The Markov Switching Multi-Fractal Model of Asset Returns: Estimation and Forecasting of Dynamic Volatility with Multinomial Specifications

Hwa Taek Lee

Department of Economics
Faculty of Economics, Business, and Social Science
Christian-Albrechts-University of Kiel

July 2007

The Markov Switching Multi-Fractal Model of Asset Returns: Estimation and Forecasting of Dynamic Volatility with Multinomial Specifications

Inaugural-Dissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Diplom-Volkswirt Hwa Taek Lee
aus Busan, Südkorea

Kiel, Juli 2007

Christian-Albrechts-Universität zu Kiel
Institut für Volkswirtschaftslehre
Lehrstuhl für Geld, Währung und Internationale Finanzmärkte
Wilhelm-Seelig-Platz 1
24118 Kiel
Germany

Gedrukt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan:
Prof. Dr. Helmut Herwartz

Erstberichterstattender:
Prof. Dr. Thomas Lux

Zweitberichterstattender:
Prof. Dr. Roman Liesenfeld

Tag der Abgabe der Arbeit:
30. Juli 2007

Tag der mündlichen Prüfung:
2. November 2007

To My Parents

And to my family without whom
I had never finished this project.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis we study the development of multi-fractal model for financial asset returns: from fractals over multi-fractal process to the multi-fractal model of asset returns (MMAR) [Mandelbrot, Calvet, and Fisher (1997)]. In particular we extend the Markov switching multi-fractal model (MSMF) [Calvet and Fisher (2004)] from binomial to trinomial and quadronomial MSMF and investigate the effects of the binomial - and the trinomial Student-t MSMF on the estimates of the multi-fractal parameters.

In chapter 1, we introduce the concept of fractal, the fractal dimension and information dimension with a typical example of the geometric fractal, i.e. Sierpinski Gasket. As stochastic fractal processes, we illustrate fractional Brownian motion (FBM), fractional Gaussian noise (FGN) and autoregressive fractional integrated moving average process (ARFIMA). The Hurst exponent plays a centre role to control long range dependence of the underlying process. To detect this Hurst effect in real time series, we shortly review ten different methods. The Hurst exponents are then estimated and evaluated.

In chapter 2, we explain how multi-fractal processes can be generated. As an example, we construct a trinomial multi-fractal process by means of a multiplicative cascade. The scaling function and the multi-fractal spectrum are derived and the features are discussed.

The multi-fractal model of asset returns has been prompted by a series of papers by Mandelbrot et al. (1997), Calvet and Fisher (2001) and Lux (2000). The multi-fractal model of asset returns (MMAR) is a product of multi-fractal process and standard Brownian motion. This model has generic multi-scaling and can generate the long range dependence of multi-fractal process. The multi-fractality of financial time series will be investigated using multi-fractal spectrum and partition function and the related multi-fractal parameters estimated.

Apart from the problem of consistency in the estimation, the multi-fractal process is generated originally in a bounded time interval via the multiplicative cascade. Despite this " bounded process", the concept of multi-fractal process may still be useful to simulate the stylised facts of financial time series (fat tails, long range dependence, volatility clustering, etc).

To overcome the problem of the *bounded* processes, there are some suggestions such as iterative random multi-fractal [Lux (2003)] and Poisson multi-fractal process [Calvet and Fisher (2001)]. However, these models do not provide a concrete mechanism by which the parameters can be estimated.

Recently, motivated by work on Markov switching models [Hamilton (1989)], Markov switching multi-fractal model (MSMF) is also proposed by Calvet and Fisher (2004): the multi-fractal process is now governed by a first order Markov process. In chapter 3, we construct Markov switching multi-fractal process and derive the moments of multi-fractal and of the compound process that

will be used later to estimate the multi-fractal parameters. Hereby, we will extend the binomial MSMF in the multinomial MSMF: trinomial and quadronomial Markov switching multi-fractal models of asset returns and investigate the features of these models.

We discuss some problems of application when using the Generalised Method of Moments (GMM). The necessary conditions are derived: conditional and unconditional moments of log-transformed processes. The maximum likelihood method (ML) of Markov switching multi-fractal model of asset returns is reviewed shortly. In contrast to two state Markov switching model, Calvet and Fisher (2004) constrain the transition matrix in which the effect of multi-scaling is preserved over hierarchical structure of multipliers. Switching of volatility component does not suffer from this constrained probability and it enhances the whole volatility dynamics. Using Bayesian updating, we will estimate the essential parameters of MSMF. For GMM and ML, we conduct Monte Carlo simulation and report the small sample properties of these methods.

There is the problem to select the number of multipliers in the discrete MSMF. In the subsection of 3.6 the problem of model selection is discussed with the tests by Vuong (1989), Clarke (2003) and Markov switching criterion. We execute these tests for the real financial data and evaluate the optimal number of multipliers in the binomial and trinomial MSMF. We observe some trade-offs between the multi-fractal parameters (m_i) and the scaling factor (σ) depending on how large the numbers of multipliers are chosen.

To capture the excess kurtosis, we specify trinomial MSMF with i.i.d innovations from a fat-tailed distribution, Student-t distribution which is a natural choice to accommodating such stylised facts. The pertinent moments of this distribution are derived. The small sample properties of GMM and ML are compared and investigated how the estimated values of multi-fractal parameters are changed due to the inclusion of the degree of freedom.

In the fifth chapter we conduct the volatility forecasting using the trinomial MSMF and compare the performance of trinomial MSMF with those of generalised autoregressive conditional heteroscedastistic (GARCH) and of fractional integrated GARCH (FIGARCH). In the first section of this chapter we introduce the Levinson - Durbin procedure which is employed for volatility forecasting to use the autocovariance function of trinomial MSMF with the estimates from GMM. Lux (2006) proposes this algorithm to forecast the volatility of exchange rate returns using binomial MSMF. After Monte Carlo experiments of the trinomial MSMF with respect to the performance of volatility forecasting, the binomial and trinomial MSMF are competing with each other as well as with the benchmark models such as GARCH and FIGARCH.

As last, we summarise briefly our major extension: trinomial and quadronomial MSMF and Student-t MSMF in chapter 6. We mention further research area like multivariate MSMF and Markov chain Monte Carlo (MCMC) for continuous MSMF.

