EARLY WARNING ON STOCK MARKET BUBBLES VIA ELLIPSOIDAL
CLUSTERING AND INVERSE PROBLEMS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EFSUN KÜRÜM

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
FINANCIAL MATHEMATICS

FEBRUARY 2014

Approval of the thesis:

EARLY WARNING ON STOCK MARKET BUBBLES VIA ELLIPSOIDAL CLUSTERING AND INVERSE PROBLEMS

submitted by **EFSUN KÜRÜM** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Department of Financial Mathematics, Middle East Technical University** by,

Prof. Dr. Bülent Karasözen
Director, Graduate School of **Applied Mathematics** _____

Assoc. Prof. Dr. A. Sevtap Kestel
Head of Department, **Financial Mathematics** _____

Prof. Dr. Gerhard Wilhelm-Weber
Supervisor, **Financial Mathematics, METU** _____

Assist. Prof. Dr. Cem İyigün
Co-supervisor, **Industrial Engineering, METU** _____

Examining Committee Members:

Assoc. Prof. Dr. A. Sevtap Kestel
Actuarial Sciences, METU _____

Prof. Dr. Bülent Karasözen
Scientific Computing, METU _____

Assoc. Prof. Dr. Ömür Uğur
Financial Mathematics, METU _____

Prof. Dr. Gerhard Wilhelm-Weber
Financial Mathematics, METU _____

Assoc. Prof. Dr. Egemen Asım Yılmaz
Electrical and Electronics Engineering, Ankara University _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: EFSUN KÜRÜM

Signature   :

ABSTRACT

EARLY WARNING ON STOCK MARKET BUBBLES VIA ELLIPSOIDAL CLUSTERING AND INVERSE PROBLEMS

Kürüm, Efsun

Ph.D., Department of Financial Mathematics

Supervisor   : Prof. Dr. Gerhard Wilhelm-Weber

Co-Supervisor   : Assist. Prof. Dr. Cem İyigün

February 2014, 87 pages

When a financial bubble bursts, not only a large number of people suffer directly in society, but it also affects the entire economy. Therefore, it is important to develop an early warning using mathematics-supported tools that aims at a detection of bubbles. We introduce a new method which approaches the bubble concept geometrically by determining and evaluating ellipsoids. In fact, we generate a volume-based index via minimum-volume covering ellipsoid clustering method, and in order to visualize these ellipsoids, we utilize Radon transform from the theory of the inverse problems, in the form of figures. Analyses were conducted for US, Japan and China stock markets. In our study, we have observed that when the time approaches bubble-burst time, the volumes of the ellipsoids gradually decrease and, correspondingly, the figures obtained by Radon transform are becoming more "brilliant", i.e., more strongly warning. The thesis ends with a conclusion and an outlook to future investigations.

Keywords : Financial bubbles, early-warning, clustering, inverse problems, ellipsoid.

ÖZ

ELİPSOİDAL KÜMELEME VE TERS PROBLEMLAR ARACILIĞIYLA HİSSE SENETLERİ PİYASASINDAKİ BALONLAR İÇİN ERKEN UYARI YAKLAŞIMI

Kürüm, Efsun

Doktora, Finansal Matematik Bölümü

Tez Yöneticisi : Prof. Dr. Gerhard Wilhelm-Weber

Ortak Tez Yöneticisi : Assist. Prof. Dr. Cem İyigün

Şubat 2014, 87 sayfa

Bir finansal balon patladığında, toplumdaki pek çok kişiye doğrudan zarar verdiği gibi, aynı zamanda, bu durum tüm ekonomik yapıları da etkiler. Bu nedenle, matematik destekli araçlar kullanarak balonların saptanmasını amaçlayan bir erken uyarı yaklaşımı geliştirmek önemlidir. Balon kavramına, elipsoitlerin belirlenmesi ve değerlendirilmesi ile geometrik olarak yaklaşan yeni bir yöntem getiriyoruz. Minimumhacim kaplayan elipsoit kümeleme yöntemi ile hacim-tabanlı bir endeks oluşturuyor ve bu elipsoitleri görselleştirmek için, ters problemler teorisinden Radon dönüşümü'nden, figürler şeklinde, yararlanıyoruz. Analizler, ABD, Japonya ve Çin hisse senedi piyasaları için yapılmıştır. Çalışmamızda, balonun patlama zamanı yaklaştığında elipsoitlerin hacimlerinin giderek küçüldüğü ve buna paralel olarak Radon dönüşümü'nden elde edilen figürlerin daha "parlak", yani daha güçlü bir uyarı, haline geldiği gözlemlenmiştir. Tez, sonuç ve gelecekte yapılacak araştırmalara bir bakış açısı ile sona ermektedir.

Anahtar Kelimeler: Finansal balonlar, erken uyarı, kümeleme, ters problemler, elipsoit.

x

In memory of Prof. Dr. Ahmet Hayri Körezlioğlu

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Stock market bubbles take place when the prices of stocks increase overmuch their *fundamental value* (*intrinsic value*). According to the fundamental value of a security, an investor can decide whether the security is *overvalued* or *undervalued*. If the intrinsic value of a security is higher than its current market price, it is classified as *overvalued*, if not, it is called as *undervalued*. Hence, he decides about to buy or to sell the security. An investor is willing to purchase or hold an asset at a price that is above its fundamental value if he thinks that the asset can be resold at an even higher price in a later trading round (11). This trading behavior is called "speculation" by *Harrison* and *Kreps*, 1978 (37). As a result, the fundamental value can be thought as the price which an investor is eager to pay if he is compelled to hold the asset forever, i.e., if he is not allowed to retrade (11).

The most common calculation method of the fundamental value of a stock or any security is the *discounted cash flow* (*DCF*) analysis that is defined as follows (2):

$$\text{Fundamental Value of a stock} = DCF := \frac{CF_1}{(1+d)^1} + \frac{CF_2}{(1+d)^2} + \ldots + \frac{CF_n}{(1+d)^n}. \tag{1.1}$$

This formula benefits from *weighted average cost of capital* (*WACC*) as a discount variable to explain the time value of money. Here, d is a discount rate and CF_n ($n \in \mathbb{Z}_+$) is a free cash flow in period n calculated in this form (2):

$$\text{Free cash flow} = \text{Net income} + \text{Amortization \& Depreciation} - \text{Changes in} \\ \text{Working Capital} - \text{Capital Expenditures}.$$

Since stock market bubbles are only realized during continuous bull market circumstance, the traders' confidences are very high (14). They believe that the demand for the stocks will never end and stocks will always become profitable. This belief about the stock market causes irrational expectations and escalates the stock prices upward and inflates the size of the stock market bubble as well. This trend ends when some investors recognize that the prices have risen unrealistically, thus they begin selling their stocks before the prices go down. Then, other traders follow this attitude; hence, panic selling starts. Not always but mostly, this process is completed by a sharp decline in price as in Figure 1.1 and when this acute drop occurs, it is called that the *bubble bursted*.
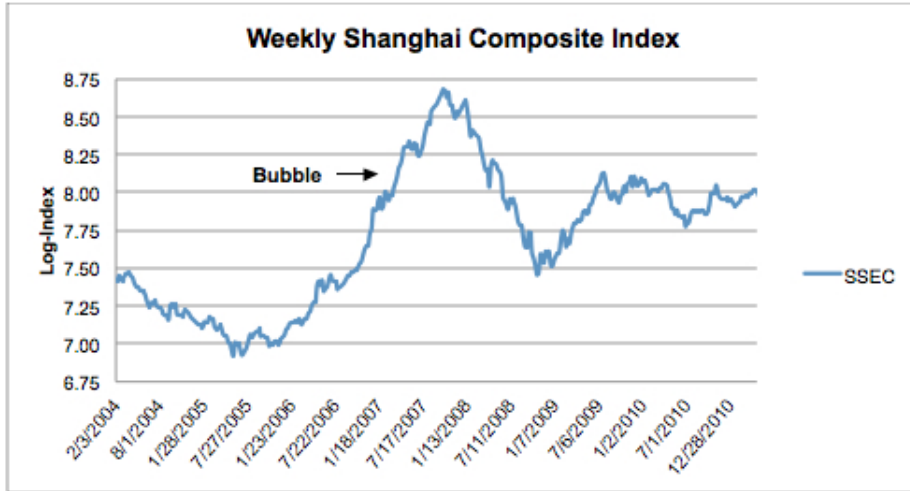
Figure 1.1: Weekly Shanghai Composite Index for the 2007 crisis (23).

If a bubble affected only a few investors, it would be of a limited importance because it has a little effect on the overall economy. Bubbles can cause major problems when they originate in an asset that is commonly held. Then, when the bubble bursts, not only a large number of people suffer directly, but it also interacts with the economy. Therefore, dangerous bubbles are generally the ones in a stock market (12).

Most of the companies that grow exponentially through the stock market bubble, when the bubble bursts, they go bankrupt. This gives rise to an increase of the unemployment rates (14). Business and consumer consumptions diminish and this may cause to commence the economic recession (14). Because of these negative effects of bursting of a bubble, it is important to develop an *early-warning* to detect it, timely.

The scientific contribution of this thesis is to introduce an early-warning signalling for bubbles, for the first time by using inverse problems and clustering methods. Here, we suggest a new approach where the logarithmic price process is approximated and quantitatively represented with ellipsoids which have a minimum volume (55). The *shape* of the price process before and after a bubble burst time inspired us to deal with this issue, firstly, via ellipses in the 2-dimensional case as shown in Figure 1.2. Here, as a first and intuitive approach, we plotted the bubble as a "curve-filling" static ellipse, parallel to the ordinate axis. However, in the two dimensions, we could only consider their coordinates as time and logarithmic price/index. In order to reflect the features of the price process better, we generalize the ellipses to higher dimensions. Hence, we generate ellipsoids.

Figure 1.2: Geometric interpretation of a bubble.

Generally, the definition of a speculative bubble is known, in consequence of positive feedback among economists: The index or the price of an asset increases considerably and becomes overvalued. Then, a crash usually occurs in a bear market with a acute declined prices for months or years, but this case does not occur every time (46). Therefore, though we became inspired by Figure 1.1 to deal with the bubble concept via ellipsoids, actually, the ellipse's position in Figure 1.2 is a special case. We can monitor the strong increase and the sharp decline in the price process like in Figures 5.1-5.4, i.e. , an increasing and decreasing price process not always looks like ellipses. However, we are interested in the dynamics of a (finite) sequence of ellipses. In fact, since our aim is to develop an early warning, we focus on the *dynamics* and the *structure* of ellipsoids *before* the crash time, especially, on their volumes in time dependence. Regarding the structure, in order to image this approach, we construct Figures 1.3-1.4 in the 2-dimensional case.
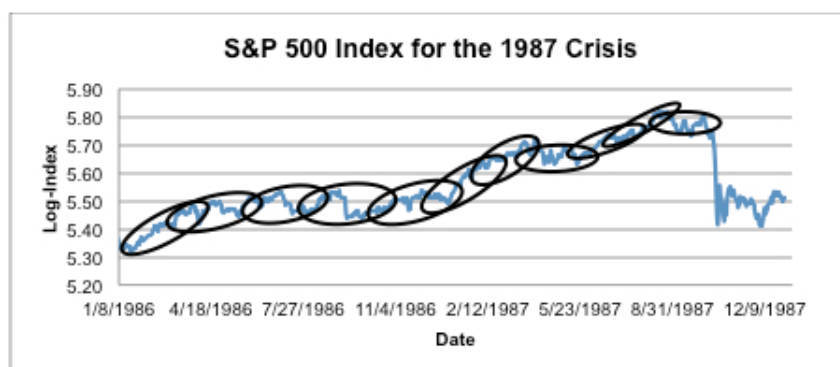
Figure 1.3: Structure of ellipsoids before a crash time.

In fact, in Figure 1.3, the dynamics of the ellipsoids does not show a strong change of the volume before the bubble bursts. In Figure 1.4, the situation is quite the opposite.

We emphasize that we are interested in ellipsoids (generated by data) "over" interval like in Figure 1.3 "between" intervals as we had it in Figure 1.2. This dynamical approach will be supported by a sliding window technique.

In Figure 1.4, shortly before the crash time, the strong drift dominates the whole price process, and the accumulated "energy" given by the volatility (oscillation) before turns into the process' rising up. Therefore, the ellipsoids become shrinking, then, dropping in terms of their volume; in other words, they become looking like a "needle" as shown in Figure 1.4. Regarding ellipsoids in general, and needles in particular, and their underlying covariances of variables, we refer to (4).
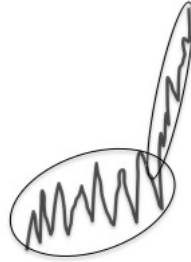
Figure 1.4: Geometric interpretation of a pronounced bubble.

In order to obtain each ellipsoid, seven factors, namely, date, log price/index process, weekly mean, standard deviation, median, growth rate of the price process and weekly log-iterated return of this process, are employed as a vector, which corresponds to each week. In fact, the log price process is divided into intervals and the data of each interval are described by an ellipsoid, determined by *minimum-volume covering ellipsoids clustering* method, and the volumes of ellipsoids are monitored with the *sliding-window* technique which we will explain. Moreover, we call a series of volumes a "volume-based index". In other words, the volume is used to measure the size of each ellipsoid. Then, to visualize these volumes, we benefit from a method from the theory of inverse problems which is called the *Radon Transform*. Here, Radon transform yields the color (representing the volume) and its change for any regarded ellipsoid, surveyed and monitored in time dependence. The behavior of the volumes of the ellipsoids and their change of color can be used as an early-warning signalling to detect bubbles and to be protected against to destructive outcomes of them.

As for the definition of *log-iterated return*: Let $\log p_1, \log p_2, \ldots, \log p_n$ be logarithmic prices of any asset or index for the first, the second and, eventually, the nth days. The return is calculated as $\log p_n - \log p_{n-1} = \log(p_n/p_{n-1})$. However, when the usual definition of return is used, we observe that Radon transform did not indicate definite and pronounced signals, i.e., most of the figures have the same color, as we later on see in Figures 5.16 and 5.18. Therefore, we generate the following variable by taking the logarithm of prices two times:

$$\log(\log p_2/\log p_1) = \log(\log p_2) - \log(\log p_1).$$

4

We call this new variable the *log-iterated return*. It means a (discrete) change-rate or first-order information which, as so often in calculus and applied mathematics, implies and determines the first-order information of the log-price itself.

As for the volume based index, since the logarithm is a monotonic function, there is no qualitative change path of the index when we compare the common definition of return with our log-iterated return, as we presented in the example of in Figure 1.5.
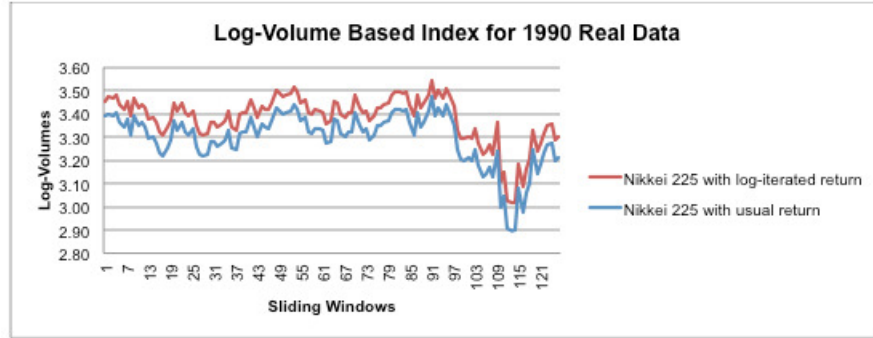
Figure 1.5: Comparison of volume based index with usual return and log-iterated return.

Here, we analyse US, Japan and China stock market bubbles and use S&P 500, Nikkei 225, Shanghai Composite (SSEC) indices. In addition to real data, we also produce *fitted data* according to *Johansen-Ledoit-Sornette model* which will be explained in detail, later. This model claims that during a speculative bubble, an economic index follows a power law adorned with a log-periodic oscillation so-called *Log-Periodic Power Law* (*LPPL*). The LPPL is a function of time *t*, stated by (47):

$$y(t) \approx A + B(t_c - t)^z + C(t_c - t)^z \cos\left[w\log(t_c - t) + \phi\right],$$
(1.2)

where t_c shows the most probable time of the crash, z is the exponential growth, the amplitude of the oscillations is controlled by w, and A has to be greater than 0 ($A > 0$) and equals the estimated price of the asset at the time of the crash $p(t_c)$. Furthermore, B has to be less than 0 ($B < 0$) and C is a number different from 0 to guarantee the log-periodic behavior. As for ϕ, it is assumed to be in the interval $[0, 2\pi]$ (46).

The rest of the thesis is organized as follows. In Chapter 2, we introduce famous historical bubbles. Then, we explain Johansen-Ledoit-Sornette model in detail in Chapter 3. A brief introduction of our methodological tools, *Minimum-Volume Covering Ellipsoids* clustering method and *Radon transform*, are provided in Chapter 4. Chapter 5 presents the details of our approach and its treatment. Chapter 6 consists of future research studies related to two main areas which are *ellipsoidal calculus*, and *optimizing the "area under the ROC curve"*. The thesis ends with Chapter 7 by a conclusion and an outlook.

CHAPTER 2

FAMOUS HISTORICAL BUBBLES

Here, we present famous financial bubbles which occurred in the history, in order to perceive destructive results of a possible forthcoming bubble in present times.

2.1   Tulip Mania

The *tulip mania* is considered as the first recorded financial bubble; it occurred in the 1630s in Holland. In 1593, tulips were introduced by a botanist *Carolus Clusius* who brought them from Ottoman Empire. He planted his garden to examine them for medicinal purposes. Then, his neighbors stole some of his bulbs in order to make some quick money. Hence, they gave rise to start the process of Dutch bulb trade (76).

In 1634, the madness of having tulips became too much and daily work was ignored. Even the lowest members of the society joined the tulip trade. Until 1635, the mania ascended, the prices increased and many people invested fortunes to possess tulip bulbs. Then, to sell tulips by their weight, in *perits* became necessity. Perit was a small weight which is less than a grain. In fact, 480 grains equalled 1 ounce. Prices for different varieties were as follows:

Admiral Liefken, weighing 400 perits = 4,400 florins; *Admiral Von der Eyk*, weighing 446 perits = 1,260 florins; *Shilder* of 106 perits = 1,615 florins; *Viceroy* of 400 perits = 3,000 florins; *Semper Augustus*, weighing 200 perits = 5,500 florins (22). In order to perceive the value of a single tulip, Table 2.1 can be useful. It was recorded in terms of items and their values by one of the authors of that time, *Munting*.

Table 2.1: Items and their values which were traded for a single Viceroy bulb (22).

Item	Value (florins)
Two lasts of wheat	448
Four lasts of rye	558
Four fat oxen	480
Eight fat swine	240
Twelve fat sheep	120
Two hogsheads of wine	70
Four casks of beer	32
Two tons of butter	192
A complete bed	100
A suit of clothes	80
A silver drinking cup	60
Total	**2,500**

As with all gambling mania, at the beginning, confidence was high because everybody was gaining. Tulip stocks were speculated in the rise and fall by the tulip-brokers. They earned much money, purchasing the stocks when prices dropped, and selling when they arose. Many people suddenly became rich. Everyone imagined that this process would last forever. Eventually, wise investors began to recognize that this imaginary world could not last forever (22). After that, as it happens in many cases of speculative bubbles, some prudent people started to sell and froze their profits. As everyone tried to sell while not many were buying, a domino effect became realized and prices became lower and lower. This caused people to panic and sell regardless of losses (7). Confidence was destroyed and the tulip market collapsed in February 1637, abruptly as shown in Figure 2.1.
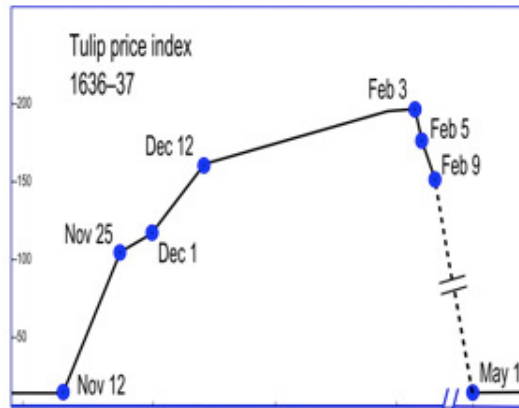
Figure 2.1: A standardized price index for tulip bulb contracts (21).

2.2 South Sea Bubble

The South Sea Company was established in 1711 by Earl of Oxford *Harley*. As a result of the war between Spain and Britain, British Government left 10 million pounds in debt. Harley offered the government paying back 10 million pounds of debt to ameliorate the government's financial condition. In turn, the government proposed 6% interest to the company (72). In addition to 6% interest, the government suggested privileged trading rights with Spain's colonies in the South Seas, today known as South America (16).

To finance its operations, the company issued stock to investors and the company's shares were snatched by investors, rapidly. After their first share issue success, the company issued more shares. Although the company had an inexperienced administration team, investors gathered this stock, competitively (16). It was thought that this company "could never fail".

The war between Britain and Spain began again in 1718; therefore all trading occasions stopped. However, this event did not deter the investors for buying stocks. After a while, management of the South Sea Company recognized that they were not generating much profit from company's operations. For this reason, to generate revenues, they decided to place more importance on issuing stock instead of making actual trade. Meanwhile, the leaders of the company decided to sell their shares since they noticed that the company's stocks were fabulously overvalued relative to its profit. At that time, the other investors did not realize that, actually, the firm was scarcely profitable (16).

Eventually, rumors spread among investors about the firm's administration had sold all of their shares in the company. Then, panic selling immediately started and soon the South Sea Company's shares became worthless as is evident from the Figure 2.2. *Isaac Newton* lost over 20,000 pounds. He stated "*I can calculate the motion of heavenly bodies, but not the madness of people*" (16). After the South Sea Bubble busted,

9

Britain's economy collapsed in spite of the government's endeavors and to heal completely almost took a century.
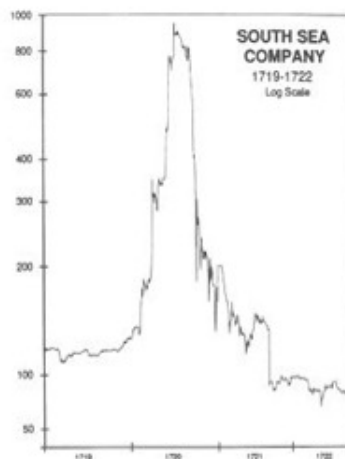
Figure 2.2: Log-scale price for south sea company stocks (18).

2.3 1929 Great Depression

The end of World War I brought a new epoch to America. An era of confidence, optimism and welfare were being experienced in the United States during 1920s. After World War I, industrialization and evolution of new technologies such as radio, automobile, air flight bolstered the economic and cultural boom. The *Dow Jones Industrial Average*, *DJIA*, increased throughout the 1920s and because of the country's strong economic conditions, most of the economists thought that shares were the most confident investment. These caused many investors to buy stocks, greedily (17). After a while, investors purchased stocks *on margin* which means that the buyer would invest some of his own money, but the rest were compensated by the broker. In those years, only 10 to 20% of the stock price had to be paid by the buyer and, hence, 80 to 90% of the cost of the stock price would be paid by the broker. If the price of the shares declined lower than the amount of the loan, the broker would probably issue a *margin call*, i.e., the buyer must pay back his loan as cash immediately. Therefore, to buy shares on margin could be very risky. However, in the 1920s, many people who expected to make a lot of money on the stock market easily, called *speculators*, purchased stocks on margin. They supposed that this rising process in prices would never end; so they could not recognize how serious the risk was which they were taking (66).

The Dow Jones bounced from 60 to 400 from 1921 to 1929. This generated a lot of new millionaires. Many people mortgaged their homes and invested their savings into stock market. However, few people really had knowledge about the companies in which they invested (17).

10

In 1929, from June through August, stocks prices reached their highest levels. Economist *Irving Fisher* stated that "*Stock prices have reached what looks like a permanently high plateau*", this was the statement many speculators wanted to hear and believe. On 3rd of September 1929, the DJIA closed at 381.17; hence, the stock market had reached its peak. Two days later, the market commenced falling. Stock prices fluctuated throughout September and into October (66).

As indicated in Figure 2.3, a strong bear market, a market condition in which the prices of shares were decreasing, had started by October 1929. On 24th of October 1929 which is known as *Black Thursday*, panic selling started since investors distinguished the stock market boom was actually an over inflated speculative bubble (17). Rumors spread that people commit suicide (66). Although the Federal Reserve Bank increased interest rates several times to relieve stock market and overheated economy in 1929, this could not prevent sad end.

When the stock market crashed on 28th and 29th of October, millionaire margin investors went bankrupt instantly. In November 1929, DJIA sharply declined from 400 to 145. Over $5 billion worth of market capitalization had been vanished from stocks that were trading on the New York Stock Exchange in just three days. The stock market crash of 1929 caused a great economic crisis, known as the *Great Depression* (17).

Figure 2.3: Dow Jones Industrial Average index from 1915 to 1942 (65).

As we observe from the famous bubble examples above, irrational expectations always trigger destructive financial crashes. Indeed, this big and yet vaguely understood phenomenon of bubbles asks for a high academic excellence. Academicians and central bankers still try to find a method to prevent from bubbles and to develop a model which estimates bubbles, in advance. If we go through the previous studies, in 1988, as one of the early articles that mentioned bubbles via a theoretical model, *Diba* and *Grossman* described a rational bubble as follows (25):

"*A rational bubble reflects a self-confirming belief that an asset's price depends on a variable (or a combination of variables) that is intrinsically irrelevant-that is, not part*

of market fundamentals-or on truly relevant variables in a way that involves parameters that are not part of market fundamentals."

