experienced large price movements, applying our method to individual days might be problematic, since for most days, the price level is not varying much. Therefore, we would have liked to consider the price movements during an entire week to allow for more movement, and thus potentially better estimates. However, if we start doing this, we will introduce jumps in the data, due to the spread between the closing value and opening value between each day. As a result, the estimator will suffer due to the jumps in the data. Consequently, we will settle on a daily level, even though it is not ideal.

Thus for each day, we will employ a dynamic value of θ for the spot variance estimator dependent on the estimated noise variance. More precisely, for each trading day, the variance of the noise will first be estimated by (4.1), and then depending on the estimate, θ will be chosen as follows;

$$\theta_i = \begin{cases} 6 \text{ if } \widehat{\omega}^2 \in (0, 0.00005], \\ 5.75 \text{ if } \widehat{\omega}^2 \in (0.00005, 0.0005], \\ 3.5 \text{ if } \widehat{\omega}^2 \in (0.0005, 0.005], \\ 1.75 \text{ if } \widehat{\omega}^2 \in (0.005, \infty), \end{cases} \tag{4.2}$$

for $i = 1, \ldots, m$ where m represents the amount of trading days we have data available for. In this thesis, $m = 2766$.

It should be noted, as in the simulation studies during Chapter 3, that the theoretical part covered in Chapter 2 works for the log prices. During the simulation studies, we started by presenting a model for the original prices, then used Itô's formula to derive the model of the log prices. Afterwards, we used the spot variance estimator for the log prices, and at last transformed those estimates back to estimates for the original prices by multiplying with the squared prices. We will use the same method throughout the data analysis, which is why we also derived the dynamics for the log price process. With this, we are now able to present how the data will be handled during this chapter. For each day, the following procedure will be applied.

1. Transform the prices into log prices.
2. Estimate the variance of the noise process by (4.1).
3. With the estimated variance of the noise process, use (4.2) to choose the optimal θ, and then use (2.10) to estimate the spot variance process for each timestamp.
4. Transform the spot variance estimates of the log prices to estimates of the prices by multiplying with the squared price level at each timestamp.
5. Estimate the function $S \mapsto \sigma(S)^2$ by a neural network.
6. Evaluate whether a bubble is present or not using extrapolated values from the trained network.
    - We will do this in the same way as in the simulation studies, by using a threshold of 0.01 and let $s = 1,000,000$.

Another thing to keep in mind is that during a trading day, some price levels will be visited several times. Hence, for some price levels, we will have different estimates of the spot variance. Thus, we can use the several estimates obtained at a given price level and then calculate the average of them, leading to a more precise estimate of the spot variance at the given price level. We will use this to our advantage similar to what we did in Chapter 3. There, prices were rounded to 1 digit, but since we are working with real data, the price movement will be lower. Consequently, if we were to do the same here, not a lot of price observations would be available. Therefore, we instead work with rounding the prices to 2 digits throughout this chapter. We will apply this procedure in the following section.

4.2 Results

Applying our method presented in Section 3.2 with the adjustments mentioned in the Section 4.1, we detected bubbles on 0 out of 2766 days by using the same threshold as in the simulation studies, i.e. the criterion

$$\frac{s}{\widehat{\sigma}(s)^2} < 0.01.$$

However, if the threshold is increased, we will find bubbles. Below is a table that summarizes the 10 days with the lowest estimated asymptotic value of the fraction $\frac{s}{\widehat{\sigma}(s)^2}$.

Date	$\frac{s}{\widehat{\sigma}(s)^2}$
1999-05-12	0.130
2008-10-10	0.145
2008-09-29	0.148
2008-03-14	0.178
2008-10-29	0.203
2008-10-24	0.273
2008-11-21	0.285
2008-07-11	0.301
2008-11-14	0.310
2008-11-11	0.318

Table 4.1. Lowest estimated values of the ratio $\frac{s}{\widehat{\sigma}(s)^2}$ with $s = 1,000,000$.

It can be seen that the lowest value of the ratio $\frac{s}{\widehat{\sigma}(s)^2}$ is 0.130 in Table 4.1, which is more than 10 times larger than the threshold proposed. This is not ideal, since not a single day was classified as a bubble, and we expected at least to detect some bubbles since the housing bubble happened throughout the time period analyzed.

