Master Thesis

Multifractal Finance

By Thijs Hoebers
Studentnumber 3118266

Under the supervision of Karma Dajani
Master: Stochastics and Financial Mathematics
Utrecht, June 2011

Universiteit Utrecht

Abstract

A striking feature of the prices of financial assets is that their statistical properties are to some degree universal across different assets, regions and epochs. There is a vast amount of literature on modelling these so-called stylized facts of financial data, but relatively recently multifractal processes have been proposed as a new formalism for financial modelling. Their main power lies in the fact that they capture many of the main statistical properties of financial time series in an effective way. The goal of this thesis is to present two multifractal models, the Multifractal Model of Asset Returns and the Markov-Switching Multifractal, and to study them in a more complete and rigorous way than in the literature.

Contents

1 Introduction

The random fluctuations of asset prices have a very rich and complicated statistical structure. A striking feature of these market price dynamics is that the statistical properties are to some degree universal across different assets, regions and epochs. There has been a vast amount of research papers on formulating a probabilistic model that reflect these properties. In 1900, Louis Bachelier introduced Brownian motion as a model for the price fluctuations. This was a revolutionary work and it was not until 1963 that a new model was proposed by Mandelbrot. He studied the fluctuations of cotton prices and established that extreme price movements are far more present than in the Gaussian world of Brownian motion. He proposed a different model in which prices follow the more heavy tailed Lévy distribution. Since then, many models were proposed to capture the properties of price dynamics. In this thesis we will present two very promising models, which were developed relatively recently and are based on Mandelbrot's work on (multi)fractals.

1.1 Empirical properties of financial returns

The interaction of thousands of individuals and institutions makes that financial markets are uniquely complicated systems, where fortunes can be made and lost in sudden bursts of activity. Especially the market crisis in 1987 and the recent credit crisis showed that financial markets can become very volatile, with big implications for economic and social welfare. These extreme events made it very clear that there was a need for better risk models. Since the accuracy of financial risk models mainly depends on the assumptions about the statistical properties of financial returns, it is very important to have realistic models. To develop such models, we have to know the statistical properties of asset price fluctuations. Although financial data has complicated structure, the result of more than half a century of empirical studies indicates that all financial returns share the same statistical properties. These so-called stylized facts are independent of the specific market, kind of asset and time. We will now give an overview of the most important stylized facts:

1. **Absence of autocorrelations:** Financial time series display only significant linear correlations in price increments on very small time scales of a few minutes, and can safely be assumed to be zero for all practical purposes. This absence of linear correlations means that the sign of the increments is independent of the past. So whether the price of an asset goes up or down, is not influenced by past behaviour.

2. **Heavy tails:** The unconditional distribution of returns has heavy tails, which means that they have a power-law-like tail and have infinite

moments. This basically means that significant deviations from the mean are much greater than in the case of the normal distribution.

3. **Volatility clustering:** High volatility events tend to cluster in time. When absolute returns are substantial on a given day, large movements are likely to follow. This clustering of volatility is apparent at intradaily, daily, weekly, monthly, yearly, and decennial epochs. This means that volatility clustering happens at many scales.

4. **Slow decay of autocorrelations in absolute returns:** The autocorrelation function of absolute returns decays more slowly than an exponential decay, and follows typically a power-like decay. This property is usually called long range dependence in volatility or long memory in squared returns. It means that the sizes of returns in the past still influence current volatility, and that this dependence decays relatively slowly in time.

5. **Multifractal moment scaling:** The (empirical) moments of financial returns vary as a power law in time. The power exponent of the q'th moment is different for different q and can be described by a strictly concave function.

6. **Discontinuous changes in volatility:** Financial data display sudden bursts of volatility, which means that volatility changes often discontinuously.

7. **Aggregational Gaussianity:** At short horizons, the distribution of returns is very non-normal, but as the time-horizon increases the distribution looks more and more like a normal distribution.

8. **Leverage effect:** Falling asset prices generate more volatility than upward movements in asset prices. A large downward move of the price tends to increase the volatility much more than a large upward move.

9. **Gain/loss asymmetry:** The unconditional distribution of asset returns is negatively skewed, which means that large downward movements in financial markets are more common than large upward movements.

Remark that although financial time series are characterized by an absence of autocorrelations, this does not mean that price increments are independent. Non-linear functions of the returns are highly dependent on each other, which is resembled by the fact that financial data display volatility clustering and long memory in squared returns. Multifractal scaling is often not mentioned as one of the stylized facts, which is due to the fact that it does not seem to have much practical importance. It does however give rise to the use of multifractal models for financial modelling. For more information about multifractal scaling in financial data, we refer to Mandelbrot (1997c), Galluccio (1997), Calvet and Fisher (2002a). A thorough review of the rest of the statistical properties can be found in Cont (2001) and Pagan (1996).

1.2 Financial modelling and multifractality

The main challenge of financial modelling is to capture all the stylized facts in a single model. Modelling these stylized facts is however far from trivial. To quote R. Cont: "the stylized facts are so constraining that it is not easy to exhibit even an (ad hoc) stochastic process which possesses the same set of properties and one has to go to great lengths to reproduce them with a model". The number of existing models, continuous or discrete, that try to capture as many of these stylized facts is wide. Models from the GARCH family seem to be the most popular models, and their main characteristic is that they are capable of capturing volatility clustering.

In this thesis we will present a relatively new class of stochastic processes, which will use the concept of multifractality. These multifractal processes form an alternative to ARCH-type presentations that in the last two decades have been the focus of financial modelling. We will present two models, the Multifractal Model of Asset Returns (MMAR) and the Markov-Switching Multifractal (MSM). Both models capture thick tails, volatility clustering, long memory in squared returns, multifractal moment scaling, discontinuous changes in volatility and aggregational Gaussianity. So the models contain, except for asymmetry properties, all the stylized facts mentioned in the previous section. Compared to the best performing models of the GARCH family, such as GARCH(1,1) and FIGARCH(1,d,0), the main advantage of the multifractal processes lies in their ability to generate volatility persistence (long memory and clustering) at different frequencies in a parsimonious way. It is showed in Mandelbrot (1997c); Calvet and Fisher (2002a); Di Matteo, Liu and Lux (2008) and Calvet and Fisher (2008), that MMAR and MSM outperform GARCH, FIGARCH and other well performing GARCH models on both short and large horizons.

The series of papers Mandelbrot et al (1997a), (1997b, (1997c) introduced for the first time the concept of multifractality to financial modelling. Multifractal theory was originally developed by Mandelbrot in the context of turbulence, but multifractality has been reported in many other systems such as rainfall distribution, sea surface temperature, heartbeat dynamics, distribution of chemical fields, etc. Multifractals model the self-similarity and scale invariance of those systems by requiring that the statistical properties of small regions of the system are the same as those of the whole system. So each 'subsystem' has the same properties as the original system. This self-similarity is also characteristic for fractals, but multifractals have a much richer structure since they are formed by a multiple of fractal sets. Mandelbrot argues in Mandelbrot (2005) that (multi)fractals are so effective in modelling many different systems because they intrinsically measure the roughness in those systems. He states that, since financial markets are also characterized by roughness, fractals tools (and especially multifractal tools) are an inescapable need for financial modelling.

1.3 Structure of the thesis

The main goal of this thesis is to give an overview of the properties and theory behind the two multifractals models MMAR and MSM. In chapter 2 and 3 we will present the Multifractal Model of Asset Returns, partly because this model is very interesting on its own, but also to give an introduction to multifractal processes and their characteristic properties. In chapter 4 we will introduce the Markov-Switching Multifractal, which is an improvement over the MMAR. In the appendix we will include proofs that might otherwise distract from the main ideas of this thesis. Chapter 6 will give an overview of which proofs in this thesis are own work and which are due to others.

Chapter 2 is devoted to introduce grid based multifractal measures and study their properties. We will first show how these measures are constructed and that this construction leads to non-degenerate measures, that are continuous but singular, and have infinite moments and long memory features. Then we will use the local Hölder exponent to study the typical local behaviour of these measures. We will see that grid based multifractals are characterized by a continuum of fractal sets, satisfy the multifractal formalism and have all their mass concentrated on a fractal set. Compared to the main financial literature about these grid based multifractals (Mandelbrot (1997a), (1997b), (1997c); Calvet and Fisher (2002a), we prove most of the characteristic properties of these multifractal measures in a more complete and rigorous way.

In chapter 3 we present the MMAR and show how the properties of grid based multifractal measures can be used for financial modelling. The properties of grid based multifractals will be passed on to the MMAR by compounding Brownian motion with a multifractal trading time. Although the MMAR is successful in capturing the stylized facts of financial time series, it is difficult to use for forecasting.

These shortcomings of the MMAR are overcome by introducing the MSM in chapter 4. This model will have a similar construction as the MMAR, but will now lead to a model that is very well suited for forecasting. To provide some intuition about how the MMAR and MSM are related to each other, chapter 4 starts by presenting a construction that is similar to the construction of grid based multifractal measures. Then in the rest of the chapter we will use a somewhat simpler but equivalent construction, and use this to study the properties of MSM and derive that there exists a usefull discretized version.

In contrast to chapter 2 and 3, we will not study the local properties of MSM. This is partly because for MSM the multifractal formalism is much harder to prove, but also because the MSM can be estimated with maximum likelihood estimation and therefore does not need the multifractal formalism for empirical modelling. So whereas in the first two chapters the idea is to introduce the concept of multifractality and compounding, chapter 4 focuses more on studying some important convergence results for MSM and its applications concerning empirical modelling.

2 Grid Based Multifractal Measures

In multifractal finance an important concept is to compound a stochastic process $B(t)$ with a time-deformation $\theta(t)$, i.e. we replace the clock time t with a trading time $\theta(t)$:

Definition: *Let B(t) be a stochastic process and $\theta(t)$ an increasing function. We call*

$$X(t) = B(\theta(t))$$

a compound process with trading time or time deformation process $\theta(t)$.

The main idea of introducing this trading time is that it can speed up or slow down the process B and therefore change the volatility and other properties of the process. In particular when the process B is a martingale this can be done without affecting its direction. In multifractal finance trading time will be a (usually random) multifractal process. For the MMAR we will define it as the cumulative distribution function of a grid based multifractal measure, a concept that will be introduced in the next section and studied in the rest of the chapter.

2.1 Multifractal measures

2.1.1 Conservative measures

Grid based multifractal measures are built by iterating a simple transformation, called a multiplicative cascade. The binomial measure on the interval $[0, 1]$ is the simplest example, and will be presented first. The cascade starts with a uniform probability measure μ_0 on the interval[1] [0,1]. Then in the first stage of the cascade the measure μ_1 is defined by uniformly spreading a mass m_0 on the interval [0,1/2] and the mass $1 - m_0$ on the interval [1/2,1], with $m_0 \in (0, 1/2) \cup (1/2, 1)$.

In the second stage of the cascade both intervals are split into two subintervals of equal length. The measure μ_2 is constructed as follows: the interval [0,1/4] will receive the fraction m_0 of the mass $\mu_1([0, 1/2])$, and the interval [1/4,1/2] will receive the fraction m_1 of $\mu_1([0, 1/2])$. When we apply the same procedure to [1/2,1], we obtain the following mass distribution for μ_2:

$$\mu_2([0, 1/4]) = m_0m_0 \qquad \mu_2([1/4, 1/2]) = m_0m_1$$
$$\mu_2([1/2, 3/4]) = m_1m_0 \qquad \mu_2([3/4, 1]) = m_1m_1$$

[1]We could also take the interval $[0, T]$ with arbitrary $T > 0$, but for simplicity we will take in this chapter the interval $[0, 1]$.

We can iterate this procedure infinitely many times, resulting in an infinite sequence of measures (μ_n). The binomial measure μ is now defined as the weak limit of this sequence as n goes to inifinity. Note that since $m_0 + m_1 = 1$ the mass is preserved at each interation step. So each measure μ_n and the limit measure μ have a total mass of one. This is why the measure μ is called a conservative measure (the mass is conserved). In this construction we could also have allocated the mass m_0 randomly to the left or right interval with equal probability. The resulting limit measure is called the randomized binomial measure. FIG.1 shows a simulation of the construction of such a measure.

To study the (randomized) binomial measure μ we introduce the dyadic intervals $[t, t+2^{-n}]$ where $t = \sum_{i=1}^{n} \eta_i 2^{-i}$ for some $\eta_1, \ldots, \eta_n \in \{0, 1\}$. Next we define the relative frequencies ϕ_0 and ϕ_1 of 0's and 1's in $(\eta_1, \ldots, \eta_n)$ as $\phi_0 = \frac{1}{n}\#\{i : \eta_i = 0\}$ and $\phi_1 = 1 - \phi_0$. The measure of a dyadic interval then becomes

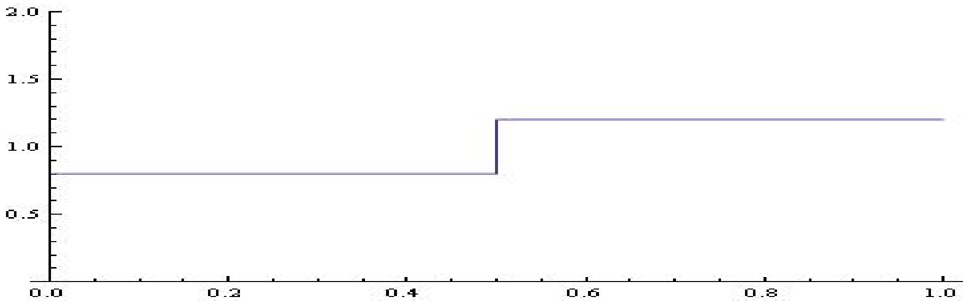
$$\mu([t, t+2^{-n}]) = m_0^{n\phi_0} m_1^{n\phi_1}$$

Because the mass of the dyadic intervals ranges between m_0^n and m_1^n, the mass of the binomial measure is far from uniformly distributed (even when $m_0 \approx m_1$ increasingly large differences arise as n goes to infinity). In fact, the distribution of mass of the measure behaves in such an irregular way, that it has no density. Because next to this the binomial measure also has zero point mass, it is a continuous but singular measure. These properties are common to many multifractals and will be studied in section 2.4.

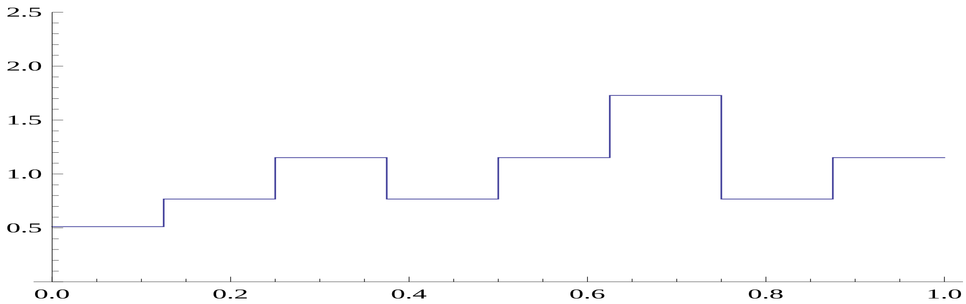
The binomial measure can easily be extended to the more general multinomial measures, which at every stage of their construction allocate mass over $b \geq 2$ cells. Again at each stage intervals are divided in b subintervals of equal length, which receive fractions of the total mass of that particular interval equal to $m_0, m_1, \ldots, m_{b-1} \in (0,1)\backslash\{b^{-1}\}$. Mass is preserved by imposing that these fractions, also called multipliers, satisfy $\sum_{\beta=0}^{b-1} m_\beta = 1$. We can extent this construction by randomly dividing the fractions $m_0, m_1, \ldots, m_{b-1}$ over the subintervals. Another extension is achieved by defining the multiplier of each cell as a discrete random variable M_β, which takes values $m_0, m_1, \ldots, m_{b-1}$ with probabilities $p_0, p_1, \ldots, p_{b-1}$. In this way a random measure is generated, where again mass is preserved by imposing that $\sum_{\beta=0}^{b-1} M_\beta = 1$.

Instead of sampling from a discrete distribution, we can also consider nonnegative multipliers M_β drawn from an arbitrary distribution M. In this construction it is assumed that the multipliers at different stages are independent. However, the multipliers at the same stage will be highly dependent on each other as we will again impose that they have to satisfy the mass conservation constraint $\sum_{\beta=0}^{b-1} M_\beta = 1$. To ensure that we won't get a trivial uniform measure, we exclude the case $M = b^{-1}$ a.s..

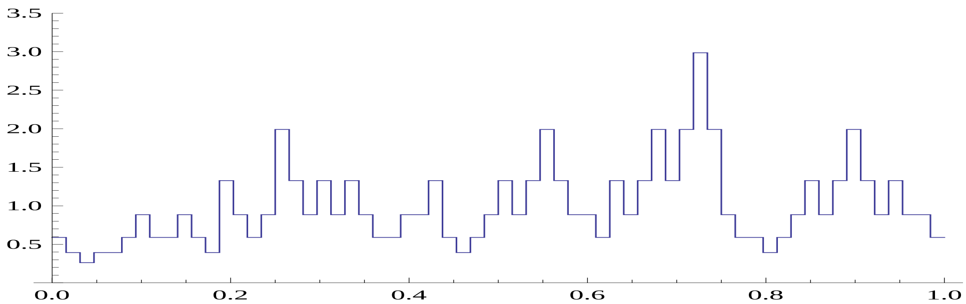
To be able to write the increments of the measure μ as a function of the



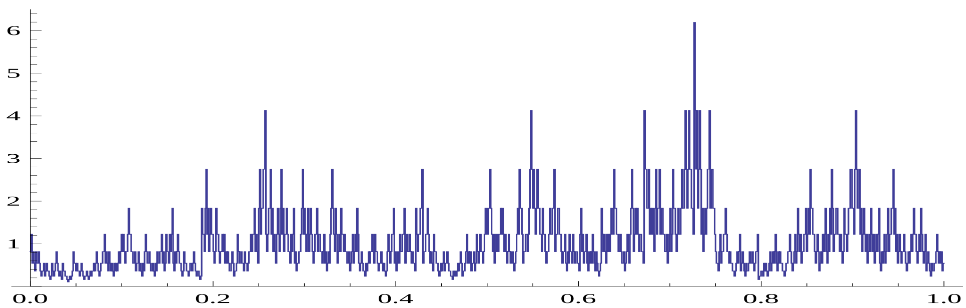
Binomial measure (stage n =1)



Binomial measure (stage n =3)



Binomial measure (stage n =6)



Binomial measure (stage n =10)

FIGURE 1 **Construction of Binomial measure**. Above graphs show simulations of a random binomial measure after respectively 1, 3, 6 and 10 stages, with $m_0 = 0.6$.

multipliers M, we define a b-adic interval of length $\Delta t = b^{-n}$ by $[t, t + b^{-n}]$ with $t = \sum_{i=1}^{n} \eta_i b^{-i}$ and $\eta_1, \ldots, \eta_n \in \{0, \ldots, b-1\}$. The measure of this interval is the product of n random multipliers:

$$\mu(\Delta t) = M(\eta_1)M(\eta_1, \eta_2) \ldots M(\eta_1, \ldots, \eta_n)$$

where we used the notation $\mu(\Delta t) = \mu([t, t + \Delta t])$.

2.1.2 Canonical measures

Canonical measures are constructed in the same way as multinomial measures, except that the random multipliers do not have to satisfy the strict mass conservation constraint, but will only have to conserve mass on average. This means that the nonnegative random multipliers are now independent and have to satisfy $\mathbb{E} \sum_{\beta=0}^{b-1} M_\beta = 1$, or equivalently $\mathbb{E}M = 1/b$. Since we also still exlude $M = b^{-1}$ a.s., a major consequence is that the total mass of a canonical measure does not have to be equal to one, but is now a random variable denoted by $\Omega = \mu([0, 1]) \geq 0$. More generally, the measure of a b-adic cell of length $\Delta t = b^{-n}$ becomes

$$\mu(\Delta t) = M(\eta_1)M(\eta_1, \eta_2) \ldots M(\eta_1, \ldots, \eta_n)\Omega(\eta_1, \ldots, \eta_n) \tag{1}$$

Note that each b-adic cell is also subjected to the same procedure as has been described for $[0, 1]$, so $\Omega(\eta_1, \ldots, \eta_n) \stackrel{d}{=} \Omega$.

As in section 2.1.1, the canonical measure μ is defined as the weak limit of the measures μ_n. We will now show that μ is well-defined, in the sense that this weak limit indeed exists:

Proposition 2.1 *For both the conservative and canonical construction, the sequence of measures μ_n has almost surely a unique weak limit μ.*

Proof. Let us first introduce some notation: define the b-adic intervals

$$J_n(\eta_1, \ldots, \eta_n) = \Big[\sum_{i=1}^{n} \eta_i b^{-i}, \sum_{i=1}^{n} \eta_i b^{-i} + b^{-n} \Big]$$

and for each n the collection of b-adic intervals

$$\mathcal{J}_n = \big\{ J_n(\eta_1, \ldots, \eta_n) : \eta_i \in \{0, 1, \ldots, b-1\} \big\}$$

Furthermore define

$$\phi_n(x) = M(\eta_1) \ldots M(\eta_1, \ldots, \eta_n) \quad \text{for } x \in J_n(\eta_1, \ldots, \eta_n)$$

Now we can write μ_n in the following way:

$$\mu_n(dx) = \mu_n([x, x + dx]) = \phi_n(x) b^n dx \qquad \text{for } dx \text{ such that } [x, x + dx] \subset J_n$$

Note that this also confirms that μ_n is indeed a measure, because μ_n is a finite sum of Lebesque measures λ:

$$\mu_n(B) = \sum_{J_n \in \mathcal{J}_n} \int_{B \cap J_n} \phi_n(x) b^n dx \quad \text{for arbitrary Borel sets } B$$

In order to show the weak convergence we will have to show that for all continuous and bounded functions f there exists a measure μ such that the integral $\bar{\mu}_n(f) = \int f d\mu_n$ converges to the integral $\int f d\mu$. We can write $\bar{\mu}_n(f)$ as:

$$\bar{\mu}_n(f) = \sum_{J_n \in \mathcal{J}_n} \int_{J_n} f(x) \mu_n(dx) = \sum_{J_n \in \mathcal{J}_n} \int_{J_n} f(x) \phi_n(x) b^n dx$$

If we use the natural filtration $\mathcal{F}_n$ of μ_n, defined by $\mathcal{F}_n = \sigma(\{M(\eta_1, \ldots, \eta_k) : \eta_i \in \{0, 1, \ldots, b-1\}, 1 \leq k \leq n\})$, we can show that $\bar{\mu}_n(f)$ is a L^1 martingale:

$$\begin{aligned}
\mathbb{E}[\bar{\mu}_n(f)|\mathcal{F}_{n-1}] &= \mathbb{E}\Big[\sum_{J_n \in \mathcal{J}_n} \int_{J_n} f(x) \phi_n(x) b^n dx \Big| \mathcal{F}_{n-1} \Big] \\
&= \sum_{J_n \in \mathcal{J}_n} \int_{J_n} f(x) b^n \mathbb{E}[\phi_n(x)|\mathcal{F}_{n-1}] dx \\
&= \sum_{J_n \in \mathcal{J}_n} \int_{J_n} f(x) b^{n-1} \phi_{n-1}(x) dx = \bar{\mu}_{n-1}(f)
\end{aligned}$$

where the use of the conditional Fubini's theorem was allowed, because f is bounded and we integrate over a compact interval. It also follows immediately that $\bar{\mu}_n(f) \in L^1$. Note also that since f is bounded we can assume without loss of generality that f is nonnegative. It follows that $\bar{\mu}_n(f)$ is a nonnegative martingale and hence we obtain from the martingale convergence theorem that there is a $\bar{\mu}(f)$ such that $\bar{\mu}_n(f)$ converges almost surely to $\bar{\mu}(f)$.

Since the operators $\bar{\mu}_n$ are linear ($\bar{\mu}_n(f+g) = \bar{\mu}_n(f) + \bar{\mu}_n(g)$), the limit operator $\bar{\mu}$ is also linear. Now we can use the Riesz Representation Theorem, which states that *for a compact metric space X and a linear functional $\tau : \mathcal{C}(X) \to \mathbb{R}$ which takes nonnegative values for nonnegative functions, there exists a unique finite measure μ defined on the Borel σ-algebra $B(X)$ such that $\tau(f) = \int f d\mu$ for all $f \in \mathcal{C}(X)$.* See for a proof for instance Sunder (2008). Now applying this theorem to the functional $\bar{\mu}$ gives that there is a unique measure μ such that

$$\bar{\mu}(f) = \int_{[0,1]} f(x) d\mu(x)$$

So we conclude that there is indeed a unique measure μ such that $\int f d\mu_n$ converges a.s. to the integral $\int f d\mu$ for every continuous and bounded function f, and hence we have established that there exists a measure μ such that μ_n converges weakly to μ a.s.. $\diamond$

Now we have proved that the limit measure μ indeed exists, we will study in the rest of this section the limit random variable $\mu([0,1]) = \Omega$. We will prove that this random variable is non-degenerate in the sense that $\mathbb{E}\Omega > 0$, which means that the sequence of measures μ_n don't degenerate to a trivial measure with zero mass. In proposition 2.2 it is proved that this doesn't happen as long as $\mathbb{E}M \log_b M < 0$. First we will introduce some notation and a recursive equation for $\mu_n([0,1])$.

We denote the total mass at stage n of the construction by the random variable $\Omega_n = \mu_n([0,1])$. Then Ω_n satisfies:

$$\Omega_n = \sum_{\eta_1,\dots,\eta_n} M(\eta_1)\dots M(\eta_1\dots\eta_n) \tag{2}$$

Let $\mathcal{F}_n$ denote the natural filtration of Ω_n, then it's easy to see that Ω_n is a positive martingale:

$$\begin{aligned}\mathbb{E}[\Omega_n|\mathcal{F}_{n-1}] &= \sum_{\eta_1,\dots,\eta_n} M(\eta_1)\dots M(\eta_1\dots\eta_{n-1})\mathbb{E}M(\eta_1\dots\eta_n) \\ &= b\mathbb{E}M \sum_{\eta_1,\dots,\eta_{n-1}} M(\eta_1)\dots M(\eta_1\dots\eta_{n-1}) = \Omega_{n-1}\end{aligned}$$

It follows that $\mathbb{E}\Omega_n = \mathbb{E}\Omega_1 = \mathbb{E}\sum_{i=1}^{b} M(i) = 1$. It also follows, as a consequence of the martingale convergence theorem, that Ω_n converges a.s. to the random variable Ω with $\mathbb{E}\Omega < \infty$. So, in addition to the weak convergence, we have also almost sure convergence of Ω_n to a tight limit random variable Ω.

Besides being a nonnegative martingale, Ω_n also satisfies a stochastic equation. As a consequence of the recursive nature of the construction of the measures (μ_n) (and as also can be derived directly from (2)), it holds that Ω_n satisfies the following recursive relation:

$$\Omega_n \stackrel{d}{=} \sum_{j=1}^{b} M_j \Omega_{n-1}(j) \tag{4}$$

where $\Omega_n, \Omega_{n-1}(1), \dots, \Omega_{n-1}(b)$ are independent random variables and the $\Omega_{n-1}(j)$'s have the same distribution as Ω_{n-1}.

Now we will prove that under $\mathbb{E}M \log_b M < 0$ the random variable Ω is non-degenerate.

Proposition 2.2 *If the multipliers M satisfy $\mathbb{E}M \log_b M < 0$, then $\mathbb{E}\Omega > 0$.*

Proof. First we introduce some more notation. We define the function[2] τ:

$$\tau(q) = -\log_b \mathbb{E}M^q - 1 \tag{5}$$

[2]This function will play a central role in multifractal analysis and will be formally introduced in section 2.2

Note that this function is well defined on the interval $[0, 1]$, because for all $h \in [0, 1]$ we have $\mathbb{E}M^h \leq 1 + \mathbb{E}M = 1 + b^{-1} < \infty$. Furthermore, in Mandelbrot et al(1997a) it is proved that this function is concave. It follows that the left- and right-hand derivatives exist, hence we can take the left-hand derivative in the point $h = 1$:

$$\tau^{'}(1) = -b\mathbb{E}M \log_b M > 0 \tag{6}$$

We will also need the following inequality, which will be proved in the appendix.

$$(x+y)^h \geq x^h + y^h - 2(1-h)(xy)^{\frac{h}{2}} \quad \forall x > 0, y > 0, 0 < h < 1$$

If we apply this result b times and use that the function x^h is subadditive: $\left(\sum_{j=i}^{b} x_j\right)^{h/2} \leq \sum_{j=i}^{b} x_j^{h/2}$ for nonnegative x_j, we get for $0 < h < 1$:

$$\begin{aligned}
\Big(\sum_{j=1}^{b} x_j\Big)^h &\geq x_1^h + \Big(\sum_{j=2}^{b} x_j\Big)^h - 2(1-h)\Big(x_1 \sum_{j=2}^{b} x_j\Big)^{h/2} \\
&\geq x_1^h + \Big(\sum_{j=2}^{b} x_j\Big)^h - 2(1-h)\sum_{j=2}^{b}(x_1 x_j)^{h/2} \\
&\geq x_1^h + x_2^h + \Big(\sum_{j=3}^{b} x_j\Big)^h - 2(1-h)\Big(\sum_{j=3}^{b}(x_2 x_j)^{h/2} + \sum_{j=2}^{b}(x_1 x_j)^{h/2}\Big) \\
&\geq \ldots \\
&\geq \sum_{j=1}^{b} x_j^h - 2(1-h)\sum_{i=1}^{b-1}\sum_{j=i+1}^{b}(x_i x_j)^{h/2}
\end{aligned}$$

If we apply this inequality to $\Omega_n \stackrel{d}{=} \sum_{j=0}^{b-1} M_j \Omega_{n-1}(j)$ and we use the notation $x_{j+1} = M_j \Omega_{n-1}(j)$ we get:

$$\begin{aligned}
\Omega_n^h \stackrel{d}{=} \Big(\sum_{j=0}^{b-1} M_j \Omega_{n-1}(j)\Big)^h &= \Big(\sum_{j=1}^{b} x_j\Big)^h \geq \sum_{j=1}^{b} x_j^h - 2(1-h)\sum_{i=1}^{b-1}\sum_{j=i+1}^{b}(x_i x_j)^{h/2} \\
&= \sum_{j=0}^{b-1} M_j^h \Omega_{n-1}(j)^h - 2(1-h)\sum_{i=0}^{b-2}\sum_{j=i+1}^{b-1} M_i^{h/2} M_j^{h/2} \Omega_{n-1}(i)^{h/2} \Omega_{n-1}(j)^{h/2}
\end{aligned}$$

And taking expectations gives:

$$\mathbb{E}\Omega_n^h \geq b\mathbb{E}M^h \mathbb{E}\Omega_{n-1}^h - b(b-1)(1-h)(\mathbb{E}M^{h/2})^2(\mathbb{E}\Omega_{n-1}^{h/2})^2 \tag{7}$$

where we used $\sum_{i=0}^{b-2}\sum_{j=i+1}^{b-1} 1 = \sum_{i=0}^{b-2}(b-i-1) = b(b-1) - \frac{b(b-1)}{2} = \frac{b(b-1)}{2}$. Aboves inequality can be rewritten as

$$\mathbb{E}\Omega_n^h - b\mathbb{E}M^h \mathbb{E}\Omega_{n-1}^h \geq -b(b-1)(1-h)(\mathbb{E}M^{h/2})^2(\mathbb{E}\Omega_{n-1}^{h/2})^2$$

Jensen's inequality applied to the concave function x^h gives that Ω_n^h is a supermartingale and hence $\mathbb{E}\Omega_n^h \leq \mathbb{E}\Omega_{n-1}^h$. If we apply this to aboves inequality we get:

$$\mathbb{E}\Omega_n^h(1-b\mathbb{E}M^h) \geq -b(b-1)(1-h)(\mathbb{E}M^{h/2})^2(\mathbb{E}\Omega_{n-1}^{h/2})^2$$

If we use that $b^{-\tau(h)} = b\mathbb{E}M^h$ we can write aboves inequality as

$$(\mathbb{E}\Omega_{n-1}^{h/2})^2 \geq \frac{\mathbb{E}\Omega_n^h}{b(b-1)(\mathbb{E}M^{h/2})^2}\frac{b^{-\tau(h)}-1}{1-h}$$

Now we want to take the limit $h \uparrow 1$. First note that since $\Omega_n^h \leq 1+\Omega_n$ and $\mathbb{E}\Omega_n = 1 < \infty$ we can use the Dominated convergence theorem to obtain $\lim_{h\uparrow 1}\mathbb{E}\Omega_n^h = \mathbb{E}\Omega_n$. In a similar way we also have $\lim_{h\uparrow 1}\mathbb{E}\Omega_n^{h/2} = \mathbb{E}\Omega_n^{1/2}$ and $\lim_{h\uparrow 1}\mathbb{E}M^{h/2} = \mathbb{E}M^{1/2}$. Now taking the limit $h \uparrow 1$ and using l'Hôpitals rule gives:

$$\begin{aligned}
(\mathbb{E}\Omega_{n-1}^{1/2})^2 &\geq \lim_{h\uparrow 1}\frac{\mathbb{E}\Omega_n^h}{b(b-1)(\mathbb{E}M^{h/2})^2}\frac{b^{-\tau(h)}-1}{1-h} \\
&= \lim_{h\uparrow 1}\frac{1}{b(b-1)(\mathbb{E}M^{h/2})^2}\frac{-\ln(b)\tau'(h)b^{-\tau(h)}}{-1} \\
&= \frac{\ln(b)\tau'(1)}{b(b-1)(\mathbb{E}M^{1/2})^2} \quad > \ 0
\end{aligned}$$

So we have established that there is a $\epsilon > 0$ such that $\mathbb{E}\Omega_n^{1/2} > \epsilon$ for all n. If we could show that $\mathbb{E}\Omega_n^{1/2}$ converges to $\mathbb{E}\Omega^{1/2}$, then this would imply that $\mathbb{E}\Omega^{1/2} \geq \epsilon$. Then by applying Jensen's inequality to the convex function x^2 we obtain the required result $\mathbb{E}\Omega \geq (\mathbb{E}\Omega^{1/2})^2 > 0$. So we need to show that $\lim_{n\to\infty}\mathbb{E}\Omega_n^{1/2} = \mathbb{E}\Omega^{1/2}$:

To obtain the required convergence we first need to show that the collection $\{\Omega_n^{1/2} : n \in \mathbb{N}\}$ is uniform integrable. To do this we use a theorem due to De La Vallée Poussin, which states that *a subset* $\{X_\alpha : \alpha \in A\}$ *of* L^1 *is uniformly integrable if there is a function* G *with* $\lim_{t\to\infty}\frac{G(t)}{t} = \infty$ *and* $\sup\{\mathbb{E}G(|X_\alpha|) : \alpha \in A\} < \infty$. If we take $G(t) = t^2$, then $\frac{G(t)}{t}$ goes to infinity and $\sup\{\mathbb{E}G(\Omega_n^{1/2}) : n \in \mathbb{N}\} = 1 < \infty$, hence $\Omega_n^{1/2}$ is uniformly integrable.