Chapter 2

Fractal and Multi-fractal Processes

What is a fractal? [1]

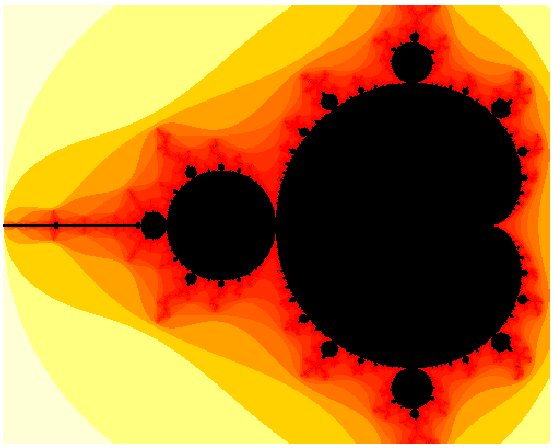

Figure 2.1: Mandelbrot Set

A fractal is a fragmented geometric object that can be subdivided into parts, each of which is a reduced replica of the whole [Mandelbrot (1983)]. Fractals are generally self-similar and independent of scale, i.e. scale-free. An object is said to be fractal if it combines the following characteristics: its parts have the same form or structure or moments as the whole, except that they are at different scale.

See the figure (2.1): Mandelbrot set. We can select randomly segments of this set. Independent of the length of segment, the Mandelbrot set looks similar. This chosen part possesses the same statistical properties over all scales of magnifications. We refer to scale invariance for that. There

[1] As the father of fractal geometry, Benoit Mandelbrot gave the word fractal in 1975, from the Latin adjective "fractus" and verb "frangere", meaning to break.

is no characteristic scale to define it.

In general, it is not enough to describe fractals with only Euclidean dimension, i.e. in terms of lines, squares, cubes, etc. We need a dimension which is not integer to describe the artificially generated Mandelbrot set as well as real objects, such as trees, meandering of rivers, lightning and coastlines.

The fractal dimension is the basic concept for describing structures that exhibit scaling. Scaling means self-similarity of the considered object on varying scale of magnification. Note that scaling invariance describes a unchanged property of geometrical object or statistical distribution under translation of temporal or spatial scales. The concept of scaling invariance is very important in the theory of critical phenomena. For example, a system or process is described by power laws, which are a direct consequence of the scale invariance. This scaling rate describes growth rates of moments as the time (or spatial) scale increases, e.g. from day to week or month.

Due to the property of scale invariance and self-similarity, fractals are used by a wide range of academic disciplines (physicist, chemists, biologists and economist). Fractals have been used to describe not only natural phenomena, but to provide new insights and improved applications. Mainly motivated by power laws, we introduce shortly fractals and related concepts : self-similarity and fractal dimension[2]. This section provides an introductory review of some concepts related with fractals and self-similarity. After illustrating the Sierpinski Gasket as a typical fractal, we introduce as example of fractals that are stochastically self-similar: fractional Brownian motion, fractional Gaussian noise and autoregressive fractional integrated moving average process. The relation between self-similarity and long range dependency is also discussed. As a parameter of long range dependence, the Hurst exponent is introduced and estimated for real financial data using various methods.

2.1 Fractals as Geometric objects

Fractals are geometric objects exhibiting an complex, highly irregular appearance on all resolutions [Mandelbrot (1983)]. In a deterministic setting, this imposes strong restrictions on the generation of fractals and the easiest way to obtain such an object is to apply a simple geometrical rule iteratively to obtain details up to infinitely fine resolution. Consequently, deterministic fractals consist of highly repetitive patterns: the whole and its parts cannot be geometrically distinguished from each other.

Let's begin to illustrate the basic features of fractals on a simple example: the so-called Sierpinski Gasket. We start with a square. In the first step the original shape is split into 4 smaller squares. In the second step we shade out one square in the north-east of the original square. The side lengths of three squares are cut into halves. In this way we can see that at each iteration, one quarter of the original square is removed. That is, three quarters of the area of the original square is left after the first iteration.

We think of this L-form as initiator at a level of iteration 1. In the third step the remaining squares are split into 12 equilateral squares and we remove the north-eastern part of each square as before.

In the next step the same procedure then applies to the remaining nine squares and so on. We see the result from iteration step 2,3,4,5. So after an infinite number of iterations, you would find there was no area at all: we reach the last step allowed with the accuracy of a square, the size of a pixel. It is the Sierpinski Gasket.

If we look at the Sierpinski Gasket using higher and higher degree of magnification, we can see

[2]For a good introduction to fractals, we recommend the books of Mandelbrot (1983) and Gleick (1987).

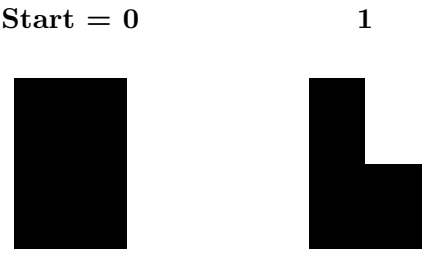

Figure 2.2: Sierpinski Gasket: first iteration

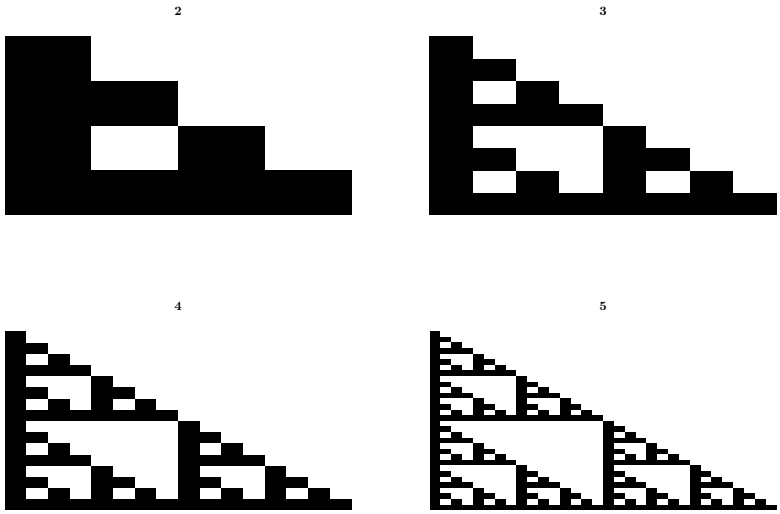

Figure 2.3: Sierpinski Gasket: 2,3,4,5 iteration

that its form is repeated on finer and finer scales. The Sierpinski Gasket looks the same under rescaling. It has strict self-similarity as a typical example of fractal.

Fractals are characterised by their fractal dimension, which is usually but not always a non-integer dimension less than its Euclidean dimension. The Euclidean dimension D_e is easily defined by the number of coordinates required to specify the object. A point is 0 dimensional, a line is 1 dimensional, a plane is 2 dimensional and a volume is 3 dimensional. However, fractals have fractional dimensions that fall in between 1 and 2 or between 2 and 3. Therefore, the fractal dimension is a more general concept for describing structures that exhibit scaling.