Since we did not detect any bubbles, we will still analyze the 10 days shown in Table 4.1 and later build a trading strategy based on those dates in order to show what was initially planned for the days deemed bubbles.

We will refer to the price path of the dates in Table 4.1 as “bubbles”, even though they technically are not according to our results. Notice that several of the days captured by our method are dates in 2008, which is when the housing bubble occurred. The only date not taking place in 2008 was instead in 1999, which is around the time of the “Dot com” bubble mentioned in Chapter 1. This further indicates that our method, even though it technically did not detect any bubble, is still finding dates that potentially are bubbles which is promising.

For one particular date shown in Table 4.1, the 29th of September 2008, we will investigate more in depth. The intraday price movement for that date can be seen in the following figure.

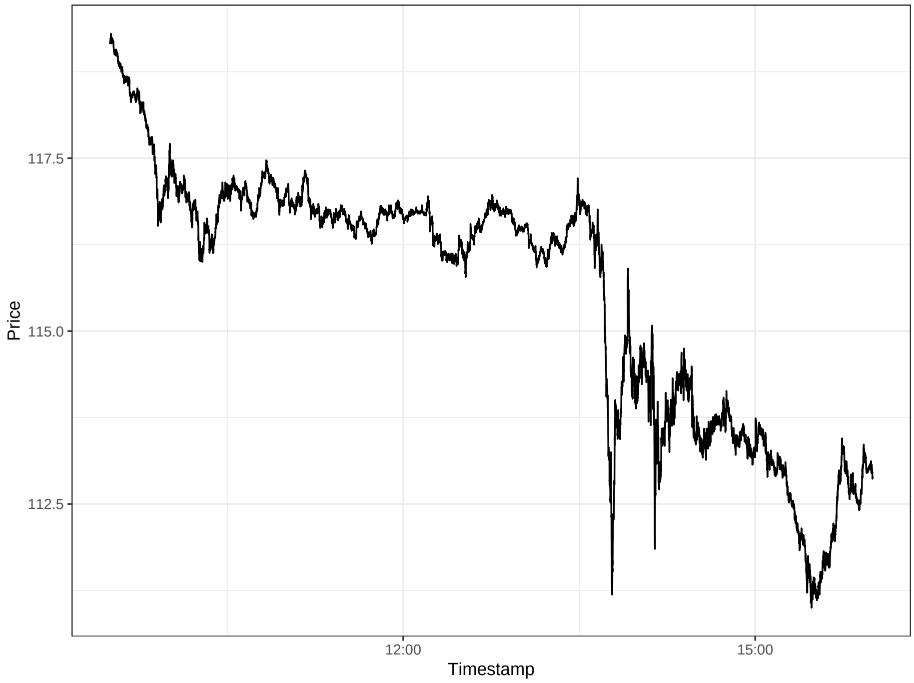

Figure 4.1. The intraday price movements for SPY on the 29'th of September 2008.

It is revealed by Figure 4.1 that the price movements during the 29th of September 2008 are very volatile, especially during the end of the day. During that same day, the S&P 500 lost 8.8% compared to the previous close value, marking one of the worst days ever for the index at the time. It was reported that investors was very concerned about the election in the USA, and thus feared that the congress would not be able to come with a solution to the problem of frozen credit markets [Twin, 2008]. Ultimately, this decline would be followed by multiple others, which in the end was a byproduct of the financial crisis in 2008.

Because we find bubble tendencies around 2008, as one would expect, it does not look like the results obtained are completely irrelevant, even though the theory states that technically, no bubbles were detected. Thus, with some minor adjustments to our method, we could perhaps have detected bubbles legitimately.

4.3 A Simple Trading Strategy

With the results from Section 4.2 in mind, we will during this section introduce a very simple trading strategy. Afterwards, the profitability of the established strategy will be investigated. In essence, the trading strategy tries to take a short position the day after a bubble has been detected, and then exit the position based on some criteria. In essence, the strategy is as follows.

1. Start by staying out of the market.
2. Short the day after the first bubble signal.
   - The reason we short the day after comes from the fact that we need the entire price movements throughout a day in order to determine whether the price process was a bubble or not.
3. Close the short position at a 10% profit or a maximum of 15% loss.
4. Repeat the process to determine the next entry point.

