Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Large Deviations of Stock Market Returns

Author:
Konstantinos EVANGELOU

Supervisor:
Prof. Didier SORNETTE

Master Thesis for the Master of Science in Physics

Entrepreneurial Risks

ETH Zürich

April 17, 2013

Abstract

We study the application of the large deviations theory (LDT) to fat tailed distributions and in particular to power law distributions. According to the large deviations theory, when a sample exhibits a large deviation in the mean, or some other moment, then it should be because one of the variables takes an extensive value, that can almost alone result to the deviation.

We test the theory's assumptions in the context of stock market returns, whose distribution's tail can be approximated with a power law. Our findings include that returns are "democratically" distributed in larger scales, but that large deviations from an expected, or fundamental average return are plentiful when zooming in smaller time intervals. We show that these deviations can usually be associated with a few extremely large, outlier returns and also examine if the large deviations theory can be used to identify these outliers. Furthermore we explore the possibilities of utilizing large deviations for predicting the index's behaviour and discuss the shortcomings of such an attempt.

Oh, the hand of a terrible croupier is that touch on the sleeves of his dreams: all in his life of what has looked free or random, is discovered to've been under some Control, all the time, the same as a fixed roulette wheel-where only destinations are important, attention is to long-term statistics, not individuals: and where the House always does, of course, keep turning a profit...

Thomas Pynchon, Gravity's Rainbow

Acknowledgements

I would like to express my gratitude to Prof. Didier Sornette for trusting me with this master thesis topic and thus introducing me to many subjects that were hitherto unknown to me. His devotion to expanding the boundaries of human knowledge will guide my future endeavours. Comments from Dr. Peter Cauwels were also really helpful, especially those concerning the 6th chapter.

A big special thanks goes to my parents and to my sister, who stood by me throughout the course of my studies in every conceivable way. Friends of old times and new should also know that I am grateful to have them.

This master thesis was conducted with the support of the Greece's State Scholarships Foundation (I.K.Y.). I am indebted to them, for they offer a service of great value in difficult times.

Contents

List of Figures

Chapter 1

Introduction

Suppose we throw a die n times and calculate the mean value of the n random values, $\overline{x} = \sum_{i=1}^{6} i f_i$, which according to the law of large numbers should be $\overline{x} \to 3.5$, since $f_i = 1/6$ and $i = 1, ..., 6$. Now we may go on to think about calculating the probability of observing a mean value other than 3.5 and in particular a considerably larger value, such as $\overline{x} \geq 4$ and also ask what will be the numbers to which the frequencies f_i converge. This now conditional probability will concentrate in the neighbourhood of a specific point, that can be computed through the minimization of a functional $I(f)$, which in this case will be

$$I(f) = \sum_{i=1}^{6} f_i ln\Big(\frac{f_i}{1/6}\Big) \tag{1.1}$$

Then the conditional probability is given by

$$p \simeq e^{-nI(f^*)},$$

meaning that when n is large, p is concentrated where $I(f)$ is minimal. When no constraints are imposed, $I(f)$ is minimal for $I(f) = 0$ and we go back to the law of large numbers. But when a constraint -for instance coming from our observations- exists, then a different set of frequencies will minimize $I(f)$. In our case minimizing equation (1.1) under e.g. an observed deviation of the mean $\overline{x} = \sum_{i=1}^{6} i f_i = 4$ yields $f_1 = 0.103$, $f_2 = 0.122$, $f_3 = 0.146$, $f_4 = 0.174$, $f_5 = 0.207$, $f_6 = 0.246$ [1].

This is a basic example of the problems that large deviations theory deals with, which was first introduced as a unified theory by Varadhan [2] in 1966. Large deviations theory is concerned with the study of probabilities of rare events and provides a set of methods to derive such results. Since extremal events have been found to play an increasingly major role in many settings [3], large deviations theory has become a field of active study and finds numerous applications, including financial ones.

Recently the problem of large deviations of power law distributions has been addressed [4], concluding to the very interesting result that when a large deviation is observed, the frequency distributions should remain unchanged, except for the appearance of a single large event that will bear almost alone the responsibility for this large deviation. The study of power laws is also at the heart of research interest, as they are part of a class of distributions that are termed fat-tailed, that have been associated with many phenomena where extreme events make their presence felt. Thus the large deviations theory seems to provide a really intriguing theoretical explanation as to why such extreme events occur.

An interdisciplinary approach of finance from economists, physicists and mathematicians have set it as a prominent area among those where power law distributions have a role to play. In this thesis we will focus on the distribution of returns of stock prices, whose tail is characterized by a power law decay, with an exponent that is usually estimated to be close to $\mu \simeq 3$ [5]. This trait of stock market returns is believed to be associated with many interesting features that the market exhibits, including extreme events such as bubbles and crashes [6].

If the power law description of the distribution of returns is justified, then the predictions of the large deviations theory should at some extent be realized at stock markets. In a nutshell, what we expect to find is that whenever returns appear to deviate from a fundamental return, that this should be because of one or few extremely large returns. Further inquiry includes the direct linking between theory and empirical evidence and an attempt to find practical uses, like predicting potential crashes. There is a number of points to consider when making such an attempt, from how firmly the power law behaviour is established to what one should identify as fundamental return. Throughout the rest of this study we tackle each of these questions from several prospectives, in an attempt to examine the large deviations theory within a real context.

Beginning with the next two chapters we provide the mathematical foundation for our study. This includes an introduction to power law distributions and the formulation of large deviations theory with respect to them. Also we devote a few pages to the likelihood function, which through its numerous applications will be used in the chapters that follow. Chapter 4 introduces power law distributions in finance and examines the conditions that need to be satisfied, in order for the large deviations theory to be applied. The main part of our study is presented in chapter 5, where we detect large deviations in financial data and compare the results with the predictions of the large deviations theory. At the next chapter, we attempt a different approach and through logical reasoning try to find potential applications for large deviations of fat tailed distributions. We also deal with the shortcomings of our study. Chapter 7 concludes.

Chapter 2

Power laws and large deviations

2.1 PDF's and power laws

A probability distribution function (pdf) $P(x)$ of X is defined such that the probability to find X in an interval Δx around x is $P(x)\Delta x$. The probability to find X between a and b is then given by

$$\mathcal{P}(a < X < b) = \int_a^b P(x)dx, \tag{2.1}$$

We also define the cumulative distribution function as

$$F_X(x) = \mathcal{P}(X \leq b) = \int_{-\infty}^b P(x)dx, \tag{2.2}$$

Here, we are mainly interested in power law distributions of the kind

$$P(x) \propto \frac{C_\mu}{x^{1+\mu}}, \tag{2.3}$$

Power laws are of particular importance to the study of complex systems [7] and their behaviour is largely defined by the value of μ. Depending on the existence or not of the distribution's mean and variance we find three different regimes:

- $\mu > 2$: Both the mean and the variance are finite.

- $1 < \mu \leq 2$: The mean is finite and the variance is not defined.
- $\mu \leq 1$: Both the mean and the variance are not defined.

Power law distributions with $\mu \geq 2$ obey the Central Limit Theorem, which roughly states that *"the sum, normalized by $1/\sqrt{N}$ of N random independent and identically distributed (i.i.d.) variables of zero mean and* [1]*finite variance σ^2 is a random variable with a pdf converging to the Gaussian distribution with variance σ^2"* [8]. On the other hand power laws with $0 < \mu < 2$, which are termed heavy tail distributions, converge to another class of distributions, called Lévy laws.

2.2 Large Deviations Theory

The Central Limit Theorem, when applicable, provides a good approximation (the Gaussian law) only for the center of the pdf of a sum of a large number N of random variables, leaving its tail subject to further investigation. Large deviations theory deals with the problem, calculating the probabilities of rare events at the tail of the pdf[2]. Considering the sum of N i.i.d. random draws from a distribution $Q(x)$, we are interested in calculating the probability of events

$$E = \frac{1}{N}\sum_{i=1}^{N} g(X_i), \tag{2.4}$$

where $g(x)$ is a function that satisfies the law of large numbers (e.g. by substituting $g(x) = x$ in the above sum we calculate the sample mean) and $E \in [\overline{g}, \overline{g} + \delta g]$. If $\langle g \rangle = \int dx g(x) Q(x) \notin [\overline{g}, \overline{g} + \delta g]$, then the event is considered not typical. For $N \to \infty$ the probability of finding such events is given by

$$P(E) \propto \exp^{NI(\overline{g})} dx, \tag{2.5}$$

where $I(\overline{x})$ is the so called Cramér or rate function. We can consider that the large deviation is realized as a typical i.i.d. sample from a modified pdf

$$P^*(x) = \frac{1}{Z(\beta)} Q(x) e^{-\beta g(x)}, \tag{2.6}$$

[1]This condition can be relaxed to include cases like power laws with $\mu = 2$

[2]We provide here only an outline of Large deviations theory, mostly following [4]. For a more thorough treatment, see [8] and [9]

where $Z(\beta)$ is a normalization constant (known in statistical physics as the partition function) and β is chosen such that $\int dx P^*(x) g(x) = \overline{g}$. To illustrate the previous statements, consider the simple case where $g(x) = x$, then equation (2.5) gives the probability of observing a sample mean different than the mean ($\overline{x} \neq \langle x \rangle$), while equation (2.6) states that such a large deviation in the mean can be thought of as a sequence drawn independently from $P^*(x)$. The Cramér function is given by

$$I(\overline{x}) = D_{KL}(P^*||G),$$

where $D_{KL}(P^*||G)$ is the Kullback-Leibler divergence for the distribution $P^*(x)$ that minimizes it [10]. It is always $I(\overline{x}) \geq 0$, where $I(\overline{x}) = 0$ is realized when the distribution remains unchanged, $\langle g \rangle = \overline{g}$. This description is sufficient if the pdf of $g(X)$ decays at least as fast an exponential, for $N \to \infty$. In the case of pdf's that decay slower than an exponential, called fat tail distributions, we have to deploy a different strategy. The reason is that for distributions of this kind the quantity $e^{\beta x} P(x)$ (see function (2.6)) diverges for all $\beta > 0$, when $x \to \infty$.

2.3 Large Deviations for Power Laws

We again consider a power law distribution

$$P(x) \propto \frac{C_\mu}{x^{1+\mu}},$$

then (2.6) becomes

$$P^*(x) = \frac{1}{Z(\beta)} \frac{C_\mu}{x^{1+\mu}} e^{-\beta g(x)}, \tag{2.7}$$

with a normalization constant

$$Z(\beta) = \sum_{x=0}^{\infty} \frac{C_\mu}{x^{1+\mu}} e^{-\beta g(x)}, \tag{2.8}$$

We can easily show that $Z(\beta)$ is finite only if $\beta \geq 0$. We now consider the expected value of $g(X)$ under distribution (2.7), using the -well known from statistical physics- relation

$$E_\beta(g(X)) = -\frac{d}{d\beta} log Z(\beta), \tag{2.9}$$

from where we find that $E_\beta(g(X))$ is a decreasing function of β and $E_{\beta=0}(g(X)) = \langle g(x) \rangle$. Those results are reminiscent of a phase transition [4] and suggest that there are two regimes, above and below $\langle g(x) \rangle$, which need to be treated separately.

- $\overline{g(x)} \leq \langle g(x) \rangle$

When the conditional value is smaller than the unconditional, then there is always a value of $\beta(x)$ that satisfies $E_{\beta=0}(g(X)) = \overline{g(x)} \leq \langle g(x) \rangle$, meaning that deviations can occur in a "democratic" way. To achieve this, one may truncate the pdf by an exponential function (cut-off), in order to reduce its expected value.

For example consider the pdf $N = 1/M^{1+2/3}$. Introducing instead the gamma distribution $N = 1/M^{1+2/3} exp(-\beta M)$ leads to a cut-off at sufficiently large values. This is exactly the case where we wish to impose the finiteness of the maximal energy that may be released in the Earth, as a condition to the Gutenberg-Richter power law distribution of earthquake seismic moment releases([8] sec. 3.3.5).

- $\overline{g(x)} > \langle g(x) \rangle$

For conditional values larger than the unconditional, equation (2.9) would require a value $\beta < 0$, something that is not acceptable in our analysis since (2.8) is finite only if $\beta \geq 0$. A different approach is required.

It turns out that such large fluctuations can be realized if a single variable X_i^* accumulates the largest part of the average, by taking an extensive value [4]. Then

$$g(X_i^*) \simeq \langle g(x) \rangle + N(\overline{g(x)} - \langle g(x) \rangle), \tag{2.10}$$

The rest of the variables are typical (meaning that if we only considered them, then it would be $\overline{g(x)} = \langle g(x) \rangle$). Thus, the probability to observe $\overline{g(x)}$ depends only on the probability $P(X_i^*)$ (as the non-deviating values will appear with probability one) and since there are N different ways to choose i^*, is at least

$$NP(X_i^*) = \frac{NC_\mu}{(X_i^*)^{1+\mu}}, \tag{2.11}$$

or in the simple case of $g(x) = x$, where we can directly substitute X_i^*

$$NP(X_i^*) = \frac{NC_\mu}{(N(\overline{x} - \langle x \rangle) + \langle x \rangle)^{1+\mu}}, \tag{2.12}$$

The Cramér function of equation (2.5) is shown to be zero. This means that the pdf is unchanged, but there is one large event that is singular and of sufficient amplitude to change $\langle g(x)\rangle$ into $\overline{g(x)}$. This application of large deviations theory to power law distributions appears to provide a plausible explanation for the appearance of extreme events that dominate the weight of the power law distribution and are found at the tail of the pdf. These events are also known as Dragon-Kings [11], a concept that is discussed later.

2.3.1 Power Laws with $\mu < 1$

Equation (2.12) is problematic if $\mu < 1$, since the standard natural conditioning moment (the mean) is not defined. To tackle this problem we first remember that the invariant quantity is the probability $P(x)\Delta x$ and not just the pdf. Therefore $P(x)\Delta x = P(y)\Delta y$ and we can proceed with a change of variables. If for example we set $y = x^a$, then we get

$$\int_b^c \frac{1}{x^{1+\mu}} dx = \int_{b'}^{c'} \frac{1}{a} \frac{1}{y^{1+\mu/a}} dy, \tag{2.13}$$

We then define the coefficient $r_n = \frac{\sum_{i=1}^{N} x_i}{M_N}$, where M_N is the largest observation. For $\mu < 1$, r_n is on the order of unity [12]. This means that the major contribution to the sum is made by the maximal observation (a sort of spontaneously occurring large deviation) or in other words we observe a "condensation" for a single x-realization. That allows us to assume that this condensation of a large amplitude for a single x-realization also occurs in the case of upward conditioning, so we can choose a, such that it corresponds to some moment of order less than μ (e.g. $a = 1/2$) and calculate $\langle x^a\rangle$ instead of $\langle x\rangle$, as

$$\langle x^a\rangle = \int_b^c \frac{x^a}{x^{1+\mu}} dx = \int_{b'}^{c'} \frac{1}{a} \frac{y}{y^{1+\mu/a}} dy \tag{2.14}$$

We therefore impose a conditional moment of order $q = a, (\overline{x^a})$ instead of $q = 1, (\overline{x})$ and compare $\overline{x^a}$ with $\langle x^a\rangle$, which can be computed analytically through (2.14). Following the same methodology as before we again have two distinct cases:

- $\overline{x^a} > \langle x^a\rangle$

The largest event again accommodates the large deviation, where we now consider that $g(x) = x^a = y$, instead of $g(x) = x$. The equivalent of equation (2.11) for $y = x^a$ will be

$$NP(Y_i^*) = \frac{NC_\mu}{(Y^*)^{1+\mu/a}} = \frac{NC_\mu}{((X_i^*)^a)^{1+\mu/a}},$$

because, as seen from (2.14), the pdf after the change of variables is no longer the same. Then, through (2.10) we get $(X^*)^a = \langle x^a \rangle + N(\overline{x^a} - \langle x^a \rangle)$ and subsequently $X^* = (\langle x^a \rangle + N(\overline{x^a} - \langle x^a \rangle))^{1/a}$. Finally, we substitute X^*, getting

$$NP(Y_i^*) = \frac{NC_\mu}{(\langle x^a \rangle + N(\overline{x^a} - \langle x^a \rangle))^{1+\mu/a}}, \tag{2.15}$$

- $\overline{x^a} \leq \langle x^a \rangle$

Here once more the distribution needs to be truncated by an exponential cut-off, i.e.

$$P(y) = \frac{1}{a} \frac{e^{-\beta y}}{y^{1+\mu/a}},$$

leading to

$$P(x) = P(y) \frac{dy}{dx} = \frac{e^{-\beta x^a}}{x^{a+\mu}} x^{a-1} = \frac{e^{-\beta x^a}}{x^{1+\mu}}$$

Chapter 3

Likelihood

3.1 The likelihood function

Statistical inference refers to the concept of inducing general statements or laws, on the basis of data observations. Given a specific dataset, one tries to find those probabilistic models that can better describe her observations, but which can also satisfactorily forecast the outcome of future events.