In case of the Sierpinski Gasket, we are always taking the original L-form, initiator, and replacing it with three smaller L-forms. Each side length of the L-form after iteration is 1/2 of the side length of the original. So, we are using a scale factor of 2. The scale and dimension are related by the formula: $S^D = N$, where S is the scale factor, D is the dimension, and N is the number of similar copies obtained.[3] When we apply this formula to the Sierpinski Gasket, the

[3]The Euclidean dimension of a square is 2. Consider the square. Now let's build similar squares at a smaller

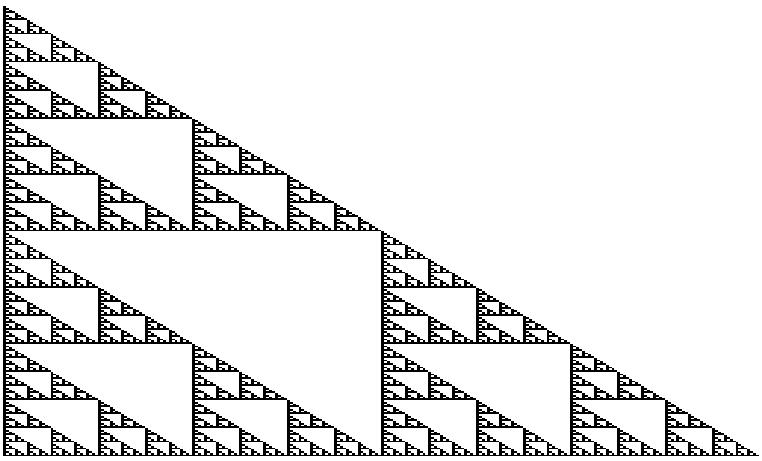

Figure 2.4: Sierpinski Gasket

scale-dimension formula gives $2^D = 3$. What power of 2 gives 3? The answer can be found directly using a property of logarithms:

$$\begin{aligned} S^D &= N \\ D\log(S) &= \log(N) \\ D &= \frac{\log(N)}{\log(S)} \end{aligned} \tag{2.1}$$

So we calculate log(3)/log(2) to get 1.584962. This is the fractal dimension ($D = 1.584962$). In short, a simple fractal structure in a D-dimensional space can be quantified by covering the structure with D-dimensional neighbourhoods of linear size ϵ. If the minimum number of such neighbourhoods needed to cover the whole structure is $N(\epsilon)$ then this number will scale as

$$N(\epsilon) \sim \left(\frac{1}{\epsilon}\right)^D \tag{2.2}$$

The fractal dimension D derived in this way is also referred to as the box dimension. This dimension concept goes back to the German mathematician Felix Hausdorff. The self-similarity dimension is a simplified version of the Hausdorff dimension. We introduce as a general concept for the self-similarity dimension the Hausdorff dimension which is one of the most important methods of characterising fractals. Felix Hausdorff introduced the dimension, providing a measure for filling space which allows for the possibility of non-integral dimensions. A fractal dimension has the following definition: Let M be partial set of the n-dimensional real number space. Now we cover his n-dimensional space with cubes having edges of length ϵ and count this cubes $N(\epsilon)$, which contains a piece of the set M. The limit

scale. We'll use a scale factor of 3 (the side of the large square is 3 times longer than the small ones). Notice that we can get 9 "self-similar" squares out of the original square at this scale factor. So: Dimension = 2, Scale = 3, and Number of self-similar objects = 9.

$$D_F = \lim_{\epsilon \to 0} \frac{\ln N(\epsilon)}{\ln \epsilon} \tag{2.3}$$

is called fractal dimension or Hausdorff dimension. The Hausdorff dimension is often referred to as box-counting dimension.

This definition has a definite resemblance to that used to define the self-similarity dimension. However, the calculation of the self-similarity dimension requires that exactly self-similar parts of the fractal are identified, where as for the Hausdorff dimension allows for a greater flexibility in the type of object that can be investigated. But in practice box-counting dimensions are much easier to work with. We cover the object with covering elements or boxes of the side length ϵ. The number of boxes $N(\epsilon)$ required to cover the object is related to ϵ through its box counting dimension D_f.

Another type of fractal dimension is the information dimension. It is similar in concept to the Hausdorff dimension. However it tries to accommodate the frequency with which a trajectory visits each covering cube or sphere. The information dimension is calculated in the following manner. The number of points N_i in each of the N cells is counted and the probability of being the points in that cell $P_i = N_i/N_0$ is denoted as P_i. Expressing all this in equation form: $P_i = N_i/N_0$, where $\sum_i^N P_i = 1,\ N_0 \neq N$. N_0 is the total number of points in the set. The information entropy is thus defined to be: $I(\epsilon) = -\sum_i^N P_i \log(P_i)$. Using this definition of entropy the information dimension is defined to be:

$$D_I = \lim_{\epsilon \to 0} \frac{I(\epsilon)}{\ln(1/\epsilon)} = \lim_{\epsilon \to 0} \frac{-\sum_i^N P_i \ln(P_i)}{\ln(1/\epsilon)} = \lim_{\epsilon \to 0} \frac{\sum_i^N P_i \ln(P_i)}{\ln(\epsilon)} \tag{2.4}$$

This information dimension is closely related to the fractal dimension. Investigated different strange attractors like deterministic chaos and random noise, Grassberger and Procaccia (1983) have demonstrated that the information dimension is a lower bound to the fractal dimension in general. That is $D_I \leq D_F$. In summarising the above relationships, it is clear that the fractal dimension does not consider the distribution of points between covering cells. However the information dimension measures the probability of including points in a cell.

A fractal does not have an integer but a fractional dimension.[4] Since a fractal like the Sierpinski Gasket is self-similar, there is no characteristic scale. A fractal is to be characterised by the fractal dimension. It is a measure of roughness. The higher the values of fractal dimension, the more rougher the surfaces of profiles. The smaller the values of fractal dimension, the more smoother the surfaces of profiles.

Patterns like the Sierpinski Gasket can rarely be found in empirical finance. Nevertheless, similarity on all scales sometimes holds in a statistical sense, leading to the notion of stochastic fractals or self-similar stochastic process.

[4]Such a fractal is also referred to as strange, if the fractal dimension is not an integer number. Strange sets always have gaps or holes. Hence, the fractal dimension gives some information about the fragmentation of a set. Note that the fractal dimension is only identical with a topological space dimension, if the set does not contain any gaps or holes.

2.2 Stochastic Processes with fractal properties

In this section we introduce stochastic fractals. Stochastic fractals are not deterministically self-similar like the Sierpinski Gasket. Note that the self-similarity associated with the Sierpinski Gasket can not be found in the financial data. However, a financial time series looks the same or behaves the same when viewed at different time scales on a time dimension. This means that parts of the process are similar to the whole process. Thus, the whole time series and its samples cannot be statistically distinguished from each other.[5]

We consider stochastic fractals as a stochastic process which is self-similar statistically. As a typical example of stochastic fractals, we will illustrate fractional Brownian Motion (FBM), fractional Gaussian noise (FGN) and autoregressive fractional integrated moving average process (ARFIMA). In principle these processes can be used to describe the price or return behaviour of financial assets. Furthermore, we discuss the connection of self-similarity with long range dependence. With respect to the stylised facts of financial time series, stochastic fractals often incorporate long rang dependence and volatility clustering.