Now we can use Vitali's convergence theorem, which states that *if a sequence of random variables* X_n *in* L^1 *is uniformly integrable and converges almost surely to a* X*, then it also converges to this* X *in* L^1. By the continuous mapping theorem we have $\Omega_n^{1/2} \overset{a.s.}{\to} \Omega^{1/2}$. So we can apply Vitali's theorem, and obtain that $\Omega_n^{1/2}$ converges in L^1 to $\Omega^{1/2}$, hence $\lim_{n\to\infty}\mathbb{E}\Omega_n^{1/2} = \mathbb{E}\Omega^{1/2}$. ◇ (7)

The implication in the other direction, $\mathbb{E}\Omega > 0 \Rightarrow \mathbb{E}M\log_b M < 0$, does also hold (Guivarc'h (1987)). So when this condition on M is not satisfied, the measures μ_n converge indeed to the trivial zero measure.

To provide some intuition on why this happens, consider for instance the products $\frac{2}{6}\frac{4}{6} = \frac{8}{36}$, $\frac{1}{6}\frac{5}{6} = \frac{5}{36}$. Here the factors represent the multipliers in the construction of the binomial measure at the second stage. Note that although in both cases the sum (and thus the average) of the factors is equal, the product with the largest difference between its factors is significantly smaller than the other one. Now note that $\mathbb{E}M \log_b M \geq 0$ only occurs when M can take values above one, and because $\mathbb{E}M = b^{-1}$ these high values of M have to be compensated by a high probability of taking values below b^{-1}. As can be seen in aboves example, these low values have more effect on the product than the values above one, in the sense that they 'take away' more mass than the large values contribute. So we can interpret $\mathbb{E}M \log_b M < 0$ as the critical condition which prevents the described effect from becoming so significant that the cascade will die out. Because in that case we would get the trivial zero measure which is not interesting, we assume in the rest of this thesis that the condition $\mathbb{E}M \log_b M < 0$ is satisfied.

We have established that $\mathbb{E}\Omega > 0$, but because $\mathbb{E}\Omega_n = 1$ it is natural to expect that $\mathbb{E}\Omega = 1$. Kahane and Peyriere (1976) proved, as a corollary of proposition 2.2, that this is indeed the case. The result $\mathbb{E}\Omega = 1$ also means that next to the a.s. (martingale) convergence of Ω_n to Ω, we also have convergence of the expectations $\mathbb{E}\Omega_n$ to $\mathbb{E}\Omega$ (convergence in L^1).

We will end this section by remarking that the total mass Ω of grid based multifractal measures is characterized by an interesting stochastic equation, which can be used to improve on proposition 2.2. The fact that the random variables Ω_n satisfy the stochastic equation (4) implies, together with the weak convergence of Ω_n to Ω, that Ω satisfies the asymptotic version of (4), the so-called star equation:

$$\Omega \stackrel{d}{=} \sum_{j=1}^{b} M_j \Omega_j$$

where $\Omega, \Omega_1, \ldots, \Omega_b$ are independent and the Ω_i's have the same distribution as Ω. The unknown in this equation is the distribution of Ω, and its far from trivial to find the distribution of Ω, given the distribution of M. However, in section 2.3.2 we will show that the fact that Ω satisfies the star equation implies that Ω has thick tails, but first we will use it to improve on the result of proposition 2.2:

Corollary 2.1 $\mathbb{P}(\Omega > 0) = 1$ *if and only if* $\mathbb{P}(M > 0) = 1$.

Proof. First assume $\mathbb{P}(M > 0) = 1$, then we get:

$$p := \mathbb{P}(\Omega > 0) = \mathbb{P}\Big(\sum_{j=1}^{b} M_j \Omega_j > 0\Big) = \mathbb{P}(M\Omega > 0)^b = \mathbb{P}(\Omega > 0)^b$$

Since $p = 0$ or $p = 1$ are the only two possible solutions of the equation $p = p^b$ (with $b \geq 2$), and $p = 0$ can not hold since $\mathbb{E}\Omega > 0$, we obtain that $p = 1$.

To prove the implication in the other direction we also use the star equation:

$$\mathbb{P}(\Omega = 0) = \mathbb{P}\Big(\sum_{j=1}^{b} M_j \Omega_j = 0\Big) \leq \mathbb{P}(M_1 = 0, \ldots, M_b = 0) = \mathbb{P}(M = 0)^b$$

So $\mathbb{P}(M = 0) > 0$ implies $\mathbb{P}(\Omega = 0) > 0$. It follows that $\mathbb{P}(\Omega = 0) = 0$ implies $\mathbb{P}(M = 0) = 0$, from which the result follows easily. $\diamond$

2.2 The scaling function

In this section we will introduce a global scaling relationship and the scaling function for multifractal measures. First note that the measures we have constructed so far are grid based, in the sense that the relationship $\mu(\Delta t) = M(\eta_1)M(\eta_1, \eta_2) \ldots M(\eta_1, \ldots, \eta_n)\Omega(\eta_1, \ldots, \eta_n)$ only holds for b-adic intervals. A major drawback of this is that grid based multifractals are not stationary. To solve this problem of nonstationarity we can consider grid free multifractal measures. These measures will be introduced and studied in chapter 4.

Since for grid based measures the way in which mass is distributed is the same for all stages of the construction, grid based measures satisfy a form of scale-invariance, which can be described by a moment scaling relationship. This scale-invariance is present at all stages, and as such will prove to be very important for the local properties of multifractals. However, we will first study the effect of scale-invariance on a more global property: scaling in the moments of the measure.

If we define the deterministic function $\tau : \mathbb{R} \to \mathbb{R}$ by

$$\tau(q) = -\log_b \mathbb{E}M^q - 1$$

then grid based multifractal measures satisfy the following global scaling relationship

$$\mathbb{E}\mu(\Delta t)^q = c(q)(\Delta t)^{\tau(q)+1} \tag{9}$$

for b-adic intervals of length Δt and $c(q)$ also a deterministic function. The relationship follows immediately from (1), with $c(q) = \mathbb{E}\Omega^q$, $c(q) = 1$ for canonical, conservative measures respectively.

The most important feature of the moment scaling relationship is the function $\tau(q)$, which is called the scaling function. It will prove to play a central role in the analysis of multifractals and therefore we will in this section study some properties of this function. We can immediately see, by setting $q = 0$, that all scaling functions have the same intercept $\tau(0) = -1$.

We will now prove that, under the conditions $\mathbb{E}M^q < \infty \ \forall \ q \in \mathbb{R}$ and $\mathbb{P}(M > 0) = 1$, the scaling function $\tau(q)$ is twice differentiable.

Proposition 2.3 *Assume $\mathbb{E}M^q < \infty$ for all $q \in \mathbb{R}$ and $\mathbb{P}(M > 0) = 1$, then the scaling function $\tau(q) = -\log_b \mathbb{E}M^q - 1$ is twice differentiable with the following derivatives:*

$$\begin{aligned}\tau'(q) &= -(\ln b)^{-1}(\mathbb{E}M^q)^{-1}\mathbb{E}M^q \ln M & (10)\\ \tau''(q) &= (\ln b)^{-1}(\mathbb{E}M^q)^{-2}((\mathbb{E}M^q \ln M)^2 - \mathbb{E}M^q\mathbb{E}M^q \ln^2 M) & (11)\end{aligned}$$

Proof. We have to show that we can interchange integration and differentiation:

$$\frac{\partial}{\partial q}\int_0^\infty x^q dF_M(x) = \int_0^\infty \frac{\partial}{\partial q} x^q dF_M(x)$$

and

$$\frac{\partial}{\partial q}\int_0^\infty x^q \ln x \ dF_M(x) = \int_0^\infty \frac{\partial}{\partial q} x^q \ln x \ dF_M(x)$$

We will use the following differentiability lemma, which is proved in R.Schilling (2005):

Let $(c, d) \subset \mathbb{R}$ be a bounded interval and let $f(q, x)$ be a differentiable function in q on $(c, d) \times \mathbb{R}$, which satisfies

- *$\mathbb{E}|f(q, M)| < \infty$ for all $q \in (c, d)$ and*
- *there is function $g(x)$ such that $|\frac{\partial}{\partial q} f(q, x)| \leq g(x)$ for all $q \in (c, d)$ and $\mathbb{E}g(M) < \infty$*

then $\int_0^\infty f(q, x)dF_M(x)$ is differentiable on (c, d) with derivative:

$$\frac{\partial}{\partial q}\int_0^\infty f(q, x)dF_M(x) = \int_0^\infty \frac{\partial}{\partial q} f(q, x)dF_M(x)$$

Take an arbitrary interval (c, d). Since we want to prove twice differentiability of $\tau(q)$ we have to use this lemma twice. First define $f : (c, d) \times \mathbb{R}_+ \to \mathbb{R}_+$ by $f(q, x) = x^q$. Then $\mathbb{E}|f(q, M)| < \infty$ is obviously satisfied, and if we take the function $g(x) = (1 + x^c + x^d)|\ln x|$, then $|\frac{\partial}{\partial q} f(q, x)| = |x^q \ln x| \leq g(x)$ for all $q \in (c, d)$ and, as we will show now, $\mathbb{E}|g(M)| < \infty$. We will use the logarithmic inequalities $|\ln y| \leq y^{-1} - 1 < y^{-1}$ for $0 < y < 1$ and $\ln y \leq y - 1 < y$ for $y > 0$, which are both proved in the appendix.

$$\begin{aligned}\mathbb{E}|g(M)| &= \mathbb{E}(1 + M^c + M^d)|\ln M|\mathbf{1}_{\{M<1\}} + \mathbb{E}(1 + M^c + M^d)|\ln M|\mathbf{1}_{\{M\geq 1\}}\\ &\leq \mathbb{E}(1 + M^c + M^d)M^{-1} + \mathbb{E}(1 + M^c + M^d)M < \infty\end{aligned}$$

We can conclude that the conditions of the lemma are satisfied. Thus, $\frac{\partial}{\partial q}\mathbb{E}M^q = \int_0^\infty \frac{\partial}{\partial q} x^q dF_M(x) = \mathbb{E}M^q \ln M \quad \forall \ q \in (c, d)$. Then (10) follows easily by straightforward differentation.

To also prove the twice differentiability and (11), we have to show that $\frac{\partial}{\partial q}\int_0^\infty x^q \ln x \; dF_M(x) = \int_0^\infty \frac{\partial}{\partial q} x^q \ln x \; dF_M(x)$. First note that we have already established the first condition $\mathbb{E}|M^q \ln M| < \infty$. The second condition of the lemma can be proved in a similar way as above: since the derivative of $x^q \ln x$ is equal to $x^q \ln^2 x$, we can use the same logarithmic inequalities to prove that $\frac{\partial}{\partial q} x^q \ln x$ is bounded by a L^1 function. Now, given that we can interchange the integration and differentation, it is again some straightforward differentation to show that $\tau''(q)$ is given by (11).

Since the above results hold for arbitrary intervals (c, d), we complete the proof by letting $c \to -\infty$ and $d \to \infty$, and obtain that the results hold on the whole of $\mathbb{R}$. $\diamond$ (11)

(12)

Now we have established the twice differentiability of the scaling function, it can be shown that the second derivative is strictly negative:

Proposition 2.4 *Assume* $\mathbb{E}M^q < \infty$ *for all* $q \in \mathbb{R}$ *and* $\mathbb{P}(M > 0) = 1$*, then* $\tau''(q) < 0$.

Proof. Since $\mathbb{E}M^q < \infty$ for all q and $\mathbb{P}(M > 0) = 1$, the second derivative exists and is given by:

$$\tau''(q) = (\ln b)^{-1}(\mathbb{E}M^q)^{-2}((\mathbb{E}M^q \ln M)^2 - \mathbb{E}M^q\mathbb{E}M^q \ln^2 M)$$

To obtain $\tau''(q) < 0$ we will have to show that $(\mathbb{E}M^q \ln M)^2 < \mathbb{E}M^q\mathbb{E}M^q \ln^2 M$. To prove this we will use the well-known Cauchy-Schwarz inequality:

(13)

Let $f, g \in L^2$*. Then* $\|fg\|_1 \leq \|f\|_2 \|g\|_2$*. Equality occurs if, and only if,* $\|f\|_2 \|g\|_2 = 0$ *or* $f^2 / \|f\|_2^2 = g^2 / \|g\|_2^2$ *a.e.* (14)

(15)

If we use the functions $f(x) = x^{q/2}$, and $g(x) = x^{q/2} \ln x$ we get:

$$\mathbb{E}M^q \ln M \leq \left(\mathbb{E}M^q\right)^{\frac{1}{2}} \left(\mathbb{E}M^q \ln^2 M\right)^{\frac{1}{2}}$$

This implies $(\mathbb{E}M^q \ln M)^2 \leq \mathbb{E}M^q\mathbb{E}M^q \ln^2 M$. Since we want this inequality to be strict, we will consider the case that the equality holds. According to Cauchy-Schwarz's result, the equality occurs only if either

$$\mathbb{E}|M^{q/2}| = 0, \; \mathbb{E}|M^{q/2} \ln M| = 0 \; \text{ or } \; \frac{(M^{q/2})^2}{\mathbb{E}(M^{q/2})^2} = \frac{(M^{q/2} \ln M)^2}{\mathbb{E}(M^{q/2} \ln M)^2} \text{ a.s.}$$

holds. Because $\mathbb{E}M = b^{-1} > 0$ and M is nonnegative, the first two options can not hold. Therefore we will consider the third option, which can be rewritten as:

$$\ln^2 M = \frac{\mathbb{E}M^q \ln^2 M}{\mathbb{E}M^q} \text{ a.s.}$$

This implies that M has to be almost surely equal to a constant. Since $\mathbb{E}M = b^{-1}$ this means $M = b^{-1}$ a.s., but this option was excluded (since this would give a trivial measure). So the equality can not hold and hence we have a strict inequality, and thus $\tau''(q) < 0$. $\diamond$

An immediate corollary of aboves result is that the scaling function is strictly concave.[3] Another important corollary is that the Legendre transform of $\tau(q)$ is also strictly concave, which will be proved and used in sections 2.6 and 2.7. In the rest of this thesis we will assume that the conditions $\mathbb{E}M^q < \infty \; \forall q \in \mathbb{R}$ and $\mathbb{P}(M > 0) = 1$, under which the concavity properties hold, are satisfied. Note that according to corollary 2.1 the second condition also guarantees that $\Omega > 0$ almost surely. These assumptions exclude some possible distributions for M, but it will turn out that they leaves us enough freedom to choose distributions which are effective for empirical modelling.

2.3 Properties

2.3.1 Continuous but singular

Grid based multifractal measures have, due to the recursive nature of their construction, besides scaling properties, some other very special properties. The measures are different from commonly used (probability) measures in the sense that they do not have a density, but also have no point mass. The fact that they don't have a density means that they are not absolutely continuous with respect to the Lebesque measure, and thus are singular measures. In fact, it will turn out that all the mass of a multifractal measure is concentrated on a set of Lebesque measure zero. The proof of this last statement will have to wait until section 2.8. In this section however we will make it intuitively clear why multifractal measures have no density and prove that they have zero point mass.

In the familiar case where a measure μ has a continuous density $\mu'(t)$, an approximation of this density is given by $\mu([t, t + \Delta t])/\Delta t$. As $\Delta t \to 0$ we expect the approximate density to go to the true $\mu'(t)$. For multifractal measures, however, this behaviour it totally different. Due to the recursive nature of the construction, we have that all time scales Δt the mass distribution over the intervals $[t, t + \Delta t]$ is approximately as irregular as in Fig.1. So as Δt goes to zero, the value of the approximate density $\mu([t, t+\Delta t])/\Delta t$ keeps changing dramatically. For example if Δt is halved, the sharing of the mass $\mu([t, t + \Delta t])$ between the two halves is often very unequal. So as $\Delta t \to 0$ the approximate density will remain to behave very wiggly and hence does not become an increasingly close approximation to some density function. It follows that the measure fails to have a density and is thus not

[3]In Mandelbrot et al (1997a) it is showed that the scaling function is concave. We improve on this result by showing that the scaling function is strictly concave, which will be used in the proofs of proposition 2.6 and 2.7, and theorem 2.1.

absolutely continuous with respect to the Lebesque measure. This implies that the mass of the measure is at least partly concentrated on a set of Lebesque measure zero. This is for instance also the case for discrete measures as their mass is concentrated on a countable set. However, multifractal measures have no point mass, as we will show now:

Proposition 2.5 *$\mu(\{x\}) = 0$ a.s. for all $x \in [0, 1]$.*

Proof. Take $x \in [0, 1)$. Define $k_n = k_n(x) = \lfloor xb^n \rfloor$, where $\lfloor y \rfloor$ is defined as the largest integer smaller or equal to y. Then k_n is the unique integer such that

$$x \in I_{k_n}^{(n)} := [k_n b^{-n}, (k_n + 1)b^{-n}] \quad \text{for all } n$$

The sets $I_{k_n}^{(n)}$ form a sequence of nested intervals, which will be proved in the appendix. The intersection of these intervals is equal to x: $\bigcap_{n=1}^{\infty} I_{k_n}^{(n)} = \{x\}$. Furthermore, note that $k_n b^{-n}$ is a dyadic number, i.e. there is a sequence $\eta_1, \ldots, \eta_n \in \{0, 1, \ldots, b-1\}$ such that $k_n b^{-n} = \sum_{i=1}^{n} \eta_i b^{-i}$. This all leads to:

$$\begin{aligned}
\mu(\{x\}) &= \mu\Big(\bigcap_{n=1}^{\infty} [k_n b^{-n}, (k_n + 1)b^{-n}]\Big) \\
&= \lim_{n\to\infty} \mu\big([k_n b^{-n}, (k_n + 1)b^{-n}]\big) \\
&= \lim_{n\to\infty} \prod_{i=1}^{n} M(\eta_1, \ldots, \eta_i)\ \Omega(\eta_1, \ldots, \eta_n) \\
&= \lim_{n\to\infty} \prod_{i=1}^{n} M_i\ \Omega
\end{aligned}$$

First remember that $\mathbb{E}\Omega < \infty$, hence $\Omega < \infty$ a.s.. So we need to show that $\lim_{n\to\infty} \prod_{i=1}^{n} M_i$ is almost surely equal to zero. We will use the following result from measure theory: *Let $X, X_1, X_2, \ldots$ be random variables. If for all $\epsilon > 0$ the series $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \epsilon)$ is convergent, then $X_n \overset{a.s.}{\to} X$.*

By an application of Markov's inequality we have:

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\Big|\prod_{i=1}^{n} M_i - 0\Big| > \epsilon\Big) \le \sum_{n=1}^{\infty} \frac{\mathbb{E}\prod_{i=1}^{n} M_i}{\epsilon} = \frac{1}{\epsilon}\sum_{n=1}^{\infty} b^{-n} < \infty$$

This holds for all $\epsilon > 0$, so we conclude that $\lim_{n\to\infty} \prod_{i=1}^{n} M_i = 0$ almost surely. This gives the result for $x \in [0, 1)$. For $x = 1$ everything can be proved similarly with $k_n(x) = \lceil xb^n \rceil$ and the intervals $[(k_n - 1)b^{-n}, k_n b^{-n}]$.$\diamond$

A simple consequence of the above result is that the cumulative distribution function θ of the measure μ is continuous. We conclude that a grid based multifractal is a continuous but singular measure.

If no mass is concentrated on sets with a nonzero Lebesque measure and neither is concentrated in single points, one can wonder where the mass went. The answer is that the mass is concentrated on sets with a fractal dimension smaller than one. This makes sense because the Lebesque measure of a fractal set with dimension smaller than one, is equal to zero. In fact, as we will show in section 2.7, all the mass of μ is concentrated on a single set with a fractal dimension $D < 1$. An introduction to fractal theory can be found in section 2.5.3.

2.3.2 Infinite moments and heavy tails

In this section we will take a closer look at the distribution of the total mass Ω. So far we have showed it is non-degenerate, it satisfies the star equation and we mentioned that its expectation is equal to one. In addition to this we will show that Ω has infinite moments, which reflects that Ω has heavy tails. This is an important property of canonical multifractal measures, as this also an intrinsic property of the distribution of asset returns.

By definition, conservative measure have a fixed mass. Consequently, conservative measures will have finite moments of all orders. The corresponding compound process will therefore also have finite moments of all orders. Conservative measures thus generate compound processes with relatively thin tails. Canonical measures, however, can generate processes with infinite moments, and thus heavy tails. In this case the tail behaviour of the compound process $B(\theta(t))$ will depend on the random variable Ω. Overall, it can be seen that multifractal measures may have a variety of tail behaviours.

As mentioned, conservative measures have finite moments of all orders. This is because $\mu([0,1]) = 1$ and hence $\mathbb{E}\mu([0,1])^q = 1$ for all q. But, when mass is not conserved, there exists a $q_{crit} < \infty$ such that for all $q > q_{crit}$ we have $\mathbb{E}\mu([0,1])^q = \mathbb{E}\Omega^q = \infty$. We will prove this as a corollary to the following result.

Proposition 2.6 *Let $q > 1$, then $\mathbb{E}\Omega^q < \infty$ if and only if $\mathbb{E}M^q < b^{-1}$.*

Proof. First we assume $\mathbb{E}\Omega^q < \infty$. If we use that Ω satisfies the star equation $\Omega \overset{d}{=} \sum_{j=1}^{b} M_j\Omega_j$ and that the function x^q is superadditive, which means that $(x_1 + x_2)^q \geq x_1^q + x_2^q$, we get:

$$\Omega^q \overset{d}{=} \left(\sum_{j=1}^{b} M_j\Omega_j \right)^q \geq \sum_{j=1}^{b} M_j^q \Omega_j^q \tag{17}$$

If we can show that there is a positive probability that this inequality holds strictly, we would get:

$$\mathbb{E}\Omega^q > \mathbb{E} \sum_{j=1}^{b} M_j^q \Omega_j^q = b\mathbb{E}M^q\mathbb{E}\Omega^q$$

Then since $\mathbb{E}\Omega^q < \infty$, dividing by $b\mathbb{E}\Omega^q$ gives the required result $\mathbb{E}M^q < b^{-1}$. So its left to show that there is positive probability that (17) holds strictly.

In the appendix we show that the equality $\big(\sum_{i=1}^b x_i\big)^q = \sum_{i=1}^b x_i^q$ can only hold if there is at most one nonzero x_i. This implies, together with the fact that $b \geq 2$ and the superadditivity of x^q, that there are only two nonzero x_i needed to guarantee that the strict inequality holds: $x_1, x_2 > 0$ implies $\Big(\sum_{i=1}^b x_i\Big)^q > \sum_{i=1}^b x_i^q$. Using this we get

$$\begin{aligned}\mathbb{P}\Bigg(\Big(\sum_{j=1}^b M_j\Omega_j\Big)^q > \sum_{j=1}^b M_j^q\Omega_j^q\Bigg) &\geq \mathbb{P}(M_1\Omega_1 > 0, M_2\Omega_2 > 0)\\ &= \mathbb{P}(M > 0)^2\mathbb{P}(\Omega > 0)^2 > 0.\end{aligned}$$

As explained above, this establishes $\mathbb{E}M^q < b^{-1}$.

Now assume $\mathbb{E}M^q < b^{-1}$ and take the integer k such that $k < q \leq k+1$. Because the function $x^{\frac{q}{k+1}}$ is subadditive, we have the following inequality:

$$\Big(\sum_{j=1}^b x_j\Big)^{\frac{q}{k+1}} \leq \sum_{j=1}^b x_j^{\frac{q}{k+1}}$$

If we define the sets $L = \{l = (l_1, \ldots, l_b) \in \mathbb{N}^b : \sum_{i=1}^b l_i = k+1\}$ and $L^* = \{l \in L : l_i = k+1, i = 1, \ldots, k+1\}$ and use the multinomial theorem, which states that $\big(\sum_{j=1}^b x_j\big)^{k+1} = \sum_{l\in L}\frac{(k+1)!}{l_1!\ldots l_b!}\prod_{j=1}^b \big(x_i\big)^{l_j}$, then the above inequality becomes:

$$\begin{aligned}\Big(\sum_{j=1}^b x_j\Big)^q \leq \Big(\sum_{j=1}^b x_j^{\frac{q}{k+1}}\Big)^{k+1} &= \sum_{l\in L}\frac{(k+1)!}{l_1!\ldots l_b!}\prod_{j=1}^b \big(x_j^{\frac{q}{k+1}}\big)^{l_j}\\ &= \sum_{j=1}^b x_j^q + \sum_{l\in L\setminus L^*}\frac{(k+1)!}{l_1!\ldots l_b!}\prod_{j=1}^b x_j^{\frac{ql_j}{k+1}}\end{aligned}$$

If we apply this to $\Omega_n = \sum_{j=1}^b M_j\Omega_{n-1}(j)$ with $x_j = M_j\Omega_{n-1}(j)$ and take expectations we get:

$$\begin{aligned}\mathbb{E}\Omega_n^q &= \mathbb{E}\Bigg[\Big(\sum_{j=1}^b M_j\Omega_{n-1}(j)\Big)^q\Bigg]\\ &\leq \sum_{j=1}^b \mathbb{E}M^q\mathbb{E}\Omega_{n-1}^q + \sum_{l\in L\setminus L^*}\frac{(k+1)!}{l_1!\ldots l_b!}\prod_{j=1}^b \mathbb{E}M^{\frac{ql_j}{k+1}}\mathbb{E}\Omega_{n-1}^{\frac{ql_j}{k+1}}\\ &\leq b\mathbb{E}M^q\mathbb{E}\Omega_{n-1}^q + \sum_{l\in L\setminus L^*}\frac{(k+1)!}{l_1!\ldots l_b!}\prod_{j=1}^b \big(\mathbb{E}M^{l_j}\mathbb{E}\Omega_{n-1}^{l_j}\big)^{\frac{q}{k+1}}\end{aligned}$$

$$\begin{aligned}
&\leq \quad b\mathbb{E}M^q\mathbb{E}\Omega_{n-1}^q + \sum_{l\in L\setminus L^*} \frac{(k+1)!}{l_1!\dots l_b!}\left(\mathbb{E}M\mathbb{E}\Omega_{n-1}\mathbb{E}M^k\mathbb{E}\Omega_{n-1}^k\right)^{\frac{q}{k+1}} \\
&= \quad b\mathbb{E}M^q\mathbb{E}\Omega_{n-1}^q + (b^{k+1}-b)b^{-\frac{q}{k+1}}\left(\mathbb{E}M^k\mathbb{E}\Omega_{n-1}^k\right)^{\frac{q}{k+1}}
\end{aligned}$$

where we used $\sum_{l\in L}\frac{(k+1)!}{l_1!\dots l_b!} = b^{k+1}$ and $\mathbb{E}U^{\frac{q}{k+1}} \leq (\mathbb{E}U)^{\frac{q}{k+1}}$, which follows from Jensen's inequality and the concavity of $x^{\frac{q}{k+1}}$. We also used that $\prod_{j=1}^b \mathbb{E}U^{l_j} \leq \mathbb{E}U\mathbb{E}U^k$ for $U \geq 0$ and l_j such that $\sum_{j=1}^b l_j = k+1$ and at least two l_j are nonzero.

Jensen's inequality applied to the convex function x^q gives that Ω_n^q is a submartingale and hence $\mathbb{E}\Omega_n^q \geq \mathbb{E}\Omega_{n-1}^q$. If we apply this to aboves inequality we get:

$$\mathbb{E}\Omega_n^q(1-b\mathbb{E}M^q) \leq (b^{k+1}-b)b^{-\frac{q}{k+1}}\left(\mathbb{E}M^k\mathbb{E}\Omega_{n-1}^k\right)^{\frac{q}{k+1}} \tag{18}$$

Note that if we assume $\lim_{n\to\infty}\mathbb{E}\Omega_n^k < \infty$, we obtain by taking limits on both sides that $\lim_{n\to\infty}\mathbb{E}\Omega_n^q$ is also finite, where we used the assumption $\mathbb{E}M^q < b^{-1}$.

We will now first discuss the case $1 < q \leq 2$ (so $k = 1$). Remember that we had $\mathbb{E}\Omega < \infty$ and L^1 convergence of Ω_n to Ω. This implies

$$\lim_{n\to\infty}\mathbb{E}\Omega_n = \mathbb{E}\Omega < \infty$$

If we now take limits in (18), we obtain $\lim_{n\to\infty}\mathbb{E}\Omega_n^q < \infty$. Since Ω_n^q is a submartingale this is equivalent to $\sup_n \mathbb{E}\Omega_n^q < \infty$, and hence the sequence is bounded in L^1. Since this implies that the sequence Ω_n^q is uniformly bounded, we can use the following result (which is for instance proved in R.Schilling (2005)):

(18)

Let X_n be a submartingale that is uniformly bounded, then $X = \lim_{n\to\infty} X_n$ exists almost surely with $\lim_{n\to\infty}\mathbb{E}X_n = \mathbb{E}X < \infty$ (19)

(20)

This result establishes the required result $\mathbb{E}\Omega^q < \infty$ for $1 < q \leq 2$, and establishes next to this also that Ω_n^2 converges in L^1 to Ω^2.

