RAFAEL KLEIN RIVERA

EXTREME PRICE DRAWDOWNS IN FINANCIAL MARKETS WITH TIME-VARYING VOLATILITY

MASTER'S THESIS

ETH Zurich
Chair of Entrepreneurial Risks
Prof. Dr. Didier Sornette
January 2011

Klein Rivera, Rafael: *Extreme price drawdowns in financial markets with time-varying volatility,* Master's thesis, ETH Zurich, January 2011.

CONTACT:
krafael@ethz.ch

ABSTRACT

Drawdowns offer a more natural measure of the financial market dynamics than fixed time-scale measures, as they take sudden persistences of successive daily drops into account. While the vast majority of drawdowns follow fat-tailed distributions, previous studies report evidence that the very largest drawdowns in a variety of financial markets are much more unrestrained than expected and belong to a different statistical distribution than the bulk of the drawdowns. We extend this concept of "Dragon Kings" by putting drawdowns into the financial context of their time, taking the observed time-varying volatility of returns into account. We define realised volatility as the standard deviation of returns in a moving time window, and extend upon this by estimating the standard deviation with robust estimators of scale, which are more resistant to extreme outlier returns. We find that Rousseeuw and Croux's estimator S_n [42] has both theoretical advantages and performs very well on real financial data.

Our study comprises eight time series (three stock market indices, one currency, two government bonds and two commodities) and we follow two approaches to describe drawdowns with respect to the volatility at their time: we adjust drawdowns by the volatility, and we segregate drawdowns happening during the same volatility regimes, i.e. we group drawdowns with similar volatility levels together. We do not find any significant new properties of the distributions by adjusting drawdowns. However, segregating drawdowns a) *confirms* Dragon Kings found in the whole population of drawdowns; b) makes these Dragon Kings *more pronounced*, i.e. distributions of drawdowns during different volatility regimes exhibit in most cases more obvious deviations and outliers than the distributions of the whole population; c) possibly finds *new* Dragon Kings, which are not visible in the distributions of the whole population. This tells us that studying distributions of drawdowns separately during different volatility regimes reinforces the idea of the existence of Dragon Kings, with the extension that they do not need to be necessarily absolute extreme drawdowns, but "relative extreme" in their financial context.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

"Errors of Nature, Sports and Monsters correct the understanding in regard to ordinary things, and reveal general forms. For whoever knows the ways of Nature will more easily notice her deviations; and, on the other hand, whoever knows her deviations will more accurately describe her ways."

FRANCIS BACON, *NOVUM ORGANUM*, 1620[1]

Systems with a large number of interconnected parts self-organise their dynamics and internal structures to the extent that they exhibit — as a whole — surprising new properties not obvious from the properties of their individual parts. The complex system approach takes these "emergent" properties into account by studying the interconnections and relationships of the mutually interacting individual parts and is nowadays ubiquitous in most of the scientific disciplines, such as in biology (e.g. ecology, evolution, neurobiology), geology (e.g. earthquakes, weather, erosion), economy and social sciences (e.g. markets, cognition, interacting agents).

1 Book II, aph. 29. Cited in [24].

A Gaussian world called Mediocristan

Mediocristan, an idea conceived by Taleb [51], is a mythical land inhabited by economists and other social scientists. They live in this land because they accept the crucial idea that the world's events fit neatly beneath a Gaussian bell curve of outcomes — not because they are mediocre themselves. In fact, they believe that extreme events, such as market crashes and other major discontinuities in our economy and society, are so rare that we can ignore them. Inhabitants of Mediocristan start by *discarding* extreme outcomes, such as in the lively discussion about the world's peak oil production [35] : "Remember that barring any unforeseen tragedy [...] the supply will not drop suddenly, meaning that the price will not rise suddenly." On the other hand, those concerned about the possible severe consequences resulting from peak oil *focus* on extreme outcomes.

Black swans in Extremistan

In Taleb's mythical world the peak oil believers, who know that it is worth taking out insurance against seemingly unlikely events, if their impact could be very severe, are placed in the land of Extremistan. In that world extreme events — "Black Swans" — are still rare, but are very important to take into account, since their consequences are so considerable that they could change the course of history.

This is how one of the most remarkable emergent properties of natural and social systems is the punctuated occurrence of rare large events, which often dominate the organisation and lead to huge alterations. The pervasive statistical feature of these systems and assumed signature of self-organising mechanisms at the origin of a hierarchy of scales is a probability distribution function with a power law dependence as a function of event sizes [44, 45].

A probability distribution function of returns $P(x)$ exhibiting a power law tail is such that

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

for large x, possibly up to some large limiting cut-off. The exponent α, the so-called "tail index", characterises the nature of the tail. In the case of financial returns, the tail index is estimated to fall into the range of 2 to 4 [8, 14, 15, 31, 38]. For $\alpha < 2$, one speaks of a "fat tail" for which the variance is theoretically not defined.

Power law distributions incarnate the notion that tail events, i.e. events of large impact, are not exceptional events, rather they belong to the same statistical distribution as their smaller siblings — like the belief that a great earthquake is just a smaller one that grows larger. This suggests common generating mechanisms of a population described by power law distributions. Consequently, there is no way to predict large events as they share the same attributes with the smaller ones. This is the view embraced by Taleb, where extreme events are reduced — rather pessimistically — to "unknown unknowns", and by Bak et al. in their formulation of self-organised criticality [4, 5].

Dragon Kings beyond power laws

However, evidence suggests that there is more beyond mere power law distributions, since in a wide range of complex systems extreme events are much more unrestrained than expected by the extrapolation of the power law distribution in their tails and belong to a different statistical distribution — different from the bulk of the distribution. These anomalous events can be termed genuine "outliers", events to be removed to obtain reliable statistical estimations, or even "kings" [28] or "dragons" [46], in analogy to the fortune of kings, which appear to exist beyond the Zipf law wealth distribution of their subjects, and to stress coexistence of a completely different kind of species, whose presence has profound significance.

Empirical evidence of the existence of these "Dragon Kings" has been found in the distribution of city sizes, material failure processes, hydrodynamic turbulences, epileptic seizures in humans, earthquakes and — most interestingly for our study — in the distribution of financial market drawdowns (see [46] for a general overview on the concept of Dragon Kings conceived by Sornette et al.). These Dragon Kings reveal the existence of mechanisms of self-organisation and are often associated with a neighbourhood of what can be called equivalently a phase transition, bifurcation, catastrophe or tipping point. Their status emerges in general from the existence of positive feedbacks that amplify the role of certain events and carry a unique fingerprint in form of a log-periodic power law [20, 25, 30]. That latter property opens the door to methods for the prediction of phase transitions, such as the forecast of the termination of financial bubbles [43, 50].

Drawdowns: Transient bursts of dependence

The characterisation of anomalous large financial market moves is of profound importance for risk management and portfolio allocation. Although most of the time, negative and positive changes balance each other out, "extreme variations are so high that they represent a significant part of loss or profit at the end of any period", as Mandelbrot pointed out [29].

It is widely accepted that the tails of the distribution of asset returns follow approximately a power law. But this is only part of the truth, since the claim is to characterise the statistics of extreme events. Financial returns defined at fixed time scales, based on the assumption that consecutive daily returns are independent, are revealing only a part of the variability of financial time series, while a major risk component is missing. In contrast, financial crashes (extreme events) are transient bursts of dependence between successive large losses.

A simple example clarifies the claim: Consider a crash as a sequence of three successive drops of 10% each, summing up to a total loss of 30%. A single drop of 10% can be seen to happen on average every four years or with a probability of 10^{-3} (with data from the Nasdaq composite index, assuming 250 trading days a year). Accordingly, the probability of three such events in a row is $(10^{-3})^3 = 10^{-9}$, an excessively rare recurrence of roughly four million years.

We see that decomposing large crashes into small independent events at fixed time scales misses the underlying dynamics of the market. However, the analyses in [21–25] report evidence that the very largest *drawdowns* in exchange markets, major world stock markets, bond markets and commodity markets are Dragon Kings (outliers), notwithstanding the fact that the very largest drops at fixed time scales (e.g. daily returns) are *not* outliers most of the time. Therefore, drawdowns offer a more natural measure of the financial market dynamics than fixed time-scale measures, as they take sudden persistences of successive daily drops with additional correlated amplification of the drops into account. In addition, for two-thirds of the identified Dragon Kings, it has been found that crashes can be considered as possible ends of bubble regimes [23], which are preceded by a faster-than-exponential unsustainable growth regime. Mechanisms leading to this growth regime with positive feedbacks include among others portfolio insurance trading, option hedging, momentum investment and imitation-based herding.

The implication for risk management is that common techniques, such as Value-at-Risk and Extreme Value Theory that focus only on one-day

extreme events occurring in a given period of time may not be the most important or most relevant measures of large risks. More light should be shed on the distribution of drawdowns and its outliers, since large losses are often the result of transient correlations leading to large cumulative losses that make these outlier-drawdowns much more frequent than expected from the extrapolation of the return distributions in their tails.

Previous research by Sornette and co-workers on the detection of Dragon Kings in financial time series was focused on the distribution of drawdowns in unadjusted series of daily returns. The results have been found to be robust with respect to change in various measures of drawdowns, in particular which allow for a certain degree of fuzziness in the definition of cumulative losses [22, 23].

In early works by Johansen and Sornette [21], drawdowns were defined as continuous decreases of the price at the close of each trading day (daily close). Hence, a drawdown was exclusively composed of negative returns, since *any* positive return marked the end of a drawdown, no matter how small the increase. "Pure drawdowns" composed by this definition were highly sensitive to noise and failed to account for the full intensity of cumulative market drops. Coarse-grained drawdowns as defined in [23] ignore increases below a certain threshold, which would terminate a "pure drawdown". Hence, these "ϵ-drawdowns" allow fluctuations below a threshold ϵ in the time series. The threshold is fixed over time and proportional to the volatility, defined as standard deviation σ_r over the entire time series of returns, with a coefficient ϵ_0 chosen from the observation of data. Formally, the threshold is defined as $\epsilon = \epsilon_0 \cdot \sigma_r$.

Recent attempts [34] to empirically determine whether Dragon Kings in financial returns are present at all time scales, concluded that there is no evidence for extreme outliers in high frequency data at one minute resolution. However, at daily time scales outliers are clearly present. These results suggest that the feedback mechanisms leading to Dragon Kings require a certain time for escalation and confirm the significance of the previous research on drawdowns and Dragon Kings.

Crashes and their financial context

In another recent approach [29] the question of identifying crashes in financial markets is further discussed with respect to the observed instability of the standard deviation of returns over time. Here, Le Bris' innovative approach is to put crashes into the financial context of their time. In other words, all previous attempts investigated the presence of drawdowns

and outliers time-independently, implying a constant volatility, and thus, a constant risk scenario over time. However, as a strong drawdown in tough, highly volatile market conditions has less negative impact for the investors than in stable conditions, a crash should represent a significant discrepancy with what was previously observed, not with what was "on average" throughout the time series observed.

A simple (exaggerated) example illustrates the point: Consider a drop of 20% when the volatility is 10%, and a drop of 2% when the volatility is 0.1%. The former is just a 2 standard deviations event, whereas the latter is a gigantic 20 standard deviations event.

Le Bris presents a method to adjust the time series of returns relative to its financial context by defining the adjusted return $r_{t,adjusted}$ at time t as the difference between the (unadjusted) return r_t and the average return $\mu_{t-1,t-T}$ previously observed over the time context T, divided by the standard deviation $\sigma_{t-1,t-T}$ previously observed over the time context T. Thus:

$$r_{t,adjusted} = \frac{r_t - \mu_{t-1,t-T}}{\sigma_{t-1,t-T}}$$

This adjustment of returns is equivalent to a "rolling standardisation" leading to a standard score, which allows a direct comparison of returns with different averages and standard deviations. The unit of measurement of the adjusted returns is the standard deviation, or "risk".

Our present study attempts to systematically extend this simple idea to the Dragon King detection methodology by investigating the impact of volatility that varies over time. To do this, we will follow two different approaches: We will adjust drawdowns by volatility at their time, similar to Le Bris' approach, and we will group drawdowns that happened during similar "volatility regimes" together, to investigate only distributions of drawdowns at similar market conditions. For both approaches we will employ the same methodology to detect outliers as successfully done in previous studies.

We will start in chapter 2 with a general discussion on volatility, in order to define a robust method to construct a time series of time-varying volatility. In chapter 3 we will reinvestigate the definitions of pure and coarse-grained drawdowns, particularly to take into account a variable level of volatility and noise. In chapter 4 we will present a methodology to detect Dragon Kings in distributions of drawdowns, either taking volatility into account or ignoring it, and adopting some tools previously used for other cases than drawdowns. In chapter 5 we will apply these concepts

of volatility, drawdowns and Dragon King detection to various financial time series. In chapter 6 we will conclude with the following question in mind: whether Dragon Kings in distributions of drawdowns taking a variable volatility into account are in line with the results from previous studies or give an even better insight into the nature of returns, drawdowns, Dragon Kings and extreme risks, with the goal to develop more adequate measures of risk in the future.

Just as Francis Bacon observed in the 17th century, we propose that the distribution of drawdowns and its outliers reveal fundamental mechanisms and properties of the financial markets.

2 VOLATILITY

Evolving volatility is a dominant feature observed in most financial time series and a key parameter used by market participants in many financial risk analyses. It is a measure of the variability of financial returns and a very popular proxy for risk. There is an extensive literature on the estimation of parametric volatility models (e.g. GARCH, MIDAS and others, see [6, 10, 16, 52]). However, these approaches have in common the fact that they are modelled to give ex ante estimates of the volatility and strongly rely on assumptions about distributions and other attributes of the time series.

Our aim is to define a robust estimator for volatility to capture the market dynamics and risk investors were faced with over the time of the time series. Since drawdowns (further discussed in chapter 3) carry only information about loss in absolute terms, a measure for volatility at the time drawdowns occurred will put the magnitude of absolute loss in relation to the actual volatility regime, i.e. to the actual market risk scenario at the time of the drawdowns. Hence, we need approaches to estimate the evolving volatility that are robust to distributional assumptions and have a sound statistical basis with reasonable precision properties, because "[in] analysing data, we do not want to even attempt to represent its stochastic behaviour accurately; rather we wish to choose techniques that spare us this essentially impossible task," as Morgenthaler and Tukey pointed out [36]. However, literature on these simpler, for our study more appropriate, non-parametric models is scarce, and usually the standard deviation is chosen as the conventional non-parametric volatility estimator.

Recent attempts by Randal [39] to construct more sophisticated non-parametric volatility estimates have shown that there is room for improvement on the accuracy of the results, in particular with respect to the resistance to outliers. These approaches are claimed to be simple, cheap to compute and tailored to fat-tailed distributions, since the standard deviation as reasonable estimator of scale for the volatility may be inefficient for non-gaussian distributions.

With Morgenthaler and Tukey in mind, we begin in section 2.1 with an overview of the problem of robust scale estimation, and then present in

section 2.2 a simple, powerful and non-parametric approach to construct estimators of time-varying historical volatility. This is a general purpose technique for estimating volatility, no attempt has been made to build predictive or stochastic volatility estimates. The approach is non-parametric, since it can be applied to time series without modelling underlying price processes. For a more detailed discussion on this and more advanced techniques please refer to Randal's study [39]. The approach we consider here is based on a moving time window of standard deviation estimates — the "realised" or "historical" volatility. In section 2.3 we introduce the concept of volatility regimes and section 2.4 concludes this chapter on volatility with a brief discussion of the time-varying volatility using our approaches we have presented in this chapter. As an actual example we will look at the time around the October 1987 crash ("Black Monday").

2.1 ROBUST SCALE ESTIMATION

Volatility is a measure of the variability of financial returns. In statistical terms, variability is the dispersion of some sample observations, quantified by a robust measure of scale.

If a robust estimator of scale is largely unaffected by a small number of large changes in the data, i.e. by outliers, and by any number of small errors, e.g. rounding errors, it is called resistant, at the cost of lower statistical efficiency when outliers or errors are not present [39]. The resistance to outliers is usually of greater interest, and, in addition, a robust estimator is a suitable estimator for non-normal data. Hence, robust estimators of scale are particularly applicable to financial data, which often features the three properties we want to protect against: occasional outlier values, many small errors (induced by properties of financial markets such as discrete price intervals and discontinuous trading) and underlying non-normality. In general, robust estimators of scale are used to complement or replace conventional estimates such as the sample standard deviation.

Below, we will present different estimators of scale for the sample observations $X = (X_1, \ldots, X_n)$. First, the non-robust, but most commonly used standard deviation in section 2.1.1; second, Gini's mean difference, another non-robust but commonly used estimator in section 2.1.2; third, the sample interquartile range (IQR) in section 2.1.3; fourth, the median absolute deviation (MAD) and a modification in section 2.1.4; and fifth, the trimmed sample in section 2.1.5. In section 2.1.6 finally, the statistical

performance of the estimators presented here will be briefly assessed on their efficiency.

2.1.1 Sample standard deviation

The most commonly used scale estimator is the sample standard deviation s_X defined as

$$s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} \tag{2.1}$$

with the sample mean

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.2}$$

The sample standard deviation is closely related to the sample variance s_X^2, which is the minimum variance unbiased estimator for the variance parameter σ^2 for Gaussian data. Despite these excellent qualities for well-behaved data, and the fact that the sample variance will be unbiased for the underlying variance generally for random samples, evidence shows us that this estimator is not robust, since it is heavily affected by single outliers [39].

2.1.2 Gini's mean difference

Gini's mean difference with absolute inter-point differences is a similar estimator as the sample standard deviation and forms the basis of robust estimation of risk in the financial literature [39]. It is defined as

$$G_X = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|X_i - X_j\right| \tag{2.3}$$

This estimator is also not very robust, since the absolute differences between every pair of observations are taken into account. Outlying values will result in many inter-point distances being large, with G_X inflated as a result. However, using the absolute differences between every pair rather than the squared differences to the mean in the standard deviation reduces the impact of large differences.

In order to use Gini's mean difference as a consistent estimator $\hat{s}_X$ for the estimation of the standard deviation (normal distribution), it has to be scaled with the factor k_G, such that $\hat{s}_X = k_G \cdot G_X$. It can be derived that $k_G = 0.8862$.

2.1.3 Sample interquartile range

The most commonly used robust scale estimator is probably the interquartile range (IQR), which measures the difference between a distribution's upper and lower quartiles. It is also called the "midspread" or "middle fifty", and is defined as the difference between the third and the first quartiles

$$IQR_X = Q_3 - Q_1 \tag{2.4}$$

where Q_3 is the 75th percentile and Q_1 the 25th percentile. In order to use the IQR as a consistent estimator, the scaling factor is $k_{IQR} = 0.7413$, such that $\hat{s}_X = k_{IQR} \cdot IQR_X$.

The sample interquartile range is resistant to outliers, since it ignores the most extreme 25% of each tail in the observations and is the simplest estimator of scale considered here, since it is very cheap to compute.

2.1.4 Median absolute deviation and S_n

The median absolute deviation (MAD) is another robust scale estimator and is along with the interquartile range, one of the most commonly used robust scale estimators. It is defined as the median of the absolute deviations of the observations X about their median

$$MAD_X = \text{med}_i \left| X_i - \text{med}_j X_j \right| \tag{2.5}$$

where $\text{med}_j X_j$ denotes the median of the sample observations X_j. In order to use the MAD as a consistent estimator, the scaling factor is $k_{MAD} = 1.4826$, such that $\hat{s}_X = k_{MAD} \cdot MAD_X$.

Like the sample interquartile range, the median absolute deviation is also very resistant to outliers and cheap to compute. However, contrary to the interquartile range where the quartiles do not need be equally far away from the centre, the MAD takes a symmetric view on dispersion, since it first estimates a central value (the median) and then attaches equal importance to positive and negative deviations from it [42]. In other words, the MAD corresponds to finding the symmetric interval around the median that contains 50% of the observations. For asymmetric distributions this does not seem to be a natural approach. Because of this drawback, Rousseeuw and Croux [42] introduce a new scale estimator, which can be seen as an analog of Gini's mean difference, where the mean values are replaced

by the median values. We will denote this estimator here as $S_{n,X}$ and define it as

$$S_{n,X} = \text{med}_i \left\{ \text{med}_j |X_i - X_j| \right\} \tag{2.6}$$

where we compute for each X_i the median med_j of $|X_i - X_j|$, for each X_j ($j = 1, \ldots, n$). This yields n medians med_j, the median med_i of which gives our final estimator $S_{n,X}$.[1] The scaling factor for this estimator is $k_{S_n} = 1.1926$, such that $\hat{s}_X = k_{S_n} \cdot S_{n,X}$.

Like the MAD, the new estimator $S_{n,X}$ is a simple combination of medians and absolute values, but without the need for any location estimate of the sample. Instead of measuring how far away the observations are from a central value, $S_{n,X}$ takes the median of the inter-point distances between every pair of observations. This is valid for asymmetric distributions, too.

2.1.5 Trimmed sample

As noted before, the sample standard deviation and Gini's mean difference are not robust estimators of scale, since they can be heavily influenced by extreme values. As a simple approach to make these estimators more robust, we introduce here a method that compensates for this, by dropping a certain lower and upper percentile from the sample of observations. For example, in a 98% trimmed sample, the smallest and largest 1% of the observations would be ignored for the estimation. We will denote a p% trimmed sample as $X^{(p)}$.