In their study, they claim that a rational bubble begins just on the first date of trading of an asset. Moreover, they assert that a rational bubble that bursts cannot start again.

In 1989, *Camerer* represented theories of asset prices' deviations from fundamental value by classifying into three categories, which are *rational growing bubbles*, *fads* and *information bubbles* (13).

In 1990, *De Long* et al. investigated the role of rational speculators in markets described by positive feedback investors. In a short period of time, their model produces a positive correlation of stock returns when positive feedback investors' reply to past price rises by flowing into the market. However, a negative correlation of stock return is generated a long period of time. In addition to this, the model estimates the stock market reactions to news since news triggers positive feedback trading (24).

In 1993, Allen and Gorton presented a theoretical model in order to prove that settings can exist where rational behavior is consistent with stock price bubbles. They presume that all agents are rational but they populate an imperfect world that is identified by asymmetric information. In their model, an important property is that it can lead risk-loving behavior (1).

As we mentioned in Chapter 1, *Anders Johansen*, *Olivier Ledoit* and *Didier Sornette* represented a theoretical model, the so-called *Johansen-Ledoit-Sornette* (*JLS*) *model* in 1999 (49). In this model, the bubble process has a log-periodic behavior. Since the JLS model is successful to catch the bubbles, in our study, we benefitted from it to generate fitted data. However, in our study, we shall prefer to address the original data rather from the fitted data.

CHAPTER 3

THE JOHANSEN-LEDOIT-SORNETTE MODEL

At a symposium sponsored by the Federal Reserve Bank of Kansas City, in 2002, the former Federal Reserve chairman *Alan Greenspan* said: "... *we recorganized that, despite our suspicions, it was very difficult to definitively identify a bubble until after the fact - that is, when its bursting confirmed its existence*". In his speech, he declared that "*It seems reasonable to generalize from our recent experience that no low-risk, low-cost, incremental monetary tightening exists that can reliably deflate a bubble. But is there some policy that can at least limit the size of a bubble and, hence, its destructive fallout? From the evidence to date, the answer appears to be no. But we do need to know more about the behavior of equity premiums and bubbles and their impact on economic activity*" (35). As we understood from his statement, there is no exact theory which can estimate the bubbles previously, and as he underlined that how it important to identify them before they blow up.

Anders Johansen, *Olivier Ledoit* and *Didier Sornette* have proposed a theoretical model, known as the *Johansen-Ledoit-Sornette (JLS)* model, to designate bubbles in advance. By analyzing the cumulative human behaviors, the price dynamics is defined by the model during a time interval related with a bubble process (also called a *regime*). Also, after a bubble regime, an estimation for the most probable crash time has been provided by the JLS model. This model actually does not describe bubbles by exponential prices but rather by *faster-than-exponential growth* of price. The reason for this arises from imitation and herding behavior of noise traders, called as *irrational investors* (78).

3.1 Evolvement of the Price

In the JLS model, only a speculative asset is considered that does not pay dividends. The interest rate, risk aversion, information asymmetry, and the market-clearing condition are ignored for simplicity. Rational expectations are equivalent to the martingale hypothesis (70):

$$E_t\left[p(t^{'})\right] = p(t) \quad \forall t^{'} > t. \tag{3.1}$$

Here, the price of the asset at time t and the expectation conditional on information up to time t are shown by $p(t)$ and $E_t[\cdot]$, respectively. The solution of Eqn. (3.1) is a constant, if the asset price is not permitted to fluctuate under the effect of noise as follows:

$$p(t) = p(t_0).$$

Here, t_0 displays some initial time and the meaning of this equality is that there is no change in the price. Subsequently, when we study equations related with trajectories and solutions, we refer to the times $t \geq t_0$. Since the asset price pays no dividend, without loss of generality and for simplicity, its fundamental value is $p(t) = 0$. Thus, a positive value of $p(t)$ generates a speculative bubble (47). In the model, the crash is the jump, each jump is defined as a bubble and small jumps are not considered. A jump process can be indicated by j and the value zero is attained before the crash, and the value one after the crash, as represented in the following:

$$j = \begin{cases} 0, & \text{before crash,} \\ 1, & \text{after crash.} \end{cases}$$

As depicted in Figure 3.1, the cumulative distribution function (cdf) of the time of the crash, t_c, is called $1 - Q(t)$ and the probability density function (pdf) is

$$q(t) := \frac{dQ}{dt}(t) \approx \frac{\Delta Q}{\Delta t}.$$

The hazard rate of the crash is expressed as (47):

$$h(t) = q(t)/[1 - Q(t)],$$

where $Q(t) := Pr(t_c > t)$. The hazard rate $h(t)$ compares that change of probability with the likelihood $1 - Q(t) = Pr(t_c \leq t)$.
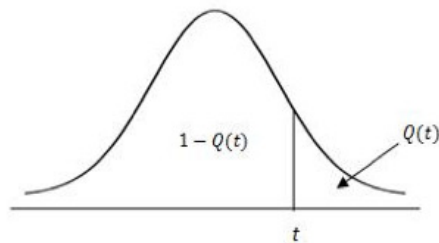
Figure 3.1: The cumulative distribution function (cdf) of the time of the crash.

According to the JLS model, the asset price is accepted to drop by a certain percentage $\varkappa \in (0, 1)$ in case of a crash, and the dynamics of the asset price before the crash is stated by (70):

$$dp(t) = \mu(t)p(t)dt - \varkappa p(t)d_j(t). \tag{3.2}$$

In order to satisfy the martingale condition for the price process, in Eqn. (3.2), the time dependent drift, $\mu(t)$ ($t \geq t_0$) is selected, namely,

$$E_t[dp(t)] = \mu(t)p(t)dt - \varkappa p(t)h(t)dt = 0.$$

Here, we would like to briefly mention the definition of a Martingale.

Definition 1: The concept of a martingale based on the conditional expectation, where $\mathbb{E}(Y|G)$ is interpreted as (usually optimized) numerical assesment of Y, given the information G. This information is expressed in terms of σ-algebras (or σ-fields). Let us assume that we have a rising family of σ-algebras, such that $F_s \subset F_t$, $t > s$, and for each $t \geq 0$ there exists a random variable Y_t which is F_t-measurable; $(Y_t)_{t \geq 0}$ is a process which is adapted to the filtration $(F_t)_{t \geq 0}$. The process $(Y_t)_{t \geq 0}$ is called a *martingale* according to $(F_t)_{t \geq 0}$ if $\mathbb{E}(Y_t|F_s) = Y_s$ for all $t > s$, almost everywhere in a probability space Ω, with respect to the given probability measure $\mathbb{P}$. Here, the main idea is that Y_t is a random variable such that the best estimation of its future value is given today's information/value (27). We might call it an "unbiased" estimation with respect to the underlying time. For further information, we refer to (68).

and, herewith,

$$\mu(t) = \varkappa h(t). \tag{3.3}$$

Substituting Eqn. (3.3) into Eqn. (3.2) will give us Eqn. (3.4) below:

$$dp(t) = \varkappa h(t)p(t)dt - \varkappa p(t)d_j(t). \tag{3.4}$$

While $j = 0$ being the case before the crash, we get an Ordinary Differential Equation, ODE:

$$dp = \varkappa h(t)p(t)dt \Rightarrow \frac{dp}{dt} = \varkappa h(t)p(t) \Rightarrow p'(t) = \varkappa h(t)p(t),$$

i.e.,

$$\frac{p'(t)}{p(t)} = \varkappa h(t). \tag{3.5}$$

Integrating the Eqn. (3.5) from both sides gives

$$\int_{t_0}^{t} \frac{p'(s)}{p(s)} ds = \int_{t_0}^{t} \varkappa h(s) ds.$$

Thus, the following expression is obtained

$$\log\left[\frac{p(t)}{p(t_0)}\right] = \varkappa \int_{t_0}^{t} h(s) ds \quad \text{before the crash.} \quad (3.6)$$

According to this new statement, under the condition that there is no crash, any increase of the probability of a crash implies an increase of the price speed in order to fulfill the martingale property. With (47) we note that the hazard rate $h(t)$ of the crash drives the price; that rate is not restricted by any bound so far.

To clarify the investors' behaviors, the model considers a *network* of agents. Each of the *agents* is indexed by an integer $i = 1, 2, \ldots, I$, and the set of agents is linked with i is indicated by $N(i)$. According to some graph like in Figure 3.2, the set of agents, $N(i)$, is directly linked with agent i. An assumption in the model is that agent i can have only one of two possible states, $s_i \in \{-1, +1\}$. Here, the values -1 and $+1$ can be interpreted as buy or sell, bullish or bearish, or, in technical or Boolean terms, "on or off". The state of investor i is decided by the model as follows (70):

$$s_i = \text{sgn}\left(K \sum_{j \in N(i)} s_j + \sigma \epsilon_i\right). \quad (3.7)$$

Here, K is a positive constant and ϵ_i is independently distributed with the standard normal distribution. The parameter K, called the *coupling strength*, controls the tendency towards imitation and σ governs the tendency towards *idiosyncratic* behavior. The parameter σ can be interpreted as an *environmental* factor. Bigger values of K reflect *strong imitation* (47). If $K = 0$, only environmental effects determine the decision of the trader.

The ratio K/σ represents the result of the order and disorder decision; in other words, it symbolizes the probability of a crash. Also, as long as the average of all values K was strictly positive, the model allows some of the values of K even to be negative. The meaning of the negative K is that the investor does not accept other investors' decisions.

As for the *susceptibility* of the system, it can be explained by adding a global influence term G to the Eqn. (3.7):

$$s_i = \text{sgn}\left(K \sum_{j \in N(i)} s_j + \sigma \epsilon_i + G \right). \qquad (3.8)$$

In Eqn. (3.8), $K \sum_{j \in N(i)} s_j + \sigma \epsilon_i$ corresponds to individual decision and G can be explained as an average group effect to the investor's decision. The parameter G will converge to the state $+1$ if $G > 0$; G will approximate the favour state -1 if $G < 0$ and, if $G = 0$, no global influence will exist.

The average state is defined as $M = (1/I) \cdot \sum_{i=1}^{I} s_i$. When there is no global influence, expectation of the average state is zero, $E(M) = 0$, i.e., agents are equally divided among the two states. Provided that a positive (negative) global influence is given, agents in the positive (negative) state will predominate the others, $E(M) \cdot G \geq 0$. With this notation, the *susceptibility* of the system is determined as (47):

$$\chi = \frac{d(E(M))}{dG}\Big|_{G=0}. \qquad (3.9)$$

The sensitivity of the average state to a small global influence is measured by this susceptibility. It also has a second interpretation, namely, as the expected squared deviance (variance) of the average state M around its mean of zero, induced by the random idiosyncratic shocks ϵ_i. The susceptibility is affected by the structure of the network. At the basis of the JLS model, there are two kinds of network structures: The 2-*dimensional grid* and the *hierarchical diamond lattice*.

3.2 Two-Dimensional Grid

In this network structure, the agents are placed on a 2-dimensional grid in the Euclidean plane. Each agent has 4 nearest neighbors: one to the North, one to the South, one to the East and one to the West. The related parametric ratio K/σ evaluates the propensity towards imitation relative to the tendency towards idiosyncratic behavior (47).

The properties of the system are arranged by a critical point K_c. While $K < K_c$, *disorder* decision wins. The meaning of this, the sensitivity is small for a small global influence, imitation only propagates, like in a *chain reaction* or even *cascading*, between close neighbors and the susceptibility χ of the system is finite. When K rises and approaches K_c, *order* decision reigns, i.e., the system is extremely sensitive to a small global perturbation, imitation propagates over long distances. In this situation, the susceptibility χ of the system approaches to infinity with respect to the power law. Regarding the definition and meaning of a *power law*, we refer to the works (15; 30; 75).

This behavior of the χ can be mathematically explained as follows (70):

$$\chi \approx A(K_c - K)^{-\gamma}, \qquad \gamma > 1. \tag{3.10}$$

Here, the parameter K does not even need to be deterministic, but it could also be a stochastic process, as long as it proceeds slowly enough. The value t_c is called the *first time* such that $K(t_c) = K_c$. Then, before the critical date t_c, the following approximation is obtained from Taylor expansion:

$$K_c - K(t) \approx \text{constant} \cdot (t_c - t).$$

Since the hazard rate of the crash behaves similar to the susceptibility around the critical point, this approximation fulfills the following statement:

$$h(t) \approx B \cdot (t_c - t)^{-\alpha}. \tag{3.11}$$

Here, B is a positive constant, and if the bubble has not bursted yet, to prevent the price from going to infinity when approaching t_c, α should be between 0 and 1. Since the crash could take place at some time prior to t_c, although the crash is not most probably, as stated in (47), t_c is not necessarily the time of the crash. Finally, substituting Eqn. (3.11) into Eqn. (3.6) gives us the following relation:

$$\log p(t) = \log p(t_c) + \varkappa B \int_{t_c}^{t} (t_c - s)^{-\alpha} ds,$$

i.e.,

$$\log p(t) = \log p(t_c) - \frac{\varkappa B}{1 - \alpha}(t_c - t)^{1-\alpha}.$$

Hence,

$$\log p(t) = \log p(t_c) - \frac{\varkappa B}{\beta}(t_c - t)^{\beta} \quad \text{before the crash.} \tag{3.12}$$

Since $\alpha \in (0, 1)$, $\beta := 1 - \alpha \in (0, 1)$ and $p(t_c)$ is the price at the critical time t_c. The logarithm of the price before the crash also obeys the power law. As the price process approaches the critical date, the slope of the log price (the expected return per unit of time) becomes limitless. This helps to compensate boundless probability values of the crash at the next instant (70).

18

3.3 Hierarchical Diamond Lattice

Another network structure is the hierarchical diamond lattice. To obtain this lattice, the process starts with a pair of investors who are connected to each other. Then, these connections are changed according to a *diamond* structure. On the diamond, two original investors reside two opposite vertices and where the two new investors occupy on the other two vertices. There are 4 links included by this diamond. For each one of these 4 links, it becomes altered by a diamond in the aforementioned way, and this procedure is iterated (71). Then, a diamond lattice is formed, as illustrated in Figure 3.2.
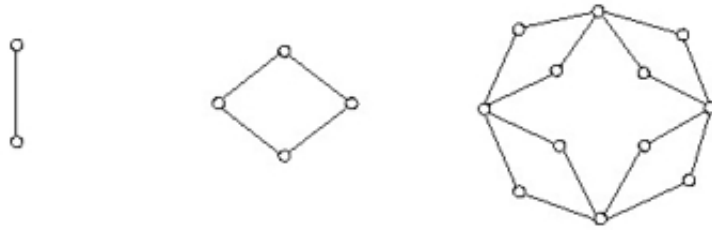
Figure 3.2: The relationship between traders and links.

Actually, in Figure 3.2, circles and edges show traders and links, respectively. After l iterations, $(2/3)(2 + 4^l)$ investors and 4^l connections between them are obtained (71). This can be proven by principle of mathematical induction.

The basic properties of the hierarchical diamond lattice are similar to the 2-dimensional grid. In this network structure, there is also a critical point K_c. The susceptibility is restricted while $K < K_c$, and it goes to infinity as K increases towards K_c. Unlike from a 2-dimensional grid, here, in the hierarchical diamond lattice, the critical exponent can be a complex number. The general solution for the susceptibility is a sum of terms like the one in $\chi \approx A(K_c - K)^{-\gamma}$, but now with complex exponents $-\gamma + iw$, etc. (47). This approximation is evaluated gradually as follows:

19

$$\chi \approx \text{Re}\left[A_0(K_c - K)^{-\gamma} + A_1(K_c - K)^{-\gamma+iw} + \ldots\right], \tag{3.13}$$

$$\chi \approx \text{Re}\left[A_0(K_c - K)^{-\gamma}\right] + \text{Re}\left[A_1(K_c - K)^{-\gamma+iw}\right] + \ldots,$$

$$\chi \approx \text{Re}\left[A_0(K_c - K)^{-\gamma}\right] + \text{Re}\left[A_1(K_c - K)^{-\gamma}(K_c - K)^{iw}\right] + \ldots,$$

$$\chi \approx \text{Re}\left[A_0(K_c - K)^{-\gamma}\right] + \text{Re}\left[A_1(K_c - K)^{-\gamma}e^{iw\log(K_c-K)}\right] + \ldots,$$

$$\chi \approx A_0(K_c - K)^{-\gamma} + A_1(K_c - K)^{-\gamma}\text{Re}\left[e^{iw\log(K_c-K)}\right] + \ldots,$$

$$\chi \approx A_0(K_c - K)^{-\gamma} +$$
$$A_1(K_c - K)^{-\gamma}\text{Re}\left[\cos(w\log(K_c - K)) + i\sin(w\log(K_c - K))\right] + \ldots,$$

$$\chi \approx A_0(K_c - K)^{-\gamma} + A_1(K_c - K)^{-\gamma}\cos\left[w\log(K_c - K) + \psi\right] + \ldots.$$

Here, $\text{Re}[\cdot]$ indicates the real part of a complex number and γ, A_0, A_1, w, ψ are real numbers. Hence, the *power law* is corrected by oscillations whose frequency busts as the price process approach the critical time. These accelerating oscillations are named as *log-periodic*, and their frequency $\lambda = w/2\pi$ is called the *log-frequency* (47). If the same steps are applied, Eqn. (3.14) is obtained for the hazard rate of a crash:

$$h(t) \approx B_0(t_c - t)^{-\alpha} + B_1(t_c - t)^{-\alpha}\cos\left[w\log(t_c - t) + \psi'\right]. \tag{3.14}$$

Substituting Eqn. (3.14) into Eqn. (3.12), we get:

$$\log\left[p(t)\right] = \log\left[p(t_c)\right] - \frac{\varkappa}{\beta}\left\{B_0(t_c - t)^{\beta} + B_1(t_c - t)^{\beta}\cos\left[w\log(t_c - t) + \phi\right]\right\}, \tag{3.15}$$

and Eqn. (3.15) can be rewritten as

$$y(t) \approx A + B(t_c - t)^{\beta} + C(t_c - t)^{\beta}\cos\left[w\log(t_c - t) + \phi\right], \tag{3.16}$$

where $A = \log\left[p(t_c)\right]$, $B = -\varkappa B_0/\beta$, $C = -\varkappa B_1/\beta$ and ϕ is another phase constant.

Eqn. (3.16) could be interpreted in terms of a *Stochastic Differential Equation* (*SDE*), equipped with an initial value, governing $y(t)$ (69):

$$\begin{cases} dY(t) = \mu(t, Y(t))\,dt + \sigma(t, Y(t))\,dW(t), \\ \\ Y_0 = y_0, \end{cases} \tag{3.17}$$

where $\mu(\cdot,\cdot) : [0, +\infty) \times \mathbb{R} \to \mathbb{R}$ and $\sigma(\cdot,\cdot) : [0, +\infty) \times \mathbb{R} \to \mathbb{R}$ are the coefficient functions which belong to drift part with its increment dt and to the diffusion part

with $dW(t)$, respectively. The initial value y_0 is an $\mathbb{R}$-valued random variable and the continuous-time stochastic process $W = (W(t)\ t \geq 0)$ subtends to a *Brownian motion*. An $\mathbb{R}$-valued stochastic process $Y = (Y(t)\ t \geq 0)$, if it exists, is a solution of the SDE defined in Eqn. (3.17). By $Y(t)$ we mean anyone of the trajectories $(y(t)) = (Y(t))_{(w)\ (w \in \Omega)}$; usually, we suppress the mentioning of w. The SDE which are derived by a *Brownian motion* is also called an *Itô Stochastic Differential Equation* (69). We recall that, usually, Lipschitz and growth conditions are assumed on the problem in Eqn. (3.17), for a "global" and unique solution. In fact, the right-hand side of Eqn. (3.16) involves a smoothening of the solution trajectory of the problem in Eqn. (3.17), where the SDE itself carries a structure which becomes inherited by its solution and its approximate structure in Eqn. (3.16). Here, we do not provide the technical details.

Because of the randomness that arises in the evolution of the stock prices as time passes, the dynamics of the stock prices can be modeled through SDEs. They are comprised of two parts: *drift* part and *diffusion* part. The random fluctuations in the evolution of the stock prices results from the initial condition and the noise generated by a Brownian motion (and possible jump processes). The random noise term belongs to the diffusion part of the SDE. On the other hand, the drift-term includes the deterministic movements in the stock prices (69):

$$\underbrace{A + B(t_c - t)^{\beta}}_{I} + \underbrace{C(t_c - t)^{\beta} \cos\left[w\log(t_c - t) + \phi\right]}_{II}.$$

In this respect, the first part of Eqn. (3.16), I, can be understood as the drift term and the second part, II, may be expressed as the diffusion term (possibly including an approximation or smoothening of impulses also)of an SDE. For closer information on SDEs that may also include jump process, we refer to (19).

If we reconsider the JLS model a bit differently and disregard some factors, then we note that the parameter β which particularly appears in the diffusion part may be regarded as a so-called *Hurst parameter*, H, of a *fractional Brownian motion*. From this point of view, the phenomena of less or more memory, can be represented by the cases $2H < 1$ (long-range dependence) and $2H > 1$ (short-range dependence), respectively, compared with the case of a standard Brownian motion, where $2H = 1$. For the interested reader, we refer to work (9; 63) for more details on fractional Brownian motions.

3.4 The Fitting Process

In the fitting process, so as to diminish the number of unknown parameters, three linearly embedded variables, A, B and C, have been optimally adjusted and computed from the obtained values of the nonlinear parameters. For this reason, approximation of Eqn. (3.16) is rewritten as (46)

21

$$y(t) \approx A + Bf(t) + Cg(t), \tag{3.18}$$

where

$$f(t) = \begin{cases} (t_c - t)^\beta, & \text{for a speculative bubble,} \\ (t - t_c)^\beta, & \text{for an antibubble,} \end{cases}$$

and

$$g(t) = \begin{cases} (t_c - t)^\beta \cos\left(w\log(t_c - t) + \phi\right), & \text{for a speculative bubble,} \\ (t - t_c)^\beta \cos\left(w\log(t - t_c) + \phi\right), & \text{for an antibubble.} \end{cases}$$

An *antibubble* is the opposite of the speculative bubble. Like a speculative bubble, it also follows a log-periodic power law (LPPL) but, of course, with decelerating oscillations and generally being bearish inclined instead of bullish. The term antibubble comes from the term *antiparticle* in physics, since an antiparticle is similar to its sibling particle, except that it conveys opposite charges and demolishes its sibling particle when it comes across with it (48).

The best values of the linear parameters are calculated for each choice of the nonlinear parameters by applying ordinary least-squares (OLS) method with the following Gaussian normal equations (46):

$$Xb = y \implies (X^T X)b = X^T y,$$

namely,

$$\begin{pmatrix} N & \sum_{i=1}^{N} f(t_i) & \sum_{i=1}^{N} g(t_i) \\ \sum_{i=1}^{N} f(t_i) & \sum_{i=1}^{N} f(t_i)^2 & \sum_{i=1}^{N} f(t_i)g(t_i) \\ \sum_{i=1}^{N} g(t_i) & \sum_{i=1}^{N} f(t_i)g(t_i) & \sum_{i=1}^{N} g(t_i)^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} \log p_i \\ \sum_{i=1}^{N} \log p_i \, f(t_i) \\ \sum_{i=1}^{N} \log p_i \, g(t_i) \end{pmatrix},$$

where

$$X := N \begin{pmatrix} 1 & f(t_1) & g(t_1) \\ \vdots & \vdots & \vdots \\ 1 & f(t_N) & g(t_N) \end{pmatrix}, \; b := \begin{pmatrix} A \\ B \\ C \end{pmatrix} \text{ and } y := \begin{pmatrix} \log p_1 \\ \vdots \\ \log p_N \end{pmatrix}.$$

Here, t_i ($i = 1, 2, \ldots, N$) are the times of the price or recording sampling (t_i, p_i). The general solution of this equation is given by

$$\hat{b} = (X^T X)^{-1} X^T y.$$

Of course, we may assume that the number of data is greater than the number of unknows, q, i.e., $N \geq q = 3$. Herewith, the design matrix X has a full rank. Now, the solutions of A, B and C become inserted into problem (3.18). In case of ill-conditionedness, we may apply a regularization technique, such as Tikhonov regularization (4). We are very careful before using such a technique here, in order not to regularize "off" just those details of information which interest us here most: the bubbles. Later on, we shall come to this important aspect more closely, (cf Subsection 5.1).

Fitting a function to data is a nonlinear estimation problem of the residuals sum of squares, RSS, where the objective function is defined by (46)

$$\min_{\theta} \quad F(\theta) := \sum_{i=1}^{N} (y_{\theta}(t_i) - y_i)^2 .$$

Here, $\theta = (t_c, \phi, w, \beta)^T$ is the vector of unknown parameters, and $y_{\theta}(\cdot) := y(\cdot)$, depending on θ.

The function of the residual sum of squares, F, is an analytically complicated or a strongly nonconvex function, and it can comprise multiple local minima with quite similar values. Here, the aim is to find the global minimum and to optimize the objective function with methods like the *downhill-simplex method* or the *quasi-Newton method*. However, this could be hazardous since these methods can be tricky by directing a local minimum rather than the global minimum (46). Therefore, more model-free *global optimization* solution methods were considered, such as *Simulated Annealing*, *Taboo Search* and *Genetic Algorithm*, etc.. Here, we used the Genetic Algorithm (32; 62; 64). Therefore, we would like to introduce this methodology, briefly.