In some cases, we can experience problem when wanting to close our position. Imagine that we have opened a short position, and we currently are at a loss of 14.5%. If the shorted asset is up another 2% at the start of the next trading day, the loss will greater than the 15%. Thus even though our strategy aims to short, and at a maximum have a 15% loss, we can sometimes be forced to close our position at a larger loss.

When opening a short position the day after a bubble was detected, it will be assumed that it is possible to short at the first price level observed the following day, and similarly when we want to close the short, it is assumed that it can done at the first price that would have made the strategy lose more than 15%, or yield a return of more than 10%.

Moreover, it will be assumed that we can short sell 10 shares at a time. Additionally, if we already have a short position, and another bubble detection signal is obtained, we will close our initial position. For the existing position, it is assumed that it can be closed at the opening price the following day, and start a new short at the same price, thus resetting the requirements according to the proposed strategy.

4.3.1 Results From the Strategy

By following the strategy, the first entry point was detected on the 12th of May 1999. Then, we short sell the following day at the first observed price level which was 137.25$. The position is then held until the 18th of October 1999, because during that day, the price drops to 123.5$, i.e. we have made more than a 10% return by our trading strategy, and thus the position will be closed. This means that by short selling 10 shares, 137.5$ was made on this trade. The results obtained for the entire trading strategy is summarized in the following table.

Entry Date	Close Date	Entry Price	Close Price	Profit
1999-05-13	1999-10-18	137.25	123.5	137.5
2008-03-17	2008-07-11	126.35	125.29	10.6
2008-07-14	2008-09-29	125.29	112.74	125.5
2008-09-30	2008-10-06	113.62	102.25	113.7
2008-10-13	2008-10-24	94	84.05	99.5
2008-10-27	2008-10-29	85.97	95.84	-98.7
2008-10-30	2008-11-11	95.84	88.22	76.2
2008-11-12	2008-11-14	88.22	86.36	18.6
2008-11-17	2008-11-20	86.36	77.72	86.4
2008-11-24	2009-01-06	81.91	94.2	-122.9

Table 4.2. Results from applying the trading strategy on the 10 price processes closest to be bubbles.

Displayed in Table 4.2 is the entry date, namely the date where the short position is opened at the first available price level. Recall that the entry date is the date after the bubble has been detected. Next is the close date, which is the date of which the position is closed. The position is only closed in case that we either made a profit 10%, lost 15%, or a new bubble was detected. Moreover, the price level of the entry and close of the position is also displayed, along with the profit from shorting 10 shares.

We can see that following the strategy throughout the 10 year period was profitable. In fact, the strategy had positive profits in 8 out of the 10 periods. The total profit over the entire period ends up being 446.4$, which might not seem like a lot, but no large short positions were taken either.

Dwelling deeper into the strategy, one could suggest taking a long position after the first short closed, since more than 8 years pass between the first and second short position. However, if this was to be done, one must take great care, since after the closing of first position, the market was very unstable, by first increasing, and then declining due to the “Dot com” bubble.

It could be of interest to test how the strategy compares to a simple buy and hold strategy. Assuming an initial capital of $10,000$$, and that we can take a long position at the first observed price on 1st of January 1999, hold it until our last data point on 31th of December 2009, we could have bought at a price level of 123.375$ a share. This results in effectively 81 shares, with a leftover of 6.625$.

At the last observation on the 31th of December 2009, the price level is 111.44$. Thus, our initial investment at that period in time would be worth $9,033.265$$, taking the leftover 6.625$ into account as well. Therefore, by following the buy and hold strategy, one would have lost almost 10%. It should be emphasized that the assumption of no dividends makes the buy and hold strategy look worse than it actually is. In fact, if dividends were included, and they were reinvested, the buy and hold strategy would not have been as bad as depicted here. Moreover, dividends would also remove some of the profit from the short seller, since the short seller is inclined to also pay the dividend back, but for simplicity, those were excluded.