Now consider the case $q > 2$. First note that $\mathbb{E}M^q < b^{-1}$ is equivalent to $\tau(q) > 0$. Since the scaling function $\tau(q)$ is strictly concave and $\tau(1) = 0$, the assumption $\tau(q) > 0$ implies that $\tau(l) > 0$ for $1 < l \leq q$. Hence $\mathbb{E}M^q < b^{-1}$ implies $\mathbb{E}M^l < b^{-1}$. Using this result, we obtain by applying the same reasoning as above, the following implications for $l = 2,\dots,k$:

$$\lim_{n\to\infty}\mathbb{E}\Omega_n^{l-1} = \mathbb{E}\Omega^{l-1} < \infty \quad \Rightarrow \quad \lim_{n\to\infty}\mathbb{E}\Omega_n^l = \mathbb{E}\Omega^l < \infty$$

By iteratively applying this implication we obtain $\lim_{n\to\infty}\mathbb{E}\Omega_n^k = \mathbb{E}\Omega^k < \infty$, which in turn implies the required result $\mathbb{E}\Omega^q < \infty$. $\diamond$

Corollary 2.2 *If $\mathbb{P}(M > 1) > 0$ there exists a q_{crit} with $1 \leq q_{crit} < \infty$ such that*

$$\mathbb{E}\Omega^q = \infty \quad \forall\ q > q_{crit}$$

Proof. Since $\mathbb{P}(M > 1) > 0$ there are $x > 1$, $\epsilon > 0$ such that $\mathbb{P}(M \geq x) > \epsilon$. Note that $x\epsilon < b^{-1}$ has to hold, because if $x\epsilon \geq b^{-1}$, then with the use of Markov's inequality we arrive at the contradiction $\mathbb{E}M \neq b^{-1}$:

$$\mathbb{E}M \geq x\mathbb{P}(M \geq x) > x\epsilon \geq b^{-1}$$

Now take $q^* = -\log_x(\epsilon b)$, then $q^* > 1$ because $x\epsilon < b^{-1}$:

$$(\epsilon b)^{-1} > x \quad \Rightarrow \quad -\log(\epsilon b) > \log x \quad \Rightarrow \quad -\log_x(\epsilon b) > 1$$

Now we will show that $\mathbb{E}M^{q*} > b^{-1}$:

$$\mathbb{E}M^{q*} = \int_0^\infty y^{q*} d\mathbb{P}^M(y) \geq \int_x^\infty y^{q*} d\mathbb{P}^M(y) \geq x^{q*}\mathbb{P}(M \geq x) > (\epsilon b)^{-1}\epsilon = b^{-1}$$

Note that this inequality also holds for all $q > q^*$, hence by proposition 2.6 $\mathbb{E}\Omega^q = \infty\ \forall q \geq q^*$. So there is a $q_{crit} < \infty$ such that $\forall\ q > q_{crit}\ \mathbb{E}\Omega^q = \infty$. We complete the proof by noting that since $\mathbb{E}\Omega < \infty$ it holds that $q_{crit} \geq 1$.$\diamond$ (21)

In the rest of this thesis we will use the definition $q_{crit} = \sup\{q : \mathbb{E}\Omega^q < \infty\}$, which is the smallest q for which the corollary holds.

As mentioned earlier, the fact that Ω can have diverging moments, reflects that Ω can have heavy tails. This is made more precise in Guivarc'h (1990), which shows that under the assumptions $\mathbb{E}M^q < \infty\ \forall\ q \geq 0$ and $\mathbb{P}(M > 1) > 0$, the total mass Ω has a Paretian right tail:

$$\mathbb{P}(\Omega > x) \sim cx^{-q_{crit}} \text{ as } \ x \to \infty$$

where c is a positive constant.[4]

2.3.3 Long range dependence

Long range dependence, also called long memory, is a property of a stochastic process related to a slow decay of the statistical dependence between different increments of this process. For a stochastic process X_t we expect the autocovariance $Cov[X_t, X_{t+\tau}]$ to decrease when τ increases, but in the case of long range dependence this decrease in covariance is slower than 'usual'. More formally, short range dependent processes are characterized by an autocorrelation function which decays exponentially fast, and processes with long range dependence have a much slower decay of the correlations. So

[4]The notation $\sim$ is defined as: let $f(x) \sim g(x)$ as $x \to \infty$, then $\lim_{x\to\infty} f(x)/g(x) = 1$

processes with long memory will show more persistence (or antipersistence) over long time spans than short memory processes.

If we define $\theta(t) = \mu([0,t])$, then $\theta(t)$ is a stochastic process which exhibits long range dependence in the size of its increments. Usually the concept of long range dependence is defined in an asymptotic sense: a stochastic process exhibits long memory when there is a power-like decay in its autocovariance as the time lag goes to infinity. However since $\theta(t)$ is only defined on a bounded interval, we need to define the concept of long memory in a slightly different way. Furthermore, because $\theta(t)$ has long memory in the absolute value of its increments, we have to use the autocovariance of the sizes of the increments instead of $Cov[\theta(t), \theta(t+\tau)]$. This covariance of the absolute values of the increments is called the autocovariance in levels and for a stochastic process $Z(t)$ it is defined by

$$\delta_Z(\tau, q, \Delta t) = Cov[|Z(t, \Delta t)^q|, |Z(t+\tau, \Delta t)^q|] \quad \text{for} \quad \tau > 0$$

where $Z(t, \Delta t) := Z(t+\Delta t) - Z(t)$ are the sizes of the increments. In our definition of long range dependence we will however only use the part of the covariance function that mainly 'quantifies' the dependence between the size of the increments:

$$\bar{\delta}_Z(\tau, q, \Delta t) = \mathbb{E}|Z(t, \Delta t)^q Z(t+\tau, \Delta t)^q|$$

This leads to the following definition:

Definition 2.1 *A stochastic process $Z(t)$ has long memory in the size of its increments if for every $q > 0$ with $\mathbb{E}Z(t)^q < \infty$ there exist strictly positive constants C_1, C_2, α and a strictly positive function $h(\Delta t)$, such that for small Δt the covariance function follows a power law in τ:*

$$C_1 h(\Delta t)\tau^{-\alpha} \leq \bar{\delta}_Z(\tau, q, \Delta t) \leq C_2 h(\Delta t)\tau^{-\alpha} \quad \textit{as} \quad \Delta t \to 0$$

Then for small Δt the decay in the autocovariance is slower than an exponential decay as τ increases. Using this definition we get the following proposition:

Proposition 2.7 *When μ is a canonical grid based multifractal measure, the time deformation process $\theta(t) = \mu([0,t])$ has long memory in the size of its increments for $0 < q < q_{crit}$.*

Proof. Consider a canonical measure in the n'th stage of its construction. We define the intervals $I_1 = [t_1, t_1 + b^{-n}]$ and $I_2 = [t_2, t_2 + b^{-n}]$ where $t_1 = \sum_{i=1}^{n} \eta_i b^{-i}$ and $t_2 = \sum_{i=1}^{n} \zeta_i b^{-i}$ are different b-adic numbers. We assume that the first $l \geq 1$ terms are equal in the b-adic expansions of t_1 and t_2, so $\eta_1 = \zeta_1, \ldots, \eta_l = \zeta_l$. For the $(l+1)$'th stage we assume that the intervals are at least one $(l+1)$'th stage b-adic interval apart: $\zeta_{l+1} - 1 < \eta_{l+1} < \zeta_{l+1} + 1$.

Then the distance $\tau = |t_2 - t_1|$ satisfies $b^{-(l+1)} < \tau < b^{-l}$, as we will show now.

Observe that

$$\begin{aligned}
\Big|\sum_{i=l+1}^{n}(\eta_i - \zeta_i)b^{-i}\Big| &\le \sum_{i=l+1}^{n}|\eta_i - \zeta_i|b^{-i} \\
&\le (b-1)\sum_{i=l+1}^{n} b^{-i} \\
&= (b-1)\Big(\sum_{i=0}^{n} b^{-i} - \sum_{i=0}^{l} b^{-i}\Big) \\
&= (b-1)\Big(\frac{1-b^{-(n+1)}}{1-b^{-1}} - \frac{1-b^{-(l+1)}}{1-b^{-1}}\Big) \\
&= b\big(1 - b^{-(n+1)} - (1 - b^{-(l+1)})\big) \\
&= b^{-l} - b^{-(n+1)} < b^{-l}
\end{aligned}$$

This establishes $\tau = \big|\sum_{i=l+1}^{n}(\eta_i - \zeta_i)b^{-i}\big| < b^{-l}$. To show $\tau > b^{-(l+1)}$, note that

$$\tau = \Big|\sum_{i=l+1}^{n}(\eta_i - \zeta_i)b^{-i}\Big| = \Big|(\eta_{l+1} - \zeta_{l+1})b^{-(l+1)} + \sum_{i=l+2}^{n}(\eta_i - \zeta_i)b^{-i}\Big|$$

Since $|\eta_{l+1} - \zeta_{l+1}|b^{-(l+1)} \ge 2b^{-(l+1)}$ and $\big|\sum_{i=l+2}^{n}(\eta_i - \zeta_i)b^{-i}\big| < b^{-(l+1)}$ we also establish $\tau > b^{-(l+1)}$.

To compute $\bar{\delta}_\theta(\tau, q, b^{-n})$, we first consider the product $\mu(I_1)^q\mu(I_2)^q$:

$$\begin{aligned}
\mu(I_1)^q\mu(I_2)^q \quad = \quad & (\Omega^q_{\eta_1,\dots,\eta_n}\Omega^q_{\zeta_1,\dots,\zeta_n})(M^{2q}_{\eta_1}\dots M^{2q}_{\eta_1,\dots,\eta_l}) \\
& (M^q_{\eta_1,\dots,\eta_{l+1}}\dots M^q_{\eta_1,\dots,\eta_n})(M^q_{\zeta_1,\dots,\zeta_{l+1}}\dots M^q_{\zeta_1,\dots,\zeta_n})
\end{aligned}$$

If we take expectations we obtain:

$$\begin{aligned}
\bar{\delta}_\theta(\tau, q, b^{-n}) = \mathbb{E}\mu(I_1)^q\mu(I_2)^q &= (\mathbb{E}\Omega^q)^2(\mathbb{E}M^{2q})^l(\mathbb{E}M^q)^{2(n-l)} \\
&= (\mathbb{E}\Omega^q)^2(\mathbb{E}M^q)^{2n}(\mathbb{E}M^{2q})^l(\mathbb{E}M^q)^{-2l}
\end{aligned}$$

We can write this in a different way by noting that $-\tau_\theta(q) - 1 = \log_b \mathbb{E}M^q = \log_{b^n}(\mathbb{E}M^q)^n$ which gives

$$(\mathbb{E}M^q)^n = (b^n)^{-\tau_\theta(q)-1} = (b^{-n})^{\tau_\theta(q)+1}$$

With this we obtain:

$$\begin{aligned}
\bar{\delta}_\theta(\tau, q, b^{-n}) &= (\mathbb{E}\Omega^q)^2(b^{-2n})^{\tau_\theta(q)+1}(b^{-l})^{\tau_\theta(2q)+1}(b^{2l})^{\tau_\theta(q)+1} \\
&= (\mathbb{E}\Omega^q)^2(b^{-n})^{2\tau_\theta(q)+2}(b^{-l})^{\tau_\theta(2q)-2\tau_\theta(q)-1}
\end{aligned}$$

If we define $\alpha = -(\tau_\theta(2q) - 2\tau_\theta(q) - 1)$, then $\alpha > 0$ since $\tau_\theta(q)$ is strictly concave:

$$-\alpha = \tau_\theta(2q) - 2\tau_\theta(q) - 1 = \tau_\theta(2q) + \tau_\theta(0) - 2\tau_\theta(q) < 0$$

If we also use that $b^{-(l+1)} < \tau < b^{-l}$ implies $b\tau > b^{-l} > \tau$, we get:

$$(\mathbb{E}\Omega^q)^2 (b^{-n})^{2\tau_\theta(q)+2} (b\tau)^{-\alpha} < \bar{\delta}_\theta(\tau, q, b^{-n}) < (\mathbb{E}\Omega^q)^2 (b^{-n})^{2\tau_\theta(q)+2} \tau^{-\alpha}$$

So if we take $h(\Delta t) = (\Delta t)^{2\tau_\theta(q)+2}$, $C_1 = (\mathbb{E}\Omega^q)^2 b^{-\alpha}$ and $C_2 = (\mathbb{E}\Omega^q)^2$, then

$$C_1 h(b^{-n}) \tau^{-\alpha} < \bar{\delta}_\theta(\tau, q, b^{-n}) < C_2 h(b^{-n}) \tau^{-\alpha} \qquad \diamond$$

2.4 Local Hölder exponents

In section 2.3 we introduced a global scaling relationship for multifractal measures. In the rest of this chapter we will study the local behaviour of μ. But first we will introduce in this section our main tool to study these local properties, the so-called local Hölder exponent. This is an important concept that describes the local behaviour of the paths of a function.

Definition 2.2 *Let g be a function defined on the neighbourhood of a given date t. The local Hölder exponent of g at time t is:*

$$\alpha(t) = \sup\{\beta : |g(t + \Delta t) - g(t)| = \mathcal{O}(|\Delta t^\beta|) \text{ as } \Delta t \to 0\} \tag{23}$$

This Hölder exponent, sometimes called the singularity exponent, exists always in $[-\infty, \infty]$. Its definition can be extended to measures defined on the real line: at a given date t, the local exponent of a measure is defined as the local exponent of its distribution function.

We can interpret local exponents in the following way: when the function g satisfies the scaling relation

$$|g(t + \Delta t) - g(t)| \sim C_t (\Delta t)^{\alpha(t)} \quad \text{as } \Delta t \to 0$$

with a positive constant C_t, then this $\alpha(t)$ is indeed the local Hölder exponent of g at time t. From this equation, we can easily compute the local exponents of several examples. For instance, for a point t where g is discontinuous $|g(t + \Delta t) - g(t)|$ will not go to zero, and hence $\alpha(t) = 0$. For t where g is differentiable with $g'(t) \neq 0$ the local exponent is equal to one, and for constant functions we have $\alpha(t) = \infty$. So when the function behaves smoothly at t the Hölder exponent is relatively high and when there is more irregularity the local exponent will be smaller. So the Hölder exponent can be viewed as a sort of 'measure' of the local regularity or smoothness of the paths of a function.

Since Brownian motion has sample paths which are continuous but not differentiable, it is a good example of a (random) function which has a local exponent between 0 and 1. In fact, the paths of Brownian motion are characterized by a local exponent equal to 1/2, which we will show now:

$$\begin{aligned}\mathbb{P}\Big(\lim_{\Delta t\to 0}\frac{|B(t+\Delta t)-B(t)|}{|\Delta t|^{\beta}}<\infty\Big) &= \mathbb{P}\Big(\lim_{\Delta t\to 0}\frac{|B(\Delta t)|}{|\Delta t|^{\beta}}<\infty\Big) = \\ \mathbb{P}\Big(\lim_{\Delta t\to 0}\frac{|\Delta t|^{1/2}|B(1)|}{|\Delta t|^{\beta}}<\infty\Big) &= \begin{cases} 1 & \text{if } \beta\le 1/2 \\ 0 & \text{if } \beta > 1/2\end{cases}\end{aligned}$$

From definition 2.2 it follows that $\alpha(t) = \sup\{(-\infty, 1/2]\} = 1/2$ almost surely. So the paths of Brownian motion (and therefore also all continuous-time Itô processes) have a unique local exponent equal to 1/2 for every time instant. In more generality, all continuous-time stochastic processes commonly used to model financial prices can each be characterized by a unique Hölder exponent. In contrast, multifractal measures contain a multiplicity of local exponents. The distribution of these exponents within a measure will be studied in the next sections, where we will use that local Hölder exponents can be computed in a more direct way:

Proposition 2.8 *Let g be a function defined on the neighbourhood of a given date t, then the local Hölder exponent of function g at date t is equal to:*

$$\limsup_{\Delta t\to 0}\frac{\ln(|g(t+\Delta t)-g(t)|)}{\ln(|\Delta t|)}$$

Proof. Define $\bar{\alpha}(t) := \limsup_{\Delta t\to 0}\frac{\ln(|g(t+\Delta t)-g(t)|)}{\ln(|\Delta t|)}$ and note that the $\alpha(t)$ defined in (23) can equivalently be written as:

$$\alpha(t) = \sup\big\{\beta : \exists M,\delta>0\ :\ \forall_{|\Delta t|<\delta}\ \ \frac{|g(t+\Delta t)-g(t)|}{|\Delta t|^{\beta}}\le M\big\} \tag{24}$$

To show $\bar{\alpha}(t)=\alpha(t)$ we will prove the following implications:

$$1)\ \ \beta<\bar{\alpha}(t)\ \Rightarrow\ \beta\le\alpha(t)\ \ \text{and}\ \ 2)\ \ \beta>\bar{\alpha}(t)\ \Rightarrow\ \beta\ge\alpha(t)$$

Since 1) implies $\bar{\alpha}(t)\le\alpha(t)$ and 2) implies $\bar{\alpha}(t)\ge\alpha(t)$, the required result $\bar{\alpha}(t)=\alpha(t)$ follows immediately.

First we will prove implication 1) by contradiction. Assume $\beta<\bar{\alpha}(t)$ and $\beta>\alpha(t)$. Then the latter assumption implies that for $0<\delta_0<1$ the following holds:

$$\exists_{|\Delta t|<\delta_0}\ :\ \frac{|g(t+\Delta t)-g(t)|}{|\Delta t|^{\beta}}>1$$

Which implies

$$\exists_{|\Delta t|<\delta_0}\ :\beta>\frac{\ln(|g(t+\Delta t)-g(t)|)}{\ln(|\Delta t|)} \tag{25}$$

Because $\beta < \bar{\alpha}(t)$, we also have:

$$\beta < \lim_{\delta \downarrow 0} \sup_{|\Delta t| < \delta} \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)} \leq \sup_{|\Delta t| < \delta_0} \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)}$$

It follows that there is no $|\Delta t| < \delta_0$ such that $\beta > \frac{\ln(|g(t+\Delta t)-g(t)|)}{\ln(|\Delta t|)}$. This contradicts with (25), hence $\beta < \bar{\alpha}(t) \;\Rightarrow\; \beta \leq \alpha(t)$.

Next we will show, again by contradiction, the second implication. Assume $\beta > \bar{\alpha}(t)$ and $\beta < \alpha(t)$. The first assumption means that:

$$\exists_{\eta_0 > 0} \;:\; \beta - \eta_0 = \lim_{\delta \downarrow 0} \sup_{|\Delta t| < \delta} \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)}$$

Take $0 < \epsilon < \eta_0$, then there is a $0 < \delta < 1$ such that:

$$0 \leq \sup_{|\Delta t| < \delta} \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)} - (\beta - \eta_0) < \epsilon$$

$$\text{Hence} \quad \beta > \sup_{|\Delta t| < \delta} \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)} + \eta_0 - \epsilon$$

And we obtain:

$$\exists_{0<\delta<1} \, \exists_{\eta_1 > 0} \;:\; \forall_{|\Delta t| < \delta} \;\; \beta > \frac{\ln(|g(t+\Delta t) - g(t)|)}{\ln(|\Delta t|)} + \eta_1$$

Now take this δ, η_1. Because $\beta > \beta'$ implies $\frac{1}{|\Delta t|^{\beta}} > \frac{1}{|\Delta t|^{\beta'}}$ for $|\Delta t| < 1$, it follows from aboves statement that:

$$\forall_{|\Delta t| < \delta} \quad \frac{|g(t+\Delta t) - g(t)|}{|\Delta t|^{\beta}} > \frac{|g(t+\Delta t) - g(t)|}{|\Delta t|^{\frac{\ln(|g(t+\Delta t)-g(t)|)}{\ln(|\Delta t|)} + \eta_1}} = \frac{1}{|\Delta t|^{\eta_1}}$$

Hence there is no M, δ such that $\forall_{|\Delta t| < \delta} \; \frac{|g(t+\Delta t)-g(t)|}{|\Delta t|^{\beta}} \leq M$. Because we also assumed $\beta < \alpha(t)$, according to (24) there should be such M, δ. So we have a contradiction and hence $\beta > \bar{\alpha}(t) \;\Rightarrow\; \beta \geq \alpha(t)$. $\diamond$

2.5 Coarse exponents

2.5.1 Coarse Hölder exponents

In this section we will use the local Hölder exponent to study the local behaviour of multifractal measures. We want to determine the probability that a randomly chosen point will have a given Hölder exponent and proposition 2.8 suggests a way to do this. We define the coarse Hölder exponent as follows:

Definition 2.3 *Let g be a (possibly random) function defined on the neighbourhood of a given date t and let $(\Delta_n t)$ be a sequence such that $\Delta_n t \to 0$. Then the coarse Hölder exponent of the function g over the interval $[t, t + \Delta_n t]$ is defined as:*

$$\alpha_n(t) = \frac{\ln(|g(t + \Delta_n t) - g(t)|)}{\ln(|\Delta_n t|)}$$

Note that when $\lim_{\Delta t \to 0} \ln(|g(t + \Delta t) - g(t)|)/\ln(\Delta t)$ exists, $\alpha_n(t)$ converges to the local Hölder exponent $\alpha(t)$. In the rest of this thesis we will assume that for the measure μ the coarse exponents $\alpha_n(t)$ indeed converge to the local Hölder exponent. Although we will be able to prove that the coarse exponents converge, it is difficult to prove that the limit of $\ln(|g(t + \Delta t) - g(t)|)/\ln(\Delta t)$ exists and is often assumed in multifractal literature. So although the assumption that $\alpha_n(t)$ converges to $\alpha(t)$ is not rigorously proven, we will in the rest of the thesis assume that this indeed the case. Remark however that even in the case that $\alpha_n(t)$ would not converge to $\alpha(t)$, it still makes sense to speak about the limit of $\alpha_n(t)$ as local exponents, since we expect the limit coarse exponents to have similar properties as the local exponents.

When we take $\Delta_n t = b^{-n}$, the coarse exponent of the measure μ over the b-adic interval $[t_i, t_i + b^{-n}]$ becomes $\alpha_n(t_i) = \ln(\mu([t_i, t_i + b^{-n}]))/\ln(b^{-n})$. So for each b-adic interval the coarse exponent can be computed and we can study which values they take. To do this we define $N_n(\alpha, \epsilon)$ as the number of coarse Hölder exponents in the interval $[\alpha - \epsilon, \alpha + \epsilon)$ in the n'th stage:

$$N_n(\alpha, \epsilon) = \#\{i = 0, \ldots, b^n - 1 : \alpha_n(t_i) \in [\alpha - \epsilon, \alpha + \epsilon)\}$$

The ratio $N_n(\alpha, \epsilon)/b^n$ can be viewed as the relative frequency of the coarse exponents approximately equal to α. In the limit, as $n \to \infty$, we have the following heuristic relation:

$$\frac{N_n(\alpha, \epsilon)}{b^n} \sim \mathbb{P}(\alpha_n(t) \in [\alpha - \epsilon, \alpha + \epsilon)) \quad \text{as } n \to \infty \tag{26}$$

If we also take the limit $\epsilon \to 0$, this ratio converges to the probability that a randomly selected point t has Hölder exponent α. Note that this probability is the same as the Lebesque measure of the set of points having Hölder exponent α. To determine this probability we will study the asymptotic statistical properties of the coarse Hölder exponent.

Let μ be a canonical measure, then for b-adic numbers t and $\Delta t = b^{-n}$ remember that we had the following equality:

$$\mu([t, t + \Delta t]) = M(\eta_1) \ldots M(\eta_1, \ldots, \eta_n)\, \Omega$$

Using this we can express the coarse Hölder exponent as a function of Ω and the multipliers M:

$$\begin{aligned}\alpha_n(t) &= \frac{\ln(\mu([t, t+\Delta t]))}{\ln(\Delta t)} \\ &= -\frac{1}{n}\big(\log_b M(\eta_1) + \ldots + \log_b M(\eta_1, \ldots, \eta_n) + \log_b \Omega\big) \qquad (27)\end{aligned}$$

Since we are interested in the asymptotic properties of the coarse exponents, the factor $-1/n \log_b \Omega$, which converges to zero almost surely, does not play a significant role. Therefore we may assume that μ is conservative ($\Omega = 1$), but all results obtained in this section and the next will hold for canonical measures as well.

In the preceeding we have implicitly assumed that we can view the coarse Hölder exponents at stage n as realisations of a random variable α_n. We will now specify how this can be done for both deterministic and random measures. For deterministic measures (such as the binomial) we consider the mass of a random b-adic cell. So although the allocation of mass over all the cells is deterministic, we can by randomly drawing $\eta_1, \ldots, \eta_n$ for the b-adic number t, consider the coarse exponents $\alpha_n(t)$ as draws of a random variable α_n.

When the measure is randomly generated, the coarse exponents $\alpha_n(t)$ are identically distributed across al b-adic cells (since the multipliers M are independent and identically distributed). So all the cells are essentially the same and we can choose a fixed cell $[t, t+b^{-n}]$. The corresponding measure $\mu([t, t+b^{-n}])$ and coarse exponent $\alpha_n(t)$ will be random. So the exponents $\alpha_n(t)$ can again be viewed as draws of a random variable α_n.

This random variable α_n is according to (27) the sum of n independent and identically distributed random variables. If we write $-\log_b M(\eta_1, \ldots, \eta_i)$ as V_i, then α_n becomes:

$$\alpha_n = \frac{1}{n}\sum_{i=1}^{n} V_i$$

Since α_n is the sum of i.i.d. random variables, there are many techniques available to study them. One important technique is the familiar Strong Law of Large Numbers (SLLN), but next to this we can also use the possibly less familiar Large Deviation Theory. This theory provides information on the tail of sums of i.i.d. random variables, and will turn out to play a central role in the analysis of these coarse exponents.

Let us first put the Law of Large Numbers to use. According to the SLLN, the random variable α_n converges in the almost sure sense to $\mathbb{E}V_1 = -\mathbb{E}\log_b M$, which will be denoted by α_0. So as n goes to infinity we expect that almost all coarse Hölder exponents are contained in an increasingly small neighbourhood of α_0. This implies that the ratio $N_n(\alpha, \epsilon)/b^n$ collapses and converges as $n \to \infty$ and $\epsilon \to 0$ to the indicator function $\mathbf{1}_{\alpha_0}(\alpha)$. So it follows that $\mathbb{P}(\alpha(t) = \alpha_0) = 1$ and hence the Lebesque measure of instants t with local exponent α_0 is equal to one: $\lambda(\{t \in [0,1] : \alpha(t) = \alpha_0\}) = 1$. This means that we can conclude that the single Hölder exponent α_0 dominates, in the sense that the set of instants with exponent α_0 carries all of the Lebesque measure.

2.5.2 Mass concentration on sets of Lebesque measure zero

Since we established $\mathbb{P}(\alpha(t) = \alpha_0) = \lambda(\{t \in [0,1] : \alpha(t) = \alpha_0\}) = 1$, it might seem that this is the end of the story and we are done. But remember that we argued (but not yet proved) in section 2.4.1 that grid based multifractal measures are not absolotely continuous with respect to the Lebesque measure. This means there is mass concentrated on sets with Lebesque measure zero. So although the exponent α_0 carries all of the Lebesque measure, the other exponents do matter. In fact, as we will prove now, the mass concentrates on sets with local exponents that are bounded away from α_0. This also proves that multifractal measures are not absolutely continuous with respect to the Lebesque measure.

Proposition 2.9 *All the mass of a multifractal measure μ is concentrated on sets of Lebesque measure zero.*

Proof. First we will show that $\alpha_0 > 1$ (which is already interesting on it self). Since the function $-\log_b(x)$ is strictly convex, we can use Jensen's inequality and get:

$$\alpha_0 = -\mathbb{E}\log_b M > -\log_b \mathbb{E}M = -\log_b b^{-1} = 1$$

Let T_n be the set of b-adic numbers t such that the b-adic cell $[t, t+b^{-n}]$ has Hölder exponent greater than $(1+\alpha_0)/2$. Since $N_n(\alpha_0,\epsilon)/b^n$ goes to one and $(1+\alpha_0)/2 < \alpha_0$, "almost all" cells belong to T_n for large values of n. However, the mass of the measure concentrates on cells with local exponents that are bounded away from α_0:

$$\sum_{t\in T_n} \mu([t, t+b^{-n}]) = \sum_{t\in T_n} \left(b^{-n}\right)^{\alpha_n(t)} \le \sum_{t\in T_n} \left(b^{-n}\right)^{\frac{1+\alpha_0}{2}} \le$$
$$b^n \left(b^{-n}\right)^{\frac{1+\alpha_0}{2}} = b^{-n\frac{\alpha_0-1}{2}} \to 0 \quad \text{as } n \to \infty$$

It follows that the mass is not concentrated on the set $\{t \in [0,1] : \alpha(t) = \alpha_0\}$ which has Lebesque measure 1. Hence the mass has to be concentrated on sets of Lebesque measure zero. $\diamond$ (27)

Remark that the above proposition means that events that occur on sets of Lebesque measure zero are responsible for all of the total variation of the measure μ and hence also the trading time $\theta(t)$. This is by itself not that special, for instance Poisson processes or other discontinuous processes also have this property. However, it may be a suprise that, for a continuous stochastic process, events occuring on sets of Lebesque measure zero can contribute all of the total variation. This is the case with multifractals.

2.5.3 Multifractal spectrum and fractal dimension

So far we have used the intuitive quantity $N_n(\alpha, \epsilon)/b^n$ to determine the distribution of the Hölder exponents within a path and found that multifractals have Lebesque-almost surely one unique local exponent. In section 2.5 we stated however that multifractals contain a multiplicity of local exponents, which distinguished them from unifractals such as Brownian motion. So the method using $N_n(\alpha, \epsilon)/b^n$ fails to distinguish between multifractals and unifractals. Somehow we have to find a method to study the other Hölder exponents, that will lie on sets of Lebesque measure zero. Mandelbrot proposed in Mandelbrot (1989) a so-called renormalization which solves this problem. Instead of studying the limit of $N_n(\alpha, \epsilon)/b^n$, we will study the limit of this quantity after taking logarithms. This leads to a concept called the multifractal spectrum:

Definition 2.4 *The limit*

$$f(\alpha) = \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{\ln N_n(\alpha, \epsilon)}{\ln b^n}$$

represents a renormalized probability distribution of local Hölder exponents, and is called the multifractal spectrum.

Note that aboves definition is defined for multifractal measures, but may with some minor modifications be extended to functions or random processes. The multifractal spectrum may also be used to give a definition of multifractality: if $f(\alpha)$ is well-defined (in the sense that the double limit exists) and is positive on a support larger than a single point, then we say that the corresponding measure or function is a multifractal.

To provide some intuition on why this renormalization is useful, consider the example with $N_n(\alpha, \epsilon) = 3^n$ and $b = 4$. Then the frequency $N_n(\alpha, \epsilon)/b^n = (3/4)^n$ will converge to zero. However, the expression for the multifractal spectrum $\ln N_n(\alpha, \epsilon)/\ln b^n = \ln 3/\ln 4$ does not vanish as $n \to \infty$. So by taking a logarithmic transform of the frequency $N_n(\alpha, \epsilon)/b^n$, we can study events that happen at a vanishing frequency, but nonetheless happen very many times.

Besides that we can interpret the method of taking a logarithmic transform of the frequency representation $N_n(\alpha, \epsilon)/b^n$ as a way to identify some (otherwise) vanishing events, the multifractal spectrum has also an intuitive interpretation as the fractal dimension of the set of points having a local exponent α. This interpretation was introduced in Frisch and Parisi (1985), for a class of multifractals which satisfy $f(\alpha) \geq 0$ (negative dimensions are not defined). For the reader who might not be familiar with fractal sets and fractal dimensions, we will now give an introduction.

In 1982 Benoit Mandelbrot introduced the concept of fractal dimension in his book *The fractal geometry of nature.* Fractal geometry considers irregular and non-smooth structures, called fractals, which are not well defined by their Euclidian length and the usual topological dimension. Many phenomena in nature, such as mountains, coastlines, the structures of plants, blood vessels, the clustering of galaxies and Brownian motion are better described using a non-integer (and thus fractal) dimension.

For instance when we measure the length of a coastline we find that the measured length might increase dramatically when the precision of the measurement is increased. In fact, as the length scale that is used for the measurement goes to zero, the measured length will diverge to infinity. This means first of all that the Euclidian length is not the proper tool to compare different coastlines. And secondly, since the length of these coastlines is infinite on any interval, the topological dimension also does not give a good characterization of this graph. Therefore new concepts of dimension were introduced which are capable of taking non-integer values. One famous dimension is the Hausdorff dimension, and fractals are usually defined as objects which have a noninteger Hausdorff dimension.