To this end the likelihood function is an important statistical tool, that can be used to summarize data. It is defined such that the likelihood of a hypothesis/distribution after observing a set of data, is equal to the probability of the data given the hypothesis. If $f(x|\theta)$ denotes the joint pdf of the sample $X = (X_1, ..., X_N)$, given a distribution that is a function of θ, then the likelihood function, given that $x = X$ is observed, is

$$\mathcal{L}(\theta|x) = f(x|\theta), \tag{3.1}$$

thus being the same in form as a pdf, where the only distinction is which variable is considered fixed and which is varying. The way of using the likelihood function as a means of data reduction is provided by the *likelihood principle*, which states that for two different samples $X = (X_1, ..., X_N)$ and $X' = (X'_1, ..., X'_N)$, for which there exists a constant $C(x, y)$, such that

$$\mathcal{L}(\theta|x) = C(x, y)\mathcal{L}(\theta|x'), \tag{3.2}$$

then the conclusions drawn from x and x' should be identical. If $\mathcal{L}(\theta_1|x) = 2\mathcal{L}(\theta_2|x)$, then in some sense θ_1 is 2 times as plausible than θ_2[1]. Then if (3.2) holds true, $\mathcal{L}(\theta_1|x') = 2\mathcal{L}(\theta_2|x')$, so whether we observe the sample X

[1] This is a consequence of the *Law of Likelihood* which is discussed in section (3.3)

or X' (or any other) we will conclude that θ_1 is 2 times as plausible as θ_2 ([13], chapter 6.3). This allows us to use the likelihood function to compare different models, in order to see which one is most likely to describe our observations. This idea of comparing different models based on their relative likelihoods is further discussed at the third part of this chapter.

3.2 Maximum-likelihood estimation

Maximum-likelihood estimation (MLE) is a method used when a model is believed to satisfactorily describe our data observations and we are interested in estimating its specific parameters. We can rewrite equation (3.1) such that

$$\mathcal{L}(\theta|x) = \mathcal{L}(\theta_1, ..., \theta_k, |x_1, ..., x_N) = \prod_{i=1}^{N} f(x_i|\theta_1, ..., \theta_k) \tag{3.3}$$

where $\theta_1, ..., \theta_k$ are the parameters we want to estimate. For each sample point, let $\hat{\theta}$ be a parameter value at which $\mathcal{L}(\theta|x)$ attains its maximum as a function of θ, then a maximum-likelihood estimator of the parameter θ based on a sample X is $\hat{\theta}$. The maximum-likelihood estimator corresponds to the parameter point for which the observed sample is most likely. For a differentiable likelihood function, the maximum-likelihood estimator can be found through

$$\frac{\partial}{\partial \theta_i}\mathcal{L}(\theta|x) = 0 \tag{3.4a}$$

$$\frac{\partial^2}{\partial \theta_i^2}\mathcal{L}(\theta|x) < 0 \tag{3.4b}$$

with $i = 1, ..., k$. It is sometimes easier to compute the maximum-likelihood estimator of a likelihood function by finding the maximum of the logarithm of $\mathcal{L}$, known as the log-likelihood function:

$$ln\mathcal{L}(\theta_1, ..., \theta_k, |x_1, ..., x_N) = \sum_{i=1}^{N} lnf(x_i|\theta_1, ..., \theta_k) \tag{3.5}$$

The peaks of the likelihood and of the log-likelihood will coincide, as the log function is monotonically increasing.

3.2.1 MLE for a power-law distribution

We will consider data that are believed to be power law distributed. In this case they are expected to obey a pdf that has the form

$$P(x) = \frac{C_\mu}{x^{1+\mu}}$$

For the distribution to be fully described, the two parameters C_μ and μ have to be determined. For that purpose, the maximum likelihood estimator $\hat{\mu}$ is generally preferred over alternatives like graphical methods, because in the limit of asymptotically large samples it has desirable features like being unique, consistent and asymptotically efficient [14]. The above equation can be rewritten as

$$P(x) = \frac{\mu}{x_{min}} \left(\frac{x}{x_{min}} \right)^{-\mu-1} \tag{3.6}$$

where x_{min} is the minimum value for which power-law behaviour holds. The log-likelihood of the data given such a power law model is then

$$ln\mathcal{L}(\mu|x) = \prod_{i=1}^{N} \frac{\mu}{x_{min}} \left(\frac{x_i}{x_{min}} \right)^{-\mu-1}$$

Using the maximum likelihood estimation that was described earlier we find the maximum likelihood estimator (also known as Hill estimator [15])

$$\hat{\mu} = N \left[\sum_{i=1}^{N} ln \frac{x_i}{x_{min}} \right]^{-1} \tag{3.7}$$

This is the value of μ that corresponds to the power-law that is most likely to have generated the data.

3.2.2 Determining the lower bound x_{min}

The task of determining the threshold x_{min} is not trivial and one has to consider the boundaries of power law behaviour. As can be seen from equation (3.7), the selection of x_{min} also affects the scaling parameter μ. The simplest way to find this lower bound is to plot the PDF (or CDF) of the distribution on a log-log diagram and to place x_{min} there where a straight line begins, but this method is subjective and sensitive to small fluctuations [16].

As an alternative we can choose x_{min}, such that it makes the probability distribution of the data and the best-fit power-law model as similar as possible above x_{min}[17]. This is achieved by calculating the Kolmogorov-Smirnov statistic, which is the maximum distance between the empirical CDF corresponding to the data for a given x_{min} and that of the power law that best fits the data for $x \geq x_{min}$,

$$D_n = \max_{x \geq x_{min}} |S(x) - F_X(x)|,$$

and then selecting the x_{min} that minimizes D_n. The Kolmogorov-Smirnov statistic provides good results and its only shortcoming is that it is sensitive to the number of data points in the distribution tail. Good results can be achieved if $\sim$ 1000 or more observations are available in this part of the distribution[17], otherwise yielding slightly larger estimates for x_{min}. We should have that in mind when calculating the two parameters C_μ and μ for stock market returns in the next chapter.

3.3 Comparing the fit of different models

An important related concept, that was already used in (3.1) is the *law of likelihood* which roughly states that "*When comparing two hypotheses for a particular dataset (i.e. that a random random variable X takes the value x according to the pdf's $f_{1,2}$), then $f_1(x) > f_2(x)$ suggests that the first hypothesis is more likely and the likelihood ratio gives the degree to which the observation x supports the first hypothesis against the second one*" [18]:

$$\Lambda = \frac{\mathcal{L}(H_1|X=x)}{\mathcal{L}(H_2|X=x)} = \frac{f_1(X=x|H_1)}{f_2(X=x|H_2)}. \tag{3.8}$$

In practice the above likelihood ratio is considered a reliable statistical test to compare the fit of two models, if one of them (the null model) is a special case of the other (the alternative model), usually with the latter having more degrees of freedom. The null hypothesis is rejected if the value of the likelihood ratio is too small, otherwise there is no support to consider a more complicated model.

When the models under consideration are not related (i.e. one is not a nested case of the other), the likelihood ratio does not suffice to make an assumption on their relative goodness of fit and other methods need to be used. One such method is given by the Akaike Information Criterion, which is grounded in information entropy and the concept of Kullback-Leibler information [20]:

$$AIC = 2k - 2ln\mathcal{L}_m, \tag{3.9}$$

where k corresponds to the independently adjusted parameters within the mode and $\mathcal{L}_m$ is the maximized likelihood. Given a set of candidate models for the observed data, the one with the minimum AIC value is to be preferred.

3.4 Distribution of large deviations

We now turn again to the case of power-law pdf's. Supposing that we observe a large deviation in the mean or in some other moment, of the general form $\overline{g(x)} > \langle g(x) \rangle$, then we may use the likelihood ratio to estimate the model that best describes the data. Potential candidates are a power law where the deviation is equally distributed, or if we follow the reasoning of section (2.2), then there will be one event that is accommodating the largest proportion of the deviation. In the following we present how to calculate the likelihood in the case of no or one outliers.

1 The large deviation is equally distributed

All N observed variables are typical and subject to a power-law distribution. The likelihood for such a configuration is then

$$\mathcal{L}_0 = \prod_{i=1}^{N} \frac{C_\mu}{x_i^{1+\mu}} \tag{3.10}$$

2 The large deviation concentrates on the largest event

The $N-1$ smaller variables are typical, but the largest one X^* causes the large deviation and is described by the formula $X^* = \langle x \rangle + N(\overline{x} - \langle x \rangle)$. The likelihood that such a configuration describes our observations is then

$$\mathcal{L}_1 = \frac{NC_\mu}{(N(\overline{x} - \langle x \rangle) + \langle x \rangle)^{1+\mu}} \prod_{i=1}^{N-1} \frac{C_\mu}{x_i^{1+\mu}} \tag{3.11}$$

where the first RHS term is given by equation (2.12), while the rest variables obey to the initial power-law pdf.

Since we are only interested in seeing which of the likelihoods is larger -and not in their absolute values-, we can ignore the terms that are the same in both (3.11) and (3.10) as they cancel out, and compare only the following likelihoods:

$$\mathcal{L}'_0 = \frac{C_\mu}{X_N^{1+\mu}} \tag{3.12a}$$

$$\mathcal{L}'_1 = \frac{NC_\mu}{(N(\overline{x} - \langle x \rangle) + \langle x \rangle)^{1+\mu}} \tag{3.12b}$$

for which we get the following likelihood ratio:

$$\Lambda = \frac{(N(\overline{x} - \langle x \rangle) + \langle x \rangle)^{1+\mu}}{NX_N^{1+\mu}} \tag{3.13}$$

Chapter 4

Distribution of stock market returns

The large deviations theory that was described in the second chapter could find an interesting application at financial markets. For instance we should expect that a market crash is more likely to be the result of one or a few large returns, as opposed to many smaller ones. If this is true, then an observed deviation from the fundamental return per unit period of an asset could be deemed as the development of a bubble and could eventually lead to a crash. We explore these possibilities in the next chapters, while here we present all the conditions that would make such an approach valid.

As the whole of the previous discussion presupposes that the large deviations theory is to be applied to power law distributions, we should investigate if a power law can describe sufficiently a particular dataset. We therefore comment on the appearance of power laws in finance and then focus on the dataset that was used for this study, examining how could well it could be fit by a power-law. We also develop an equivalent to the well-known Gini coefficient for stock returns as a means of measuring the inequality of returns in large scales and test it on the particular dataset.

4.1 Power laws in finance

Our interest in power laws stems from the fact that they seem to be ubiquitous in nature, appearing in plenty of physical, economic and social systems [21]. The latter are examples of complex systems, which arise from the collaborative action of a large number of mutually interacting parts, leading to some distinct macroscopical behaviour, in this case a power law. A celebrated example comes by Pareto, who discovered that the extreme distribution of income and wealth can be described by a power law.

In recent decades many properties of financial markets have been studied and were found to be approximated by power laws. This is particularly true for high frequency economic variables, while it appears to be an endogenous property of markets [22]. Examples include the distribution of trading volume of the largest stocks at several stock markets [23] and the autocorrelation of signs of trading orders [22].

The vast majority of research on the subject has aimed at understanding the distribution of empirical returns of financial time series. Here we will consider log-returns, defined as

$$r = log(pr(t + \Delta t)) - log(pr(t)) \tag{4.1}$$

where $pr(t)$ refers to the stock market index (it could also refer to the transaction price of a particular stock, etc.) at a certain time. Log-returns are generally believed to obey a power law, although this description is adequate only for the tail of the distributions. The main body of the distribution is usually better described by a log-normal or exponential distribution. In particular a stretched-exponential distribution is found to be better over the whole quantile range, but a power law suffices for the upper quantiles [3]. Depending on what are the lower quantiles of the distributions of returns that are taken into consideration, the tail index μ of the cumulative distribution seems to take two different values, either $\mu = 1.4 - 1.7$ or $\mu \simeq 3$ [3]. This suggests that the tails of the returns distribution are certainly heavier than the tails of a Gaussian or of an exponential distribution.

All of the above indicate that the application of the large deviations theory on power laws is a tricky task when it comes to returns of financial time series. Returns that are not sufficiently described by a power-law (usually that is small returns) should be filtered out in order to achieve better results.

4.2 Data used

For the needs of this thesis we have studied the S&P500 (Standard & Poors 500) index and in particular the E-Mini S&P500 index, which is a stock market index for futures contracts. The notional value of one contract is 50 times the value of the S&P500 stock index. Our data span from January 1, 1998 to June 15, 2012 and this particular dataset was chosen because it has been studied numerous times and a power law behaviour at the tales is now well-established.

We have chosen to study 5-min and 60-min returns. Although one could suggest that studying high-frequency (e.g. 5-min) returns and then just performing 12 convolutions of the pdf of returns at the 5-min time scale should give us the pdf of the 60-min returns, this is not true at it leads to less fat tails than what is actually observed. This happens because the returns are not independent and identically distributed, but are subject to volatility correlations [3].

One usually does not expect to observe dissimilarities in the distribution of positive and negative returns [24], although there are certain phenomena which are mainly linked only with one of the two categories of returns (such as the leverage effect, which is discussed later in greater detail). Here, we have chosen to treat negative and positive returns separately, thus we have studied 313475 positive 5-min returns and 268239 negative 5-min returns, yielding a total of 581714 5-min returns. Correspondingly there are 31133 positive 60-min returns, 29879 negative 60-min returns and a total of 61012 60-min returns. In the next two sections of this chapter the dataset is subjected to a number of tests with respect to large deviations.

4.3 Power law fitting to the data

A first, rather rough way to determine if an empirical distribution follows a power law is to sketch its complementary cumulative distribution function (CCDF) in a log-log diagram. If part of the distribution falls on a straight line, then one can assert that a power law holds for that part, with the scaling parameter μ given by the absolute slope. This crude method is usually not sufficient, but we can still obtain an estimate of which returns are of interest, as one can see in figure (4.1). It is only returns above a certain threshold.

We next attempt to determine the threshold, as well as the scaling parameter that corresponds to the slope of the straight line in the diagram. It should be noted that our intention is not to strictly determine if a power law is the best distribution to fit the data, but to approach the data from different viewpoints and then reason on what part of it should be tested for heavy-tail distributions large deviations. This will become clearer in next chapters.

To calculate the scaling parameter we employ the maximum likelihood estimation method that was described in the second chapter. It is for the maximum likelihood values that the distance between the true model and the assumed model reaches its minimum [25]. We will examine three cases: 1. No minimum threshold r_{min}, i.e. all returns are considered to be power law

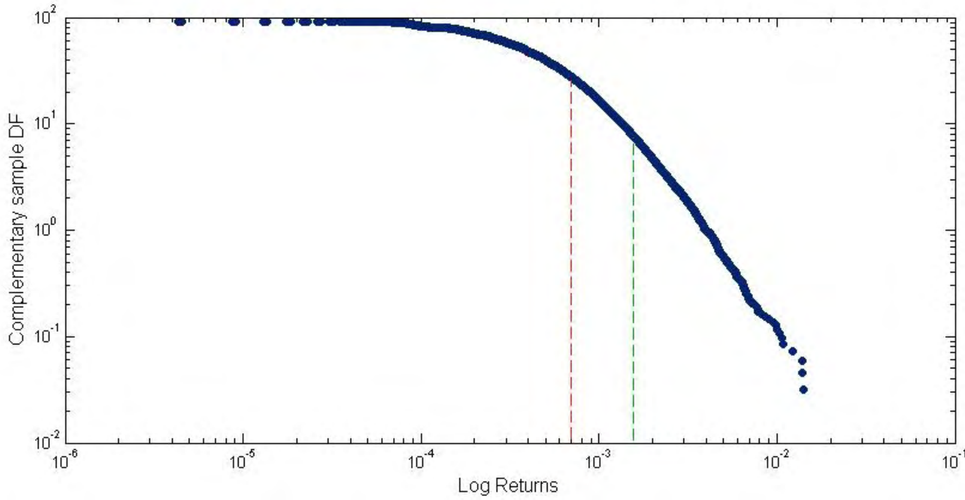

Figure 4.1: Complementary cumulative distribution for the E-Mini S&P500 5-min positive returns. The red line separates the 95% smaller returns from the 5% largest ones and the green line corresponds to the threshold determined by the method proposed by Clauset et al.

distributed, 2. we only consider returns at the 95% quantile (since returns above the 95% quantile are sufficiently described by a power law [3]), 3. the minimum threshold is determined according to the method developed by Clauset et al. [17] described at section 3.2.2. The results are presented in the following table:

	Positive Returns		Negative Returns	
	5-min	60-min	5-min	60-min
No r_{min}	1.26	1.28	1.26	1.28
95% quantile	3.71	3.63	3.64	3.73
Clauset	4.02	3.77	4.26	3.91

Table 4.1: Estimation of the scaling parameter μ with the three different methods

The scaling parameters that are estimated for no r_{min} describe a distribution with a much heavier tail than the one that is actually observed and can be easily discarded. The estimation given by the Clauset method determines r_{min} too strictly and fits a power law only to the ∼1% larger returns, thus neglecting a large part of the straight line of the CCDF. This can be a problem when there are not enough data points, therefore for the rest of this study we will adopt the scaled parameter values that are found for the 95% quantile, which are slightly smaller[1].