Self-similarity of a stochastic process is defined as follows:[6] Let $X(t)$ be a stochastic process in continuous time t. $X(t)$ is self-similar with self-similarity parameter $H > 0$, if the rescaled process with time scale t/c, $c^H X(t/c)$ is equal in distribution to the original process $X(t)$.

$$X(t) \stackrel{d}{=} c^H X(t/c), \forall c, H > 0, \tag{2.5}$$

where $\stackrel{d}{=}$ means equality for all finite dimensional distributions. Hence, for a positive dividing factor c, the distribution of $c^H X(t/c)$ is identical to that of $X(t)$.[7] In terms of a financial time series this means that whether the data in question is intra-day, daily, weekly, or monthly data, the resulting processes look similar. We remark that fractals do not have a characteristic scale at which its features occur because they occur at all scales equally. The stochastic fractals or stochastic fractal processes do not possess strict self-similarity, but possess statistical self-similarity. As a scaling factor, the so-called Hurst exponent H plays a central role in stochastic fractals.

Assumed that the unconditional moments of $X(t)$ exist, they behave as power-laws of time:

$$E[|X(t)|]^q = E[|X(1)|^q]|t|^{qH}. \tag{2.6}$$

As can be seen from the equation (2.6), the self-similar process with $H > 0$ is non-stationary in general. To obtain a stationary process, we take the increments of self-similar process ($Y(t) = X(t) - X(t-1)$). For example, FGN is the self-similar processes with stationary increments. The corresponding self-similar process $X(t)$ is called FBM. FBM is an attractive model to describe the scaling effect because it is mathematically well-defined.

2.2.1 Fractional Brownian motion

As the first example of stochastic fractals, we consider fractional Brownian motion (FBM). In the work of Mandelbrot and Van Ness (1968), FBM is defined by its stochastic equation

[5]The absence of such scales means that standard techniques based on characteristic times for example simple Markov models fail adequately to analyse and to model such processes and new processing tools are needed for that [Abry, Baraniuk, Fladrin, Riede, and Veitch (2002)].

[6]For the definition and the property of FBM and FGN, we follow Beran (1994).

[7]Provided that a self-similar process is properly rescaled, the property of distribution is not changed by a translation or dilation of this process. It is of particular interest in the stable distribution: the sum of this process has the same shape as the initial distribution. It is called stable laws (see Bouchaud and Potters (2000) for details). The Gaussian and Lévy distributions belong to the family of stable distribution.

$$B_H(t) = \frac{1}{\Gamma(H+\frac{1}{2})}\left(\int_{-\infty}^{0}\Big[(t-s)^{H-1/2} - (-s)^{H-1/2}\Big]dB(s) + \int_0^t (t-s)^{H-1/2}dB(s)\right), \quad (2.7)$$

where Γ denotes the Gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-s)dx$ and $0 < H < 1$ is called the Hurst exponent or the self-similar parameter. We use the Hurst exponent to characterise the temporal development of a stochastic process. By taking H = 0.5, we get a standard Brownian motion. These are some properties of fractional Brownian motion:

1) $B_H(t)$ has a Gaussian distribution with stationary increments,

2) $B_H(0) = 0$,

3) $B_H^2(t) = Var(B_H(1))t^{2H}$ for $t \geq 0$

4) $Cov(B_H(t), B_H(s)) = \frac{1}{2}\big(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\big)Var(B_H(1))$.

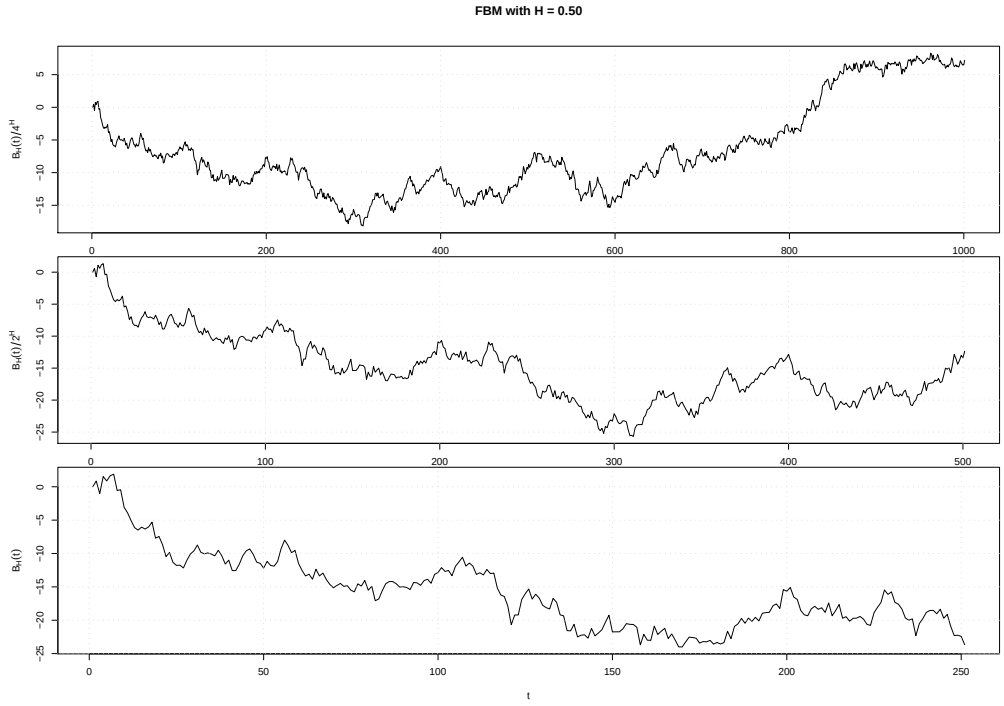

Figure 2.5: Fractional Brownian Motion (FBM) with H = 0.5 as an example of self-similar process. A sample path is generated for $0 < t < 1000$. The middle panel shows the normalised sample path for $0 < t < 500$. The bottom panel shows the normalised sample path for $0 < t < 250$.