3.5   Genetic Algorithm

The Genetic Algorithm (GA) is an evolutionary algorithm inspired by Charles Darwin's "description of natural selection". In first place, the theory of GA was developed by John Holland with his pioneer study in 1975. The principal utilities of this algorithm are, first of all, that there is no need of any particular information about the solution surface, such as slope or curvature (46). The GA can be performed to various optimization problems which are not suitable for standard optimization algorithms, e.g. , the GA is applicable when the objective function is discontinuous, non differentiable, stochastic, or highly nonlinear. Table 3.1 expresses, shortly, the differences of the genetic algorithm from a standard optimization algorithm in three principal manners (32).

Table 3.1: A basic comparison of Standard and Genetic Algorithm.

Standard Algorithm	Genetic Algorithm
At each iteration, a single point is produced.	At each iteration, a population of points is produced.
The iteration values approximate an optimal solution.	The population approximates an optimal solution.
Chooses the next point in the sequence by a deterministic calculation.	Chooses the next population by calculations which contain random selections.

The set of candidate solutions are generated contingently in GA; hence, it is also called as a *stochastic algorithm*. These randomly produced solutions are collected in a pool and all individual ones are in the vector form called *chromosomes*. After a fitness criterion is implemented to the solution space, the new generation is created from actual population with respect to *reproduction*, *mutation* and *crossover* (62).

3.5.1 Coding

In this step of the GA, generally, each individual solution is coded in strings which are so-called *chromosomes*. *Binary coding* is one of the common forms of the coding. It is comprised from components of the genes. Binary values are presented to each gene indicating, for instance, genes' functions in metabolic pathways (64).
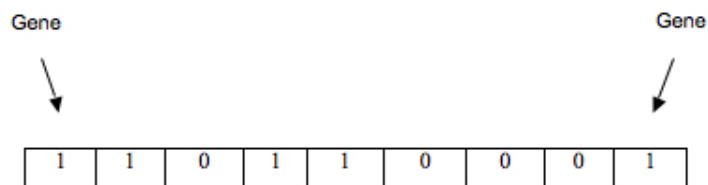
Figure 3.3: Representation of a chromosome where each box identifies one gene of chromosome.

3.5.2 Creating the Next Generation

At each step, the GA employs a current population to produce the *children* which bring out the next generation. The algorithm chooses a group of individuals in the current population which is called as *parents*, who give their genes- the entries of their vectors- to their children. Generally, the individuals that have better fitness values are taken as parents by the algorithm. The GA derives three types of children for the next generation (32). The first are *elite children*. They have the best fitness values in the

current generation. Automatically, these individuals survive to the next generation. The second are *crossover children*. These individuals are generated by combining the vectors of a pair of parents. The third group are *mutation children*. They are produced by representing random modifications, or mutations, to a single parent. Figure 3.4 displays these three types of children (32).
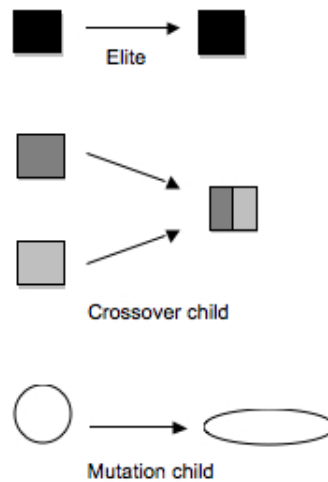
Figure 3.4: Diagram of three types of children created by the GA for the next generation.

3.5.3 Reproduction

In this step, the GA chooses the pairs of chromosomes to create offsprings which are the children of mated pairs. Reproduction take places randomly where the couples that abide by the fitness criteria more than the others; hence, they have a greater chance to be chosen (62). The algorithm sequences each chromosome according to its fitness in the population where the one with minimum error is going to be the first candidate to be selected. If this individual satisfies the given conditions, e.g., a tolerance value, then this vector of genes is selected by the algorithm. Otherwise, the GA proceeds the mutation or crossover operators (64).

Figure 3.5: A binary-valued, randomly produced pool.

In our study, in the initial population, each element of the population is a vector of the parameters (t_c, ϕ, w, z). The first population comprises elements with parameters drawn randomly from a uniform distribution with a pre-determined range.

3.5.4 Crossover

In the crossover step, the genetic characteristics between the individuals are redistributed (62). Variation arises within the two couples of strings which are cut off randomly selected points. Then, these cutted pairs exchange each other. Those exchanges can be performed in two ways:

1. The first part of the first string can be put at the place of the first part of the second, or the first parts can remain the same but the tails can be changed by each other.
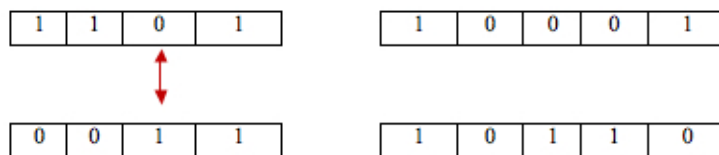
Figure 3.6: A crossover 1.

2. The selected parts of the cut strings can be crossovered.
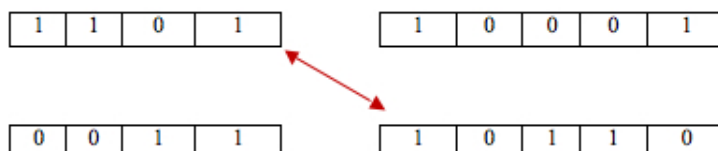
Figure 3.7: A crossover 2.

26

By randomly drawing two parents, an offspring is generated, without replacement, and taking the arithmetic mean of them.

3.5.5 Mutation

In nature, mutations happen when offspring are born with a variation within the gene pool and the outcomes of it can be beneficial or less favorable characteristics. Beneficial mutations are very crucial in the evolution process since it enhances the fitness of the species. In order to be able to survey the new regions of the search space, mutation operator of the GA perturbs the solutions. Hence, this mechanism obstructs premature convergence in a local minimum. Before running the mutation process, firstly, the range of maximum and minimum values of each coefficient are arranged in the current population. The parameters are free to drift away from the pre-set range of the parameter solutions thanks to the mutation mechanism. This could be a reparation if one took a wrong estimate related to the initial intervals in the solution space (46). Similar to the other operators, after chromosomes are mutated, they are controlled whether they match to the solution or not, and the algorithm proceeds until the stopping criterion is satisfied (64).

The following pseudo code sums up the whole steps of the GA.

Algorithm GA (64):

Step 1. Construct the population by the help of randomly generated vectors of solution candidates (reproduction).
Step 2. Check whether the candidate solutions in the pool fit the model.
Step 3. By placing the solution with a minimum error as a first candidate, order the candidate solutions with respect to their fitness.
Step 4. Perform mutation and crossover.
Step 5. Check a recent created pool within Step 4, whether it is fitting the model or not.
Step 6. Go to Step 3.
Step 7. If there exists a solution that confirms the stopping criteria: STOP; otherwise, go to Step 4.

CHAPTER 4

METHODOLOGICAL TOOLS FROM DATA MINING AND INVERSE PROBLEMS

In our model, we benefit from data mining by clustering methods. We redefine log price process/index and some variables which are obtained from this log price process/index, geometrically, by the help of ellipsoids. The most important property of the considered ellipsoids is that their volume is not an ordinary volume, but they have minimum volumes which guaranteed by *Minimum-Volume Covering Ellipsoids Clustering method*. This minimum-volume approach is fully compatible and supportive with respect to our aim of detecting volume changes of ellipsoids, including "needles" (almost pure steep "drifts") before a bubble bursts.

Hence, we partition the price process/index into the data clusters which are given over the time axis and represented by ellipsoids with a minimum volume. Then, we observe the behavior of the volumes via the sliding-window technique. Before expressing the minimum-volume covering ellipsoids method, firstly, we shall introduce the meaning of clustering in the following paragraph.

In fact, the clustering problem is to partition any data set by disjoint clusters which include the points that are considered to be similar. Mathematically represented, given a data set D with N points, $D = \{x_1, x_1, \ldots, x_N\} \subset \mathbb{R}^n$, suppose that the number of cluster is K ($1 \leq K \leq N$) and $D = C_1 \cup C_2 \cup \ldots \cup C_K$, where $C_i \cap C_j = \emptyset$ if $i \neq j$, i.e., the points of any two clusters are dissimilar. The meaning of *similar* is being close in terms of distances $d(x,y)$ between points $x, y \in \mathbb{R}^n$. In particular, if $K=1$ or $K=N$, the whole set D is one cluster or every point is a separate cluster, respectively (45).

To cope with difficulties such as clusters of equal size or of spherical shapes, Mahalanobis distance can be used instead of Euclidean distance. That measure is scale-invariant and it can cope with asymmetric, nonspherical clusters. A promising alternative scale-invariant metric of cluster quality is represented by minimum-volume ellipsoids, where data elements are allocated into clusters such that the volumes of the covering ellipsoids for every cluster are minimal (45).

Traditional distance-based clustering methods such as, *k*-means or *k*-median methods are not scale-invariant. However, clustering employing minimum-volume ellipsoids, that employs the minimum-volume covering ellipsoid for covering all points of any

regarded cluster and minimizes the total volume of those covering ellipsoids, has the so-called linear-transformation-invariance property (73).

4.1 Minimum-Volume Covering Ellipsoids Clustering Method

In the *minimum-volume covering ellipsoids* (*MVCE*) clustering method, the aim is to cover m given points $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ with an ellipsoid of minimum volume. In our study, each point refers to any week together with 7 different variables which are *date*, *logarithm of the price/index*, *weekly mean of log-price process*, *weekly standard deviation of log-price process*, *weekly median of log-price process*, *weekly growth rate of log-price process* and *weekly log-iterated return of the log price process*. That is to say, the vectors, a_i ($i = 1, 2, \ldots, m$) are corresponding to each week in a year, and as their coordinates they include date (*date*), log-price process (*log pr*), weekly mean (*mean*), weekly standard deviation (*std.dev*), weekly median (*med*), weekly growth rate (*gr rate*) and weekly log-iterated return of the logarithmic price process (*rtn*). Mathematically represented, this can be stated as follows:

$$
\begin{aligned}
a_1 &= (1, \log pr_1, \text{mean}_1, \text{std. dev}_1, \text{med}_1, \text{gr rate}_1, \text{rtn}_1)^T, \\
a_2 &= (2, \log pr_2, \text{mean}_2, \text{std. dev}_2, \text{med}_2, \text{gr rate}_2, \text{rtn}_2)^T, \\
&\vdots \\
a_m &= (m, \log pr_m, \text{mean}_m, \text{std. dev}_m, \text{med}_m, \text{gr rate}_m, \text{rtn}_m)^T.
\end{aligned}
$$

Here, the indices $1, 2, \ldots, m$, display the number of the week, i.e., 1 is the 1st week, 2 is the 2nd week, and so on.

We construct a matrix $\mathcal{A}$ whose format is $n \times m$ and its columns include each week $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$, i.e.,

$$
\mathcal{A} := \begin{bmatrix} a_1| a_2| \ldots| a_m \end{bmatrix}.
$$

This matrix $\mathcal{A}$, by its transposed matrix $\mathcal{A}^T$, is the design matrix of an inverse problem. Then, the ellipsoids are constituted by using the above matrix, and we apply *sliding-window* technique to consecutively observe the volumes for monitoring the behavior of the price process. In order to perform sliding-windows, the size of each window should be determined, previously. At this point, we benefit from the following assumption which guarantees that each ellipsoid, including the points $a_1, a_2, \ldots, a_m$, has a positive volume (73).

Assumption 1 The affine hull of the points $a_1, a_2, \ldots, a_m$, spans $\mathbb{R}^n$. In mathematical symbols:

$$
\text{rank} \begin{bmatrix} \mathcal{A} \\ e^T \end{bmatrix} = n + 1,
$$

where $e := (1, 1, \ldots, 1)^T \in \mathbb{R}^m$; hence $n < m$. We recall that $[e \,|\, \mathcal{A}^T]$ is the design matrix of a linear regression model, where $\mathcal{A}^T$ comprises the input data and e is related to the intercept.

In order to constitute the *sliding-windows*, we basically divide the matrix $\mathcal{A}$ according to the above assumption that warrants ellipsoids which have positive volumes and we also choose the intervals lengths small enough, to stay under a possible time range which include a bubble-burst time. For instance, if the optimal length (number of columns) of all the windows is $n + 1 = 8$, we partition the matrix $\mathcal{A}$ as 1:8, 2:9, 3:10, and so on. Also, if 9 was another alternative optimal length of each window which again gave positive volumes of the ellipsoids, we would prefer to choose 8 because it means a shorter interval than 9. The meaning of this is that, we will observe the price process with sections of 8 weeks instead of 9 or more weeks. Here, again, we should underline that these are not ordinary volumes, but by the use of MVCE clustering method, we guarantee that these volumes are minimal. By this property we intend to approximate and, in fact, neatly reflect, typical behaviors of the regarded process, e.g., high oscillation behavior followed by a sharp increase before a bubble bursts.

An ellipsoid is defined as

$$E_{Q,\mu} := \left\{ x \in \mathbb{R}^n \,\middle|\, (x - \mu)^T Q (x - \mu) \leq 1 \right\}.$$

Here, $\mu \in \mathbb{R}^n$ denotes the center of the ellipsoid and the general shape, directions and size of the ellipsoid are determined by $Q \in S_{++}^n$. The notation S_{++}^n displays the convex cone of the symmetric positive definite $n \times n$ matrices Q, in a symbol: $Q > 0$. By " $>$ ", we indicate positive definiteness. In addition, throughout the thesis, by " $\geq$ " and " $\leq$ ", we will display positive and negative semi-definiteness, respectively. The volume of the ellipsoid $E_{Q,\mu}$ is given by the formula $(\pi^{n/2} / \Gamma(\frac{n}{2} + 1)) \cdot (1/\sqrt{\det Q})$ (73).

Under Assumption 1, the formulation of the minimum-volume covering ellipsoid (MVCE) problem is given as follows (73):

$$MVCE_1 \qquad \min_{Q,\mu} \qquad \det\ Q^{-1/2}$$

$$\text{subject to} \quad (a_i - \mu)^T Q (a_i - \mu) \leq 1 \quad (i = 1, 2, \ldots, m),$$

$$Q > 0.$$

We recall that in the constraints of the optimization problem $MVCE_1$, " $>$ " has been used to represent the partial ordering induced by the cone S_{++}^n. Here, $MVCE_1$ is a *nonconvex* program. In order to transform this optimization problem into a *convex* program, the variables are changed by the following transformation:

$$\mathcal{M} := Q^{1/2} \quad \text{and} \quad z := \mathcal{M}\mu.$$

Hence, the program $MVCE_1$ is equivalently rewritten as

$MVCE_2$ $\quad \min_{\mathcal{M},z} \quad \varphi(\mathcal{M},z) := -\ln \det \mathcal{M}$

$$\text{subject to} \quad (\mathcal{M}a_i - z)^T(\mathcal{M}a_i - z) \leq 1 \quad (i = 1, 2, \ldots, m), \quad (4.1)$$

$$\mathcal{M} > 0.$$

Now, $MVCE_2$ is a convex program. If $(\overline{\mathcal{M}}, \overline{z})$ is a solution of the problem $MVCE_2$, the solution of problem $MVCE_1$ can be found by setting $(\overline{Q}, \overline{\mu}) := (\overline{\mathcal{M}}^2, \overline{\mathcal{M}}^{-1}\overline{z})$ (73).

For the MVCE problem, in 2002, *Freund* and *Sun* introduced an algorithm which is called *Dual Reduced Newton* (*DRN*). In the Newton step of this algorithm, the following optimization problem obtained by adding slack variables t_i and a related logarithmic barrier function to the $MVCE_2$ (73):

$MVCE_3$ $\quad \min_{\mathcal{M},z,t} \quad -\ln \det \mathcal{M} - \theta \sum_{i=1}^{m} \ln t_i$

$$\text{subject to} \quad (\mathcal{M}a_i - z)^T(\mathcal{M}a_i - z) + t_i = 1 \quad (i = 1, 2, \ldots, m),$$

$$\mathcal{M} > 0,$$
$$t > 0.$$

Here, $t := (t_1, t_2, \ldots, t_m)^T$ and $t > 0$ is understood coordinate wise. Likewise, throughout the thesis, we mean " $\geq$ ", " $\leq$ " and " $<$ " in the sense given by the coordinates. Besides, as θ takes its values between 0 and ∞, the parametrized solutions to $MVCE_3$ describe the *central trajectory* of the problem $MVCE_2$. The optimality conditions for $MVCE_3$ are expressed by the Eqns. (4.2)-(4.7) referring to *dual multipliers* u_i $(i = 1, 2, \ldots, m)$ that are related with the equality constraints in $MVCE_3$, look as follows (73):

$$\sum_{i=1}^{m} u_i \left[(\mathcal{M}a_i - z)a_i^T + a_i(\mathcal{M}a_i - z)^T\right] - \mathcal{M}^{-1} = 0, \quad (4.2)$$

$$\sum_{i=1}^{m} u_i(z - \mathcal{M}a_i) = 0, \quad (4.3)$$

$$(\mathcal{M}a_i - z)^T(\mathcal{M}a_i - z) + t_i = 1 \quad (i = 1, 2, \ldots, m), \quad (4.4)$$

$$Ut = \theta e, \quad (4.5)$$

$$u, t \geq 0, \quad (4.6)$$

32

$$\mathcal{M} > 0. \tag{4.7}$$

Eqns. (4.2)-(4.7) are solved for $(\mathcal{M}, z, t, u)$ directly by using Newton's method. This is done by resolving Eqn. (4.3) for z and obtained Eqn. (4.8). Furthermore, in Eqns. (4.2)-(4.7), $e = (1, 1, \ldots, 1)^T$, and U stands for a diagonal matrix whose diagonal components are entries of the vector u, i.e., $U :=\text{diag}(u)$.

$$z = \frac{\mathcal{M} \mathcal{A} u}{e^T u}. \tag{4.8}$$

Substituting Eqn. (4.8) into Eqn. (4.2), the following condition for the matrix $\mathcal{M}$ is found:

$$\left(\mathcal{A} U \mathcal{A}^T - \frac{\mathcal{A} u u^T \mathcal{A}^T}{e^T u} \right) \mathcal{M} + \mathcal{M}^T \left(\mathcal{A} U \mathcal{A}^T - \frac{\mathcal{A} u u^T \mathcal{A}^T}{e^T u} \right) = \mathcal{M}^{-1}. \tag{4.9}$$

An important feature of the matrix appearing in Eqn. (4.9) is defined in Proposition 1 (73).

Proposition 1 (73) Under Assumption 1, in the case of $u > 0$, then $\left(\mathcal{A} U \mathcal{A}^T - \mathcal{A} u u^T \mathcal{A}^T /(e^T u) \right) > 0$.

For the proof of Proposition 1, we refer to (73). The following remark introduces a closed form solution for Eqn. (4.9).

Remark 1 For a given $S \in S_{++}^n$, $X = S^{1/2}$ denotes the unique positive definite matrix solving the following equation (73):

$$\frac{1}{2}(X^T S + S X) = X^{-1}.$$

For closer information on $S^{1/2}$ cf. (80). By the help of Proposition 1 and Remark 1, the unique solution of Eqn. (4.9) is derived as follows (73):

$$\mathcal{M} := \mathcal{M}(u) := \left[2 \left(\mathcal{A} U \mathcal{A}^T - \frac{\mathcal{A} u u^T \mathcal{A}^T}{e^T u} \right) \right]^{-1/2}. \tag{4.10}$$

Substituting Eqn. (4.10) into Eqn. (4.8), the following proposition is obtained.

Proposition 2 (73) Under Assumption 1, provided that $u > 0$, then the unique solver of Eqns. (4.2)-(4.3) and Eqn. (4.7) in $\mathcal{M}$ and z is given by:

$$\mathcal{M} := \mathcal{M}(u) := \left[2 \left(\mathcal{A} U \mathcal{A}^T - \frac{\mathcal{A} u u^T \mathcal{A}^T}{e^T u} \right) \right]^{-1/2} \tag{4.11}$$

and

33

$$z := z(u) := \frac{\left[2\left(\mathcal{A}U\mathcal{A}^T - \frac{\mathcal{A}uu^T\mathcal{A}^T}{e^T u}\right)\right]^{-1/2}\mathcal{A}u}{e^T u}. \tag{4.12}$$

Substituting Eqns. (4.11) and (4.12) into the optimality conditions of Eqns. (4.2)-(4.7), the variables $\mathcal{M}$ and z be explicitly eliminated from the optimality conditions. Hence, the subsequent reduced optimality conditions which just involve the variables (u,t) are obtained (73):

$$\begin{aligned} h(u) + t \quad &= e, \\ Ut \quad &= \theta e, \\ u, t \quad &\geq 0. \end{aligned} \tag{4.13}$$

Here, $h = (h_i)_{i=1,\dots,m}^T$, and each h_i is a nonlinear function of u, defined as follows:

$$\begin{aligned} h_i(u) \quad &:= (\mathcal{M}(u)a_i - z(u))^T(\mathcal{M}(u)a_i - z(u)) \\ &= (a_i - \tfrac{\mathcal{A}u}{e^T u})^T \left[2\left(\mathcal{A}U\mathcal{A}^T - \tfrac{\mathcal{A}uu^T\mathcal{A}^T}{e^T u}\right)\right]^{-1}(a_i - \tfrac{\mathcal{A}u}{e^T u}). \end{aligned} \tag{4.14}$$

Eqn. (4.13) is solved by employing Newton's method for a given value of the barrier parameter θ. The solution of the following system of linear equations gives the Newton direction $(\Delta u, \Delta t)$ for Eqn. (4.13) at the point (u, t):

$$\begin{aligned} \nabla_u h(u)\Delta u + \Delta t \quad &= e - t - h(u) =: r_1, \\ T\Delta u + U\Delta t = \quad &\theta e - Ut =: r_2. \end{aligned} \tag{4.15}$$

where $\nabla_u h(u)$ is the Jacobian matrix of $h(u)$, T demonstrates a diagonal matrix whose diagonal entries are the entries of the vector t, i.e., $T :=\text{diag}(t)$, and the unique solution $(\Delta u, \Delta t)$ of the system is given in the subsequent form:

$$\begin{aligned} \Delta u \quad &= (\nabla_u h(u) - U^{-1}T)^{-1}(r_1 - U^{-1}r_2), \\ \Delta t \quad &= U^{-1}r_2 - U^{-1}T\Delta u. \end{aligned} \tag{4.16}$$

In order to perform the above methodology, $\nabla_u h(u)$ and the inverse of the matrix $\nabla_u h(u) - U^{-1}T$ must be calculated explicitely. Firstly, the matrix $\nabla_u h(u) - U^{-1}T$ should be guaranteed to be nonsingular in order to be invertible; this will be done by Corollary 1. As a preparation, the following theorem supplies a formula to compute the Jacobian matrix $\nabla_u h(u)$ and $h(u)$.

Theorem 1 (73) Under Assumption 1,

(i) $\nabla_u h(u) = -2\left(\sum(u)/(e^T u) + \sum(u) \otimes \sum(u)\right),$

(ii) $\nabla_u h(u) \leq 0,$

(iii) $h(u) = \text{diag}(\sum(u)).$

Here, $A \otimes B$ displays the Hadamard product of two matrices A and B, i.e., $(A \otimes B)_{ij} := A_{ij}B_{ij}$ for $i, j = 1, \ldots, m$, and the matrix function $\sum(u)$ is defined as in Eqn. (4.17) as follows:

$$\sum(u) := \left(\mathscr{A} - \frac{\mathscr{A}ue^T}{e^T u}\right)^T \mathscr{M}^2 u \left(\mathscr{A} - \frac{\mathscr{A}ue^T}{e^T u}\right). \tag{4.17}$$

This matrix $\sum(u)$ can be interpreted as a variance-covariance matrix. Soon, its regularity properties will interest us very much. Furthermore, we could express it in terms of the *Fisher* information content; for further information cf. (20; 38)

Proof: Let $C(u) := \mathscr{M}^2(u) = \left[2\left(\mathscr{A}U\mathscr{A}^T - \frac{\mathscr{A}uu^T\mathscr{A}^T}{e^T u}\right)\right]^{-1}$ and $\tilde{a}_i(u) := a_i - \frac{\mathscr{A}u}{e^T u}$. From the definition of $\sum(u)$, we have

$$\sigma_{ij}(u) := [\sum(u)]_{ij} = (\tilde{a}_i(u))^T \mathscr{M}^2(u)((\tilde{a}_j(u)) \tag{4.18}$$

and, thus, $h_i(u) = (\sum(u))_{ii}$ from Eqn. (4.14), which displays part (iii) of Theorem 1. To calculate $\nabla_u h(u)$, the chain rule is applied. Hence, the following statement is obtained.

$$\frac{\partial \tilde{a}_i(u)}{\partial u_j} = \frac{\mathscr{A}u}{(e^T u)^2} - \frac{a_j}{e^T u} = \frac{-\tilde{a}_j}{e^T u}$$

and

$$\frac{\partial C(u)}{\partial u_j} = -2C(u)\left[a_j a_j^T + \frac{\mathscr{A}uu^T\mathscr{A}^T}{(e^T u)^2} - \frac{a_j u^T \mathscr{A}^T}{e^T u} - \frac{\mathscr{A}ua_j^T}{e^T u}\right]C(u)$$

$$= -2C(u)\tilde{a}_j(u)(\tilde{a}_j(u))^T C(u).$$

Then, we invoke the chain rule on Eqn. (4.18):

$$\frac{\partial h_i(u)}{\partial u_j} = \frac{-2}{e^T u}(\tilde{a}_i(u))^T C(u)\tilde{a}_j(u) - 2(\tilde{a}_i(u))^T C(u)\tilde{a}_j(u)(\tilde{a}_j(u))^T C(u)\tilde{a}_i(u)$$

$$= -2\left(\frac{\sigma_{ij}(u)}{e^T u} + (\sigma_{ij}(u))^2\right).$$

Therefore, $\nabla_u h(u) = -2\left(\sum(u)/e^T u + \sum(u) \otimes \sum(u)\right)$, proving part (i). Let us note $\sum(u) \succeq 0$, $e^T u > 0$, and $\sum(u) \otimes \sum(u) \succeq 0$ since the Hadamard product of symmetric and positive semidefinite matrices is symmetric and positive semidefinite (73). Therefore, $\nabla_u h(u) \leq 0$, proving part (ii).