4.4 Gini coefficient for log-returns

In our attempt to assess the impact that the largest return(s) have when large deviations from the fundamental return occur, it is worthwhile to first examine if such large returns exist. As a means to do so we introduce an equivalent to the well-known Gini coefficient, which is a measure of statistical dispersion; It measures the inequality among values of a frequency distribution, in our case among returns. Schematically it is the area enclosed by the line of equality and the Lorentz curve.

Returns r_i are arranged in non-decreasing order ($r_i \leq r_{i+1}$). The Gini coefficient is then defined as [26]

$$G = \frac{2\sum_{i=1}^{N} i r_i}{N\sum_{i=1}^{N} r_i} - \frac{N+1}{N} \tag{4.2}$$

where the sum is performed over all N returns. The value of the Gini coefficient is bounded by the values 0 and 1 and as one can easily observe lower values correspond to greater equality, meaning that the total growth (decline) of the index is more democratically distributed among positive (negative) returns.

A first step is to calculate the Gini coefficient for the whole timespan of 15 years that we examine, as well as for each of the individual years. This way we can find an estimate of the relative dispersion of the returns and also see whether the coefficient remains constant, or if it fluctuates depending on the state of the market. We get the following extremal values:

	Positive Returns		Negative Returns	
	5-min	60-min	5-min	60-min
Maximum annual value	0.538	0.575	0.534	0.569
Minimum annual value	0.367	0.464	0.368	0.474
15-year value	0.469	0.527	0.472	0.536

Table 4.2: Gini coefficient. Summary of the calculations for yearly distribution of returns.

[1]For a discussion on the continuous growth of the Hill estimator please refer to Malevergne and Sornette [25].

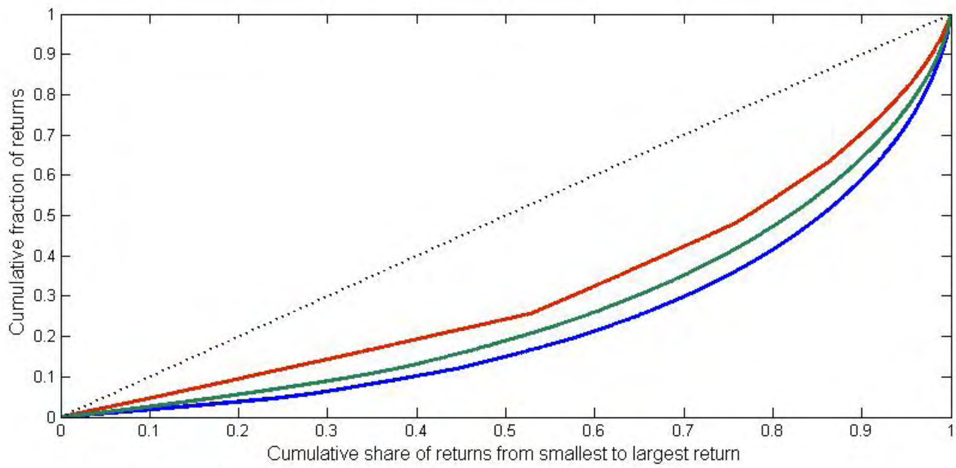

Figure 4.2: The Gini coefficient(here for 5-min positive returns) corresponds to the area between the line of perfect equality and the Lorenz curve. Blue line: Lorentz curve for the maximum annual Gini coefficient value, red line: Lorentz curve for the minimum annual Gini coefficient value, green line: Lorentz curve for the total span of 15 years

Here we should remember that we are interested in power-law distributed returns and this is not the case for small returns. If we are to test the validity of the large deviations theory we should concentrate mainly on the tails. Therefore we will also consider the Gini coefficient for the 95% quantile of the returns, which yields the results:

	Positive Returns		Negative Returns	
	5-min	60-min	5-min	60-min
15-year value	0.214	0.217	0.203	0.188

Table 4.3: Gini coefficient for the 95% quantile of returns

Two comments should be made here: 1. The Gini coefficient is larger for the whole body of returns, than for those at the tail of the distribution, hinting towards the well-known fact (as can be confirmed by looking at figure (4.1) (CCDF)) that small returns contribute little to the average index growth. 2. The relative equality in size among larger returns suggests that -at least at large scale- there are no outliers with a large enough magnitude, so as to dominate above others. This somewhat disheartening remark obliges us to zoom into smaller scales (than at the whole distribution of returns) and this is what we will do in the next chapters.

Before doing so, it is worth calculating the Gini coefficient of a power-law with an index index $\mu = 3.71$ which we found in the previous section for 5-min positive returns. Power-law distributed random numbers can be generated from a uniform distribution of real random numbers r through [17]

$$r = r_{min}(1-x)^{\frac{-1}{\mu-1}} \tag{4.3}$$

where $r_{min} = 0.000288$. We generated 5×10^6 i.i.d. power law distributed numbers and found that the Gini coefficient in this case is $G = 0.226$, very close to the value $G = 0.214$ for the 95% upper 5-min positive returns. After plotting the corresponding Lorentz curves (since the two distributions may have similar Gini coefficients but still different significantly) we see that they almost coincide, even for the largest returns, despite the fact that returns are correlated and not iid. The conclusion is that if we consider only returns in high quantiles, then the conditions for testing the large deviations theory on power laws are valid.

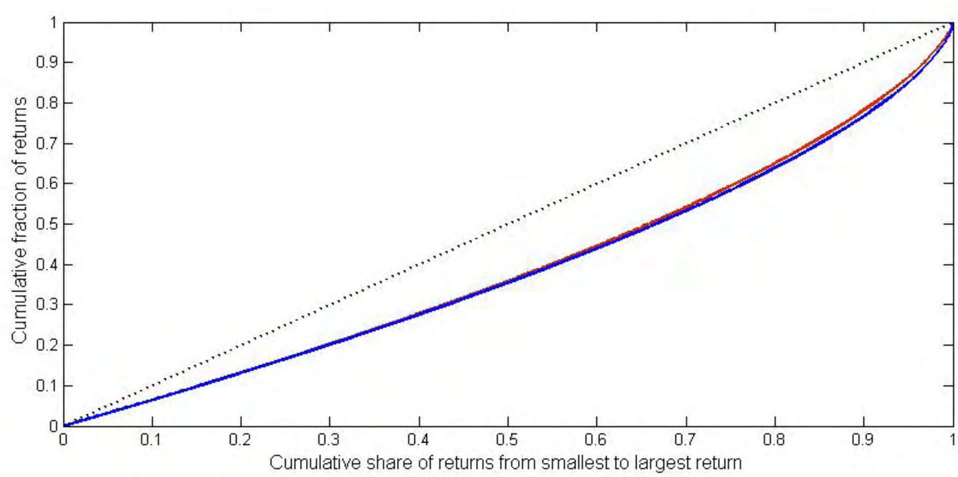

Figure 4.3: Comparison of the Lorentz curves for (blue) power law-distributed synthetic data and for (red) the 95% quantile of 5-min positive returns

Chapter 5

Deviations of stock market returns

5.1 Empirical and fundamental rates of return

Financial markets and in particular stock markets offer fertile ground to test the large deviations theory that was discussed in the first chapter. Suppose that an investor makes an estimate about the average growth of a stock market index $\langle r \rangle$, of say 5% annual growth. Instead she notices that the realized rate of return $\overline{r}$ greatly surpasses this value, so for instance a 10% annual growth is observed. As discussed previously, returns approximately follow a power law distribution, thus by applying the reasoning of the first chapter she should expect that this abnormal growth rate is probably due to one or only a few extremely large gains. The same also applies for negative returns, so one should expect that a crash will more likely be the result of only one (or a few) large loss(es).

Let's examine more carefully a few details in the above stream of arguments. Firstly, as we already know the whole body of the distribution of returns cannot be described by a power law, which does so sufficiently only for the tail. So when we observe a large deviation of the kind that was just described, how can one be sure that a power law is there, allowing the application of the large deviations theory? The obvious solution would be to discard all returns that do not fall in the tail, the ones that are below the 95% quantile in accordance with what we accepted earlier. But this could lead to strange situations, where each of the accepted last N returns (see for instance equation (2.12)) at a specific time frame is too far apart from the others and it is doubtful whether we can associate a large growth with a large return that occurred much older in time.

Secondly we tacitly accepted that a fundamental, or average stock market return exists, or at least that it is easy for an investor to approximately calculate it. In practice, defining a fundamental return is a multifold problem, that reflects the difficulty of using some average of past returns to make long term return forecasts. The simplest approach is to accept as fundamental return the empirical return on stocks corresponding to a very long period, although it is now accepted that stock returns should not be expected to remain constant over time [27]. Problems with this approach include that the returns are very noisy and the difficulty of choosing between the geometric average and the arithmetic average [28].

The above remarks are valuable when considering large deviations of stock market returns. We say that a large deviation is observed at a specific time t when the empirical average of the last N returns is (much) larger than the expected average return, or

$$\overline{r} > \langle r \rangle$$

5.1.1 Values for $\langle r \rangle$ and $\overline{r}$

- Expected average return $\langle r \rangle$

We use three methods to estimate $\langle r \rangle$ for the E-Mini S&P index. The process is repeated for both positive and negative (with inverted sign) 5-min and 60-min returns:

1. We consider $\langle r \rangle$ to be the arithmetic average for the 15-years period that is studied. Returns that are not in the tail of the distribution are therefore also taken into consideration.

2. We equate $\langle r \rangle$ with the mean of the power law, whose pdf is determined by the values of μ and x_{min} that were found in the previous chapter for the 95% quantile, or

$$\langle r \rangle = \int_{r_{min}}^{\infty} \mu \Big(\frac{r}{r_{min}} \Big)^{-\mu} = \frac{\mu r_{min}}{\mu - 1},$$ [1]

3. We consider $\langle r \rangle$ on a rolling basis, as a simple moving average spanning the last 1000 returns. This way a deviation is defined as a departure from the recently observed trend.

[1] If $\mu < 1$ then we have to employ the method that was described in chapter 2.3.1

The first method has the advantage of providing conclusions about the whole range of results, which can be an asset when one does not know if recent returns can be considered to be power-law distributed, an approach that in practice is indeed not feasible. If we want to be strict we have to consider only deviations above the mean of the power-law distribution that describes returns in higher quantiles. Thus, the second method allows us to focus on only large deviations and examine potential differences in behaviour at this regime. A problem with the above two approaches is that they consider the distribution of returns to remain steady as time passes. In fact financial indexes rise faster than exponential [29] and the average return will be considerably different when considering two periods of the same length that are e.g. 10 years apart. To deal with that problem we employ the last method. Here, observing that most of the last 1000 returns were relatively large means that we can assume that we are close to the power law regime.

- Empirical average of the last N return $\langle r \rangle$

The empirical average is calculated as a simple moving average for the last $N = 50$, or $N = 100$ returns. As N becomes larger, finding a large return far away in the tail that will be almost solely responsible for the large deviation from the fundamental price becomes all the more unlikely and as we will shortly encounter this is already visible for the above values of N.

5.2 Examining large deviations

When discussing the Gini coefficient for stock market returns we found that they are fairly equally distributed, when considering all returns for the 15-years examined period. As a benchmark we note that the largest 1% 5-min positive returns amount to about 5.4% of all returns, while when considering only the 95% quantile the largest 1% returns correspond to 4.4% of all returns at this quantile. This hints towards the fact that at larger scales, the market rise or fall is not dominated by only a few returns. We now turn to smaller scales.

We begin by calculating the quantities $R_1 = \frac{r_{max}}{N\overline{r}}$ and $R_2 = \frac{r_{max}+r_{max-1}}{N\overline{r}}$ (with r_{max-1} being the second largest among the last N returns) on moving time windows, for $\overline{r} > \langle r \rangle$. We then plot R_1 and R_2 against $\frac{\overline{r}-\langle r \rangle}{\langle r \rangle}$. This way we will see if large deviations occur in the way that they are described in the second chapter. Here the results for 5-min positive returns are presented, while the rest of the plots are found in the appendix. Only datapoints that are at least $N/2$ time steps apart are presented, to avoid correlation effects.

A first look at the plots shows that there is indeed a difference at smaller scales, where in most cases the largest return is responsible for a significant percentage of the observed deviation. When considering the largest among the 100 returns that lead to a deviation (figures 5.4-5.6), we can directly compare it with the largest 1% of all 5-min positive returns and find it to be at least as large and usually larger (more than 10% of the total sum of returns). For $N = 50$ (figures 5.1-5.3) the largest return accounts for an even larger percentage of the deviation and the explanation is simple, in that the latter has to be only half as big as the one for 100 returns to amount for the same percentage of the deviation, thus it need not be at the far tail of the power law distribution. Apart from this remark the plots for $N = 50$ and $N = 100$ are qualitatively similar.

Small deviations from the mean $\langle r \rangle$ are quite common, while large deviations -depending on how the mean $\langle r \rangle$ was calculated- can reach up to 9 times the size of the expected average return. We observe that in general most deviations are characterized by $R_1 \sim 0.1$, but in numerous cases R_1 is much bigger, with the largest return being almost solely responsible for the deviation. These returns-outliers are more common close to small values of $\frac{\overline{r}-\langle r \rangle}{\langle r \rangle}$ and the trend is that they become less important as the deviation becomes larger.

This means that our theory is only partially verified by the data, since large outliers become more uncommon at larger deviations, a behaviour that is at odds with what we expected in the first place. However, a look at figure (5.2), where we only consider deviations above the mean of the power-law distribution that describes returns in the 95% quantile shows that the trend for decreasing magnitude of R_1 is no longer present, an interesting remark since as we have argued, it is these returns that can be better approximated by a power law. But is there a direct way to test the theory in this regime? We recall equation (2.10) from where we get

$$r_{max} = N(\overline{r} - \langle r \rangle) + \langle r \rangle,$$

and consequently

$$R_1 = \frac{r_{max}}{N\overline{r}} = \frac{N(\overline{r} - \langle r \rangle) + \langle r \rangle}{N\overline{r}} = 1 - \frac{\langle r \rangle}{\overline{r}} + \frac{\langle r \rangle}{N\overline{r}} =$$

$$1 + \frac{\langle r \rangle(1/N - 1)}{\overline{r}} = 1 + \frac{1/N - 1}{\frac{\overline{r}-\langle r \rangle}{\langle r \rangle} + 1}, \tag{5.1}$$

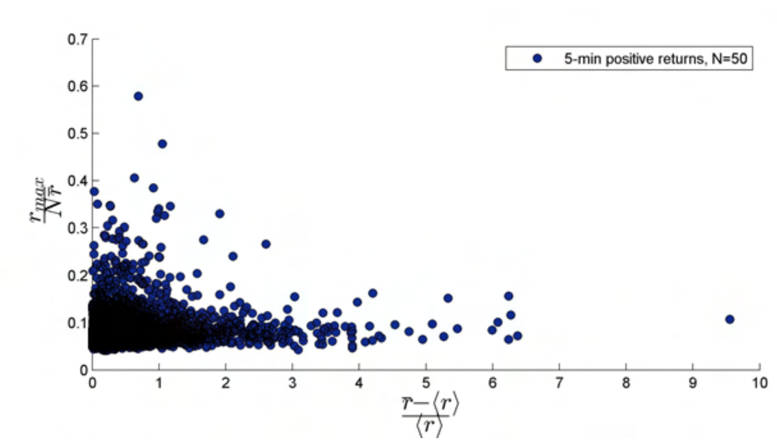

Figure 5.1: The figure shows the magnitude of the largest return among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the whole body* of returns.