We will give a somewhat informal definition of the Hausdorff dimension. Consider the minimal number of $\bar{N}(r)$ balls of radius r required to cover a fractal set or fractal curve completely. In many cases $\bar{N}(r)$ satisfies a scaling law as r goes to zero:

$$\bar{N}(r) \sim C_r r^{-D}$$

For example for the unit square in $\mathbb{R}^2$, $\bar{N}(r)$ grows as r^{-2}. When objects are very irregular, $N(r)$ will typically increase much faster as r decreases than for smoother objects. Hence this D in the power law gives an indication of the degree of irregularity. If a object satisfies the relation $\bar{N}(r) \sim C_r r^{-D}$ as $r \to 0$, then D is called the Hausdorff dimension of this object. It is easy to derive that we can compute D more directly by taking logarithms:

$$D = \lim_{r \downarrow 0} \frac{\ln \bar{N}(r)}{\ln\left(\frac{1}{r}\right)}$$

The paths of a Brownian motion are a nice example of paths which behave very roughly and have infinite length on any interval. This suggests that Brownian motion is a fractal, which is indeed the case as its fractal Hausdorff dimension is equal to 1,5. Another classical example of a fractal is the Cantor set. For those who are not familiar with this set, it is generated by removing the open middle third of the interval $[0, 1]$, and then removing the middle third of each of the two remaining pieces. This procedure is continued into infinity and generates a set with some special properties. Since the

Cantor set is a subset of the interval $[0,1]$ and behaves very irregularly, one can expect that it has a fractal dimension between zero and one. And indeed, the Hausdorff dimension is equal to $\log 2/\log 3 < 1$. Next to this it can also be shown that this set is uncountable, but yet contains no intervals. So although it has the same cardinality as the interval $[0,1]$, it has Lebesque measure equal to zero. In more generality, the Lebesque measure of sets in $\mathbb{R}$ with Hausdorff dimension between zero and one is equal to zero.

Now we gave a brief introduction into the concept of fractal geometry, we will make it intuitively clear why the multifractal spectrum can be interpreted as a fractal dimension. Under the assumption that μ contains a multiplicity of Hölder exponents, it seems very unlikely that there are intervals in $[0,1]$ where all points have the same Hölder exponent. This is because the recursive procedure for the construction of μ continues into infinity and is the same for all b-adic intervals. It follows that one might expect that, in this limit, the b-adic intervals which have coarse Hölder exponent α will contain only one point with local Hölder exponent α. So in the limits as $n \to \infty$, the quantity $N_n(\alpha, \epsilon)$ can be considered as the minimal number of balls $\bar{N}(r)$ with length $r = b^{-n}$ needed to cover the points with local exponent in $[\alpha - \epsilon, \alpha + \epsilon)$. Now if we take logarithms and let ϵ and b^{-n} go to zero we obtain:

$$f(\alpha) = \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{\ln N_n(\alpha, \epsilon)}{\ln b^{-n}} = \lim_{r \downarrow 0} \frac{\ln \bar{N}(r)}{\ln \left(\frac{1}{r}\right)}$$

Hence $f(\alpha)$ can be interpreted as the fractal dimension of the set of points with Hölder exponents α. However, this interpretation is not correct for α for which $f(\alpha)$ is negative, because fractal dimensions can not be negative. So although this interpretation is not always correct, it is proved in Frisch and Parisi (1985) and Halsey et al (1986) that for the class of multifractals with $f(\alpha) \geq 0$ for all α, the fractal dimension of the sets $T(\alpha) = \{t : \alpha(t) = \alpha\}$ coincides with the multifractal spectrum $f(\alpha)$.

To conclude this section we mention that fractals are in general characterized by some form of self-similarity. An object is said to be self-similar when it is exactly or approximately similar to a part of itself, in the sense that the whole has the same shape as one or more of its parts. Since the level sets of the local exponents are fractal sets we might expect that they also have a self-similar structure. This is indeed the case, since multifractal measures also satisfy a form of scale-invariance: μ has the same statistical properties across all b-adic cells, and thus across all scales.

2.6 Multifractal formalism

In this section we will show that the multifractal spectrum can be directly related to the scaling function of section 2.2. They are equivalent in the sense that they both contain the same amount of information of the measure μ

and one can obtain $f(\alpha)$ from $\tau(q)$, and $\tau(q)$ from $f(\alpha)$. The relation they satisfy is called the multifractal formalism and this formalism is said to hold when the multifractal spectrum coincides with the Legendre transform of the scaling function:

$$f(\alpha) = \inf_q [\alpha q - \tau(q)]$$

In this section we will use a result from Large Deviation Theory to prove the above identity. There are however other methods to do this. In fact, there is an extensive literature initiated by Frisch and Parisi (1985) and Halsey et al (1986), which define the multifractal spectrum directly as a fractal dimension. Using this definition they also prove that the multifractal formalism holds, but we will follow the proof given in Mandelbrot (1997b). This proof uses the statistical properties of α_n and renormalization, which allows for a more intuitive understanding. In this proof some properties of the Legendre transform will be used, which we will prove first:

Lemma 2.1 *The Legendre transform $g(\alpha) = \inf_{q\in\mathbb{R}}[q\alpha - \tau(q)]$ of a twice differentiable function $\tau(q)$ with $\tau''(q) < 0$ is continuous and strictly concave.*

Proof. First note that since the second derivative $\tau''(q)$ is strictly negative, $-\tau(q)$ is strictly convex and the first derivative $\tau'(q)$ is strictly decreasing. To study the Legendre transform $\inf_{q\in\mathbb{R}}[q\alpha - \tau(q)]$ we will first show that we can write it in a more direct way. To accomplish this we have to find the minimum of $q\alpha - \tau(q)$. Since $\tau''(q) < 0$, the second derivative of $q\alpha - \tau(q)$ is strictly positive, so a minimum is attained at the point where its derivative is zero. Since $q\alpha - \tau(q)$ is the sum of two convex functions, $q\alpha - \tau(q)$ itself is also convex, and thus the minimum is a global minimum. So the global minimum of $q\alpha - \tau(q)$ can be found by setting its derivative equal to zero:

$$\frac{\partial}{\partial q} q\alpha - \tau(q) = \alpha - \tau'(q) = 0$$

Because $\tau'(q)$ is strictly decreasing and differentiable with $\tau''(q) < 0$, it has a differentiable inverse[5]. So the minimum is located at the point $q = \tau'^{-1}(\alpha)$, the inverse of the derivative at α, which gives

$$g(\alpha) = \tau'^{-1}(\alpha)\alpha - \tau(\tau'^{-1}(\alpha))$$

This expression makes it possible to compute the derivative of $g(\alpha)$:

$$\begin{aligned}
g'(\alpha) &= \frac{\partial}{\partial\alpha}\Big(\tau'^{-1}(\alpha)\Big)\alpha + \tau'^{-1}(\alpha) - \tau'\big(\tau'^{-1}(\alpha)\big)\frac{\partial}{\partial\alpha}\Big(\tau'^{-1}(\alpha)\Big) \\
&= \frac{\partial}{\partial\alpha}\Big(\tau'^{-1}(\alpha)\Big)\alpha + \tau'^{-1}(\alpha) - \alpha\frac{\partial}{\partial\alpha}\Big(\tau'^{-1}(\alpha)\Big) \\
&= \tau'^{-1}(\alpha)
\end{aligned}$$

[5]The inverse $h^{-1}(x)$ of a one-to-one and differentiable function $h(x)$ is differentiable at points x such that $h'(x) \neq 0$.

Since $\tau'(q)$ is strictly decreasing it follows that its inverse is also strictly decreasing. So we conclude that $g'(\alpha)$ is strictly decreasing, which implies that $g(\alpha)$ is strictly concave. We complete the proof by noting that the continuity of $g(\alpha)$ follows easily from $g(\alpha) = \tau'^{-1}(\alpha)\alpha - \tau(\tau'^{-1}(\alpha))$, because τ and the inverse τ'^{-1} are continuous. $\diamond$ (28)
(29)

Next to these properties the main ingredient of the proof of the multifractal formalism is a result from Large Deviation Theory. In section 2.5.2 we showed that the mass of a multifractal concentrates on sets with Hölder exponents that are bounded away from the predominant exponent α_0. Information on these sets is contained in the asymptotic tail probabilities of the coarse exponent α_n. This is where the Large Deviation Theory comes in to play, since it concerns the asymptotic tail behaviour of sums of random variables. To prove the multifractal formalism we will use the following result of Large Deviation Theory, which was established in 1938 by H. Cramér under conditions that were gradually weakened.

Cramér's theorem. *Let $\{X_k\}$ denote a sequence of i.i.d. random variables. Then for any $\alpha > \mathbb{E}X_1$ and $c > 0$*

$$\frac{1}{k}\log_c \mathbb{P}\Big(\frac{1}{k}\sum_{i=1}^{k} X_i > \alpha\Big) \to \inf_{q\in\mathbb{R}} \log_c \mathbb{E}e^{q(\alpha - X_1)} \quad \text{as } k \to \infty$$

There are many proofs of this theorem in the literature, including for instance in Deutschel and Stroock (1989). Note that since the coarse exponent α_n is the sum of i.i.d. random variables V_i, we can apply Cramér's theorem to α_n to study its asymptotic tail properties.

Theorem 2.1 *The multifractal spectrum $f(\alpha)$ is the Legendre transform of the scaling function $\tau(q)$:*

$$f(\alpha) = \inf_{q\in\mathbb{R}}[\alpha q - \tau(q)]$$

Proof. We start by showing that the multifractal spectrum is the double limit of $\frac{1}{n}\log_b \mathbb{P}(\alpha_n \in [\alpha - \epsilon, \alpha + \epsilon)) + 1$. One can easily derive the following equality:

$$\frac{\ln N_n(\alpha,\epsilon)}{\ln b^n} = \frac{1}{n}\log_b \frac{N_n(\alpha,\epsilon)}{b^n} + 1$$

Furthermore, remember that we had the heuristic relation (26) which can be rewritten as

$$\frac{1}{n}\log_b \frac{N_n(\alpha,\epsilon)}{b^n} \sim \frac{1}{n}\log_b \mathbb{P}(\alpha_n \in [\alpha - \epsilon, \alpha + \epsilon)) \quad \text{as } n \to \infty$$

By combining the above two relations we obtain

$$f(\alpha) = \lim_{\epsilon \downarrow 0} \lim_{n \to \infty} \frac{1}{n} \log_b \mathbb{P}(\alpha_n \in [\alpha - \epsilon, \alpha + \epsilon)) + 1 \tag{31}$$

We will study the right-hand side by considering $\frac{1}{n} \log_b \mathbb{P}(\alpha_n > \alpha - \epsilon)$. First assume that $\alpha > \alpha_0$ and take ϵ small enough such that $\alpha - \epsilon > \alpha_0$. Now we can apply Cramér's theorem to the above probability. Using that $\alpha_n = 1/n \sum_{i=1}^n V_i$, $V_i = -\log_b M_i$ and $\alpha_0 = \mathbb{E}V_1$, we obtain:

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \log_b \mathbb{P}(\alpha_n > \alpha - \epsilon) &= \lim_{n \to \infty} \frac{1}{n} \log_b \mathbb{P}\Big(\frac{1}{n} \sum_{i=1}^n V_i > \alpha - \epsilon\Big) \\ &= \inf_{q \in \mathbb{R}} \log_b \mathbb{E} e^{q(\alpha - \epsilon - V_1)} \\ &= \inf_{q \in \mathbb{R}} \log_b \mathbb{E} e^{q(\alpha - \epsilon - V_1) \ln b} \end{aligned}$$

where in the last step we substituted $q \ln b$ for q.

For the case $\alpha < \alpha_0$ and ϵ small enough such that $\alpha + \epsilon < \alpha_0$ we will consider $\frac{1}{n} \log_b \mathbb{P}(\alpha_n < \alpha + \epsilon)$. If we take $X_i = -V_i$ and $x = -(\alpha + \epsilon)$, then $x > -\alpha_0 = \mathbb{E}X_1$ and we can apply Cramér's theorem again:

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} \log_b \mathbb{P}(\alpha_n < \alpha + \epsilon) &= \lim_{n \to \infty} \frac{1}{n} \log_b \mathbb{P}\Big(\frac{1}{n} \sum_{i=1}^n X_i > x\Big) \\ &= \inf_{q \in \mathbb{R}} \log_b \mathbb{E} e^{q(x - X_1)} \\ &= \inf_{q \in \mathbb{R}} \log_b \mathbb{E} e^{q(\alpha + \epsilon - V_1) \ln b} \end{aligned}$$

where in the last step we substituted $-q \ln b$ for q.

This all gives that $1/n \log_b \mathbb{P}(\alpha_n > \alpha - \epsilon)$ and $1/n \log_b \mathbb{P}(\alpha_n < \alpha + \epsilon)$ for respectively $\alpha - \epsilon > \alpha_0$, $\alpha + \epsilon < \alpha_0$ converge to:

$$\begin{aligned} \delta(\alpha \pm \epsilon) := \inf_{q \in \mathbb{R}} \log_b \mathbb{E} e^{q(\alpha \pm \epsilon - V_1) \ln b} &= \inf_{q \in \mathbb{R}} \log_b \mathbb{E} b^{q(\alpha \pm \epsilon - V_1)} \\ &= \inf_{q \in \mathbb{R}} [q(\alpha \pm \epsilon) + \log_b \mathbb{E} M^q] \\ &= \inf_{q \in \mathbb{R}} [q(\alpha \pm \epsilon) - \tau(q)] - 1 \end{aligned} \tag{32}$$

where we used that $\tau(q) = -\log_b \mathbb{E} M^q - 1$.

Let us again consider the case $\alpha > \alpha_0$. Note that we already obtained the Legendre transform expression, but we derived it for $\mathbb{P}(\alpha_n > \alpha - \epsilon)$ instead of $\mathbb{P}(\alpha - \epsilon < \alpha_n < \alpha + \epsilon)$. Therefore we will also have to show the following relation:

$$\mathbb{P}(\alpha - \epsilon < \alpha_n < \alpha + \epsilon) \sim \mathbb{P}(\alpha_n > \alpha - \epsilon) \tag{33}$$

We can write the left-hand side as follows:

$$\begin{aligned}\mathbb{P}(\alpha-\epsilon<\alpha_n<\alpha+\epsilon) &= \mathbb{P}(\alpha_n>\alpha-\epsilon)-\mathbb{P}(\alpha_n>\alpha+\epsilon)\\ &= \mathbb{P}(\alpha_n>\alpha-\epsilon)\left(1-\frac{\mathbb{P}(\alpha_n>\alpha+\epsilon)}{\mathbb{P}(\alpha_n>\alpha-\epsilon)}\right)\end{aligned}$$

Because $\mathbb{P}(\alpha_n > x) = (b^n)^{\frac{1}{n}\log_b \mathbb{P}(\alpha_n>x)} \sim (b^n)^{\delta(x)}$ we get:

$$\frac{\mathbb{P}(\alpha_n>\alpha+\epsilon)}{\mathbb{P}(\alpha_n>\alpha-\epsilon)} \sim \frac{(b^n)^{\delta(\alpha+\epsilon)}}{(b^n)^{\delta(\alpha-\epsilon)}} = b^{n(\delta(\alpha+\epsilon)-\delta(\alpha-\epsilon))}$$

If we show that $\delta(\alpha+\epsilon) < \delta(\alpha-\epsilon)$, then $b^{n(\delta(\alpha+\epsilon)-\delta(\alpha-\epsilon))} \to 0$ and we establish (33). First we show $\delta(\alpha+\epsilon) \leq \delta(\alpha-\epsilon)$:

$$\delta(\alpha+\epsilon) = \lim_{n\to\infty}\frac{1}{n}\log_b \mathbb{P}(\alpha_n>\alpha+\epsilon) \leq \lim_{n\to\infty}\frac{1}{n}\log_b \mathbb{P}(\alpha_n>\alpha-\epsilon) = \delta(\alpha-\epsilon)$$

To get a strict inequality, we use that since the second derivative of $\tau(q)$ is strictly negative. Then lemma 2.1 states that the Legendre transform of $\tau(q)$ is strictly concave. This means that for small enough ϵ we have the following inequality:

$$\delta(\alpha+\epsilon) = \inf[q(\alpha+\epsilon)-\tau(q)]-1 \neq \inf[q(\alpha-\epsilon)-\tau(q)]-1 = \delta(\alpha-\epsilon)$$

Note however that aboves inequality is only guaranteed to hold when α is not the point of maximum of the function δ. Since δ is strictly concave it has a unique maximum and since $\delta(\alpha_0) \geq \delta(\alpha)$ for $\alpha > \alpha_0$, the function δ can not attain its maximum at any $\alpha > \alpha_0$.

Now we have established (33) we can combine this with (32) and get:

$$\lim_{n\to\infty}\frac{1}{n}\log_b \mathbb{P}(\alpha-\epsilon<\alpha_n<\alpha+\epsilon)+1 = \inf_{q\in\mathbb{R}}[q(\alpha-\epsilon)-\tau(q)]$$

Since it is proved in the appendix that the Legendre transform of $\tau(q)$ is continuous, we can take the limit $\epsilon \downarrow 0$ and use (31) to obtain the required result for $\alpha > \alpha_0$:

$$f(\alpha) = \inf_{q\in\mathbb{R}}[q\alpha-\tau(q)]$$

In the same way as for $\alpha > \alpha_0$ one can show that a similar version of (33) also holds for $\alpha < \alpha_0$, and thus the above results hold also for $\alpha < \alpha_0$. And finally, the continuity of the Legendre transform implies that the required result also holds for $\alpha = \alpha_0$. $\diamond$

(33)

Since the Legendre transform of the Legendre transform of a function g, is again equal to this function g, we also have that $\tau(q)$ is the Legendre transform of $f(\alpha)$: $\tau(q) = \inf_{\alpha\in\mathbb{R}}[\alpha q - f(\alpha)]$. So we can indeed obtain

$f(\alpha)$ from $\tau(q)$, and $\tau(q)$ from $f(\alpha)$. This means that, since $f(\alpha)$ and $\tau(q)$ describe the local and global behaviour respectively, the local and global scaling properties contain an equivalent amount of information about the measure μ.

In the above proof we have established that the multifractal spectrum is the Legendre transform of $\tau(q)$ and thus strictly concave, but actually more can be said. In section 2.5.3 we defined $f(\alpha)$ as the limit of the renormalized histogram of coarse Hölder exponents, but in the proof we showed that $f(\alpha)$ can equivalently be defined as the limit of:

$$\frac{1}{n}\log_b \mathbb{P}(\alpha_n > \alpha) + 1 \quad \text{if } \alpha > \alpha_0$$
$$\frac{1}{n}\log_b \mathbb{P}(\alpha_n < \alpha) + 1 \quad \text{if } \alpha < \alpha_0$$

Using this definition one can easily see that $f(\alpha)$ increases for $\alpha < \alpha_0$ and decreases for $\alpha > \alpha_0$. Next to this it is also easy to see that $f(\alpha) \leq 1$, which makes sense because we do not expect the fractal dimension of a set on $[0, 1]$ to be larger than 1. Furthermore, since the set of instants with local exponent α_0 has Lebesque measure 1, and the Lebesque measure of sets with a fractal dimension smaller than one is zero, we would also expect that the fractal dimension of the set of points with exponent α_0 is equal to one: $f(\alpha_0) = 1$. This is indeed the case as we will show now:

First note that $\tau'(0) = -\mathbb{E}\log_b M = \alpha_0$, so

$$f(\alpha_0) = f(\tau'(0)) = \inf_{q\in\mathbb{R}}[q\tau'(0) - \tau(q)]$$

Since $\tau''(q) < 0$, the second derivative of $q\tau'(0) - \tau(q)$ is strictly positive, so the minimum is attained at the point where its derivative is zero: $\tau'(0) - \tau'(q) = 0$. Since $\tau'(q)$ is strictly decreasing this equation has a unique solution given by $q = 0$. Now substituting $q = 0$ in aboves expression for $f(\alpha_0)$ gives $f(\alpha_0) = -\tau(0) = 1$.

2.7 The carrier of the mass

In section 2.5.2 we showed that the mass of a multifractal measure concentrates on sets with local exponents that are bounded away from α_0. In this section we will make this result more precise, by proving that in fact all the mass concentrates on a set characterized by a single Hölder exponent D. This result was rigorously proved in Kahane and Peyriere (1976), but their proof was very technical. We will give a less rigorous proof, but it will be more intuitive and relates to the concept of multifractal concentration which we will discuss at the end of this section. To prove the statement we will first prove the following lemma, which states that there is a set such that its fractal dimension and its corresponding local exponent are equal.

Lemma 2.2 *There is a unique α such that $f(\alpha) = \alpha$, and all other α's satisfy $f(\alpha) < \alpha$.*

Proof. First we will show that $\alpha = \tau'(1)$ is a solution of $f(\alpha) = \alpha$. According to lemma 2.1 we can write $f(\alpha)$ as follows:

$$f(\alpha) = \tau'^{-1}(\alpha)\alpha - \tau(\tau'^{-1}(\alpha))$$

This gives:

$$f(\tau'(1)) = \tau'^{-1}(\tau'(1))\tau'(1) - \tau\big(\tau'^{-1}(\tau'(1))\big) = \tau'(1) - \tau(1) = \tau'(1)$$

where we used the definition of an inverse and used that $\tau(1) = -\log_b \mathbb{E}M - 1 = -\log_b b^{-1} - 1 = 0$. So $\tau'(1)$ is a solution.

Now assume there is another $\alpha^* \neq \tau'(1)$ such that $f(\alpha^*) \geq \alpha^*$. Then the strict concavity of f implies that the derivative of f in points such that $f(\alpha) = \alpha$ can not be equal to one. However, when we look at the point $\tau'(1)$ and use that $f'(\alpha) = \tau'^{-1}(\alpha)$ (which was shown in lemma 2.1), we arive at the contradiction $f'(\tau'(1)) = \tau'^{-1}(\tau'(1)) = 1$. Hence $f(\alpha) < \alpha$ has to hold for all $\alpha \neq \tau'(1)$. $\diamond$

It follows from aboves proof that the unique solution of $f(\alpha) = \alpha$ is $\tau'(1) = -b\mathbb{E}M \log_b M$, which will be denoted by D. Using aboves lemma we can prove the following proposition:

Proposition 2.10 *All the mass of the measure μ is concentrated on the set of instants with local Hölder exponent $D = -b\mathbb{E}M \log_b M$, where this D is also the fractal dimension of this set.*

Heuristic Proof. We will consider again the coarse exponents α_n and the histogram $N_n(\alpha, \epsilon)$, and define $N_n(\alpha) = N_n(\alpha, 0)$. Then $N_n(\alpha)$ is the number of b-adic intervals with coarse exponent equal to α. Since $f(\alpha)$ is the double limit of $\ln N_n(\alpha, \epsilon)/\ln b^n$, we have the following relation:

$$N_n(\alpha, \epsilon) \sim (b^n)^{f(\alpha)} \quad \text{as } n \to \infty \text{ and } \epsilon \to 0$$

Which can also be written for $N_n(\alpha)$:

$$N_n(\alpha) \sim (b^n)^{f(\alpha)} = (b^{-n})^{-f(\alpha)} \quad \text{as } n \to \infty$$

Since $\alpha_n = \ln \mu(\Delta_n t)/\ln \Delta_n t$ with $\Delta_n t = b^{-n}$, we have $\mu(\Delta_n t) = (\Delta_n t)^{\alpha_n}$. Under the assumption that the coarse exponent α_n converges to α, we have the following asympotic relation:

$$\mu(\Delta_n t) \sim (\Delta_n t)^{\alpha} = (b^{-n})^{\alpha}$$

Now consider the cells that in the limit have local exponent α. If we multiply the mass of these cells with the number of cells, we get:

$$\lim_{n\to\infty} N_n(\alpha)\mu(\Delta_n t) = \lim_{n\to\infty} \left(b^{-n}\right)^{-f(\alpha)} (b^{-n})^{\alpha} = \lim_{n\to\infty} (b^{-n})^{\alpha - f(\alpha)}$$

If $f(\alpha) = \alpha$, then $(b^{-n})^{\alpha - f(\alpha)} \to 1$, but in all other cases we have $\alpha - f(\alpha) > 0$ and thus $\left(b^{-n}\right)^{\alpha - f(\alpha)} \to 0$. So the α for which the product $N_n(\alpha)\mu(\Delta_n t)$ does not converge to zero is $\alpha = D$. This means that in the limit all mass is concentrated on intervals with local Hölder exponent D. Hence all mass of the multifractal measure is concentrated on a set with local exponent D, which according to lemma 2.2 is equal to $-b\mathbb{E}M \log_b M$. $\diamond$

(34)

This D can be directly related to the non-degeneracy condition of section 2.1.2. In this section it was stated that the multifractal measure μ was non-degenerate if and only if $\mathbb{E}M \log_b M < 0$. Since $D = -b\mathbb{E}M \log_b M$ this condition is actually equivalent with $D > 0$. Note that when the condition $\mathbb{E}M \log_b M < 0$ is not satisfied, we would have $D \leq 0$. Remark however that a strictly negative dimension does not make sense, but also $D = 0$ is not a option, since this would mean that the set of points with exponent D is differentiable, which can not be true since μ does not have a density. The seemingly contradiction $D \leq 0$ can easily be solved by noting that when the condition is not satisfied, the measure μ has no mass, and has thus also no carrier of the mass.

We will finish this chapter by introducing and discussing the concept of concentration, and especially multifractal concentration. If we consider the increments of a certain process or measure, we say that there is absence of concentration when the number of increments required to get a prescribed proportion of the total sample variance is of order N^1. In this case even the largest of N increments has a negligible contribution to the overall sample variance. Brownian motion is an example of such a process, since each increment's relative contribution is of order N^{-1}.

There are also processes which possess a 'hard' form of concentration, which means that a significant proportion of the overall sample variance comes from a very small number of large contributions. In this case a prescribed proportion of the total variance, requires a number of increments of the order N^0. An example of processes with hard concentration are the discontinuous Lévy jump processes.

Both absent and hard concentration are not in agreement with financial data, which display a more intermediate form of concentration. However, it turns out that multifractals possess a special intermediate form of concentration, which will be called 'soft' concentration or multifractal concentration. This is a new and very flexible form, where also, as in the case of absent concentration, the single largest contribution to the sample variance is asymptotically negligible. However, any prescribed proportion of the total variance is contributed by a number of increments of the order N^D, with $0 < D < 1$. Note that as N increases N^D will also increase, but the relative number N^D/N decreases to zero.

With the proof of proposition 2.9 in mind it is easy to see that multifractal processes posses this intermediate 'soft' form of concentration. If we take $N_n = b^n$, then the proof shows that in the limit as n goes to infinity there are N_n^D intervals with a mass N_n^{-D}, for $D = -b\mathbb{E}M \log_b M$. This means that the number of contributions N_n^D is exactly large enough to insure that their total mass-contribution is nearly equal to the whole mass. In this case the relation between frequency and size is exactly right to prevent the product of the number contributions and their masses to go to zero. For all other powers not equal to D, the contributions are either large but too few to matter, or very numerous but so small that their contribution is negligible as well.

3 Multifractal Model of Asset Returns

In this chapter we will present the Multifractal Model of Asset Returns (MMAR) which was introduced by Mandelbrot, Calvet and Fisher in 1997. The MMAR will use the concept of compounding a Brownian motion[6] with a time deformation $\theta(t)$, where this time deformation process will be the cumulative distribution function of a grid based multifractal measure. The special characteristics of the multifractal measure, such as long range dependence, heavy tails and a multiplicity of local Hölder exponents will be passed on to the price process through the method of compounding. The main reason for using the concept of compounding is that it allows us to model a processes' variability without affecting the direction of its increments.

Before we give a formal definition of the MMAR, we will define in the next section what it means for a stochastic process to be multifractal. Then in section 3.2 we will show that the MMAR is indeed multifractal and study its properties.

3.1 Multifractal and unifractal processes

In section 2.2 we stated that grid based multifractal measures satisfy the moment scaling relationship $\mathbb{E}\mu(\Delta t)^q = c(q)(\Delta t)^{\tau(q)+1}$. In this section a similar scaling relationship is used to characterize multifractal processes:

Definition 3.1 *A stochastic process $X(t)$ is called a discrete multifractal if there is an integer $b \geq 2$ such that the following moment scaling rule holds for all b-adic intervals $[t, t + \Delta t]$:*

$$\mathbb{E}|X(t + \Delta t) - X(t)|^q = c_X(q)(\Delta t)^{\tau_X(q)+1} \tag{36}$$

with a nonlinear scaling function $\tau_X(q)$.

The condition that $\tau_X(q)$ should be nonlinear is the main feature of the above definition. When $\tau(q)$ is linear, the corresponding process is said to be unifractal. To give an example of such a process, consider the class of the so-called self similar processes, which are commonly used to model financial data. A stationary process $X(t)$ is called self-similar if there is an $H > 0$ such that for all $c, k, t_1, \ldots, t_k \geq 0$ the process satisfies the scaling relation

$$\{X(ct_1), \ldots, X(ct_k)\} \stackrel{d}{=} \{c^H X(t_1), \ldots, c^H X(t_k)\}$$

[6]Instead of Brownian motion we may also use fractional Brownian motion B^H, which is also a continuous-time Gaussian process. The main feature of this process is that it may have, depending on the value of H, persistent or antipersistent increments with long memory.

From this definition follows immediately that $X(t) = t^H X(1)$, and since $X(t+\Delta t) - X(t) \stackrel{d}{=} X(\Delta t)$ this gives:

$$\mathbb{E}|X(t+\Delta t) - X(t)|^q = \mathbb{E}|X(\Delta t)|^q = (\Delta t)^{Hq}\mathbb{E}|X(1)|^q$$

Hence self-similar process satisfy the moment scaling relation (36) with the linear scaling function $\tau_X(q) = Hq - 1$ and are thus unifractal. The restriction that $\tau(q)$ has to be nonlinear therefore excludes a lot of the self-similar processes commonly used in finance, such as Itô processes and Lévy processes. However, the processes provided by the MMAR have a strictly concave scaling function (as we will show in the following section) and as such form a fundamentally new class of stochastic processes for financial applications.

3.2 The Multifractal Model of Asset Returns and its properties

The MMAR provides a new model for the price of a financial asset $\{P(t) : 0 \leq t \leq T\}$. We introduce the notation:

$$X(t) = \ln P(t) - \ln P(0)$$

This expression is called the log price or log excess return and is very common in finance. At first sight it might be more intuitive to use the so-called simple excess returns $R_t = (P(t) - P(0))/P(0)$, but the log returns have some advantages over the simple returns. The main advantage for our discussion here is that the statistical properties of these log excess returns are more tractable, but they have next to this also some convenient additive properties.

Note that we can have a more intuitive understanding of the log price by writing it as:

$$\ln P(t) - \ln P(0) = \ln\Big(1 + \frac{P(t) - P(0)}{P(0)}\Big) \approx \frac{P(t) - P(0)}{P(0)}$$

Where the approximation follows from $\ln(1+x) \approx x$ for small x. So when the simple excess return $R_t = (P(t) - P(0))/P(0)$ is very small, the log price is approximately equal to R_t.

The main ingredient of our model for the log-price process $X(t)$ is the cumulative distrbitution $\theta(t)$ of a grid based multifractal measure μ. This may be either a conservative or canonical measure, but the best choice seems to be a canonical measure, since these have thick tails and are therefore more in agreement with financial data. In chapter 2 we defined these measures on the interval $[0,1]$, but if we use the 'T, b-adic' intervals $[tT, (t+\Delta t)T]$ instead of the b-adic intervals $[t, t+\Delta t]$, the multifractal measure μ may in a similar way be defined on the interval $[0,T]$ with $\mu([0,T]) = 1$. So we may also define the MMAR process on an arbitrary bounded interval $[0,T]$:

Definition 3.2 *Let B be a Brownian motion and $\theta(t)$ the cumulative distribution function of a grid based multifractal measure defined on $[0, T]$, where the processes $B(t)$ and $\theta(t)$ are assumed to be independent. Then the logprice $X(t)$ is defined as the compound process:*

$$X(t) = B(\theta(t))$$

The trading time $\theta(t)$ is the most important aspect of the MMAR and it might be expected that it passes its multifractal properties on to the log-price process $X(t)$. Next to this we also expect that the scaling function and multifractal spectrum of both processes are closely related, and that the multifractal formalism also holds for $X(t)$. These multifractal properties will be studied and proved in the following subsections.

3.2.1 Global multifractal properties

First we will prove that $X(t)$ is indeed a discrete multifractal according to definition 3.2 and that the scaling functions satisfy a very simple relation.