Corollary 1 (73) *Under the Assumption* 1, *provided that $u > 0$ and $t > 0$, the matrix* $\nabla_u h(u) - U^{-1}T$ *is nonsingular.*

Proof: From Theorem 1(ii) and the fact that $U^{-1}T > 0$, it is obvious that $\nabla_u h(u) - U^{-1}T$ is nonsingular.

The following procedure is performed to compute the Newton direction $(\Delta u, \Delta t)$ for the reduced optimality conditions of Eqn. (4.13) at a regarded point (u, t).

Algorithm for DRN-DIRECTION (73): Given (u, t, θ) which satisfy $u, t, \theta \geq 0$:

Step 1. Build and factorize (by Cholesky method) the matrix

$$\mathcal{M}^{-2}(u) := \left[2 \left(\mathcal{A}U\mathcal{A}^T - \frac{\mathcal{A}uu^T\mathcal{A}^T}{e^T u} \right) \right].$$

Step 2. Form the matrix

$$\sum(u) = \left(\mathcal{A} - \frac{\mathcal{A}ue^T}{e^T u} \right)^T \mathcal{M}^2 u \left(\mathcal{A} - \frac{\mathcal{A}ue^T}{e^T u} \right).$$

Step 3. Form

$$\nabla_u h(u) = -2 \left(\frac{\sum(u)}{e^T u} + \sum(u) \otimes \sum(u) \right) \text{ and factorize (by Cholesky method)}, \nabla_u h(u) - U^{-1}T.$$

Step 4. Resolve Eqn. (4.16) for $(\Delta u, \Delta t)$.

The variable (u, t) is called the *DRN direction* for *Dual Reduced Newton method* (73).

4.1.1 Algorithm DRN

In order to solve the optimization problem $MVCE_3$, for the $MVCE$ method, the following basic interior-point method is constructed based on the Newton step procedure (73). We refer to preselected tolerances $\epsilon_1, \epsilon_2 > 0$.

Algorithm DRN

Step 1. Initialization.
Put $r \leftarrow 0.99$. Select initial values of (u^0, t^0) satisfying $u^0, t^0 > 0$. Put $(u, t) \leftarrow (u^0, t^0)$.

Step 2. Control Stopping Criteria.
If $\|e - h(u) - t\|_2 \leq \epsilon_1$ and $u^T t \leq \epsilon_2$, STOP. Return u, $Q := [\mathcal{M}(u)]^2$, and $c := [\mathcal{M}(u)]^{-1} z(u)$.

Step 3. Calculate Direction.
Put $\theta \leftarrow (u^T t)/(10m)$. Calculate $(\Delta u, \Delta t)$ by Algorithm DRN-DIRECTION (u, t, θ).

Step 4. Step-Size Calculation and Step.
Calculate $\bar{\beta} \leftarrow \max\{\beta \,|\, (u, t) + \beta(\Delta u, \Delta t) \geq 0\}$ and $\tilde{\beta} \leftarrow \min\left\{ r\bar{\beta}, 1 \right\}$.

Put $(u, t) \leftarrow (u, t) + \tilde{\beta}(\Delta u, \Delta t)$. Go to **Step 1**.

In order to maintain the iteration value of (u, t) strictly positive, the condition $r < 1$ is applied (73); this can be seen in Step 4. In Step 1, the stopping criteria are controlled. The tolerances ϵ_1 and ϵ_2 are used for feasibility and optimality, respectively. The barrier parameter value θ is redefined and the DRN direction is calculated in Step 2. Also, the norm $\|d\|_2$ in Step 2 is the Euclidean norm identified as $\sqrt{d^T d}$. In Step 3, the calculation of step size done by a standard min-ratio test augmented by a fraction $r \in (0, 1.0)$ which maintains the new iteration value of (u, t) being strictly positive. To take $r = 0.99$ has been found to work best (73).

4.1.1.1   Initialization Strategies

The Algorithm DRN is started choosing a pair (u^0, t^0) that fulfills $u^0, t^0 > 0$. For instance, $(u^0, t^0) = (\alpha e, \alpha e)$ for some appropriate positive scalar α. The pair (u^0, t^0) is selected such that these values warrant the initial primal feasibility of $\mathcal{M}(u^0), z(u^0)$. The algorithm begins by arranging $u^0 = (n/2m)e$; the factor $n/2m$ has been selected empirically. Then, $\mathcal{M}(u^0)$ and $z(u^0)$ are calculated by Eqns. (4.11)-(4.12). For the strict primal feasibility, the condition $h(u^0) \leq (0.95)e$ is tested. If this condition is satisfied, then t^0 is set as $t^0 := e - h(u^0) > 0$. Hence, the positivity of (u^0, t^0) and initial feasibility of the equations $h(u)+t = e$ at $(u, t) = (u^0, t^0)$ are guaranteed. If the condition $h(u^0) \leq (0.95)e$ is not satisfied, i.e., $h(u^0) \not\leq (0.95)e$, because of $h(\alpha u) = (1/\alpha)h(u)$ by Eqn. (4.14), u^0 is rescaled to guarantee strict feasibility of the algorithm as follows:

$$\alpha = (1/0.95) \max \left\{ h_1(u^0), \ldots, h_m(u^0) \right\},$$

$$u^0 \leftarrow \alpha u^0,$$
$$t^0 \leftarrow e - h(u^0). \tag{4.19}$$

This initialization strategy then assures strict positivity of (u^0, t^0) also initial feasibility of the equations $h(u) + t = e$ at $(u, t) = (u^0, t^0)$ (73).

4.1.1.2   Stopping Criteria

The following result is the basis of the stopping criteria of Algorithm DRN.

Lemma 7. *Under Assumption 1, assume that $u > 0$. If $h(u) \leq e$, then $(\mathcal{M}, z) = (\mathcal{M}(u), z(u))$ is feasible for program $MVCE_2$ and $\varphi(\mathcal{M}, z) - u^T t$ is a lower bound on the optimal objective function value of $MVCE_2$.*

Lemma 7 explains that the optimality gap of a feasible solution $(\mathcal{M}, z) = (\mathcal{M}(u), z(u))$ of $MVCE_2$ is at most $u^T t$, where $t = e - h(u) \geq 0$. The stopping criteria of Algorithm DRN, specified in Step 1, is to control that primal feasibility is fulfilled to a prespecified tolerance ϵ_1, and then to control whether the relative optimality gap is not greater than

ϵ_2, where ϵ_2 is the prespecified tolerance. We take $\epsilon_1 = \epsilon_2 := 10^{-7}$, in the calculations (73).

4.2   Radon Transform

In order to view the ellipsoids, we benefit from Radon transform which is one of the important image transform methods (39). Common image transforms contain Radon transform, Hough transform or discrete Fourier transform. Actually, Hough transform is a special case of Radon for straight lines, and since Radon transform has a solid mathematical basis and is more accurate (33; 26), we prefer to apply Radon transformation.

With our new method, we aim at an *early-warning signalling* which, eventually, will have the form of a *graphic user interface* (*GUI*) too (8). In this respect, the theory of inverse problems provides a scientific tradition and, indeed, practical approaches such as the use of "reshape" of vectors and matrices, of squares and rectangles, and of colors. Spikes and more complex data or model structures can become much more appealing, intuitive and informative (4). For our problem from finance and economics, this new approach to that area becomes most powerful and convincing when choosing the representation tool of Radon transform.

The *Radon transform* (*RT*) entitled after the Austrian mathematician *Johann Karl August Radon* (1887 - 1956) who worked on the properties of this transform (29). As in our study on MVCE clustering, Radon transform generally addresses an inverse problem, namely, in image processing, tomography, discrete tomography, and signal processing (40). To implement the Radon transform on an image $f(x, y)$ for a definite set of angles can be thought of as detecting the projection of the image along the given angles (29). The resultant projection is the sum of the "intensities" of the pixels in any regarded direction, i.e., by line integrals, such as straight "beams". In other words, for a given function f determined in the space, for any pair of coordinate values ρ and θ, f is integrated along a different line. Hence, the result of this transformation is a new function (or a new image), $R(\rho, \theta)$, where the inputs are the values of ρ and θ and the output is the value of the integral of f along the corresponding line $l_{\rho,\theta}$. This can be explained by the following scheme (29):

$$\text{input } f \longmapsto \text{output} \left\{ (\rho, \theta) \longmapsto \int_{l_{\rho,\theta}} f ds \right\}.$$

This may be written mathematically by defining

$$\rho := x \cos\theta + y \sin\theta. \tag{4.20}$$

Referring to this and to $\delta(\cdot)$ which is the Dirac delta impulse, the Radon transform can be defined as (42)

38

$$R(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\delta(\rho - x \cos\theta - y \sin\theta)dx \, dy, \qquad (4.21)$$

or

$$R(\rho, \theta) = \int_{l_{\rho,\theta}} f ds = \int_{-\infty}^{\infty} f(\rho\cos\theta - s \sin\theta, \rho\sin\theta + s \cos\theta)ds. \qquad (4.22)$$

We notice in Eqn. (4.21) that (i) both $f(x, y)$ and $R(\rho, \theta)$ are functions; (ii) f is a function of the Cartesian coordinates x and y, while $R(\rho, \theta)$ is a function of the polar coordinates ρ and θ; (iii) $R(\rho, \theta)$ is a finite number (the value of an integral) and (iv) in Eqn. (4.22), within of the integral, the integration variable is s, whereas values ρ and θ are preselected. Therefore, ρ and θ are considered as constants when calculating the integral (29).

The Radon transform is a mapping from the Cartesian rectangular coordinates (x, y) to a distance and an angle (ρ, θ) (polar coordinates). An example of the Radon transform of an image is given in Figure 4.1 for a specific angle. There, the source and sensor is rotated around the center of the object. For each angle θ, the rays from the source which pass through are collected at the sensor. This is iterated for a given set of angles, generally, $\theta \in [0, 180)$. Since the outcome would be the same with the case of angle 0, the angle 180 is not included. At (b), for each angle θ and each distance ρ, the intensity of the matter a ray perpendicular to the ρ axis crosses are summed up at $R(\rho, \theta)$ (42).

(a) (b)
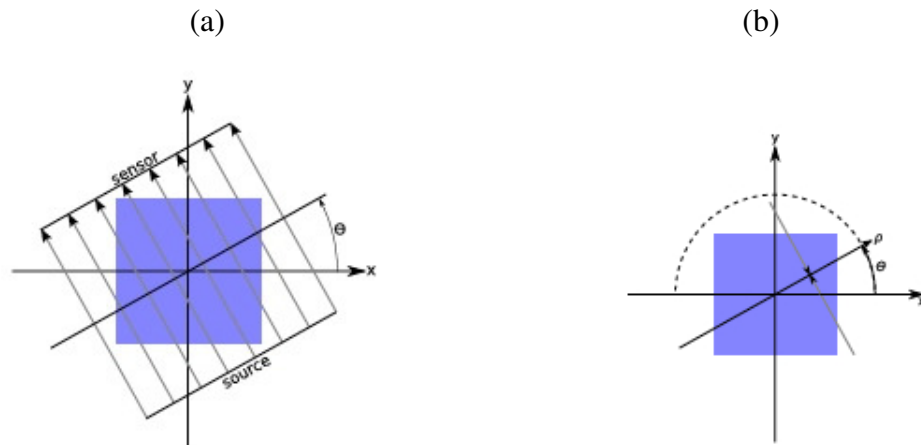
Figure 4.1: An example of the Radon transform of an image for a specific angle θ (42).

The transform for a set of angles can be depicted as in Figures 4.2 and 4.3. In Figure 4.2 (c), the angle θ is taken as $19°$ and the corresponding Radon transform with $\theta = 19°$ is shown in (d); as we see, there is no peak at the transform. Yet, in Figure 4.3 (f), we observe a definite peak for Radon transform which subtends to $\theta = 64°$. Other

observations in (*d*) and (*f*) are that for $\theta = 64°$ the line is distributed over a very small interval unlike for $\theta = 19°$ for which the line is distributed over a larger interval.

(c)   (d)
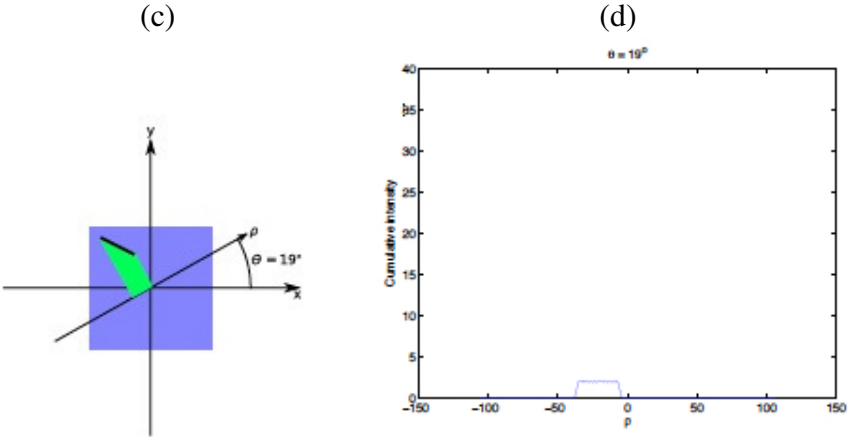
Figure 4.2: An example of the Radon transform for the angle $\theta = 19°$ (42).

(e)   (f)
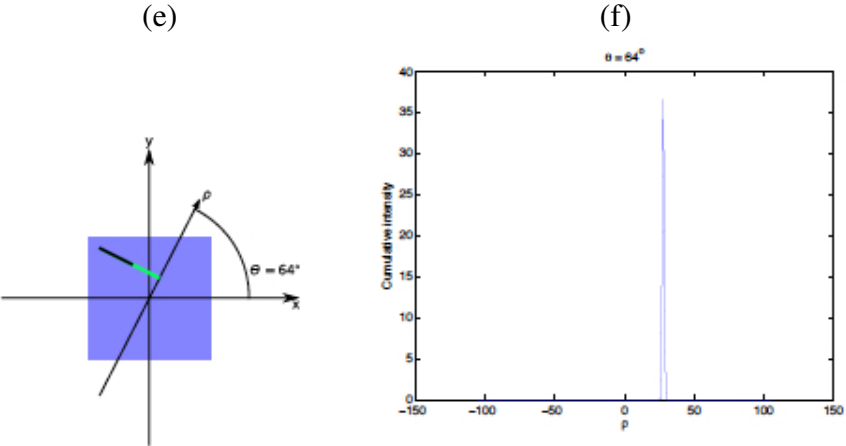
Figure 4.3: An example of the Radon transform when $\theta = 64°$ (42).

The complete Radon transform of an image is indicated in Figure 4.4. This means that, now, there is not only a specific θ, but θ ranges between 0 and 180°, $\theta \in [0, 180)$. The horizontal axis displays θ (degrees) and the vertical axis shows the corresponding coordinates when the Cartesian x-axis rotates according to these angles. In addition, at the right-hand side of Figure 4.4, we see a colorbar which indicates the color scale when we color a figure with the Matlab command *colormap*.
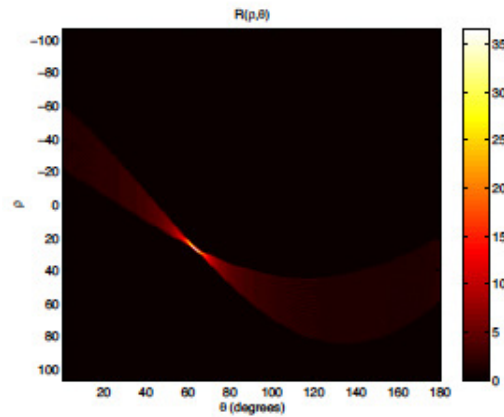
Figure 4.4: The complete Radon transform of the image (42).

CHAPTER 5

DETAILS OF MODELING AND MODEL TREATMENT

5.1 Details of General Approach

Firstly, we constitute the matrix $\mathcal{A}$ by using the variables *date*, *logarithm of the price*, *weekly mean of log-price process*, *weekly standard deviation of log-price process*, *weekly median of log-price process*, *weekly growth rate of log-price* and *weekly log-iterated return of the log-price process*. In the matrix $\mathcal{A}$, rows and columns correspond to variables and weeks, respectively. In order to observe the change of ellipsoids's volumes, when the bubble-burst time approaches, we apply sliding-window technique on the $\mathcal{A}$, and we divide the matrix $\mathcal{A}$ into submatrices with the "smallest size", $\mathcal{A}_k$, such as, $\mathcal{A}_1$ includes the columns from 1 to 8, $\mathcal{A}_2$ contains the columns from 2 to 9, and so on. Then, we obtain ellipsoids with minimum volumes by using these submatrices. In fact, this is done according to the rank criterion in Assumption 1 which guarantees that the ellipsoids have positive volumes. Here, the important issue is that all the windows have the same length, i.e., all the submatrices have the same dimensions, and the reason why we choose the smallest length, i.e., the smallest size of the submatrices is for giving a chance to observe the possibility of any burst of a bubble as early as possible.

As for application, we found the possible smallest size of each submatrix, i.e., where each submatrix corresponds to a certain window, to be 8 which is also satisfied by the rank condition in Assumption 1. This means that the size of the each window is 7×8. Here, 7 is the number of variables (*date*, *logarithm of the price*, *weekly mean of log-price process*, *weekly standard deviation of log-price process*, *weekly median of log-price process*, *weekly growth rate of log-price* and *weekly log-iterated return of the log price process*) and 8 is the number of weeks. We proceed with the weeks beginning from 1 to 8, then 2 to 9, then 3 to 10, and so on, i.e.; by this we employ the sliding-window technique.

In order to implement this procedure to the fitted data, firstly, as in Figures 5.1-5.5, we obtain fitted log-prices according to *Johansen-Ledoit-Sornette* (*JLS*) *model* (47; 71; 70) for the crises of US, Japan and China stock markets respectively. In the figures, t_c demonstrates the estimated bubble-burst time by *JLS* model and t_R stands for the real bubble-burst time.
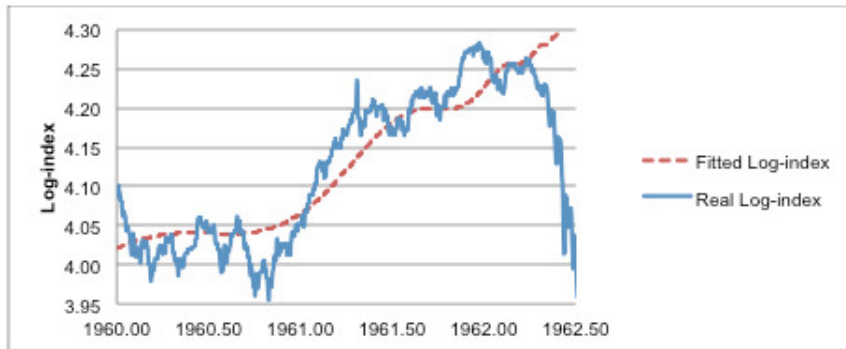
Figure 5.1: Fitted Log-index process for the year 1962 at S&P 500 index.

For Figure 5.1, A=73.48, B=-9.96, C=1.46, t_c=1962.4074, z=0.83, w=13.63, ϕ=1.07 and t_R=1962.4416. $R^2 = 0.8330$.
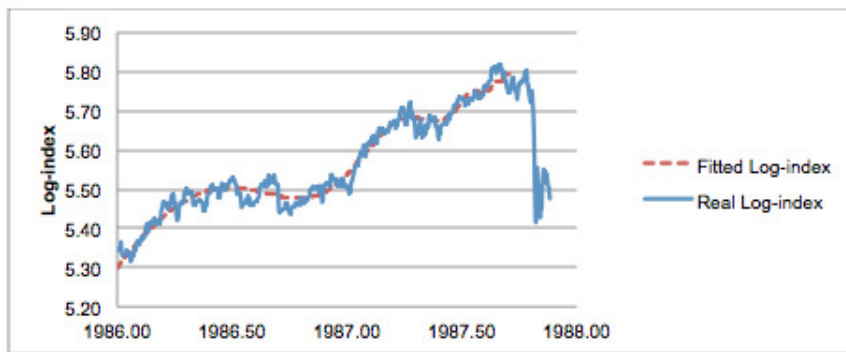
Figure 5.2: Fitted Log-index process for the year 1987 at S&P 500 index.

For Figure 5.2, A=328.63, B=-85.26, C=-17.47, t_c=1987.7055, z=0.82, w=13.64, ϕ=1.40 and t_R=1987.8222. $R^2 = 0.9694$.

44

Figure 5.3: Fitted Log-index process for the year 2008 at S&P 500 index.

For Figure 5.3, A=1265.47, B=-187.44, C=78.18, t_c=2008.7285, z=0.77, w=12.77, ϕ=0.28 and t_R=2008.7888. $R^2 = 0.7993$.
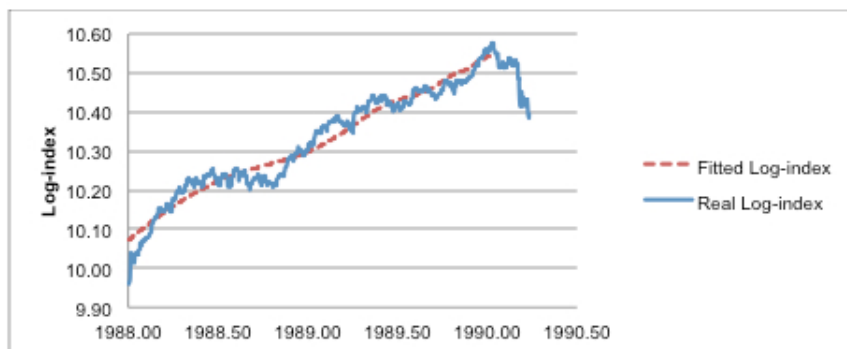
Figure 5.4: Fitted Log-index process for the year 1990 at Nikkei 225 index.

For Figure 5.4, A=38228.82, B=-7907.31, C=-481.65, t_c=1990.0315, z=0.85, w=14.58, ϕ=0.26 and t_R=1990.0416. $R^2 = 0.9580$.

Figure 5.5: Fitted Log-index process for the year 2007 at Shanghai Composite index (SSEC).

For Figure 5.5, A=5374.28, B=-3146.27, C=-429.87, t_c=2007.8105, z=0.82, w=12.85, ϕ=0.28 and t_R=2007.8138. $R^2 = 0.9671$.

Since the *JLS* model generates daily data, these data are transformed into weekly data, and another matrix $\mathcal{B}$ and submatrices $\mathcal{B}_k$ are constructed like the matrices $\mathcal{A}$ and $\mathcal{A}_k$ for the case of the real data. Similar to the real data, again, based on Assumption 1, the length of each window is determined as 8. Theoretically, Corollary 1 guaranteed that $\nabla_u h(u) - U^{-1}T$ in the DRN-Direction procedure, is nonsingular. Although we did not encounter singularity (or ill-conditionedness) of submatrices in the case of real data, we came across this problem for submatrices over the fitted data. We solved that problem via two strategies: (i) singular-value decomposition and regularization, and (ii) a perturmation method, respectively. As we observed from the DRN-Direction procedure, the matrix $\nabla_u h(u) - U^{-1}T$ is calculated from $\sum(u)$ and $\sum(u)$ is obtained using the related matrix $\mathcal{B}_k$. Therefore, we employed these strategies directly to matrices which face an ill-conditionedness. Before stating how we perform those methods, we shortly mention about *singular-value decomposition* method.