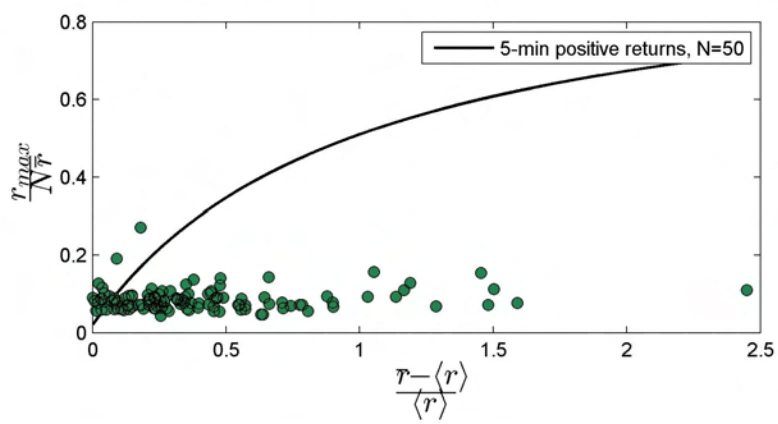

Figure 5.2: The figure shows the magnitude of the largest return among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated for the *power law distribution corresponding to the* 95% *quantile* of returns. The black line corresponds to the function $R_1(\frac{\overline{r}-\langle r\rangle}{\langle r\rangle}) = 1 - \frac{49}{50}/(\frac{\overline{r}-\langle r\rangle}{\langle r\rangle} + 1)$.

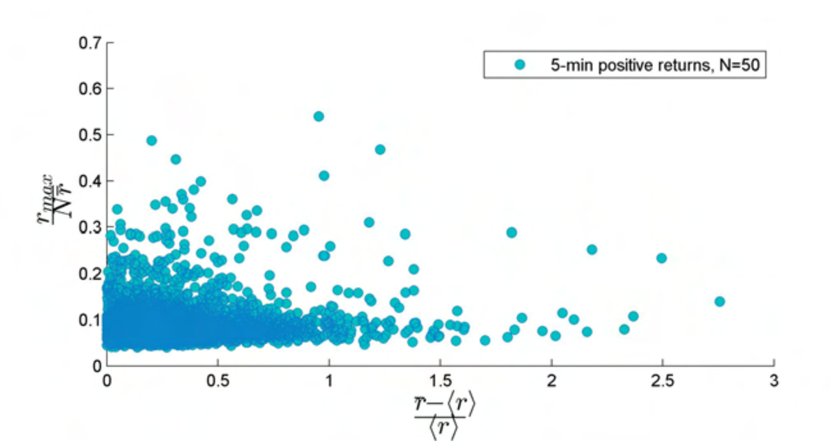

Figure 5.3: The figure shows the magnitude of the largest return among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *average over the last 1000* returns.

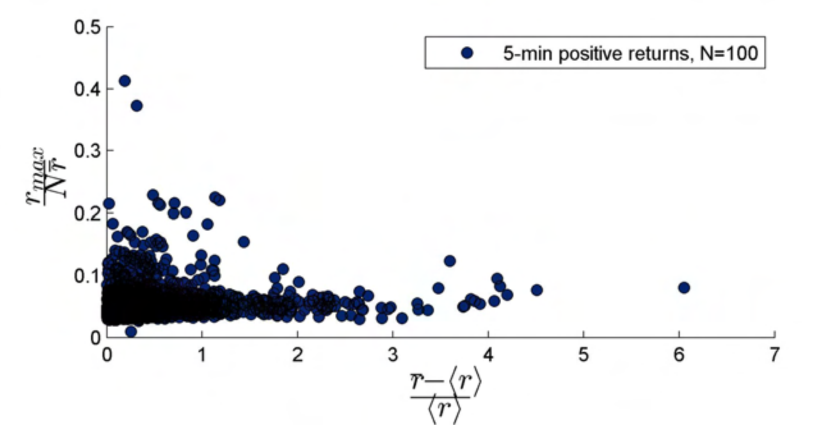

Figure 5.4: The figure shows the magnitude of the largest return among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *average over the whole body* of returns.

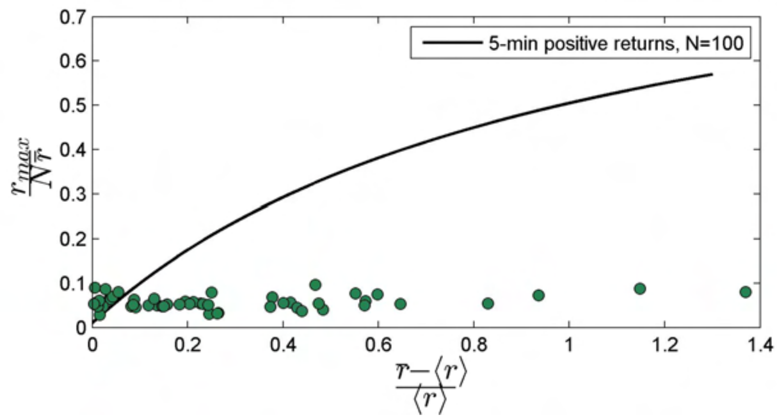

Figure 5.5: The figure shows the magnitude of the largest return among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated for the *power law distribution corresponding to the* 95% *quantile* of returns. The black line corresponds to the function $R_1(\frac{\bar{r}-\langle r\rangle}{\langle r\rangle}) = 1 - \frac{99}{100}/(\frac{\bar{r}-\langle r\rangle}{\langle r\rangle} + 1)$.

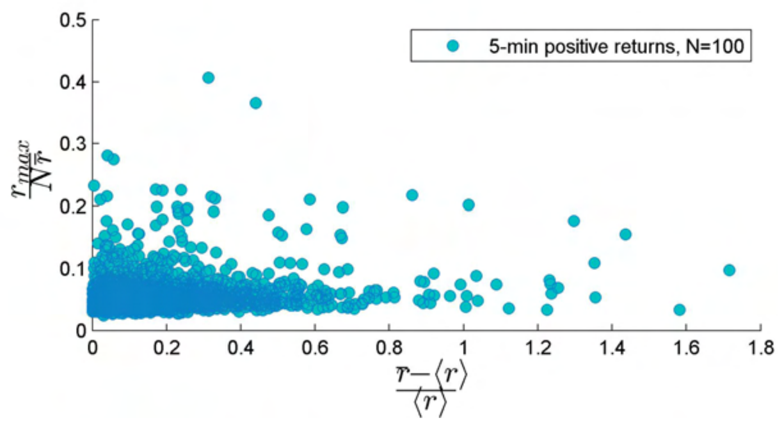

Figure 5.6: The figure shows the magnitude of the largest return among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the last 1000* returns.

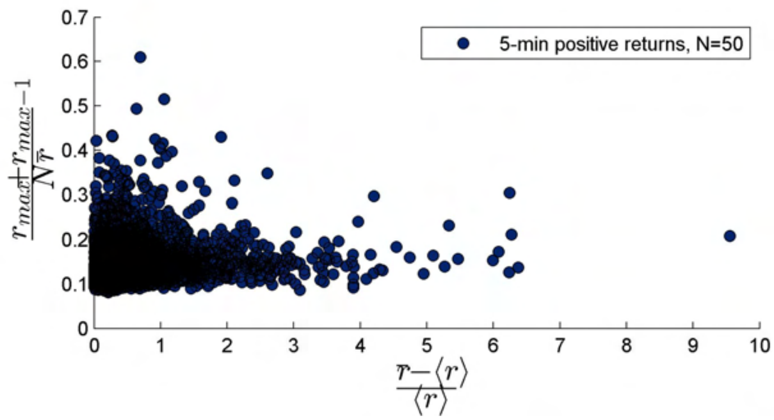

Figure 5.7: The figure shows the magnitude of the *two* largest returns among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the whole body* of returns.

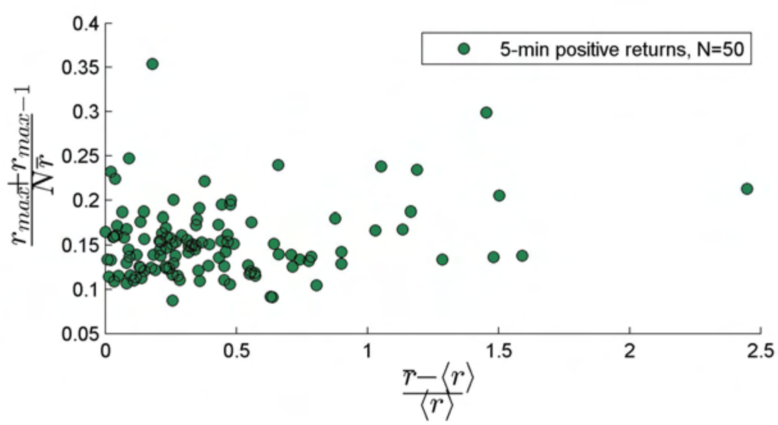

Figure 5.8: The figure shows the magnitude of the *two* largest returns among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated for the *power law distribution corresponding to the* 95% *quantile* of returns.

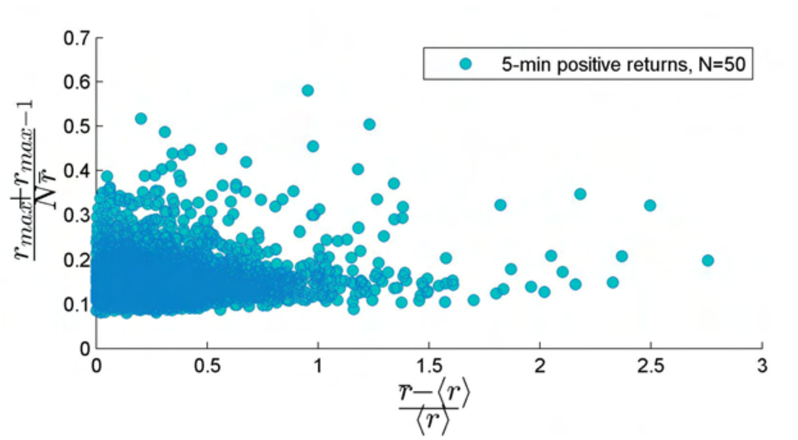

Figure 5.9: The figure shows the magnitude of the *two* largest returns among the last *50* returns versus the deviation of the last *50 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *average over the last 1000* returns.

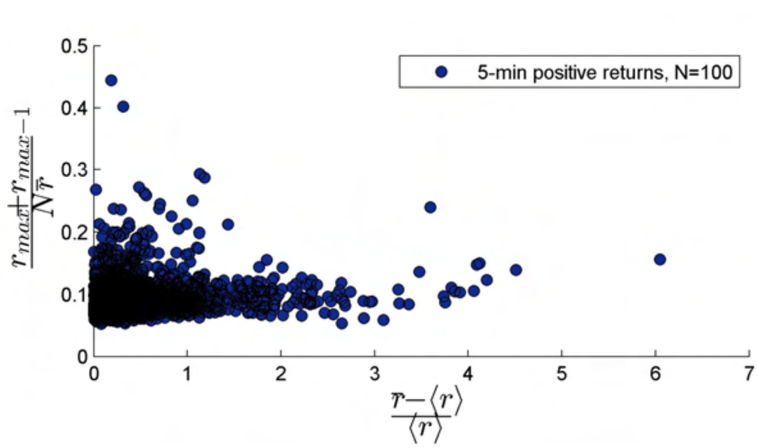

Figure 5.10: The figure shows the magnitude of the *two* largest returns among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *average over the whole body* of returns.

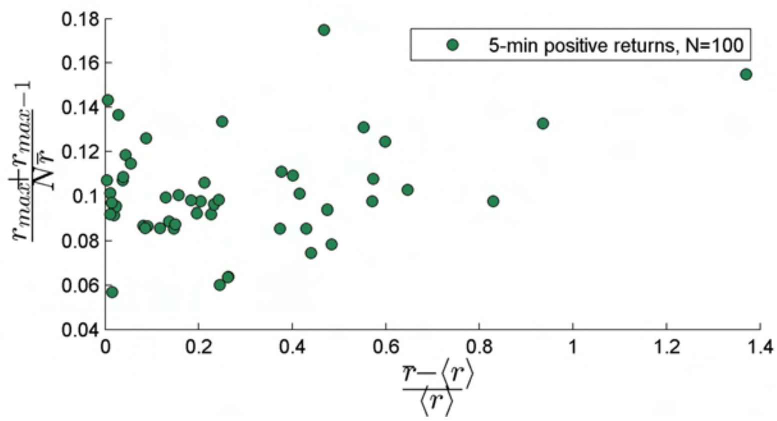

Figure 5.11: The figure shows the magnitude of the *two* largest returns among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated for the *power law distribution corresponding to the* 95% *quantile* of returns.

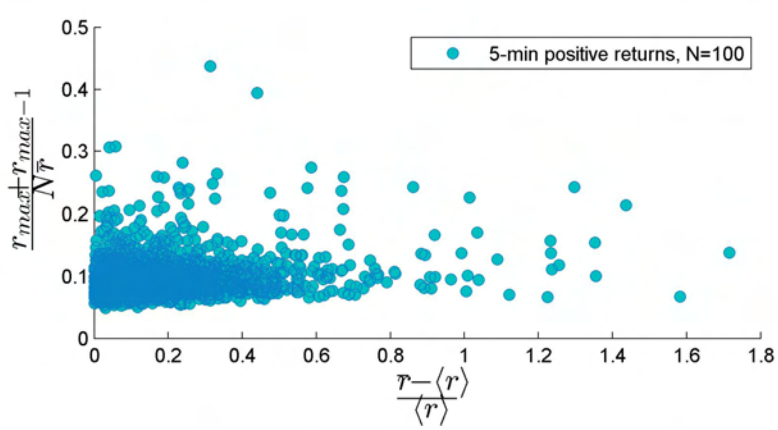

Figure 5.12: The figure shows the magnitude of the *two* largest returns among the last *100* returns versus the deviation of the last *100 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *average over the last 1000* returns.

suggesting that R_1 should be an increasing function of the relative deviation. Thus figures (5.2) and (5.5) also depict this function, for $N = 50$ and $N = 100$ respectively. Despite the existence of a noisy upper trend in the data, the function quickly rises above them and can be used to approximate them only for very small deviations. A last comment for this method, is that we notice R_1 to be relatively large, with $R_1 \sim 0.1 - 0.3$ when $N = 50$.

In the third method (figure 5.3) $\langle r \rangle$ is calculated on a rolling basis and this is probably the best approximation, since really old or future returns do not affect the average. Interestingly with such a configuration the largest deviations correspond to a realized rate of return that is at most four times as big as the expected average, or $\frac{\overline{r}-\langle r \rangle}{\langle r \rangle} > 3$. At such a setting outliers are also present, again with a small tendency to become less important as the deviation becomes larger.

Plots (5.7-5.12) show the relative magnitude of the *two* largest returns among N. A comparison between e.g. figures (5.1) and (5.7) shows that the second largest return is usually not sufficiently large to make R_2 much larger than R_1 and is a further confirmation that in general (and especially when the sample is small, e.g. $N = 50$ as opposed to $N = 100$) the largest return prevails over all others.

5.3 Large returns as outliers

As we have shown by now, power law distributions on their own do not usually allow for extreme events. However, the previous section shows that when stock-market returns at smaller scales are examined, we may notice that a large gain/loss may be responsible for the average index behaviour, sometimes making up to as high as half of the total gain/loss.

This is only one of the many cases, where a power-law distribution is accompanied by the existence of one (or a few) rare extreme event, that dominates over all others, subsequently leading to an alteration of the system's properties. There have been two approaches to understand the nature of such events. The first one holds the view that all small, large and extreme events belong to the same population and are governed by the same mechanism. This view suggests that extreme events, found at the very end of the tail of power-law distributions, are virtually unpredictable, as there is no way to diagnose early what separates them from their smaller siblings. This line of arguments can be traced back to the notion of "Self-organized criticality"[2].

[2]For example Bak and Paczuski [30] after studying the collapsing of sandpiles that

Recently, another approach has appeared suggesting that extreme events (termed Dragon-Kings) may actually be statistically and mechanistically different from the rest of the distribution [11]. The consequences of such an assumption are direct, as rare and catastrophic events (like abrupt changes in weather regimes or in our case, financial crashes) may be within the reach of predictability. Empirical evidence confirms that in many systems Dragon-Kings coexist with power-law distributions, yet the real challenge remains to understand the mechanisms that produce them and to develop tools that can detect them (see also [31]).