Chapter 5

Discussion and Conclusion

5.1 Discussion

In this thesis, financial bubbles in a high frequency setup has been studied. We opted to contribute to the current level of research by using neural networks as function approximators and hereby do non parametric estimation of the spot variance. Before we were able use the neural networks, we first had to establish a pointwise estimator of the spot variance. The estimator is based a kernel estimator in combination with the pre average estimator of the integrated variance. Other estimators are also available for this task, namely an estimator based on the two scale realized variance of the integrated variance, but this estimator has an inferior convergence rate compared to the one used throughout this thesis [Figueroa-López and Wu, 2020], [Zu and Boswijk, 2014].

Moreover, we chose to use the exponential kernel, but in reality, we could have chosen several other kernels. A major drawback of the exponential kernel is that the estimates of the endpoints are poor. An explanation as to why this is the case can be found by looking at the estimator in (2.10). Because the exponential kernel was chosen, the estimator uses data from before and after each timestamp to estimate the process. Therefore, at the left boundary, there is not a lot of data before each time stamp, and likewise at the right boundary, not a lot of data after each timestamp. To address this issue, we could have used other kernels which are more robust at the end points as in [Zhang and Karunamuni, 1998]. On the other hand, this could possible result in worse performance in the remaining part of the dataset. Ultimately, we could have tried to implement a varying kernel depending on whether we are estimating the boundaries or not, but did not do so due to lack of time.

The estimator was then tested in a simulation study. We chose a constant elasticity of variance model, CEV, without drift for our price process. The reason a drift was omitted was due to Girsanov's theorem, which roughly states that the drift can be eliminated by a suitable transformation of the original probability measure. We also chose to present a model for prices and not log prices. We could also have done it the other way around, namely by first presenting a model for the log prices X_t, and then by use of Itô's formula derived the market dynamics for the raw prices by letting $S_t = \exp(X_t)$.

During the simulation, we let the initial value of the price process be 100. The reason for this particular value was that the SPY index was trading at around those price levels in the time frame analyzed in the thesis. This particular initial value, compared with the chosen model parameters of $\gamma_0 = 1.2$ and the two different regimes of $\gamma_1 = 0.9$ and $\gamma_1 = 1.1$ led to very volatile behaviour of the simulated paths. A γ_0 value of 1.2 is also very large, but we intentionally sought very volatile behaviour, because high volatility is usually associate with bubbles. The impact of large volatility is easily identified when examining Tables 3.1 and 3.2, where the RMSE estimates in the case of $\gamma_1 = 0.9$ is much lower than those of $\gamma_1 = 1.1$, i.e. paths that were simulated to be a bubble according to the presented theory. It was to be expected that more uncertainty would arise in case of bubble behavior, but not to this extent. Because the paths are so volatile, the RMSE estimates will also become more inaccurate. The RMSE estimates obtained in the case of $\gamma_1 = 1.1$ are practically useless. Therefore, to obtain better estimates of the RMSE, we could have increased the amount of paths simulated for each case.

When doing simulation analysis of our method, we chose a very simple structure of the network, namely only one hidden layer with 50 neurons. We tried experimenting with several structures of more neurons and hidden layers, but this simple architecture performed best in the simulation studies. We also tried different activation functions, but the softplus was the one yielding the best performance.

For the comparison with the linear model, one should keep in mind that it is essentially the sign of the coefficient estimated to the cubed prices that determines whether we will have a bubble or not. A possible extension to this would be to do a t-test in order to test the significance of the three parameters in the model. If we were to find the cubic and squared terms insignificant, we would toss them out, and thus conclude that the price process was not a bubble.

At the time we evaluated whether a given price path was a bubble or not according to the extrapolated values of the trained networks, we tested whether $\lim_{s\to\infty} \frac{s}{\widehat{\sigma}(s)^2} = 0$. However, this is only a necessary condition for the integral in (2.9) to be finite. Even though this limit goes to 0, we cannot be exactly sure that the integral is finite. For a more rigorous argument, we should instead have investigated whether

$$\frac{s}{\widehat{\sigma}(s)^2} \leq C \frac{1}{s^{\alpha}},$$

where $\alpha > 1$ and $C > 0$. This can also be written alternatively as

$$\frac{s^{1+\alpha}}{\widehat{\sigma}(s)^2} \leq C. \tag{5.1}$$

This condition would instead imply that the integral in (2.9) would be finite, which is the condition we should have used, but we did not have the time to implement this due to time restraints.