Thus, for the estimation of scale using the standard deviation (2.10) or Gini's mean difference (2.3) only a trimmed sample may be used. This trimmed sample alters the sample standard deviation in two ways in order to reduce the effects of outliers. First, the sample mean as the location estimator is altered and second, the squared deviations about this new mean will be different. For Gini's mean difference the effect would be similar, since the inter-point differences would be smaller. The effect of the trimmed sample on the standard deviation and Gini's mean difference as estimators of scale is therefore similar to the sample interquartile range.

1 A straightforward algorithm for computing (2.6) would need $O(n^2)$ computation time. However, Rousseeuw and Croux [7] have constructed a more efficient $O(n \log n)$-time algorithm.

2.1.6 Tukey's trieffiiciency

To assess the statistical performance of the scale estimators introduced before, we present here the results from [39]. A comprehensive overview, discussion and a detailed description of the methodology to assess the performance of these estimators can be found there. The estimators are assessed by their minimum relative efficiency over Tukey's three corners — three sampling situations to reflect the three extreme cases of importance to robust statistics [36]:

1. STANDARD NORMAL DISTRIBUTION
   Described as "unrealistically nice" by Morgenthaler and Tukey.

2. ONE-WILD SITUATION
   Also known as 1-wider, where $n-1$ of the observations in a sample of size n are standard normal and the remaining observation has ten times the standard deviation of the others.

3. SLASH DISTRIBUTION
   An observation that is obtained by dividing a standard normal random variable by an independent random variable distributed uniformly on the interval $[0,1]$. Like the Cauchy, the slash distribution has no mean or variance due to its slowly decaying tails.

All three theoretical distributions are symmetric: the standard normal has rapidly decaying tails; the one-wild has a single outlying, but otherwise well behaved value (in the upper or lower tail with equal probability); and the slash, with its infinite mean and variance, has very slowly decaying tails. In practice, most samples from the one-wild will be highly asymmetric, with the presence of the single outlier. An estimator that copes well in all three situations can be suitably used either when the data is well behaved, or in the presence of occasional outliers, or when the data is very fat-tailed, or some combination [39].

The sample efficiency $\mathit{eff}(S)$ of a scale estimator $S(X)$ is defined as

$$\mathit{eff}(S) = \frac{\text{sample variance of } \ln \hat{\sigma}_i, \ldots, \ln \hat{\sigma}_m}{\text{sample variance of } \ln S(X)_i, \ldots, \ln S(X)_m} \cdot 100\% \qquad (2.7)$$

using m independent realisations of the observations $X = (X_1, \ldots, X_n)$, where $\hat{\sigma}_i$ is the maximum likelihood scale estimate used as an "close-to-optimal" reference and $S(X)_i$ the scale estimate for sample i.[2] The "trieffiiciency" is the minimum efficiency of the estimator over the three corners,

2 See [39] for further details on the maximum likelihood scale estimates.

Table 2.1: Average efficiencies (%) for the estimators of scale, based on 100 realisations of the efficiencies, each estimated from $m = 20000$ samples of size $n = 20$ for the three corner cases (std. normal, one-wild and slash distributions). The estimators are defined in sections 2.1.1 to 2.1.5. (Results obtained from [39], table 2.8)

ESTIMATOR	NORMAL	ONE-WILD	SLASH	TRIEFFICIENCY
s_X	100.0	11.4	7.5	7.5
G_X	98.0	26.7	11.4	11.4
IQR_X	39.4	42.4	84.0	39.4
MAD_X	37.8	40.5	87.3	37.8
$S_{n,X}$	54.7	55.9	95.8	54.7
s_X, $X^{(95)}$	80.9	88.1	42.1	42.1
s_X, $X^{(90)}$	65.0	70.8	76.1	65.0

and the best triefficient estimator will have the maximum triefficiency. We would expect the triefficiency of this estimator to be less than 100% since no single estimator will be optimal at all three corners.

The efficiencies for the sample standard deviation s_X (2.1), Gini's mean difference G_X (2.3), sample interquartile range IQR_X (2.4), median absolute deviation MAD_X (2.5), $S_{n,X}$ (2.6), and the sample standard deviation of a 95% and 90% trimmed sample ($X^{(95)}$, $X^{(90)}$) are presented in table 2.1. We notice that both the sample standard deviation and Gini's mean difference have a very poor performance for the one-wild and slash distributed data, while Gini's mean difference performs slightly better in both cases. However, both estimators are highly efficient for normal data. As expected, the robust estimators IQR, MAD and Rousseeuw and Croux's estimator $S_{n,X}$, which depend primarily on order statistics, perform in the opposite way: their efficiency for normal distributed data is the worst and for slash distributed data the best. Under the criterion of triefficiency, the IQR is more robust than the MAD, i.e. the IQR is generally more suitable. However, the MAD performs better for slash distributed data, while it performs only slightly worse than the IQR for one-wild data. The $S_{n,X}$ estimator out-performs the IQR and MAD in all three cases and has the second best triefficiency of all studied estimators. Only the sample standard deviation from a 90% trimmed sample has a better triefficiency, but still lower efficiencies for the one-wild and slash data than the robust estimators. Finally, for the sample standard deviation from a 95% trimmed sample the efficiency is very high for the one-wild, but very low for the slash.

Since the estimators are assessed only over Tukey's three corner cases and the distributions of drawdowns are somewhere in between of these corners, the estimators' performances shown in table 2.1 should not be taken as a determinant of which estimator to choose. Therefore, to clarify the effect of these estimators on the distribution of drawdowns, we will present an example with real financial data at the end of this chapter.

2.2 REALISED VOLATILITY

Volatility estimates for real data are difficult to appraise since the true volatility is unobservable. A popular non-parametric, model-free estimator of time-varying volatility that provides a useful framework for this problem is the "realised" or "historical" volatility estimator. The realised volatility is basically the sample standard deviation of returns over a rolling time window of a given size. This measure gives us a powerful ex post estimate of the volatility. Since we are interested in capturing the actual market dynamics and not the market participants' assessment of future volatility, it gives us insight into what actually took place in the market. It has been prominent in empirical studies of stock returns [37] and plays an important role used to estimate "historical" volatility as a basis for evaluation of volatility forecasting techniques [12].

We define the realised *daily* volatility $\sigma_{t,T}$ at time t over the time window of size T for the time series of n daily returns $r = (r_1, \ldots, r_n)$ (3.1) as

$$\sigma_{t,T} = s_{r(t,T)} \tag{2.8}$$

where $s_{r(t,T)}$ denotes the standard deviation (2.1) of the sample of returns falling into the time window $r(t, T) = (r_{t-T}, \ldots, r_{t-1})$. Note that in our definition we are not taking into account the return *at* time t.[3]

Using daily data in our case, a time series of model-free volatility estimates can be constructed by taking returns spanning over any number of days T into account. However, when choosing the number of past observations, an important tradeoff has to be made. If the volatility is calculated over a long time frame with a large number of observations, e.g. taking the daily returns over the past year, many local properties of volatility, such as volatility clustering and leverage effect, tend to disappear. On the other

3 In this paper, "volatility" refers to the volatility $\sigma_{t,T}$ at *daily* time scale in percentage units. Hence, the *annualised* volatility corresponds to $\sigma_{t,T,ann} = \sigma_{t,T} \cdot \sqrt{250}$, assuming 250 trading days per year.

hand, by choosing a small number of observations, the measures are subject to great statistical error. Hence, the span T of the moving window should be ideally chosen such that the volatility within the local time window be approximately constant, but the number of observations be sufficiently large to avoid statistical errors.

An important assumption underlying this estimator is that the sample standard deviation $s_{r(t,T)}$ is a reasonable estimator of scale for the variability of returns. However, as we have seen in section 2.1, the sample standard deviation may be very inefficient for non-Gaussian data. As a consequence, it will provide poor estimates of volatility. Therefore, we seek to address this problem by replacing the sample standard deviation $s_{r(t,T)}$ with other robust estimators $\hat{s}_{r(t,T)}$ introduced in section 2.1. In addition, in the following section 2.2.1 we will introduce a more advanced definition of the sample standard deviation for a moving time window than (2.1). This definition will allow us to give exponentially decaying weights to the returns $r(t, T)$. It will not change the standard deviation to become a robust estimator of scale, but it tries to enhance its performance and follows approaches used in practice.

2.2.1 Exponentially weighted standard deviation

Since the aim is to define a volatility that reflects the market dynamics at a certain time t without taking into account exceptional events, such as big rallies and slumps that happened in the close or distant past from t, we refine our definition of volatility with equal weights over the previous period T with exponentially decaying weights over that period. This approach is similar to the approach introduced by JPMorgan's RiskMetrics software [26], where the variance at time t is a weighted average of all past squared returns, with the weights decaying exponentially back through time with the recursion

$$\sigma_t^2 = \omega\sigma_{t-1}^2 + (1-\omega)r_t^2 \tag{2.9}$$

Our definition will help us to improve the characterisation of the market at every time step, giving more importance to the market events in the very close past than to the events that happened in the far past. Hence, we will extend the common definition (2.1) of the sample standard deviation in order to be able to assign different weights to each of the returns r_i falling into the time window of size T, such that $r_i = t - i$ with $i = 1, \ldots, T$. We

define the weighted sample standard deviation $s_{r(t,T),w}$ with weights w_i of the sample of returns falling into the time window $r(t, T)$ as

$$s_{r(t,T),w} = \sqrt{\frac{1}{1 - W_2} \sum_{i=1}^{T} w_i \left(r_i - \bar{r}_w\right)^2} \tag{2.10}$$

with the weighted mean

$$\bar{r}_w = \sum_{i=1}^{T} w_i r_i \tag{2.11}$$

and the sum of the squared weights

$$W_2 = \sum_{i=1}^{T} w_i^2 \tag{2.12}$$

such that $1/T \leq W_2 \leq 1$. Note that we are using normalised weights w_i such that $\sum_{i=1}^{T} w_i = 1$. For the special case with equal weights $w_i = 1/T$, the weighted standard deviation (2.10) equals the common standard deviation (2.1) with the common mean (2.2).

Exponentially decaying weights

Let us now define the exponentially decaying unnormalised weights with damping constant ω as $w'_i = \omega^{i-1}$. A usual choice for the damping constant is $\omega = 0.94$ (e.g. in RiskMetrics) or

$$\omega = 1 - \frac{2}{T + 1} \tag{2.13}$$

The sum of the unnormalised weights is

$$\sum_{i=1}^{T} w'_i = \sum_{i=1}^{T} \omega^{i-1} = \frac{1 - \omega^T}{1 - \omega} \tag{2.14}$$

By dividing the unnormalised weights w'_i by that sum, we get the the exponentially decaying normalised weights w_i for the time window T with damping constant ω:

$$w_i = \frac{1 - \omega}{1 - \omega^T} \omega^{i-1} \tag{2.15}$$

2.3 VOLATILITY REGIMES

As we will discuss in depth later in chapter 4, our aim is to (a) adjust, i.e. divide, drawdowns by the volatility at their time, and (b) to segregate drawdowns that occur at different volatility regimes. The purpose of this concept is, that we will be able to assign to every drawdown a certain volatility regime. In other words, we will be able to group drawdowns that happened during similar market or risk conditions together, and separate drawdowns that happened during different conditions.

To clarify the point, let us reconsider our example from chapter 1: Consider two events, first, a drawdown of 20% when the volatility is 10%, and second, a drawdown of 2% when the volatility is 0.1%. The first drawdown is 2 times larger than the volatility, the second is 20 times larger. Did these two drawdowns happen during similar market conditions? Probably not, since the second event happened at a time with a volatility, i.e. with a "risk", 100 times smaller. Therefore, we propose to analyse the nature of these drawdowns separately, as they seem to belong to different market and risk conditions. Note, that we are only interested in the *relative* magnitude of the volatilities at different times, the *absolute* values of volatility are not pertinent (the volatility in the second event would be still 100 times smaller than in the first).[4] We will now present an approach to determine different volatility regimes.

In this context, a volatility regime is defined as a set of volatility observations with the same level. If we have k volatility regimes V_i with k corresponding volatility levels L_i ($i = 1, \ldots, k$) in a time series of n volatility observations σ_t ($t = 1, \ldots, n$), we define

$$V_i = \{\sigma_t \,|\, \sigma_t \in L_i\} \tag{2.16}$$

as the ith volatility regime. A time series of n volatility observations could have theoretically n different values of volatility. Therefore, and to allow for a certain degree of granularity, we define a volatility level L as an interval $L_i = [l_i, u_i)$ and $L_k = [l_k, u_k]$ with a certain lower bound (l_i) and upper bound (u_i) of volatility, similar to a binning of volatility observations with

4 The exact values of the scaling factors k for the different estimators of scale to obtain a consistent estimator $\hat{s}_X$ for the standard deviation are therefore of little importance for our study, too.

granularity k to reduce the effects of observation errors. We will set the lower and upper bounds for every volatility regime such that

$$l_i = \min_t \sigma_t + (i-1)h \tag{2.17}$$

$$u_i = \min_t \sigma_t + ih \tag{2.18}$$

with the interval width $h = (\max_t \sigma_t - \min_t \sigma_t)/k$ for $t = 1, \ldots, n$. The centre value for each regime is then

$$c_i = l_i + \frac{h}{2} \tag{2.19}$$

For the sake of clarity when dealing with different granularities k, we will denote the volatility regime V_i $(i = 1, \ldots, k)$ as $V_{i/k}$.

2.4 VOLATILITY IN REAL FINANCIAL DATA

In this chapter we have presented three robust estimators of scale, namely the sample interquartile range (2.4), the median absolute deviation (2.5), and Rousseeuw and Croux's estimator $S_{n,X}$ (2.6). We have also presented two well known non-robust estimators, the sample standard deviation (2.1) and Gini's mean difference (2.3), with the possible addition of trimming the sample on both tails (section 2.1.5). In section 2.1.6 we have discussed the efficiency of these estimators with respect to Tukey's three corner cases. We then have introduced a popular method to construct a time series of time-varying volatility, the realised volatility (2.8), with the possible addition of assigning exponentially decaying weights to the returns when using the standard deviation (2.10), and have finally defined the term "volatility regime" in section 2.3.

As mentioned earlier, an important assumption underlying the realised volatility estimator is that the sample standard deviation $s_{r(t,T)}$ is a reasonable estimator of scale for the variability of returns. However, the sample standard deviation will provide poor estimates of volatility for non-Gaussian data. Since the distribution of returns is fat-tailed, this will most certainly be the case, even more around points in the time series of returns with exceptional outlier returns. Therefore, we are going to address this problem by replacing the sample standard deviation $s_{r(t,T)}$ with other estimators $\hat{s}_{r(t,T)}$ and with the standard deviation $s_{r(t,T),w}$ with exponentially decaying weighted returns. We will *qualitatively* compare the effects

of these three robust and three non-robust estimators (additionally with trimmed samples) on the time series of the realised volatility around the Black Monday crash (October 1987), in addition to the *quantitative* comparison testing Tukey's triefficiency we have done in section 2.1.6.

Figure 2.1 shows the volatilities using the different non-robust estimators of scale (Gini's mean difference and exponentially weighted standard deviation) for the S&P 500 index from 1950 to 2000 in panel (a) and for a detail around the October 1987 crash from mid-1983 to mid-1992 in panel (b). Figure 2.2 shows the volatilities using the same estimators of scale as in figure 2.1 but using 90% trimmed samples of returns. Figure 2.3 shows the volatilities using the different robust estimators of scale (IQR, MAD and $S_{n,X}$). In all cases the sample standard deviation (2.1) is shown as a reference (black line) and the rolling time window is set to a size of $T = 125$ trading days (roughly half a year).

Let us start with the estimators in figure 2.1. Since the standard deviation is statistically not a robust estimator of scale, it can be heavily influenced by extreme values. We can see this undesirable effect on the estimated volatility in figures 2.1 to 2.3 (b) (black lines), showing the computed volatility using the sample standard deviation before and after the crash in October 1987. Due to a single extreme value in the sample of returns within the moving time window at Black Monday the estimated volatility jumps up to a level almost three times higher than just before the crash and remains at this elevated level for the span of the time window. This kind of pattern is also clearly visible at the end of 1989. One could tentatively conclude that this volatility behaviour does not truly reflect the risk perception during the time of these "artefacts". The exponentially weighted standard deviation (figure 2.1, blue line) exhibits a noisier behaviour by over- and undershooting with respect to the equally weighted standard deviation. Especially at the time of extreme values it strongly overreacts, since the large returns at the very edge of the time window are much stronger weighted than the past (smaller) returns. However, as intended, the volatility decays faster from its elevated level. Gini's mean difference (figure 2.1, red line) has a behaviour like the standard deviation, exhibiting a similar effect of artefacts. However, these elevated volatility levels are not as high as for the standard deviation, but still clearly visible due to a lack of fast decay.

The volatilities computed using the same estimators as just discussed, but with 90% trimmed samples of returns (i.e. with samples of size $0.9T = 112$), are almost identical to each other (figure 2.2). This is especially true for the standard deviation (green line) and Gini's mean difference (red

line) with trimmed samples. As expected, the overall volatility levels for the trimmed samples are lower than for the standard cases, since the deviations about the mean are smaller and the trimmed estimators are not corrected with a scaling factor k like the robust estimators. However, apart from that, the instantaneous jumps followed by constant elevated levels have disappeared, as clearly visible for the artefact at the crash of 1989, such that the three volatility times series appear to be smoother. The volatility estimations with the standard deviation and Gini's mean difference from trimmed samples are almost identical with the estimation using the exponentially weighted standard deviation from trimmed samples (blue line), such that the previously identified artefact- and overshooting-behaviour have disappeared, respectively. Thus, one can deduce, that these estimates reflect more accurately the risk perception than the non-trimmed estimates.

Let us turn now to the volatility estimates using the robust estimators of scale (figure 2.3). Like the non-robust estimators with trimmed samples, the volatility estimates using the sample interquartile range (blue line), the median absolute deviation (red line), and Rousseeuw and Croux's estimator $S_{n,X}$ are very similar. However, there are still particular features visible for each estimator. The volatilities using the IQR and MAD are more noisy than $S_{n,X}$ and exhibit very small, but visible peaks.

In general, at first glance, the volatility distributions of the robust estimators seem to follow very closely the distributions using the non-robust estimators from trimmed samples. Coming back to Tukey's triefficiency, let us remember (table 2.1), that the exponentially weighted standard deviation with 90% trimmed samples ($s_X, X^{(90)}$) and Rousseeuw and Croux's estimator $S_{n,X}$ have the two highest triefficiencies (65.0 and 54.7, respectively), and, as noted before, more accurate for the distribution of returns, the efficiency for the slash distributed data is the overall highest for $S_{n,X}$ (95.8) and the highest for $s_X, X^{(90)}$ among the non-robust estimators (76.1).

Figure 2.5 (a) directly compares the volatilities using the two estimators $s_X, X^{(90)}$ (blue line) and $S_{n,X}$ (red line). For this case, we have scaled both distributions to fit the interval $[0, 1]$ to compare the distributions independently from their overall levels. As we can see, the volatility distributions of the two estimators have indeed a remarkable similarity.[5] In panel (b) we have plotted the corresponding volatility regimes. For this purpose, we have set $k = 10$, i.e. we show the distribution of 10 volatility regimes

5 Appendix A: See figure A.2 for the same plot as in figure 2.5 (a), but with time windows $T = 20$ (one month) and $T = 250$ (one year). Here, the volatility distributions for the two estimators are very similar, too (low absolute deviations).

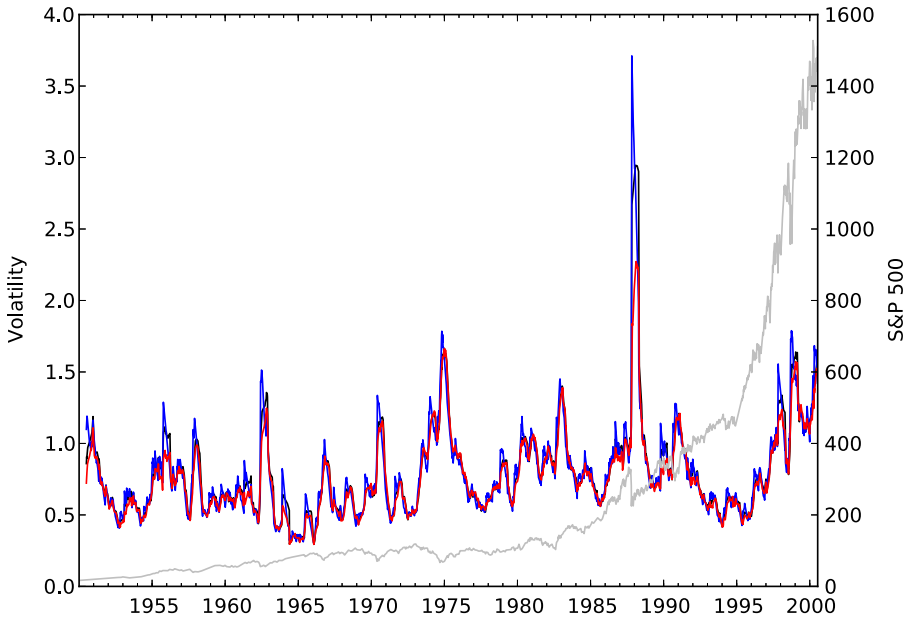
$V_1, ..., V_{10}$ for the two estimators of scale. Again, apart from a few dates, almost all dates fall into the same volatility regime for both estimators.

This corroborates our quantitative and qualitative assessment of both estimators as suitable estimators of scale for the distributions of financial returns, and thus as appropriate estimators for realised volatility. As we will discuss in chapter 4, we will be able to use both estimators to adjust drawdowns by the volatility at their time, and to segregate drawdowns that occur during different volatility regimes. Therefore, and due to the similarity of the volatilities with both estimators, we will tentatively pick $S_{n,X}$ as our scale estimator of choice.