By applying the large deviations theory on the power law that describes them, we see that a few returns can significantly move quantities like the mean of the distribution upwards, something which is partially confirmed in the previous section. In particular, by claiming that a few gains/losses should be responsible for large deviations, we implicitly suggest that these returns are special and should be considered as outliers. Therefore, we propose that the few extreme returns that we noticed are Dragon-Kings and as such, they are not described by the same power law as the rest of the distribution, but rather by equation (2.12).

$$\Lambda = \frac{\mathcal{L}(DK|Data)}{\mathcal{L}(PL|Data)}. \tag{5.2}$$

We present the plots for 5-min positive returns, since the plots for other types of returns are qualitatively similar. The colorbar shows the relative magnitude of the largest return (among the last N) as calculated in the previous section. It is evident that for small deviations the likelihood ratio is in many cases $\Lambda > 1$, suggesting that they should be the result of the appearance of a single large return. This reflects the form of equation (2.12) and suggests that -according to our previous discussion- a relatively small, largest return can alone result to a small deviation from the mean. We remember (e.g. figure (5.1) that indeed it was for small deviations that the higher values of R_1 are observed.

Since the two cases are related, we can apply the likelihood ratio method that was presented earlier[3] as a way to examine if a large return should be

follows a power law $P(s) \propto s^{-\tau}$, (where s is the number of sand grains at which a collapse is observed) argue that long-term predictions are not possible and that avalanches may occur at any time, being an unavoidable and intrinsic part of the sandpile dynamics.

[3]We will find that the two likelihoods may be many orders of magnitude different and given that the two models are related, the likelihood ratio test suffices for our purposes. A more rigorous approach would require to use the Akaike information criterion or similar, but here it is not deemed necessary.

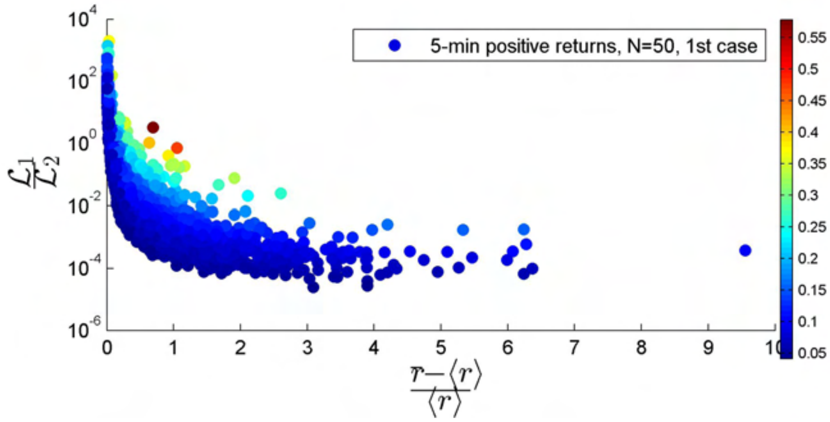

Figure 5.13: Likelihood ratio against the deviation of the last *50 5-min positive* returns from the mean with $\langle r \rangle$. The mean is calculated as the *average over the whole body* of returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *50*).

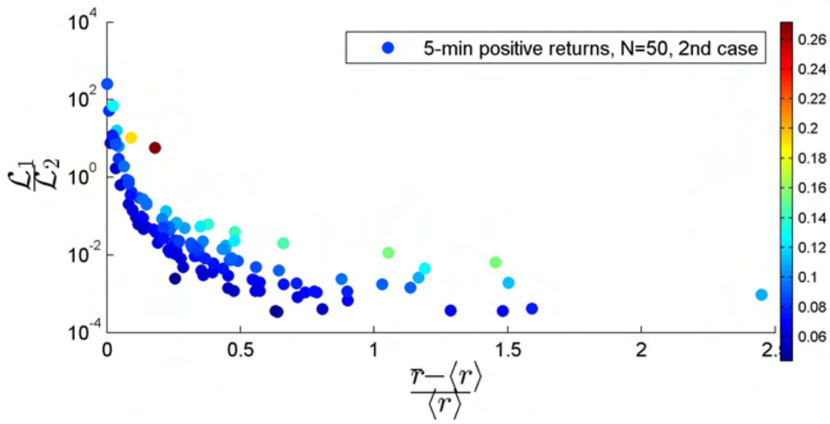

Figure 5.14: Likelihood ratio against the deviation of the last *50 5-min positive* returns from the mean $\langle r \rangle$. The mean is calculated as the *power law distribution corresponding to the* 95% *quantile* of returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *50*).

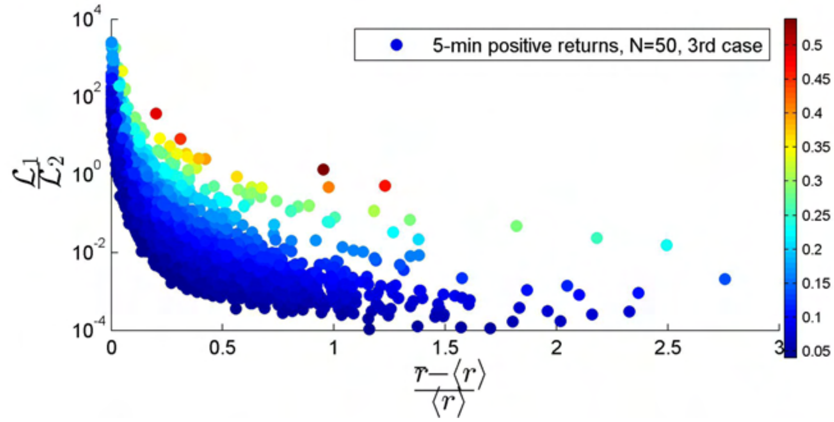

Figure 5.15: Likelihood ratio against the deviation of the last *50 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the last 1000* returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *50*).

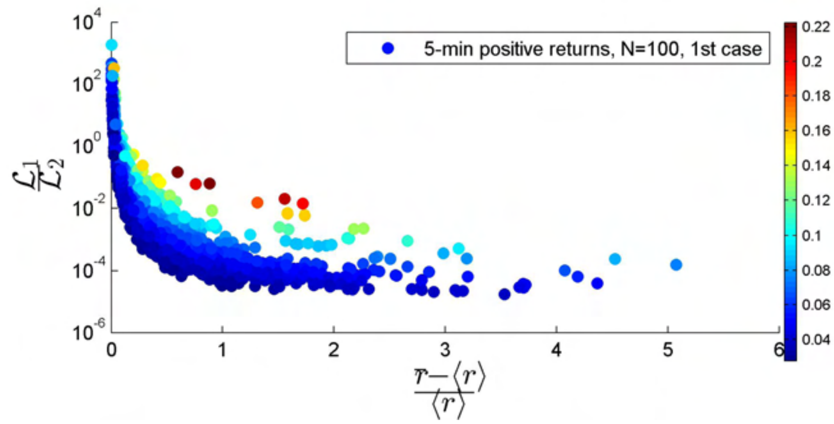

Figure 5.16: Likelihood ratio against the deviation of the last *100 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the whole body* of returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *100*).

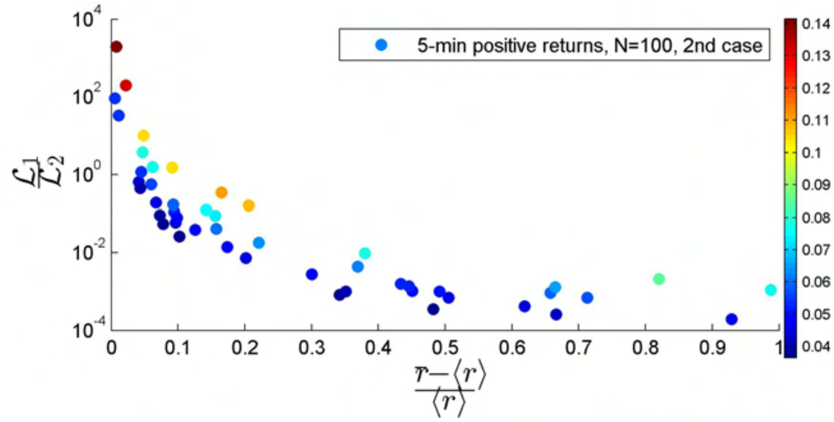

Figure 5.17: Likelihood ratio against the deviation of the last *100 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *power law distribution corresponding to the* 95% *quantile* of returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *100*).

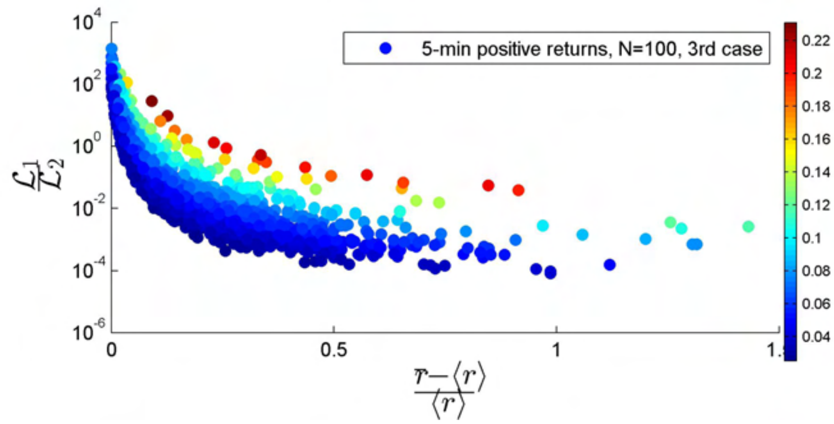

Figure 5.18: Likelihood ratio against the deviation of the last *100 5-min positive* returns from the mean $\langle r\rangle$. The mean is calculated as the *average over the last 1000* returns. The colorbar shows R_1 (the relative magnitude of the largest return among the last *100*).

considered an outlier among its neighbours, or if it comes from the same distribution as the rest of the returns. If the ratio is $\Lambda > 1$ then it is more likely that the largest return is a Dragon-King:

If we want to be more strict and study only the largest deviations (i.e. $\frac{\bar{r}-\langle r \rangle}{\langle r \rangle} > 0.5$), then Dragon-Kings are more clearly identifiable. The method we suggest depends highly on how $\langle r \rangle$ was calculated, but in general suggests that for large deviations a return could be classified as a Dragon King when $R_1 > 0.2 - 0.4$ (see figures 5.13-5.15). As we move to larger deviations the likelihood of observing a Dragon-King decreases faster than the one of having just a regular large return (that comes from a power law) and this makes it more difficult to observe one.

What is worth noting, is the formation of two different clusters among deviations, that is observable in most of the plots (e.g. figures 5.15 and 5.18). Deviations with a relatively large ratio Λ -that is close to 1 or larger- group together, as opposed to those having $\Lambda \ll 1$ which form a different cluster. This fact cannot be omitted as it suggests the existence of two different market regimes. However, an unambiguous distinction is not possible.

Having the knowledge of the two last sections we may now conclude that the assumption of outlier returns leading to large deviations is partially confirmed by the data, where such returns are present even if a distinction from normal returns is not always easy. In particular, this behaviour occurs especially when $\frac{\bar{r}-\langle r \rangle}{\langle r \rangle}$ is not big, but instead we observe extremely large deviations ($\frac{\bar{r}-\langle r \rangle}{\langle r \rangle} \gg 0.5$) to be the product of the collaborative effort of more than one returns. This could be evidence of special market behaviour (other than the power law distribution of returns) that causes the largest deviations to occur and we will refer to such possibilities at the end of the next chapter.

Chapter 6

Can large deviations be utilized?

6.1 Large deviations and crashes

We will now attempt to make use of previous discussions and calculations. In particular if the large deviations theory for power laws holds true then a large loss at the market should be expected to be concentrated on one or a few negative returns. The empirical study of the previous chapter cannot back up that assumption[1] when the observed mean is far from the expected average rate of return, but it does so for deviations near it. Thus, it is worth investigating if smaller crashes can be related to previous deviations.

Another way of looking at large positive deviations is as if it represents a bubble under development. The fundamental return is of course only known approximately and captured by some distribution, so one can suppose that the realized returns are just fluctuating, but as time goes by it becomes more and more clear that they are deviating upwards[2] and it is probable that investors and other parties of interest will believe that a bubble is being formed. Therefore, the combination that i) a bubble has begun developing and ii) that the agents realize the deviation from the expected fundamental value and attribute it to a bubble, yields a high probability that they will try to sell their assets before the bubble crashes, which eventually leads to mechanisms such as herd-selling [34]. This is simply a manifestation of the efficient markets hypothesis [35], with the exception that investors have not

[1](see the appendix for negative returns which are of particular interest here)

[2]Several models have been proposed to explain such overpricing on behaviouristic terms [32] or other amplification mechanisms [33]

access to perfect information.

If such a causal relation exists, then an observed short-term deviation should be followed by really large negative returns and the correction will appear in the form of one or a few finite steps. This is tested in the next section. Additionally this stream of arguments could also explain why extremely big outliers were not observed in the previous section. If no crash has occurred while the bubble is small, the case may be that it may never happen, so the market enters a systemic large deviation phase. A larger correction is now less probable according to the power law distribution of returns and this mechanism could explain the existence of long lived bubbles.

6.2 Large deviations as precursors

Let us now answer the question if large deviations can be used as precursors of crashes. To this end, two methods are developed that will lead to a more or less definitive conclusion. In the following two methods we will use the Spearman's rank correlation coefficient as a measure of correlation [36]. It measures if there is a monotonic relationship between two variables and is defined as

$$\rho = \frac{\sum_{i=1}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}(x_i - \overline{x})^2 \sum_{i=1}(y_i - \overline{y})^2}}$$

where x_i,y_i are the respective ranks of each pair of observations, among other components of sets X and Y.

6.2.1 1st Method

The E-Mini S&P500 index is expected to peak at the time that the large deviation occurs, as previously it was rising and afterwards negative returns should follow. If a large positive deviation is likely to be followed by a crash, or at least a considerable decline, then it should be close to a local maximum and this should become increasingly observable for larger deviations. We consider the index value at $N+1$ time frames, with $N/2$ of them before the large deviation and $N/2$ after it and compute the large deviation's index rank among its neighbours. This is then plotted against $\frac{\overline{r}-\langle r\rangle}{\langle r\rangle}$. We also calculate the correlation between the rank and the deviations.

By observing the plots, we can say that there is no discernible pattern and the rank of the large deviation is more or less evenly allocated among all N values. There is also no dependence on N or on the magnitude of the

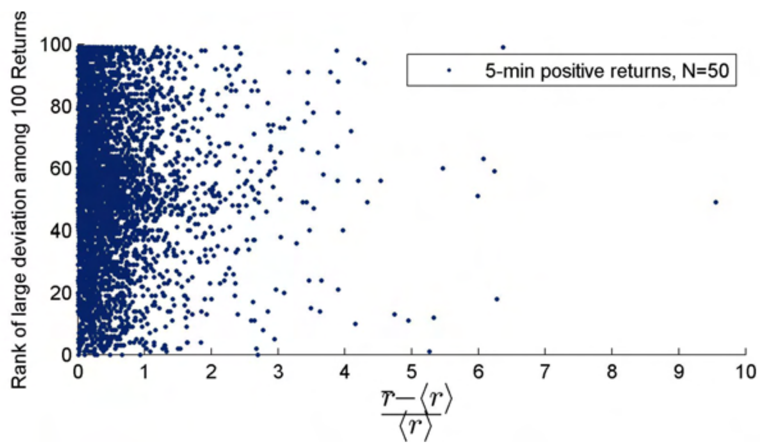

Figure 6.1: The plot depicts the rank among the *100* neighbouring returns of the return for which the deviation occurs, against the corresponding deviation. $\langle r\rangle$ is calculated as a 15-years average.

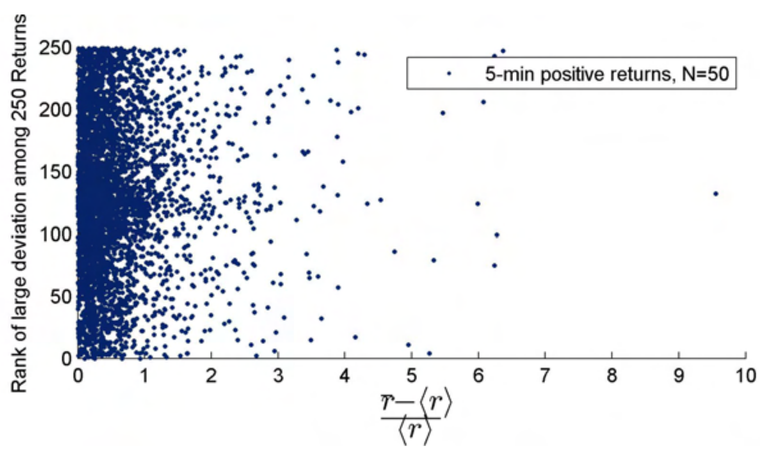

Figure 6.2: The plot depicts the rank among the *250* neighbouring returns of the return for which the deviation occurs, against the corresponding deviation. $\langle r\rangle$ is calculated as a 15-years average.