Proposition 3.1 *The process $X(t)$ is a discrete multifractal with scaling function $\tau_X(q) = \tau_\theta(q/2)$.*

Proof. We assume without loss of generality that $T = 1$. We will compute the expectation $\mathbb{E}|X(t+\Delta t)-X(t)|^q$ for an arbitrary b-adic interval $[t, t+\Delta t]$. Conditioning on $\mu(\Delta t) = \mu([t, t + \Delta t]) = u$ gives:

$$\begin{aligned}
\mathbb{E}[|X(t+\Delta t) - X(t)|^q \; |\mu(\Delta t) = u] &= \mathbb{E}[|B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)|^q \; |\mu(\Delta t) = u] \\
&= \mathbb{E}[|B\big(\theta(t+\Delta t) - \theta(t)\big)|^q \; |\mu(\Delta t) = u] \\
&= \mathbb{E}[|B\big(\mu(\Delta t)\big)|^q \; |\mu(\Delta t) = u] \\
&= \mu(\Delta t)^{q/2}\mathbb{E}|B(1)|^q
\end{aligned}$$

This gives:

$$\begin{aligned}
\mathbb{E}[|X(t+\Delta t) - X(t)|^q] &= \mathbb{E}[\mathbb{E}[|X(t+\Delta t) - X(t)|^q \; |\mu(\Delta t) = u]] \\
&= \mathbb{E}\mu(\Delta t)^{q/2}\mathbb{E}|B(1)|^q \qquad (37) \\
&= c_\theta(q/2)(\Delta t)^{\tau_\theta(q/2)+1}\mathbb{E}|B(1)|^q
\end{aligned}$$

Hence the process $X(t)$ satisfies (36) with $c_X(q) = c_\theta(q/2)\mathbb{E}|B(1)|^q$ and $\tau_X(q) = \tau_\theta(q/2)$. ◇ (37)

Note that in this proof we established (37) by conditioning on $\mu(\Delta t)$. In a similar way we can get

$$\mathbb{E}|X(t)|^q = \mathbb{E}\theta(t)^{q/2}\mathbb{E}|B(1)|^q$$

by conditioning on $\theta(t)$. It follows from this relation that the q-th moment of $X(t)$ exists if and only if the trading time θ has a moment of order $q/2$, or equivalently:

$$q_{crit}(X) = 2q_{crit}(\theta)$$

Since corollary 2.2 stated that $q_{crit}(\theta) < \infty$ when the multipliers M satisfy $\mathbb{P}(M > 1) > 0$, it follows that the multifractal process $X(t)$ may also have infinite moments and thus thick tails. We can however also choose the multipliers M such that $q_{crit}(\theta) = \infty$ to obtain a process without fat tails. Furthermore note that since $\mathbb{E}\theta(t) < \infty$, we have $\mathbb{E}X(t)^2 = \mathbb{E}\theta(t)\mathbb{E}|B(1)|^2 < \infty$. So although $X(t)$ might have infinite moments, its second moment will always exist. We conclude that the MMAR can obtain a variety of tail behaviours, but will always have finite variance.

3.2.2 Local multifractal properties

Proposition 3.1 showed that the global scaling properties of $X(t)$ and $\theta(t)$ are closely related. Next we will prove that also the local multifractal properties of both processes are related, via the corresponding multifractal spectra:

Proposition 3.2 *The multifractal spectrum of the process $X(t)$ is given by*

$$f_X(\alpha) = f_\theta(2\alpha)$$

Proof. We assume without loss of generality that $T = 1$. In this proof we will use the following scaling relation for the normal distribution with random variance. Let the random variable Y have the normal distribution $N(0, S^2)$ where S is an a.s. strictly positive random variable. Then we have $Y \stackrel{d}{=} S^2 Z$ with $Z \stackrel{d}{=} N(0,1)$ independent of S, as we will show now:

$$\begin{aligned}
\mathbb{P}(Y \in B) &= \int_0^\infty \mathbb{P}(Y \in B | S = \sigma) dF_S(\sigma) \\
&= \int_0^\infty \mathbb{P}(\sigma^2 Z \in B | S = \sigma) dF_S(\sigma) = \mathbb{P}(S^2 Z \in B)
\end{aligned}$$

where B is an arbitrary Borel set and we used that $Y|S = \sigma \stackrel{d}{=} \sigma^2 Z$.

Now to prove the proposition, consider the coarse exponents of the process $X(t)$ for the sequence $\Delta_n t = b^{-n}$ and the b-adic intervals $[t, t + \Delta_n t]$:

$$\begin{aligned}
\alpha_n^X(t) = \frac{\ln(|X(t + \Delta_n t) - X(t)|)}{\ln \Delta_n t} &= \frac{\ln(|B\big(\theta(t + \Delta_n t)\big) - B\big(\theta(t)\big)|)}{\ln \Delta_n t} \\
&\stackrel{d}{=} \frac{\ln(|B(\mu(\Delta_n t))|)}{\ln \Delta_n t} \\
&\stackrel{d}{=} \frac{\ln(\mu(\Delta_n t)^{1/2} |B(1)|)}{\ln \Delta_n t}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\frac{\ln(\mu(\Delta_n t))}{\ln \Delta_n t} + \frac{\ln(|B(1)|)}{\ln \Delta_n t} \\
&= \frac{1}{2}\alpha_n^\theta(t) + o_p(1)
\end{aligned}
$$

Hence it follows from Slutksy's lemma that $\alpha_n^X(t) \overset{d}{\sim} \frac{1}{2}\alpha_n^\theta(t)$ as $n \to \infty$. This means that we could reproduce the proof of theorem 2.1 with

$$
\alpha_n^X(t) = 1/n \sum_{i=1}^{n} V_i' \quad \text{with } V_i' = 1/2V_i = -1/2\log_b M(\eta_1, \ldots, \eta_i)
$$

In that proof we established that:

$$
f_\theta(\alpha) = \lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \frac{1}{n}\log_b \mathbb{P}(\alpha_n^\theta \in [\alpha - \epsilon, \alpha + \epsilon)) + 1
$$

The following is proved in the appendix: provided that $\lim_{n\to\infty}\mathbb{P}(Y_n \in B) > 0$, we have that $X_n \overset{d}{\sim} Y_n$ implies $\mathbb{P}(X_n \in B) \sim \mathbb{P}(Y_n \in B)$. If we apply this to $\alpha_n^X(t) \overset{d}{\sim} 1/2\alpha_n^\theta(t)$ we get:

$$
\begin{aligned}
f_X(\alpha) &= \lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \frac{1}{n}\log_b \mathbb{P}(\alpha_n^X \in [\alpha - \epsilon, \alpha + \epsilon)) + 1 \\
&= \lim_{\epsilon \downarrow 0} \lim_{n\to\infty} \frac{1}{n}\log_b \mathbb{P}(1/2\alpha_n^\theta \in [\alpha - \epsilon, \alpha + \epsilon)) + 1 \\
&= \lim_{\epsilon' \downarrow 0} \lim_{n\to\infty} \frac{1}{n}\log_b \mathbb{P}(\alpha_n^\theta \in [2\alpha - \epsilon', 2\alpha + \epsilon')) + 1 \\
&= f_\theta(2\alpha)
\end{aligned}
$$

◇

Let α_0^X, α_0^θ be the predominant Hölder exponents of respectively $X(t)$ and $\theta(t)$, then aboves proposition implies

$$
f_X(1/2\alpha_0^\theta) = f_\theta(\alpha_0^\theta) = 1 \qquad \text{and hence} \qquad \alpha_0^X = 1/2\alpha_0^\theta
$$

Note that since $\alpha_0^\theta > 1$ we obtain $\alpha_0^X > 1/2$. This means that the compounded process $X(t) = B(\theta(t))$ has at almost every instant a local exponent that's larger than the Hölder exponent of Brownian motion. So although the sample paths of $X(t) = B(\theta(t))$ have some apparent irregularity, its sample paths are Lebesque almost surely smoother than the paths of Brownian motion. This means that at infinitely small scales the trading time runs sometimes very fast, but most often very slow.

3.2.3 Multifractal formalism and its application to empirical modelling

Now we have established $\tau_X(q) = \tau_\theta(q/2)$ and $f_X(\alpha) = f_\theta(2\alpha)$, it is a simple corollary that the multifractal formalism also holds for the process $X(t)$:

Corollary 3.1

$$f_X(\alpha) = \inf_{q\in\mathbb{R}}[q\alpha - \tau_X(q)]$$

Proof.

$$\begin{aligned} f_X(\alpha) = f_\theta(2\alpha) &= \inf_{q\in\mathbb{R}}[q2\alpha - \tau_\theta(q)] \\ &= \inf_{q\in\mathbb{R}}[q\alpha - \tau_\theta(q/2)] \\ &= \inf_{q\in\mathbb{R}}[q\alpha - \tau_X(q)] \end{aligned}$$ ◇

The multifractal formalism does not only give us information about the local properties of $X(t)$, it also presents a method for empirical modelling. If we want to find the distribution M that best characterizes some multifractal data, we can first estimate the scaling function. This can be done by computing the empirical moments for different q and fit the estimate $\hat{\tau}_X(q)$ such that it satisfies the relationship (36), where the expectation is replaced by an empirical average. Next we can obtain an estimate for the multifractal spectrum by taking the Legendre transform of $\hat{\tau}_X(q)$.

In Mandelbrot et al (1997b) the multifractal spectrum is computed for some different distributions of M (by using a different version of Cramér's theorem) and it is shown that the multifractal spectrum is very sensitive to the distribution of the multipliers. So when we have an estimate for the multifractal spectrum, we can choose the distribution M that best agrees with this estimate. In Mandelbrot et al (1997c) it is shown that the lognormal distribution gives good results for modelling the behaviour of the Deutschemark and US dollar exchange rates. Remark that a lognormal distributed random variable is almost surely strictly positive, which together with the fact that it has moments of all orders guarantees that all results of this thesis hold. Secondly, the condition of corollary 2.2 is satisfied because the support of the lognormal distribution is the whole of $[0, \infty)$. It follows that the corresponding price process has thick tails.

3.2.4 Volatility persistence

In section 3.2.2 we established that at infinitely small scales, the trading time runs sometimes very fast, but most often very slow. However, this does not occur only in the limit, but happens at all scales, as we will explain now. The fact that in each stage of the construction of the measure μ the multipliers take different values on each b-adic cell, makes that the mass of the multifractal becomes very concentrated. The multipliers in the first stage of the construction are the most important and determine more or less the overall mass concentration. Then the multipliers in the second stage of the construction determine the way mass is concentrated on these 'first stage' b-adic cells. Since this recursive construction continues into infinity, the resulting multifractal measure has mass concentration at all scales. Because for intervals where μ has much mass the corresponding trading time $\theta(t)$ runs faster, the process $X(t) = B(\theta(t))$ has also a higher volatility on these intervals. It follows that the concentration of mass in the multifractal measure, results in a process $B(\theta(t))$ with volatility clustering at all scales.

The fact that the MMAR displays volatility clustering means that the sizes of increments that are close to each other are heavily correlated. Next to this volatility persistence on small time scales, the MMAR has also strong persistence in volatility on larger time scales. We will show that the MMAR has long range dependence in absolute returns.

Proposition 3.3 *The MMAR process $X(t)$ has long memory in the size of its increments (as in definition 2.2) for $0 < q < q_{crit}(X)/2$.*

Proof. Take a $0 < q < q_{crit}(X)/2$. We will show that

$$\bar{\delta}_X(\tau, q, \Delta t) = \bar{\delta}_\theta(\tau, q/2, \Delta t)\mathbb{E}[|B(1)|^q]^2$$

Then since $\theta(t)$ has long memory in the size of its increments the required result follows. We compute $\bar{\delta}_X(\tau, q, \Delta t)$ by conditioning on $\theta(t)$:

$$\begin{aligned}
&\mathbb{E}|X(0,\Delta t)X(\tau,\Delta t)|^q = \\
&\mathbb{E}[\mathbb{E}[|X(0,\Delta t)X(\tau,\Delta t)|^q|\theta(\Delta t),\theta(\tau+\Delta t),\theta(\tau)]] = \\
&\mathbb{E}[\mathbb{E}[|B\big(\theta(\Delta t)\big)B\big(\theta(\tau+\Delta t)-\theta(\tau)\big)|^q|\theta(\Delta t),\theta(\tau+\Delta t),\theta(\tau)]] = \\
&\mathbb{E}[\theta(\Delta t)^{q/2}(\theta(\tau+\Delta t)-\theta(\tau))^{q/2}]\mathbb{E}[|B(1)|^q]^2 = \\
&\mathbb{E}[\theta(0,\Delta t)^{q/2}\theta(\tau,\Delta t)^{q/2}]\mathbb{E}[|B(1)|^q]^2 = \\
&\bar{\delta}_\theta(\tau, q/2, \Delta t)\mathbb{E}[|B(1)|^q]^2
\end{aligned}$$

$\diamond$

We will show in the rest of this section that although $\theta(t)$ changes the speed of the Brownian motion, the variance of the increments of the process is still equal to Δt and thus the same as for Brownian motion. In the previous section we showed that most innovations $|X(t + \Delta t) - X(t)|$ are of order[7] $(\Delta t)^{\alpha_0^X}$. One might expect that the standard deviation of the process is also of this order, but it turns out that the exponents $\alpha < \alpha_0^X$ appear sufficiently frequent to give that on average the innovations are of order $(\Delta t)^{1/2}$. Or in other words, the standard deviation of the increments $X(t + \Delta t) - X(t)$ is equal to $(\Delta t)^{1/2}$, as we will show now:

Proposition 3.4 *Let $[t, t + \Delta t]$ be a b-adic interval, then*

$$Var[X(t + \Delta t) - X(t)] = \Delta t$$

[7] The fact that almost all time points have local Hölder exponent α_0^X implies that increments are in the limit $\Delta t \downarrow 0$ of order $(\Delta t)^{\alpha_0^X}$.

Proof Since $\mathbb{E}[X(t+\Delta t) - X(t)] = \mathbb{E}[B(\theta(t+\Delta t)) - B(\theta(t))] = 0$ we get:

$$Var[X(t+\Delta t) - X(t)] = \mathbb{E}(X(t+\Delta t) - X(t))^2$$

From proposition 3.1 we obtain

$$\mathbb{E}(X(t+\Delta t) - X(t))^2 = c_X(2)(\Delta t)^{\tau_X(2)+1} = c_X(2)(\Delta t)^{\tau_\theta(1)+1}$$

where we used that $\tau_X(q) = \tau_\theta(q/2)$. Since

$$\tau_\theta(1) = -\log_b \mathbb{E}M - 1 = -\log_b b^{-1} - 1 = 0$$

and

$$c_X(2) = c_\theta(1)\mathbb{E}B(1)^2 = \mathbb{E}\Omega = 1$$

we obtain the required result:

$$\mathbb{E}(X(t+\Delta t) - X(t))^2 = \Delta t \qquad \diamond$$

It can be concluded from the above result that the MMAR is a good example of a continuous process, where the fact that the standard deviation is of order $(\Delta t)^{1/2}$, does not imply that all innovations are also of the same order, as is for instance the case for Itô processes.

3.2.5 Martingale property

The fact that the increments of a Brownian motion B have unpredictable signs, makes Brownian motion a martingale. Since for the compound process $X(t)$ the trading time does not influence the direction of the Brownian motion, we obtain that $X(t)$ is also a martingale.

Proposition 3.5 *$X(t)$ is a martingale with uncorrelated increments.*

Proof. Let $\mathcal{F}_t$ and $\mathcal{F}_t^{'}$ denote the natural filtrations of $\{X(t)\}$ and $\{X(t), \theta(t)\}$. For any t, s and $u \geq t$, the independence of B and θ implies that

$$\mathbb{E}\big[B\big(\theta(t+s)\big)|\mathcal{F}_t^{'}, \theta(t+s) = u\big] = \mathbb{E}[B(u)|\mathcal{F}_t^{'}] = B(\theta(t))$$

where in the second equality we used that $B(t)$ is a martingale. We now infer that

$$\begin{aligned}
\mathbb{E}[X(t+s)|\mathcal{F}_t] &= \mathbb{E}\big[\mathbb{E}\big[B\big(\theta(t+s)\big)|\mathcal{F}_t^{'}, \theta(t+s) = u\big]|\mathcal{F}_t\big] \\
&= \mathbb{E}[B(\theta(t))|\mathcal{F}_t] \\
&= X(t)
\end{aligned}$$

This establishes that $X(t)$ is a martingale. Now we will show that this martingale property implies that the increments of $X(t)$ are uncorrelated.[8]

[8]Actually this property holds for all square integrable martingales.

If we use the notation $X(t, \Delta t) = X(t + \Delta t) - X(t)$ and take $t' \geq t + \Delta t$, consider the covariance

$$Cov[X(t, \Delta t), X(t', \Delta t)] = \mathbb{E}[X(t, \Delta t)X(t', \Delta t)]$$

where we used that $\mathbb{E}X(t, \Delta t) = 0$. The martingale property of $X(t)$ implies that $\mathbb{E}[X(t', \Delta t)|\mathcal{F}_{t'}] = 0$. Using this we obtain

$$\begin{aligned} \mathbb{E}[X(t, \Delta t)X(t', \Delta t)] &= \mathbb{E}[\mathbb{E}[X(t, \Delta t)X(t', \Delta t)|\mathcal{F}_{t'}]] \\ &= \mathbb{E}[X(t, \Delta t)\mathbb{E}[X(t', \Delta t)|\mathcal{F}_{t'}]] = 0 \end{aligned}$$

So we conclude that $Cov[X(t, \Delta t), X(t', \Delta t)] = 0$. $\diamond$ (38)

It follows from this proposition that $\ln P(t)$ is martingale. If we now apply Jensen's inequality to the convex function e^x we obtain that the price process $P(t) = P(0)e^{X(t)}$ is a submartingale and hence a semi-martingale. The concept of a semi-martingale is very important in finance. It means for instance that we can use the Itô stochastic integral to calculate the gains from trading multifractal assets with price $P(t)$. In particular, stochastic integration may be used to develop portfolio selection and option pricing theory. The semi-martingale property of $P(t)$ also implies that there are no arbitrage opportunities in a two-asset eonomy with a price $P(t)$ and a risk-free bond with constant rate of return.

3.3 MMAR's performance

Since the Multifractal Model of Asset Returns incorporates fat tails, volatility clustering and long memory in volatility, it captures many of the most important stylized facts of financial time series. It also presents a new class of continuous-time processes that bridge the gap between Itô diffusions and jump-diffusions by allowing a multiplicity of local Hölder exponents. Next to this the MMAR is a model where prices follow a semi-martingale, which makes it possible to use Itô calculus for option pricing and portfolio selection.

In contrast with the earlier model Mandelbrot (1963), which used Lévy-stable distribution with infinite variance, the MMAR has returns with finite variance which is more in agreement with financial data. Next to this the highest finite moment of these returns can take any value larger than two, which means that the model has some flexibility in matching data. Because the trading time displays more and more irregular behaviour at smaller time scales, the MMAR also has the property that the tails of the unconditional distribution of returns become thinner as the time scale increases. So, to a certain extent, the distribution of returns looks at larger time scales more and more like a normal distribution, which means that the MMAR also captures the aggregational Gaussianity of financial returns.

Since the MMAR accounts for most of the stylized facts, it seems very promising in modelling financial data. This was indeed confirmed in Mandelbrot et al (1997c). They investigated the Deutschemark/US Dollar currency exchange rates and found that these large data sets exhibit multifractal moment scaling with nonlinear scaling functions as in (36). After finding this evidence of multifractal moment scaling, the multifractal spectrum was estimated with use of the multifractal formalism and it was found that the MMAR outperforms other statistical models as GARCH and FIGARCH when it comes to simulating financial data.

Although the MMAR captures many of the main stylized facts and is successful in modelling financial data, it has some drawbacks from a more practical perspective. The model has, as a consequence of its combinatorial construction, some restrictions when it comes to some typical financial applications such as volatility forecasting. The main drawback of the MMAR is that the multipliers change at predetermined points in time, which makes the model non-stationary. To be able to use the model for forecasting, the distribution of the time before a new change in the multipliers occurs, should be the same for all time points. But the fact that all changes happen at predetermined time points, implies that the time that it takes for a new change to occur is in general different for different points in time. Another problem that is closely related to this non-stationarity is that the model is also non-causal. To able to use the model for forecasting, you would have to know for an arbitrary point the time[9] that it takes for a new change to occur. This means that you would have to know the future in order to be able to forecast the future. We can conclude that this noncausility and nonstationarity is very inconvenient, since it makes it very difficult to use MMAR for forecasting.

Next to the above mentioned problems, another drawback of the MMAR is that it involves two arbitrary fixed parameters: the scale ratio b and the finite horizon T. Since these parameters do not seem to have an important empirical meaning, but for empirical work still have to be estimated or chosen, it would be better to have a model without these parameters. It is however obvious that we need a scale ratio for the construction of the trading time $\theta(t)$, so the MMAR really needs the constant b. In the rest of this section, we will show that the MMAR also really needs the finite horizon T, in the sense that it is not possible to define the MMAR on the whole of $[0, \infty)$ without it losing its multifractal properties. We will show that in the limit as $T \to \infty$, the scaling function of the MMAR becomes linear and hence the process becomes unifractal.

Proposition 3.6 *Let $X(t)$ be a discrete multifractal process on $[0, T]$, then if T goes to infinity, the corresponding scaling function $\tau(q)$ becomes linear.*

Proof. We will show that in the limit as $T \to \infty$ the following holds (from

[9]Normally this would be the 'distribution of time', but since the construction of time points is deterministic this distribution reduces to a constant.

which the linearity of $\tau(q)$ follows immediately):

$$\forall_{q_1,q_2\in\mathbb{R}}\ \forall_{u\in(0,1)} \qquad \tau(uq_1+(1-u)q_2) = u\tau(q_1)+(1-u)\tau(q_2)$$

To prove this we will use Hölder's inequality:

Let $p_1, p_2 \in [1,\infty), f \in L^{p_1}, g \in L^{p_2}$. *If* $\frac{1}{p_1}+\frac{1}{p_2} = 1$, *then* $\|fg\|_1 \leq \|f\|_{p_1}\|g\|_{p_2}$. (39)

with $\|f\|_p = \left(\mathbb{E}|f(X)|^p\right)^{\frac{1}{p}}$. Take arbitrary $q_1, q_2 \in \mathbb{R}$ and arbitrary $u \in (0,1)$. Define $f(x) = x^{uq_1}$, $g(x) = x^{(1-u)q_2}$ and take $p_1 = \frac{1}{u}, p_2 = \frac{1}{1-u}$, then $p_1, p_2 \in [1,\infty)$ and $\frac{1}{p_1}+\frac{1}{p_2} = u+(1-u) = 1$. So we can use Hölder's inequality and apply this to the measure of the intervals $[t,(t+\Delta t)]$:

$$\begin{aligned}
\mathbb{E}[\mu(\Delta t)^{uq_1+(1-u)q_2}] &= \mathbb{E}\big[f\big(\mu(\Delta t)\big)g\big(\mu(\Delta t)\big)\big] \\
&\leq \left(\mathbb{E}\big[f\big(\mu(\Delta t)\big)^{\frac{1}{u}}\big]\right)^{\frac{1}{(\frac{1}{u})}}\left(\mathbb{E}\big[g\big(\mu(\Delta t)\big)^{\frac{1}{1-u}}\big]\right)^{\frac{1}{(\frac{1}{1-u})}} \\
&= \left(\mathbb{E}\big[\big(\mu(\Delta t)^{uq_1}\big)^{\frac{1}{u}}\big]\right)^{u}\left(\mathbb{E}\big[\big(\mu(\Delta t)^{(1-u)q_2}\big)^{\frac{1}{1-u}}\big]\right)^{1-u} \\
&= \big(\mathbb{E}[\mu(\Delta t)^{q_1}]\big)^{u}\big(\mathbb{E}[\mu(\Delta t)^{q_2}]\big)^{1-u}
\end{aligned}$$

Now using the notation $q = uq_1+(1-u)q_2$ and using $\mathbb{E}\mu(\Delta t)^q = c(q)(\Delta t)^{\tau(q)+1}$ we get:

$$c(q)(\Delta t)^{\tau(q)+1} \leq \left(c(q_1)(\Delta t)^{\tau(q_1)+1}\right)^{u}\left(c(q_2)(\Delta t)^{\tau(q_2)+1}\right)^{1-u}$$

By taking logarithms we obtain:

$$\begin{aligned}
&\ln c(q)+\tau(q)\ln\Delta t \leq \\
&\qquad u(\ln c(q_1)+\tau(q_1)\ln\Delta t)+(1-u)(\ln c(q_2)+\tau(q_2)\ln\Delta t) \quad (41)
\end{aligned}$$

Now we can take b-adic intervals $[t,t+\Delta t]$ such that the length Δt of these intervals goes to zero. Hence we can take the limit $\Delta t \to 0$ (along a discrete sequence), and by also dividing aboves inequality by $\ln\Delta t < 0$ we obtain

$$\begin{aligned}
\tau(q) &= \lim_{\Delta t\to 0}\left(\frac{\ln c(q)}{\ln\Delta t}+\tau(q)\right) \\
&\geq \lim_{\Delta t\to 0}\left(u\Big(\frac{\ln c(q_1)}{\ln\Delta t}+\tau(q_1)\Big)+(1-u)\Big(\frac{\ln c(q_2)}{\ln\Delta t}+\tau(q_2)\Big)\right) \\
&= u\tau(q_1)+(1-u)\tau(q_2)
\end{aligned}$$

Hence $\tau(q)$ satisfies:

$$\tau(uq_1+(1-u)q_2) \geq u\tau(q_1)+(1-u)\tau(q_2) \qquad (42)$$

Because we take the limit $T\to\infty$ we can also take b-adic intervals $[t,t+\Delta t]$ such that the length Δt is arbitrarily large. So we can also divide expression

(41) by $\ln \Delta t > 0$ and take the limit $\Delta t \to \infty$, which gives the reverse of (42). It follows that $\tau(uq_1 + (1-u)q_2) = u\tau(q_1) + (1-u)\tau(q_2)$ and the proof is complete. $\diamond$

We can conclude that the MMAR is successful in modelling financial data, but it has some restrictions: (i) it is non-stationary, (ii) it is non-causal, (iii) it involves a fixed scale ratio b, (iv) it can not be defined on $[0, \infty)$ and thus also involves an arbitrary fixed horizon T. In the next chapter we will introduce the Markov-Switching Multifractal which will get rid of (i), (ii) and (iv).

4 Markov-Switching Multifractal

In this chapter we will present the Markov-Switching Multifractal (MSM), which provides a causal and fully stationary version of the Multifractal Model of Asset Returns. The Markov-Switching Multifractal will also model the log price $X(t) = \ln P(t) - P(0)$ for a financial asset with price $P(t)$, and will parsimoniously capture many of the most important stylized facts of financial time series, such as fat tails, volatility persistence, jumps in volatility and moment scaling. It can be interpreted as a Markov-Switching model where different volatility components change at different frequencies. The stationary construction delivers a model for which a closed-form likelihood function is available, which makes it possible to use a standard econometric toolkit for estimating and forecasting. There are two versions of the Markov-Switching Multifractal, a continuous-time and a discrete-time version. We will first present the continuous version and later show that there also exists a convenient discretized version, which as the grid step size goes to zero, converges weakly to the continuous-time version.

The Markov-Switching Multifractal was introduced and studied by Calvet and Fisher in different works. First, in their paper Calvet and Fisher (2001), they improved on the MMAR by introducing a multifractal model, called the Poisson Multifractal, which used Poisson arrivals to construct the multifractal measure. They constructed a continous-time process and showed that there was a discrete-time version that converged weakly to the continuous-time version. Their construction was however still based on a bounded interval $[0, T]$ and different stages in the construction were not independent of each other. In later works, Calvet and Fisher (2002b); Calvet and Fisher (2004); Calvet and Fisher (2008), an improved version of the Poisson Multifractal was introduced and studied. This model uses Markov-Switching and was as such named the Markov-Switching Multifractal (although it still uses Poisson arrivals).

4.1 Continuous-time MSM

In this section we will introduce a grid free multifractal measure in continuous-time. The combinatorial construction of the MMAR will be replaced by a construction where the instants, at which new multipliers are drawn, follow a Poisson process. So the time that a certain multiplier lasts is not fixed anymore (as it was in the construction of the MMAR), but will now be exponentially distributed. The fact that the Poisson process is stationary and Markov will result in a causal process which is also stationary and Markov.

In sections 4.1.1 and 4.1.2 we will introduce and study a construction that is similar to the construction of the Poisson Multifractal in Calvet and Fisher (2001), but differs from this construction in two ways: we define the measures on the unbounded interval $[0, \infty)$ and we will achieve that the 'arrivals of new multipliers' at a particular stage are independent of all other stages (which was not the case for the Poisson Multifractal). Then in section 4.1.3 we will introduce the original and simpler construction of the Markov-Switching Multifractal and we will show that it is equivalent to the construction of section 4.1.1.. The main reason for first presenting a somewhat more complicated construction is because this construction provides more intuition about how the MMAR and the MSM relate to each other.

4.1.1 Construction of MSM

The Markov-Switching Multifractal also uses the concept of compounding a Brownian motion $B(t)$ with a trading time $\theta(t)$ to model the log price process $X(t) = P(t) - P(0)$. We assume that $B(t)$ and $\theta(t)$ are independent and that $\theta(t)$ is the cumulative distribution function of the weak limit μ of a sequence of random measures μ_n. In this section we will give the construction for these measures μ_n, which will be defined on the whole of $\mathbb{R}_+$.

For the first stage measure μ_1, consider the infinite sequence $\{T_{1,k}\}_{k=1}^{\infty}$ of independent random variables which are exponentially distributed with intensity λ. The random variables $T_{1,k}$ will be used to randomize the time instants at which the multipliers change. We define the first stage random instants $\{S_j\}_{j=0}^{\infty}$ as:

$$S_0, = 0, \quad \text{and} \quad S_j = \sum_{k=1}^{j} T_{1,k} \text{ for } j \geq 1$$

Given these $\{S_j\}$, the intervals $\{I_j = [S_j, S_{j+1}] : j \in \mathbb{N}\}$ form a random partition of $[0, \infty)$. We now define the measure μ_1 by drawing independent and identically distributed nonnegative multipliers M_j for each interval I_j and uniformly spread the mass over each interval:

$$\mu_1(I_j) = M_j \ell(I_j)$$

where ℓ is the Lebesque measure denoting the length of a given interval: $\ell([t, s]) = t - s$. To obtain a non-degenerate limit we want to impose that on any interval I with $\ell(I) = 1$ we have $\mathbb{E}\mu(I) = 1$. We do this by considering the interval $[0, T]$ and then imposing $\mathbb{E}\mu([0, T]) = T$. Because we look at the interval $[0, T]$ we need to define the random variable $N_T = \max\{m : \sum_{k=1}^{m} T_{1,k} < T\}$. Then the random partioning of $[0, T]$ is given by $\{[0, S_1], [S_1, S_2], \ldots, [S_{N_T - 1}, S_{N_T}], [S_{N_T}, T]\}$. Note that since the mass was uniformly spread we also have $\mu([S_{N_T}, T]) = M_{N_T} \ell([S_{N_T}, T])$.

This all gives:

$$\begin{aligned}
\mathbb{E}[\mu([0,T])] &= \mathbb{E}\Big[\sum_{j=0}^{N_T-1}\mu([S_j,S_{j+1}])+\mu([S_{N_T},T])\Big] \\
&= \mathbb{E}[\mathbb{E}\Big[\sum_{j=0}^{N_T-1}M_j\ell([S_{j+1}-S_j])+M_{N_T}\ell([S_{N_T},T])|\{T_{1,k}\}_{k=1}^{\infty}\Big]] \\
&= \mathbb{E}\Big[\sum_{j=0}^{N_T-1}S_{j+1}-S_j+(T-S_{N_T})\Big]\mathbb{E}M \\
&= T\mathbb{E}M
\end{aligned}$$

So to conserve mass we require $\mathbb{E}M = 1$

Now to construct μ_2, we consider a new sequence $\{T_{2,k}\}_{k=1}^{\infty}$ of i.i.d. exponentially distributed random variable, but now with intensity $b\lambda$ for a real number $b > 1$. So the intensity increases and it follows that the average number of arrivals becomes denser. We assume that each $T_{2,k}$ is independent of all other random variables. Now consider the intervals $I_{j_1} = [S_{j_1}, S_{j_1+1}]$, which were generated in the first stage, and define for each interval the total number of arrivals up to time S_{j_1+1}:

$$N^{j_1} = \max\Big\{m : \sum_{k=1}^{m} T_{2,k} < S_{j_1+1}\Big\}$$

Using this N^{j_1} we define the new random instants S_{j_1,j_2} as follows:

$$S_{j_1,j_2} = \begin{cases} S_{j_1} & \text{if } j_2 = 0 \\ \displaystyle\sum_{k=1}^{N^{j_1-1}+j_2} T_{2,k} & \text{if } 1 \le j_2 \le N^{j_1} - N^{j_1-1} \\ S_{j_1+1} & \text{if } j_2 = N^{j_1}+1 \end{cases} \tag{43}$$

So every interval I_{j_1} has now the random partition

$$\{[S_{j_1}, S_{j_1,1}], [S_{j_1,1}, S_{j_1,2}], \ldots, [S_{j_1,N_{j_1}-1}, S_{j_1,N_{j_1}}], [S_{j_1,N_{j_1}}, S_{j_1+1}]\}$$

On each subinterval $I_{j_1,j_2} = [S_{j_1,j_2}; S_{j_1,j_2+1}]$ a random multiplier M_{j_1,j_2} is drawn and again the mass is uniformly spread:

$$\mu_2(I_{j_1,j_2}) = M_{j_1}M_{j_1,j_2}\ell(I_{j_1,j_2})$$

For $j_1 \ge 1$ we choose for the intervals $[S_{j_1,0}; S_{j_1,1}]$ the same multiplier as in the previous interval. So the intervals $[S_{j_1,N^{j_1}}; S_{j_1,N^{j_1}+1}]$ and $[S_{j_1+1,0}; S_{j_1+1,1}]$

are linked in the sense that they have the same second stage multiplier (but do not share the same first stage multiplier). This somewhat complicated construction achieves that arrivals of new multipliers at the second stage are independent of the arrivals in the first stage and follow a Poisson process.