From the above properties, we conclude that the two processes share the same finite distribution:

$$B_H(at) = a^H B_H(t)$$

for $0 \leq t \leq \infty$. Independent of the constant a, a fractional Brownian motion ($B_H(at)$) with the Hurst exponent H is self-similar with its scaled fractional Brownian motion ($a^H B_H(t)$) with the same Hurst parameter H. Brownian motion is just a special case of fractional Brownian motion with the Hurst exponent (H = 0.5). The figure (2.5) presents a typical example of stochastic self-similar process. In contrast to the deterministically generated fractal as an exactly self-similar

object, a sample process of FBM ($H = 0.5$) seems to be *stochastically* similar as we zoom in the process. The time window of the middle panel is just the first half of the upper panel and the time window of the bottom panel is the first half of the middle panel. Each process is rescaled corresponding to the time window. Note that fractional Brownian motion is the Gaussian process with a unique covariance structure (see the property 4)). For any $a > 0$, we can easily extend the self-similarity to the autocovariance function (see Beran (1994)):

$$Cov(Y(at), Y(as)) = a^{2H} Cov(Y(t), Y(s)) \tag{2.8}$$

2.2.2 Fractional Gaussian Noise

We introduce as the second example of stochastic fractals fractional Gaussian noise. Let's define the incremental process, $Y(t)$ $t = 0, \ldots$, of fractional Brownian motion, which is called fractional Gaussian noise (FGN), by

$$Y(t) = B_H(t+1) - B_H(t).$$

Since the process $Y(t)$ has independent and stationary increments, FGN is self-similar with the Hurst exponent due to

$$Y(at) \stackrel{d}{=} a^H Y(t).$$

It has a standard normal distribution for every t. The autocovariance function of FGN is given by

$$\gamma(t) = \frac{1}{2}\sigma^2 \Big[|t+1|^{2H} - 2|t|^{2H} + |t-1|^{2H} \Big]. \tag{2.9}$$

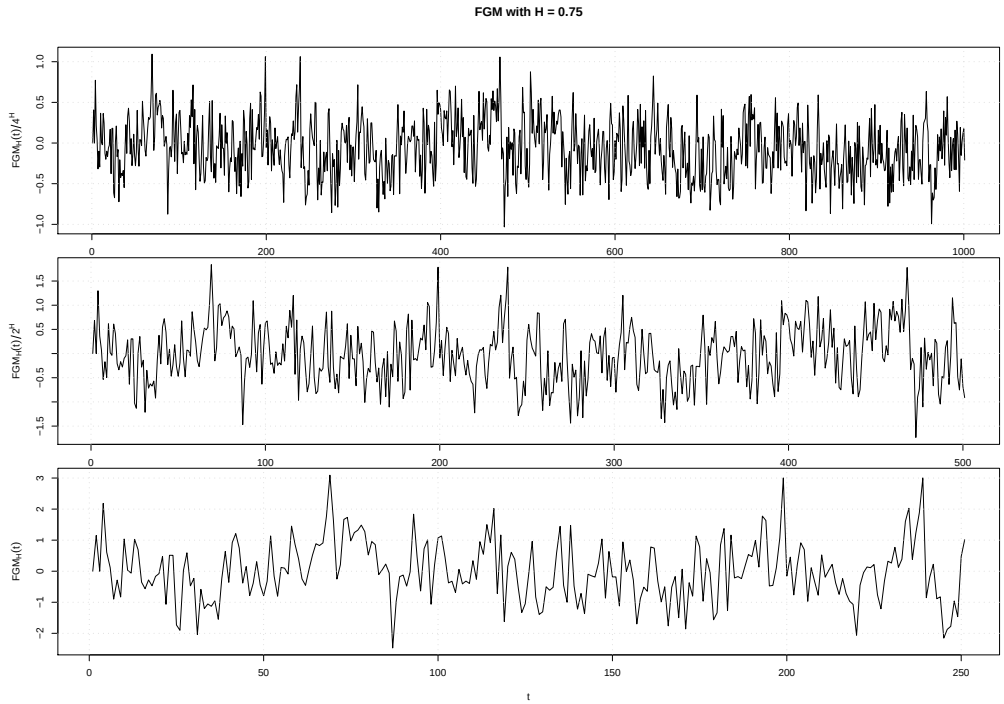

Figure 2.6: Fractional Gaussian Noise (FGN) with H = 0.75 as an example of self-similar process. A sample path is generated for $0 < t < 1000$. The middle panel shows the normalised sample path for $0 < t < 500$. The bottom panel shows the normalised sample path for $0 < t < 250$.

If H = 0.5, all the covariances are 0 except for t = 0. Since fractional Gaussian noise is a Gaussian process, this implies independence. This agrees with the properties of standard Brownian motion, which has independent increments. By writing down the Taylor expansion at the origin of the function $g(x) = (1-x)^{2H} - 2 + (1+x)^{2H}$ and noting that the autocorrelation is given by $\rho(t) = \frac{1}{2}t^{2H}g(1/t)$ for $t \geq 1$, it follows from (2.9) that

$$\rho(t) \sim H(2H-1)t^{2H-2}, \tag{2.10}$$

as $t \to \infty$.

We divide this process in the three categories:

i) Setting H = 0.5 in the autocorrelation function gives

$$\rho(t) = 0$$

for any t, which shows that the classical Brownian motion is a fractional Brownian motion with Hurst exponent H = 0.5. It is just a Gaussian identical independent distributed (i.i.d) process where all autocorrelations are zero at non-zero lags. The observations $Y(t)$ are uncorrelated and there is no memory.

ii) For $0 < H < 0.5$, this process is called anti-persistent. Such a process reverses itself more frequently than a purely random one. the autocorrelation function can be summed so that it holds:

$$\sum_{k=-\infty}^{\infty} \rho(k) = 0.$$

Mandelbrot called it anti-persistent.

iii) For $0.5 < H < 1$, this Gaussian process is called persistent. A persistent process is characterised by long range dependence or long-memory effects. Its autocorrelation function decay so slowly to zero that they can not be summed:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty.$$

In the presence of long-memory effects, all daily returns are correlated with all future daily returns, and all weekly returns are correlated with all future weekly returns, and so on. Note that there is no characteristic time scale for such fractal time series. But we can easily characterise self-similar process with stationary increments by the so-called Hurst exponent. This Hurst exponent H is important because it determines the fractal dimension of the sample paths of the process and then the intensity of long-range dependence.[8] Note that the fractal dimension measures the degree of irregularity or roughness of a stochastic process [Falconer (2003)].