	Rank (100 returns)	Rank (250 returns)
15-years average	-0.0053	0.0116
Power-law mean for 95% quantile	-0.0058	0.0733
Average over last 1000 returns	0.0248	0.0257

Table 6.1: Correlation coefficient for the rank of the index at the time that the large deviation occurs, among 100 and 250 neighbouring returns, with respect to the corresponding deviation size. 5-min positive returns are examined. The left column describes the method that was used to estimate $\langle r \rangle$.

deviation. The correlation coefficient estimations further support that, so the first method does not give rise to believe that previous considerations were true.

6.2.2 2nd Method

If a large deviation signals the late phase of a bubble development, then some time later the index should have decreased, therefore a large negative return is to be observed after some time (if we consider the log-return of the index between the appearance of the large deviation and several time steps later, e.g. after N=50 5-min time steps). We calculate these long-term returns and then plot them against $\frac{\bar{r}-\langle r \rangle}{\langle r \rangle}$. We also compute the correlation coefficient between the returns and the deviations, expecting it to be negative (large positive deviations should be followed by negative returns).

	10 time steps	50 time steps
15-years average	0.0022	0.0264
Power-law mean for 95% quantile	0.0683	0.0497
Average over last 1000 returns	-0.0303	0.01865

Table 6.2: Correlation coefficient for the size of the return that occurs 10 and 50 5-min time steps after the large deviation, against the corresponding deviation. The left column describes the method that was used to estimate $\langle r \rangle$.

Again, the case is that the index can equiprobably either rise or fall after a large deviation, as no sign of order is found in the plots. The close-to-zero correlation coefficients also support this. We can conclude that the expectation that the market should mostly go down after a large deviation

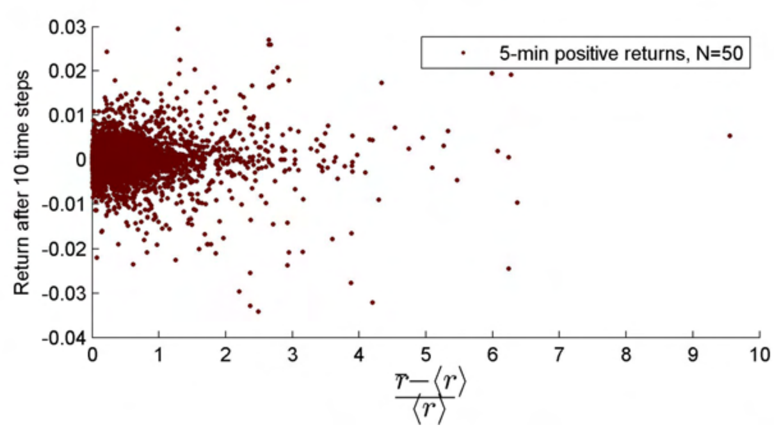

Figure 6.3: Plot of the size of the return that occurs *10* timesteps return after the large deviation has occurred, with respect to the corresponding deviation size. 5-min positive returns are examined. $\langle r\rangle$ is calculated as a 15-years average.

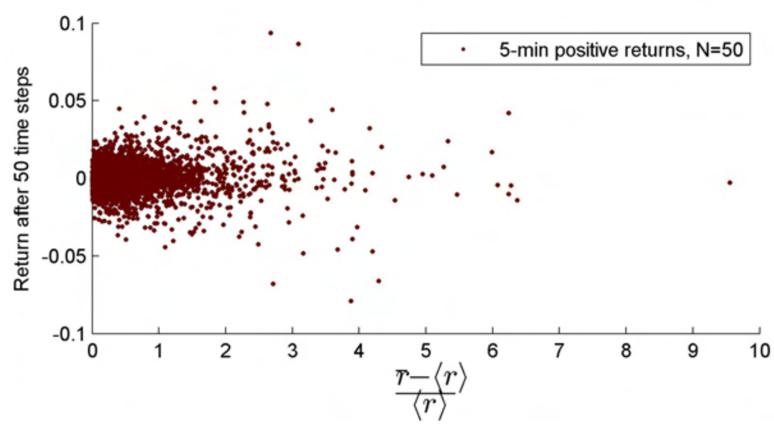

Figure 6.4: Plot of the size of the return that occurs *50* timesteps return after the large deviation has occurred, with respect to the corresponding deviation size. 5-min positive returns are examined. $\langle r\rangle$ is calculated as a 15-years average.

is not justified by the data.

6.3 Large deviations and other market properties

The failure to find support for our assumptions forces us to reconsider whether conditions at stock markets are appropriate to test the large deviations theory. And since we made no claim against it, we have accepted that returns and thus large deviations are homogeneously distributed (similar to synthetic power law data) and that there is no particular reason to expect a large deviation at a particular moment in time.

Nevertheless, plotting large deviations against time and comparing it with the market status at the time shows the opposite. As figure (6.5) shows, large deviations are heavily concentrated at periods of big drawdowns and this behaviour is the same for 5-min and 60-min positive returns, hinting towards the systemic nature of the phenomenon. What is more interesting (compare figures (6.5) and (6.6)) is the striking resemblance of the distribution of both positive and negative deviations, which indicates that there is something special about the periods that the market is going down.

This can be associated with the so-called leverage effect, or volatility asymmetry. It has been observed that the volatility of stock prices (and thus of returns) increases when its price drops [37]. Several attempts have been made to understand the leverage effect, like the retardation of price updates, as the reference price seems to reflect a moving average of the price over the past few months [38], a clear departure from our consideration that returns are i.i.d. Furthermore, it was shown that the leverage effect is stronger, although it decays faster, in indexes (and that is the focus of our study) than in individual stocks [39].

So when considered in comparison with large deviations, it is natural to assume that the leverage effect prevails over the presumed effect that we would expect to observe because of them. It is much clear now, that the increased volatility of returns when there are losses in the market leads to numerous large (both positive and negative) returns, that in turn lead to a large aggregate deviation. This deviation however has now considerably different characteristics than the ones we studied, since it is the result of many events and cannot be associated to just one large return. This could also provide an explanation as to why we found most large deviations to have a low ratio $R_1 = \frac{r_{max}}{N\bar{r}}$ (see figures (5.1-5.6)).

More empirical evidence against our original claim that one return can be enough to lead to a large deviation comes from time-varying correlations

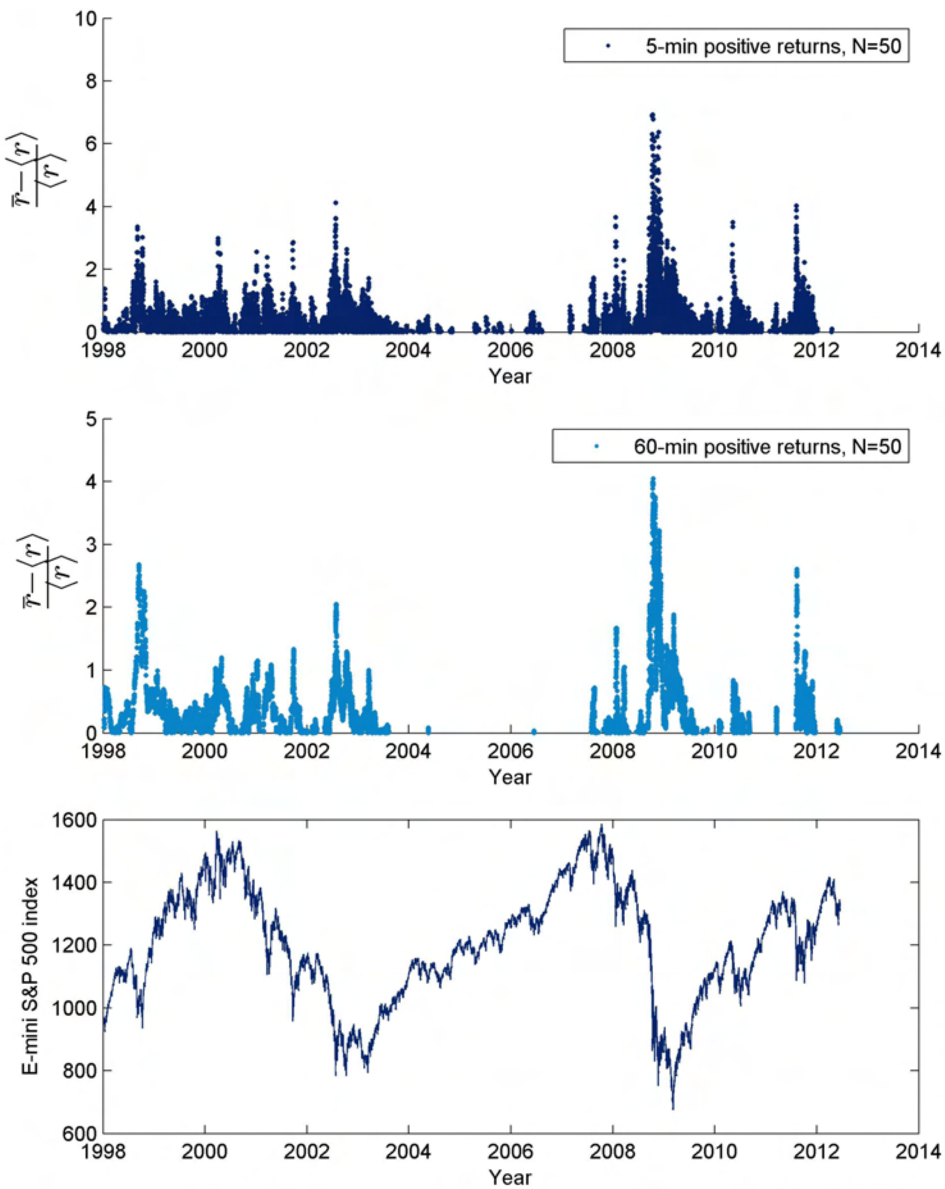

Figure 6.5: The first two graphs show the distribution of large deviations for 5-min and 60-min positive returns respectively, over the 15-years period examined. The bottom graph shows the E-Mini S&P500 index for the same period.

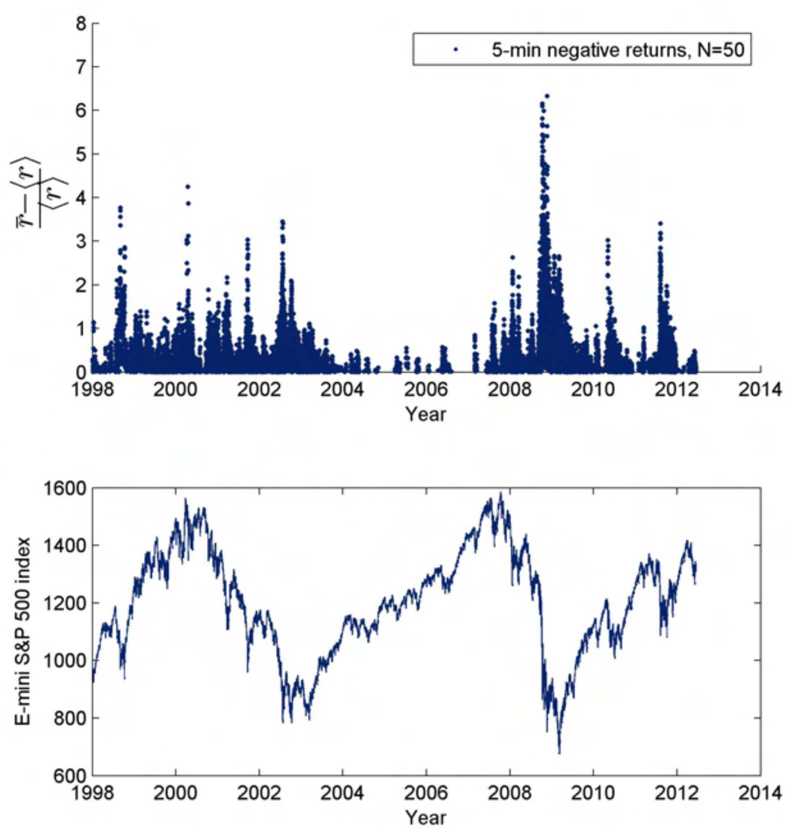

Figure 6.6: The first graph shows the distribution of large deviations for 5-min negative returns (if we treat them with inverted sign), over the 15-years period examined. The bottom graph shows the E-Mini S&P500 index for the same period.

between returns. Such correlations have been observed to increase substantially during periods of high market volatility (such as when the market over- or underperforms) [40]. This suggests that spotting a return that differs considerably from its counterparts is more unlikely than usual when $\overline{r} \gg \langle r \rangle$.

Another phenomenon that may have a counter-effect to what we expected from our analysis is the market momentum. Having mostly empirical support, is refers to the tendency for rising (falling) prices to rise (fall) even further, so that a portfolio (which could represent the index) is expected to be profitable following periods of market gains [41]. In other words, momentum suggests that when certain assets overperform (or deviate in our terminology) we should expect them to continue to rise and not to lead to large drawdowns. Although largely understood as a behavioural effect, it is still observable when agents are rational and markets are efficient [42], thus providing a substantial obstacle to the capabilities of large deviations as predecessors of crashes.

Chapter 7

Conclusions

In this paper we examined large deviations of power law distributions, traced such deviations in the distribution of stock market returns and examined both empirical facts and potential applications. In principal what we tested was whether *the concentration of large deviations for fat tailed distributions may lead to counterintuitive results, showing that phenomena such as sharp changes or strongly uneven fluctuations can arise as a result of pure randomness*[4], in the form of one large outlier or Dragon-King.

To this end we first pointed out the difficulties of transforming the mathematical formulation of the large deviations theory into a guide for studying stock market returns. Our findings could then be concluded in that large deviations of stock markets exist and that they can be partially attributed to only one large outlier. The proposed mechanism that we considered as responsible for large deviations dominates at smaller deviations, but usually does not suffice to explain the largest ones. An attempt to make use of our assumptions proved largely unsuccessful, but led to interesting remarks about the markets' characteristics, that reduce the ability of the large deviations theory to extract practical conclusions.

Of course, since this was only an early attempt, further inquiry could shed light to perspectives that were omitted from this study. One such would be to consider other power law distributions, for instance should a sudden overall increase in internet traffic (which is also approximated by a power law distribution [21]) be attributed to just a couple of large outlier sites? And if we stick to stock market returns, many parameters, such as the choice of this particular dataset, the way that the expected average was approached, what type of returns were considered etc. are subject to revision. Probably even the fact that we favoured a power law distribution instead of some other fat-tailed one (e.g. a stretched exponential distribution) could lead to remarkably different results.

Notwithstanding these changes, we saw in the previous chapter that the nature of stock markets poses additional difficulties and such obstacles need to be filtered out, in a way that is far from trivial. Yet, even with the existence of such problems, large deviations of stock market returns are present and their implication is definitely worth investigating, including the mechanisms that create them, the ability to predict and utilize them and foremost the consequences that they may lead to.