When checking the limit $\lim_{s\to\infty} \frac{s}{\widehat{\sigma}(s)^2} = 0$, we chose to implement a threshold of 0.01 and decided to put $s = 1,000,000$ for the price level to represent infinity. We felt that a price level that high would be satisfactory in representing a price at infinity, since the current price level of S&P 500 is much lower than this.

Performance wise, the networks were exceptional when tasked with classifying simulated strict local martingales, since it had an accuracy of 98.4%. For the linear model, it was not as good, only classifying 23.2% correctly. In the case where the simulated paths were not bubbles, the linear model performed substantially better than the networks, classifying 56.8% of the simulated paths correctly, compared to 38.6% from the networks. However, neither of these results are very promising. The reason for this could, as already mentioned for the linear model, stem from missing inference on the estimated parameters. It could also be due to the relative close values of γ_1 in the two scenarios. Had we experimented with γ_1 equal to 0.7 and 1.3, the results could have been different, but that would also change the RMSE estimates in Tables 3.1 and 3.2.

Had we implemented the condition (5.1) in the simulation studies, we would possibly gotten different results for our network method. The reason is that (5.1) is a more strict condition than $\lim_{s\to\infty} \frac{s}{\widehat{\sigma}(s)^2} = 0$. Consequently, we would most likely had been less successful in detecting simulated bubble processes, but possibly also better at detecting paths that were not bubbles.

Another reason our own method did not provide more satisfactory results could come from the fact that the type of networks we have decided to work with has difficulties doing extrapolation [Xu et al., 2021]. Therefore, better results could possibly have been obtained by working with other types of networks than the feedforward ones.

Importantly, the method we presented is a non parametric way of detecting bubbles. The main implications of this is that it will most likely perform worse compared to a parametric method, under the assumption that the spot volatility process and the form of the parametric estimator is the same. However, non parametric methods are more general and less sensitive to the underlying dynamics.

During the data analysis, we decided to apply our method to each trading day. We could also have grouped each trading week into one data set, and then tried applying our method to each week instead. This would, as mentioned during Section 4.1, introduce jumps in the data due to the spread between the close and opening prices for two different days. Since our estimator of the spot variance is not robust to jumps, we would get misleading results. However, we could have chosen to implement a jump robust version of the estimator used, which essentially excludes returns that are larger than a threshold, but we did not have the time to explore this further.

For the result obtained when trying to detect bubbles in the S&P 500, we failed to detect a single bubble with the methods we proposed, even though the housing bubble is present in the data. This is not promising, but as already mentioned, the type of the networks used throughout the thesis has problems extrapolating.

The reason no bubbles were detected might be a matter of how well we can estimate not only the function $S \mapsto \sigma(S)^2$, but also a matter of how great the point estimates of the spot variance are. We concluded during Chapter 3 that the errors from the estimator explodes when we simulated bubbles, so the estimates might have been too bad to learn the true map.

Due to no bubbles being detected, we thought it would be interesting to see which dates had the lowest value of the ratio $\frac{s}{\hat{\sigma}(s)^2}$, since we deemed those days the one closest to being bubbles. The lowest value of $\frac{s}{\hat{\sigma}(s)^2}$ was 0.130, which is more than 10 times the threshold for being classified as a bubble. Therefore, we could have tried to vary the threshold more, to see how it would end up affecting the classification.

In particular, we investigated the 10 days with the lowest ratio, which showed that 9 of the 10 days deemed potential bubbles occurred in 2008, just around the time of the financial crisis. The last day was in the 1999s, just before the "Dot com" bubble. This can indicate that the methods we used throughout the thesis can work, but perhaps with some additional fine tuning in order to actually detect bubbles.

Using the same 10 days, we built a trading strategy to see whether relying on our findings could be beneficial financially. The strategy was based on shorting immediately after discovering a bubble, and then keep the position until either a 15% loss, a 10% gain, or detection of a new bubble. This strategy beat the classic buy and hold period for the entire period investigated throughout the thesis, earning 446.4\$, under the assumption that we could short 10 shares a time.