5.2   Singular-Value Decomposition

Systems of linear equations sometimes are rank-deficient or ill-conditioned; these properties are examined by the help of *Singular-Value Decomposition* (*SVD*) (4). This methodology can be explained as follows:

Definition 2: Given an $p \times q$ matrix A of rank r, A can be factorized in the subsequent manner:

$$A = U \ S \ V^T,$$

where U and V are $p \times p$ and $q \times q$ orthogonal matrices, respectively, containing the

"singular vectors". Also, $U^T U = I_p$ and $V^T V = I_q$; the columns of U are orthonormal eigenvectors of AA^T and the columns of V are orthonormal eigenvectors of $A^T A$. Furthermore, S is a diagonal matrix including the square roots of the eigenvalues from AA^T or $A^T A$ in decreasing order (5), i.e., the singular values of A placed in the leading diagonal of S are arranged in a decreasing order, $s_1 \geq s_2 \geq \ldots \geq s_{\min\{p,q\}} \geq 0$ (4). Since just the first r singular values are different from zero, S can be written as the following partitioned form, given by block matrices (4; 57):

$$\begin{pmatrix} \begin{matrix} s_1 & & \\ & \ddots & \\ & & s_r \end{matrix} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix} 0 & & \\ & \ddots & \\ & & 0 \end{matrix} \end{pmatrix} \tag{5.1}$$

Eigen-decomposition (*ED*) can also be used in place of SVD. However, ED requires A to be a square matrix, and even if A was a square matrix, it does not have to have an eigen-decompositons in the field of real numbers. Unlike ED, SVD always exists and it can be applied for any matrices (57): large, small, square, rectangular, singular, nonsingular, sparse or dense. Because of these advantages of SVD, we prefer to employ it. In MATLAB, we used *svd* command to calculate singular values of a matrix.

5.2.1 SVD and Regularization

If at the first submatrix of $\mathscr{B}$ (cf. Subsection 5.1), called $\mathscr{B}_1$, we could not find a positive volume, then we separated the regarded 7×8 matrix into 2×2 matrices, exempting the first variable, which is the date. Hence, we got a block-structured matrix whose entire format is 6×8, as follows:

2×2	2×2	2×2	2×2
2×2	2×2	2×2	2×2
2×2	2×2	2×2	2×2

.

We calculate the singular values of each 2×2 matrix above and order these values in a descending manner. By starting from the smallest one which is very close to zero, we obtain new 2×2 matrices according to the following formula:

$$\mathscr{B}_k^{new} = U \, S_{new} \, V^T,$$

where

$$S_{new} = S_{old} + P$$

and

$$P = \begin{pmatrix} 0 & 0 \\ 0 & 0.005 \end{pmatrix}.$$

Since we partitioned the matrix by 2×2 submatrices, the format of P is 2×2. After finding the matrix $\mathcal{B}_k^{new}$, it is substituted in place of the old version of the matrix, $\mathcal{B}_k$. Then, we compute the volume of the ellipsoid which is constituted by this matrix. If we found a positive "sufficiently nonvanishing" volume, then we continue with the next window [2:9]. If we could not calculate the volume, then we pass to the another 2×2 matrix which has the second smallest singular value. Until we have found a positive volume; this process continues. If we could not find any positive volume, even if we tried all 2×2 submatrices, then we apply our *perturbation method*, explained subsequently.

5.3 Perturbation Method

5.3.1 Foundation of Perturbation Method

Again according to descending order of the singular values, i.e., starting our technique from any 2×2 submatrix which has the smallest singular value, we apply a perturbation according to the following formula:

$$\mathcal{B}_k^{new} := \mathcal{B}_k + \epsilon\, I_2. \tag{5.2}$$

Here, I_2 is the 2×2 identity matrix, and, generally, we take $\epsilon = 0.001$. After making various computations and trials, we found that the number 0.001is the most appropriate value since we observed that when we gradually increase the value for the perturbation factor ϵ such as 0.003, 0.005, 0.01, 0.02,..., 0.09, the volume decreased. This phenomenon may cause an artificial global minimum point produced. So, it can give rise to confuse the artificial global minimum points with real global minimum points observed when the bubble-burst time approaches. However, if we could not obtain a positive volume with $\epsilon = 0.001$, we choose a suitable smaller number such as, in the case of Japan stock market, 0.008.

When we compute the first volume by using the submatrix $\mathcal{B}_1$, then we pass to the second submatrix (or window), $\mathcal{B}_2$, whose columns are between 2 and 9. If we found a positive volume, we continue to the next submatrix, $\mathcal{B}_3$, whose columns are between 3 and 10. However, if could not find a positive volume at the matrix 2:9, we again applied

SVD method to the 9th and 10th columns by dividing them into 2×2 submatrices as follows:

2×2	2×2
2×2	2×2
2×2	2×2

.

Again, we order the singular values of each 2×2 submatrix of the columns 9th and 10th, and we start from the 2×2 submatrix which has the smallest singular value; then we get a new matrix according to $\mathcal{B}_k^{new} = U\ S_{new}\ V^T$ and substitute it at the place of the previous one. If we found a positive volume after this regularization, we pass to the next matrix which is between the columns 3 and 10 and so on. After the calculation of the volumes of ellipsoids by sliding-windows has been completed, we observe that when a bubble-burst time approaches, a global minimum point is obtained at the volume-based indices for real and fitted data which are displayed in Figures 5.6-5.10. For the real data, this global minimum point in the volume-based index is observed more clearly than for the fitted data. The fitting can be interpreted as a first kind of regularization in the sense that the data become smooth with reference to the real data. Because of the smoothness of the fitted data, sometimes we cannot observe a sharply decreasing process in volume-based index. By the fitting such a temporary process, which may just be a bubble, could be mildened or even treated as an outlier. But we are explicitly interested in such outlier-kind of phenomena in this thesis, in financial bubbles. So, the reason for the less pronounced signal in case of the fitted data lies in the fact that they are "smoothed" already, with mildened outliers and with thinner ellipsoids in general "spanned" up by those data, over the windows. This leads to smaller ellipsoidal volumes and, hence, to weaker signals of warning. In addition to this, every bubble has its own characteristics. For example, the 1987 October crash was an outstanding and pronounced one, but the crash that occurred in 1962 May was a small one which is named as a *flash* crash. Therefore, especially, at the fitted data, sometimes, it is not easy to observe that global minimum point.

Let us recall that we offered a 2-stage approach in overcoming illposedness of the inverse problem of our study: (*i*) To see whether the use of SVD works well, if so: to conduct it, and if it is not the case, (*ii*) to apply the perturbation technique which we introduced. In fact, that we employed submatrices of type 2×2, in order to detect and, if needed, overcome the illposedness of the entire design matrix of the present (sliding) window, might seems surprising at the first glance. But for the following reason, the 2×2 format turns out to be a very reasonable choice: (*a*) When we go from one to the next sliding window, we take away one column while adding a new column which seems "closest" and most related to its previous column (the last one over the previous window of time). This possible linear dependence of 2 vectors could be looked at and become rank-updated. (*b*) If we looked at larger submatrices with 3 or more rows and columns, the entire data and, hence, the solution of our model and decision support system, could become changed too much. In fact, our data are informations of financial and probabilistic kind, exposing high random fluctuation often. To make bigger changes which actually would lead to a smoothening, could corrupt both our problem

and its result. By this we underline that our study of financial bubbles is interested in "outlier" phenomena in the financial sector, so that all approximations and changes need to be done with great care, such to say, with a very fine "needle", such as the 2×2-kind of correction matrices are.

Any changes in the conditioning of the design matrix, i.e., on overcoming illposedness, work via improvements, in fact, increases of (*i*) singular values or (*ii*) eigenvalues, unless truncation methods were applied. Those changes can be conducted explicitly on those values, or implicitly through matrices whose roots of characteristic polynomials they are. Now, any addition of a small increment Δs_i or $\Delta \lambda_i$ to a singular value s_i or an eigenvalue λ_i changes the *condition number* of the design matrix, i.e., the differences between smallest and largest of those numbers are diminished, respectively. Since the product of the condition number and the relative error in any (response) data vector majorizes the relative error in the corresponding model vector of our solution, a diminished condition number leads to a smaller upper bound of the relative error of the inverse problem. For closer information on the aforementioned *sensitivity analysis*, we refer to the works (4; 58; 67) from the theory of inverse problems and to the works (34) from the numerical analysis of linear algebra.

Since bubbles are rare events and sensitivity analysis necessitate too many data, for the rare event analysis, sensitivity analysis may not be appropriate. Nevertheless, if the sensitivity analysis is considered the following issues may be studied.

The sensitivity analysis may play a role for a refined control of singular and eigenvalues and, by this, of shape and volume of our ellipsoids. This will also include an optimized choice of the dimensions of the subintervals which we address when regularizing our inverse problem over time windows, whenever needed. In this respect, sensitivity analysis can become another model-based module in our entire approach towards an early warning about financial bubbles and their near-coming bursting time. Herewith, we are on the way towards a mathematics-based risk management and ellipsoidal uncertainty quantification, where approximative heuristics are gradually replaced by model-based techniques.

As mathematical methods of sensitivity analysis, we mention (i) Tikhonov regularization, which bases on a strong use of SVD, and (ii) Projection or iterative methods in cases where the design matrix of inverse problem is sparse (4). If the problem consists of a differential equation of any kind, a system of such equations, (iii) refined perturbation methods are sometimes used to solve them. Here, the perturbation can lead to a reduction of the problem: "complexity"; eventually, when withdrawing the perturbation gradually, the approximate solution converges to the real solution. We mention that also the celebrated "interior point methods" from conic optimization belong to the perturbation methods (10; 61; 77).

From the perspective of statistics and sensitivity measures, we add the following methods: (iv) *partial derivatives*, *importance index*, *variation of inputs by one standard deviation* and *by* 20%, *relative deviation of output distribution*, *partial rank correlation coefficients*, *relative deviation ratio*, *standardized coefficient of regression*, *rank*

50

regression coefficients, Gramer-von Mises test, Smirnov test, Mann-Whitney test and *squared-ranks test* (36).

(i)   (ii)
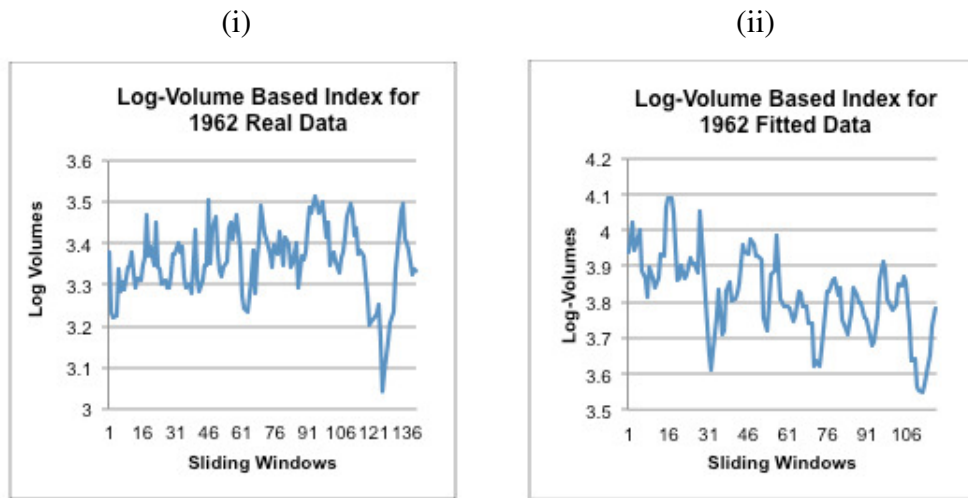
Figure 5.6: Volume-based index for real and fitted data of 1962 at S&P 500 index.

(i)   (ii)
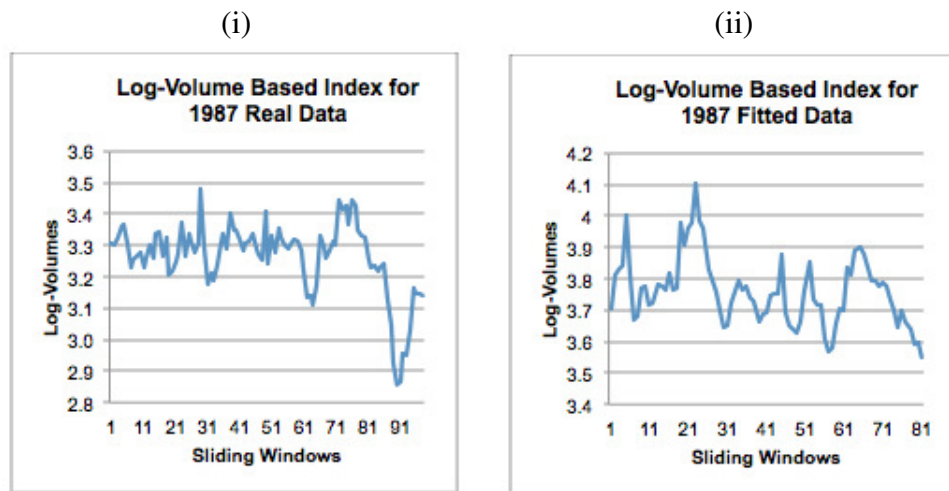
Figure 5.7: Volume-based index for real and fitted data of 1987 at S&P 500 index.

(i)   (ii)
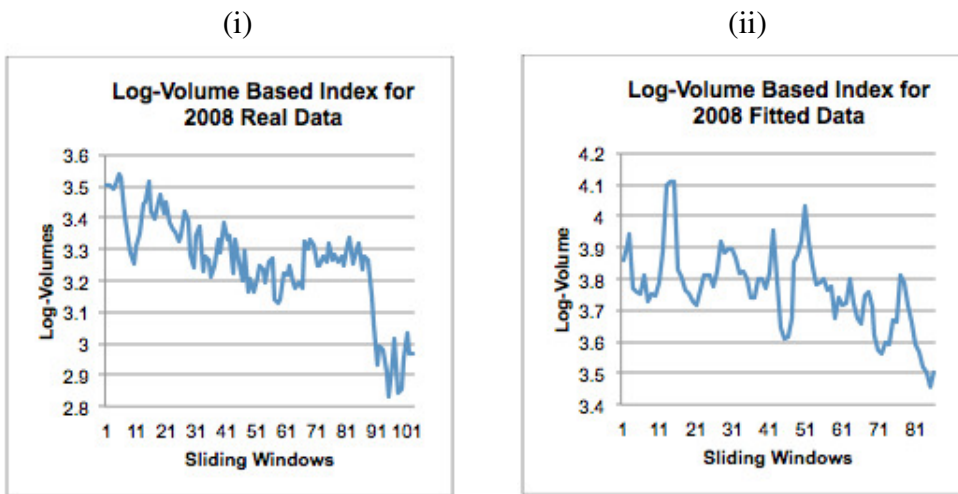
Figure 5.8: Volume-based index for real and fitted data of 2008 at S&P 500 index.

(i)   (ii)
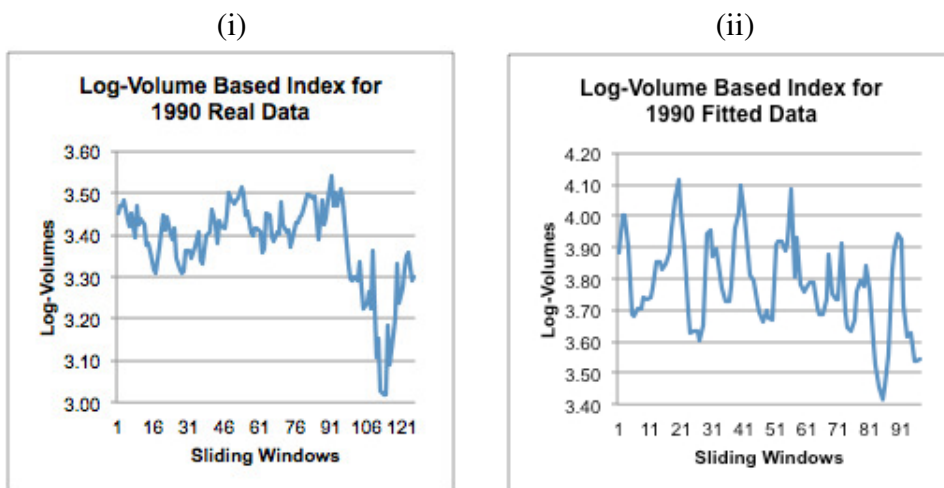
Figure 5.9: Volume-based index for real and fitted data of 1990 at Nikkei 225 index.

Although, we observe the regular decreasing process and occurrence of a global minimum point when the bubble-burst time approaches for S&P 500 and Nikkei 225 indices; However, for the Shanghai Composite index, we saw that there are more than one global minimum points and no regular diminishing process as shown in Figure 5.10.
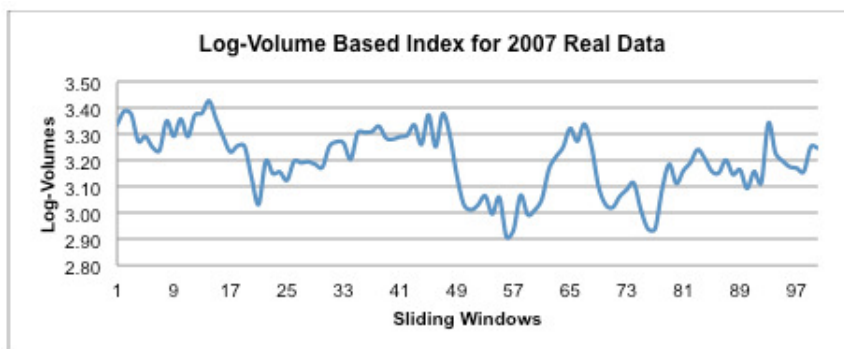
Figure 5.10: Volume-based index for real data of 2007 at Shanghai Composite index (SSEC).

The reason of this case is the volatility (whose measurement is a standard deviation) of US and Japan markets takes values within the definite intervals, but in Chinese stock market the range of the volatility varies significantly, as the following Figures 5.11-5.15 show.
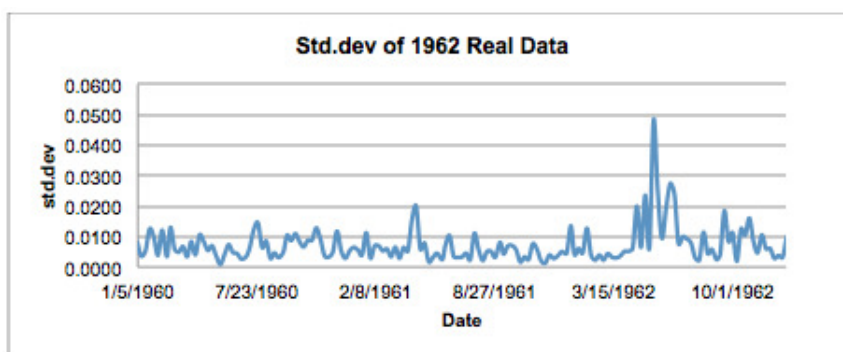
Figure 5.11: Volatility of US stock market before 1962 crises.
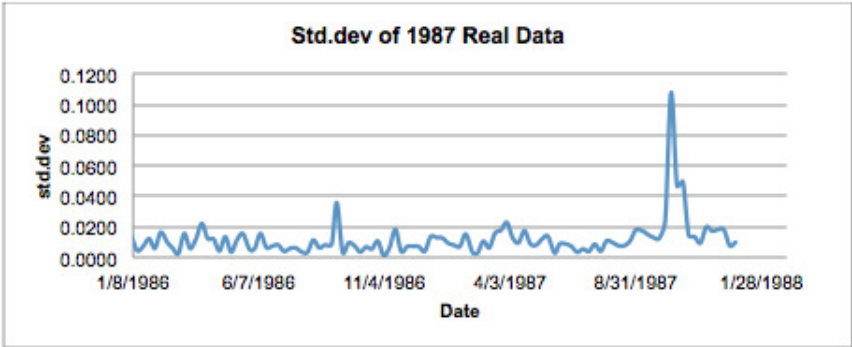
Figure 5.12: Volatility of US stock market before 1987 crises.
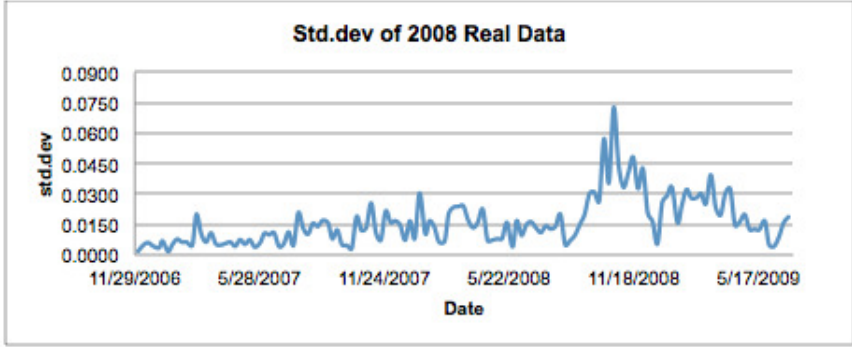
Figure 5.13: Volatility of US stock market before 2008 crises.
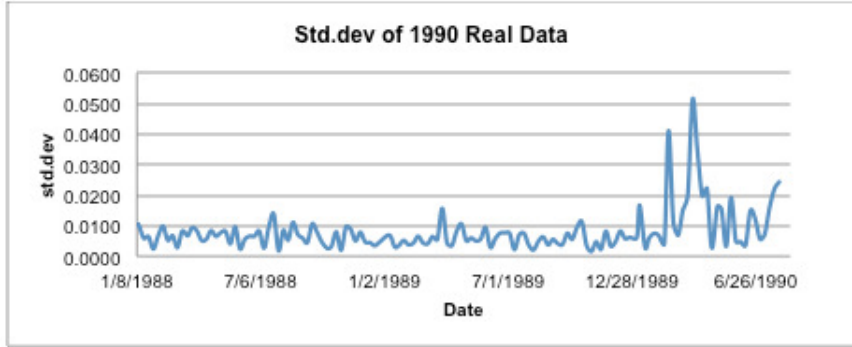
Figure 5.14: Volatility of Japan stock market before 1990 crises.
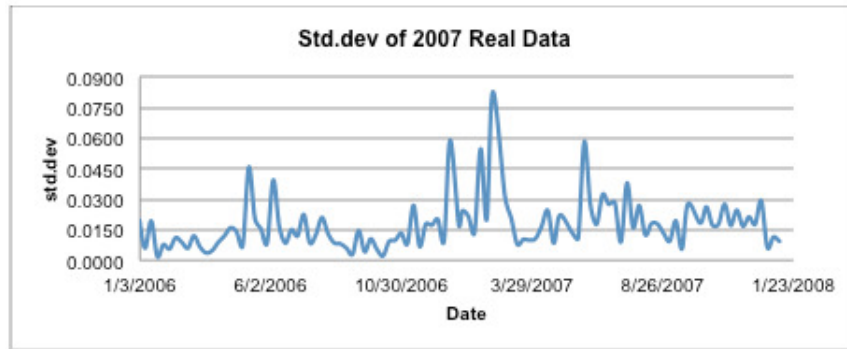
Figure 5.15: Volatility of China stock market before 2007 crises.

When the volume index drops and we decide to give an early warning, we particular have in mind the behavior of the logarithmic asset price and its incremental changes. It is the idea of a steep "drift" which dominates the diffusion part and might even resemble an impulse or a jump, with some time consumption, however. But the reason of the diminishment of the ellipsoid's volume could also come from a diminishment of some other of our variables. Not surprisingly, as explained before, the standard deviation could tend to 0, but there might also be other variables "collapsing". Since the purpose of this thesis consists in an early warning in the sense of *not coming too late*, we leave the possibility of other variables' approach to 0, and the decision make and his/her team could look into their values as soon as possible, but in any way, financial measurements of control of the price process should be taken in time. In future work, we want to include a closer analysis and representation of the other processes rather than of the logarithmic price into the early warning signalling.

In order to visualize the ellipsoids, as we mentioned previously, we benefit from Radon transform. Our study bases on the evaluation of ellipsoids by a Radon transform applied on the given data over the regarded interval, which actually come from measurements that are now, rather than by Cartesian coordinates, represented by polar coordinates. We recall that the data, or measurements, are comprised in the matrix $\mathcal{A}$ they enters the definition of the ellipsoid $E_{Q,\mu}$ (cf. Section 4.1) through the matrix Q, which is the inverse of a (empirical) variance-covariance matrix. This inverse matrix is characterized by the reciprocal values of the eigenvalues of the variance-covariance matrix. Herewith, the Radon transform represents information about these values. In particular, a smallest eigenvalue which could be close to 0 contributes to an "impulse" by the Radon transform through its very high reciprocal value, and such an almost vanishing eigenvalue also means a very small semiaxis length of the ellipsoid, i.e., a degeneracy of $E_{Q,\mu}$ to a lower dimension almost. Then, the volume of $E_{Q,\mu}$ comes close to 0, indeed.

We observe that Radon Transform generates different colors according to up or down movements of the volume-based index. We classified these colors as, yellow, orange, red, grey and black as shown in Figures 5.16-5.20. We notice that, generally, when the direction of the volume-based index goes up, the color of the RT becomes more

brilliant and when its direction turns downwards, RT responds with more dark color. So, the RT yields a signal when the volume index increases or decreases. Volumes of ellipsoids are obtained from the log-price and 5 other components mentioned by us, which are derived from the log-price, and RT results are acquired from these ellipsoids. So, RT can be related with a logarithmic price process.