Appendix A

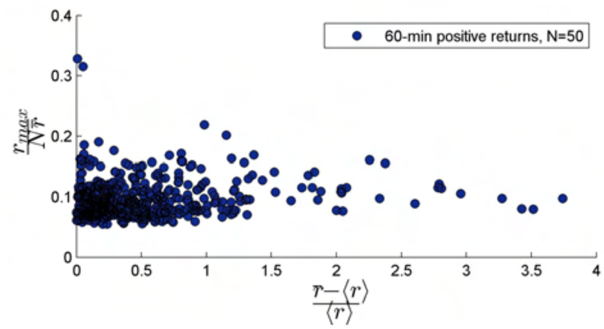

Figure A.1: Largest return vs large deviation - 60-min positive, N=50, $\langle r\rangle$: long-term average

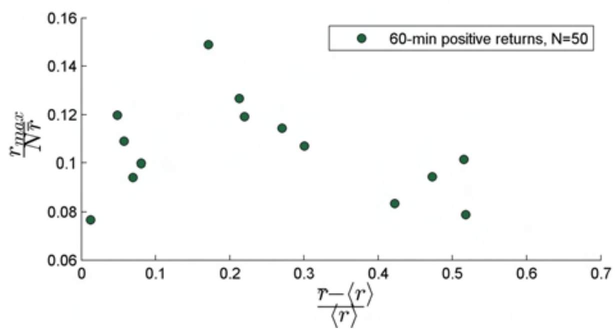

Figure A.2: Largest return vs large deviation - 60-min positive, N=50, $\langle r\rangle$: power law mean

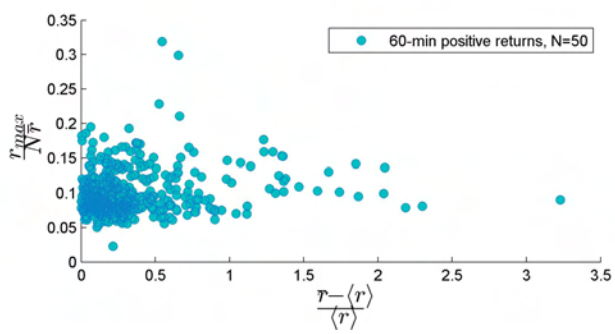

Figure A.3: Largest return vs large deviation - 60-min positive, N=50, $\langle r\rangle$: short-term average

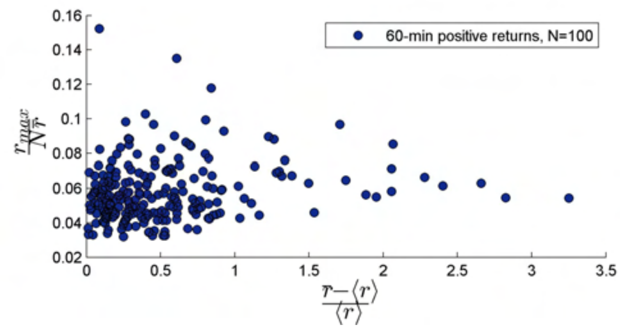

Figure A.4: Largest return vs large deviation - 60-min positive, N=100, $\langle r \rangle$: long-term average

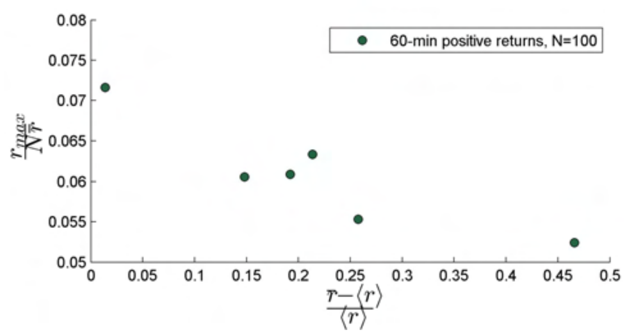

Figure A.5: Largest return vs large deviation - 60-min positive, N=100, $\langle r \rangle$: power law mean

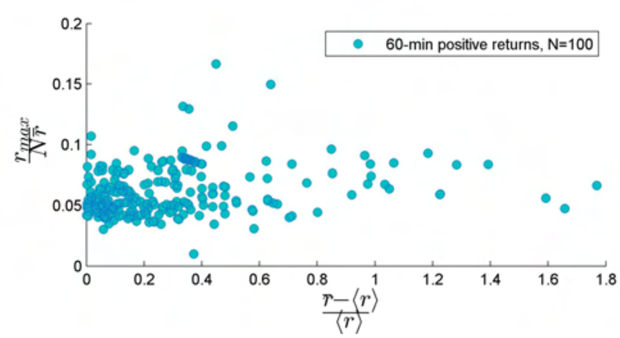

Figure A.6: Largest return vs large deviation - 60-min positive, N=100, $\langle r \rangle$: short-term average

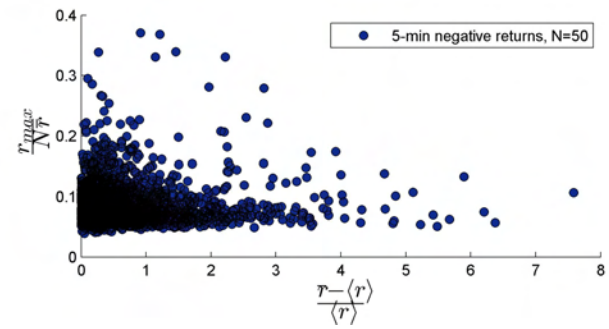

Figure A.7: Largest return vs large deviation - 5-min negative, N=50, $\langle r\rangle$: long-term average

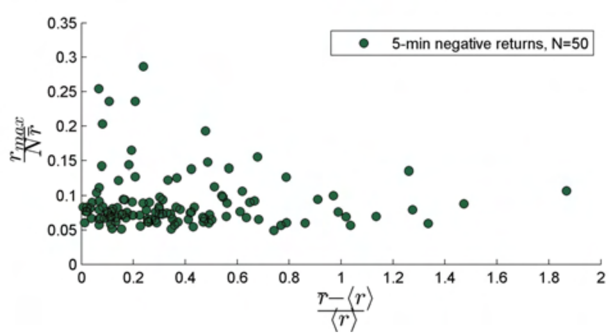

Figure A.8: Largest return vs large deviation - 5-min negative, N=50, $\langle r\rangle$: power law mean

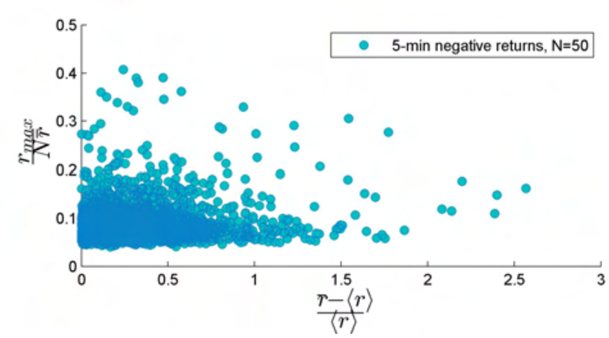

Figure A.9: Largest return vs large deviation - 5-min negative, N=50, $\langle r\rangle$: short-term average

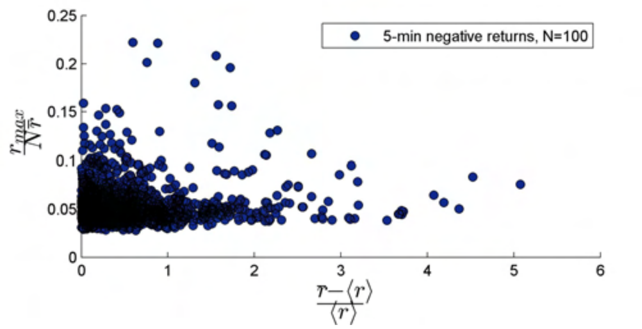

Figure A.10: Largest return vs large deviation - 5-min negative, N=100, $\langle r\rangle$: long-term average

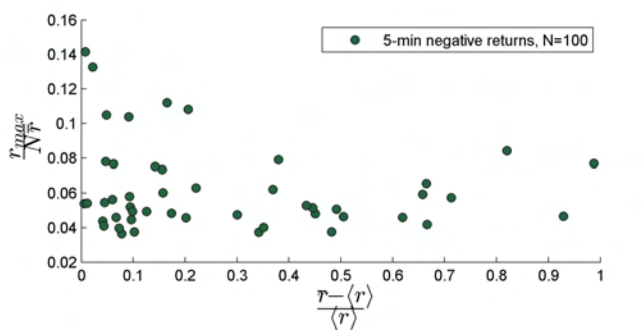

Figure A.11: Largest return vs large deviation - 5-min negative, N=100, $\langle r\rangle$: power law mean

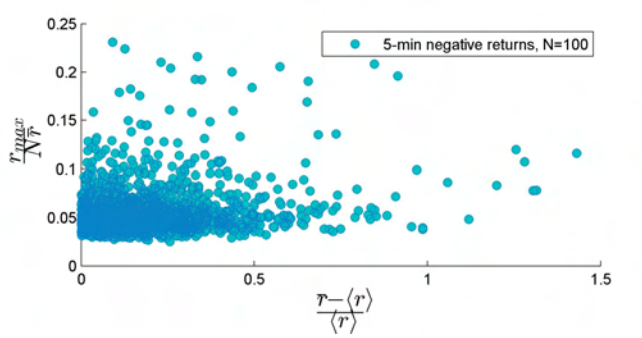

Figure A.12: Largest return vs large deviation - 5-min negative, N=100, $\langle r\rangle$: short-term average

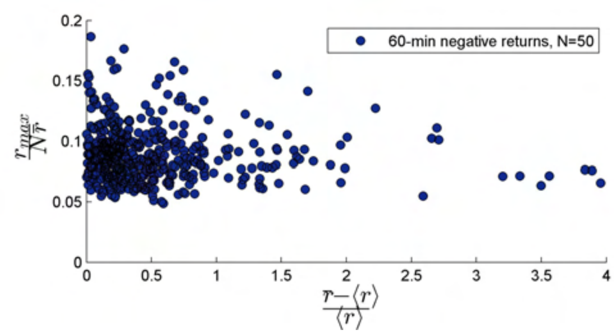

Figure A.13: Largest return vs large deviation - 60-min negative, N=50, $\langle r\rangle$: long-term average

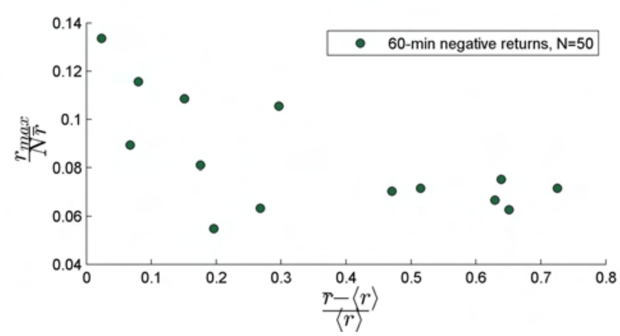

Figure A.14: Largest return vs large deviation - 60-min negative, N=50, $\langle r\rangle$: power law mean

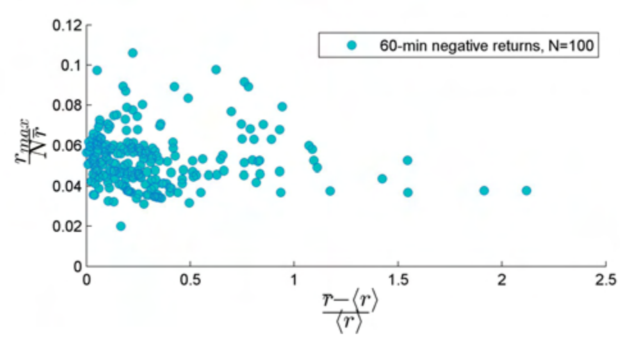

Figure A.15: Largest return vs large deviation - 60-min negative, N=50, $\langle r\rangle$: short-term average

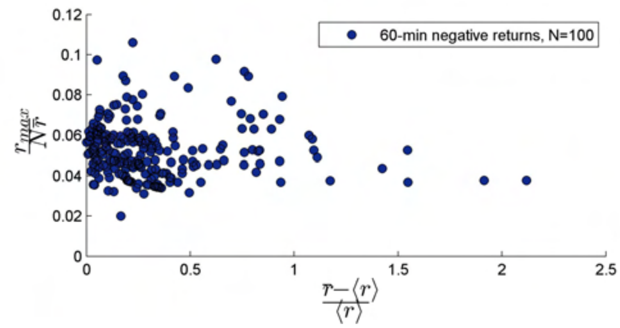

Figure A.16: Largest return vs large deviation - 60-min negative, N=100, $\langle r\rangle$: long-term average

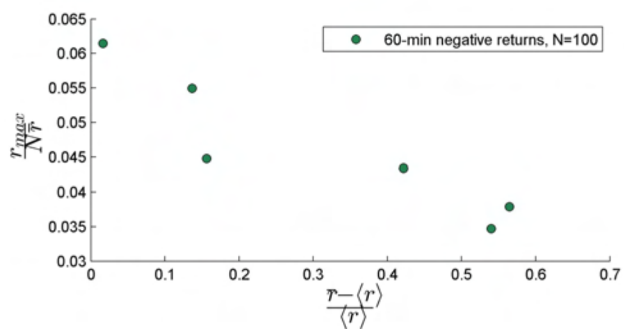

Figure A.17: Largest return vs large deviation - 60-min negative, N=100, $\langle r\rangle$: power law mean

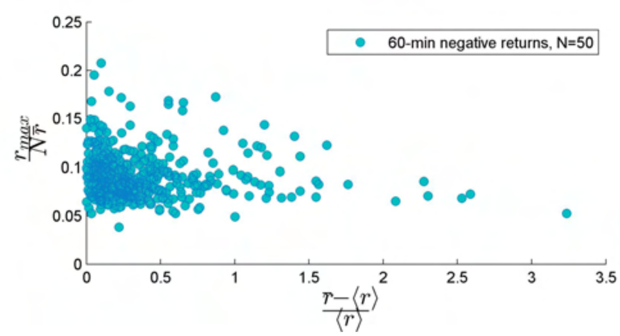

Figure A.18: Largest return vs large deviation - 60-min negative, N=100, $\langle r\rangle$: short-term average

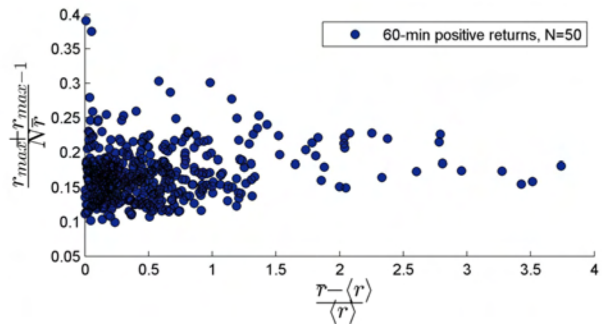

Figure A.19: Two largest returns vs large deviation - 60-min positive, N=50, $\langle r \rangle$: long-term average

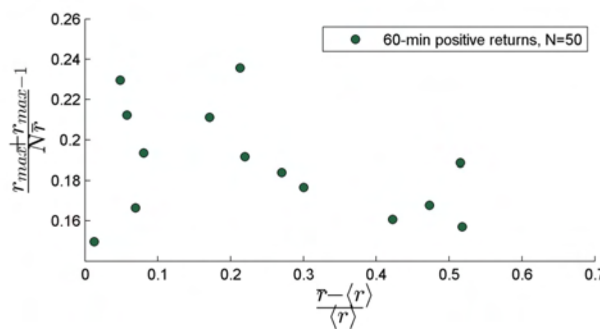

Figure A.20: Two largest returns vs large deviation - 60-min positive, N=50, $\langle r \rangle$: power law mean

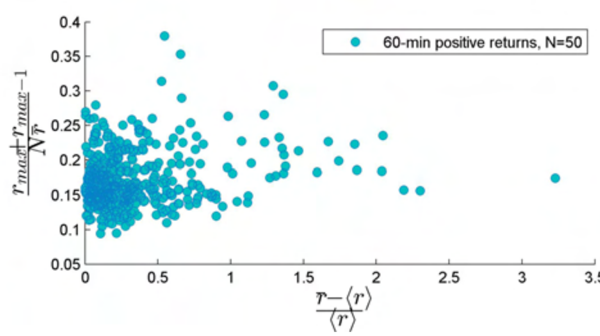

Figure A.21: Two largest returns vs large deviation - 60-min positive, N=50, $\langle r \rangle$: short-term average

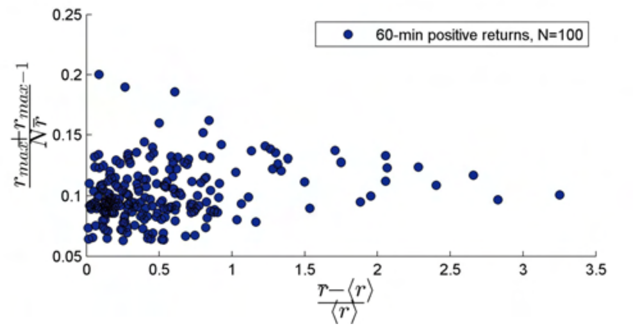

Figure A.22: Two largest returns vs large deviation - 60-min positive, N=100, $\langle r \rangle$: long-term average