Figure (2.7) consists of simulated sample paths for three different values of $H = 0.25, 0.5, 0.75$. It shows us self-similarity of fractional Gaussian noise with the Hurst exponents $0 < H < 1$. We can easily check from equation (2.9) that the autocovariances are negative for $H < 0.5$ and positive for $H > 0.5$. This behaviour is also recognised in figure (2.7), in which samples of fractional Gaussian noise in the right panel are depicted for the values of the Hurst exponent ($H = 0.25, 0.5, 0.75$). For $H = 0.25$, the sample path is clearly different from the fractional Gaussian noise with $H = 0.5$ (see also the standard Brownian motion). The negative autocorrelation accounts for the high variability,

[8]For finite variance process, the relation between H and d is $H = d + 0.5$. d is the fractional difference operator of ARFIMA(0,d,0): $\phi(B)(1-B)^d Y(t) = \theta(B)\epsilon(t)$, where $-0.5 < d < 0.5$

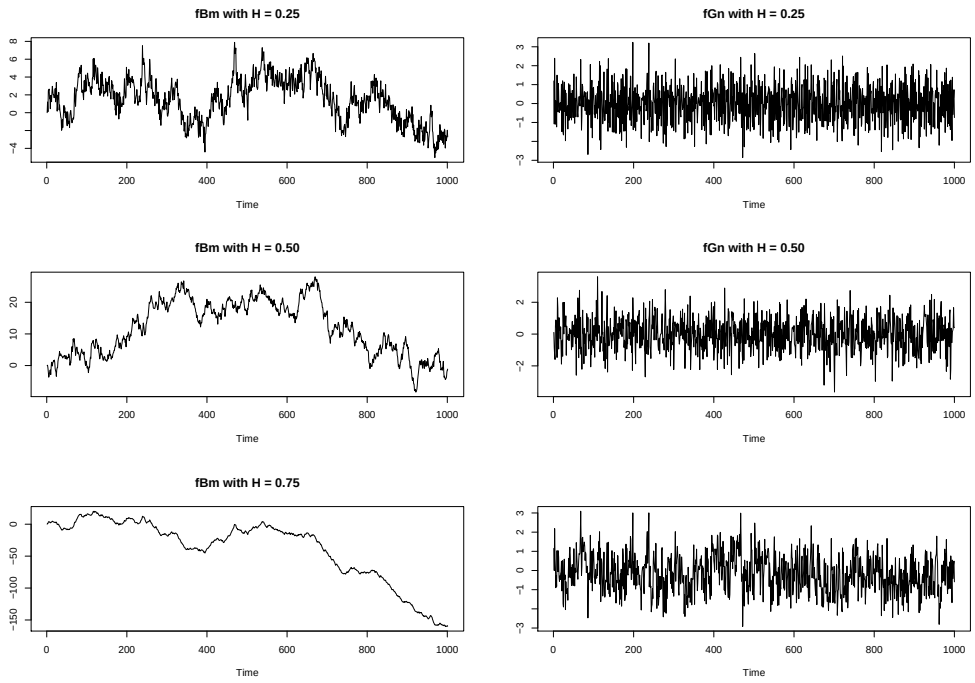

Figure 2.7: Fractional Brownian Motion (FBM) and Fractional Gaussian Noise (FGN): H = 0.25,0.5,0.75.

whereas for $H = 0.75$ the positive autocorrelation is responsible for the long range dependence: there are clearly periods in which the sample path increases and periods in which it decreases.

The left panels of figure (2.7) plot the cumulative sums of the same samples of fractional Gaussian noise with the same Hurst exponents. We observe typically the negative correlations of the fractional Brownian motion with H = 0.25 in the left-top panel. It is also called anti-persistent. The sample path of the left-bottom panel is persistent for H = 0.75 and more smooth due to the positive autocorrelations.

For self-similar processes, the local properties are reflected in the global ones and vice versa. Hence, self-similar processes are *invariant* stochastic processes in distribution under suitable scaling of time.[9] To characterise such invariant stochastic process in a global sense, we use the Hurst exponent (H) as a parameter of self-similar process, whereas fractal dimension is used as a measure of roughness.[10] We often refer to the Hurst effect of time series when long-range dependence in the time series is associated with power correlations.

2.2.3 ARFIMA

We discuss another example of a self-similar process: autoregressive fractional integrated moving average process (ARFIMA)[11](p,d,q). For only $-0.5 < d < 0.5$, the ARFIMA(p,d,q) process $X(t)$ is a self-similar process that it is the solution of

[9]Mandelbrot (1963) and Mandelbrot and Van Ness (1968) maintained that the distribution of the price changes for a long-term time scale can be calculated from a distribution for a shorter time scale $\tau < T$:

$$P_T(x) = \frac{1}{\lambda} P_\tau(\frac{x}{\lambda}) \;\; with\; \lambda = (\frac{T}{\tau})^H,$$

where H is the self-similarity parameter.

[10]Gneiting and Schlather (2004)

[11]see Beran (1994)

$$\phi(L)(1-L)^d X(t) = \psi(L)\epsilon(t), \tag{2.11}$$

where $\epsilon(t)$ is a white noise process, the polynomial terms $(\phi(x), \psi(x))$ are defined by $\phi(x) = 1 - \sum_{j=1}^{p} \phi_j x^j$ and $\psi(x) = 1 + \sum_{j=1}^{q} \psi_j x^j$ respectively. As a self-similar process, the normalised partial sums of a ARFIMA process have the same limiting distribution (see Beran (1994)). The ARFIMA(p,d,q) process allows a fractional value for the parameter d $(-0.5 < d < 0.5)$. The figure (2.8) illustrates the self-similarity of this increment process. We refer to this special order of ARFIMA (0,d,0) process as fractionally integrated noise. The fractional differencing operator $(1-L)^d$ is defined for non-integer[12] d by an infinite binomial expansion