For the bubble, when we see the bright colors, i.e., yellow, orange or red, we should hesitate to detect that the bubble-burst time could approach. Actually, that signal could be wrong which may then convey a *fake bubble*. The signal of RT becomes meaningful when it is combined with the volume-based index. Together with a decreasing process of the volume-based index and with the color of RT being very brilliant, i.e., yellow, orange or red, we say that the burst time of a bubble could be approached. The following Figures 5.16-5.20 are some examples of RT colors for the years 1962, 1987, 2008 crisis in the US and the 1990 crisis in the Japan stock markets. That is to say, we classified the colors which look like in Figure 5.16 as *orange*, or in Figure 5.17 as *red*, and so on.
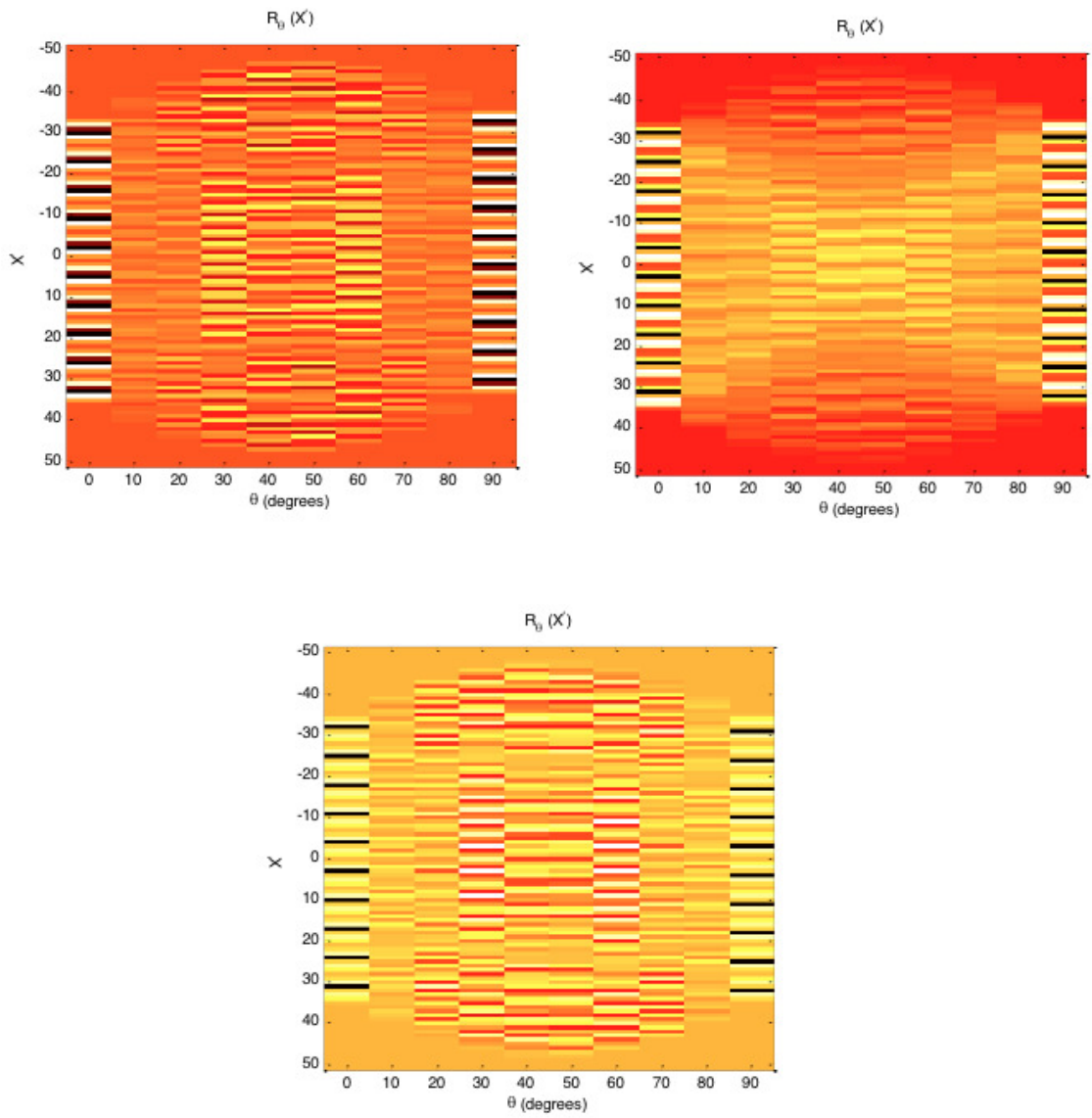
Figure 5.16: Examples of Radon signals for the color Yellow.
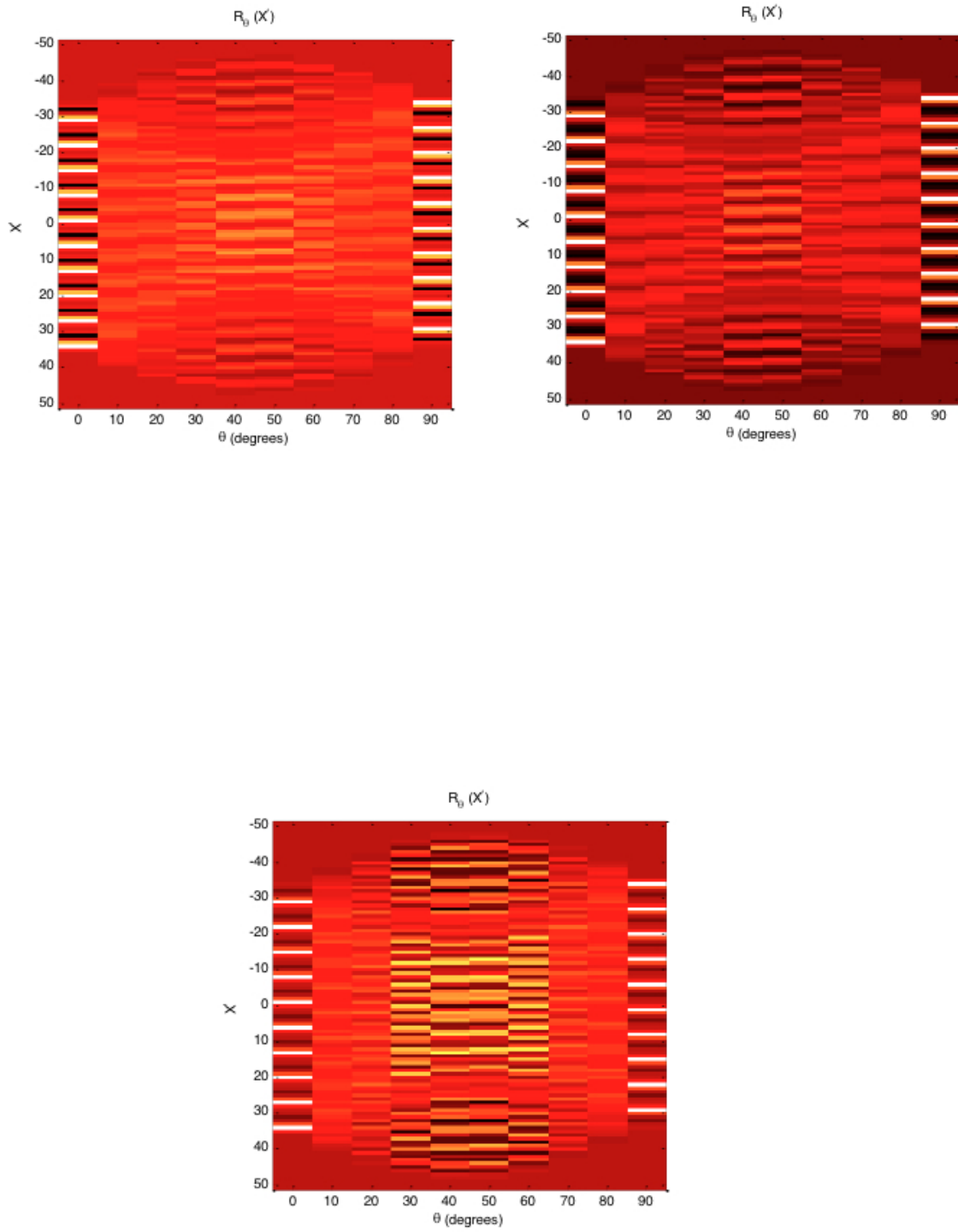
Figure 5.17: Examples of Radon signals for the color Orange.
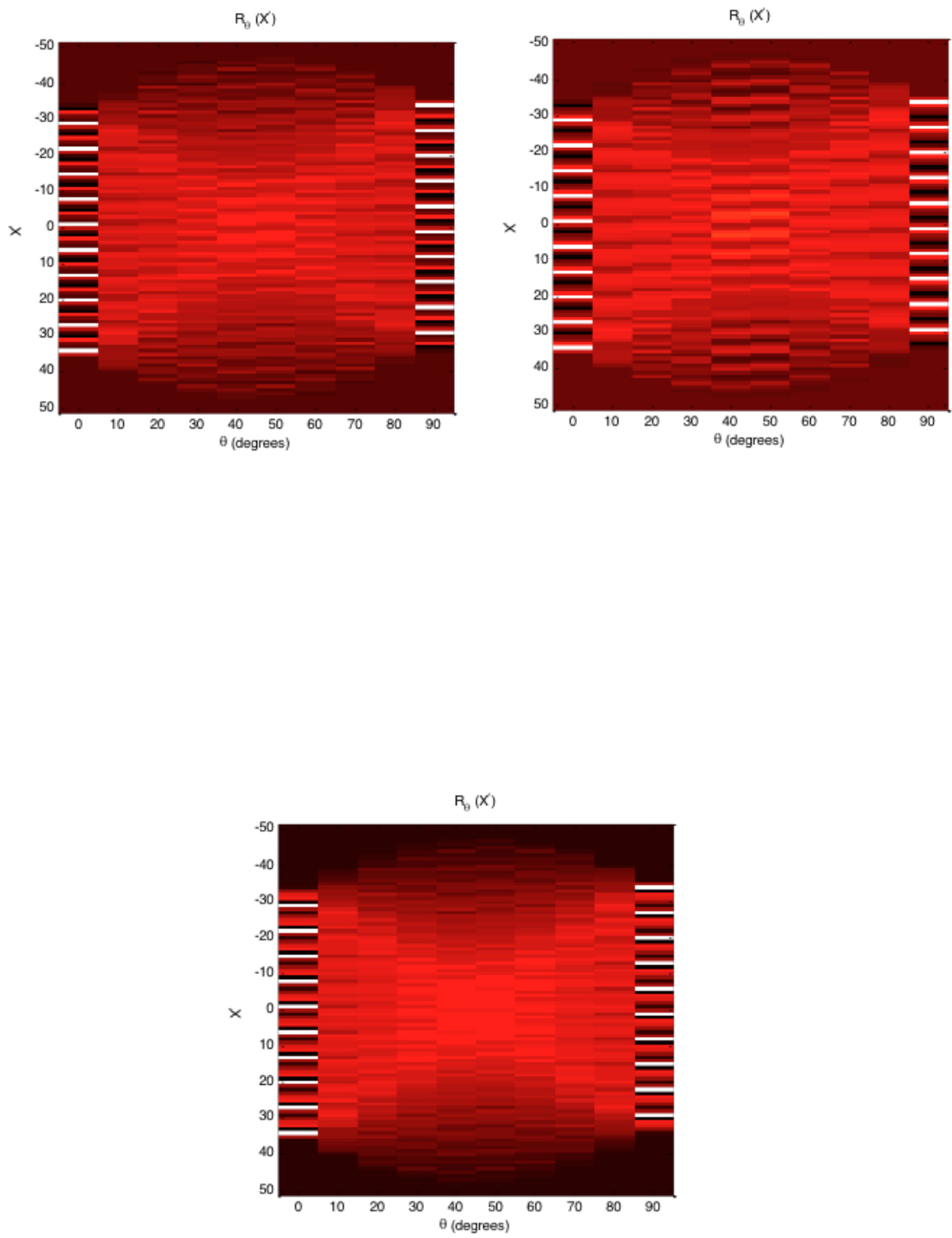
Figure 5.18: Examples of Radon signals for the color Red.
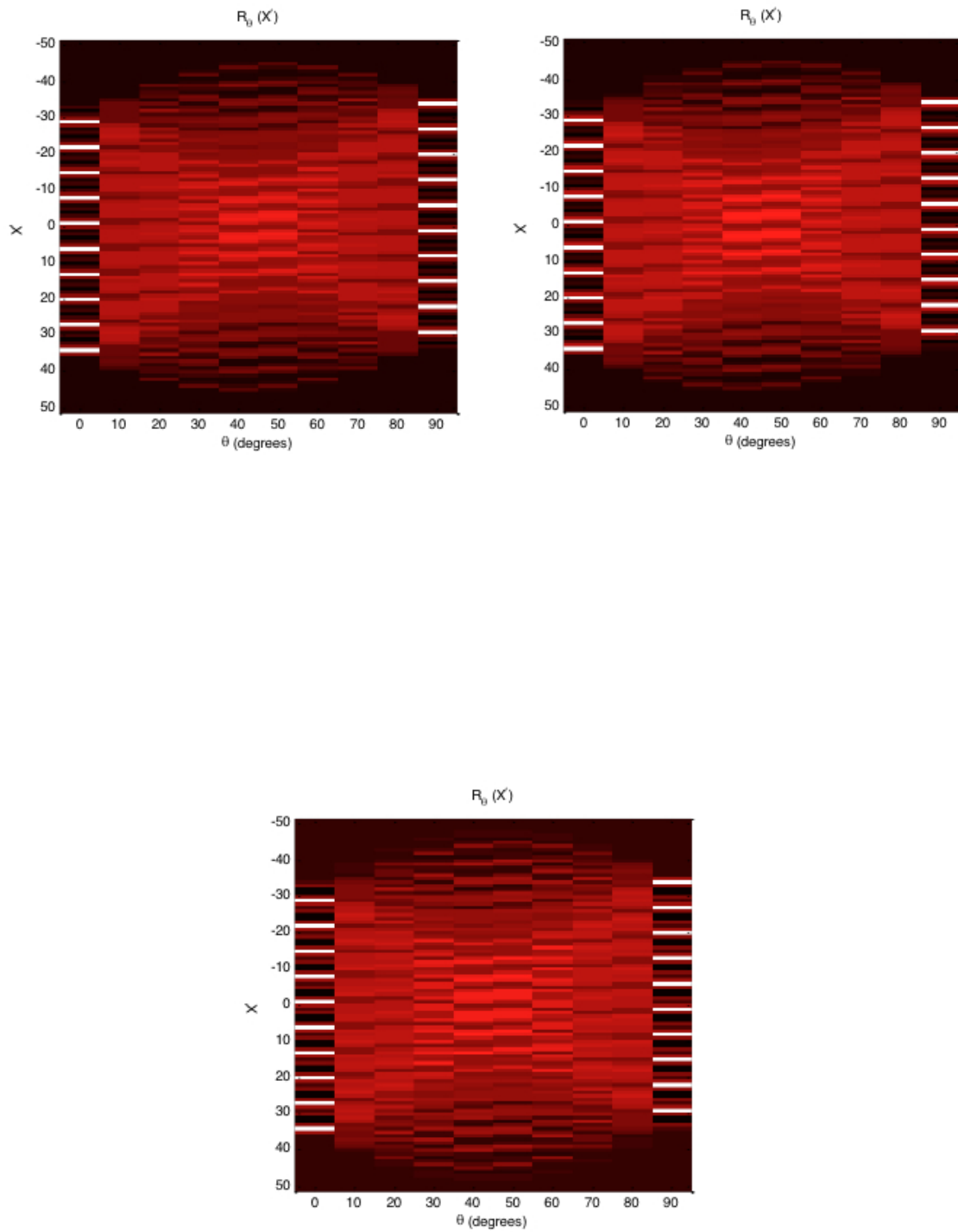
Figure 5.19: Examples of Radon signals for the color Gray.
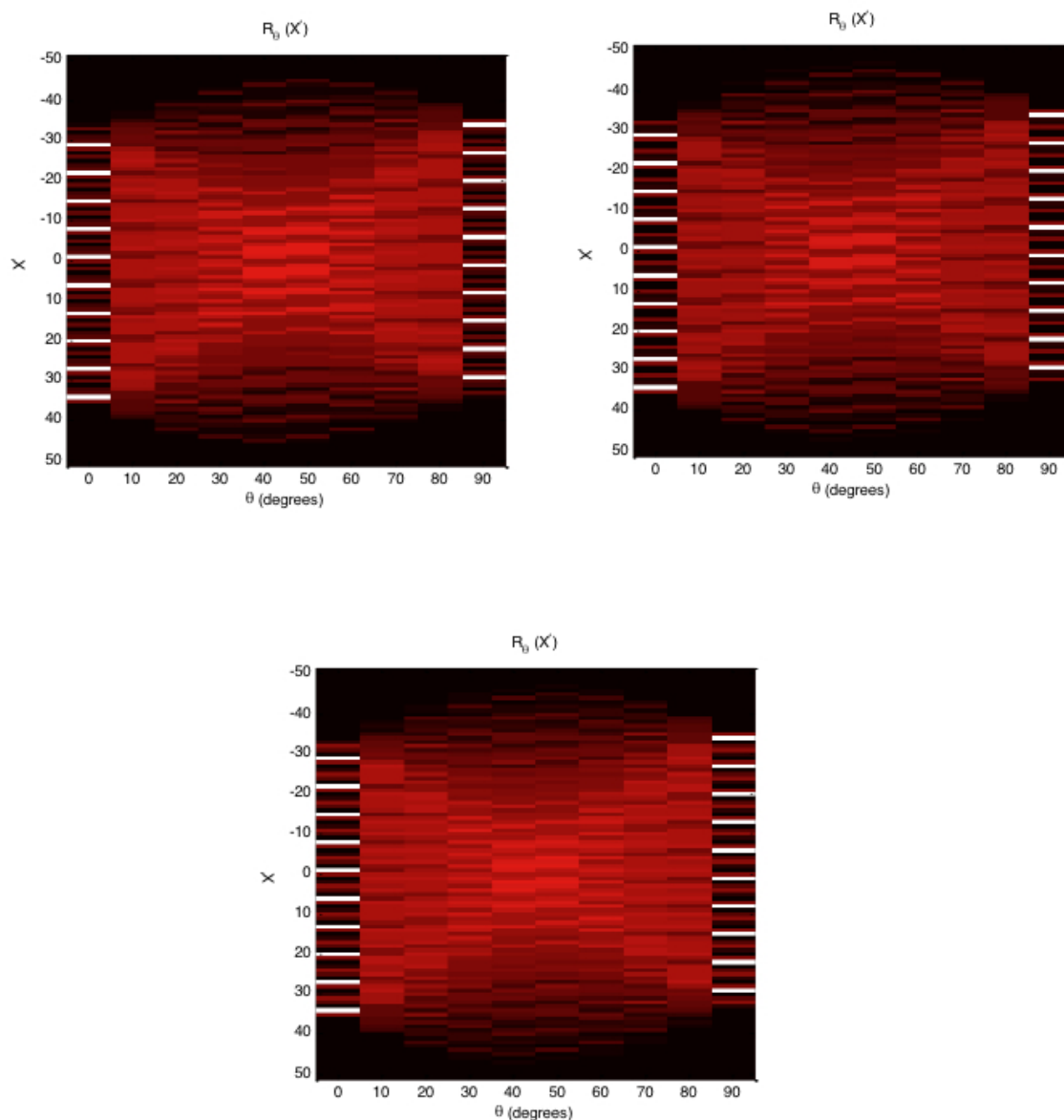
Figure 5.20: Examples of Radon signals for the color Black.

Figures 5.21 and 5.22 summarize the RT colors and corresponding sliding windows for the real and the fitted data of the years 1962, 1987, 2008 and 1990. When a local maximum point occurred and then a regular decreasing process commences, we measure a decreasing rate and control whether this decline in the volume-based index is temporary or permanent. If it is permanent, that maximum point is a starting point for the Figures 5.21 and 5.22. For instance, this local maximum point took place in 110th sliding window for 1962 crisis in the real data; therefore, the table which belongs to 1962 began from the 110th sliding window and ended with the first sliding window which does not include the bubble-burst time. This procedure is the same for the fitted data.

1962 Real Data	
SW	RT
110.	Black
111.	Orange
112.	Gray
113.	Gray
114.	Black
115.	Gray
116.	Gray
117.	Gray
118.	Gray
119.	Orange
120.	Red
121.	Red
122.	Red
123.	Yellow
124.	Gray
125.	Red
126.	Black

1987 Real Data	
SW	RT
76.	Red
77.	Red
78.	Gray
79.	Red
80.	Red
81.	Gray
82.	Red
83.	Orange
84.	Orange
85.	Yellow
86.	Red
87.	Red
88.	Red
89.	Gray
90.	Red
91.	Red
92.	Red
93.	Red
94.	Gray
95.	Red

2008 Real Data	
SW	RT
82.	Orange
83.	Red
84.	Red
85.	Red
86.	Orange
87.	Red
88.	Red
89.	Yellow
90.	Yellow
91.	Orange
92.	Yellow
93.	Orange
94.	Yellow
95.	Black
96.	Black
97.	Black

1990 Real Data	
SW	RT
95.	Gray
96.	Gray
97.	Gray
98.	Gray
99.	Black
100.	Black
101.	Black
102.	Black
103.	Black
104.	Gray
105.	Gray
106.	Orange
107.	Orange
108.	Orange
109.	Red
110.	Gray
111.	Red
112.	Gray
113.	Gray

Figure 5.21: Colors of RT and corresponding sliding windows (SW), related to the real data, for the crisis in the US and the Japan stock markets.

The 1962 US stock market bubble bursted at the last week of May and this date corresponds to the 126th week of our data. Since the length of each sliding window is 8, the first sliding window which contains the 126th week is the 119th. As we see from Figure 5.21, the corresponding RT color for the 119th window is *orange* and until we reach the bubble-burst time, we observe the colors *red* and *yellow* that are the classified as strong alarms for the bubble-burst time. The 1987 US stock market bubble bursted in the middle of October, in the 95th week of our data. The first sliding window which includes the 95th week is the 88th. As we observe from Figure 5.21, the corresponding RT color for the 88th window is *red* and until we reach bubble-burst time, we observe the color *red* again and again. For the 2008 crisis, the burst of the bubble is realized at the beginning of October which is the 97th week of the data, and the initial window encompasses that week is the 90th one. If we look at Figure 5.21, the corresponding RT color for that window is monitored as *yellow* and also *orange*. For the cases of 1987 and 2008, we see strong warnings before the 88th and 90th windows by the colors *orange* and *yellow*. The reason for this lies in the characteristics of those bubbles. As we mentioned previously, while the 1962 bubble was a flash bubble, the 1987 was outstanding one, also the 2008 bubble. As for the 1990 Japan stock market bubble, it bursted at the beginning of January which matches with the 103rd week in our data, and the window which for the first time contains this week is the 96th window. Our approach almost catches this bubble; however, it gives the signal with four weeks delay.

We also compute the decline rates with respect to the local maximum points in volume-based indices; the rate can be thought as first-order or finite difference information.

For the year 1962, we observe that the highest decline occurs as - 12.956% when the bubble-burst time approaches. For 1987, the maximum decline is -16.792%. As for 2008, the highest decrease rate is -10.330%. Also, for the 1990 Japanese stock market bubble, we observed an decrease in volume-based index by -15.956%. Thus, we can conclude that if this decline exceeds 10%, the decision maker should be careful for the bubble-burst time could get very close. In other words, 10% can be thought as a decline threshold for the volume-index of the real data.

1962 Fitted Data		1987 Fitted Data		2008 Fitted Data		1990 Fitted Data	
SW	RT	SW	RT	SW	RT	SW	RT
97.	Gray	65.	Gray	77.	Black	72.	Gray
98.	Gray	66.	Black	78.	Black	73.	Black
99.	Black	67.	Black	79.	Black	74.	Black
100.	Gray	68.	Black	80.	Black	75.	Gray
101.	Gray	69.	Black	81.	Black	76.	Orange
102.	Red	70.	Black	82.	Black	77.	Orange
103.	Red	71.	Black	83.	Gray	78.	Orange
104.	Red	72.	Black	84.	Black	79.	Red
105.	Red	73.	Black	85.	Gray	80.	Yellow
106.	Red	74.	Gray	86.	Red	81.	Black
107.	Gray	75.	Black			82.	Black
108.	Gray	76.	Red			83.	Gray
109.	Gray	77.	Red			84.	Black
110.	Gray	78.	Red			85.	Gray
111.	Gray	79.	Red				
112.	Gray	80.	Red				
113.	Yellow	81.	Gray				
114.	Yellow						
115.	Yellow						
116.	Red						
117.	Red						

Figure 5.22: Colors of RT and corresponding sliding windows (SW), related to the fitted data, for the crisis in the US and the Japan stock markets.

At Figure 5.22, because of the term $(t_c - t)$ in JLS model, we cannot proceed beyond the week of bubble-burst time. Therefore, the last line is the first sliding window which includes the week of bubble bursted. For instance, for the year 1962, the bubble bursts at the 124th week and the first sliding window that contains this week is 117th window; however, we can not observe the RT color (also volume of an ellipsoid) for the 118th window. This case is the same for the fitted data which belong to the years 1987, 2008 and 1990. Nevertheless, we monitor strong signals with the colors *yellow*, *orange* and *red* before the burst time of the bubble, except of the year 2008. We recall here a main reason is given by the smoothness of the fitted data which lead to small volumes and small changes of the colors in time. The highest decline rates in volume-based index for the years 1962, 1987, 2008 and 1990, related to the fitted data, are, -9.305% , -8.952% , -9.359% and -12.755%, respectively.

We would like to emphasize that our entire approach means an approximation and does not promise an exact, error-free prediction. Actually, we are not providing a detection in the sense of a deterministic "machine model", "algorithm", or even "machine" or

"apparatus"; otherwise, the financial sector was not in the class of NP-hard decision problems where it, however, can be found. For more information of machines of different complexity classes we refer to (31). We also do not send our signal or information at a bubble-burst time which, in fact, is unknown, but we give a warning at a moment when there is a strong indication that a bubble-burst time is close to appear. Therefore, we call it an *early warning*, and since we do not think in terms of a (deterministic) machine or apparatus, we prefer not to add the word "system" to "early warning", but we add "signalling". This just means a warning sent to decision makers from economy, finance or politics. As this warning is given before a possible burst time of a bubble, those responsible managers are asked to revisit their variables and parameters and to take precaution in order to control asset prices or indices into more "healthy", more reasonable value intervals, to prevent from heavy collapses of prices, indices, companies and economies, respectively.