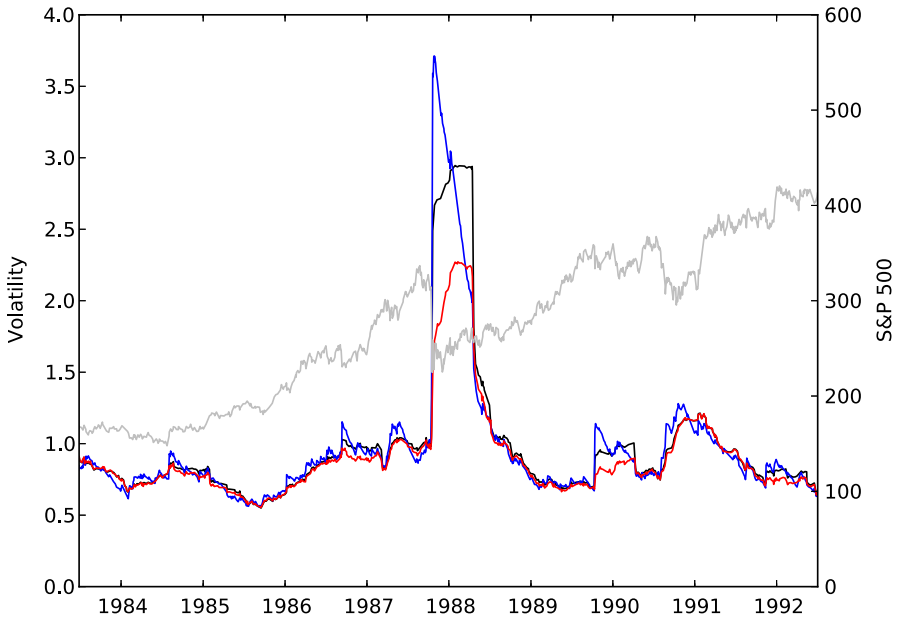
The choice of the size of the time window T is a bit more subtle. As we can see by comparing figures 2.5 (a), A.2 (a) and A.2 (b), the volatilities with a very small time window $T = 20$ are very noisy with strong peaks. Figure 2.4 shows the corresponding volatility regimes for $T = 20$ and $T = 250$ ($S_{n,X}$). We see that for the small time window $T = 20$, most of the events fall into the first two volatility regimes, quickly alternating between the two, without showing a moving tendency as for the bigger time window $T = 250$. Another important issue to keep in mind when choosing the size of the time window is the concept of drawdowns we are going to work with (see chapter 3). During a small time window $T = 20$ (approximately one month) only around 5 to 10 drawdowns, i.e. essential market events, occur (assuming a typical drawdown length of around 2 to 4 days). This small number of essential market events seems to be too small to represent and generate the general market dynamics. On the other hand, a long time window $T = 250$ has the potential to wash out short-term market dynamics, being influenced by market events at the far edge of the window.

Hence, we conclude that realised volatility time series constructed with very small time windows T do not reflect general the market dynamics and tendencies we are looking for in the distribution of the volatility. A robust and adequate compromise in terms of accuracy of short-term market dynamics and long-term tendencies is a time window of around half a year.[6]

6 Appendix A: Figure A.1 shows the relative deviations of the volatilities with $T = 20$ and $T = 250$ from our reference with $T = 125$, representing the errors we are accepting by choosing $T = 125$.

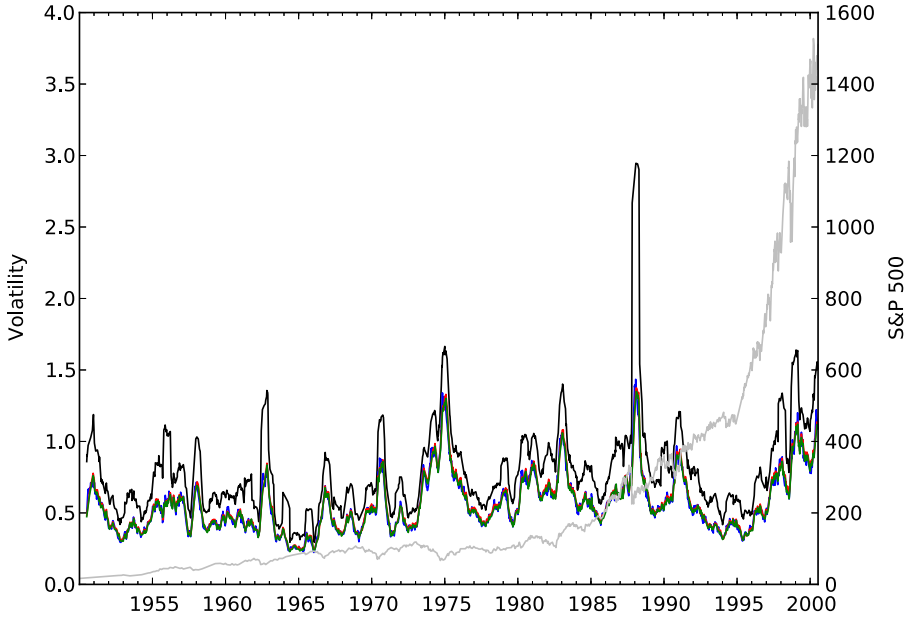


(a)

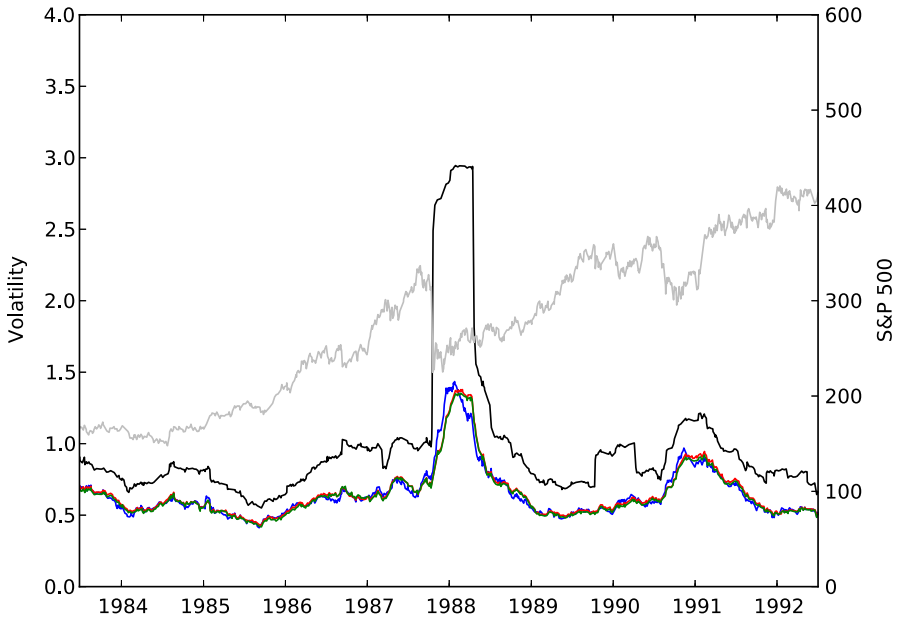


(b)

Figure 2.1: S&P 500 index (grey). Realised volatilities using different estimators of scale (time window $T = 125$): Sample standard deviation (black), exponentially weighted standard deviation with ω (2.13) (blue), and Gini's mean difference (red).

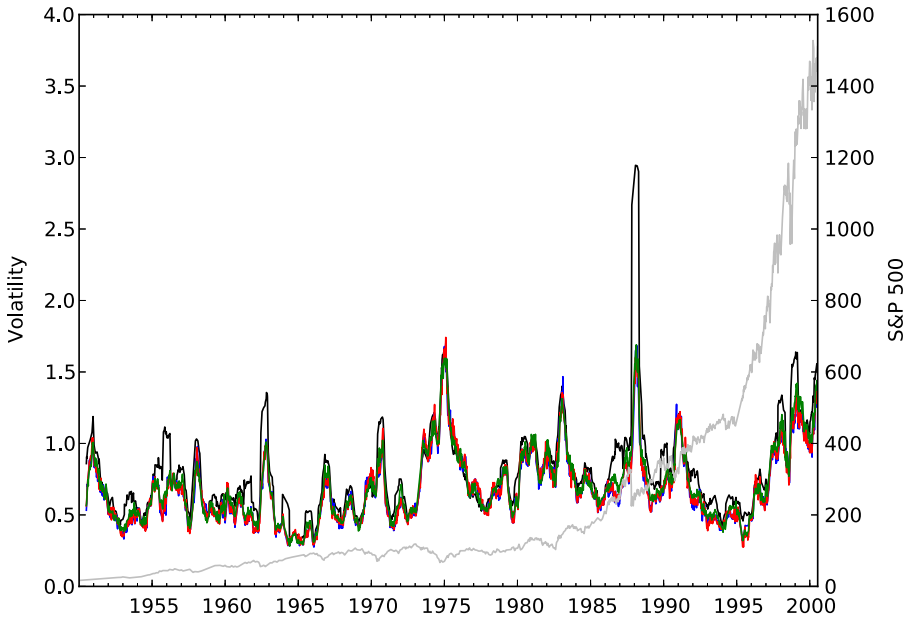


(a)

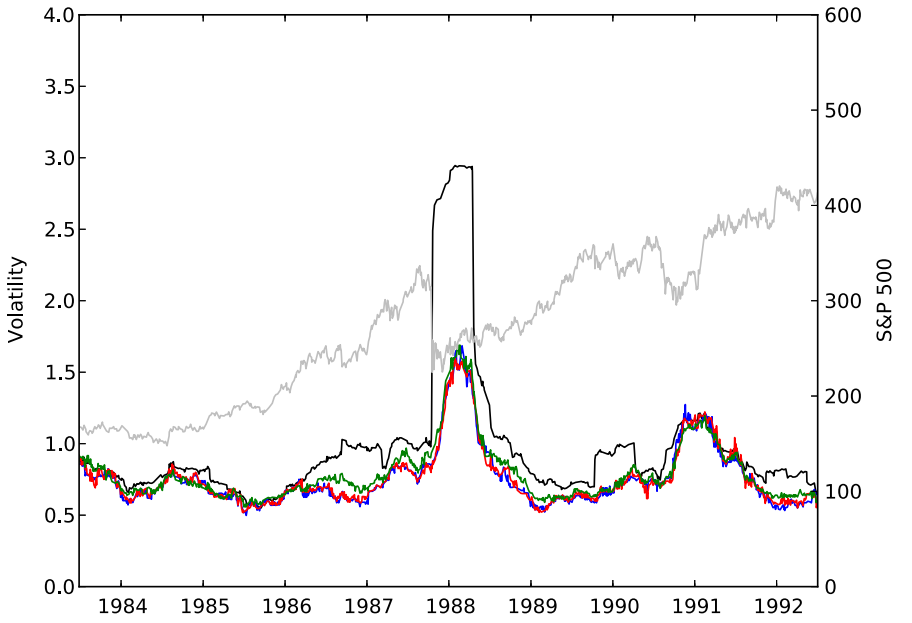


(b)

Figure 2.2: S&P 500 index (grey). Realised volatilities using different estimators of scale (time window $T = 125$) with 90% trimmed samples of returns: Sample standard deviation (non-trimmed, black), exponentially weighted standard deviation (blue), Gini's mean difference (red), and sample standard deviation (green).



(a)



(b)

Figure 2.3: S&P 500 index (grey). Realised volatilities using different estimators of scale (time window $T = 125$): Sample standard deviation (black), IQR (blue), MAD (red), and $S_{n,X}$ (green).

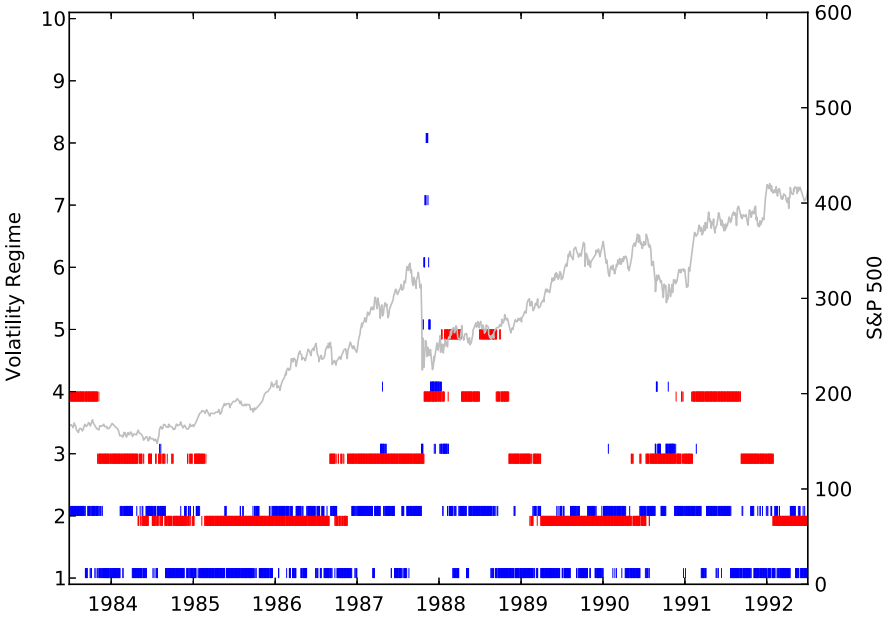
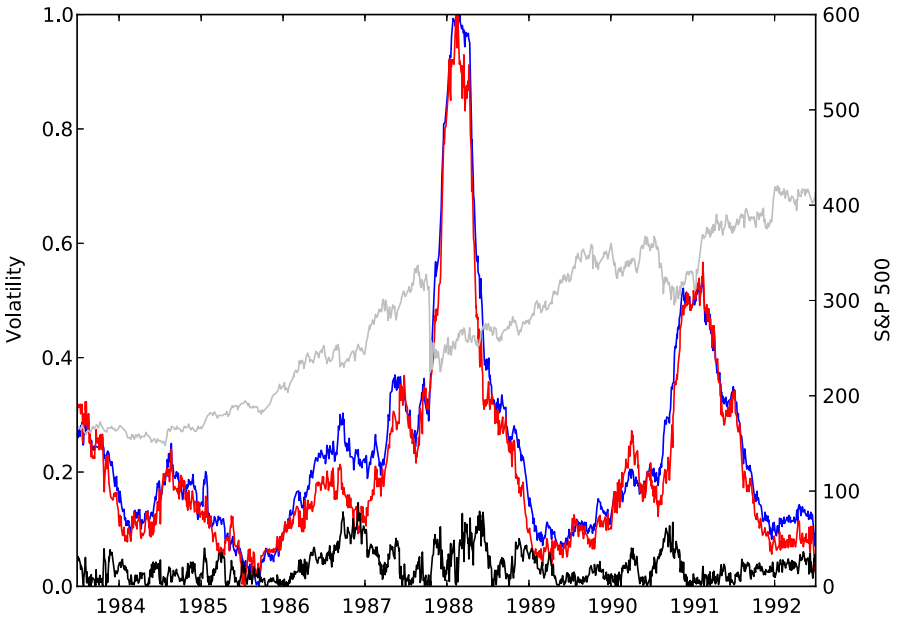
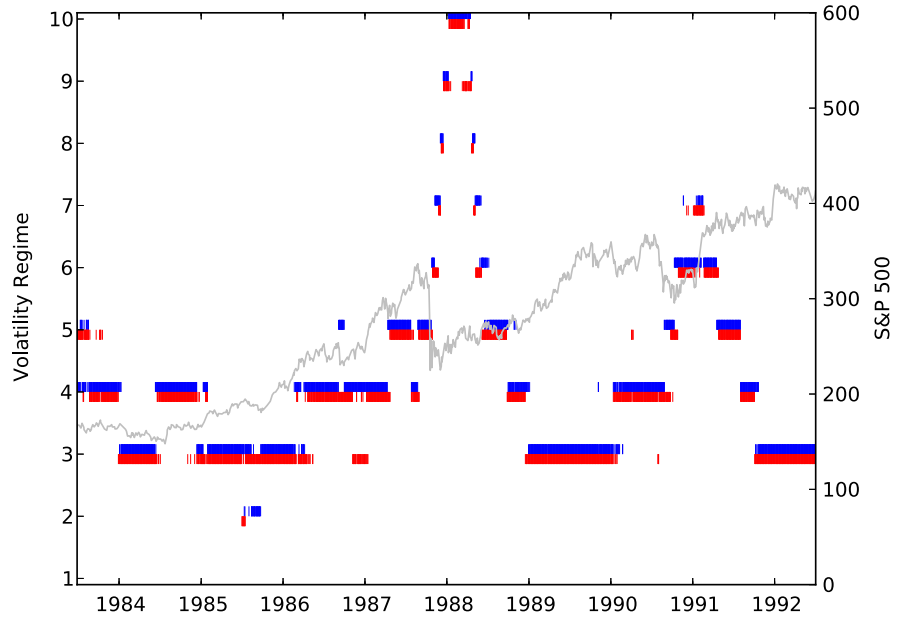


Figure 2.4: S&P 500 index (grey). Volatility regimes ($k = 10$) using the $S_{n,X}$ estimator of scale: Time window $T = 20$ (blue), and $T = 250$ (red).



(a)



(b)

Figure 2.5: S&P 500 index (grey). (a) Realised volatilities using two estimators of scale (time window $T = 125$): Exponentially weighted standard deviation with 90% trimmed samples (s_X, $X^{(90)}$, blue), and $S_{n,X}$ (red), as well as the absolute deviations between the two (black). The y-axis is scaled to $[0,1]$. (b) Corresponding volatility regimes with $k = 10$.

3 DRAWDOWNS

A drawdown on a daily time scale is defined as a continuous decrease in the price over consecutive days.[1] In other words, a drawdown is the cumulative loss from the last (local) maximum of the price to the next (local) minimum of the price. Since drawdowns are composed of returns with the same sign, they embody transient dependencies of successive returns and capture the effect successive losses have on each other. The distribution of drawdowns constructed in this way is therefore a persistent process of losses. Notice that following this definition all drawdowns are followed by drawups, which in turn are followed by drawdowns. This property of strict alternation illustrates the alternating flow of losses and gains investors are faced with on an elastic time scale.

By definition, the distribution of returns in a Bachelier-Samuelson [3] world does not capture this persistence, as it measures only the returns' frequency neglecting the relative positions of returns as they reveal themselves over time. The two-point correlation function has the same lack of information, as it measures an average linear dependence over the whole time series, while transient bursts of dependence may only appear at certain times, e.g. at very large runs. Hence, that feature will be washed out by the global averaging procedure (see [18, 41, 48] for an in-depth study).

In early studies [21, 24], drawdowns were simply defined as a continuous decrease of the price at each successive trading day (daily close), terminating a drawdown by any increase of the price. However, drawdowns constructed following this definition (*pure* drawdowns) are sensitive to noise, i.e. random uncorrelated and correlated fluctuations of the price. Simulations adding noise to the time series that were analysed, indicate that the distributions of drawdowns are robust to i.d.d. noise of "reasonable" magnitude [24].

A recent study [34] investigated distributions of drawdowns at smaller than daily timescales, i.e. constructing drawdowns from high-frequency 1,

1 The explicit reference to *drawups* will be neglected in this chapter and throughout this study, since they are only different in their sign and constructed in an analogous way to drawdowns. A drawup is thus a drawdown for a market agent with a short position on that market.

15 and 60 minutes data. However, it was found that distributions of drawdowns at daily timescale exhibit more valuable information about bursts of dependence than drawdowns from high-frequency data, as no evidence of a change of regime was found for high-frequency data. The results suggest that the positive feedback mechanisms that allegedly lie at the origin of the largest drawdowns require a certain time to build up and their results cannot be observed at high levels of granularity. Hence, in this study, our investigations will continue to focus on daily timescales.

In this chapter we formally define *pure* drawdowns (section 3.1) and present the concept of *coarse-grained* drawdowns to take noise into account (section 3.2). In section 3.3, we introduce a theoretical approach to the expected distribution of drawdowns that motivates our study of Dragon Kings in chapter 4.

3.1 PURE DRAWDOWNS

We define the return at time t in this context as

$$r_t = ln(p_{t+1}) - ln(p_t) \tag{3.1}$$

with the price p_t at time t in the time series. A *pure* drawdown ("p-drawdown") $D_{t,l}$ starting at time t with length l (number of days) is thus a sum of negative returns (r^-) in absolute terms:

$$D_{t,l} = -\sum_{i=0}^{l-1} r^-_{t+i} \tag{3.2}$$

Thus, at day $t + l$ we reach a local price minimum, which is the starting date of a new drawup.

3.2 COARSE-GRAINED DRAWDOWNS

There are two straightforward ways to define coarse-grained drawdowns, price coarse-grained and temporally coarse-grained drawdowns; previous research by Sornette and co-workers focused on price coarse-grained drawdowns. Here, we will briefly introduce the concept of temporally coarse-grained drawdowns, but will continue to focus on price coarse-grained drawdowns, with the introduction of a modification to the "traditional" approach.

3.2.1 Temporally coarse-grained drawdowns

In temporally coarse-grained drawdowns ("τ-drawdowns") price increases of any size within a certain time horizon τ may be ignored, i.e. drawdowns are only terminated after more than τ consecutive price increases regardless of magnitude. Consequently, for $\tau = 0$ we obtain pure drawdowns. Since the sensitivity to noise can be adjusted only in terms of the number of days, i.e. in integers, time horizons τ greater than one or two days with increases of *any* size may seem unreasonable to justify dependencies of successive returns. Due to this lack of a finer sensitivity to noise, this method will not be further discussed in this paper.