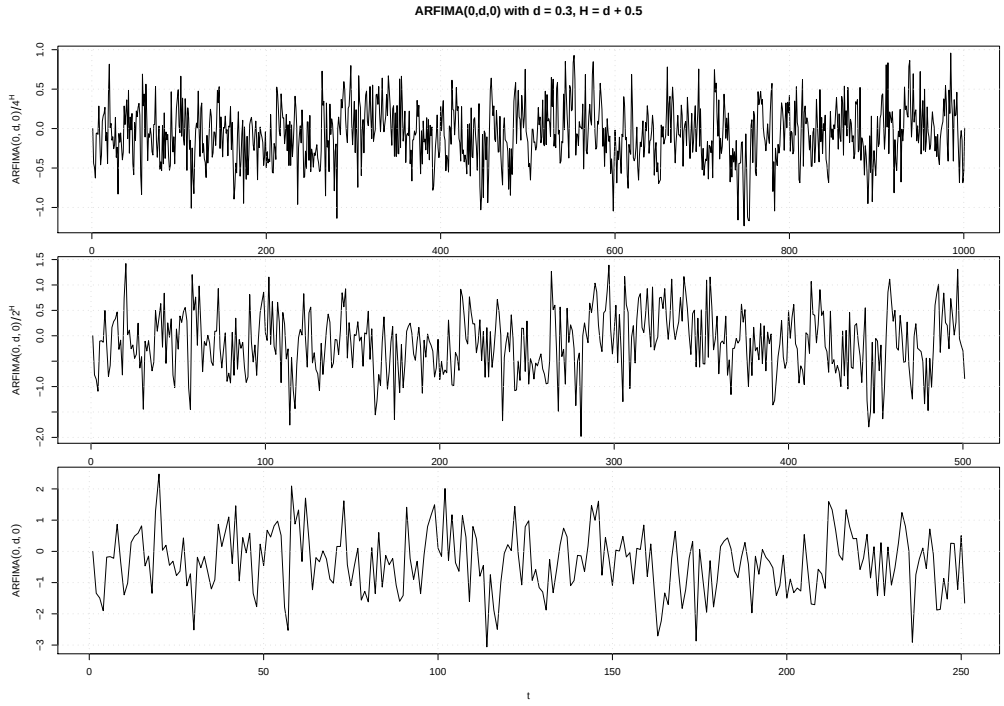

Figure 2.8: Autoregressive fractional integrated moving average process (ARFIMA(0,d,0) with d = 0.3 as an example of self-similar process. Note that $d = H - 0.5$. A sample path is generated for $0 < t < 1000$. The middle panel shows the normalised sample path for $0 < t < 500$. The bottom panel shows the normalised sample path for $0 < t < 250$. As a fractionally integrated noise, the ARFIMA(0,d,0) is is the increment process for a self-similar process.

$$\begin{aligned}(1-L)^d &= 1 - dL + \frac{d(d-1)}{2!}L^2 - \frac{d(d-1)(d-2)}{3!}L^3 + \cdots \\ &= \sum_{i=0}^{d} \left(\begin{array}{c} d \\ k \end{array} \right) (-1)^k L^{d-k}. \end{aligned} \tag{2.12}$$

To proceed from integer value of binomial coefficients to real numbers d, we can also rewrite the binomial coefficients as the following Gamma function:[13]

[12]Integer values of d lead to traditional ARIMA(autoregressive integration moving average) models.
[13]The expansion can also be presented in terms of the hyper-geometric function:

$$\left(\begin{array}{c} d \\ k \end{array} \right) = \frac{d!}{k!(d-k)!} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$$

Note that the Gamma function, denoted Γ, is defined for all real numbers. The long range properties of such processes depend on the values of d. For $d \in (0, 0.5)$ we have a stationary long memory process. This process is asymptotically second order stationary. The corresponding Hurst exponent is $H = 0.5 + d$. The autocorrelations of this process are all positive. They decay monotonically and hyperbolically to zero as the lag increases. When d=0, the process is a ARMA(p,q) process, i.e. a short-memory process.

For $d \in (-0.5, 0)$ the ARFIMA(0,d,0) process has a negative correlation: the autocorrelations are all negative, except $\rho(0) = 1$. It is also said to exhibit anti-persistent or intermediate memory. They also decay monotonically and hyperbolically to zero.

For $d \geq 0.5$ the series are no longer covariance stationary, and have infinite variance. By taking appropriate differences, we can reduce a non-stationary process with $d > 0.5$ to the stationary ARFIMA process(p,d,q) with $-0.5 < d < 0.5$. For a more detailed discussion see, for example, Baillie, Bollerslev, and Mikkelsen (1996).

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)} L^k = F(-d, 1, 1, L).$$

2.3 Tools for estimation of Hurst exponent

We have seen that a fractional stochastic process with Hurst exponent can be used to model various degrees of long-range dependence: FBM, FGN, and ARFIMA(p,d,q). In this section we review some popular methods to estimate the parameter for the intensity of self-similarity or shortly Hurst exponent. Most of the methods are described in Taqqu, Teverovsky, and Willinger (1995) and the monograph of Beran (1994). The latter gives a compact explanation of nine estimators and a numerically based comparison of their performance. In addition, we introduce the wavelet method. Note that many of them can be used to distinguish a fractional process from a purely random time series.

To exploit the long-range dependence of a time series, most of the following methods are based more or less on the same idea. They firstly form the averaged or aggregated time process with aggregation level m as $X(t)^{(m)} = \frac{1}{m}\sum_{i=tm-m}^{tm} X(i)$ and they look at the changing variance in each level of aggregation:

$$Var[X(t)^{(m)}] \sim T^{-\beta},\ N \to \infty.$$

By means of a log(m) versus $log(Var[X(t)^{(m)}])$ plot, one obtain the estimate of β from a least-square fit.

I *Aggregated Variance method :* The aggregated variance method is based on the self-similarity property of the sample time series. The time series $X(t), t = 1, \ldots, N$ is divided into blocks labelled by index k. Each block has m elements. The aggregated series is calculated as the mean of each block:

$$X^{(m)}(k) = \frac{1}{m}\left(X_{km} + \ldots + X_{(k+1)m-1}\right),$$

for $k = 0, 1, \ldots$. Since we suppose that $X^{(m)}$ has asymptotically the same distributions as $m^{H-1}X$ for large m, we obtain the rescaled relation between two variances:

$$Var(X^{(m)}(k)) = m^{2H-2}Var(X(k))$$

when using the self-similar property. Due to the stationary assumption, the expectation of the variances of $X^m(k)$ is equal for every k:

$$\widehat{VarX^{(m)}} = \frac{1}{M}\sum_{i=1}^{M-1}\left(X^{(m)}(i) - \overline{X^{(m)}i}\right)^2,$$

where M is the integer part of N/m and $\overline{X^{(m)}}$ is the sample average of $X^{(m)}$:

$$\overline{X^{(m)}} = \sum_{i=0}^{M-1} X^{(m)}(i).$$

This procedure is repeated for different values for $m_i, 1 \geq i$. The chosen values for m should be equidistant on a log scale, i.e. $m_{i+1}/m_i = C$, where C as a constant depends on the length of the time series and the desired number of points.

Since the expectation of the variances of fractional Gaussian noise or autoregressive fractional integrated moving average process (ARFIMA) behaves like $Var(X^{(m)}(k)) = m^{2H-2}Var(X(k))$ as $m \to \infty$, we can obtain an estimate for H from a straight line with slope 2H-2 when plotting a log-log representation of m and $\hat{VarX^{(m)}}$.

II *Differencing the Variance :* Assume that there is a non-stationary time series that includes jumps in the mean and slowly decaying trends. To distinguish them from long-range depen-

dence, one can difference the variance: $\hat{Var}X^{(m_{i+1})} - \hat{Var}X^{(m_i)}$. This method provides a way of detecting the types of the non-stationarity. It is used in conjunction with the basic aggregated variance method.

III *Absolute moments of the aggregated series :* The absolute moments method is a generalisation of the aggregated variance method. It also uses the self-similar property that $X^{(m)}$ hat the same distribution as $m^{H-1}X$ for large m. Starting again with splitting and aggregating the series, one obtains the sum of the absolute values of the aggregated series, $\frac{1}{M}\sum_{i=0}^{M-1} |X^{(m)}(i) - \overline{X^{(m)}}|$. The logarithm of this statistic is plotted versus the logarithm of m. The result yields a line with slope $H-1$ if the underlying series has long-range dependence with parameter H.