The buy and hold strategy instead, with an initial capital of $10,000$\$, ended up being worth $9,033.265$\$ in the same time period, effectively losing almost 10%. Thus, even though we assumed an initial capital of $10,000$\$, we lost money, while the shorting strategy did not require any capital at all. However, the profitability of the two strategies was under the assumption of no transaction cost and dividends, and especially reinvested dividends throughout a longer time period would have been noticeable in the end result for the buy and hold strategy. Similarly, the profit from the shorting strategy does not take dividends into account either, which one when shorting also has to cover.

5.2 Future Research

A natural extension to our work would be to impose a more complex model for the prices in the simulation study compared to the one we presented. Then under a more complicated model, it would be possible to investigate the performance of our method. This could for instance be a rough volatility model in order to expand upon the current research regarding these type of models.

Moreover, as already mentioned in the discussion, the networks we have worked with throughout this thesis is not the ideal ones for extrapolation tasks. Therefore, another thing going forward would be to use other type of networks, e.g. recurrent or long-short term memory neural networks. Possibly, the research conducted in this thesis could also have been done with other machine learning techniques such as trees, etc.

It would also be of great interest to work with high frequency data of individual companies instead of an index. Typically, there is more variation for individual companies compared to indices, so we would expect that more bubbles would be present in those type of assets.

It would also be possible to extend the research by allowing for more than 1 risky asset, and then investigate vectors of strict local martingales as in [Protter and Dandapani, 2019].

One could also extend the research conducted in this thesis by trying to model the lifetime of bubbles, and not only detect bubble signals. This has already been explored in [Obayashi et al., 2016], but not with the same emphasis on neural network as in this thesis.

5.3 Conclusion

Based on the results obtained throughout this thesis, we can conclude that we were successful in applying neural networks to detect financial bubbles in a simulation study. We did so in a high frequency setup where we assumed that microstructure noise was present. We then introduced a spot variance estimator that was robust to noise, and used a neural network to learn the function $S \mapsto \sigma(S)^2$ from the point estimates obtained by the estimator. Using the network, we then classified whether price paths were a bubble or not.

When applying this to real data in form of the S&P 500, we failed to detect any bubbles, but instead we examined the 10 days deemed closest to being bubbles according to our setup. We found that 9 of the days estimated to almost be bubbles were dates in the 2008, which was around the time of the financial bubble.

Finally, we developed a trading strategy using those 10 days based on shorting after detecting bubbles, and we concluded that using this strategy would have generated positive profit, compared to the buy and hold strategy which would have lost cash over the same time period.

Bibliography

Bashchenko, O. and Marchal, A. (2020). Deep learning for asset bubbles detection.

Figueroa-López, J. E. and Wu, B. (2020). Kernel estimation of spot volatility with microstructure noise using pre-averaging.

Jarrow, R. A., Kchia, Y., and Protter, P. (2011). How to detect an asset bubble.

Jarrow, R. A., Protter, P., and Shimbo, K. (2006). Asset price bubbles in complete markets.

Jarrow, R. A., Protter, P., and Shimbo, K. (2010). Asset price bubbles in incomplete markets. *Mathematical finance*, 20:145–185.

Jensen, J. B. and Mikkelsen, J. S. (2021). Forecasting the integrated volatility and vix using artificial neural networks.

Mijatović, A. and Urusov, M. (2010). On the martingale property of certain local martingales. *Probability theory and related fields*, 152:1–30.

Obayashi, Y., Protter, P., and Yang, S. (2016). The lifetime of a financial bubble.

Protter, P. and Dandapani, A. (2019). Strict local martingales and the khasminskii test for explosions.

Ross, D. (2020). The Real Story Behind the 17th-Century 'Tulip Mania' Financial Crash. `https://www.history.com/news/tulip-mania-financial-crash-holland`.

Twin, A. (2008). Stocks crushed. `https://money.cnn.com/2008/09/29/markets/markets_newyork/`.

Xu, K., Zhang, M., Li, J., Du, S. S., ichi Kawarabayashi, K., and Jegelka, S. (2021). How neural networks extrapolate: From feedforward to graph neural networks.

Zhang, S. and Karunamuni, R. J. (1998). On kernel density estimation near endpoints.

Zu, Y. and Boswijk, H. P. (2014). Estimating spot volatility with high-frequency financial data. *Journal of Econometrics*, 181:117–135.