3.2.2 Price coarse-grained drawdowns

In price coarse-grained drawdowns ("ϵ-drawdowns") relative increases of the price below a certain threshold ϵ may be ignored, i.e. drawdowns are only terminated at positive price fluctuations above the threshold. In such a sequence of negative returns and conditional positive returns the local minimum will be selected to mark the ending of a drawdown, in order to maintain a dependency of successive negative returns. Consequently, for $\epsilon = 0$ we obtain pure drawdowns. An ϵ-drawdown $D_{t,l,\epsilon}$ is thus defined as

$$D_{t,l,\epsilon} = -\left(r_t^- + \sum_{i=1}^{l-2} r_{t+i}^{\pm|\epsilon} + r_{t+l-1}^-\right) \tag{3.3}$$

with strictly negative (r^-) starting and ending returns, and negative or conditional positive $(r^{\pm|\epsilon})$ returns inside the sequence of returns.

Fixed ϵ-threshold

The magnitude of the threshold ϵ lends room to various definitions. Previous research fixed the threshold throughout the time series and defined it in units of the global volatility:

$$\epsilon^f = \epsilon_0 \cdot \sigma_{glob} \tag{3.4}$$

The threshold is thus proportional to the volatility calculated for the entire series of returns, σ_{glob} (in previous research always using the standard deviation s_X). The coefficient ϵ_0 is chosen from the observation of data, but should be kept smaller than 1, since large fluctuations contribute to the global volatility. Hence, ϵ should be smaller than σ_{glob} to avoid breaks

of transient bursts of dependence we are looking for. It was shown in past studies [23] that the choice of setting ϵ_0 between 0 (pure) and 0.5 give reasonable and robust results that improve upon pure drawdowns.

Variable ϵ-threshold

If we accept the assumption of time-varying volatility, there is one major drawback when using coarse-grained drawdowns with a fixed threshold ϵ^f that is proportional to the global volatility calculated for the entire series of returns, σ_{glob}. Absolute returns in times with a realised volatility $\sigma_{t,T}$ below σ_{glob} are expected to be smaller than in times with a volatility above σ_{glob}. Therefore, while constructing drawdowns, the threshold will be exceeded less often in times where $\sigma_{t,T} < \sigma_{glob}$ than in times where $\sigma_{t,T} > \sigma_{glob}$. Hence, we expect drawdowns constructed in times where $\sigma_{t,T} < \sigma_{glob}$ to be longer than in times where $\sigma_{t,T} > \sigma_{glob}$. In figure 3.1 (a) we see that this is actually the case. Drawdowns at times with volatility below the global volatility (black line) are much longer than drawdowns at times with volatility above the global one.

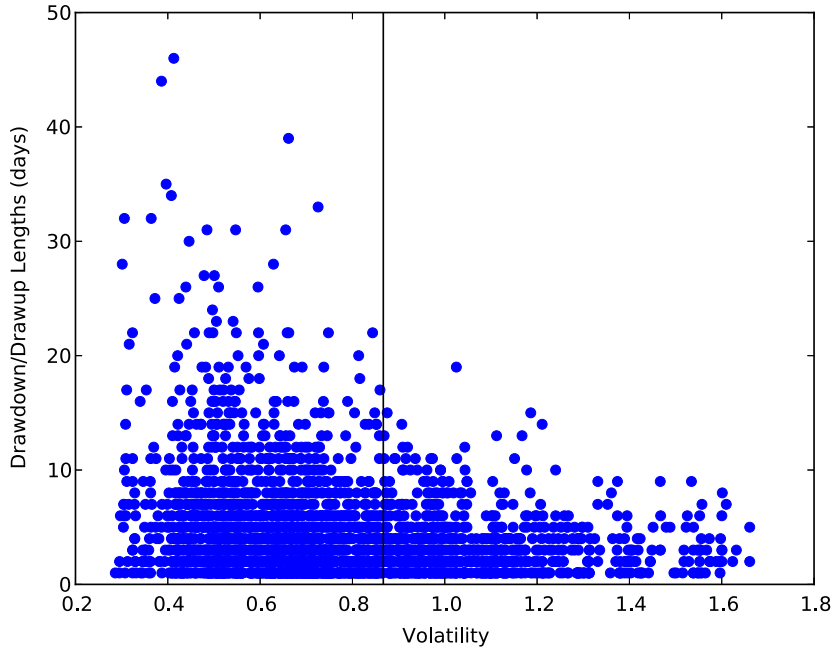
To cope with that problem, we introduce here a *variable* threshold in terms of the realised volatility at the time of a drawdown starting at time t, to adapt the threshold more accurately to the time-varying volatility:

$$\epsilon^v(t) = \epsilon_0 \cdot \sigma_{t,T} \tag{3.5}$$

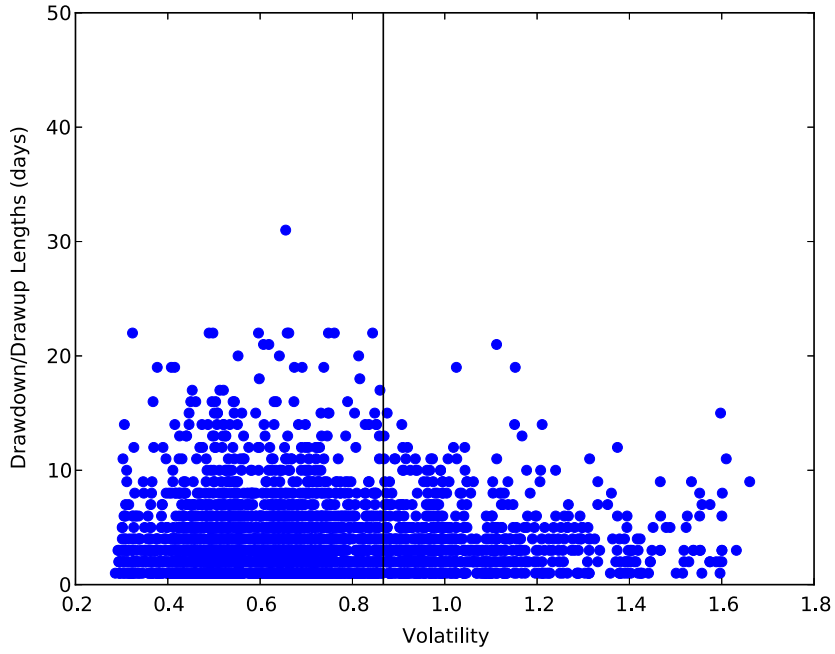
The variable threshold we define is thus proportional to the transient volatility $\sigma_{t,T}$ at time t as defined by (2.8), which indicates the magnitude of the volatility just *before* the start of a drawdown. Hence, the threshold ϵ^v will be fixed for the returns *inside* a drawdown, since these returns are dependent and should therefore correspond to the same volatility regime. The coefficient ϵ_0 will be chosen in an analogous way as for the fixed threshold.[2]

In figure 3.1 (b) we see that the lengths of the drawdowns and drawups are distributed more evenly than using the fixed threshold ϵ^f. The ϵ^v-drawdowns at times where $\sigma_{t,T} > \sigma_{glob}$ are now longer than ϵ^f-drawdowns, and shorter at times where $\sigma_{t,T} < \sigma_{glob}$. Hence, we can conclude that using the variable threshold ϵ^v to construct price coarse-grained drawdowns is an improvement upon the fixed threshold ϵ^f and coherent with the assumption of time-varying volatility.

2 We will denote the value of ϵ_0 in the index of ϵ^f and ϵ^v, e.g. $\epsilon^v_{0.5}$ is a variable threshold ϵ^v with $\epsilon_0 = 0.5$.



(a)



(b)

Figure 3.1: Lengths of fixed and variable threshold coarse-grained drawdowns and drawups at times with different realised volatility (S&P 500 index). (a) With fixed threshold $\epsilon^f_{0.5}$. (b) With variable threshold $\epsilon^v_{0.5}$. In both cases volatility $\sigma_{t,T}$ calculated with $S_{n,X}$ ($T = 125$). The black line marks the global volatility σ_{glob} (calculated with s_X).

3.3 DISTRIBUTION OF DRAWDOWNS

As discussed in this chapter, the concept of price drawdowns is salient for a direct measurement of the cumulative loss an investment decision can incur and to quantify the worst-case scenario of an investor going long at the local high (beginning of a drawdown) and going short at the next minimum (end of the drawdown). Therefore, it is important to know if there is any structure in the distribution of drawdowns. It is important to keep in mind, that, in contrast to returns, drawdowns are not defined over a fixed time scale, since they embody dependencies through the same sign variations. Drawdowns may last for only one day or for longer periods, such as a week. Their distribution measures how successive drops can influence each other and construct a persistent process.

It can be shown [24] that the distribution of drawdowns for independent returns is asymptotically exponential when the distribution of the independent returns does not decay more slowly than an exponential, i.e. belong to the class of exponential or super-exponential distributions. In contrast, for sub-exponentials (such as stable Lévy laws, power laws and stretched exponentials) the tail of the distribution of drawdowns is asymptotically the same as the distribution of the individual returns.

In several studies [1, 28, 32, 47, 49] stretched exponentials have been found to offer an accurate quantification of returns to capture a possible sub-exponential behaviour. Since stretched exponentials contain the exponential law as a special case (exponent $z = 1$ in (4.5), chapter 4) and by the fact that it has been successfully used as a model for the distribution of drawdowns, we shall take the stretched exponential law as our preliminary candidate distribution for drawdowns, with the the power law and lognormal distributions as possible alternative models [33]. In chapter 4 we will discuss the different models in more depth.

4 DRAGON KINGS

Dragon Kings in financial markets are extreme events in the distribution of drawdowns and drawups. As elaborated in chapters 1 and 3, drawdowns (and their counterpart drawups) are better adapted to capture the risk perception of market participants, and therefore they better reflect the realised market risks.

As analysed in [21–24], it can be demonstrated that the distributions of drawdowns efficiently diagnose financial crashes, i.e. special events associated with specific bubble regimes that precede them, which are seen as Dragon Kings.[1] About 99% of drawdowns can be represented nicely by a common distribution with a tail slightly fatter than an exponential distribution, whereas the remaining events have been found to be statistically different, i.e. the hypothesis that they belong to the same distribution as the 99% bulk of the population of the drawdowns is rejected at the 99.9% confidence level [22, 25]. It has been found that two-thirds of these Dragon Kings identified are stock market crashes, which were preceded by large bubbles.

It is important to stress that there is no unique methodology to diagnose Dragon Kings in general. They can be observed directly in the form of obvious breaks or bumps in the tail of size distributions, as in the example of the distribution of city sizes or material failure and rupture processes, or the distributions have to be compared at different resolution scales, as in the example of the distribution of turbulent velocity fluctuations. Another mechanism is found in the strong coupling regime of coupled heterogeneous oscillators of relaxation, where the statistics of epileptic seizures in human subjects and in animal models closely resemble that observed for earthquakes, i.e. Dragon Kings correspond here to so-called "characteristic earthquakes". One more way is the construction of new observables that capture more appropriately the dynamics of the system, as — in the case of our study — in the distribution of price drawdowns.

1 The evidence of Dragon Kings encompasses exchange markets, the major world stock markets, commodity markets and government bond markets. The results have been found robust with respect to various measures of pure and coarse-grained drawdowns.

The goal of our study is to corroborate the evidence of the existence of Dragon Kings in financial markets by taking into account the time-varying volatility. However, testing for Dragon Kings, or more generally for a change of regime in the population of drawdowns, is a subtle new problem that requires new techniques. In section 4.1 we present our approach to that problem with a set of tools to test for outliers or a change of regime. In section 4.2 we finally present the innovative approach on Dragon Kings in financial markets by taking the time-varying volatility into account. Specifically we will present two different approaches, first, a straightforward adjustment, and second, an approach by segregating and diagnosing Dragon Kings that occur at different volatility regimes.

4.1 TOOLS TO TEST FOR DRAGON KINGS

In this section we present a set of tools to test generally for a change of regime in the population of drawdowns to provide evidence for the existence of Dragon Kings. Most of these tools have already been successfully used in approaching this problem [21–24, 32, 34].

In section 4.1.1 we provide a starting point by looking at the complementary cumulative distribution function of drawdowns. In section 4.1.2 we present a parametric statistical test to discriminate between the null hypotheses stretched exponential or power law and modifications of these models that describe deviations in the tail. In section 4.1.3 we apply the uniformly most powerful unbiased test between the lognormal and the power law, which has been successfully applied in the debate about the distribution of city sizes [33] and present Hill's inverse tail index estimate.

4.1.1 Complementary cumulative distribution function

A starting point in the analysis of the distribution of drawdowns to diagnose Dragon Kings is to check for underlying models that describe the form of the distribution as closely as possible. Obvious breaks or bumps in the tail of the distributions may then be detected directly in the form of deviations and statistical outliers.

In our case, we will study the complementary cumulative distribution function (CCDF) of (the absolute size of) drawdowns. Specifically, this distribution function expresses how often the random variable of drawdowns is above a particular level, or — in terms of "survival function" — it cap-

tures the probability that the emergence of bursts of dependence that lead to the creation of drawdowns will survive beyond a specific size. The complementary cumulative distribution function is defined as

$$F_c(x) = P(X > x) = 1 - F(x) \tag{4.1}$$

where $F(x)$ is the cumulative distribution function of drawdowns. This function is monotonically decreasing and $F_c(0) = P(X > 0) = 1$.

In particular, the analysis of the complementary cumulative distribution function of drawdowns can be split into two steps, first, the classification of unknown distributions, and second, the detection of obvious deviations and outliers.

Classification of unknown distributions

The semi-logarithmic and logarithmic representations of $F_c(x)$ can be a very powerful first tool to classify the unknown distributions in order to assign appropriate underlying models. As discussed in chapter 3, possible candidate models are particularly the power law and the stretched exponential model, as well as conceivably the lognormal model.

A distribution of drawdowns following a power law model would appear as a straight line in the log-log representation and as convex curve in the log-linear representation. On the other hand, a distribution following an exponential model would appear as a straight line in the log-linear representation, while the stretched exponential with an exponent < 1 would appear as a convex curve.

Detection of obvious deviations and outliers

The complementary cumulative distribution function is the simplest tool to discriminate "by eye" different regimes in the population of drawdowns. These changes of regime can be suspicious breaks and bumps that qualify as obvious outliers and Dragon Kings. Furthermore, a change of regime could be indicated by a deviation from the bulk of the population in the distribution, suggesting that Dragon Kings obey a different distribution than the bulk. After classifying the distribution for some possible underlying model, more sophisticated and adequate (parametric) statistical tests can be applied to discriminate different regimes.

4.1.2 Likelihood-ratio test for nested hypotheses

Here, we formulate a test in the spirit of [11, 22]. This test is aimed at the question of whether there is a threshold rank of drawdowns below or above which the null stretched exponential model or power law model can be rejected. In other words, in order to qualify the existence of outliers in the distribution of drawdowns, we perform this test tailored to estimate the significance of the curvature seen in the distributions (in the log-linear or log-log representation respectively).

Let us formally introduce the general framework of hypothesis testing within a parametric formulation.

The sample $X_1, ..., X_n$ has a probability density distribution (PDF) $p(x|\theta)$, where θ is a vector corresponding to the set of free parameters in the PDF. In general, we consider two hypotheses corresponding to two sets of parameters $\theta = (\theta_1, ..., \theta_k)$:

H_1: The parameters $\theta_1, ..., \theta_k$ belong to some k-dimensional parameter space Θ_1: $\theta \in \Theta_1$

H_0: One of the parameters $\theta_1, ..., \theta_k$ is equal to zero, whereas the other parameters can vary in the same $(k-1)$-dimensional space as in H_1. We denote this subset in parameter space as Θ_0: $\theta \in \Theta_0$. Evidently, Θ_1 contains Θ_0.

Let us denote the maximum likelihood under H_i as L_i $(i = 0, 1)$:

$$L_i = \max_{\theta \in \Theta_i} [p(X_1|\theta)...p(X_n|\theta)] \tag{4.2}$$

where the maximum is taken over θ in the parametric space Θ_i.

By construction $L_0 \leq L_1$ since adding one or several parameters cannot decrease the quality of the fit to the data. A theorem by Wilks [40] states that the test statistic T

$$T = -2 \ln \Lambda \tag{4.3}$$

with the maximum likelihood ratio $\Lambda = L_0/L_1$, which is asymptotically distributed as χ^2 with one degree of freedom as n tends to infinity. Using the log-likelihood $\ln L_0$ and $\ln L_1$ we can formulate (4.3) as

$$T = -2(\ln L_0 - \ln L_1) \tag{4.4}$$

Thus, the test is based on this test statistic T by comparing H_0 and H_1. If T is large, L_1 is significantly larger than L_0 which means that adding a parameter significantly improves the quality of the fit. On the other hand, if T is

small, L_1 is not much larger than L_0 which means that the additional parameter does not much improve the fit. The corresponding p-value gives the probability of exceeding T by chance.

Let us now introduce the specific tests for our stretched exponential and power law models.

Stretched exponential model

The stretched exponential for the sample of drawdowns $X_1, ..., X_n$ (rank 1 biggest drawdown, rank n smallest drawdown) corresponds to a straight line in the log-linear representation of $\ln P$ versus x^z with the cumulative distribution

$$P(x|\theta) = P(x = 0) \exp\left(-Bx^z + Cx^{2z}\right) \tag{4.5}$$

where $\theta = (B, z, C)$. Accordingly, the "pure" stretched exponential distribution (SE) corresponds to the case where $C = 0$.

The choice of the parametrisation (4.5) with a correction Cx^{2z} where the exponent is twice that of the first term in the exponential is taken to avoid introducing two additional parameters and as the natural measure of a curvature in the log-linear plot of $\ln P$ versus x^z that would qualify the simple stretched exponential as a straight line. We will refer to that general case where $C \neq 0$ as modified stretched exponential (MSE). We thus have our hypotheses

H_0: $C = 0$ (SE)
H_1: $C \neq 0$ (MSE)

and we define

$$P_{SE}(x) = A_{SE}(t) \exp\left(-Bx^z\right) \tag{4.6}$$
$$P_{MSE}(x) = A_{MSE}(t) \exp\left(-Bx^z + Cx^{2z}\right) \tag{4.7}$$

as two complementary cumulative distribution functions of drawdowns defined for the drawdowns at rank t (here, lower threshold rank) up to the maximum rank n. This corresponds to the interval $[0, X_t]$ of drawdowns.

Hence, the corresponding density distribution functions are $p_{SE}(x) = -dP_{SE}(x)/dx$ and $p_{MSE}(x) = -dP_{MSE}(x)/dx$. The normalising factors $A_{SE}(t)$ and $A_{MSE}(t)$ are different and functions of the lower cut-off rank

t since $p_{SE}(x)$ and $p_{MSE}(x)$ must be normalised to 1 in the interval $[0, X_t]$. This normalisation condition gives

$$A_{SE}(t) = \frac{1}{1 - \exp\left(-BX_t^z\right)} \tag{4.8}$$

$$A_{MSE}(t) = \frac{1}{1 - \exp\left(-BX_t^z + CX_t^{2z}\right)} \tag{4.9}$$

The maximum likelihood estimation of the best parameters for the candidate stretched exponential model (4.6) is done as a minimisation of

$$\begin{aligned} -\ln L_0 &= -\sum_{i=t}^{n} \ln p_{SE}(X_i) \\ &= -\sum_{i=t}^{n} \left[\ln A_{SE}(t) + \ln\left(BzX_i^{z-1}\right) - BX_i^z\right] \end{aligned} \tag{4.10}$$

with respect to B and z. Similarly, the estimation for the candidate modified stretched exponential model (4.7) is done as a minimisation of

$$-\ln L_1 = -\sum_{i=t}^{n} \left[\ln A_{MSE}(t) + \ln\left(BzX_i^{z-1} - 2CzX_i^{2z-1}\right) - BX_i^z + CX_i^{2z}\right] \tag{4.11}$$

with respect to B, z and C. Finally, the maximised log-likelihood $\ln L_0$ and $\ln L_0$ will be used in (4.4) to determine the test statistic T and the corresponding p-value.

Technically, the minimisation is done with the downhill simplex minimisation algorithm. In order to secure that the maximum likelihood estimation does indeed retain the parameter values of the global maximum, the downhill simplex minimisation algorithm will be applied with a wide range of start values in the search.[2]

Power law model

In the spirit of the stretched exponential model, we now introduce a similar test based on the power law model, where a power law in the tail corresponds to a straight line in the log-log representation of $\ln P$ versus $\ln x$.

For convenience we will introduce a different notation. We take the logarithm of the drawdowns D_i, such that $X_i' = \ln D_i$. Note, that for this model we are going to use an *upper* threshold rank t, compared to the

2 We use a total of 1000 combinations of start values for B, z and C.

lower threshold for the stretched exponential model, as we are going to look only at the drawdowns in the far tail. Hence, we are going to shift the truncated sample of t log-drawdowns $X'_1, \ldots, X'_t$ to $X_i = X'_i - X'_t$, such that $X_t = 0$. Our sample of log-drawdowns is thus $X_1, \ldots, X_t$ with the cumulative distribution

$$P(x|\theta) = P(x = 0) \exp\left(-Bx + Cx^2\right) \tag{4.12}$$

where $\theta = (B, C)$. Accordingly, the "pure" power law distribution (PL) corresponds to the case where $C = 0$.