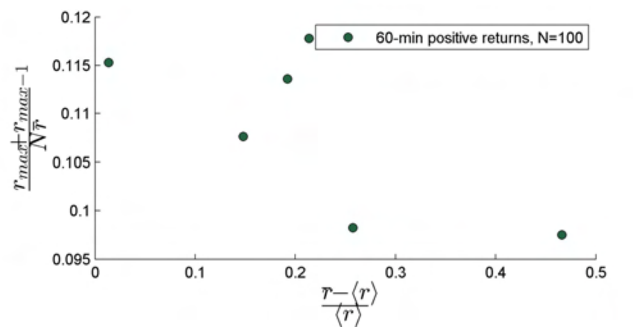

Figure A.23: Two largest returns vs large deviation - 60-min positive, N=100, $\langle r \rangle$: power law mean

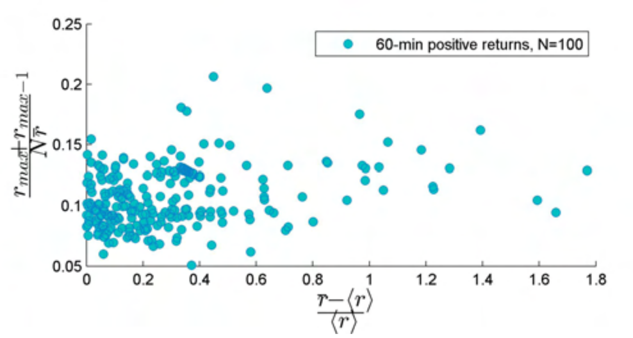

Figure A.24: Two largest returns vs large deviation - 60-min positive, N=100, $\langle r \rangle$: short-term average

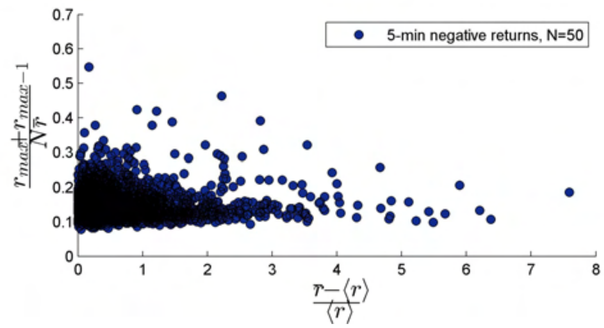

Figure A.25: Two largest returns vs large deviation - 5-min negative, N=50, $\langle r\rangle$: long-term average

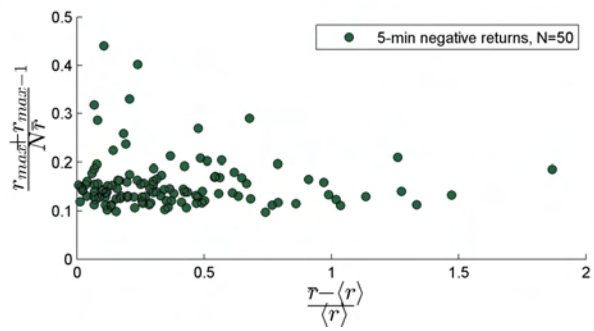

Figure A.26: Two largest returns vs large deviation - 5-min negative, N=50, $\langle r\rangle$: power law mean

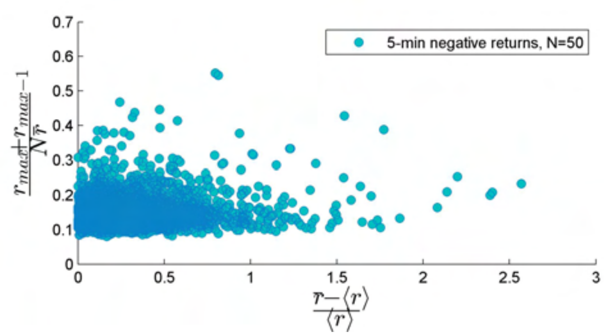

Figure A.27: Two largest returns vs large deviation - 5-min negative, N=50, $\langle r\rangle$: short-term average

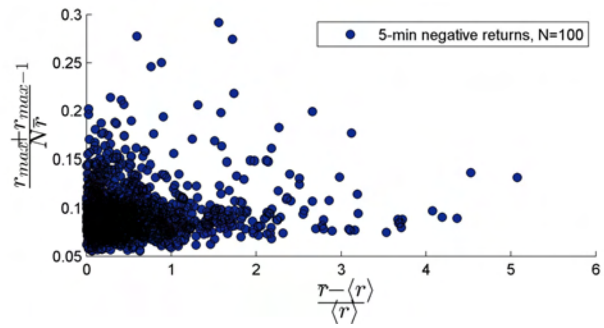

Figure A.28: Two largest returns vs large deviation - 5-min negative, N=100, $\langle r\rangle$: long-term average

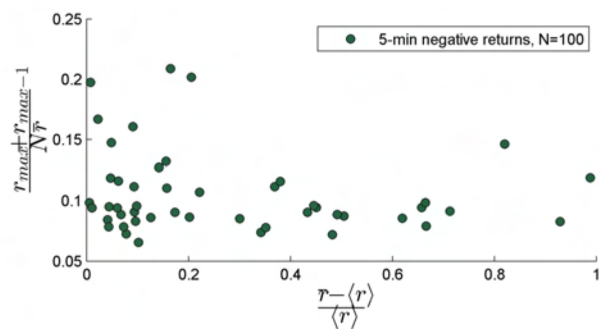

Figure A.29: Two largest returns vs large deviation - 5-min negative, N=100, $\langle r\rangle$: power law mean

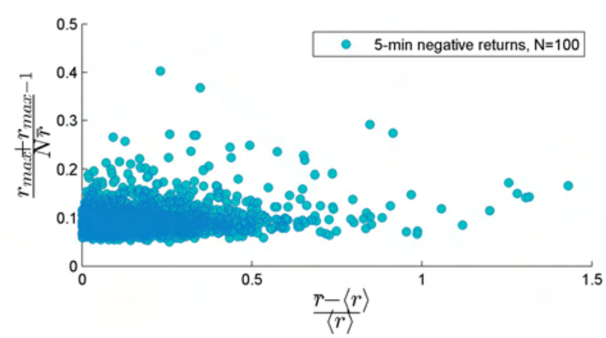

Figure A.30: Two largest returns vs large deviation - 5-min negative, N=100, $\langle r\rangle$: short-term average

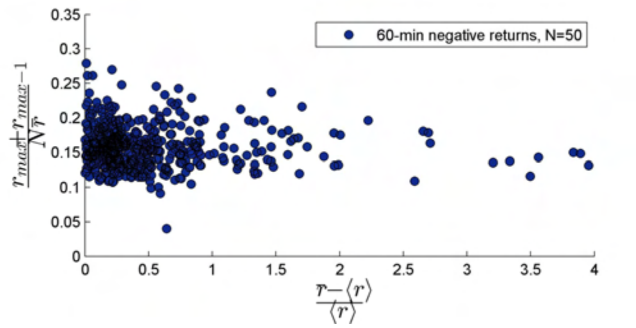

Figure A.31: Two largest returns vs large deviation - 60-min negative, N=50, $\langle r\rangle$: long-term average

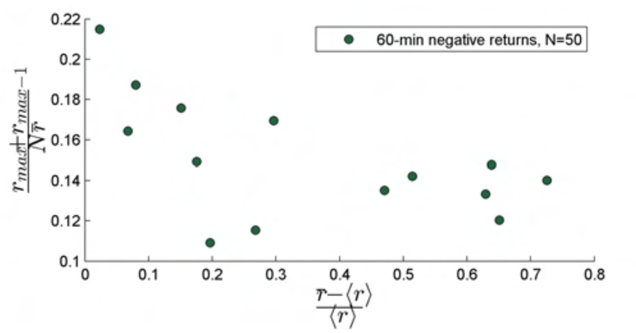

Figure A.32: Two largest returns vs large deviation - 60-min negative, N=50, $\langle r\rangle$: power law mean

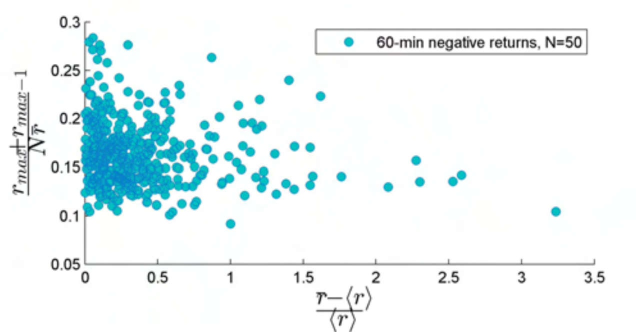

Figure A.33: Two largest returns vs large deviation - 60-min negative, N=50, $\langle r\rangle$: short-term average

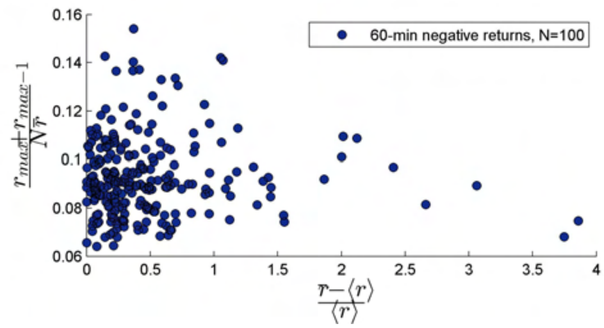

Figure A.34: Two largest returns vs large deviation - 60-min negative, N=100, $\langle r\rangle$: long-term average

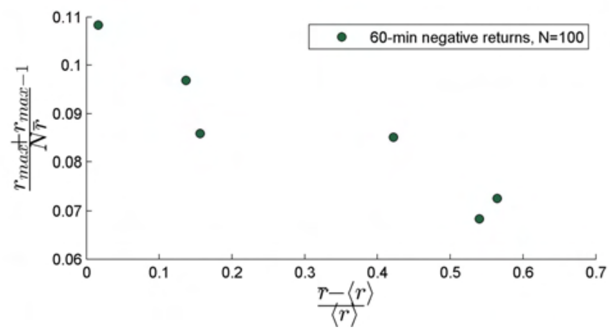

Figure A.35: Two largest returns vs large deviation - 60-min negative, N=100, $\langle r\rangle$: power law mean

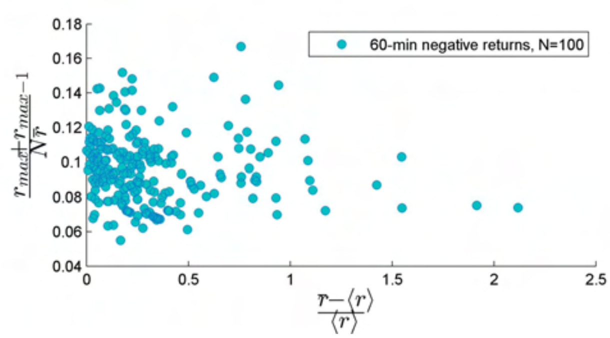

Figure A.36: Two largest returns vs large deviation - 60-min negative, N=100, $\langle r\rangle$: short-term average

Bibliography

[1] H. Pham. Large deviations in mathematical finance. Third SMAI European Summer School in Financial Mathematics, 2010.

[2] S. R. S. Varadhan. Asymptotic probabilities and differential equations. *Communications on Pure and Applied Mathematics*, 19:261–286, 1966.

[3] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns: between the stretched exponential and the power law? *Quantitative Finance*, 5 (2):379401, 2005.

[4] M. Marsili. On the concentration of large deviations for fat tailed distributions, 2012. arXiv:1201.2817.

[5] L. A. N. Amaral, V. Plerou, P. Gopikrishnan, M. Meyer, and H. E. Stanley. The distribution of returns of stock prices. *International Journal of Theoretical and Applied Finance*, 3(3):365369,, 2000.

[6] T. Kaizoji and D. Sornette. Market bubbles and crashes, 2009. arXiv:0812.2449.

[7] D. Sornette. Probability distributions in complex systems. Review article for the Encyclopedia of Complexity and System Science (Springer Science), 2009. arXiv:0707.2194.

[8] D. Sornette. *Critical Phenomena in Natural Sciences, Chaos, Fractals, Self-organization and Disorder: Concepts and Tools.* Springer Series in Synergetics, Heidelberg, 2nd edition, 2003.

[9] H. Touchette. A basic introduction to large deviations:theory, applications, simulations, 2012. arXiv:1106.4146v3.

[10] M. Mezard and A. Montanari. *Information, Physics, and Computation.* Oxford University Press, USA, 2009.

[11] D. Sornette. Dragon-kings, black swans and the prediction of crises. *International Journal of Terraspace Science and Engineering*, 2(1):1–18, 2009. arXiv:0907.4290.

[12] I. V. Zaliapin, Y. T. Kagan, and F. P. Schoenberg. Approximating the distribution of pareto sums. *Pure and Applied Geophysics*, 162: 1187–1228, 2005.

[13] G. Casella and R. L. Berger. *Statistical Inference*. Duxbury Press, California, 2nd edition, 2001.

[14] H. Bauke. Parameter estimation for power-law distributions by maximum likelihood methods. *The European Physical Journal B*, 23 (2): 167–173, 2007. arXiv:0704.1867.

[15] B. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3 (5):1163 –1174, 1975.

[16] S. Stoev, G. Michailidis, and M. Taqqu. Estimating heavy-tail exponents through max self-similarity, 2006. arXiv:0609163.

[17] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661–703, 2009. arXiv:0706.1062.

[18] I. Hacking. *Logic of Statistical Inference*. Cambridge University Press, 1965.

[19] I. Hacking. Review: Likelihood. *British Journal for the Philosophy of Science*, 23 (2):132–137, 1972.

[20] H. Akaike. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19:716–723, 1974.

[21] M. E. J. Newman. Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46:323, 2005.

[22] J.D. Farmer and J. Geanakoplos. Power laws in economics and elsewhere. *SFI Technical report*, 2008.

[23] X. Gabaix. Power laws in economics and finance. *Annual Review of Economics, Annual Reviews*, 1(1):255–294, 2009.

[24] E. Jondeau and M. Rockinger. Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10:559–581, 2003.

[25] Y. Malevergne and D. Sornette. *Extreme Financial Risks: From Dependence to Risk Management*. Springer, Berlin-Heidelberg, 2005.

[26] C. Chrisinger, E. Choi, E. Covington, and J. Rder. Indicators for equity and welfare, 2007. Computer Science and Engineering & Daniel J. Evans School of Public Affairs, University of Washington, Final Report.

[27] J. Cochrane. Discount rates. *Journal of Finance*, 66:1047–1108, 2011.

[28] J. Shoven J. Campbell, P. Diamond. Estimating the real rate of return on stocks over the long term, 2001. Presented to the Social Security Advisory Board.

[29] D. Sornette. *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton University Press, 2004.

[30] P. Bak and M. Paczuski. Complexity, contingency and criticality. *Proc. Nat. Acad. Sci.*, 92:6689–6696, 1995.

[31] D. Sornette and G. Ouillon. Dragon-kings: mechanisms, statistical methods and empirical evidence. *Eur. Phys. J. Special Topics*, 205: 1–26, 2012.

[32] T. R. Palfrey and S.W. Wang. Speculative overpricing in asset markets with information flows. *Econometrica*, 80 (5):19371976, 2012.

[33] R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 1st edition, 2000.

[34] D. S. Scharfstein and J. C. Stein. Herd behavior and investment. *The American Economic Review*, 80(3):465–79, 1990.

[35] E. Fama. The behavior of stock-market prices. *The Journal of Business*, 38:34–105, 1965.

[36] R. Walpole, R. Myers, S. Myers, and Keying E. Ye. *Probability and Statistics for Engineers and Scientists*. Pearson, 9th edition, 2011.

[37] S. Figlewski and X. Wang. Is the 'leverage effect' a leverage effect? *Technical Report, NYU Stern School of Business and City University of Hong Kong*, 2000.

[38] J.P. Bouchaud, A Matacz, and M Potters. Leverage effect in financial markets: the retarded volatility model. *Physical review letters*, 87(22): 228701(1–4), 2001.

[39] M. Potters and J.P. Bouchaud. More stylized facts of financial markets: leverage effect and downside correlations. *Physica A*, 299:6070, 2001.

[40] M. Loretan and W.B. English. Evaluating correlation breakdowns during periods of market volatility, 2000. Board of Governors of the Federal Reserve System International Finance Working Paper No. 658,.

[41] A. Hameed M. J. Cooper, R. C. Gutierrez Jr. Market states and momentum. *The Journal of Finance*, 59:13451365, 2004.

[42] J. Crombez. Momentum, rational agents and efficient markets. *The Journal of Psychology and Financial Markets*, 2(4):190200, 2001.

[43] J. Laherrre and D. Sornette. Stretched exponential distributions in nature and economy: Fat tails with characteristic scales. *European Physical Journal B 2*, pages 525–539, 1995.