We can continue this construction in a recursive way: given the measure μ_{n-1}, consider the intervals $I_{j_1,\dots,j_{n-1}} = [S_{j_1,\dots,j_{n-1}}; S_{j_1,\dots,j_{n-1}+1}]$. Here $I_{j_1,\dots,j_{n-1}}$ is the interval in which in the first stage of the construction the j_1'th interval is chosen, then in the second stage (when this interval is again subdivided) the j_2'th subinterval of $I_{j_1}$is chosen, and so on. In the n'th stage we consider the exponential random variables $\{T_{n,k}\}_{k=1}^{\infty}$ with intensity $b^{n-1}\lambda$, where each $T_{n,k}$ is assumed to be independent of all the other random variables defined up to stage n. Again we define for each interval $I_{j_1,\dots,j_{n-1}}$ the total number of arrivals up to time $S_{j_1,\dots,j_{n-1}+1}$:

$$N^{j_1,\dots,j_{n-1}} = \max\left\{m : \sum_{k=1}^{m} T_{n,k} < S_{j_1,\dots,j_{n-1}+1}\right\}$$

With these $N^{j_1,\dots,j_{n-1}}$ the intervals $I_{j_1,\dots,j_{n-1}} = [S_{j_1,\dots,j_{n-1}}; S_{j_1,\dots,j_{n-1}+1}]$ can again be subdivided in the same way as in (43):

$$S_{j_1,\dots,j_n} = \begin{cases} S_{j_1,\dots,j_{n-1}} & \text{if } j_n = 0 \\ \displaystyle\sum_{k=1}^{N^{j_1,\dots,j_{n-1}-1}+j_n} T_{n,k} & \text{if } 1 \le j_n \le N^{j_1,\dots,j_{n-1}} - N^{j_1,\dots,j_{n-1}-1} \\ S_{j_1,\dots,j_{n-1}+1} & \text{if } j_n = N^{j_1,\dots,j_{n-1}} + 1 \end{cases}$$

which for each interval $I_{j_1,\dots,j_{n-1}}$ results in a random partitioning: $\{I_{j_1,\dots,j_n} : 0 \le j_n \le N^{j_1,\dots,j_{n-1}}\}$.

On each subinterval $I_{j_1,\dots,j_n}$ we draw new multipliers $M_{j_1,\dots,j_n}$, where again (just as for μ_2) we take the n'th stage multipliers of the intervals $[S_{j_1,\dots,j_{n-1},0}; S_{j_1,\dots,j_{n-1},1}]$ equal to the multiplier of the previous interval. So we take $M_{j_1,\dots,j_{n-1}+1,0} = M_{j_1,\dots,j_{n-1},N_{j_1,\dots,j_{n-1}}}$ for all $j_{n-1} \ge 1$. This has the consequence that the multiplier $M_{j_1,\dots,j_{n-1},N_{j_1,\dots,j_{n-1}}}$ also lasts for an exponentially distributed time, which implies that arrivals of new multipliers at each stage follow a Poisson process. Note also that arrivals of new multipliers at a particular stage are independent from all other stages. Given the multipliers we allocate the mass again uniformly over each interval, which results in:

$$\mu_n(I_{j_1,\dots,j_n}) = M_{j_1} \dots M_{j_1,\dots,j_n} \ell(I_{j_1,\dots,j_n})$$

So we have constructed a sequence of measures μ_n where the instants at which the n'th stage multipliers change are exponentially distributed, with an intensity $b^{n-1}\lambda$ that increases geometrically with n. The fact that the exponential distribution is memoryless implies that the process is stationary and that the probability of an 'arrival of a new multiplier' at any date t is independent of past history. It follows that the process $M_{n,t}$, which is defined as the value of the stage n multiplier $M_{j_1,\dots,j_n}$ at time t, is Markov. We can stack the values of all these date t multipliers into the infinite sequence $Z_t = \{M_{n,t}\}_{n=1}^{\infty}$. This Z_t is also a Markov process (on $\mathbb{R}^\infty$) and also does not depend on the future, which makes it causal. Since this Z_t contains all the information about the measure at time t, it follows that the MSM measures are Markov and causal, which are both major improvements on the MMAR.

4.1.2 Martingale property

We will now show that the sequence $\mu_n(I)$ is a martingale:

Proposition 4.1 *Let I be an arbitrary bounded interval of $[0, \infty)$ and $\mathcal{F}_n$ the natural filtration of μ_n, then*

$$\mathbb{E}[\mu_n(I)|\mathcal{F}_{n-1}] = \mu_{n-1}(I) \tag{44}$$

Proof. We will show that the martingale property holds for the intervals $I_{j_1,\dots,j_{n-1}}$:

$$\begin{aligned}
&\mathbb{E}[\mu_n(I_{j_1,\dots,j_{n-1}})|\mathcal{F}_{n-1}] = \\
&\mathbb{E}\Big[\sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} \mu_n(I_{j_1,\dots,j_n})\Big|\mathcal{F}_{n-1}\Big] = \\
&\mathbb{E}\Big[\sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} M_{j_1}\dots M_{j_1,\dots,j_n}\ell(I_{j_1,\dots,j_{n-1}})\frac{\ell(I_{j_1,\dots,j_n})}{\ell(I_{j_1,\dots,j_{n-1}})}\Big|\mathcal{F}_{n-1}\Big] = \\
&\mu_{n-1}(I_{j_1,\dots,j_{n-1}})\mathbb{E}\Big[\sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} M_{j_1,\dots,j_n}\frac{\ell(I_{j_1,\dots,j_n})}{\ell(I_{j_1,\dots,j_{n-1}})}\Big|\mathcal{F}_{n-1}\Big] = \\
&\frac{\mu_{n-1}(I_{j_1,\dots,j_{n-1}})}{\ell(I_{j_1,\dots,j_{n-1}})}\mathbb{E}\Big[\sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} \ell(I_{j_1,\dots,j_n})\Big|\mathcal{F}_{n-1}\Big]\mathbb{E}M_{j_1,\dots,j_n} = \\
&\frac{\mu_{n-1}(I_{j_1,\dots,j_{n-1}})}{\ell(I_{j_1,\dots,j_{n-1}})}\mathbb{E}[\ell(I_{j_1,\dots,j_{n-1}})|\mathcal{F}_{n-1}] = \mu_{n-1}(I_{j_1,\dots,j_{n-1}})
\end{aligned}$$

Because in the construction of μ_n the mass is distributed uniformly over each of the intervals $I_{j_1,\dots,j_n}$, it can be shown in a similar way that the martingale property also holds for arbitrary intervals I. $\diamond$ (44)

In section 4.3 we will use this proposition to prove that the weak limit of the sequence of measures μ_n exists, which implies that we can define μ as the weak limit of μ_n.

4.1.3 The cumulative distribution function and weak convergence

Just as in the construction of the MMAR, the main reason for constructing the measures μ_n and μ, is to get the cumulative distribution function θ, which we can use as a trading time. In this section, however, we will introduce an alternative construction for the cumulative distribution functions θ_n, which will be simpler than the somewhat complicated construction of the measures μ_n.

In section 4.1.1 we introduced the Markov process $M_{n,t}$, which gives the value of the stage n multiplier $M_{j_1,\dots,j_n}$ prevailing at date t. It follows from the construction of the measures μ_n, that a change in $M_{n,t}$ is triggered by a Poisson arrival with intensity $b^{n-1}\lambda$, where at each change a new multiplier is drawn. It is useful to stack these $M_{n,t}$ into a vector:

$$M_t = (M_{1,t}; M_{2,t}; \dots; M_{n,t}) \in \mathbb{R}^n_+, \qquad t \in [0, \infty)$$

This vector is called the Markov state vector and contains for a given date t all the needed information up to stage n. Using this vector, we define the stochastic volatility:

$$\sigma^2_n(M_t) = \bar{\sigma}^2 \prod_{k=1}^{n} M_{k,t}$$

where $\bar{\sigma}$ is a positive constant. We will now prove that the functions θ_n can be written as an integral over the stochastic volatility $\sigma^2_n(M_t)$:

Proposition 4.2 *For $\bar{\sigma} = 1$ the cumulative distribution function θ_n of the measure μ_n satisfies:*

$$\theta_n(t) = \int_0^t \sigma^2_n(M_s)ds \tag{46}$$

Proof. To get an explicit expression for the function $\theta_n(t)$, assume that in each k'th stage of the construction the interval $I_{j_1,\dots,j_{k-1}}$ that contains t, is cut off at this date t. This allows us to write $\theta_n(t)$ as

$$\theta_n(t) = \mu_n([0,t]) = \sum_{j_1=0}^{N_t} \sum_{j_2=0}^{N^{j_1}} \dots \sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} M_{j_1} M_{j_1,j_2} \dots M_{j_1,\dots,j_n} \ell(I_{j_1,\dots,j_n})$$

Using that on each interval $I_{j_1,\dots,j_n}$ the process $M_{k,t}$ is equal to $M_{j_1,\dots,j_k}$ for $k = 1, \dots, n$, we will show that $\int_0^t \sigma^2_n(M_s)ds$ also satisfies the above expression.

$$\begin{aligned}
\int_0^t \sigma^2_n(M_s)ds &= \int_0^t \prod_{k=1}^{n} M_{k,s} ds \\
&= \sum_{j_1=0}^{N_t} \sum_{j_2=0}^{N^{j_1}} \dots \sum_{j_n=0}^{N^{j_1,\dots,j_{n-1}}} \int_{I_{j_1,\dots,j_n}} \prod_{k=1}^{n} M_{k,s} ds
\end{aligned}$$

$$= \sum_{j_1=0}^{N_t} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_n=0}^{N^{j_1,\ldots,j_{n-1}}} \int_{I_{j_1,\ldots,j_n}} M_{j_1} \ldots M_{j_1,\ldots,j_n} ds$$

$$= \sum_{j_1=0}^{N_t} \sum_{j_2=0}^{N^{j_1}} \cdots \sum_{j_n=0}^{N^{j_1,\ldots,j_{n-1}}} M_{j_1} \ldots M_{j_1,\ldots,j_n} \ell(I_{j_1,\ldots,j_n})$$

So we conclude that $\int_0^t \sigma_n^2(M_s)ds = \mu_n([0,t]) = \theta_n(t)$. $\diamond$ (46)

Observe that this proposition implies that we could equivalently have defined the functions θ_n by (46). This definition has somewhat more flexibility since it also contains the constant $\bar{\sigma}$. This constant can however also easily be added to the construction of μ_n, by just multiplying each measure with $\bar{\sigma}^2$. In the rest of this chapter we will use that θ_n satisfies (46) for a given $\bar{\sigma}$.

We will now use this alternative definition to show that, under the condition $\mathbb{E}M^2 < b$, the sequence (θ_n) converges weakly to a tight and continuous random function θ. The condition $\mathbb{E}M^2 < b$ guarantees that volatility shocks are either sufficiently small or have durations that decrease sufficiently fast.

Proposition 4.3 *Under the condition $\mathbb{E}M^2 < b$ the sequence (θ_n) converges weakly to a tight and continuous function θ.*

Proof. To prove the proposition we need to show that the continuous functions θ_n converge weakly to a function θ on the space of continuous functions $\mathcal{C}([0,\infty))$. We will first show that we have convergence on $\mathcal{C}([0,1])$ and then extent this to $[0,\infty)$. We will assume without loss of generality that $\bar{\sigma} = 1$.

It follows from proposition 4.1 that $\theta_n(t) = \mu_n([0,t])$ is a positive martingale. Hence the martingale convergence theorem implies that $\theta_n(t)$ converges almost surely to a limit $\theta(t)$. Since the almost sure convergence of each of the components of a random vector implies the almost sure convergence of the vector, we have that $\{\theta_n(t_1), \ldots, \theta_n(t_d)\}$ converge a.s. to $\{\theta(t_1), \ldots, \theta(t_d)\}$. Now theorem 7.1 in Billingsley (1999) states that, in order to guarantee the convergence of θ_n on $\mathcal{C}([0,1])$, it is left to show that the sequence θ_n is uniformly tight on $[0,1]$. To do this we will use the following result, which is proved in Billingsley (1999) (theorem 7.3): (47)

(48)

Let $w(\theta_n, \delta) = \sup_{|t-s|<\delta} |\theta_n(t) - \theta_n(s)|$, then the sequence (θ_n) is uniformly tight if and only if the following two conditions hold:

1. $\forall \eta > 0$, *there exist a and K such that* $\forall k \geq K$ $\mathbb{P}(|\theta_k(0)| \geq a) \leq \eta$.
2. $\forall \epsilon > 0$, $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{P}(w(\theta_n, \delta) \geq \epsilon) = 0$

(49)

The first condition is immediately satisfied since $\theta_n(0) = 0$. So we will concentrate on proving the second condition. Since $\limsup_{n\to\infty} \mathbb{P}(w(\theta_n, \delta) \geq$

ϵ) is increasing in δ, it is allowed to restrict the δ's to the discrete sequence δ_k with $\delta_k = 1/k$. For a given k, we partition the interval $[0,1]$ in equally spaced intervals $[t_i, t_{i+1}]$ with $i = 0, 1, \ldots, k$ and $t_i = i/k$. Now we can use the following result proved in Billingsley (1999) (theorem 7.4): (50)
(51)

Suppose that $0 = t_0 < t_1 < \ldots < t_k = T$ *with* $\min_{1\leq i\leq k}(t_i - t_{i-1}) \geq \delta$, *then for an arbitrary function* x:

$$\mathbb{P}(w(x,\delta) \geq 3\epsilon) \leq \sum_{i=1}^{k} \mathbb{P}\Big(\sup_{t_{i-1}\leq s\leq t_i} |x(s) - x(t_{i-1})| \geq \epsilon\Big)$$

This result implies:

$$\mathbb{P}(w(\theta_n, \delta_k) \geq \epsilon) \leq \sum_{i=0}^{k-1} \mathbb{P}\Big(\theta_n(t_{i+1}) - \theta_n(t_i) \geq \frac{\epsilon}{3}\Big)$$

Because θ_n is stationary, each increment $\theta_n(t_{i+1}) - \theta_n(t_i)$ is equal in distribution to $\theta(\delta_k)$, and if we also use Markov's inequality we obtain for any $q > 0$

$$\mathbb{P}(w(\theta_n, \delta_k) \geq \epsilon) \leq k\mathbb{P}\Big(\theta(\delta_k) \geq \frac{\epsilon}{3}\Big) \leq k\Big(\frac{3}{\epsilon}\Big)^q \mathbb{E}[\theta(\delta_k)^q] \qquad (53)$$

Now we need to find a q such that the right-hand side converges to zero. For $q = 1$ this does not happen, but we will show that we do have convergence for $q = 2$. Observe that for any $t \geq 0$

$$\mathbb{E}[\theta_{n+1}(t)^2] = \int_0^t \int_0^t \mathbb{E}[M_{1,u}M_{1,v}] \ldots \mathbb{E}[M_{n,u}M_{n,v}]\mathbb{E}[M_{n+1,u}M_{n+1,v}]dudv$$

Now let T_n denote an exponentially distributed random variable with intensity $b^n\lambda$, then we have the following equality for $n \geq 0$:

$$\begin{aligned}
\mathbb{E}[M_{n+1,u}M_{n+1,v}] &= \mathbb{E}[M_{n+1,u}M_{n+1,v}|M_{n+1,u} = M_{n+1,v}]\mathbb{P}(T_n > |u-v|) + \\
&\quad\ \mathbb{E}[M_{n+1,u}M_{n+1,v}|M_{n+1,u} \neq M_{n+1,v}]\mathbb{P}(T_n \leq |u-v|) \\
&= \mathbb{E}M^2 e^{-b^n\lambda|u-v|} + (\mathbb{E}M)^2(1 - e^{-b^n\lambda|u-v|}) \\
&= (Var(M)+1)e^{-b^n\lambda|u-v|} + 1 - e^{-b^n\lambda|u-v|} \\
&= 1 + Var(M)e^{-b^n\lambda|u-v|}
\end{aligned}$$

Using this equality and $\mathbb{E}[M_{i,u}M_{i,v}] \leq \mathbb{E}M^2$ for $i = 1, \ldots, n$, which follows from Jensen's inequality, we obtain for $n \geq 1$

$$\mathbb{E}[\theta_{n+1}(t)^2] \leq \mathbb{E}[\theta_n(t)^2] + Var(M)(\mathbb{E}M^2)^n \int_0^t \int_0^t e^{-b^n\lambda|u-v|}dudv$$

For $n = 0$ we get in the same way the inequality

$$\mathbb{E}[\theta_1(t)^2] \leq t^2 + Var(M) \int_0^t \int_0^t e^{-\lambda|u-v|}dudv$$

Now we will use that for any $t \geq 0$ and $\phi \in [0, 1]$

$$\int_0^t \int_0^t e^{-b^n \lambda |u-v|} du dv \leq \frac{2t^{1+\phi}}{(b^n \lambda)^{1-\phi}}$$

which is proved in the appendix. Since we have assumed that $\mathbb{E}M^2 < b$, there exists a real number $\phi \in (0, 1)$ such that $\mathbb{E}M^2 < b^{1-\phi}$. Then the integral inequality gives

$$\begin{aligned}
\mathbb{E}[\theta_{n+1}(t)^2] &\leq \mathbb{E}[\theta_n(t)^2] + Var(M)(\mathbb{E}M^2)^n \frac{2t^{1+\phi}}{(b^n \lambda)^{1-\phi}} \\
&= \mathbb{E}[\theta_n(t)^2] + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}} \left(\frac{\mathbb{E}M^2}{b^{1-\phi}} \right)^n \\
&\leq \mathbb{E}[\theta_{n-1}(t)^2] + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}} \left(\left(\frac{\mathbb{E}M^2}{b^{1-\phi}} \right)^{n-1} + \left(\frac{\mathbb{E}M^2}{b^{1-\phi}} \right)^n \right) \\
&\leq \dots \\
&\leq t^2 + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}} \sum_{k=0}^{n} \left(\frac{\mathbb{E}M^2}{b^{1-\phi}} \right)^k
\end{aligned}$$

Now taking the limit $n \to \infty$ gives

$$\limsup_{n \to \infty} \mathbb{E}[\theta_n(t)^2] \leq t^2 + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}} \sum_{k=0}^{\infty} \left(\frac{\mathbb{E}M^2}{b^{1-\phi}} \right)^k < \infty$$

It follows that $\lim_{t \to 0} \limsup_{n \to \infty} t^{-1} \mathbb{E}[\theta_n(t)^2] = 0$. This result holds also along any sequence t_k with $t_k \to 0$, so if we take $t_k = \delta_k$ we obtain $\lim_{k \to \infty} \limsup_{n \to \infty} k \mathbb{E}[\theta_n(\delta_k)^2] = \lim_{k \to \infty} \limsup_{n \to \infty} t_k^{-1} \mathbb{E}[\theta_n(t_k)^2] = 0$ and if we use (53) we obtain the required result:

$$\lim_{k \to \infty} \limsup_{n \to \infty} \mathbb{P}(w(\theta_n, \delta_k) \geq \epsilon) = 0$$

So we have convergence on the space $\mathcal{C}([0, 1]$ of continuous functions on $[0, 1]$. Now theorem 16.8 in Billingsley (1999), which is an analogue of the tightness theorem 7.3 used above, implies that the sequence is also tight on $\mathcal{C}([0, \infty])$, which together with the martingale convergence implies that the sequence θ_n converges to a limit process θ with continuous sample paths. $\diamond$

So we have shown that when $\mathbb{E}M^2 < b$ the functions $\theta_n(t)$ converge weakly to a continuous random function $\theta(t)$. Therefore we will assume in the rest of this chapter that $\mathbb{E}M^2 < b$ indeed holds. The proposition states also that under this condition the limit function $\theta(t)$ is continuous, which means that the corresponding multifractal measure μ has no point mass, just as in the MMAR.

4.1.4 Non-degeneracy of $\theta(t)$

In this section we will show that the random function $\theta(t)$ is non-degenerate in the sense that $\mathbb{E}\theta(t) = \bar{\sigma}t^2$. The fact that $\theta(t)$ is non-degenerate is not obvious. Consider for instance the integrand of the function $\theta_n(t) = \int_0^t \sigma_n^2(M_s)ds$. It follows from the Law of Large Numbers that $\sigma_n^2(M_s)$ converges almost surely to zero as n goes to infinity:

We can write $\sigma_n^2(M_s)$ as

$$\sigma_n^2(M_s) = \bar{\sigma}^2 \prod_{k=1}^{n} M_{k,t} = \bar{\sigma}^2 e^{\sum_{k=1}^{n} \ln M_{k,s}} = \bar{\sigma}^2 \left(e^{\frac{1}{n}\sum_{k=1}^{n} \ln M_{k,s}}\right)^n$$

The Law of Large Numbers and Jensen's inequality imply that as $n \to \infty$

$$\frac{1}{n}\sum_{k=1}^{n} \ln M_{k,s} \overset{a.s.}{\to} \mathbb{E} \ln M < \ln \mathbb{E}M = 0$$

So $\exists_F$ with $\mathbb{P}(F) = 1$ such that $\forall_{\omega \in F}\ \exists_{N(\omega)} : \forall_{n \geq N(\omega)}\ \frac{1}{n}\sum_{k=1}^{n} \ln M_{k,s} < 0$. It follows that for each $\omega \in F$ $\left(e^{\frac{1}{n}\sum_{k=1}^{n} \ln M_{k,s}}\right)^n \to 0$ as $n \to \infty$, and hence $\sigma_n^2(M_s) \to 0$ almost surely.

The fact that $\sigma_n^2(M_s)$ converges almost surely to zero might suggest that the limit function $\theta(t)$ also degenerates to zero. However, we will prove that this does not happen:

Proposition 4.4 *The random function $\theta(t)$ is non-degenerate in the sense that*

$$\mathbb{E}\theta(t) = \bar{\sigma}^2 t \tag{54}$$

Proof. We will first show that $\sup_n \mathbb{E}[\theta_n(t)^2] < \infty$. In the proof of proposition 4.3 we established that for all $n \geq 1$

$$\mathbb{E}[\theta_{n+1}(t)^2] \leq \mathbb{E}[\theta_n(t)^2] + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}}\left(\frac{\mathbb{E}M^2}{b^{1-\phi}}\right)^n \tag{55}$$

And for $n = 0$

$$\mathbb{E}[\theta_1(t)^2] \leq t^2 + \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}} \tag{56}$$

If we denote $a_n := \frac{2Var(M)t^{1+\phi}}{\lambda^{1-\phi}}\left(\frac{\mathbb{E}M^2}{b^{1-\phi}}\right)^n$, then inequality (55) is equivalent to $\mathbb{E}[\theta_{n+1}(t)^2] - \mathbb{E}[\theta_n(t)^2] \leq a_n$ and (56) gives $\mathbb{E}[\theta_1(t)^2] \leq t^2 + a_0$. We obtain

$$\mathbb{E}[\theta_n(t)^2] = \mathbb{E}[\theta_1(t)^2] + \sum_{k=2}^{n}\left(\mathbb{E}[\theta_k(t)^2] - \mathbb{E}[\theta_{k-1}(t)^2]\right) \leq t^2 + \sum_{k=0}^{n-1} a_k$$

Since the ϕ in (55) can be chosen such that $\mathbb{E}M^2 < b^{1-\phi}$, we have $\sum_{k=1}^{\infty} a_k < \infty$ and we obtain:

$$\sup_n \mathbb{E}[\theta_n(t)^2] \leq \sup_n \Big(t^2 + \sum_{k=0}^{n-1} a_k\Big) = t^2 + \sum_{k=0}^{\infty} a_k < \infty$$

Now observe that

$$\mathbb{E}[\theta_n(t)^2] = \mathbb{E}\int_0^t \bar{\sigma}^2 \prod_{k=1}^{n} M_{k,s} ds = \int_0^t \bar{\sigma}^2 \prod_{k=1}^{n} \mathbb{E}M_{k,s} ds = \bar{\sigma}^2 t$$

where the use of Fubini's theorem was allowed since the integrand $\bar{\sigma}^2 \prod_{k=1}^{n} M_{k,s}$ is nonnegative. To show that $\mathbb{E}\theta(t) = \lim_{n\to\infty} \mathbb{E}\theta_n(t)$, note that since $\sup_n \mathbb{E}[\theta_n(t)^2] < \infty$ the sequence (θ_n) is bounded in L^2 and hence uniform integrable. Since martingale convergence implies $\theta_n(t) \overset{a.s.}{\to} \theta(t)$ for every $t \geq 0$, we can use Vitali's convergence theorem to obtain that $\theta_n(t)$ also converges in L^1 to $\theta(t)$. Hence

$$\mathbb{E}\theta(t) = \lim_{n\to\infty} \mathbb{E}\theta_n(t) = \bar{\sigma}^2 t \qquad \diamond$$

So we can conclude that although the integrands $\sigma_n^2(M_s)$ converge almost surely to zero, the integral $\theta(t)$ does not vanish. Apparently, on any finite interval, large realizations of $\sigma_n^2(M_s)$ appear sufficiently frequently to guarantee that the integral (46) remains positive.

Note that by taking $\bar{\sigma} = 1$ in (54) we obtain $\mathbb{E}\theta(t) = t$. This means that for the compound process $B(\theta(t))$ the trading time $\theta(t)$ can speed up or slow down the Brownian motion B, but on average has the same speed as the usual time function t. By taking another $\bar{\sigma}$ we can change the overall speed of the trading time $\theta(t)$ and change the unconditional variance of the compound process.

It also follows from proposition 4.4 that the increments of the compound process $B(\theta(t))$ are of order $(\Delta t)^{1/2}$ (just as for the MMAR).

Corollary 4.1

$$Var\big[B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\big] = \bar{\sigma}^2 \Delta t$$

Proof. Since $\mathbb{E}\big[B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\big] = 0$, we have

$$Var\big[B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\big] = \mathbb{E}\big[\{B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\}^2\big]$$

So we need to compute $\mathbb{E}\big[\{B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\}^2\big]$. Let $\mathcal{F}_t^{'}$ denote the natural filtration of $\theta(t)$, then conditioning gives

$$\begin{aligned}
\mathbb{E}\big[\{B\big(\theta(t+\Delta t)\big) - B\big(\theta(t)\big)\}^2 \big| \mathcal{F}_{t+\Delta t}^{'}\big] &= \mathbb{E}\big[B\big(\theta(t+\Delta t) - \theta(t)\big)^2 \big| \mathcal{F}_{t+\Delta t}^{'}\big] \\
&= \big(\theta(t+\Delta t) - \theta(t)\big)\mathbb{E}B(1)^2
\end{aligned}$$

Using this we get

$$\begin{aligned}\mathbb{E}\big[\{B\big(\theta(t+\Delta t)\big)-B\big(\theta(t)\big)\}^2\big] &= \mathbb{E}\big[\mathbb{E}\big[\{B\big(\theta(t+\Delta t)\big)-B\big(\theta(t)\big)\}^2\big|\mathcal{F}'_{t+\Delta t}\big]\big] \\ &= \mathbb{E}[\theta(t+\Delta t)-\theta(t)]\mathbb{E}B(1)^2 = \bar{\sigma}^2\Delta t \qquad \diamond\end{aligned}$$

Note that for $\bar{\sigma}=1$, this corollary shows that compounding Brownian motion with the trading time $\theta(t)$, does not alter the standard deviation of the increments of Brownian motion. So although the trading time $\theta(t)$ changes the local speed of the Brownian motion, the innovations of the compound process are on average still of order $(\Delta t)^{1/2}$.

4.1.5 Multifractal moment scaling

In chapter 3 we defined a stochastic process to be discrete multifractal when it satisfied a moment scaling relationship for all b-adic intervals. In this section we will give a similar definition for continuous multifractality and prove that MSM is indeed a continuous multifractal. In the literature it is common to define the multifractality of a stochastic process in the following way:

Definition 4.1 *A stochastic process $X(t)$ is called (continuous) multifractal if it has stationary increments and satisfies the following asymptotic moment scaling relationship:*

$$\mathbb{E}[|X(t)|^q] \sim c_X(q)t^{\tau_X(q)+1} \quad \text{as } t \to 0 \tag{57}$$

with a nonlinear scaling function $\tau_X(q)$.

Note that the above scaling relation defines the continuous multifractality of X by considering the q'th power of $X(t)$, but since X has stationary increments we could equivalently define multifractality using the q'th power of the increments $X(t+\Delta t)-X(t)$ (as we did for discrete multifractals). In the literature continuous and discrete multifractals are both just called multifractals, but in this thesis discrete and continuous multifractality is introduced to distinguish between the asymptotic moment scaling on infinitely small time scales and discrete moment scaling on b-adic intervals.

To prove that the process $B(\theta(t))$ satisfies this asymptotic scaling relationship, we will first need to prove some lemma's, where each lemma is actually also interesting on its own. The following lemma shows that the critical moment $q_{crit}=\sup\{q:\mathbb{E}[\theta(t)^q]<\infty\}$ does not depend on t.

Lemma 4.1 *If $\mathbb{E}[\theta(t)^q]<\infty$ for some instant $t>0$, then also $\mathbb{E}[\theta(t')^q]<\infty$ for every $t'\in[0,\infty)$.*

Proof. For $t' \leq t$ we know that $\theta(t') \leq \theta(t)$, which implies $\mathbb{E}[\theta(t')^q] \leq \mathbb{E}[\theta(t)^q]$ and hence $\mathbb{E}[\theta(t')^q]$ is also finite. For $t' > t$ we will show that $\mathbb{E}[\theta(nt)^q] < \infty$ for every $n \geq 2$. Note that we can write $\theta(nt)^q$ as:

$$\theta(nt)^q = \left(\sum_{i=1}^{n} [\theta(it) - \theta((i-1)t)] \right)^q$$

If we use that $\left(\sum_{i=1}^n x_i\right)^q \leq \max(n^{q-1}, 1) \sum_{i=1}^n x_i^q$ for $q \geq 0$, $n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbb{R}_+^n$, which is proved in the appendix, then we obtain

$$\theta(nt)^q \leq \max(n^{q-1}, 1) \sum_{i=1}^{n} [\theta(it) - \theta((i-1)t)]^q$$

By taking expectations and using $\theta(it) - \theta((i-1)t) \stackrel{d}{=} \theta(t)$ we obtain

$$\mathbb{E}[\theta(nt)^q] \leq \max(n^q, n)\mathbb{E}[\theta(t)^q] < \infty$$

Since for every t' there is a n such that $t' \leq nt$, we obtain in the same way as for $t' \leq t$ that $\mathbb{E}[\theta(t')^q] < \infty$. $\diamond$

We will also need the following lemma, which shows how the total mass of the unit interval $\Omega(\lambda)$ depends on the intensity λ and how $\theta(t)$ relates to $\Omega(\lambda)$.

Lemma 4.2 *The trading $\theta(t)$ satisfies the invariance property*

$$\theta(t) \stackrel{d}{=} t\Omega(t\lambda)$$

for all $t \in [0, \infty)$ and $\lambda > 0$.