The choice of the parametrisation (4.12) with a correction Cx^2 is analogous to the one for the stretched exponential model. We will refer to the general case where $C \neq 0$ as modified power law (MPL). We thus have the hypotheses H_0: $C = 0$ (PL) and H_1: $C \neq 0$ (MPL). We define

$$P_{PL}(x) = A_{PL} \exp\left(-Bx\right) \tag{4.13}$$

$$P_{MPL}(x) = A_{MPL} \exp\left(-Bx + Cx^2\right) \tag{4.14}$$

as two complementary cumulative distribution functions defined for the log-drawdowns at rank 1 up to the cut-off rank t. This corresponds to the interval $[X_t = 0, X_1]$ of log-drawdowns with the normalising factors

$$A_{PL} = \frac{1}{1 - \exp\left(-BX_1\right)} \tag{4.15}$$

$$A_{MPL} = \frac{1}{1 - \exp\left(-BX_1 + CX_1^2\right)} \tag{4.16}$$

Thus, the maximum likelihood estimation of the best parameters of the candidate models (4.13) and (4.14) is done as a minimisation of

$$-\ln L_0 = -\sum_{i=1}^{t} \left(\ln A_{PL} + \ln B - BX_i\right) \tag{4.17}$$

$$-\ln L_1 = -\sum_{i=1}^{n} \left[\ln A_{MPL} + \ln\left(B - 2CX_i\right) - BX_i + CX_i^2\right] \tag{4.18}$$

with respect to B, and B, C, respectively.

The minimisation will be done with the same minimisation algorithm as for the stretched exponential model. Likewise, the test statistic T and the corresponding p-value will be determined with (4.4) using the maximised log-likelihood $\ln L_0$ and $\ln L_0$.

4.1.3 Power law versus lognormal distribution

The power law and the lognormal distribution are often difficult to distinguish. Although both distributions exhibit a distinct behaviour in the tails, the lognormal can easily be mistaken for a power law over a range which can cover several decades [33]. Both distributions may be generated by Gibrat's law of proportional growth, which can be generalised by a general class of self-similar fragmentation processes [27], yet with some additional apparently innocuous but actually profound intricacy for the power law. Power law distributions are regularly varying, whereas the limit behaviour of lognormal distributions characterises a rapidly decreasing function at infinity. Therefore, both distributions exhibit qualitatively different behaviours in their upper tails. The lognormal density in the upper tail goes to zero faster than any power law density. However, writing the lognormal density as

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\mu^2}{2\sigma^2}} \cdot x^{-1+\frac{\mu}{\sigma^2}-\frac{\ln x}{2\sigma^2}} \tag{4.19}$$

we observe that the lognormal distribution is superficially like a power law with a slowly increasing effective exponent

$$\alpha(x) = \frac{1}{2\sigma^2} \ln\left(\frac{x}{e^{2\mu}}\right) \tag{4.20}$$

This shows us that the lognormal distribution decays at infinity faster than any power law, since the apparent exponent $\alpha(x)$ diverges with x. Hence, with σ^2 large enough, $\alpha(x)$ varies so slowly as to give the impression of constancy over several decades in x.

Since we are interested in extreme drawdowns, i.e. in the behaviour of the tail of the distributions, we present in this section Hill's inverse tail index estimate $\hat{\alpha^{-1}}$ and a uniformly most powerful unbiased test in the spirit of [33] aimed to discriminate between the power law and the lognormal distribution at a certain threshold. A tail in the distribution of drawdowns following a power law suggests that the tail is fatter than in the case of a lognormal distribution and that particularly large drawdowns occur far more often than a lognormal distribution would suggest. In this case, Dragon Kings would follow a power law, whereas the bulk of drawdowns would follow a lognormal distribution.

Hill's inverse tail index estimate

Hill's inverse tail index estimate $\hat{\alpha^{-1}}$ for power law distributions is the best unbiased estimator for the inverse of the tail index [17].[3] It is defined as

$$\hat{\alpha^{-1}} = \frac{1}{t}\sum_{i=1}^{t} \ln D_i - \ln D_{t+1} \tag{4.21}$$

with the upper threshold drawdown rank t.

Plotting Hill's estimate as a function of the threshold rank t will help us to discriminate between the power law and the lognormal distribution in the tail up to the threshold. An approximately constant estimate $\hat{\alpha^{-1}}$, i.e. a plateau in the plot, will then reinforce the validity of a power law. On the other hand, a decaying estimate will reinforce a deviation from the power law in the tail or in the higher ranks.

Uniformly most powerful unbiased test

The test that addresses the question, whether the power law or lognormal model holds in the tail, considers the null hypothesis that, beyond some threshold t, the upper tail of the distribution is power law (PL) distributed, against the alternative that it is (truncated) lognormal (LN) distributed:

$$H_0\colon\ f_0(x) = f_{PL}(x;\alpha)$$

$$H_1\colon\ f_1(x) = f_{LN}(x;\alpha,\beta)$$

with

$$f_{PL}(x;\alpha) = \alpha\frac{t^\alpha}{x^{\alpha+1}} \tag{4.22}$$

for $x \geq t$ with $\alpha > 0$, and

$$f_{LN}(x;\alpha,\beta) = \left[\sqrt{\frac{\pi}{\beta}}e^{\frac{\alpha^2}{4\beta}}\left(1-\Phi\left(\frac{\alpha}{\sqrt{2\beta}}\right)\right)\right]^{-1}\cdot\frac{1}{x}e^{-\alpha\ln\frac{x}{t}-\beta(\ln\frac{x}{t})^2} \tag{4.23}$$

for $x \geq t$ with $\alpha \in \mathbb{R}$, $\beta > 0$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the normal distribution.

Note, that this is equivalent to testing the null hypothesis that the upper tail of the distribution of the logarithm of the drawdowns D_i is exponential against the alternative that it is a (truncated) normal. For this problem the clipped sample coefficient of variation $\hat{c} = \min(1, c)$ provides the

3 It is not possible to get an unbiased estimate for α.

uniformly most powerful unbiased (UMPU) test [33] (for a detailed derivation of this test, please refer to [9, 13]). The sample coefficient of variation $c = s_X/\bar{x}$ is the ratio of the sample standard deviation to the sample mean, with the sample of log-drawdowns $X_i = \ln D_i$.

The critical point of the test can be derived with very high accuracy by a saddle point approximation. The likelihood equations can be reduced to

$$c^2 = \frac{-\gamma h(\gamma) + \gamma^2 + 0.5}{(h(\gamma) - \gamma)^2} - 1 \tag{4.24}$$

with $\gamma = \alpha/\left(2\sqrt{\beta}\right)$ and

$$h(x) = \frac{\exp(-x^2)}{2\sqrt{\pi}\left(1 - \Phi\left(\sqrt{2}x\right)\right)} \tag{4.25}$$

The left part of (4.24), c^2, is the empirical squared coefficient of variation, whereas the right part is the model squared coefficient of variation C^2. The solution of (4.24) is $\hat{\gamma} = \hat{\gamma}(c)$. The test statistic W^* can then be derived as

$$W^* = W(\hat{\gamma}) + 2L(\hat{\gamma}) + \frac{L(\hat{\gamma})^2}{W(\hat{\gamma})} \tag{4.26}$$

with

$$W(\hat{\gamma}) = n\left[2\ln\left(2h(\hat{\gamma})\left(h(\hat{\gamma}) - \hat{\gamma}\right)\right) + 2\hat{\gamma}^2 - 2\hat{\gamma}h(\hat{\gamma}) + 1\right] \tag{4.27}$$

for $c < 1$, $W(\hat{\gamma}) = 0$ for $c > 1$, and

$$L(\hat{\gamma}) = \frac{1}{2}\ln\left[\frac{\hat{\gamma}h(\hat{\gamma})\left(2\hat{\gamma}^2 - 4\hat{\gamma}h(\hat{\gamma}) + 2h(\hat{\gamma})^2 + 3\right) - 3h(\hat{\gamma}) + 1}{\frac{4}{n}\left(h(\hat{\gamma}) - \hat{\gamma}\right)^2 W(\hat{\gamma})}\right] \tag{4.28}$$

It can be established that the asymptotic distribution of W^* is a 50:50 mixture of the constant 0 and χ^2 distribution with one degree of freedom.

4.2 DRAGON KINGS AND VOLATILITY

In this section we systematically extend the methodology presented to diagnose Dragon Kings to empirically answer the question whether the effect of time-varying volatility on drawdowns and Dragon Kings has the

potential to enhance the previous methodology or is in line with previous results. As mentioned in section 4.1, the methodology and tools we presented can be performed independently from our innovative approaches of taking time-varying volatility into account or following the "traditional" approach. They can be generally applied to any kind of distributions of drawdowns. Therefore, we will apply these tools to two approaches. First, in section 4.2.1 we introduce a straightforward approach by adjusting drawdowns, second, in section 4.2.2 we introduce the concept of grouping drawdowns that happened during the same volatility regimes together.

4.2.1 Distributions of adjusted drawdowns

The most straightforward approach, and similar to Le Bris' idea we have shown in chapter 1, is to construct distributions of volatility-adjusted drawdowns. For this case, we define adjusted drawdowns D_t^* as

$$D_t^* = \frac{D_t}{\sigma_{t,T}} \tag{4.29}$$

with the unadjusted drawdown D_t starting at time t and the realised volatility $\sigma_{t,T}$ at that time over the time window T as defined in (2.8). Since we defined the realised volatility at time t over a time window T without taking the return *at* time t into account, the drawdown D_t will be adjusted by the volatility just before the starting return at time t. Thus, the volatility used to adjust a drawdown will not be "contaminated" by the proper drawdown.[4] The idea of this adjustment is very simple: A drawdown at a time of high volatility will be smaller after adjustment than a comparable drawdown at a time of low volatility, since a high level of volatility indicates a generally higher level of risk, and thus, an acclimatised perception of smaller loss. The aim of this is to detect Dragon Kings we could not detect without adjustment, i.e. drawdowns of absolute smaller size with low volatility that disappeared in the bulk of the distribution. After adjustment, these drawdowns would appear to be outliers of bigger event size (lower ranks), and thus, qualify for Dragon Kings. We see this effect of the adjustment for drawdowns with different volatilities in figure 4.1. Panel (a) shows the drawdowns per volatility before adjustment, and panel (b) shows the same drawdowns after adjustment, which can be regarded as a division by the abscissa.

4 The subtraction of the average return μ_r (in Le Bris' normalisation, chapter 1) from D_t can be safely ignored here, since it is close to zero and only a very small fraction of a drawdown composed of several returns.

Hence, constructing distributions of adjusted drawdowns D_t^* and analysing them with the tools from section 4.1, will enable us to directly compare our results, where we take the volatility in form of adjustment into account, with the previous results.

4.2.2 Drawdowns during different volatility regimes

A different approach, but with the same aim, is to look separately at drawdowns during different volatility regimes. Here, we do not just adjust the size and rank ordering of all drawdowns to better reflect their different levels of volatility, but we completely segregate them according to their volatility. What we called before the population of drawdowns, will be subdivided into smaller populations composed of drawdowns with the same level of volatility, i.e. occurring during the same volatility regimes. In these sub-populations then, we will perform our Dragon King detection methodology to find changes of (population) regimes *per* volatility regime. In other words, we ask if each population of drawdowns occurring during the same volatility regime (our sub-population) is composed of a bulk and a Dragon King population. On the other hand, this would show us that the Dragon King regimes we find in the entire population of drawdowns are just composed of a few extreme cases — mostly during high volatility regimes where extreme cases have to be relatively bigger in size. As a result, these few extreme cases just happen to appear as a different regime that is not composed of all extreme cases, including relatively "smaller" extreme cases during low volatility regimes.

Formally we will assign to every drawdown D_t a volatility regime V_i, as defined in (2.16), such that the volatility σ_t at the time of the beginning of the drawdown belongs to V_i ($i = 1, \ldots, k$). We will then construct k distributions of drawdowns that have the same volatility regime V_i assigned.

If the simple adjustment is not powerful enough to make the relatively smaller cases more pronounced as outliers, and the distribution of drawdowns just gets washed out or flattened, we hope that the separation of drawdowns by volatility regime will be able to provide support for the concept of Dragon Kings in financial markets, by clearly showing us that certain events of absolute smaller size are in fact extreme events, too. However, a major drawback of this approach is that the distributions will be composed of fewer drawdowns, and as a consequence, the Dragon King regimes would be composed of a much smaller number of drawdowns than in the case where we take the entire population. Our parametric tools (sections 4.1.2 and 4.1.3), where we check for a particular underlying model in

Table 4.1: Quantities (#) and ratios (%) of drawdowns in each volatility regime V_i ($k = 3, 5, 10$) with centre volatility value c_i (S&P 500 index). In all cases $\epsilon^v_{0.5}$ and volatility calculated with $S_{n,X}$ ($T = 125$).

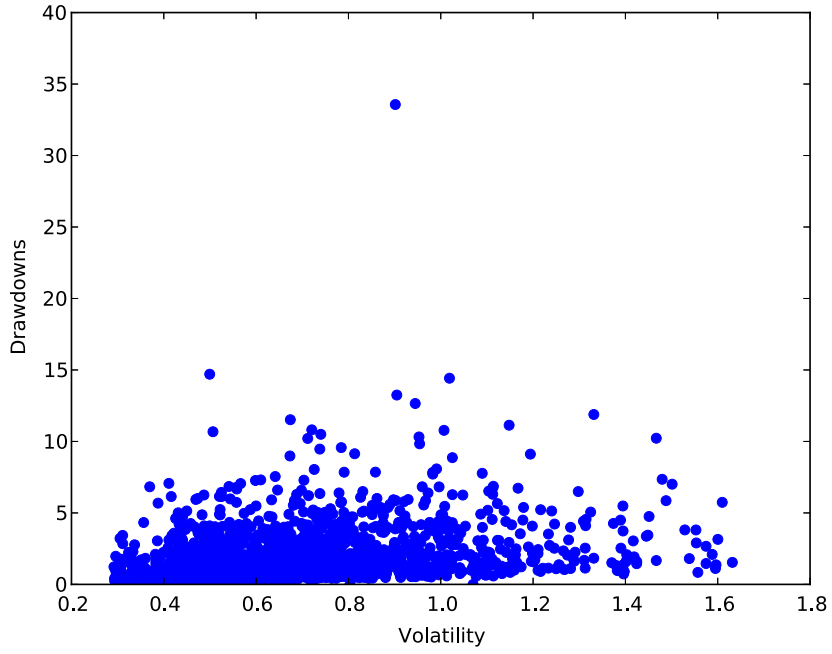
VOLATILITY REGIMES		DRAWDOWNS	
V_i	c_i	#	%
1	0.52	1126	66.7
2	0.99	494	29.3
3	1.46	68	4.0
1	0.42	591	35.0
2	0.70	687	40.7
3	0.99	302	17.9
4	1.27	83	4.9
5	1.55	25	1.5
1	0.35	135	8.0
2	0.49	456	27.0
3	0.63	394	23.3
4	0.77	293	17.4
5	0.92	195	11.6
6	1.06	107	6.3
7	1.20	51	3.0
8	1.34	32	1.9
9	1.48	14	0.8
10	1.62	11	0.7

the upper tail beyond a certain threshold, could simply fail because of too small samples. Hence, we expect to obtain suitable results only through the visual inspection of the complementary cumulative distribution function (section 4.1.1).

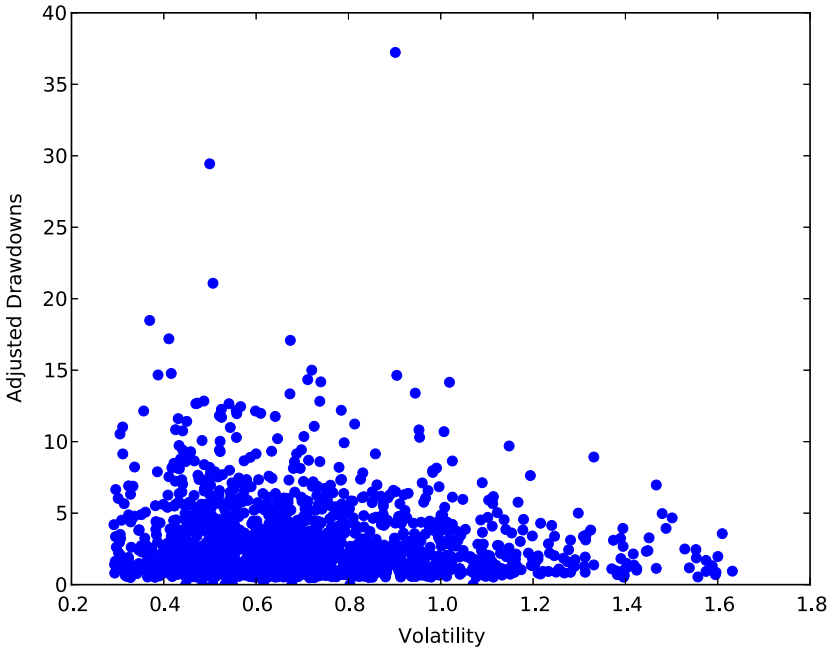
In this procedure, the number k of volatility regimes has to be freely chosen. The number should be as big as necessary to have enough different levels that appropriately reflect the variability of the volatility, but as small as possible to have enough drawdowns in each volatility regime. We will tentatively choose $k = 5, 10$. As a possible extension to this approach, and to overcome the aforementioned major drawback, we could chose a very small number of volatility regimes, e.g. $k = 3$, to avoid very small numbers of drawdowns in the volatility regimes, and then adjust the drawdowns within each volatility regime. This would in fact be a combination of our two approaches, by first coarsely separating all drawdowns into

"low", "middle" and "high" volatility regimes ($k = 3$), and then adjusting them on an individual basis by their proper volatility.

Table 4.1 lists the quantities and ratios of drawdowns in each volatility regime, for the three cases $k = 3, 5, 10$ (S&P 500 index from 1950 to 2000, sample size of 12718 returns, with a total of 1688 drawdowns, compare also with figure 4.1). As we can see, already for $k = 3$ the high volatility regime (V_3) consists only of a very small number of drawdowns (68, 4.0%), but two-thirds (1126, 66.7%) of all drawdowns belong to the low volatility regime (V_1). Here, an adjustment within the three regimes could be appropriate to further pronounce outliers within a big sub-population of 1126 drawdowns. For $k = 10$ the upper four regimes consist of less than 100 drawdowns. The choice $k = 5$ seems to be the best compromise of quantities per volatility regime and distribution of different volatility levels, where a maximum of 40.7% belong to one regime (V_2). Hence, we conclude that the choice of k gives room for adjustment to specific time series, taking $k = 5$ as a starting point. With small numbers of volatility regimes, i.e. large quantities of drawdowns within one volatility regime, an adjustment within the regimes could be appropriate.



(a)



(b)

Figure 4.1: Non-adjusted and adjusted drawdowns at times with different realised volatility (S&P 500 index). (a) Before adjustment. (b) After adjustment by the volatility. In both cases $\epsilon^{\nu}_{0.5}$ and volatility $\sigma_{t,T}$ calculated with $S_{n,X}$ ($T = 125$).

5 DATA ANALYSIS

After having introduced the theoretical frameworks for volatility (chapter 2), drawdowns (chapter 3) and Dragon Kings (chapter 4), we will apply them on real financial data.

The focus of the data analysis will be to investigate the effect of taking time-varying volatility into account on the detection of Dragon Kings. We will compare these results with results obtained without taking volatility into account by comparing the quantities and the actual Dragon Kings diagnosed. An important question we will try to answer is how pronounced these outliers appear in the distributions of drawdowns, i.e. how distinctive the changes of regime manifest themselves in the populations of drawdowns, since the more prominent the outliers are, or the more pronounced the deviations appear in the tails of the distributions, the more it will help us to develop a better perception for extreme financial risks. Before discussing the results of the analysis in section 5.3, we will describe the time series data we will analyse in section 5.1, and define the combinations of parameters and tools we use for the analyses in section 5.2.

5.1 DATA

In our study to diagnose Dragon Kings taking time-varying volatility into account, we will use a selection of eight daily time series, covering a large part of the financial markets, i.e. major stock market indices, foreign exchange, government bonds and commodities. Below, we will briefly describe the time series we consider for this study. In table 5.1 the time series are listed with starting and ending dates, as well as the sample sizes (number of price observations).

Stock market indices

We will study three major stock market indices (daily close) from around the world, namely the S&P 500 (U.S.), FTSE 100 (U.K.) and the Hang Seng Index (Hong Kong). Out of these indices, the HSI is the most volatile

index, while the S&P 500 appears to be the most stable index with the lowest average volatility. The HSI's average volatility is around 75% and 45% higher than the S&P 500 and FTSE 100, respectively.[1] It is worthwhile to note, that the HSI is composed of only 43 companies, compared to the other two indices with 100 and 500 constituent companies.

Foreign exchange

The foreign exchange market is by far the largest market in terms of volume. Of all possible pairs of currencies we will study the Japanese Yen (JPY) in currency units per U.S. dollar (USD), acknowledging the leading role of the U.S. dollar. Under the Bretton Woods system established after World War II, only the U.S. dollar had a direct gold parity, whereas the gold content of the other currencies was established only indirectly, by means of a fixed parity with the dollar. Fluctuations were to be confined to a narrow 1% band. Due to international pressure in the early 1970s, the dollar was devalued with respect to gold, and then by 1973 the Bretton Woods system collapsed completely [23]. Hence, the time series we use start in the late 1970s to ignore the turmoils before the collapse of the Bretton Woods system.[2]

Government bonds

We will study the prices of two government bonds, a U.S. Treasury security (*T-Note*, 10 years maturity), which has become the security most frequently quoted when discussing the performance of the U.S. government bond market and is used to convey the market's take on longer-term macroeconomic expectations, and a German government bond (*Bundesanleihe*, 30 years maturity).