CHAPTER 6

FUTURE RESEARCH STUDIES

As future research challenges related to this thesis, two main areas are introduced and also offered to the scientific community. They are related with an intensified use of *ellipsoidal calculus*, and with *optimizing the "area under the ROC curve"* as another alternative for a signalling of an early-warning, respectively.

6.1 Ellipsoidal Calculus

In this thesis, the variables which are employed to develop an early-warning signalling, were produced from log-prices/log-index processes. So, the external effects, such as economic indicators, have not yet been considered. In order to deal with the interaction between burst time of a bubble and economic variables and also to understand the embedded behavior, the so-called *system dynamics* of the bubble process, discrete-time *regulatory target-environment systems* (*TE-regulatory systems*) in the presence of ellipsoidal uncertainty can be used. In other words, a specific description of the system's multivariate uncertain states are decided on and identified with ellipsoidal calculus (53). By this approach, it is also possible to explain one country's effect on another country related to financial crises.

The *TE-regulatory systems* consist of two different groups of data. The first one includes the *targets*; which are the most crucial variables in the system and they hinge on the second group which is named as the *environmental items*. Despite those groups of data are strongly related, they reflect entirely different behaviors. Through *clustering* and *classification* methods, the structure of the data can be better perceived and these methods allow to describe groups of data whose elements jointly act on further clusters of target and environmental elements. The uncertain states of those clusters become identified by ellipsoids; and ellipsoidal calculus is performed to assess the dynamics of the TE-regulatory system; i.e., the states of both target and environmental variables of the TE-model are presented via ellipsoids (53). Therefore, firstly, we will shortly repeat the mathematical definition of an ellipsoid and its volume.

An *ellipsoid* in $\mathbb{R}^q$ is defined with its center $\mu \in \mathbb{R}^q$ and a symmetric nonnegative definite *configuration matrix* $\Sigma \in \mathbb{R}^{q \times q}$ as

$$\mathcal{E}(\mu, \Sigma) = \{\Sigma^{1/2}u + \mu \mid u \in \mathbb{R}^q, \|u\|_2 \leq 1\},$$

where $\Sigma^{1/2}$ is a matrix square root which fulfills $\Sigma^{1/2}(\Sigma^{1/2})^T = \Sigma$. If Σ reveals a full rank, the nondegenerate ellipsoid $\mathcal{E}(\mu, \Sigma)$ is described with the following way:

$$\mathcal{E}(\mu, \Sigma) = \{x \in \mathbb{R}^q \mid (x - \mu)^T \Sigma^{-1}(x - \mu) \leq 1\}.$$

The eigenvectors of the matrix Σ indicate directions of the principal semiaxes of $\mathcal{E}$. The semiaxes lengths of the ellipsoid $\mathcal{E}(\mu, \Sigma)$ are calculated with the eigenvalues of Σ, $\sqrt{\lambda_i}$, where λ_i $(i = 1, 2, \ldots, q)$. The volume of $\mathcal{E}(\mu, \Sigma)$ can be represented by $\operatorname{vol}\mathcal{E}(\mu, \Sigma) = V_q\sqrt{\det(\Sigma)}$, where V_q is the volume of the unit ball in $\mathbb{R}^q$, i.e.,

$$V_q = \begin{cases} \dfrac{\pi^{q/2}}{(q/2)!} & \text{, for even } q, \\[3mm] \dfrac{2^q \pi^{(q-1)/2}((q-1)/2)!}{q!} & \text{, for odd } q. \end{cases}$$

A discrete-time version of TE-regulatory systems in presence of ellipsoidal uncertainty is predicated on clustering the sets of targets or environmental items, respectively, which means the unification of variables mostly exerting effect on other groups of system variables. The uncertain states of those clusters are defined by ellipsoids; these provide a more detailed definition of uncertainty which represents the correlation of data items. Clusters of unknown parameters, all of them directly depending on the structure of the set of system variables, decide on the dynamical behavior of the clusters and their interactions (53).

In the thesis, we particularly have one ellipsoid over each time interval k and we follow its dynamics via a transition to interval $k + 1$, then $k + 2$, etc.. Here, the ellipsoids are preprocessed by the minimum-volume clustering approach, based on the data that relate to the regarded intervals. With the *ellipsoidal calculus* and *model*, we can identify that transition process, and can go beyond the thesis' main framework.

6.1.1 The Discrete-Time Model

In a discrete-time TE-regulatory system, the number of target factors and environmental variables are taken into account as n and m, respectively. Here, functionally related groups of targets or environmental items are assumed to be described in a preprocessing stage of clustering and data analysis. Especially, the set of targets can be partitioned into R disjoint or overlapping clusters $C_r \subset \{1, 2, \ldots, n\}$ $(r = 1, 2, \ldots, R)$. Likewise, the set of all environmental items can be grouped into S (disjoint or overlapping) clusters $D_s \subset \{1, 2, \ldots, m\}$ $(s = 1, 2, \ldots, S)$. In a condition of disjointness of the clusters the properties $C_{r_1} \cap C_{r_2} = \emptyset$ for all $r_1 \neq r_2$, and $D_{s_1} \cap D_{s_2} = \emptyset$ for all $s_1 \neq s_2$, are satisfied. Furthermore, since every cluster matches to a functionally related group of data elements, the uncertain states of these clusters are described by ellipsoids (53).

Ellipsoids may be designated with intervals if clusters are singletons, i.e., 1-dimensional. Additionally, flat ellipsoids $\mathcal{E}(\mu_r, \Sigma_r) \subset \mathbb{R}^{|C_r|}$ and $\mathcal{E}(\rho_s, \Pi_s) \subset \mathbb{R}^{|D_s|}$ would refer to data sets where at least one of the variables is definitely known; however, we can avoid this by an artificial extension in the corresponding coordinate directions of length $\varepsilon > 0$. That is to say, one can impose lower bounds on the semiaxes lengths. Likewise, the extension can be controlled by imposing sufficiently large upper bounds and, hence, one can keep away from "needles" (cf. Chapter 1 and 4) or even degenerate (lower-dimensional) ellipsoids.

The *target items* in *a regulatory system* are described by

  (i) the interactions among the clusters of target items
(defined by an $n \times n$-*interaction matrix* A^{TT} and a vector of *n-intercept vector* V^{TT}),

 (ii) the influences of the clusters of environmental items onto the target clusters
(defined by an $n \times m$ *interaction-matrix* A^{TE} and an *n-intercept vector* V^{TE})

The unknown parameters of the regulatory system are composed by the entries of the interaction matrices A^{TT}, A^{TE} and the intercept vectors V^{TT}, V^{TE}. Clusters of parameters, represented by particular submatrices and subvectors of A^{TT}, A^{TE} and V^{TT}, V^{TE}, describe affine-linear coupling rules. To define the interactions among the clusters of target items, a submatrix $\Gamma_{jr}^{TT} \in \mathbb{R}^{|C_j| \times |C_r|}$ of A^{TT} is designated to each pair C_j and C_r (the components of C_j and C_r decide about the indices of rows and columns). This submatrix can in turn be taken into account as a *connectivity matrix* among the clusters C_j and C_r that corresponds to the (uncertain) degree of connectivity among the elements of the two clusters of targets. Later on, an additional shift (intercept) will be inserted by the subvector $\Phi_j^{TT} \in \mathbb{R}^{|C_j|}$ of V^{TT}. In case of overlapping clusters, the submatrices Γ_{jr}^{TT} and subvectors Φ_j^{TT} will be partially comprised of the same elements (53).

The impact of the clusters of environmental items on the target clusters can be explained, similarly.

For every pair of target clusters C_j and environmental clusters D_s, a submatrix $\Gamma_{js}^{TE} \in \mathbb{R}^{|C_j| \times |D_s|}$ (the elements of C_j and D_s determine the indices of rows and columns) and a subvector $\Phi_j^{TE} \in \mathbb{R}^{|C_j|}$ of V^{TE} are determined. The submatrix Γ_{js}^{TE} behaves as a *connectivity matrix* among the clusters C_j and D_s and Φ_j^{TE} acts as a shift (53).

In addition to the regulatory system of target variables, we look at an *environmental regulatory system* which can be identified by

 (iii) the interactions among the clusters of environmental items
(defined via an $m \times m$ interaction-matrix A^{EE} and an m-intercept vector V^{EE}),

 (iv) the influences of the target clusters on the environmental clusters
(defined via an $m \times n$ interaction-matrix A^{ET} and an m-intercept vector V^{ET}).

The degree of connectivity between the pairs of environmental clusters D_i and D_s or a pair of environmental and target clusters, D_i and C_r, are represented through the

submatrices $\Gamma_{is}^{EE} \in \mathbb{R}^{|D_i| \times |D_s|}$ of A^{EE} and $\Gamma_{ir}^{ET} \in \mathbb{R}^{|D_i| \times |C_r|}$ of A^{ET}, and also the subvectors $\Phi_i^{EE} \in \mathbb{R}^{|D_i|}$ of V^{EE} and $\Phi_i^{ET} \in \mathbb{R}^{|D_i|}$ of V^{ET}.

The discrete-time model, which allows to compute the estimations $\hat{X}_r^{(k)}$ and $\hat{E}_s^{(k)}$ of the ellipsoidal states of target and environmental variables, is presented by the following pseudo model.

TE Model (53):

For $k = 0, 1, 2, \ldots$; and for $j = 1, 2, \ldots, R$:

Step 1. Interactions among the clusters of targets

(A) Effect of cluster C_r onto cluster C_j:

$$G_{jr}^{(k)} = \Gamma_{jr}^{TT} \cdot X_r^{(k)} + \Phi_j^{TT} \ (r = 1, 2, \ldots, R).$$

(B) Cumulative effect of all clusters of targets onto cluster C_j:

$$G_j^{(k)} = \bigoplus_{r=1}^{R} G_{jr}^{(k)}.$$

Step 2. Effects of the environmental clusters onto the clusters of targets

(A) Effect of environmental cluster D_s onto target cluster C_j:

$$H_{js}^{(k)} = \Gamma_{js}^{TE} \cdot E_s^{(k)} + \Phi_j^{TE} \ (s = 1, 2, \ldots, S).$$

(B) Cumulative effect of all environmental clusters onto cluster C_j:

$$H_j^{(k)} = \bigoplus_{s=1}^{S} H_{js}^{(k)}.$$

Step 3. Sum of effects onto the target clusters

$$X_j^{(k+1)} = G_j^{(k)} \oplus H_j^{(k)}.$$

For $i = 1, 2, \ldots, S$:

Step 1. Interactions among the clusters of environmental items

(A) Effect of cluster D_s onto cluster D_i:

$$M_{is}^{(k)} = \Gamma_{is}^{EE} \cdot E_s^{(k)} + \Phi_i^{EE} \ (s = 1, 2, \ldots, S).$$

(B) Cumulative effect of all environmental clusters onto cluster D_i:

$$M_i^{(k)} = \bigoplus_{s=1}^{S} M_{is}^{(k)}.$$

Step 2. Effects of the target clusters onto the clusters of environmental items

(A) Effect of target cluster C_r on environmental cluster D_i:

$$N_{ir}^{(k)} = \Gamma_{ir}^{ET} \cdot X_r^{(k)} + \Phi_i^{ET} \ (r = 1, 2, \ldots, R).$$

(B) Cumulative effect of all target clusters onto environmental cluster D_i:

$$N_i^{(k)} = \bigoplus_{r=1}^{R} N_{ir}^{(k)}.$$

Step 3. Sum of effects on clusters of environmental items

$$E_i^{(k+1)} = M_i^{(k)} \oplus N_i^{(k)}.$$

For closer and refined information on how the fusion "$\oplus$" is defined and how the possible ellipsoidal regression problems are defined, we refer to the works (53; 54).

Since $\Gamma_{jr}^{TT} \cdot X_r^{(k)} + \Phi_j^{TT}$, $\Gamma_{js}^{TE} \cdot E_s^{(k)} + \Phi_j^{TE}$, $\Gamma_{is}^{EE} \cdot E_s^{(k)} + \Phi_i^{EE}$ and $\Gamma_{ir}^{ET}{}^{(k)}_r + \Phi_i^{ET}$ are affine-linear transformations, the sets $G_{jr}^{(k)}$, $H_{js}^{(k)}$, $M_{is}^{(k)}$ and $N_{ir}^{(k)}$ are ellipsoids. Additionally, $G_j^{(k)}$, $H_j^{(k)}$, $M_i^{(k)}$ and $N_i^{(k)}$ are described by sums of ellipsoids and, thus, form ellipsoids again. Hence, the model given here allows us to obtain and evaluate predictions $(\hat{X}_1^{(k+1)}, \ldots, \hat{X}_R^{(k+1)}, \hat{E}_1^{(k+1)}, \ldots, \hat{E}_S^{(k+1)})^T$ of the ellipsoidal states of targets or environmental items. In future research, we shall investigate the core properties of the ellipsoids and determine their centers and configuration matrices.

For example, for the time step $k \in \mathbb{N}_0$, predictions of $\hat{X}_r^{(k)}$ and $\hat{E}_s^{(k)}$ are expressed by the ellipsoids, i.e., $\hat{X}_r^{(k)} = \mathcal{E}(\mu_r^{(k)}, \Sigma_r^{(k)})$ and $\hat{E}_s^{(k)} = \mathcal{E}(\rho_s^{(k)}, \Pi_s^{(k)})$.

In the frame of ellipsoidal calculus, the most important problem is how the target and environmental clusters will be organized according the bubble concept. We recall that the dynamics investigated in our thesis, where "single ellipsoid" at time interval k is transferred into a single ellipsoid at time interval $k + 1$, can be considered as a special interpretation and application of that general model from ellipsoidal calculus, and both suit into a wider framework. In fact, our further suggestion is that log-prices can be used as one cluster of target variable, and economic indicators, such as GDP, inflation or interest rate, may be considered, namely, for environmental clusters one by one or all together or constituted by two or more indicators together within one cluster.

6.2   Receiver Operating Characteristics Analysis

6.2.1   Optimizing the Area Under the Curve

To compute the *probability* of bubble-burst time in the future, *Area Under the* (*ROC*) *Curve* may be used as a tool. It can be thought a kind of an early-warning signalling.

Receiver Operating Characteristics (*ROC*) analysis is designed for two-state problems such as default or non-default, dead or alive, since such problems are from the real world (of observations) which often expresses itself as a two-state world, and an ROC analysis is performed to assess the distinction power of a model. In order to develop an early-warning, we can benefit from ROC analysis by determining a threshold between two classes which could be "*bubble*" and "*no bubble*", respectively. These classes are

understood in the sense of a time of a bubble burst existing or not existing. Hence, the early-warning signalling displays an alarm when a measurement surpasses the threshold. In addition to this, we can develop *multi classes* instead of only two classes, and we can determine risk levels related to how much we are close to the bubble-burst time. As we remember, these levels and classes can be naturally identified with a finite number of colors. (cf. Subsection 5.3.1)

An ROC curve reflects the relationship among the "true positive fraction" (fraction of actually positive cases correctly classified as positive) and the "false positive fraction" (fraction of actually negative cases incorrectly classified as positive) (59), as represented in Figure 6.1 (43). Here, the words *positive* and *negative* correspond to the *bubble* and *no-bubble* cases, respectively. In terms taken over from medical diagnosis, *true positive fraction* (*TPF*) is equivalent to *sensitivity*, while *false positive fraction* (*FPF*) is equivalent to 1-*specificity*, as the sum of the FPF and TNF is equal to 1. Hence, an ROC curve displays the connection between sensitivity and specificity (59). The relationship between *FPF* and *TPF* may be explained by a *contingency* table, stated in Figure 6.1.

		truth	
		p	*n*
model outcome	$p^!$	True Positive Fraction (TPF)	False Positive Fraction (FPF)
	$n^!$	False Negative Fraction (FNF)	True Negative Fraction (TNF)
	total	1	1

Figure 6.1: Model estimation versus actual case of relative frequency of the bubble and no-bubble cases, represented by a *contingency table* or a *confusion matrix*.

According to the bubble concept, the contingency table can be interpreted as follows:

TPF: The model estimated as bubble and in reality bubble (*sensitivity*).
FPF: The model estimated as bubble and in reality no-bubble (1-*specificity*).
FNF: The model estimated as no-bubble and in reality bubble.
TNF: The model estimated as no-bubble and in reality no-bubble (*specificity*).

We introduce random variables S_p and S_n as model score distributions of *bubble* and *no-bubble* case, respectively, and suppose that they obey a normal distribution with different means and standard deviations. Mean and standard deviation of the no-bubble distribution are named by μ_n and σ_n, and for the bubble distribution by μ_p and σ_p, respectively. We suppose that the parameters σ_n and σ_p are nonzero. The shape of the populations is shown in Figure 6.2.

cut-off value

actually no-bubble case | actually bubble case

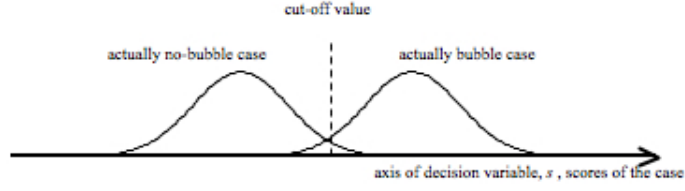axis of decision variable, s, scores of the case

Figure 6.2: Two normal distributions associated with actually bubble and actually no-bubble case.

The cumulative distribution function of the decision variable s, called the *score*, for the *no-bubble* and the *bubble* case are named as $F_n(s)$ and $F_p(s)$, respectively. Both functions can be described by (59):

$$F_n(s) := \Phi\left(\frac{s - \mu_n}{\sigma_n}\right) \quad \text{and} \quad F_p(s) := \Phi\left(\frac{s - \mu_p}{\sigma_p}\right). \tag{6.1}$$

The ROC curve is a function of TPF(c) versus FPF(c), drawn for each cut-off value c. The case can be classified as bubble if $s \leq c$, and as no-bubble if $s > c$. It is important to notice that we presume our cut-off values to increase as from left to right on the decision variable axis.

At first, the smallest score received after applying the model is investigated. The logit model can be applied in order to compute the scores. Here, the meaning of a score is forecasting the probability that an event takes place or not. That is to say, by using a set of independent variables we estimate a binary dependent outcome. For our case, the occurrence of a bubble burst can be included as a dependent variable in a relationship with macroeconomic indicators, and the linear probability model which is used to calculate the probability of occurrence of the bubble burst can be expressed as the following model (44):

$$\log\left(\frac{P(Y = 1 \,|X = x)}{P(Y = 0 \,|X = x)}\right) = \beta_0 + \beta_1 \cdot x_1 + \ldots + \beta_p \cdot x_p + \epsilon. \tag{6.2}$$

Here, $x = (1, x_1, x_2, \ldots, x_p)^T \in \mathbb{R}^{p+1}$ and β_m ($m = 0, 1, 2, \ldots, p$) are predictor vector and coefficients, respectively. By ϵ, we mean a *noise term*. The predictor X and its vectorial values x corresponds to the input variables of our bubble research or macroeconomic indicators. Moreover, Y represents a binary response, i.e., Y follows a Bernoulli distribution. Thus, for a Bernoulli distributed variable we have the relationship $P(Y = 1 \,|X) = E(Y \,|X)$ (60). Furthermore, the values 0 and 1 express the bubble and no-bubble case, respectively. As seen from Eqn. (6.2), $P(Y = 1 \,|X = x)$ obeys a logistic distribution, i.e.,

71

$$P(Y = 1 \mid X = x) = E(Y \mid X = x) = \left(\frac{\exp(X^T\beta)}{1 + \exp(X^T\beta)} \right)\Bigg|_{X=x}. \tag{6.3}$$

where $P(Y = 1 \mid X = x)$ represents the likelihood of the occurrence of the bubble burst.

The values of *FPF* and *TPF* corresponding to each setting of the cut-off point c can be stated by (59):

$$\text{TPF}(c) := Pr(s \le c \mid p) = \Phi\left(\frac{c - \mu_p}{\sigma_p}\right) \text{ and } \text{FPF}(c) := Pr(s \le c \mid n) = \Phi\left(\frac{c - \mu_n}{\sigma_n}\right). \tag{6.4}$$

The expressions in Eqn. (6.4) include cumulative distribution functions to represent how we can obtain each ROC curve point with respect to any different cut-off value. Thus, we can conclude that the ROC axes are comprised by the two cumulative distribution functions that are introduced in Eqn. (6.1). In order to make a decision about the model, i.e., whether its discriminative power of between two classes is good or bad, one of the most appealing ways is to compute the *Area Under the (ROC) Curve*, to observe the differences between the corresponding curves. This integral area is defined by the following term (52):

$$\text{Area Under the Curve} = AUC := \int_0^1 F_p(s) \, dF_n(s). \tag{6.5}$$

Ideally, an excellent model, i.e., a model that differentiates two classes very effectively, has an *AUC* with value very close the value 1. In other words, any model with a rather steep ROC curve is a better one, whereas an area value of 0.5 means a random forecasts. Accordingly, a greater *AUC* value can be regarded as performing better when it comes to the discrimination of two classes (74).

ROC curves produced from experimental visual data approximate closely theoretical ROC curves generated from normal probability distributions. Hence, they can be plotted as a straight line beyond binormal axes whose coordinates scale the normal-deviate value of the probability linearly (41). That transformation can be visualized in the form of Figure 6.3.
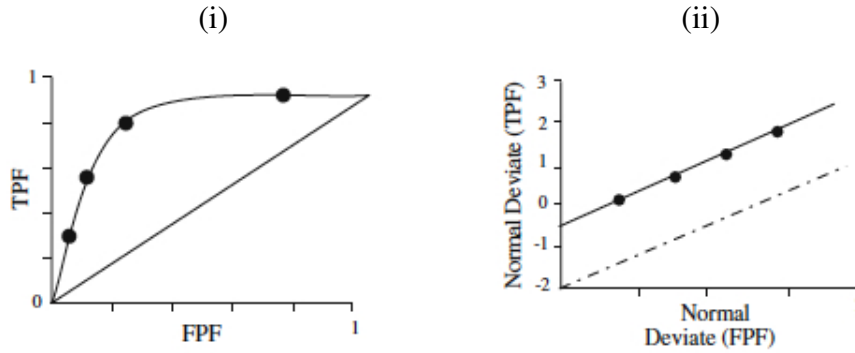
Figure 6.3: (i) ROC curve of *TPF* compared with *FPF* (79), and (ii) Normal-deviate plot of ROC curve (79).

Such an ROC curve can be identified through two parameters: a, being the ordinate intercept, and b, which is the slope of the straight-line plot on normal-deviate axes. If an ROC curve is plotted as a straight line over the normal-deviate axes, the vertical and horizontal coordinates of each point on this line are coupled by (59):

$$\Phi^{-1}(TPF) = a + b \cdot \Phi^{-1}(FPF). \tag{6.6}$$

Here, the function Φ^{-1} takes values in $(-\infty, \infty)$. Herewith, the coordinates of the ROC plot are received by

$$TPF = \Phi(a + b \cdot \Phi^{-1}(FPF)). \tag{6.7}$$

By the expressions in Eqn. (6.4) , a cut-off value can be formulated in terms of *FPF* and *TPF* as the following expression:

$$c = \sigma_n \cdot \Phi^{-1}(\mathrm{FPF}(c)) + \mu_n = \sigma_p \cdot \Phi^{-1}(\mathrm{TPF}(c)) + \mu_p. \tag{6.8}$$

If we drawn the statement $\Phi^{-1}(\mathrm{TPF}(c))$ from Eqn. (6.8), we obtain the following Eqn. (6.9) which will be used to get the ROC parameters:

$$\Phi^{-1}(\mathrm{TPF}(c)) = \frac{\mu_n - \mu_p}{\sigma_p} + \frac{\sigma_n}{\sigma_p}\Phi^{-1}(\mathrm{FPF}(c)). \tag{6.9}$$

Here, Eqn. (6.9) is accepted uniformly, i.e., for all $c \in \mathbb{R}$. If we compare Eqns. (6.6) and (6.9) on TPF(c) and FPF(c), which take the values into the interval [0,1], then we can decide the ROC parameters a and b as in Eqn (6.10):

$$a = \frac{\mu_n - \mu_p}{\sigma_p} \quad \text{and} \quad b = \frac{\sigma_n}{\sigma_p}. \tag{6.10}$$

In order to rewrite the ROC curve axes by using ROC parameters, we can assign a random variable v such that $s - \mu_n = \sigma_n \cdot v$, $\sigma_n \neq 0$. Via this linear combination between s and v, we can rewrite our cumulative distribution functions as follows:

$$F_n(s) = \Phi((s - \mu_n)/\sigma_n) = \Phi(v), \tag{6.11}$$

where $v := (s - \mu_n)/\sigma_n$. Substituting $s = \sigma_n \cdot v + \mu_n$ into $F_n(s)$ and $F_p(s)$, gives the distribution functions which will be used to explain the multi cut-off values case:

$$F_n(v) = \Phi(v) \quad \text{and} \quad F_p(v) = \Phi(a + b \cdot v). \tag{6.12}$$

As we mentioned at the beginning of this section, we can use ROC analysis for the multiple cut-off values, as seen in Figure 6.4 (56). These cut-off values can be employed as thresholds of risk levels which can indicate how we close the bubble-burst time.
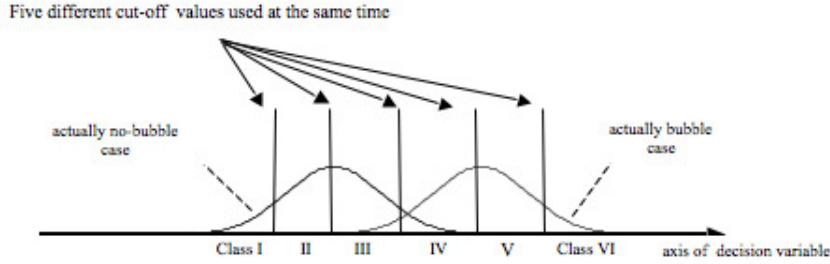

Figure 6.4: In a multiple cut-off values situation, the resulting classes are indicated with respect to the normal distribution curves.