Proof. We will first prove that this property holds for $\theta_n(t) = \int_0^t \sigma_n^2(M_s)ds$. If we use the change of variables $u = s/t$ we obtain

$$\theta_n(t) = \int_0^t \sigma_n^2(M_s)ds = t\int_0^1 \sigma_n^2(M_{ut})du$$

Now note that the state vector $M'_u = M_{ut}$ is driven by the arrivals of a Poisson proces with intensities $t\lambda, \ldots, tb^{n-1}\lambda$. It follows that $\theta_n(t) \stackrel{d}{=} t\Omega_n(t\lambda)$, where $\Omega_n(\lambda)$ is the random variable that denotes the total mass of the function θ_n on the unit interval. Since the sequence θ_n converges weakly to θ we obtain $\theta(t) \stackrel{d}{=} t\Omega(t\lambda)$. $\diamond$

The following lemma is consistent with the intuition that when the intensity λ is very low, there is a high probability that the first stage multiplier is constant on the unit interval, which suggests $\mathbb{E}\Omega(\lambda)^q \approx \mathbb{E}M^q\mathbb{E}\Omega(b\lambda)^q$.

Lemma 4.3 $\mathbb{E}\Omega(\lambda)^q \sim \mathbb{E}M^q\mathbb{E}\Omega(b\lambda)^q$ *as $\lambda \to 0$.*

Proof We define the function $g(\lambda) = \mathbb{E}\Omega(\lambda)^q$ and define $t_0 < t_1 < \ldots < t_{N_1} < t_{N_1+1}$ by $t_i = S_i$ for $i \leq N_1$ and $t_{N_1+1} = 1$, where $S_0, S_1, \ldots, S_{N_1}$ are the arrival times of new multipliers. Conditioning on N_1 gives:

$$\begin{aligned} g(\lambda) &= \mathbb{E}[\mathbb{E}[\Omega(\lambda)^q | N_1]] \\ &= \sum_{n=0}^{\infty} \mathbb{E}[\Omega(\lambda)^q | N_1 = n]\mathbb{P}(N_1 = n) \\ &= \mathbb{E}M^q \mathbb{E}\Omega(b\lambda)^q e^{-\lambda} + \sum_{n=1}^{\infty} \mathbb{E}[\Omega(\lambda)^q | N_1 = n] \frac{\lambda^n e^{-\lambda}}{n!} \end{aligned} \tag{58}$$

We concentrate on $\mathbb{E}[\Omega(\lambda)^q | N_1 = n]$ and observe that as in the proof of lemma 4.1, the relation $\Omega(\lambda) = \sum_{j=1}^{N_1+1} [\theta(t_j) - \theta(t_{j-1})]$ implies

$$\Omega(\lambda)^q \leq \max\left((N_1+1)^{q-1}, 1\right) \sum_{j=1}^{N_1+1} [\theta(t_j) - \theta(t_{j-1})]^q \tag{59}$$

Now note that on any interval $[t_j, t_{j-1}]$ the first stage multiplier is constant, so $\mathbb{E}[(\theta(t_j) - \theta(t_{j-1}))^q] \leq \mathbb{E}M^q \mathbb{E}\Omega(b\lambda)^q$. By conditioning on N_1 in (59) we get:

$$\begin{aligned} \mathbb{E}[\Omega(\lambda)^q | N_1 = n] &\leq \max\left((n+1)^{q-1}, 1\right) \sum_{j=1}^{n+1} \mathbb{E}[(\theta(t_j) - \theta(t_{j-1}))^q] \\ &\leq \max\left((n+1)^q, n+1\right) \mathbb{E}M^q \mathbb{E}\Omega(b\lambda)^q \\ &= (n+1)^{\max(q,1)} \mathbb{E}M^q g(b\lambda) \end{aligned}$$

Now if we apply this inequality to (58), rearrange the terms and use that $g(\lambda) \geq \mathbb{E}M^q \mathbb{E}\Omega(b\lambda)^q e^{-\lambda}$, then we obtain

$$e^{-\lambda} \leq \frac{g(\lambda)}{\mathbb{E}M^q g(b\lambda)} \leq e^{-\lambda}\Big(1 + \sum_{n=1}^{\infty} \frac{\lambda^n}{n!}(n+1)^{\max(q,1)}\Big) < \infty$$

Now taking the limit $\lambda \to 0$ establishes $\frac{g(\lambda)}{\mathbb{E}M^q g(b\lambda)} \to 1$, and hence $\mathbb{E}\Omega(\lambda)^q \sim \mathbb{E}M^q \mathbb{E}\Omega(b\lambda)^q$ as $\lambda \to 0$. $\diamond$

Now we are ready to prove that the trading time $\theta(t)$ satisfies the asymptotic moment scaling (57):

Proposition 4.5 *The q'th moment of the random function $\theta(t)$ satisfies for $0 < q < q_{crit}$*

$$\mathbb{E}\theta(t)^q \sim c_q t^{\tau_\theta(q)+1} \quad \textit{as } t \to 0$$

where $\tau_\theta(q) = -\log_b \mathbb{E}M^q + q - 1$ and c_q is a positive constant.[10]

[10] Note that since the MMAR uses $\mathbb{E}M = b^{-1}$ and MSM uses $\mathbb{E}M = 1$, MSM has a slightly different scaling function $\tau(q)$.

Proof. We will first prove that

$$\mathbb{E}\theta(t)^q \sim \frac{\mathbb{E}M^q}{b^q}\mathbb{E}\theta(bt)^q \quad \text{as } t \to 0$$

Lemma 4.2 implies that $\mathbb{E}\theta(t)^q = t^q\mathbb{E}\Omega(t\lambda)^q$ and $\mathbb{E}\theta(bt)^q = (bt)^q\mathbb{E}\Omega(t\lambda)^q$, which gives

$$\frac{\mathbb{E}\theta(t)^q}{\mathbb{E}\theta(bt)^q} = \frac{1}{b^q}\frac{\mathbb{E}\Omega(t\lambda)^q}{\mathbb{E}\Omega(bt\lambda)^q}$$

Since by lemma 4.3 $\mathbb{E}\Omega(t\lambda)^q/\mathbb{E}\Omega(bt\lambda)^q$ converges to $\mathbb{E}M^q$ we obtain

$$\frac{\mathbb{E}\theta(t)^q}{\mathbb{E}\theta(bt)^q} \to \frac{\mathbb{E}M^q}{b^q} \quad \text{as } t \to 0$$

Now note that since $b^{\tau_\theta(q)+1} = (\mathbb{E}M^q/b^q)^{-1}$, the function $c_q t^{\tau_\theta(q)+1}$ satisfies

$$\frac{c_q t^{\tau_\theta(q)+1}}{c_q(bt)^{\tau_\theta(q)+1}} = \frac{\mathbb{E}M^q}{b^q}$$

In Calvet and Fisher (2008) it is proved that the above implies

$$\frac{\mathbb{E}\theta(t)^q}{c_q t^{\tau_\theta(q)+1}} \to 1 \quad \text{as}\, t \to 0$$

◇

Now we have established the asymptotic moment scaling of $\theta(t)$, we can by mimicking the proof of proposition 3.1, easily establish that the compound process $X(t) = B(\theta(t))$ also satisfies asymptotic moment scaling:

$$\mathbb{E}X(t)^q \sim c_q t^{\tau_X(q)+1} \quad \text{as } t \to 0$$

with $\tau_X(q) = \tau_\theta(q/2)$. This confirms that the log-price $X(t)$ is indeed a continuous multifractal process. Next to this asymptotic scaling, Calvet and Fisher (2008) used Monte Carlo simulations to show that this moment scaling also holds remarkably well for finite time increments, which is consistent with the moment scaling exhibited by many financial time series.

The scale invariant construction of MSM and the fact that it satisfies asymptotic moment scaling suggest that MSM has similar local properties as the MMAR. It is stated in Calvet and Fisher (2008) (without proof) that "the local variations of MSM are almost everywhere smoother than the $(dt)^{1/2}$ variations of an Itô process". This would suggest that the sample paths of MSM also contain a multiplicity of local Hölder exponents and satisfy the multifractal formalism. However, to my knowledge there are no results available in the literature about the local behaviour of MSM. So it is still open for more research to confirm that the local variations of MSM have the same characteristic multifractal properties as described in chapter 2.

4.1.6 Discussion of continuous-time MSM

By compounding a Brownian motion with a trading time driven by Poisson arrivals, MSM improves on the MMAR in several ways. First of all, MSM is a causal process with stationary increments on an unbounded time domain in which volatility components change at random instants generated by a Poisson process, and not at predetermined points of time as in the MMAR. In addition, MSM is a process with a latent Markov state vector for which the likelihood function exists in closed form (for discrete multipliers). This is a major improvement over the MMAR, which due to its combinatorial constructing was not well suited to standard techniques of econometrics.

Next to these improvements, MSM has also much of the MMAR's appealing statistical properties. It displays (asymptotic) moment scaling and has thick tails, as was shown by simulations in Calvet and Fisher (2008). These simulations also showed that MSM captures long memory in volatility. It turned out that the autocovariances of squared returns of the simulated data decline hyperbolically. In addition to this volatility persistence on large scales, MSM also displays volatility persistence at smaller scales. When a multiplier changes, the volatility changes discontinuously and has strong persistence. Since these multipliers change at infinitely many stages, this construction results in a process with bursts of volatility at all scales.

The fact that MSM assumes that volatility is the product of multi-frequency volatility components with different durations and discontinuous changes, has also an attractive economic interpretation. For instance news innovations in an economy have often a direct impact on the volatility in that economy, and as such, jumps in volatility are needed to account for those shocks. Furthermore, economic intuition suggests that different types of volatility shocks in an economy have different degrees of persistence. This idea is in agreement with the way in which MSM generates volatility clustering. The lowest frequencies at which multipliers change might correspond to business cycles and technological shocks, while higher frequencies could correspond to for instance news shocks which often last for a shorter period of time. So MSM captures the impact of different economic shocks with different degrees of persistence. By contrast, most standard models treat all volatility innovations as the same.

Next to the statistical properties, it can also be shown (in the same way as for the MMAR) that the MSM process $X(t)$ is a martingale. It follows that the price process $P(t)$ is a semi-martingale. This has important consequences for practical work, since the semi-martingale property makes it possible to apply Itô calculus to obtain results for option pricing and portfolio selection.

We can conclude that MSM is major advance over the MMAR since first of all, it gets rid of the nonstationarity, noncausality and the restriction to bounded intervals, secondly it has the same appealing statistical properties and also satisfies the martingale property, and finally it is the first multifractal model that can be estimated using standard econometric techniques. However, we might still expect to be able to improve on the MSM since it still involves an arbitrary fixed scale ratio b. This scale parameter b does not seem to have a clear empirical meaning, but is still an extra parameter that has to be estimated. To solve this, some alternative multifractal models have been developed, that provide more mathematical elegance by not requiring a scale ratio b. We will now give a brief overview of those models.

Barral and Mandelbrot (2002) introduced a multifractal model which incorporated a continuum of time scales and in which trading time can be represented as an integral over a cone in the space of scales and time. These results were further generalized and extended in a series of works: Bacry, Delour and Muzy (2000); Bacry, Delour and Muzy (2001); Bacry, Muzy (2002); Bacry, Kozehemyak and Muzy (2008); Bacry, Duvernet, Muzy (2010). In these works the Multifractal Random Walk (MRW) was introduced, which is a multifractal model that does not require a scale parameter b. But the MRW does have the same characteristic properties as the multifractal models discussed in this thesis, since it also displays long memory in squared returns, thick tails, volatility clustering and multifractal moment scaling. However, MRW is less tractable than MSM in the sense that it is not based on Markov-switching and as such does not have the advantages of those models, including for instance a closed-form likelihood function. Whether MRW provides meaningful empirical differences compared to MSM remains an open question.

4.2 Discrete-time MSM

This section will introduce a discretized version of the continuous-time MSM. The main reason for doing this is that a discrete-time model is more convenient to use for forecasting. The discretized model will share much of the same properties as its continuous counterpart. By similarly defining volatility as the product of random Markov components, the discrete model also captures volatility clustering, jumps in volatility, thick tails and long memory in squared returns. The components will be drawn from the same distribution, but the transition probabilities of each component are different. These transition probabilities will be related to each other by some restrictions, which are inspired by the continuous-time model.

We will present the discrete-time version of MSM, by first giving a definition of this discretized model in section 4.2.1. Then in section 4.2.2 we will show that this definition is consistent with our earlier model, in the sense that the discrete-time MSM weakly converges to the continuous-time MSM as the grid step size goes to zero. In the last section we will discuss its properties and performance. Our discussion of the discrete-time model is mainly based on Calvet and Fisher (2001), Calvet and Fisher (2008).

4.2.1 Construction of discrete MSM

Consider the log-returns $r_t = \ln P_t - \ln P_{t-1}$, where P_t is a financial time series on the regular grid $t = 0, 1, 2, \ldots$. These log returns are modelled by

$$r_t = \sigma(M_t)\epsilon_t$$

where the sequence of random variables $\{\epsilon_t\}$ are i.i.d. innovations with zero mean and variance equal to one. The most common used distribution for ϵ_t is the normal distribution, but we may also use a more heavy tailed distribution like the student-t distribution (see for instance Lux, Morales-Arias (2009)).

The innovations ϵ_t play the same role as the Brownian motion does in the continuous model. $\sigma(M_t)$ determines the volatility of the process r_t and the innovations ϵ_t determine (more or less) the sign of the returns. It follows that the main aspect of the product $\sigma(M_t)\epsilon_t$ is the volatility $\sigma(M_t)$, where $M_t = (M_{1,t}, \ldots, M_{\bar{k},t})$ is again the Markov state vector of the process with $\bar{k}$ componenents. Just as in the construction of the continuous-time model, the stochastic volatility $\sigma(M_t)$ is again the product of the nonnegative components $M_{k,t}$:

$$\sigma^2(M_t) = \bar{\sigma}^2 \prod_{k=1}^{\bar{k}} M_{k,t}$$

Note that the discrete-time model makes use of a finite number of volatility components, denoted by $\bar{k}$. The multipliers $M_{k,t}$ have the same distribution M with $\mathbb{E}M = 1$, but evolve at different frequencies. For each $k \in \{1, \ldots, \bar{k}\}$, the time that a certain multiplier M_k remains unchanged is geometrically distributed with parameter γ_k. This means that, assuming that the Markov state vector has been constructed up to date $t-1$, the k'th multiplier remains unchanged at time t with probability γ_k. And the probability that the multiplier $M_{k,t-1}$ at time t will be replaced by a new draw from the distribution M, is equal to $1 - \gamma_k$. We can summarize this as follows:

$$\begin{array}{cl} M_{k,t} \text{ drawn from distribution } M & \text{with probability } \gamma_k \\ M_{k,t} = M_{k,t-1} & \text{with probability } 1 - \gamma_k \end{array}$$

The switching events and new draws from M are assumed to be independent across k and t, such that the variables $M_{k,t}$ and $M_{k',t}$ are independent if k differs from k'.

In the continuous-time MSM, the fact that interarrival times of new multipliers are exponentially distributed, implies that the stochastic volatility process is Markov. Since the discrete model uses an geometric distribution, which is also memoryless, we obtain that the discrete stochastic volatility process is also Markov. Note that this also means that it is indeed justified to call M_t a Markov state vector.

Similarly to the continuous-time construction, where the Poisson arrivals use geometrically increasing intensities, we will introduce discretized Poisson arrivals in the following section. These discrete-time arrivals will also have a geometric distribution and inspired to define the transition probabilities $\gamma_1, \dots, \gamma_{\bar{k}}$ as follows:

$$\gamma_k = 1 - (1 - \lambda)^{(b^{k-1})} \tag{60}$$

where $\lambda \in (0, 1)$ and $b > 1$. Note that $\gamma_1 = \lambda$ and that the transition probabilities are increasing, but remain lower than one. Since the probabilities are increasing, the interarrival times between arrivals of new multipliers tend to shorten. So at higher stages the multipliers last for a shorter amount of time. Another feature of these specifications is that they are consistent with the geometric distribution of the discretized Poisson arrivals, as we will show in the following section.

The resulting process is called the discrete-time Markov-Switching Multifractal, which is denoted by MSM($\bar{k}$). The term 'Multifractal' refers to the weak convergence to the continuous-time MSM. The term 'Markov-Switching' refers to a concept introduced in Hamilton (1989). He introduced the so-called Markov-switching model, which was characterized by a switching mechanism that is controlled by an unobservable state variable that follows a Markov chain. In this model the conditional mean and variance depended on this unobservable state, and changes in their conditional distributions were triggered by the underlying state variable. The fact that MSM also contains an unobservable Markov chain that determines the conditional variance of the MSM($\bar{k}$), justifies its name.

MSM imposes only minimal restrictions on the distribution of the volatility components. Since we only require $M \geq 0$ and $\mathbb{E}M = 1$ we have a lot of freedom for choosing the distribution of the multipliers. Just as for the MMAR, a popular choice is the lognormal distribution. But also the simple binomial model, in which M takes only two values with equal probability, is very popular and gives already good results. As stated earlier, the MSM($\bar{k}$) has a relatively easy closed form likelihood function, but this is only when the distribution of M is discrete. For more general distributions the likelihood function does not exist in closed form or is computationally infeasible. The main advantage of the binomial MSM over the lognormal MSM is therefore that maximum likelihood estimation can be used to obtain good parameter estimates.

4.2.2 Convergence of discrete-time MSM

In this section we explain where the specifications (60) come from and prove that a rescaled version of MSM($\bar{k}$) weakly converges to the continuous-time MSM. We will first discretize the continuous-time MSM on a uniform grid. Then we will show that this discretized model is indeed a rescaled version of MSM($\bar{k}$) and converges to continuous-time MSM.

In order to discretize the continuous-time MSM, we need to discretize the Poisson arrivals. For each bounded interval $[0, T]$ we will discretize the arrival times in such a way, that the grid step size goes to zero as the number of volatility components $\bar{k}$ goes to infinity. First we will introduce some new notation.

Consider the sequence $\{\theta_{\bar{k}}\}_{\bar{k}=1}^{\infty}$ used in the construction of the continuous trading time $\theta(t)$. For this construction we used the sequences $\{T_{n,k}\}_{k=1}^{\infty}$ of exponential distributed interarrival times with intensity $b^{n-1}\lambda$. For stage n we define the arrival times $\{t_{n,j}\}_{j=0}^{\infty}$ by $t_{n,0} = 0$ and $t_{n,j} = \sum_{k=1}^{j} T_{n,k}$. Furthermore we denote by $M_n(j)$ the value of the multiplier over the interval $[t_{n,j}; t_{n,j+1}]$. So $M_n(j)$ can be interpreted as the $j + 1$'th multiplier in the n'th stage.

For each $\bar{k}$ we will discretize the Poisson arrival times $\{t_{n,j}\}_{j=0}^{\infty}$ on the uniform grid $0, T/c^{\bar{k}}, 2T/c^{\bar{k}}, \ldots, T$ with $c \geq 2$ a discrete integer. Let $\lfloor x \rfloor$ be the floor function (so $\lfloor x \rfloor$ is for all $x \in \mathbb{R}$ the unique integer such that $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$), then we define the discretized arrival times $\{s_{n,j}\}_{j=0}^{\infty}$ by letting $s_{n,0} = 0$, and

$$s_{n,j} = \sum_{k=1}^{j} \frac{\left\lfloor c^{\bar{k}} T_{n,k}/T \right\rfloor + 1}{c^{\bar{k}}/T}$$

The random variable $c^{\bar{k}}(s_{n,j+1} - s_{n,j})/T$ has a geometric distribution with parameter $\gamma_n = 1 - \exp(-b^{n-1}\lambda T/c^{\bar{k}})$, as we will show now:

$$\begin{aligned}
\mathbb{P}\big(c^{\bar{k}}(s_{n,j+1} - s_{n,j})/T = m\big) &= \mathbb{P}\Big(\left\lfloor c^{\bar{k}} T_{n,j+1}/T \right\rfloor + 1 = m\Big) \\
&= \mathbb{P}\bigg(\frac{(m-1)T}{c^{\bar{k}}} \leq T_{n,j+1} < \frac{mT}{c^{\bar{k}}}\bigg) \\
&= 1 - e^{-\frac{mb^{n-1}\lambda T}{c^{\bar{k}}}} - \Big(1 - e^{-\frac{(m-1)b^{n-1}\lambda T}{c^{\bar{k}}}}\Big) \\
&= \Big(e^{-\frac{b^{n-1}\lambda T}{c^{\bar{k}}}}\Big)^{m-1} - \Big(e^{-\frac{b^{n-1}\lambda T}{c^{\bar{k}}}}\Big)^{m} \\
&= (1-\gamma_n)^{m-1} - (1-\gamma_n)^{m} \\
&= (1-\gamma_n)^{m-1}\gamma_n
\end{aligned}$$

Note that this result can be rewritten as

$$\mathbb{P}\Big(s_{n,j+1} - s_{n,j} = \frac{mT}{c^{\bar{k}}}\Big) = (1-\gamma_n)^{m-1}\gamma_n$$

which means that we can interpret the above result as $s_{n,j+1} - s_{n,j}$ having a geometric distribution on the grid $0, T/c^{\bar{k}}, 2T/c^{\bar{k}}, \ldots$.

The idea is now to use the discretized Poisson arrivals $\{s_{n,j}\}_{j=0}^{\infty}$ to construct a discretized trading time $\theta^{*}_{\bar{k}}(t)$ on the interval $[0, T]$ as follows: for each stage $n \leq \bar{k}$ the multiplier on the interval $[s_{n,j}; s_{n,j+1}]$ is set equal to the value $M_n(j)$ of the multiplier on the interval $[t_{n,j}; t_{n,j+1}]$ in the continuous construction. So the continuous-time multiplier process $M^*_{n,t}$ of the discretized trading time, changes at the time instants $\{s_{n,j}\}_{j=0}^{\infty}$ and if $t \in [s_{n,j}; s_{n,j+1}]$, then $M^*_{n,t} = M_n(j)$. Now define

$$\theta^*_{\bar{k}}(t) = \int_0^t \bar{\sigma} \prod_{n=1}^{\bar{k}} M^*_{n,s} ds$$

The result is a continuous and piecewise linear process.

Note that although $\theta^*_{\bar{k}}(t)$ is defined on $[0, T]$, changes of multipliers only occur on the grid $0, T/c^{\bar{k}}, 2T/c^{\bar{k}}, \ldots, T$. As the number of volatility components $\bar{k}$ goes to infinity, the grid will become more and more dense on the interval $[0, T]$ and the transition probabilities $\gamma_n = 1 - \exp(-b^{n-1}\lambda T/c^{\bar{k}})$ will go to zero, such that the distribution of the interarrival times will converge to the exponential distribution. Before we show that $\theta^*_{\bar{k}}(t)$ indeed weakly converges to the process $\theta(t)$, we will show that $\theta^*_{\bar{k}}(t)$ is intimately related to the discrete-time MSM($\bar{k}$).

We define the function $\theta^{**}_{\bar{k}}(t)$ on the regular grid $s = 0, 1, 2 \ldots, c^{\bar{k}}$, by $\theta^{**}_{\bar{k}}(0) = 0$ and

$$\theta^{**}_{\bar{k}}(t) = \sum_{n=1}^{t} \sigma^{**}_{\bar{k}}(M_n)^2$$

where the volatility $\sigma^{**}_{\bar{k}}(M^t)^2$ is specified by $M, \bar{\sigma}$ and the transition probabilities $\gamma_n = 1 - \exp(-b^{n-1}\lambda T/c^{\bar{k}})$ (instead of (60)). We extend the domain of $\theta^{**}_{\bar{k}}(t)$ to the continuous interval $[0, c^{\bar{k}}]$ by linear interpolation.

Note that this function $\theta^{**}_{\bar{k}}(t)$ can be interpreted as the linear interpolated version of the original MSM($\bar{k}$) process on the interval $[0, c^{\bar{k}}]$, except for the fact that in the construction of MSM($\bar{k}$) the transition probabilities (60) are used, but for $\theta^{**}_{\bar{k}}(t)$ we used the alternative transition probabilities $\gamma_n = 1 - \exp(-b^{n-1}\lambda\Delta t)$ (with $\Delta t = T/c^{\bar{k}}$). However, as $\bar{k}$ goes to infinity, and Δt goes to zero we have

$$1 - \exp(-b^{n-1}\lambda\Delta t) \approx 1 - (1 - \lambda\Delta t)^{(b^{n-1})}$$

(since $e^{-x} \approx 1 - x$ for small x). This means that $\theta^{**}_{\bar{k}}(t)$ has asymptotically the same transition probabilities as the discrete-time MSM($\bar{k}$) with a grid of step size Δt and transition probabilities $1 - (1 - \lambda\Delta t)^{(b^{n-1})}$.

Now we know that the function $\theta^{**}_{\bar{k}}(t)$ is asymptotically equivalent to the linear interpolated version of the MSM($\bar{k}$) process, observe that since $c^{\bar{k}}(s_{n,j+1} - s_{n,j})/T$ has a geometric distribution with the same parameters γ_n as $\theta^{**}_{\bar{k}}(t)$, the discretized trading time $\theta^*_{\bar{k}}(t)$ satisfies

$$\theta^*_{\bar{k}}(t) = \theta^{**}_{\bar{k}}(tc^{\bar{k}}/T)$$

We conclude that $\theta^*_{\bar{k}}(t)$ is a rescaled version of the trading time $\theta^{**}_{\bar{k}}(t)$, with asymptotically the same transition probabilities as MSM($\bar{k}$) as the grid step size goes to zero. So $\theta^*_{\bar{k}}(t)$ and MSM($\bar{k}$) are related to each other through the function $\theta^{**}_{\bar{k}}(t)$.

We will now show that $\theta^*_{\bar{k}}(t)$ converges to the continuous-time MSM as the number of volatility components goes to infinity, or equivalently the grid step size goes to zero. To guarantee this convergence we have to choose the integer c such that either $b < c$ or $b\mathbb{E}M^2 < c^2$ is satisfied. The first condition requires that the number of grid points grows faster than the volatility frequencies, which makes sure that in the limit the possibility $s_{n,j} = s_{n,j+1}$ won't occur. The second condition is more difficult to understand, but note that for $b = c$ the condition reduces to the basic assumption $\mathbb{E}M^2 < b$.

Proposition 4.6 *Let c satisfy either $b < c$ or $b\mathbb{E}M^2 < c^2$, then the sequence $\{\theta^*_{\bar{k}}\}_{\bar{k}=1}^{\infty}$ of discretized trading times converges weakly to the continuous-time process θ.*

Proof We will not give a full proof, but give the main ideas. The proof is based on the following result, proved in Billingsley (1999).

(60)

If for arbitrary $p \in \mathbb{N}$

$$\{\theta^*_{\bar{k}}(t_1), \ldots, \theta^*_{\bar{k}}(t_p)\} \xrightarrow{d} \{\theta(t_1), \ldots, \theta(t_p)\} \tag{62}$$

holds for all $t_1, \ldots, t_p$, and if for each $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{\bar{k} \to \infty} \mathbb{P}(w(\theta^*_{\bar{k}}, \delta) \geq \epsilon) = 0 \tag{63}$$

with $w(\theta^*_{\bar{k}}, \delta) = \sup_{|t-s|<\delta} |\theta^*_{\bar{k}}(t) - \theta^*_{\bar{k}}(s)|$, then $\theta^*_{\bar{k}}$ weakly converges to θ.

So provided that the sequence is tight, the convergence of the finite dimensional distributions implies that $\theta^*_{\bar{k}}(t)$ weakly converges to $\theta(t)$. We will first establish (62).

We assume without loss of generality that $T = 1$ and $\bar{\sigma} = 1$. Let $H^*_n = \max\{j : s_{n,j} < 1\}$ be the highest integer j such that $s_{n,j} < 1$ in the n'th stage of the discrete-time construction. For any $j_1, \ldots, j_{\bar{k}}$ let $\Delta^*(j_1, \ldots, j_{\bar{k}})$ be the length of the largest subinterval of $[0, 1]$ over which the multipliers are given by $M_{1,j_1}, \ldots, M_{\bar{k},j_{\bar{k}}}$. Then we have the expression

$$\theta^*_{\bar{k}}(1) = \sum_{j_1=0}^{H^*_1} \cdots \sum_{j_{\bar{k}}=0}^{H^*_{\bar{k}}} M_{1,j_1} \ldots M_{\bar{k},j_{\bar{k}}} \Delta^*(j_1, \ldots, j_{\bar{k}})$$

For the continuous-time construction we define H_n and $\Delta^*(j_1, \ldots, j_{\bar{k}})$ in a similar way, and get

$$\theta_{\bar{k}}(1) = \sum_{j_1=0}^{H_1} \cdots \sum_{j_{\bar{k}}=0}^{H_{\bar{k}}} M_{1,j_1} \ldots M_{\bar{k},j_{\bar{k}}} \Delta(j_1, \ldots, j_{\bar{k}})$$

Let $V_n = \max(H_n^*, H_n)$. The intervals attached to a given set of multipliers in the paired constructions differ in length by

$$\delta_{j_1,\ldots,j_{\bar{k}}} = \Delta^*(j_1, \ldots, j_{\bar{k}}) - \Delta(j_1, \ldots, j_{\bar{k}})$$

Then the difference between the discretized and the original trading time over the interval $[0, 1]$ is given by

$$\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1) = \sum_{j_1=0}^{V_1} \cdots \sum_{j_{\bar{k}}=0}^{V_{\bar{k}}} M_{1,j_1} \ldots M_{\bar{k},j_{\bar{k}}} \delta_{j_1,\ldots,j_{\bar{k}}}$$

We will now show that either $\mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)|$ or $\mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)|^2$ converge to zero as $\bar{k}$ goes to infinity. Since both imply $|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)| \xrightarrow{d} 0$, we obtain $\theta_{\bar{k}}^*(1) \xrightarrow{d} \theta(1)$ as follows:

$$\begin{aligned} |\theta_{\bar{k}}^*(1) - \theta(1)| &\stackrel{d}{=} |\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1) + \theta_{\bar{k}}(1) - \theta(1)| \\ &\leq |\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)| + |\theta_{\bar{k}}(1) - \theta(1)| \xrightarrow{d} 0 \end{aligned}$$

This argument can be applied equally well to all time intervals $[0, t]$, so we conclude that $\theta_{\bar{k}}^*(t) \xrightarrow{d} \theta(t)$ for all t.

It is left to show that the first and second moment of $|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)|$ converge. We first discuss the first moment. Observe that

$$\begin{aligned} \mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)| &\leq \mathbb{E}\Big(\sum_{j_1=0}^{V_1} \cdots \sum_{j_{\bar{k}}=0}^{V_{\bar{k}}} M_{1,j_1} \ldots M_{\bar{k},j_{\bar{k}}} |\delta_{j_1,\ldots,j_{\bar{k}}}| \Big) \\ &= \mathbb{E}\Big(\sum_{j_1=0}^{V_1} \cdots \sum_{j_{\bar{k}}=0}^{V_{\bar{k}}} |\delta_{j_1,\ldots,j_{\bar{k}}}| \Big) \end{aligned}$$

As in the proof of the convergence of the discretized Poisson multifractal in Calvet and Fisher (2001), it can be shown that the average number of nonzero mismatches $\delta_{j_1,\ldots,j_{\bar{k}}}$ is of order $b^{\bar{k}}$, with size of order $1/c^{\bar{k}}$. From this can be inferred that $\mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)|$ is bounded from above by a multiple of $(b/c)^{\bar{k}}$. Hence under the condition $b < c$ we conclude that $\mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)| \to 0$ as $\bar{k} \to \infty$.

In a similar way it can be verified that the second moment $\mathbb{E}|\theta_{\bar{k}}^*(1) - \theta_{\bar{k}}(1)|^2$ is bounded above by a multiple of $(\mathbb{E}M^2 b/c^2)^{\bar{k}}$ and therefore converges to zero under the condition $b\mathbb{E}M^2 < c^2$.