Commodities

Commodity markets have seen an upturn in the volume of trading in recent years. Global physical and derivative trading of commodities on exchanges increased more than a third in 2007 to reach 1684 million contracts. Agricultural contracts trading grew by 32% in 2007, while precious

1 Average volatilities $\bar{\sigma} \approx 1.35$ (HSI), 0.77 (S&P 500), 0.96 (FTSE 100), robust for different scale estimators and time windows.

2 The exchange rates are noon buying rates in New York for cable transfers (U.S. Federal Reserve).

Table 5.1: Time series data for this study, indicating the starting dates, as well as the sample size. All series end on 2010-10-01.

TIME SERIES	START DATE	SIZE
STOCK MARKET INDICES		
S&P 500	1950-01-03	15286
Hang Seng Index (HSI)	1969-11-24	10660
FTSE 100	1984-04-02	6696
FOREIGN EXCHANGE		
JPY/USD	1978-01-03	8232
GOVERNMENT BONDS		
U.S. T-Note	1967-01-03	11414
German gov. bond (GER Bund)	1986-06-20	6336
COMMODITIES		
Gold	1976-08-02	8915
Wheat	1985-06-03	6610

metals trading grew by 3% [19]. Especially through the use of futures contracts and derivatives, commodities are popular as an investment and subject to speculation. Here, we will study the price time series of gold and wheat to have representatives of the precious metal and agricultural commodities. Gold is the most popular of all precious metals as an investment and is very important throughout the world as a vehicle for monetary exchange.[3] Among the grains soy, corn, oats etc., wheat is one of the grains with the highest trading volume.[4]

5.2 PARAMETERS

We discussed so far the concepts of volatility, drawdowns and Dragon King detection. Each of these concepts is bound to some kind of parameter, i.e. the realised volatility depends on the choice of the scale estimator and of the size of the time window, drawdowns depend on the method

3 Gold prices from the London bullion market, a wholesale over-the-counter market for gold and silver.

4 Wheat prices from the Kansas City Board of Trade (spot values of the 5000 bushels Hard Red Winter Wheat No. 2 contract).

and degree of coarse-graining, and Dragon King detection on the tools we presented.

For the realised volatility we will use the $S_{n,X}$ estimator of scale (2.6) as consistent estimator $\hat{s}_X$ for the standard deviation with the time windows $T = 20, 125, 250$. For each of the three volatility estimates we construct pure drawdowns and $\epsilon^v_{0.5}$-drawdowns. We both use non-adjusted ("traditional") drawdowns and volatility-adjusted drawdowns as defined by (4.29). However, in order to be able to better compare these drawdowns, we scale the non-adjusted drawdowns by the average volatility $\bar{\sigma}$, i.e. we simply divide each non-adjusted drawdown by the same number, without changing the distributions. For each of these four distributions of drawdowns (per volatility estimate) we apply our Dragon King detection tools: First, the complementary cumulative distribution functions (CCDF) in logarithmic and semi-logarithmic representation; second, the likelihood-ratio (Wilks) test for nested hypotheses (stretched exponential and power law models); third, the unified most powerful unbiased (UMPU) test to discriminate between the power law and lognormal model; and fourth, Hill's inverse tail index estimate to further check the validity of the power law model. We further apply for each of the three volatility estimates our approach of taking distributions of drawdowns during k different volatility regimes, as introduced at the end of chapter 4 (section 4.2.2). For this purpose we define three different levels of granularity for the volatility regimes, namely $k = 3, 5, 10$. Again, for each of the volatility regimes $V_{i/k}$ we separate the non-adjusted[5] and volatility-adjusted p and $\epsilon^v_{0.5}$-drawdowns. For each of these four distributions of drawdowns per volatility regime $V_{i/k}$ (i.e. $4k = 72$ distributions of drawdowns per volatility estimate) we finally generate the CCDFs.

To provide an overview, in table 5.2 we summarise in a hierarchical way the tools we apply to the distributions of drawdowns with the different combinations of parameters. For the sake of completeness, all the resulting figures and tables for the different parameters and tools for each time series can be found in appendices B to F.[6]

5 In this case, we scale non-adjusted drawdowns by the centre volatility value $c_{i/k}$, instead of the average volatility $\bar{\sigma}$, to better compare non-adjusted and adjusted drawdowns within the volatility regimes $V_{i/k}$.

6 To keep the number of figures and tables as small as possible, but without sacrificing the clarity of the information, we combine in almost all cases non-adjusted and volatility-adjusted drawdowns, keeping p- and ϵ-drawdowns separate, and in some cases, all four kinds of drawdowns, into one single figure and table.

Table 5.2: Parameters and tools for the data analysis. Parameters: Volatility (VOL) with the estimator of scale $\hat{s}_X$ and time window T; p/ϵ-drawdowns (DD), non-adjusted/adjusted (n/a); and volatility regimes (VR), whole population (–) or $V_{i/k}$. Tools, applied/non-applied (•/–): CCDF, logarithmic (LL) and semi-logarithmic (SL); UMPU test power law versus lognormal (PL-LN); Hill's estimate $\hat{\alpha^{-1}}$; and Wilks test, stretched exponential (SE) and power law model (PL).

PARAMTERS					TOOLS					
VOL		DD		VR	CCDF		UMPU	HILL	WILKS	
$\hat{s}_X$	T	p/ϵ	n/a	$V_{i/k}$	LL	SL	PL–LN	$\hat{\alpha^{-1}}$	SE	PL
$S_{n,X}$	20	p	n	–	•	•	•	•	•	•
				$V_{1\ldots3/3}$	•	•	–	–	–	–
				$V_{1\ldots5/5}$	•	•	–	–	–	–
				$V_{1\ldots10/10}$	•	•	–	–	–	–
			a	–	•	•	•	•	•	•
				$V_{1\ldots3/3}$	•	•	–	–	–	–
				$V_{1\ldots5/5}$	•	•	–	–	–	–
				$V_{1\ldots10/10}$	•	•	–	–	–	–
		$\epsilon^{v}_{0.5}$	n	–	•	•	•	•	•	•
				$V_{1\ldots3/3}$	•	•	–	–	–	–
				$V_{1\ldots5/5}$	•	•	–	–	–	–
				$V_{1\ldots10/10}$	•	•	–	–	–	–
			a	–	•	•	•	•	•	•
				$V_{1\ldots3/3}$	•	•	–	–	–	–
				$V_{1\ldots5/5}$	•	•	–	–	–	–
				$V_{1\ldots10/10}$	•	•	–	–	–	–
	125	…			…					
		…			…					
	250	…			…					
		…			…					

5.3 RESULTS

In this section we will present the most important features of the results from our analyses. We will start with a thorough presentation of the results of the S&P 500 index in section 5.3.1 and then briefly present the main results of the remaining stock market indices (section 5.3.2), foreign exchange (section 5.3.3), government bonds (section 5.3.4) and commodities (section 5.3.5).

5.3.1 S&P 500

Please refer to appendix B for the complete set of figures and tables of the S&P 500 index.

VOLATILITY The volatilities for the three time window sizes are shown in figure B.1, with average volatilities around 0.77 for all T. The non-adjusted and adjusted p- and ϵ-drawdowns with their volatilities are shown in figures B.2, B.3 and B.4 ($T = 20, 125, 250$, respectively). We see that for $T = 20$ most of the drawdowns fall into the volatility range of around 0.25–2.5, while the upper bound of this range decreases with increasing T, compressing the distribution of drawdowns closer to the average volatility. This effect arises from the fact of large peaks in the distribution of realised volatility for small T. For a larger T, the bulk of the drawdowns is more symmetrically distributed around the average volatility, reducing the skewness of the distribution of volatility. This can also be seen in tables B.4, B.26 and B.48, showing the distribution of drawdowns within different volatility regimes ($k = 3, 5, 10$). The larger the time window size, the more drawdowns fall into the volatility regimes with volatility levels around the mean volatility.

CCDF In figures B.5, B.6 and B.7 the complementary cumulative distribution functions (CCDF) in log-log (a) and log-linear representations (b) are shown. As expected, the distributions of non-adjusted p-drawdowns are identical for the three time window sizes, since they are constructed and distributed independently of the volatility. At first glance, the distribution follows a slightly concave curve in the logarithmic representation giving first evidence of a stretched exponential behaviour where we could identify two obvious outliers in the tail. By comparing p-drawdowns with ϵ-drawdowns, we observe that for $T = 20$ and 125 the distributions fol-

low a straight line in the logarithmic representation with a rather smooth convex behaviour in the semi-logarithmic representation. For $T = 250$, the distribution resembles more closely the distribution of non-adjusted p-drawdowns. This behaviour is expected, since, with increasing T, the volatility time series becomes asymptotically flat, i.e. at the extreme $T = n$, with n being the sample size of returns, the volatility would be at a constant level. The difference of the distributions of volatility-adjusted and non-adjusted drawdowns is rather subtle. However, we observe that the adjustment has a "smoothing" effect, eliminating bumps and deviations. E.g. if we take the distribution of non-adjusted p-drawdowns around the drawdown size of 10%, we see a very obvious break, which does not appear in the adjusted distribution. In the adjusted distribution this break does not appear. The bump between 15% and 20% in the distribution of non-adjusted ϵ-drawdowns with $T = 125$ disappears after adjustment, too. Hence, in this case of the S&P 500, the adjustment seems to produce qualitatively more immaculate distributions, suggesting to be a better foundation for parametric tests.

EXTREME DRAWDOWNS Tables B.1, B.2 and B.3 list the largest drawdowns up to rank 15. Taking non-adjusted p-drawdowns as reference, 12, 11 and 12 ($T = 20, 125, 250$, respectively) out of the top 15 non-adjusted ϵ-drawdowns are the same events.[7] The picture looks very different for the adjusted drawdowns, where only 2, 5 and 5 (p-drawdowns) and 1, 5 and 5 (ϵ-drawdowns) out of the top 15 are the same as for non-adjusted p-drawdowns. However, 10, 9 and 7 out of the top 15 adjusted ϵ-drawdowns are the same as for the adjusted p-drawdowns. We see that the largest drawdown events seem to be robust when comparing non-adjusted p- and non-adjusted ϵ-drawdowns, as well as comparing adjusted p- and adjusted ϵ-drawdowns. Cross-comparing non-adjusted and adjusted drawdown events reveals larger discrepancies. However, this result is expected, since all but 1–3 events ($T = 20$: 1970-04-01, $T = 125, 250$: 1962-05-15, 1970-04-01, 1978-10-11) out of the top 15 non-adjusted drawdowns happened at times with above-average volatility. Hence, the volatility-adjustment decreases the magnitude (increases the rank) of these drawdowns in a much stronger way, such that they disappear from the top 15 ranking. We see this in figure 5.3 (a), showing the probability density functions (PDF) of the

7 We regard two drawdowns as the same event, if they start on the same date, or if they overlap, e.g. the drawdown starting on 2009-02-09 with 9 days length (non-adjusted ϵ-drawdown rank 5, $T = 20$) is the same event as the drawdown starting on 2009-02-12 with 6 days length (non-adjusted p-drawdown rank 10, $T = 20$).

volatilities of the top 15 non-adjusted and adjusted ϵ-drawdowns. Comparing the four kinds of drawdowns with different T, taking $T = 125$ as reference, we see that for $T = 20, 15, 13, 8$ and 8 (non-adj. p-, non-adj. ϵ-, adj. p- and adj. ϵ-drawdowns, respectively), and for $T = 250, 15, 14, 10$ and 12 drawdowns are the same events as in the equivalent top 15 drawdowns with $T = 125$. This tells us that, first, the distributions of non-adjusted p-drawdowns are — as mentioned earlier — identical for all T, second, non-adjusted ϵ-drawdowns are rather robust for all T, third, adjusted ϵ-drawdowns are also quite robust for all T, but more robust with increasing T. The last point is in line with our previous observation, that with increasing T the volatility time series becomes asymptotically flat, decreasing the differences between the distributions of adjusted and non-adjusted drawdowns.

HILL'S ESTIMATOR As mentioned earlier, we have identified by eye a stretched exponential and power law behaviour in the tails. Figures B.8, B.9 and B.10 show Hill's inverse tail index estimates. In all cases, they fluctuate in the range of ranks 10 up to 100 around 0.3, suggesting a power law behaviour in the tail with a tail index around 3. This is in line with the literature [8, 14, 15, 31, 38] for the distributions of returns, rather than drawdowns. However, if we assume the returns to be independent with a power law tail with exponent $\alpha = 3$, then we can predict that the distribution of drawdowns is also a power law with the same exponent. More interestingly, for the first few ranks, up to around rank 10, the inverse tail index estimates exhibit a strong instability, especially for the p-drawdowns, whereas the ϵ-drawdowns seem to be more stable in the lower ranks.

UMPU TEST Figures B.11, B.12 and B.13 show the results of the UMPU test to discriminate between the power law and the lognormal distribution. In all cases, the test selects the power law up to the first 100 ranks. However, for the first few ranks, the p-value seems to be unstable, especially for adjusted p- and ϵ-drawdowns. Hence, both Hill's estimator and the UMPU test show instability for the first few ranks. This can be an indicator for some change of regime or just due to the inability of the tests to work with small samples in the far tail. In the latter case, the UMPU test would prefer the power law in the tail, and this will occur even more if there is a Dragon King regime, since then the power law is closer to a fatter tail than the lognormal. Hence, we could imagine the drawdowns to be distributed lognormal-like in the bulk, power law-like in an intermediate tail and a Dragon King regime for the few extreme drawdowns, and the

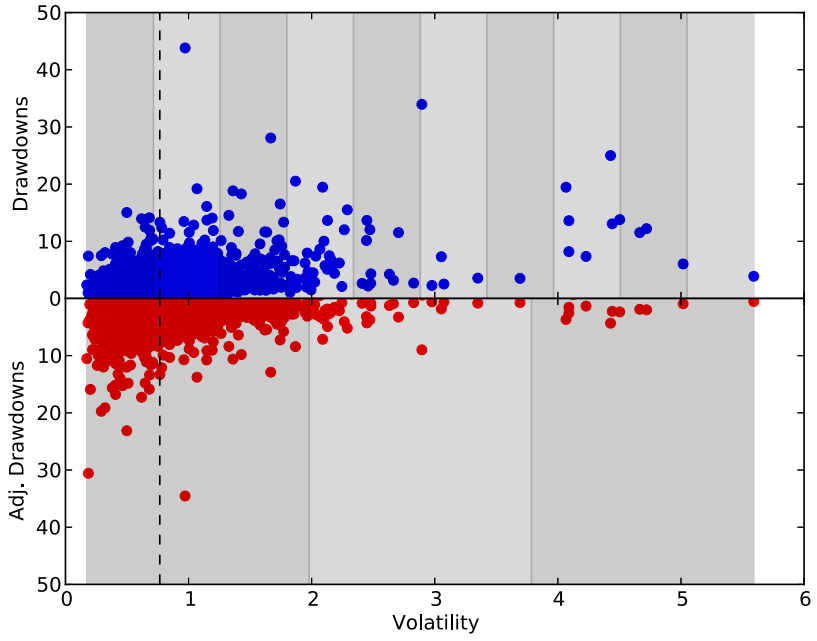
UMPU test would just confirm that the 100 smaller ranks follow a power law instead of lognormal distribution.

LIKELIHOOD-RATIO TEST Figures B.14, B.15 and B.16 show the results of the likelihood-ratio test for nested hypotheses (stretched exponential model). The test confirms the existence of two populations, one that describes the distribution of the smallest drawdowns (around rank 1000 upwards) better as stretched exponential ($C = 0$), and one that describes the larger drawdowns (around rank 1000 downwards) better as modified stretched exponential ($C \neq 0$). The test results for the power law model are shown in figures B.17, B.18 and B.19. In all cases, the p-value never decays beyond 5–10% to be able to reject the null ($C = 0$), suggesting a "pure" power law behaviour. However, the p-value sharply decays down to around 20% after the first 3 to 5 ranks. For the volatility-adjusted cases, the declines are generally stronger than for the non-adjusted cases, with the sharpest drop in the case of adjusted p-drawdowns with $T = 250$, where the p-value reaches its lowest point at just above 10%. After the p-values decay, the p-values rise again at around rank 50 and remain at levels close to 100%. We can see in the logarithmic representations of the CCDF that the increases of the p-values happen to be at a point of inflection, where the logarithmic CCDFs turns convex. Hence, although we formally cannot reject the null in order to describe the distributions as modified power laws ($C \neq 0$), we clearly see a change of regime after the first few drawdowns in form of a declining p-value. This can be seen as a strong indicator for a Dragon King regime for the first few ranks.

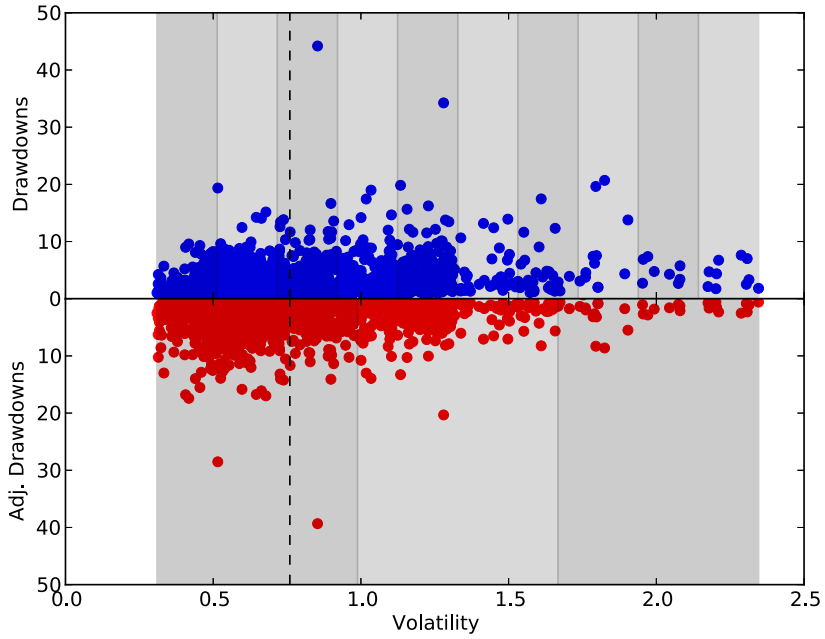
VOLATILITY REGIMES Until now, we have seen the results for the distributions of the whole population of non-adjusted and volatility-adjusted drawdowns. Let us now turn to the distributions of drawdowns during different volatility regimes. Figures B.21–B.23 ($T = 20$), B.42–B.44 ($T = 125$) and B.63–B.65 ($T = 250$) show the CCDFs for the volatility regimes with $k = 3$, tables B.6–B.8, B.28–B.30 and B.50–B.52 list the corresponding top 10 ranked drawdowns.[8] The figures starting at B.25, B.46 and B.67 show the CCDFs for the volatility regimes with $k = 5$ (tables starting at B.12, B.32 and B.54), and the figures starting at B.31, B.52 and B.73 show the CCDFs for the volatility regimes with $k = 10$ (tables starting at B.16, B.38 and B.60). Let us start with the small $k = 3$. We see in the CCDFs of almost all volatility regimes $V_{i/3}$ for all T strong indicators for outliers or changes

8 Only figures and tables for $V_{i/k}$ are shown if they consist of more than 10 drawdowns (see tables B.4, B.26 and B.48 for the quantities of drawdowns within volatility regimes).

of regime. This is especially true for our combined approach of adjusting drawdowns within volatility regimes, since the effect of adjustment is rather strong for a small k with broad ranges of volatility levels. For the first volatility regime $V_{1/3}$, we see in tables B.6, B.28 and B.50 that the top ranks coincide with the ranks of the corresponding drawdowns in the whole population (see column RK). For the adjusted p- and ϵ-drawdowns, virtually all ranks coincide very well. However, for non-adjusted p- and ϵ-drawdowns with $T = 250$ the discrepancies are quite large. This effect arises from the fact that the volatility levels get narrower with increasing T, e.g. the upper bounds of the volatility levels of $V_{1/3}$ decrease from 1.97 over 1.29 to 0.99 ($T = 20, 125, 250$, respectively). Since we identified earlier in the ranking of the whole population (tables B.1, B.2 and B.3) that most of the largest drawdowns happen at times with volatilities above 0.99, and since the volatility-adjustment generally emphasises drawdowns with small volatility, the absence of these drawdowns in $V_{1/3}$ with $T = 250$ can be explained. In other words, with increasing T, the drawdowns get more evenly and finer distributed over the k volatility regimes. We see this by taking k slices from figures B.2, B.3 and B.4, as shown in figure 5.1, where we have taken the extreme cases $T = 20$ and $T = 250$ (ϵ-drawdowns) and have drawn the slices for $k = 3$ and $k = 10$ to illustrate our point. For the volatility regimes $V_{2/3}$, as well as $V_{3/3}$, we see in tables B.7, B.29 and B.51, as well as B.8, B.30 and B.52, that there are large discrepancies in the top ranks with the ranks of the corresponding drawdowns in the whole population. However, the events are robust throughout the different kinds of drawdowns and time window sizes.



(a)



(b)

Figure 5.1: S&P 500 index. ϵ-Drawdowns at times with different realised volatility. (a) $T = 20$; (b) $T = 250$. Volatility-adjusted (red) and non-adjusted (blue), indicating $k = 3$ (bottom grey) and $k = 10$ (top grey) volatility regimes.