In order to generalize the multiple cut-off values case, we presume that we have R classes. Therefore, we are concerned with $R - 1$ cut-off values, $c_1, c_2, \ldots, c_{R-1}$, which divide the decision variable axis into R classes as shown in Figure 6.4, and these cut-off values fulfill $c_1 < c_2 < \ldots < c_{R-1}$. The coordinates of a binormal ROC curve are then computed by Eqn. (6.12) as follows (59):

$$\text{FPF}(c_i) = F_n(c_i) = \Phi(c_i) \quad \text{and} \quad \text{TPF}(c_i) = F_p(c_i) = \Phi(a + b \cdot c_i). \tag{6.13}$$

As represented in Figure 6.5, the cut-off points are transformed into thresholds values by the standard normal cumulative distribution function. Consequently, the thresholds take values between 0 and 1. Therefore, we get the first threshold value as $t_0 = 0$, and the last one as $t_R = 1$, such that $0 < t_i < 1$ for all $i = 1, 2, \ldots, R - 1$, and the most crucial subject is that all threshold values are in a strictly increasing order from the smallest one to the greatest, i.e., $t_{i-1} < t_i$. The relationship between thresholds and cut-off values is described by $\Phi(c_i) = t_i$; these values are lying on the *FPF* axis of the ROC curve.
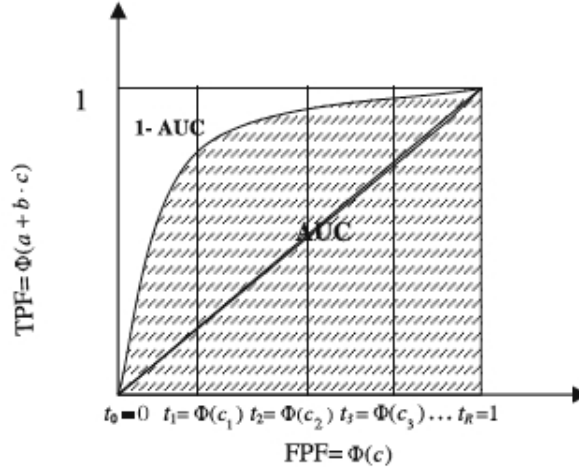
74

Figure 6.5: The relationship between thresholds and cut-off values.

In fact, this relationship can be explained more clearly and functionally by the following statement:

$$\Phi(c) = t \;\Leftrightarrow\; c = \Phi^{-1}(t). \tag{6.14}$$

Let us recall that we have introduced AUC as $AUC := \int_0^1 F_p(s)\,dF_n(s)$. We remind that we also defined $F_p(s)$ and $F_n(s)$ for the multiple cut-off values along with the Eqn. (6.13). Now, we can represent the AUC as the Eqn. (6.15) (for further technical details see (56)).

$$AUC := \int_0^1 \Phi(a + b \cdot \Phi^{-1}(t))dt. \tag{6.15}$$

By solving the following optimization problem, the thresholds and, herewith, the cut-off values can be obtained (56):

$$\max_{a,b,\tau} \quad \alpha_1 \cdot \int_0^1 \Phi(a + b \cdot \Phi^{-1}(t))dt - \alpha_2 \cdot \sum_{i=0}^{R-1}(\tfrac{\gamma_i}{n} - (t_{i+1} - t_i))^2$$

$$\text{subject to} \quad \int_{t_i}^{t_{i+1}} \Phi(a + b \cdot \Phi^{-1}(t))dt \geq \delta_i \quad (i = 0, 1, 2, \ldots, R-1), \tag{6.16}$$

$$t_0 = 0 \quad \text{and} \quad t_R = 1.$$

Here, the decision variable τ is a vector which comprises the thresholds t_i, whereas "regularity" parameters $\delta_i \geq 0$ are chosen by the modeler or decision maker beforehand, and the parameters γ_i are "sensitivity" for some particular purpose of "propor-

tionality" of the each classes. Moreover, γ_i and n indicate the number of historical data in the i th class ($i = 0, 1, 2, \ldots, R - 1$) and the total number of historical data in certain time interval, respectively. The ratio γ_i/n displays the relative frequency of the data in the i th class.

In fact, by solving the above optimization problem we realized three aims: The first one is (*i*) to compute the threshold values t_i and the ROC curve parameters *a* and *b* simultaneously, making the area under the ROC curve, *AUC*, maximal or, equivalently, since the underlying total area is equal to 1, minimizing 1-*AUC*. This is depicted in Figure 6.5. The second purpose (*ii*) is that these threshold values should be inclined to maximize the estimation accuracy of the model, i.e., that they should assure class sizes close to the data proportions directly provided by the data. That is to say, threshold values which tend to minimize the difference between some data-based value and the corresponding model estimation value for each class. The third goal is (*iii*) while realizing the first two goals, to balance the different subareas which altogether comprise the *AUC*, i.e., balancing the size of the classes, and to order the thresholds. Our third goal is approached by the inequality constraints in the above problem.

For further detailed and refined information on how this optimization becomes discretized in a balanced, regularized way and then solved as a *nonlinear regression* problem, we refer to our previous work (56).

Our new method of optimizing the *Area under the ROC Curve* that we explain here, is a module on our way to narrow down uncertainty related with the estimator which we present in this thesis. When optimizing that AUC, we maximize the discriminative power of the classification, which expresses itself by the identification of finitely many risk levels. Herewith, we are going beyond the classical ROC technique from statistics of "testing" the statistical quality of our estimator, but we are on the way of suiting that new method into the ellipsoidal methodology of this thesis. Maximizing the AUC strongly increases affirmation, bringing the estimation proposed by us *closer* to determinism but, of course, remaining in the realm of approximation. In our future work, we aim to further integrate all these methods of early warning and risk management.

CHAPTER 7

CONCLUSION AND OUTLOOK

In our thesis, we present an early-warning signalling for financial bubbles by bene-fitting from the theory of optimization, of inverse problems and clustering methods. For the time being, we consider the US, Japan and China stock markets since bubbles, mostly, occur in these markets. Yet, we will extend our work for further countries' stock markets and also for other types of assets such as gold and oil markets.

Here, we generated a new index and called it as *volume-based index*, and we have found that when the time approaches the bubble-burst moment, the volumes of the ellipsoids, i.e., the volume-based index, gradually decrease and, correspondingly, the signals obtained by Radon transform are becoming more "brilliant", i.e., more strongly warning. The meaning of more brilliant or more strongly warning is given by the color change of Radon transform. If it turns to orange, yellow or red, together with a decreasing process of the volume-based index, then that could be- and has been in our computations- an early-warning for the bubble-burst time. This tool is appropriate, particularly, for the less volatile markets. In future research, we focus on to find methods for more volatile markets.

Standard models and simulations of scenarios of extreme events are exposed to count-less sources of error and each of these sources can have a negative effect on the valid-ness of the estimations (51). In the modeling process, some of the ambiguities are under control and they generally include trade-offs between a more accurate descrip-tion and controllable computations. Other sources of errors are far beyond control as they are inherent in the modeling methodology of the particular fields (70). Even if there are remarkable counter-examples (3; 6; 28), analytical theoretical estimations are still out of reach for many complex problems and numerical resolution of the equations (when they are known) or of scenarios is credible in the center of the distribution, i.e., in the regime far from the extremes where good statistics can be accumulated (70). Hence, the two known strategies for modeling are both restricted in this regard (70). Crises with extraordinary effects are extreme events that take place seldomly, and are thus completely under-sampled and poorly constrained (70). In the light of these ex-planations, we did not make a simulation study in this thesis. We have also performed our volume-based approach on the fitted data. However, as a future work, we plan to benefit from *compound Poisson processes* (19; 50; 69) to explain bubbles and by this approach it may be possible to test the model by simulated data.

The originality of our study is that the concept of bubbles is dealt with geometrically by using the theory of inverse problems and one of the clustering methods, called minimum-volume covering ellipsoids, supported by optimization theory. This is a new and pioneering approach. Our study may contribute to a deeper theoretical understanding of financial bubbles. Investors may control their risks, and the risk management departments at the banks may use theoretical and empirical results of our investigations to manage their market risk. In addition to these groups interested, Central Banks and policy makers could benefit from our theoretical and empirical results, and they could test and apply our approach to monitor their risk levels and compute the risk associated with various economic factors. This could prevent from financial and economics crises.

Bibliography

[1] F. Allen and G. Gorton, Churning bubbles, The Review of Economic Studies, 60(4), pp. 813–836, 1993.

[2] S. Alvarez, What is the intrinsic value of a stock?, April 2013, http://www.investopedia.com/articles/basics/12/intrinsic-value.asp#ixzz2NoC3gA4z.

[3] I. Arad, L. Biferale, A. Celani, I. Procaccia, and M. Vergassola, Statistical conservation laws in turbulent transport, Phys. Rev. Lett, 87(16), pp. 62–64, 2001.

[4] A. Aster, B. Borchers, and C. Thurber, *Parameter Estimation and Inverse Problems*, Academic Press, 2nd edition, 2012.

[5] K. Baker, Singular value decomposition tutorial, http://www.ling.ohio-state.edu/˜kbaker/pubs/Singular_Value_Decomposition_Tutorial.pdf, August 2013.

[6] F. Barra, B. Davidovitch, and I. Procaccia, Iterated conformal dynamics and laplacian growth, Phys. Rev. E, 65(4), pp. 486–497, 2002.

[7] A. Beattie, Market crashes: The tulip and bulb craze, December 2012, http://www.investopedia.com/features/crashes/crashes2.asp#axzz2ED0bIWnb.

[8] F. BenHajji and E. Dybner, 3d graphical user interfaces, Technical report, Department of Computer and Systems Sciences Stockholm University and The Royal Institute of Technology, 1999.

[9] F. Biagini, Y. Hu, B. Øksendal, and T. Zhang, *Stochastic Calculus for Fractional Brownian Motion and Applications*, Springer, 2008.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

[11] M. K. Brunnermeier, *Asset Pricing under Asymmetric Information, Bubbles, Crashes, Technical Analysis, and Herding*, Oxford University Press, 2001.

[12] J. Calverley, *Bubbles and How to Survive Them*, Nicholas Brealey Publishing, 2004.

[13] C. Camerer, Bubbles and fads in asset prices, Journal of Economic Surveys, 3(1), pp. 3–41, 1989.

[14] S. L. Carty, Definition of stock market bubble, April 2013, http://www.ehow.com/about_6601375_definition-stock-market-bubble.html.

[15] A. Clauset, C. R. Shalizi, and M. E. J. Newman, Power-law distributions in empirical data, SIAM (Society for Industrial and Applied Mathematics), 51(4), pp. 661–703, 2009.

[16] J. Colombo, Stock market crash, the south sea bubble (1719-1720), December 2012, http://www.stock-market-crash.net/south-sea-bubble/.

[17] J. Colombo, Stock market crash, the stock market crash of 1929, December 2012, http://www.stock-market-crash.net/1929-crash/.

[18] J. Colombo, The bubble bubble, we may be on the verge of the next major bubble boom, January 2013, http://www.thebubblebubble.com/next-bubble/.

[19] R. Cont and P. Tankov, *Financial Modelling with Jump Processes*, Financial Mathematics Series, Chapman Hall/CRC, 2004.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley Series in Communication, New York : Wiley, 1991.

[21] Tulip mania, January 2013, http://en.wikipedia.org/wiki/Tulip_mania.

[22] The tulipomania, an investing bubble, December 2012, http://www.thetulipomania.com/.

[23] Yahoo finance, August 2013, www.finance.yahoo.com.

[24] J. B. De Long, A. Shleifer, L. H. Summers, and R. J. Waldmann, Positive feedback investment strategies and destabilizing rational speculation, The Journal of Finance, 45(2), pp. 379–395, 1990.

[25] B. T. Diba and H. I. Grossman, The theory of rational bubbles in stock prices, The Economic Journal, 98(392), pp. 746–754, 1988.

[26] What is the difference between the hough and radon transforms, June 2013, http://dsp.stackexchange.com/questions/470/whats-the-difference-between-the-hough-and-radon-transforms.

[27] The conditional expectations and martingales, January 2014, http://www.ems.bbk.ac.uk/for_students/msc_finEng/math_methods/lecture10.pdf.

[28] G. Falkovich, K. Gawedzki, and M. Vergassola, Particles and fields in fluid turbulence, Rev. Mod. Phys., 73(4), pp. 913–975, 2001.

[29] T. Feeman, *The Radon Transform*, The Mathematics of Medical Imaging, Springer Undergraduate Texts 11 in Mathematics and Technology, 2013.

[30] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, A theory of power-law distributions in financial market fluctuations, Nature, pp. 267–270, 2003.

[31] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, San Francisco : W. H. Freeman, 1979.

[32] Matlab user's guide: Genetic algorithm and direct search toolbox, April 2012, www.mathworks.com//help/releases/R13sp2/pdf_doc/gads/gads_tb.pdf.

[33] M. V. Ginkel, C. L. L. Hendriks, and L. J. V. Vliet, A short introduction to the radon and hough transforms and how they relate to each other, Technical Report QI-2004-01, The Quantitative Imaging Group Technical Report Series, 2004.

[34] G. H. Golub and P. V. Dooren, *Numerical linear algebra, digital signal processing, and parallel algorithms*, New York : Springer-Verlag, 1991.

[35] A. Greenspan, Economic volatility. remarks at a symposiums sponsored by the federal reserve bank of kansas city, jackson hole, wyoming, August 2002, www.federalreserve.gov/boarddocs/ Speeches/2002/20020830/.

[36] D. M. Hamby, A comparison of sensitivity analysis techniques, Health Physics, 68(2), pp. 195–204, 1995.

[37] J. M. Harrison and D. M. Kreps, Speculative investor behavior in a stock market with heterogeneous expectations, The Quarterly Journal of Economics, 92(2), pp. 323–336, 1978.

[38] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer, 2nd edition, 2009.

[39] S. Helgason, *Radon Transform*, Cambridge, 2nd edition, 1999.

[40] G. T. Herman and A. Kuba, *Discrete Tomography Foundations, Algorithms, and Applications*, Birkhäuser, Boston, 1999.

[41] G. A. Hermann, N. Herrera, and H. T. Sugiura, Comparison of inter laboratory survey data in terms of roc indices, J Nucl Med, 23, pp. 525–531, 1982.

[42] C. Hoilund, The radon transform, November 2007, http://www.cvmt.dk/education/teaching/e07/MED3/IP/Carsten_Hoilund_-_Radon_Transform.pdf.

[43] L. Hopley and J. V. Schalkwyk, The magnificent roc, July 2011, http://www.anaesthetist.com/mnm/stats/roc/Findex.htm.

[44] D. W. Hosmer and S. J. Lemeshow, *Applied logistic regression*, Wiley, New York, 2000.

[45] C. Iyigün, Clustering, Wiley Encyclopedia of Operations Research and Management Science, 2010.

[46] E. Jacobsson, How to predict crashes in financial markets with the log-periodic power law, Technical report, Stockholm University, 2009.

[47] A. Johansen, O. Ledoit, and D. Sornette, Crashes as critical points, International Journal of Theoretical and Applied Finance, 3(2), pp. 219–255, 2000.

[48] A. Johansen and D. Sornette, Financial "antibubbles": Log-periodicity in gold and nikkei collapses, Int. J. Mod. Phys. C, 10(4), pp. 563–575, 1999.

[49] A. Johansen, D. Sornette, and O. Ledoit, Predicting financial crashes using discrete scale invariance, Journal of Risk, 1(4), pp. 5–32, 1999.

[50] I. Karatsaz and S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer, New York, USA, 1991.

[51] W. J. Karplus, *The Heavens are Falling: The Scientific Prediction of Catastrophes in Our Time*, New York: Plenum Press, 1992.

[52] R. Korn and E. Baydar, Workshop on credit rating in view of basel ii, Technical report, Fraunhofer Institute for Industrial Mathematics, University of Kaiserslautern, Germany, 2006.

[53] E. Kropat, G. W. Weber, and C. S. Pedamallu, *Regulatory Networks under Ellipsoidal Uncertainty–Data Analysis and Prediction by Optimization Theory and Dynamical Systems*, volume Statistical, Bayesian, Time Series and other Theoretical Aspects, Springer-Verlag, Berlin, Heidelberg, 2012.

[54] E. Kropat, G. W. Weber, and J. J. Rückmann, Regression analysis for clusters in gene-environment networks based on ellipsoidal calculus and optimization, Dynamics of Continuous, Discrete and Impulsive Systems Series B, 17(5), pp. 639–657, 2010.

[55] E. Kürüm, G. W. Weber, and C. Iyigün, *Financial Bubbles*, volume Modeling, Optimization, Dynamics and Bioeconomy of *Springer Proceedings in Mathematics*, Springer, 2013.

[56] E. Kürüm, K. Yildirak, and G. W. Weber, A classification problem of credit risk rating investigated and solved by optimization of the roc curve, CEJOR (Central European Journal of Operations Research), 20(3), pp. 529–557, 2012.

[57] A. Langvillea, An introduction to singular value decomposition, May 2013, http://langvillea.people.cofc.edu/DISSECTION-LAB/Emmie'sLSI-SVDModule/p4module.html.

[58] E. Laporte and P. Le Tallec, *Numerical Methods in Sensitivity Analysis and Shape Optimization*, Birkhäuser, 2002.

[59] C. E. Metz, B. A. Herman, and J. H. Shen, Maximum likelihood estimation of roc curves from continuously distributed data, Stat Med, 17, pp. 1033–1053, 1998.

[60] M. Müller, Generalized linear models, Technical report, Fraunhofer Institute for Industrial Mathematics (ITWM), Kaiserslautern, Germany, 2004.

[61] J. Nocedal and W. Stephen, *Numerical Optimization*, New York, NY: Springer, 1999.

[62] J. M. Nougues, M. D. Grau, and L. Puigjaner, Parameter estimation with genetic algorithm in control of fed-batch reactors, Chemical Engineering and Processing, 41, pp. 303–309, 2002.

[63] I. Nourdin, *Selected Aspects of Fractional Brownian Motion*, Bocconi University Press Springer, 2012.

[64] S. Özöğür, *Mathematical Modelling of Enzymatic Reactions, Simulation and Parameter Estimation*, Master's thesis, Institute of Applied Mathematics of METU, Ankara, Turkey, 2005.

[65] The econ review, stock market crash causes depression, January 2013, http://www.econreview.com/events/stocks1929b.htm.

[66] J. Rosenberg, The stock market crash of 1929, December 2012, http://history1900s.about.com/od/1920s/a/stockcrash1929.htm.

[67] A. Saltelli, S. Tarantola, F. Campolonga, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley Sons, 2004.

[68] S. Shreve, *Stochastic calculus for finance I The Binomial Asset Pricing Model*, Springer, New York, USA, 2000.

[69] S. Shreve, *Stochastic calculus for finance II continuous-time models*, Springer, New York, USA, 2001.

[70] D. Sornette, Critical market crashes, Physics Reports, 1(378), pp. 1–98, 2003.

[71] D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems*, Princeton University Press, New Jersey, USA, 2003.

[72] The south sea bubble, a stock market bubble, December 2012, http://www.thesouthseabubble.com/.

[73] P. Sun and R. M. Freund, Computation of minimum-volume covering ellipsoids, Operations Research, 52(5), pp. 690–706, 2004.

[74] T. C. Tang and L. C. Chi, Predicting multilateral trade credit risks: comparisons of logit and fuzzy logic models using roc curve analysis, Expert Systems with Applications, 28, pp. 547–556, 2005.

[75] Ç. Tuncay and D. Stauffer, Power laws and gaussians for stock market fluctuations, Physica A, 1(374), pp. 325–330, 2007.

[76] C. Wood, The dutch tulip bubble of 1637, March 2006, http://www.damninteresting.com/the-dutch-tulip-bubble-of-1637/.

[77] S. Wright, *Primal-Dual Interior-Point Methods*, Philadelphia, PA: SIAM, 1997.

[78] W. Yan, *Identification and Forecasts of Financial Bubbles*, Ph.D. thesis, ETH, Zurich, Switzerland, 2011.

[79] A. Zemcov, L. L. Barclay, J. Sansone, and C. Metz, Receiver operating characteristic analysis of regional cerebral blood flow in alzheimer's disease, J Nucl Med, 26, pp. 1002–1010, 1985.

[80] Y. Zhang and L. Gao, On numerical solution of the maximum volume ellipsoid problem, SIAM J. Optim, 14(1), pp. 53–76, 2003.

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Kürüm, Efsun
Nationality: Turkish
Date and Place of Birth: 18 November 1978 , Ankara
Marital Status: Single
Phone: (533) 441 10 00

EDUCATION

Degree	Institution	Year of Graduation
Ph.D.	Financial Mathematics, METU	February 2014
Scientific Preparation	Financial Mathematics, METU	June 2006
M.S.	Financial Mathematics in Life , METU and Pension Insurance	August 2005
B.S.	Mathematics, Hacettepe University	June 2001

PUBLICATIONS

International Refereed Journals

E. Kürüm, K. Yildirak and G.-W. Weber, A classification problem of credit risk rating investigated and solved by optimization of the ROC curve, CEJOR (Central European Journal of Operations Research) 20, 3, 529-557, 2012.

E. Kürüm, G.-W. Weber, C. Iyigün, Early Warning on Stock Market Bubbles via Methods of Clustering, Optimization and Inverse Problems, 2014, (to be submitted).

International Refereed Proceedings

E. Kürüm, N. Azevedo, D. Pinheiro and G.-W. Weber, Stochastic Control of Hybrid Systems with Applications to Finance, 8th International Conference on

Optimization: Techniques and Applications (ICOTA8), Shanghai, China, December 10-13, 2010.

Book Chapter

E. Kürüm, G.-W. Weber, C. Iyigün, Financial Bubbles, to appear as chapter in book, Springer volume *Modeling*, *Optimization*, *Dynamics and Bioeconomy*, series *Springer Proceedings in Mathematics* (2013), D. Zilberman and A. Pinto, eds..

PARTICIPATION IN INTERNATIONAL SCIENTIFIC MEETINGS

Presentations and Scientific Events Held

E. Kürüm, K. Yildirak and G.-W. Weber, *A Classification Problem of Credit Risk Rating Investigated and Solved by Optimization of the ROC Curve*, EURO XXIII, Bonn, Germany, July 5-8, 2009.

G.-W. Weber, **E. Kürüm** and K. Yildirak *Prediction of Credit Default by Continuous Optimization*, 4th International Summer School Achievements and Applications of Contemporary Informatics, Mathematics and Physics, National University of Technology of the Ukraine, Kiev, Ukraine, August 5-16, 2009.

E. Kürüm, G.-W. Weber, *Modelling Financial Bubbles by Stochastic and Ellipsoidal Calculus and Copulas*, EURO XXIV, Lisbon, Portugal, July 11-14, 2010.

E. Kürüm, Stochastic Modelling and Financial Mathematics, EURO XXIV, Lisbon, Portugal, July 11-14, 2010; Session Organizer, Session Chair.

E. Kürüm, G.-W. Weber, *Understanding Dynamics of Financial Bubbles by Geometrical and Topological Aspect*, OR 2011, Zurich, Switzerland, August 30 - September 2, 2011.

SEMINARS

E. Kürüm, *Simple Construction of the Efficient Frontier*, Middle East Technical University, Institute of Applied Mathematics, Turkey, May, 2005.

E. Kürüm, *Annuities with Controlled Random Interest Rates*, Middle East Technical University, Institute of Applied Mathematics, Turkey, August, 2005.

E. Kürüm, *Detecting Stock Market Bubbles via Methods of Clustering and Inverse Problems*, Middle East Technical University, Institute of Applied Mathematics, Turkey, December, 2013.

PROJECTS

BAP Project, 01.01.2009 − 31.12.2009 A Classification Problem of Credit Risk Rating Investigated and Solved by New Methods of Statistical Learning and Optimization.

(BAP: Research Fund for Scientific Collaboration in the Context of Post-Graduate Studies)

TUBITAK Project, 15.02.2013 − 15.02.2014 Modeling and Understanding Dynamics of the Financial Crises via Ellipsoidal Calculus and Clustering Methods.

TUBITAK (The Scientific and Technological Research Council of Turkey).

PROFESSIONAL ACTIVITIES

Referee, Journal of Computational and Applied Mathematics (CAM), January 2013−Present.

Refree, European Journal of Operational Research (EJOR) , 2010−Present.

Refree, Machine Learning, 2010−Present.