Now it is left to show that the sequence satisfies (63). Take $\delta = c^{-l}$. As in the proof of proposition 4.3, theorem 7.4 in Billingsley (1999) implies

$$\mathbb{P}(w(\theta_{\bar{k}}^*, \delta) \geq \epsilon) \leq \delta^{-1} \mathbb{P}\Big(\theta_{\bar{k}}^*(\delta) \geq \frac{\epsilon}{3}\Big)$$

Since we already established $\theta^*_{\bar{k}}(\delta) \xrightarrow{d} \theta(\delta)$, we obtain

$$\limsup_{\bar{k}\to\infty} \mathbb{P}(w(\theta^*_{\bar{k}}, \delta) \le \epsilon) \le \delta^{-1}\mathbb{P}(\theta(\delta) \ge \epsilon/3)$$

Markov's inequality implies

$$\limsup_{\bar{k}} \mathbb{P}(w(\theta^*_{\bar{k}}, \delta) \ge \epsilon) \le \left(\frac{3}{\epsilon}\right)^2 \delta^{-1}\mathbb{E}\theta(\delta)^2$$

By applying proposition 4.5, we obtain

$$\delta^{-1}\mathbb{E}\theta(\delta)^2 \sim c_2\delta^{\tau_\theta(2)} \text{ as } \delta \to 0$$

Now note that $\tau_\theta(2) > 0$, since $\tau_\theta(2) = -\log_b \mathbb{E}M^2 + 1 > 0$, where $-\log_b \mathbb{E}M^2 + 1 > 0$ follows from our basic assumption $\mathbb{E}M^2 < b$. The fact that $\tau_\theta(2) > 0$ implies that $\delta^{\tau_\theta(2)}$ goes to zero and hence condition (63) is satisfied:

$$\lim_{\delta\to 0} \limsup_{\bar{k}\to\infty} \mathbb{P}(w(\theta^*_{\bar{k}}, \delta) \ge \epsilon) \le \lim_{\delta\to 0} \left(\frac{3}{\epsilon}\right)^2 \delta^{-1}\mathbb{E}\theta(\delta)^2 = 0$$

We conclude that $\theta^*_{\bar{k}}$ weakly converges to θ as $\bar{k} \to \infty$. $\diamond$

The fact that a rescaled version of the discrete-time MSM($\bar{k}$) weakly converges to its continuous-time counterpart, has important implications for volatility forecasting. The convergence implies that the discrete-time MSM can be used for volatility forecasting, and this forecast will be consistent for the continuous-time MSM under an appropriate sequence of increasingly refined discretizations. This result is important, since it means that researchers can move back and forth between the discrete-time model, which is more convenient for applied work, and the continuous-time model, where theory is sometimes easier.

4.2.3 Discussion of discrete-time MSM

Since the discrete-time MSM is intimately related to its continuous-time counterpart, one would expect that it also captures the same important stylized facts as continuous-time MSM. This is indeed confirmed by simulations and by the fact that MSM($\bar{k}$) performs well in modelling financial data, as was for the first time shown in Calvet and Fisher (2002b). They showed that the binomial MSM($\bar{k}$) is consistent with financial data, and compared it also to the commonly used GARCH and FIGARCH models. They found that MSM($\bar{k}$) performs equally well as those models at short horizons, but outperforms them at medium and long horizons (up to 50 days). At these horizons MSM provides significant gains in forecasting accuracy over the GARCH and FIGARCH models.

The above results were later confirmed in Calvet and Fisher (2004) and expanded (mainly) by Lux in a series of papers. First in Lux (2006) the

Generalized Method of Moments (GMM) was presented, which is an alternative for maximum likelihood estimation. Although GMM is slightly less efficient than MLE, it can be applied in all those cases where the maximum likelihood function does not exist in closed form or is computationally infeasible. This new method allowed Di Matteo, Liu and Lux (2008) to model financial data with the Lognormal MSM. Although this new specification for the multipliers did not provide much gains compared to the binomial model, the alternative estimation method also allowed for modelling data with an higher number of volatility components. In Calvet and Fisher (2002), Calvet and Fisher (2004) the number of volatility components was restricted to $\bar{k} = 10$, because for higher $\bar{k}$ the maximum likelihood method became computionally infeasible. Lux, Di Matteo and Liu (2008) however showed that using the GMM method the forecasting accuracy of MSM($\bar{k}$) is in general higher when the number of volatility components exceeds $\bar{k} = 10$.

Next to the important accomplishment of MSM of generating better volatility forecasts than the common used models from the GARCH-family, it is noteworthy that it does this in a remarkable effective way. MSM($\bar{k}$) uses only one single mechanism to capture volatility clustering, long memory features and thick tails. Previously it was thought that it was best to model those aspects separately. MSM($\bar{k}$) shows however that a single regime-switching approach can play all three of these roles in a very effective way. First, the multifrequency construction of MSM guarantees that there is volatility clustering at all levels. Second, the changes in volatility components (and especially the low-frequency multipliers) are very persistent. And finally, the high-frequency volatility components generate high-frequency switches and substantial outliers (and thus thick tails).

The fact that MSM($\bar{k}$) has only a few parameters that need to be estimated, contributes also to the effectiveness and parsimony of the model. Apart from the choice of $\bar{k}$, which can be viewed as a model selection problem, MSM($\bar{k}$) only needs four parameters. First of all, the set of transition probabilities is completely specified by (60) and (λ, b). Next to this we need the constant $\bar{\sigma}$, which is the unconditional standard deviation of the returns. It is left to choose the distribution M. As discussed before, there are two popular choices in the literature, the binomial and lognormal. Since we assume $\mathbb{E}M = 1$, both distributions require only one parameter to characterize them. So we conclude that MSM($\bar{k}$) is constructed in a very parsimonious way, as it only needs four parameters to specify the whole model.

MSM($\bar{k}$) can be extended in two ways. First, it is showed in Calvet and Fisher (2008) that the MSM($\bar{k}$) construction can be generalized to several assets. For practical work the main advance of a multivariate MSM will be that it can be used in portfolio selection (for instance computing values-at-risk for portfolios). In this multivariate model the different assets are correlated in two ways. First, the arrivals of new multipliers are positively correlated. When a market is very volatile one would expect that al the assets in this market display more volatility. The positive correlation of arrivals of new changes in volatility is thus in agreement with economic intuition. Second, next to the correlation in the values of the squared returns, the 'direction' of the returns can also be correlated by taking a multivariate normal distribution for the Gaussian innovations. This would account for the possibility that prices of different assets might tend to move in the same or opposite direction.

Eisler and Kertész (2006) presented another extension of the Markov-Switching Model. They introduced a method to also capture the asymmetry in returns. As discussed in chapter one, it is observed that in financial markets downward movements are in general larger than upward movements (gain/loss asymmetry), and falling asset prices generate more volatility than upward movements in asset price (leverage effect). Since in the product $r_t = \sigma(M_t)\epsilon_t$ the 'amplitude factor' $\sigma(M_t)$ is nonnegative, asymmetry can only originate from the 'sign process' ϵ_t. Eisler and Kertész (2006) introduced some alternative stochastic processes for ϵ_t to account for the gain/loss asymmetry and leverage autocorrelation, but could not confirm whether these models provide substantial gains in forecasting accuracy and is therefore still open for more research.

We can conclude this thesis by stating that MSM is a very promising model for financial modelling. First of all it captures many of the main stylized facts (and can be extended to also account for the leverage effect) in a very effective and parsimonious way. Next to this, MSM satisfies the martingale property and has a multivariate version, which provides possibilities to develop portfolio selection and option pricing theory. But the most promising aspect of MSM is that it performs better than existing models in modelling financial data, and can relatively easily be used in practice since there are several econometric tools available for estimating and forecasting.

5 Appendix

Lemma 5.1 *We have the following two logarithmic inequalities:*

$$\begin{array}{lll} 1) & |\ln y| \leq y^{-1} - 1 & \text{for } 0 < y < 1 \\ 2) & \ln y \leq y - 1 & \text{for } y > 0 \end{array}$$

Proof. To prove the first inequality, we will need the following inequality:

$$\ln(y) \geq 1 - y^{-1} \qquad \text{for } y > 0$$

Define $f(y) = y - e^{1-y^{-1}}$, then $\lim_{y \downarrow 0} f(y) = 0$ and $f'(y) = 1 + y^{-2} e^{1-y^{-1}} > 0$. This implies $f(y) \geq 0$, and hence $y \geq e^{1-y^{-1}}$ for all $y > 0$. By taking logarithms on both sides we obtain the above logarithmic inequality. Now note that for $y \in (0,1)$ the logarithm $\ln y$ is negative, which gives:

$$|\ln y| = -\ln y \leq y^{-1} - 1 \qquad \text{for } y \in (0,1)$$

To show the second inequality, we define the function $g(y) = e^{y-1} - y$ for $y > 0$. We compute $g'(y) = e^{y-1} - 1$ and $g''(y) = e^{y-1} > 0$. Since $g''(y) > 0$ we obtain that the function $g(y)$ is convex, and hence its global minimum is given by the solution of $g'(y) = 0$. Now note that $g'(1) = 0$ and $g(1) = 0$, so the global minimum of $g(y)$ is given by 0, hence $e^{y-1} \geq y$ for $y > 0$. If we now take logarithms on both sides we obtain $y - 1 \geq \ln y$. $\diamond$

Lemma 5.2 $(x+y)^h \geq x^h + y^h - 2(1-h)(xy)^{\frac{h}{2}} \quad \forall x > 0, y > 0, 0 < h < 1.$

Proof. First note that the above equation can be written as

$$\left(\sqrt{\frac{x}{y}} + \sqrt{\frac{y}{x}}\right)^h \geq \sqrt{\frac{x}{y}}^h + \sqrt{\frac{y}{x}}^h - 2(1-h)$$

Now define $t = \ln \sqrt{\frac{x}{y}}$, then $e^t = \sqrt{\frac{x}{y}}$. Now with the continuous function

$$f(t) = e^{th} + e^{-th} - (e^t + e^{-t})^h$$

aboves inequality is equivalent to $f(t) \leq 2(1-h)$. So we need to show that $\sup f(t) \leq 2(1-h)$ for all $0 < h < 1$. Note that because f is symmetric we only have to consider nonnegative t. First we will show that $f(t)$ goes to 0 as t goes to infinity:

$$\begin{aligned}
\lim_{t\to\infty} f(t) &= \lim_{t\to\infty} e^{th} + e^{-th} - (e^t + e^{-t})^h \\
&= \lim_{t\to\infty} e^{th} - e^{th}\Big(\frac{e^t + e^{-t}}{e^t}\Big)^h \\
&= \lim_{t\to\infty} e^{th}(1 - (1 + e^{-2t})^h) \\
&= \lim_{t\to\infty} e^{th}\Big(1 - \sum_{k=0}^{\infty} C_{h,k} e^{-2tk}\Big) \\
&= \lim_{t\to\infty} e^{th} \sum_{k=1}^{\infty} C_{h,k} e^{-2tk} \\
&= \sum_{k=1}^{\infty} C_{h,k} \lim_{t\to\infty} e^{-(2k-h)t} = 0
\end{aligned}$$

where we used the binomial series $(1+x)^h = \sum_{k=0}^{\infty} C_{h,k} x^k$ for $|x| < 1$ with $C_{h,k} = \frac{h(h-1)\ldots(h-k+1)}{k!}$ for $k > 1$ and $C_{h,0} = 0$.

Now consider the point $t = 0$. Note that $f(0) = 2 - 2^h$, and that $f(0) \leq 2(1-h)$ is equivalent to $2^h - 2h \geq 0$. Define the function $u : [0,1] \to \mathbb{R}$ by $u(h) = 2^h - 2h$. Then it is easy to compute that $u'(h) = 2^h \ln 2 - 2 < 0$, which together with $u(0) = 1$ and $u(1) = 0$ implies that $u(h) \geq 0$.

As a result we only have to look at the local maxima, i.e. at points $t \in (0, \infty)$ such that $f'(t) = 0$:

$$f'(t) = he^{th} - he^{-th} - h(e^t + e^{-t})^{h-1}(e^t - e^{-t}) = 0$$

which is equivalent to

$$(e^t + e^{-t})(e^{th} - e^{-th}) - (e^t + e^{-t})^h(e^t - e^{-t}) = 0$$

From this equation follows that for $t > 0$ such that $f'(t) = 0$ we have

$$(e^t + e^{-t})^h = \frac{e^{(1+h)t} - e^{(1-h)t} + e^{-(1-h)t} - e^{-(1+h)t}}{e^t - e^{-t}}$$

Now if we implement this expression in $f(t)$ we get:

$$\begin{aligned}
& \frac{(e^{th} + e^{-th})(e^t - e^{-t})}{e^t - e^{-t}} - \frac{e^{(1+h)t} - e^{(1-h)t} + e^{-(1-h)t} - e^{-(1+h)t}}{e^t - e^{-t}} \\
= & \frac{e^{(1+h)t} - e^{-(1-h)t} + e^{(1-h)t} - e^{-(1+h)t} - e^{(1+h)t} + e^{(1-h)t} - e^{-(1-h)t} + e^{-(1+h)t}}{e^t - e^{-t}} \\
= & 2\frac{e^{(1-h)t} - e^{-(1-h)t}}{e^t - e^{-t}}
\end{aligned}$$

So for t where $f(t)$ has a local maximum we have the expression $f(t) = 2\frac{e^{(1-h)t}-e^{-(1-h)t}}{e^t-e^{-t}}$. Now define $\epsilon = 1-h$ and consider the function

$$g(\epsilon) = e^{\epsilon t} - e^{-\epsilon t} - \epsilon(e^t - e^{-t})$$

If $g(\epsilon) \leq 0$ then it follows that $f(t) \leq 2(1-h)$ for t where $f(t)$ has a local maximum, and hence $\sup f(t) \leq 2(1-h)$. So to complete this proof we need to show that $g(\epsilon) \leq 0$ for all $\epsilon \in (0,1)$:

If we write the exponential function as its power series we get

$$g(\epsilon) = e^{\epsilon t} - e^{-\epsilon t} - \epsilon(e^t - e^{-t}) = \sum_{n=0}^{\infty}\Big(\frac{(\epsilon t)^n}{n!} - \frac{(-\epsilon t)^n}{n!} - \frac{\epsilon t^n}{n!} + \frac{\epsilon(-t)^n}{n!}\Big)$$

For even n, the terms in the sum are equal to 0, and for uneven n, the terms are equal to $(2(\epsilon t)^n - 2\epsilon t^n)/n! = (2t^n\epsilon(\epsilon^{n-1}-1))/n! \leq 0$. It follows that $g(\epsilon) \leq 0$ and the proof is complete. ◇

Lemma 5.3 *The intervals* $I_{k_n}^{(n)} = [k_n b^{-n}, (k_n+1)b^{-n}]$ *with* $k_n(x) = \lfloor xb^n \rfloor$ *satisfy* $I_{k_n}^{(n+1)} \subset I_{k_n}^{(n)}$ *for all* n.

Proof. We will first show that $k_{n+1}b^{-(n+1)} \geq k_n b^{-n}$:

$$\left\lfloor xb^{n+1} \right\rfloor b^{-(n+1)} = \lfloor xb^n + xb^n(b-1) \rfloor b^{-(n+1)} \geq$$
$$\lfloor \lfloor xb^n \rfloor + \lfloor xb^n \rfloor (b-1) \rfloor b^{-(n+1)} = \lfloor xb^n \rfloor b^{-n}$$

The proof of $(k_{n+1}+1)b^{-(n+1)} \leq (k_n+1)b^{-n}$ is more complicated. We will first prove the following property of floor functions:

$$k = \sum_{i=0}^{m-1} \left\lfloor \frac{k+i}{m} \right\rfloor \quad \text{for all positive integers } k, m \tag{64}$$

To prove this we fix $k, m \in \mathbb{N}$. There are $z_1, z_2 \in \mathbb{N}$ such that $k = z_1 m + z_2$ with $z_1 \geq 0$, $0 \leq z_2 \leq m-1$. We get:

$$\sum_{i=0}^{m-1} \left\lfloor \frac{k+i}{m} \right\rfloor = \sum_{i=0}^{m-1} \left\lfloor \frac{z_1 m + z_2 + i}{m} \right\rfloor$$
$$= \sum_{i=0}^{m-z_2-1} \left\lfloor z_1 + \frac{z_2+i}{m} \right\rfloor + \sum_{i=m-z_2}^{m-1} \left\lfloor z_1 + \frac{z_2+i}{m} \right\rfloor = \sum_{i=0}^{m-z_2-1} z_1 + \sum_{i=m-z_2}^{m-1} (z_1+1)$$
$$= (m-z_2)z_1 + (m-(m-z_2))(z_1+1) = z_1 m + z_2 = k$$

If we use that for $y \in \mathbb{R}$ we have $\left\lfloor \frac{\lfloor y \rfloor + i}{m} \right\rfloor = \left\lfloor \frac{y+i}{m} \right\rfloor$ which was proved in Graham, Knuth, and Patashnik (1994), then equality (64) gives:

$$\lfloor my \rfloor = \sum_{i=0}^{m-1} \left\lfloor \frac{\lfloor my \rfloor + i}{m} \right\rfloor = \sum_{i=0}^{m-1} \left\lfloor \frac{my+i}{m} \right\rfloor = \sum_{i=0}^{m-1} \left\lfloor y + \frac{i}{m} \right\rfloor$$

With $y = xb^n$ and $m = b$ this gives:

$$\lfloor xb^{n+1} \rfloor + 1 = \lfloor bxb^n \rfloor + 1 = 1 + \sum_{i=0}^{b-1} \left\lfloor xb^n + \frac{i}{b} \right\rfloor \leq \sum_{i=0}^{b-1} \lfloor xb^n + 1 \rfloor = b\big(\lfloor xb^n \rfloor + 1 \big)$$

If we multiply by $b^{-(n+1)}$ we get the required result:

$$\big(\lfloor xb^{n+1} \rfloor + 1 \big) b^{-(n+1)} \leq \big(\lfloor xb^n \rfloor + 1 \big) b^{-n}$$

and the proof is complete. $\diamond$

Lemma 5.4 *Let B be a Borel set such that $\lim_{n\to\infty} \mathbb{P}(Y_n \in B) > 0$, then $X_n \overset{d}{\sim} Y_n$ implies $\mathbb{P}(X_n \in B) \sim \mathbb{P}(Y_n \in B)$*

Proof. $X_n \overset{d}{\sim} Y_n$ means $X_n/Y_n \xrightarrow{d} 1$ and thus $\mathbb{P}(X_n/Y_n \in B) \to \mathbb{P}(1 \in B)$. This gives:

$$\mathbb{P}(X_n = Y_n) = \mathbb{P}(X_n/Y_n \in \{1\}) \to \mathbb{P}(1 \in \{1\}) = 1$$

So we established that $\mathbb{P}(X_n = Y_n) \to 1$ and it follows that

$$\frac{\mathbb{P}(X_n \in B)}{\mathbb{P}(Y_n \in B)} =$$

$$\frac{\mathbb{P}(X_n \in B | X_n = Y_n)\mathbb{P}(X_n = Y_n) + \mathbb{P}(X_n \in B | X_n \neq Y_n)\mathbb{P}(X_n \neq Y_n)}{\mathbb{P}(Y_n \in B)} =$$

$$\mathbb{P}(X_n = Y_n) + \frac{\mathbb{P}(X_n \in B | X_n \neq Y_n)}{\mathbb{P}(Y_n \in B)} (1 - \mathbb{P}(X_n = Y_n)) \to 1$$

And thus $\mathbb{P}(X_n \in B) \sim \mathbb{P}(Y_n \in B)$. $\diamond$

Lemma 5.5 *Let $q > 1$, then for all $k \geq 1$ and $x_1, \ldots, x_k \geq 0$ the equality*

$$\left(\sum_{i=1}^{k} x_i \right)^q = \sum_{i=1}^{k} x_i^q \tag{65}$$

implies that at most one x_i is nonzero.

Proof. First note that for $k = 1$ there is nothing to prove, so take $k \geq 2$. We will show that the equality can not hold if there is more than one nonzero x_i. Define the function $f : \mathbb{R}_+^k \to \mathbb{R}_+$ by

$$f(x_1, \ldots, x_k) = \left(\sum_{i=1}^{k} x_i \right)^q - \sum_{i=1}^{k} x_i^q$$

Note that when there is at most one nonzero x_i, then $f(x_1, \ldots, x_k) = 0$. Now define the set $G = \{x = (x_1, \ldots, x_k) \in \mathbb{R}_+^k | \; \exists_{i,j} : x_i > 0, x_j > 0, i \neq j\}$. So this is the set of $x \in \mathbb{R}_+^k$ with at least two nonzero x_i. Note that for all $x \in G$ we have that for all $i = 1, \ldots, k$, the partial derivative with respect to x_i is strictly positive:

$$\frac{\partial f(x)}{\partial x_i} = q\left(\sum_{i=1}^{k} x_i\right)^{q-1} - q x_i^{q-1} > 0 \qquad \text{for } x \in G$$

Together with $f(x) = 0$ for $x \notin G$, this gives $f(x) > 0$ for all $x \in G$. $\diamond$

Lemma 5.6 *The following equality holds for any $t \geq 0$, $h > 0$ and $\phi \in [0,1]$:*

$$\int_0^t \int_0^t e^{-h|u-v|} du dv \leq 2 \frac{t^{1+\phi}}{h^{1-\phi}}$$

Proof. We can compute the integral:

$$\begin{aligned}
\int_0^t \int_0^t e^{-h|u-v|} du dv &= \int_0^t \left(\int_0^v e^{-h(v-u)} du + \int_v^t e^{-h(u-v)} du \right) dv \\
&= \int_0^t \left(\frac{e^{-hv}}{h}(e^{hv} - 1) - \frac{e^{hv}}{h}(e^{-ht} - e^{-hv}) \right) dv \\
&= \int_0^t \frac{2 - e^{-hv} - e^{-h(t-v)}}{h} dv \\
&= \frac{2t}{h} + \frac{(e^{-ht} - 1) - (1 - e^{-ht})}{h^2} \\
&= \frac{2}{h^2}(e^{-ht} - 1 + ht)
\end{aligned}$$

Now we will show that $e^{-x} - 1 + x \leq x^{1+\phi}$ for all $x \geq 0$. To show this inequality for $x \in [0,1]$ consider the function $f : [0,1] \to \mathbb{R}$ given by $f(x) = x^2 - x + 1 - e^{-x}$. Its first and second derivative are respectively given by $f'(x) = 2x - 1 + e^{-x}$ and $f''(x) = 2 - e^{-x}$. Furthermore note that $f(0) = f'(0) = 0$ and that $f''(x) > 0$ for all $x \in [0,1]$. The fact that $f'(x)$ has a positive derivative on $[0,1]$ and $f'(0) = 0$, guarantees that $f'(x)$ is nonnegative on $[0,1]$. Since $f(x)$ also has $f(0) = 0$ it follows that $f(x)$ is nonnegative on $[0,1]$.

The inequality $f(x) \geq 0$ implies $e^{-x} - 1 + x \leq x^2 \leq x^{1+\phi}$ for $x \in [0,1]$. For $x \geq 1$ we obtain (by noting that $e^{-x} - 1 \leq 0$) that $e^{-x} - 1 + x \leq x \leq x^{1+\phi}$. If we now combine the results for $x \in [0,1]$ and $x \in [1, \infty)$ we obtain that $e^{-x} - 1 + x \leq x^{1+\phi}$ for all $x \geq 0$. It follows that

$$\int_0^t \int_0^t e^{-h|u-v|} du dv \leq \frac{2}{h^2}(ht)^{1+\phi} = 2\frac{t^{1+\phi}}{h^{1-\phi}}$$

holds for all $t \geq 0$, $h > 0$ and $\phi \in [0,1]$. $\diamond$

Lemma 5.7 *For all $q \geq 0$, $n \geq 1$ and $(x_1, \ldots, x_n) \in \mathbb{R}_+^n$ the following inequality holds*

$$\Big(\sum_{i=1}^{n} x_i\Big)^q \leq \max(n^{q-1}, 1) \sum_{i=1}^{n} x_i^q$$

Proof. For $q = 1$ the result is obvious. For $q < 1$, observe that $\max(n^{q-1}, 1) = 1$ and the inequality follows immediately by the subadditivity of the function x^q. For $q > 1$ we have $\max(n^{q-1}, 1) = n^{q-1}$ and the assertion is somewhat more difficult to prove. We define a discrete random variable X by $\mathbb{P}(X = x_i) = 1/n$ for $i = 1, \ldots, n$. So X has expectation $\mathbb{E}X = \sum_{i=1}^{n} \frac{x_i}{n}$ and since the function x^q is convex we can apply Jensen's inequality:

$$\Big(\sum_{i=1}^{n} \frac{x_i}{n}\Big)^q = (\mathbb{E}X)^q \leq \mathbb{E}X^q = \sum_{i=1}^{n} \frac{x_i^q}{n}$$

It follows that

$$\Big(\sum_{i=1}^{n} x_i\Big)^q \leq n^{q-1} \sum_{i=1}^{n} x_i^q \qquad \diamond$$

6 Credits

This section gives an overview of which proofs are original work, and which proofs are due to others.

Chapter 2

Proposition 2.1: *original work.*
Proposition 2.2: *an extended, more complete version of the proof in Kahane and Peyriere (1976).*
Corollary 2.1: *original work.*
Proposition 2.3: *original work.*
Proposition 2.4: *original work.*
Proposition 2.5: *original work.*
Proposition 2.6: *an extended, more complete version of the proof in Kahane and Peyriere (1976).*
Corollary 2.2: *original work.*
Proposition 2.7: *an adapted and extended version of the proof in Calvet and Fisher (2008).*
Proposition 2.8: *original work.*
Proposition 2.9: *based on a derivation in Mandelbrot et al (1997b).*
Lemma 2.1: *original work.*
Theorem 2.1: *an extended, more complete version of the proof in Mandelbrot et al (1997b).*
Lemma 2.2: *original work.*
Proposition 2.10: *original work, but inspired by Mandelbrot (2001).*

Chapter 3

Proposition 3.1: *Mandelbrot et al (1997a).*
Proposition 3.2: *original work.*
Corollary 3.1: *original work.*
Proposition 3.3: *Mandelbrot et al (1997a).*
Proposition 3.4: *Calvet and Fisher (2008)*
Proposition 3.5: *first part: Mandelbrot et al (1997a), second part: original work.*
Proposition 3.6: *an adapted version of a proof in Mandelbrot et al (1997a).*

Chapter 4

Proposition 4.1: *original work.*
Proposition 4.2: *original work.*
Proposition 4.3: *an extended, more complete version of the proof in Calvet and Fisher (2008).*

Proposition 4.4: *original work, but inspired by Calvet and Fisher (2008).*
Corollary 4.1: *original work*
Lemma 4.1: *Calvet and Fisher (2008).*
Lemma 4.2: *Calvet and Fisher (2008).*
Lemma 4.3: *an extended, more complete version of the proof in Calvet and Fisher (2008).*
Proposition 4.5: *Calvet and Fisher (2008).*
Proposition 4.6: *Calvet and Fisher (2008).*

Appendix

Lemma 5.1: *original work.*
Lemma 5.2: *an extended, more complete version of the proof in Kahane and Peyriere (1976).*
Lemma 5.3: *original work.*
Lemma 5.4: *original work.*
Lemma 5.5: *original work.*
Lemma 5.6: *an extended, more complete version of the proof in Calvet and Fisher (2008).*
Lemma 5.7: *original work.*

7 References

Bachelier, L. (1900), "Theorie de la Speculation", *Annales de l'Ecole Normale Superieure* 3

Bacry, E., J. Delour, and J.-F. Muzy, 2000, "Modelling fluctuations of financial time series: from cascade process to stochastic volatility model", *European physical journal B* 17: 537-548

Bacry, E., J. Delour, and J.-F. Muzy, (2001), "Multifractal random walks", *Physical review* E 64: 026103-06

Bacry, E. and J.-F. Muzy, (2002), "Multifractal stationary random measures and multifractal random walks with log-infinitely divisible scaling laws"', *Physical review* E 66: 056121

Bacry, E., A. Kozehemyak, and J.-F. Muzy, (2008), "Continuous cascade models for asset returns", *Journal of economic dynamics and control* 32(1): 156-199

Bacry, E., L. Duvernet, and J.-F. Muzy (2010), "Continuous-time skewed multifractal processes as a model for financial returns", *International Journal of Theoretical and Applied Finance*

Barral, J. and B. B. Mandelbort, (2002), "Multifractal products of cylindrical pulses", *Probability theory and related fields* 124: 409-430

Billingsley, P., (1999), "Convergence of Probability Measures", 2nd edition. *John Wiley and Sons.*

Calvet, L., and A. Fisher, (2001), "Forecasting Multifractal Volatility", *Journal of Econometrics* 105: 2758.

Calvet, L., and A. Fisher, (2002a), "multifractality in asset returns: theory and evidence", *Review of Economics and Statistics* 84: 381-406

Calvet, L. E. and A. J. Fisher, (2002b), "Regime-switching and the estimation of multifractal processes". *Available at http://www.cirano.qc.ca/realisations/grandes_conferences/risques_financiers/25-10-02/Calvet-Fisher.pdf*

Calvet, L. E. and A. J. Fisher, (2004), "How to forecast long-run volatility: regimeswitching and the estimation of multifractal processes". *Journal of Financial Econometrics* 2: 4983.

Calvet, L. E. and A. J. Fisher, (2008), "Multifractal Volatility: Theory, Forecasting and Pricing", *Academic Press Advanced Finance*

Cont, R. (2001), "Empirical properties of asset returns: stylized facts and statistical issues", *Quantatitive finance* 1: 223236

Deutschel, J. D. and D. W. Stroock, (1989), "Large Deviations", *New York: Academic Press*

Di Matteo, T., R. Liu, and T. Lux, (2008), "Multifractality and Long-Range Dependence of Asset Returns: The Scaling Behaviour of the Markov-Switching Multifractal Model with Lognormal Volatility Components", *Kiel Working Paper* 1427

Eisler, Z. and J. Kerteész, (2006), "Multifractal model of asset returns with leverage effect", *Physica A - statistical Mechanics and Its Applications* 343: 603-622

Evertsz, C. J. G., and Mandelbrot, B. B. (1992), "Multifractal Measures", *in: Peitgen, H. O., Jurgens, H., and Saupe, D. (1992), Chaos and Fractals: New Frontiers of Science* 921-953

Falconer, K. (1990) "'Fractal Geometry: Mathematical Foundations and Applications"'. *John Wiley and Sons, New York.*

Frisch, U. and G. Parisi, (1985), "Fully Developed Turbulence and Intermittency", *Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics* 84-88

Galluccio, S., G. Caldarelli, M. Marsili, and Y. C. Zhang, (1997), "Scaling in currency exchange", *Physica A* 245: 42336

Graham, R. L., D. E. Knuth, and O. Patashnik, (1994), "Concrete mathematics"

Guivarch, Y., (1987), "Remarques sur les Solutions d'une Equation Fonctionnelle Non Lineaire de Benoit Mandelbrot", *Comptes Rendus (Paris)* 3051

Guivarc'h, Y, (1990), "Sur une extension de la notion de loi semistable", *Annales de l'Institut Henri Poincaré* 2: 261-285

Halsey, T. C., M. H. Jensen, L. P. Kadanoff , I. Procaccia, and B. I. Shraiman, (1986), "Fractal Measures and Their Singularities : The Characterization of Strange Sets", *Physical Review Letters* A 33:2 1141

Hamilton, J. D. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle". *Econometrica* 57: 35784.

Kahane, J.-P. and J. Peyriere, (1976), "Sur Certaines Martingales de Benoit Mandelbrot", *Advances in Mathematics* 22: 132-145

Lux, T. (2006), "The Markov-Switching Multifractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility", *Working Paper*

Lux, T. and L. Morales-Arias, (2009), "Forecasting Volatility under Fractality, Regime-Switching, Long Memory and Student-t Innovation", *Kiel Working Paper* 1532

Mandelbrot, B. B., (1963), "The Variation of Certain Speculative Prices", *Journal of Business* 36:4 394419.

Mandelbrot, B. B. (1982), "The Fractal Geometry of Nature", *Freeman*

Mandelbrot, B. B. (1989), "Multifractal measures, especially for the geophysicist", *Pure and Applied Geophysics* 131: 4-42

Mandelbrot, B. B., A. Fisher, L. Calvet, (1997a), "The Multifractal Model of Asset Returns", *Cowles Foundation Discussion Paper 1164*

Mandelbrot, B. B., A. Fisher, L. Calvet, (1997b), "Large Deviations and the Distribution of Price Changes", *Cowles Foundation Discussion Paper 1165*

Mandelbrot, B. B., A. Fisher, L. Calvet, (1997c), "Multifractality of Deutschmark/US Dollar Exchange Rates", *Cowles Foundation Discussion Paper 1166*

Mandelbrot, B. B. (2001), "Multifractal concentration", *Quantitative Finance* 1: 641-649

Mandelbrot, B. B. (2003), "Heavy tails in finance for independent or multifractal price increments", *In: Handbook of Heavy Tailed Distributions in Finance, Edited by S.T. Rache*

Mandelbrot, B. B. (2005), "The inescapable need for fractal tools in finance", *Annals of Finance 1: 193-195*

Pagan, A. (1996), "The econometrics of financial markets", Journal of Empirical Finance 3: 15-102

Riedi, R.H. (1999), "Multifractal Processes", *Available at http://www.spin.rice.edu/PDF/MP.pdf*

Schilling, R. (2005), "Measures, Integrals and Martingales", *Cambridge University Press*

Sunder, V.S. (2008), "The Riesz representation theorem", *Available at http://www.imsc.res.in/ sunder/rrt1.pdf*