EXTREME DRAWDOWNS (VOLATILITY REGIMES) As mentioned earlier, we see in all cases strong indicators for outliers and changes of regime. Let us take from each of the four kinds of drawdowns the top $n_{i/3}$ ranked drawdowns from each volatility regime $V_{i/3}$ and combine them into an overall ranking.[9] Tables B.5, B.27 and B.49 list these 15 largest drawdowns combined into a single ranking. We see that in most cases the first 5–10 ranks coincide with the ranks from the whole population. These results are consistent four all four kinds of drawdowns and robust for all T. However, with increasing T, the ranks coincide even more with the ranks from the whole population, telling us that the very largest drawdowns in each volatility regime with $T = 250$ can be found within the very largest drawdowns of the whole population. We see that most of the drawdowns at ranks 10–15 are high-volatile events from 2008 and early 2009, which can be found only at higher ranks within the whole population. Figure 5.2 shows the S&P 500 index for the time around 2008–2009, indicating the bubble bursting in 2008. Hence, the events at ranks 10–15 are huge losses that occurred after — but not immediately after — the burst of the bubble. For the volatility regimes with $k = 5, 10$, the results are in line with the results obtained for $k = 3$. However, the effect of adjustment is much weaker with increasing k, since the volatility levels are much narrower. The distributions of non-adjusted and adjusted drawdowns are thus almost identical. In general, in almost all volatility regimes we see again strong indicators for outliers and changes of regime. Tables B.9, B.31 and B.53 ($k = 5$) and B.15, B.37 and B.59 ($k = 10$) list the 15 largest drawdowns combined into a single ranking, where, like for $k = 3$, the first 5–10 ranks coincide with the ranks of the whole population. The higher ranks are events from around 2008 and 2009, however, not as exclusively as for $k = 3$, i.e. events from the 1960s and 1970s are also present. As we can see in figure 5.3 (b), in the combined rankings drawdowns with low and high volatility are more evenly spread than in the whole population. To summarise the findings from the analysis of volatility regimes for the S&P 500, we confirm that the largest drawdowns combined from each volatility regime (especially for larger k) are — as expected — from a broader range of dates, since more dates with low and high volatility are taken into account. Again, it is important to stress, that the distributions of drawdowns during differ-

9 The choice of $n_{i/k}$ should coincide with the actual number of Dragon Kings detected in each volatility regime $V_{i/k}$. However, since we detect obvious outliers by eye, and to obtain a combined ranking of exactly 15 drawdowns, we generally choose $n_{i/3} = \{6, 5, 4\}$, $n_{i/5} = \{4, 4, 3, 2, 2\}$ and $n_{i/10} = \{2, 2, 2, 2, 2, 1, 1, 1, 1, 1\}$, with $n_{i/k} = 0$ if there are less than 10 drawdowns in $V_{i/k}$.

ent volatility regimes appear to exhibit a power law-like behaviour with strong indicators for outliers, deviations and changes of regime in the tails. In other words, it appears that there are more outliers than the 5–10 outliers we identify in the CCDFs of the whole population of drawdowns (figures B.5, B.6 and B.7). Furthermore, the results appear to be robust for p- and ϵ-drawdowns (generally a minimum of 70% of the combined top 15 drawdowns are the same events for p- and ϵ-drawdowns). There are discrepancies in the distributions of drawdowns for specific volatility regimes with different T. However, the aggregate (combined) results appear to be robust for different T, with more robust results for larger T.[10]

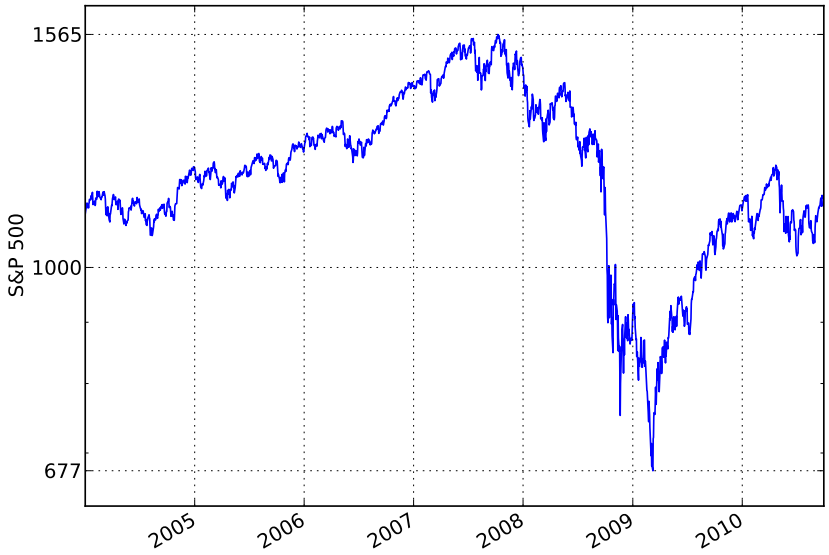
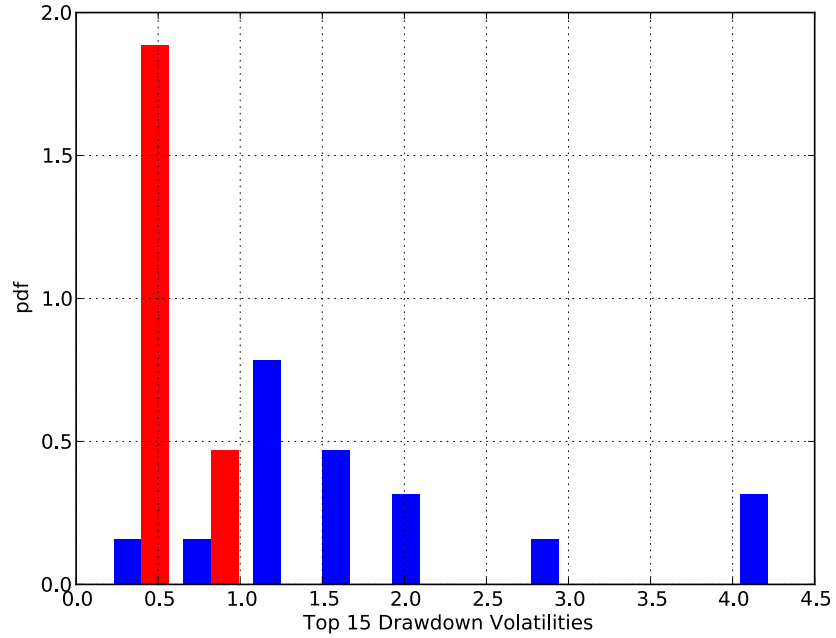
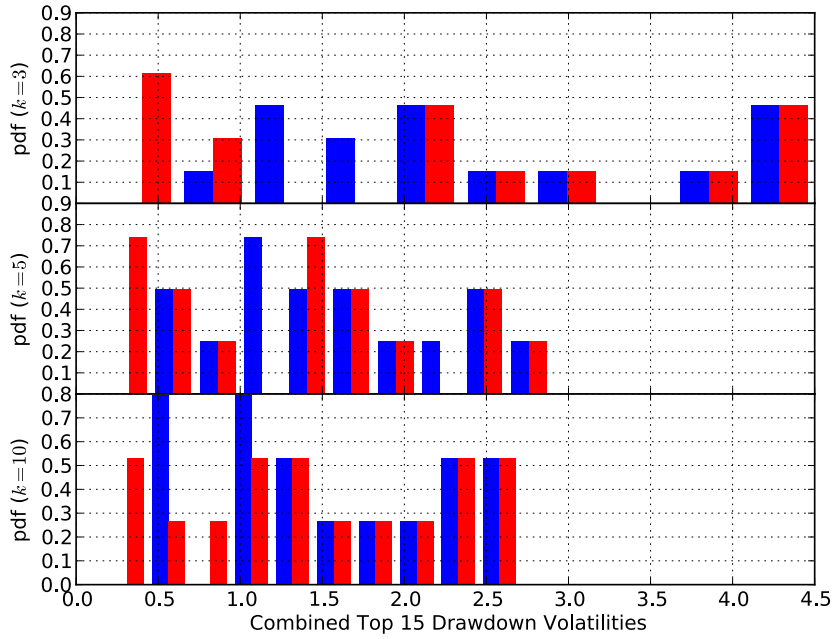


Figure 5.2: S&P 500 index from 2004 to 2010 (logarithmic price scale).

10 To avoid too much clutter in the results, we will focus the study of volatility regimes for the remaining time series on the two extreme cases $T = 20$ and $T = 250$, since the results appear to be rather robust for different time window sizes.



(a)



(b)

Figure 5.3: S&P 500 index ($T = 20$). Probability density functions (PDF) of the volatilities for the non-adjusted (blue) and adjusted (red) ϵ-drawdowns from (a) the top 15 ranks of the whole population, (b) the combined top 15 ranks of the volatility regimes with $k = 3, 5, 10$.

5.3.2 Stock market indices

Please refer to appendix C for the figures and tables of the stock market indices (Hang Seng Index and FTSE 100).

VOLATILITY As mentioned earlier, the Hang Seng Index is with a mean volatility of 1.35 the most volatile stock market index in our study. Especially in the 1970s and in the recent years, the volatility has been very high, with levels ranging from 2–4. For both the Hang Seng Index and the FTSE 100, the distributions of drawdowns across the different volatilities show the same patterns we already encountered, where — with increasing T — the distributions of drawdowns get compressed closer to the mean volatility. This behaviour confirms that for larger T the distribution of the volatility becomes less fat tail and less skewed.

WHOLE POPULATION In the logarithmic representations of the CCDF of the Hang Seng Index we identify a very straight line for all T, suggesting a power law behaviour in the tail. The adjustment seems to have a similar smoothing effect as for the S&P 500, while the shape of the distribution appears to be the same for all T (figure C.6). For the FTSE 100, the distributions are more distinct for different T, where larger deviations appear already at higher ranks (around drawdown sizes of 9%). However, the adjustment seems to have a smoothing effect on these deviations (figure C.59). The largest drawdowns are robust for different measures of noise, i.e. generally at least 10 out of the top 15 ranked drawdowns are the same events for non-adjusted p- and ϵ-drawdowns. Adjusted p- and ϵ-drawdowns are even more robust among themselves, and comparing drawdowns with the equivalent drawdowns with other T shows very high robustness, too. In general, for larger T the different kinds of drawdowns become more and more similar. This confirms the pattern we identified for the S&P 500. The inverse tail index estimates confirm in all cases a power law-like behaviour for the first 100 ranks, with tail indices around 2.5–3.5. However, the adjusted cases exhibit a very unstable behaviour for the first 10 ranks, where the inverse tail index reaches 0.5–0.6, indicating a fatter tail for the first few ranks (figure C.9). The UMPU test results are in line with Hill's estimates, choosing the power law for the first 100 ranks. However, for the FTSE 100 we notice a behaviour that we find across all non-adjusted drawdowns for all T, where around drawdown sizes 8–9% the p-value decays below 5–10% and increases again around drawdown sizes at 7–8%, indicating some kind of deviation in the tail, which is not present for ad-

justed drawdowns (figure C.65). Indeed, this coincides with the deviation and smoothing we identified earlier in the CCDF. The likelihood-ratio test results for the stretched exponential model are pretty straightforward, selecting the modified model ($C \neq 0$) for drawdowns larger than around 5%. This suggests the existence of a pure stretched exponential model ($C = 0$) for the smaller and a fatter than stretched exponential behaviour for the larger drawdowns. However, it appears that the p-values are more unstable for p-drawdowns, suggesting less disruptive deviations in the distributions around smaller drawdown sizes, or, in other words, smoother deviations for ϵ-drawdowns (figure C.15). The results for the power law model show again significant declines of the p-value after the first few ranks, and significant increases at ranks around 100, suggesting a wilder than power law-like behaviour for the very largest drawdowns (figure C.71).

VOLATILITY REGIMES The study of the volatility regimes reveals similar results as for the S&P 500. We see for almost all volatility regimes and time window sizes power law-like distributions with strong indicators for outliers and deviations, e.g. for $V_{1/10}$ (HSI, $T = 250$, figure C.44) we see a very straight line in the logarithmic representation of the CCDF with significant outliers in the tail, and $V_{7/10}$ (HSI, $T = 250$, figure C.50) seems to be a volatility regime without any significant outliers. The combined top 15 ranked drawdowns from all volatility regimes are very consistent with the rankings of the whole population, with a very robust behaviour for non-adjusted p- and ϵ-drawdowns, as well as the inclusion of drawdown events from a broader range of volatilities. Comparing the rankings between non-adjusted and adjusted drawdowns reveals a robust behaviour, too, e.g. 11 out of 15 ϵ-drawdowns are the same in the non-adjusted and adjusted cases for the combined ranking of $V_{1..5/5}$ (FTSE 100, $T = 20$, table C.51). With increasing T, we find again that the combined rankings become even more consistent with the rankings of the whole population for all kinds of drawdowns. For $V_{1..5/5}$ (FTSE 100), but with $T = 250$ (table C.70), we find 14 out 15 matching non-adjusted and adjusted ϵ-drawdown events. This confirms that the analysis of distributions of drawdowns during different volatility regimes is much more robust with larger T.

5.3.3 Foreign exchange

Please refer to appendix D for the figures and tables of the foreign exchange (JPY/USD).

VOLATILITY The exchange rate of the Japanese Yen against the U.S. dollar is low-volatile, with an average volatility at around 0.57 for all T, while the bulk of the drawdowns lies in the range of around 0.3–0.8 (figure D.3).

WHOLE POPULATION The CCDFs appear to follow a power law, with straight lines in the logarithmic representations and constant inverse tail index estimates around 0.3 for the lower ranks and for all T. Contrary to the time series already discussed, the adjustment appears to emphasise the tails of the distribution, especially for smaller T (figure D.5). The results of the UMPU test confirm a power law behaviour for the first 100 ranks, and the test appears to be sensitive to deviations in the far tail around rank 20, e.g. in figure D.12 we see a strong indicator for a change of regime around that rank. The likelihood-ratio test results for both the stretched exponential model and the power law model are in line with the results for the time series we have discussed so far. A fatter than pure stretched exponential behaviour is selected for drawdowns larger than around 3%, and a significant decline of the p-value for the power law model in the first few ranks suggests a wilder than power law-like behaviour of the very largest drawdowns. However, there are no significant different behaviours of the tests for non-adjusted and adjusted drawdowns.

VOLATILITY REGIMES We find the same patterns as for the other time series we discussed. In almost all volatility regimes the CCDFs exhibit a power law-like behaviour with strong indicators for outliers and deviations, suggesting the existence of changes of regime within different volatility regimes, e.g. for $V_{3/5}$ ($T = 250$, figure D.41). The patterns for the rankings of the combined top 15 drawdowns from each volatility regime are also in line with our previous observations, showing a robust behaviour for the equivalent rankings of the whole population. Comparing the combined rankings of p- and ϵ-drawdowns with each other reveals larger discrepancies, e.g. in the case of $k = 5$ with $T = 250$, only 8 drawdowns are the same for the 15 largest combined p- and ϵ-drawdowns (table D.28). A possible reason could be the fact that there are very long ϵ-drawdowns in this time series, e.g. the non-adjusted ϵ-drawdown at combined rank 9, staring at 2006-04-10, lasts over 26 trading days. Hence, the choice of $\epsilon_0 = 0.5$ could be too large for this time series. However, comparing the combined ranks of non-adjusted and adjusted ϵ-drawdowns reveals again a robust behaviour, sharing 12 out of 15 drawdown events.

5.3.4 Government bonds

Please refer to appendix E for the figures and tables of the government bonds (U.S. T-Note and German government bond).

VOLATILITY The average volatility of the U.S. T-Note is around 0.85 for all T, with a volatility constantly growing over the period of the time series, starting in the 1960s at a low level at around 0.5 and recently reaching its highest levels in the range of 1.5–3.5. The bulk of the drawdowns lies in the range of around 0.1–2. The German government bond is by far the lowest-volatile time series in our study, with an average volatility around 0.31, being rather stable over the recent years. The bulk of the drawdowns lies in the range of around 0–0.6, i.e. a much smaller range than for the other time series. For $T = 250$, we see a rather harsh lower bound at around 0.13, caused by the very constant low volatility spanning over several months around 1990 and 1992–1993 (compare figures E.59 and E.62).

WHOLE POPULATION For the German government bond we see in the CCDFs a very strong effect of the volatility-adjustment, emphasising a very distinctive outlier regime, especially for $T = 20$ (figure E.63). However, across all T, the smoothing effect of the adjustment we identified for other time series is not as salient (figures E.64 and E.65). The rankings of the 15 largest drawdowns reveal our usual patterns: non-adjusted p- and ϵ-drawdowns, as well as adjusted p- and ϵ-drawdowns are rather robust, non-adjusted ϵ-drawdowns are very robust across all T, and generally, consistencies increase with increasing T. The inverse tail index estimates of the U.S. T-Note are again constantly fluctuating around 0.3, with slightly higher values for the non-adjusted drawdowns, i.e. the tail indices of the non-adjusted drawdown distributions are slightly lower than the indices of the adjusted ones. Hill's estimates of the German government bond underline the particular shape of the distribution for this time series. We observe mainly two behaviours that are new in our study, first, the estimates do not fluctuate at such a constant level as in the other time series, i.e. a plateau validating a power law is not as apparent, second, the estimates for the adjusted drawdowns are with values over 0.4 much higher than for the non-adjusted drawdowns, where the values are in the range of 0.2–0.3. This large discrepancy arises from the strong effect of adjustment we identified before. Hence, strong changes of regime can be seen in the adjusted cases (figures E.66, E.67, E.68).

VOLATILITY REGIMES The distributions of drawdowns during different volatility regimes show in almost all cases strong indicators for outliers and deviations, and the rankings confirm the patterns of consistency we already identified for other time series. However, for the very low-volatile German government bond, we see the effect of choice of the time window size on segregating drawdowns into different volatility regimes, since for $T = 20$ the largest drawdowns are distributed across all volatility levels (figure E.60), but for $T = 250$ most of them are at the harsh lower bound (figure E.62). Hence, the combined rankings across different T are not as robust as for the other time series, e.g. for $k = 5$, only 7 out of 15 drawdown events are the same in the combined rankings for $T = 20$ and 250. The reason for this could arise from the fact that the volatility range for the German government bond is very narrow, obviating the claimed need of segregating drawdowns into different volatility regimes. Hence, for this time series, the results by adjusting the whole population by the volatility could be more appropriate.

5.3.5 Commodities

Please refer to appendix F for the figures and tables of the commodities (gold and wheat).

VOLATILITY The volatility of the gold price was very high in the early 1980s and was rising again over the recent years. The average volatility for the whole time series is around 0.87, with the bulk of the drawdowns lying in the range of 0.2–2. The average volatility for the wheat price is with 1.33 quite high and was rising over the recent years, while the bulk of the drawdowns lies in the range of 0.5–2. However, here we see a weakness of the estimator of scale, where some clustering appears around specific times, e.g. the artefact in the volatility time series around 1990 produces the clusters at the lower volatility bound in figure F.57. For the adjustment, this may produce distorted results, but for the study of the volatility regimes, this effect can be safely ignored.

WHOLE POPULATION Since most of the largest drawdowns for gold are at high-volatile times (figure F.3), the adjustment decreases the size of the largest drawdowns, making the tails of the CCDFs less pronounced for all T, i.e. the adjustment appears to produce smoother distributions with less obvious outliers (figure F.6). Because of this, the events in the rankings of the 15 largest adjusted drawdowns are very distinct to the non-adjusted

cases, e.g. for $T = 125$, only one drawdown event from the 15 largest adjusted drawdowns is also present within the 15 largest non-adjusted p-drawdowns. For the other time series discussed, we already found this pattern, however, not as strong as here. For wheat, the volatility-adjustment appears to have the smoothing effect we already encountered, probably because of the largest drawdowns for wheat are distributed more evenly across middle and high volatility levels (figure F.56). Hill's inverse tail index estimates for gold and wheat are again very particular compared to all time series in our study, since the estimates for the non-adjusted drawdowns do not fluctuate around a plateau in the lowest ranks and appear to decay towards 0. However, for the adjusted cases for gold, the estimates fluctuate around a constant level in the range of 0.1–0.2 in the lowest ranks. Figures F.9 (gold) and F.62 (wheat) are two remarkable examples. The UMPU test confirms this non-power law behaviour in some cases, where the null (power law) is rejected after a few ranks. However, in these cases, the p-value fluctuates, without clearly selecting the power law or lognormal distribution (figures F.12 and F.65). On the other hand, the likelihood-ratio test results for both commodities are in line with the results of the time series we already discussed. Although the p-value for the power law model does not decay as much as for the other time series, a change of regime after the first few ranks is clearly visible (figures F.18 and F.71).

VOLATILITY REGIMES The study of the volatility regimes for gold underlines the main idea of the concept. Since we observed that most of the largest drawdowns for gold are at high-volatile times, we still find outliers at the low volatility regimes. Even for the small $k = 3$, the distributions in $V_{1/3}$ indicate the existence of outliers (figures F.21 and F.35). For larger k the lower volatility regimes indicate even stronger outliers (figure F.46). Consequently, the combined rankings of the volatility regimes are less robust with respect to the rankings of the whole population, since the combined rankings include many low-volatile drawdowns, which are not included in the rankings of the whole population, due to their absolutely smaller size, e.g. ranks 11–15 in table F.34 are low-volatile drawdowns (non-adjusted ϵ-drawdowns). The results for wheat are in line with the main observations we have made so far, especially revealing some possible low-volatile outliers.

6 CONCLUSIONS

We started our study with the aim of answering the question of whether Dragon Kings in distributions of drawdowns taking a time-varying volatility into account give an even better insight into the nature of returns, drawdowns, Dragon Kings and extreme risks.

Our approach to answering that question consisted in first defining the term "volatility" to construct robust and sound time series of volatility varying over time. We then applied our concept of volatility to coarse-grained drawdowns sensitive to noise and to a set of tools to detect Dragon Kings. Our tools were mainly designed to confirm particular distributional behaviours of the bulk of the drawdowns and to find indicators for changes of regimes for the extreme drawdowns in the tail of the distributions. To test the effect of time-varying volatility we introduced two approaches. First, by "adjusting" or "normalising" drawdowns by the volatility at their time, and second, by "segregating" drawdowns into different "volatility regimes" — populations, which are characterised by the same level of volatility, or in other terms, by "conditioning" drawdowns by the volatility. Both approaches are not mutually exclusive, such that we also followed a combined approach of segregation and adjustment.

Our empirical analysis comprised eight time series, i.e. three stock market indices (S&P 500, HSI, FTSE 100), one currency (Japanese Yen), two government bonds (U.S. and Germany) and two commodities (gold and wheat). We defined a set of combinations of parameters, to test the influence these parameters have on the results obtained. To characterise the volatility, we have chosen the robust estimator of scale $S_{n,X}$ and time window sizes of $T = 20$ (one month), 125 (half year) and 250 (one year). We worked with two levels of noise for coarse-grained drawdowns, namely pure-drawdowns (no noise) and our newly defined variable ϵ^v-drawdowns with $\epsilon_0 = 0.5$. We set the granularity of volatility regimes to $k = 3, 5$ and 10, and generally worked with non-adjusted and adjusted drawdowns.

Our parametric tests confirmed the results of previous studies that distributions of drawdowns can be described very well by stretched exponentials and power laws in the tails, usually up to around rank 100 and with tail indices around 3, matching the power law behaviour of returns.

Checking the robustness of the different parameters, we found that the choice of T is the most crucial factor in the whole study, because it affects not only the adjustment and the volatility regimes, but also the ϵ^v-drawdowns. However, we found that p- and ϵ^v-drawdowns are in the same way very robust, as p- and ϵ^f-drawdowns turned out to be very robust in previous studies. Hence, we believe that due to the theoretical advantages, ϵ^v-drawdowns are a straightforward extension to coarse-grained drawdowns. The results of studying volatility regimes turned out to be more robust with both increasing k and T, with the disadvantage of decreasing sample sizes with larger k.

The effect of adjustment is manifold and non-conclusive: we found for most time series that the adjustment appears to smooth the distributions of drawdowns. Only for the Japanese Yen and for the German government bond we did find that the adjustment actually emphasised the tails, making outliers more obvious and pronounced. Interestingly, both of these times series are very low-volatile, but at this point, we are not able to confirm any connection between these characteristics.

The results of studying drawdowns separately during different volatility regimes are more promising: We find power law-like distributions with strong indicators for outliers and changes of regime for almost all volatility regimes, independent of the granularity k and time window size T. Depending on the time series studied, we find two remarkable features. The first feature follows from the high consistency we found when comparing the rankings of the whole population of drawdowns and the combined rankings of the largest drawdowns from each volatility regime, i.e. we found that especially for larger T most of the events of the 15 largest drawdowns from each ranking match with each other. This tells us that Dragon Kings we already identified in the whole population of drawdowns are not only *confirmed* by looking at combined rankings of the largest drawdowns from each volatility regime, but also *more pronounced* than in the whole population. In other words, the distributions of drawdowns during different volatility regimes exhibit in most cases more obvious deviations and outliers. The second remarkable feature we found follows from the few largest drawdowns in the combined rankings that are not present in the overall rankings. Here, we possibly found *new* Dragon Kings at times with low volatility, i.e. Dragon Kings of absolutely smaller size, but of extreme relative size in their context.

The synthesis of both features tells us that studying distributions of drawdowns separately during different volatility regimes reinforces the idea of the existence of drawdowns that appear to belong to a different popula-

tion than the bulk, with the extension that Dragon Kings do not need to be necessarily *absolute* extreme drawdowns, since we confirmed known "absolute" Dragon Kings and possibly found new "relative" Dragon Kings.

This conclusion gives rise to questions that remain unanswered. Here we propose several lines of work to continue this study:

- The adjustment appears to have some smoothing effect on the distributions of drawdowns such that parametric tests could benefit from these more immaculate distributions. Future research could investigate if there are indeed mechanisms and processes yielding smoother distributions.

- We identified that the volatility, especially through the choice of the time window size and scale estimator, is the most crucial factor when following the approach of volatility regimes. Hence, we propose to further investigate the concept of volatility with more sophisticated methods. Since literature in this area is very rare, we propose to consider Randal's [39] "iterated t-volatility estimator", which chooses the t-distribution with five degrees of freedom as a candidate for the data generation process for returns. His results appear to be promising, even if the underlying distribution of returns is not the t_5-distribution. Another approach is to use the robust scale estimators in conjunction with an adaptive time window size T, such that the volatility within the local time window be approximately constant.

- We introduced a methodology to check for Dragon Kings across different volatility regimes through the concept of combined rankings, taking a fixed number $n_{i/k}$ of the largest drawdowns within each volatility regime $V_{i/k}$ and combining them into one ranking. As we already mentioned, the actual Dragon Kings detected in each volatility regime should be combined. This is just a shortcut method, since we do not have good tests for outliers or changes of regimes for small samples in the tails. Hence, research should focus on new methods and tools to detect Dragon Kings even for small samples; we saw that the likelihood-ratio test for the power law model failed, probably because of samples in the tail being too small.

- Another focus should be to streamline the methodology to check for Dragon Kings across different volatility regimes. Here, the effect of the granularity k could be further investigated, e.g. by testing for Dragon Kings *across* different "resolution scales" k.

The results of our study can be regarded as a success, since we simply confirm the existence of Dragon Kings. However, we encourage a more optimistic take on the results, since we found evidence for possibly new, unknown Dragon Kings. Therefore, an important next step should be to link our concept of volatility regimes and Dragon Kings to the theory of exogenous and endogenous crashes, i.e. LPPL-bubbles [23], and fearful and fearless bubbles [2].

A VOLATILITY PARAMETERS

The figures in this appendix complement the analysis and figures in chapter 2 dealing with the parameters for the realised volatility.

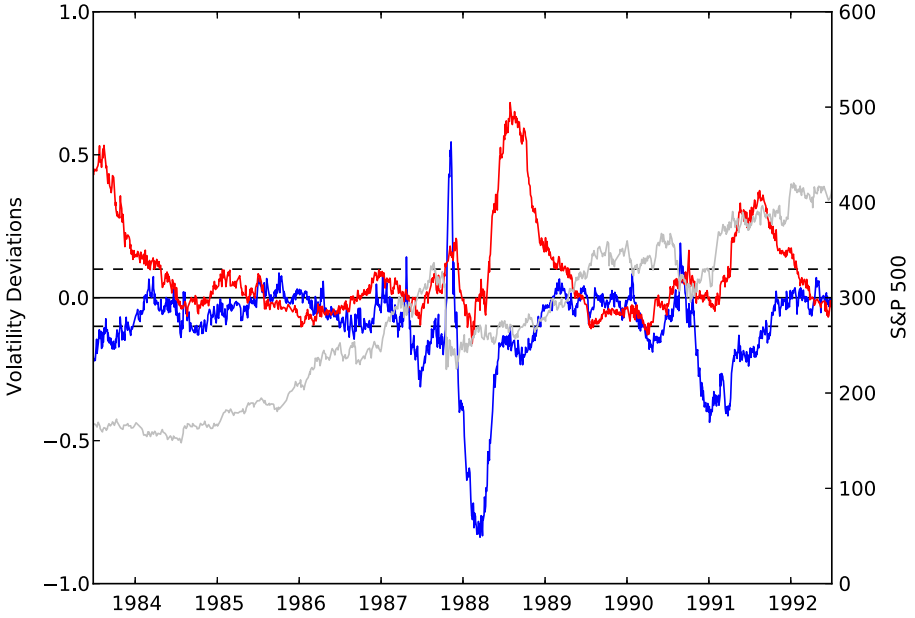
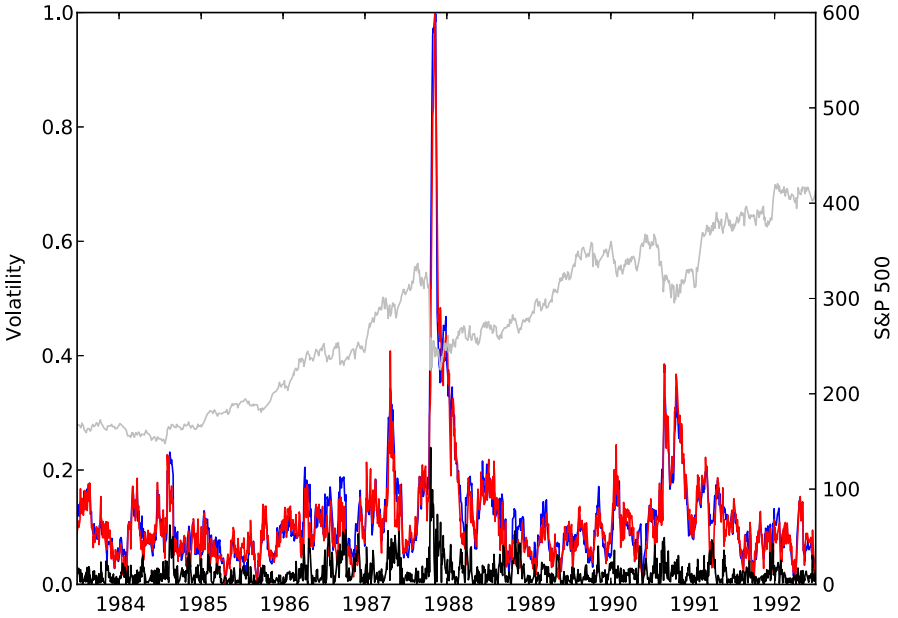
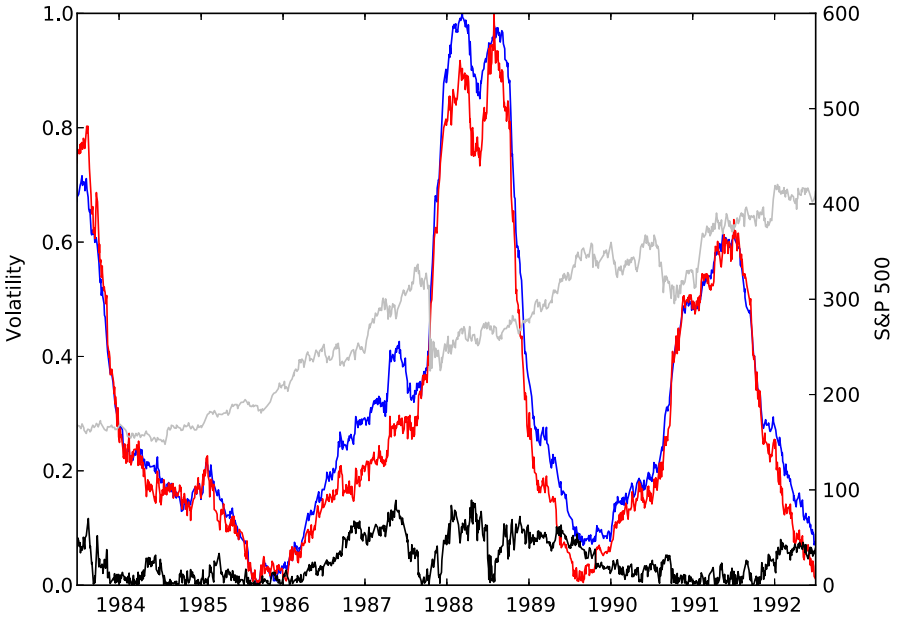


Figure A.1: S&P 500 index (grey). Relative volatility deviations of volatilities with time windows $T = 20$ (blue) and $T = 250$ (red) from the reference with time window $T = 125$ ($S_{n,X}$, scaled to $[0,1]$), showing the negative and positive 10% (black $--$).



(a)



(b)

Figure A.2: S&P 500 index (grey). Realised volatilities using two estimators of scale: Exponentially weighted standard deviation with 90% trimmed samples (s_X, $X^{(90)}$, blue), and $S_{n,X}$ (red), as well as the absolute deviations between the two (black). The y-axis is scaled to $[0,1]$. (a) Time window $T = 20$; (b) $T = 250$.

BIBLIOGRAPHY

[1] J. Andersen and D. Sornette. Have your cake and eat it, too: increasing returns while lowering large risks! *Journal of Risk Finance*, 2(3):70–82, 2001. (Cited on page 34.)

[2] J. V. Andersen and D. Sornette. Fearless versus fearful speculative financial bubbles. *Physica A*, 337(3–4):565–585, 2004. (Cited on page 74.)

[3] L. Bachelier. Théorie de la spéculation. *Annales Scientifiques de l'École Normale Supérieure*, 3(17):21–88, 1900. (Cited on page 29.)

[4] P. Bak. *How nature works: the science of self-organized criticality*. Copernicus New York, 1996. (Cited on page 3.)

[5] P. Bak and M. Paczuski. Complexity, contingency, and criticality. *Proceedings of the National Academy of Sciences of the United States of America*, 92(15):6689, 1995. (Cited on page 3.)

[6] O. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001. (Cited on page 9.)

[7] C. Croux and P. Rousseeuw. Time-efficient algorithms for two highly robust estimators of scale. *Computational Statistics*, 1(1):18, 1992. (Cited on page 13.)

[8] C. De Vries and K. Leuven. Stylized facts of nominal exchange rate returns. *Papers*, 1994. (Cited on pages 2 and 58.)

[9] J. Del Castillo and P. Puig. The Best Test of Exponentiality against Singly Truncated Normal Alternatives. *Journal of the American Statistical Association*, 94(446):529–533, 1999. (Cited on page 44.)

[10] R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4):987–1007, 1982. (Cited on page 9.)

[11] J. Feigenbaum. A statistical analysis of log-periodic precursors to financial crashes. *Quantitative Finance*, 1(3):346–360, 2001. (Cited on page 38.)

[12] S. Figlewski. Forecasting volatility. *Financial Markets, Institutions & Instruments*, 6(1):1–88, 1997. (Cited on page 16.)

[13] R. Gatto and S. Jammalamadaka. A saddlepoint approximation for testing exponentiality against some increasing failure rate alternatives. *Statistics & Probability Letters*, 58(1):71–81, 2002. (Cited on page 44.)

[14] P. Gopikrishnan, M. Meyer, L. Amaral, and H. Stanley. Inverse cubic law for the distribution of stock price variations. *The European Physical Journal B*, 3(2):139–140, 1998. (Cited on pages 2 and 58.)

[15] D. Guillaume, M. Dacorogna, R. Davé, U. Müller, R. Olsen, and O. Pictet. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and stochastics*, 1(2):95–129, 1997. (Cited on pages 2 and 58.)

[16] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *The Review of Economic Studies*, pages 247–264, 1994. (Cited on page 9.)

[17] B. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3(5):1163–1174, 1975. (Cited on page 43.)

[18] D. Hsieh. Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business*, 62(3):339–368, 1989. (Cited on page 29.)

[19] IFSL Research. *Commodities Trading 2008*. International Financial Services London, 2008. (Cited on page 53.)

[20] A. Johansen, O. Ledoit, and D. Sornette. Crashes as critical points. *International Journal of Theoretical and Applied Finance*, 3(2):219–255, 2000. (Cited on page 3.)

[21] A. Johansen and D. Sornette. Stock market crashes are outliers. *European Physical Journal B*, 1:141–143, 1998. (Cited on pages 4, 5, 29, 35, and 36.)

[22] A. Johansen and D. Sornette. Significance of log-periodic precursors to financial crashes. *Quantitative Finance 1 (4), 452-471 (2001)*, 2001. (Cited on pages 5, 35, and 38.)

[23] A. Johansen and D. Sornette. Shocks, crashes and bubbles in financial markets. *Brussels Economic Review (Cahiers economiques de Bruxelles)*, 53(2):201–253, Summer 2010, (http://ideas.repec.org/s/bxr/bxrceb.html, http://arXiv.org/abs/cond-mat/0210509). (Cited on pages 4, 5, 32, 52, and 74.)

[24] A. Johansen and D. Sornette. Large stock market price drawdowns are outliers. *Journal of Risk*, 4(2):69–110, Winter 2001/02. (Cited on pages 1, 29, 34, 35, and 36.)

[25] A. Johansen, D. Sornette, and O. Ledoit. Predicting financial crashes using discrete scale invariance. *Journal of Risk*, 1(4):5–32, 1999. (Cited on pages 3, 4, and 35.)

[26] JPMorgan. RiskMetrics technical document. *New York*, 4, 1996. (Cited on page 17.)

[27] A. N. Kolmogorov. On log-normal distribution law of particle sizes in fragmentation process. *Doklady Akad. Nauk SSSR*, 31:99–101, 1941 (in Russian). (Cited on page 42.)

[28] J. Laherrère and D. Sornette. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *The European Physical Journal B*, 2(4):525–539, 1998. (Cited on pages 3 and 34.)

[29] D. Le Bris. What is a stock market crash? 20 french crashes since 1854. September 2009. (Cited on pages 4 and 5.)

[30] L. Lin, Ren R.E., and D. Sornette. A consistent model of 'explosive' financial bubbles with mean-reversing residuals. *submitted to Quantitative Finance*, 2009, (http://arxiv.org/abs/0905.0128, http://papers.ssrn.com/abstract=1407574). (Cited on page 3.)

[31] T. Lux. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Applied Financial Economics*, 6(6):463–475, 1996. (Cited on pages 2 and 58.)

[32] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns: between the stretched exponential and the power

law? *Quantitative Finance*, 5(4):379–401, 2005. (Cited on pages 34 and 36.)

[33] Y. Malevergne, V. Pisarenko, and D. Sornette. Testing the pareto against the lognormal distributions with the uniformly most powerful unbiased test applied to the distribution of cities. *in press in Physical Review E*, 2011, (http://arxiv.org/abs/0909.1281, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1479481). (Cited on pages 34, 36, 42, and 44.)

[34] X. Meseguer. Dragon-kings in financial data, study at different time scales. Master's thesis, ETH Zurich, 2010. (Cited on pages 5, 29, and 36.)

[35] M. Moffatt. We will never run out of oil. http://economics.about.com/cs/macroeconomics/a/run_out_of_oil.htm. (Cited on page 2.)

[36] S. Morgenthaler and J. W. Tukey. *Configural Polysampling. A Route to Practical Robustness*. 1991. (Cited on pages 9 and 14.)

[37] R. Officer. The variability of the market factor of the New York Stock Exchange. *Journal of Business*, 46(3):434–453, 1973. (Cited on page 16.)

[38] A. Pagan. The econometrics of financial markets. *Journal of empirical finance*, 3(1):15–102, 1996. (Cited on pages 2 and 58.)

[39] J. Randal. *Robust volatility estimation and analysis of the leverage effect*. PhD thesis, Victoria University of Wellington, 2009. (Cited on pages 9, 10, 11, 14, 15, and 73.)

[40] C. Rao. *Linear statistical inference and its applications*. Wiley New York, 1973. (Cited on page 38.)

[41] P. Robinson. The estimation of a nonlinear moving average model. *Stochastic processes and their applications*, 5(1):81–90, 1977. (Cited on page 29.)

[42] P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993. (Cited on pages iii and 12.)

[43] D. Sornette. Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth. *Proc. Nat. Acad. Sci. USA, 99 SUPP1:2522-2529 (2002)*. (Cited on page 3.)

[44] D. Sornette. *Critical phenomena in natural sciences: chaos, fractals, selforganization, and disorder: concepts and tools*. Springer Verlag, 2004. (Cited on page 2.)

[45] D. Sornette. Probability distributions in complex systems. *Core article for the Encyclopedia of Complexity and System Science (Springer Science)*, 2009, Editor-in-chief: R.A. Meyers, 10370 p. 4300 illus., in 11 volumes, ISBN: 978-0-387-75888-6, (http://arxiv.org/abs/0707.2194). (Cited on page 2.)

[46] D. Sornette. Dragon-Kings, Black Swans and the Prediction of Crises. *International Journal of Terraspace Science and Engineering*, 2(1):1–18, 2009, (http://arXiv.org/abs/0907.4290). (Cited on page 3.)

[47] D. Sornette, J. Andersen, and P. Simonetti. Portfolio Theory for "Fat Tails". *International Journal of Theoretical and Applied Finance*, 3(3):523–536, 2000. (Cited on page 34.)

[48] D. Sornette and V. Pisarenko. Properties of a simple bilinear stochastic model: estimation and predictability. *Physica D: Nonlinear Phenomena*, 237(4):429–445, 2008. (Cited on page 29.)

[49] D. Sornette, P. Simonetti, and J. Andersen. ϕ^q-field theory for portfolio optimization. *Physics Reports*, 335(2):19–92, 2000. (Cited on page 34.)

[50] D. Sornette, R. Woodard, M. Fedorovsky, S. Reimann, H. Woodard, and W.-X. Zhou (The Financial Crisis Observatory). The financial bubble experiment: Advanced diagnostics and forecasts of bubble terminations. 2009, (http://arxiv.org/abs/0911.0454). (Cited on page 3.)

[51] N. N. Taleb. *The Black Swan: The impact of the highly improbable*. Random House Inc, 2007. (Cited on page 2.)

[52] S. Taylor. *Modelling financial time series*. World Scientific Pub Co Inc, 2007. (Cited on page 9.)