IV *Higuchi's method :* Higuchi (1988) originally proposed to calculate the fractal dimension of a time series. It is similar to the method of absolute values of the aggregated series. At first, one constructs the partial sums $Y(m) = \sum_{i=1}^{m} X(i)$ of the original time series $X(i), i = 1, \ldots, N$. Then one find the normalised length of the curve:

$$L(m) = \frac{N-1}{m^3} \sum_{i=1}^{m} \Big[\frac{N-i}{m}\Big]^{-1} \sum_{k=1}^{[(N-i)/m]} \Big|Y(i+km) - Y(i+(k-1)m)\Big|,$$

where N is the length of the time series, m is the block size and [ ] denotes the greatest integer function. Higuchi defines the curve length for the time interval of m as the average value over m sets of L(m). If $L(m) \sim C_H m^{-D}$ within the rage $m_{min} \leq m \leq m_{max}$, the curve is self-affine with fractal dimension $D = 2 - H$ in this range. A log-log plot of L(m) versus m should yield a straight line with a slope of 2 - H.

V *Detrended fluctuation analysis (DFA) :* Peng, Buldyrev, Havlin, Simons, Stanley, and Goldberger (1994) suggest DFA to detect the long-range correlations of time series. The DFA method involves the following steps. 1) The time series is divided into non-overlapping blocks of size m. Within each of the blocks, the partial sums of the series, $Y(i), i = 1, \ldots, m$, are calculated. 2) One fits a least-squares line to the Y(i) and calculates the sample variance of the residuals. 3) One repeats this procedure for each of the blocks and computes the average of the resulting sample variances. It is apparent that the sample variance increases with the block size. For large m, the resulting number is proportional to m^{2H} for FGM and ARFIMA. When the result is plotted on log-log plot versus m, we should get a straight line with a slope of 2H.

VI *R/S method :* The rescaled range (R/S) method is often used to obtain the Hurst exponent. For time series $X(t), i = 1, \ldots, N$, we divide the whole time length N into K blocks (N/K) and construct the partial sum $Y(n) = \sum_{i}^{n} X(i)$ and the corresponding sample variance $S^2(n) = (1/n)\sum_{i=1}^{n} X(i)^2 - (1/n)Y(n)^2$. The rescaled-adjusted-range statistic is given by

$$\frac{R}{S}(n) = \frac{1}{S(n)}\Big[\max_{0\leq t\leq n}\Big(Y(t) - \frac{t}{n}Y(n)\Big) - \min_{0\leq t\leq n}\Big(Y(t) - \frac{t}{n}Y(n)\Big)\Big].$$

With this procedure, we get K different estimates of $R/S(n)$ when dividing the whole series of N consecutive values into K blocks, each of size N/K. We define the starting points of each block $k_{(m)} = \frac{(m-1)N}{K} + 1, m = 1, \ldots, K$, where K is the total number of blocks and m is the current block number. For each lag n, we compute $R(n, k_{(m)})/S(n, k_{(m)})$ such that $k_{(m)} + n < N$. Plotting the $\log(R(n, k_{(m)})/S(n, k_{(m)}))$ for each block versus $\log(n)$ allows the slope of the fitted straight line to be estimated. Note that we use the values of n between some lower and higher cut-off points to estimate H.

Thus R/S is called the rescaled adjusted range as its mean is zero, and it is expressed in terms of the local standard deviation. For large n, the expected value of this statistic approaches $C_H n^H$:

$$E[R/S(n) \sim C_H n^H,$$

where C_H is a positive, finite constant that does not depend on n. The R/S method is a non-parametric tool and therefore does not require the assumption about the limiting distribution of underling time series. In case of an i.i.d Gaussian process, a Hurst exponent of H = 0.5 implies that there is no long-range dependence in the time series.

VII *GPH :* Geweke and Porter-Hudak (1983) first proposed regressing the log values of the periodogram on the logarithmic spectral density function. We refer to this as the GPH estimator. For a time series $X(j), j = 1, \ldots, N$, one calculate the spectral density of this series:

$$I(\lambda, N) = \frac{1}{2\pi N} \left| \sum_{j=1}^{N} X(j) \exp^{ij\lambda} \right|,$$

where λ is a frequency and N is the number of time series. Because $I(\lambda, N)$ is an estimator of the spectral density, a time series with long-range dependence should have a periodogram which is proportional to $|\lambda|^{1-2H}$ close to the origin. The regression is based on the following equation:

$$\ln I(\lambda, N) = a + b \ln 4sin^2(\lambda_{j,N}/2) + \epsilon(j),$$

where $\lambda = 2\pi j/N$ and $j = 1, \ldots, N/2$. A regression of the logarithm of the periodogram on the logarithm of the frequency should yield a coefficient of $\hat{b} = 1 - 2H$.

VIII *Modified periodogram method :* In case of the periodogram method, most of the frequencies fall on the far right on a log-log plot. Thus they exert a strong influence on the least-squared line fitted to the periodogram. The modified periodogram method is designed to compensate this by giving a equally averaged weights to the frequencies. The frequency axis is divided into logarithmically equally spaced boxes and the periodogram values corresponding the frequencies inside the box are averaged.[14]

XI *Whittle estimator :* The Whittle estimator[15] is also based on the periodogram. Assumed that the parametric form of the spectral density is known, it is based on the minimisation of the likelihood function defined in the frequency domain. The object function is

$$Q(\theta) = \int_{-\pi}^{\pi} \frac{I(\lambda, N)}{f(\lambda, N; \theta)} d\lambda,$$

where $I(\lambda, N)$ is the periodogram and $f(\lambda, N; \theta)$ is the spectral density at frequency λ, and θ denotes the vector of unknown parameters. When dealing with FGM or ARFIMA(0,d,0), θ is the parameter H or d. If the series is assumed to be ARFIMA(p,d,q), then θ includes the unknown coefficients in the autoregressive and moving average parts.

X *Wavelet method*[16] *:* Jensen (1999) introduced a wavelet-based estimator of H using ordinary least squares regression. It consists of three steps. First, the discrete wavelet transform

[14]Due to the inconsistency of the periodogram as an estimator of the spectrum, the GPH estimator has no satisfactory asymptotic properties. Due to the violation of i.i.d assumption of the normalised periodogram, the modified periodogram estimator can not be consistent(Jensen (1999)).

[15]Beran (1994) and Fox and Taqqu (1986)

[16]For those interested in a basic introduction to wavelet see Gençay, Selç, and Whitcher (2002) and Mallat (1989).

decorrelates the time series and produces wavelet coefficients for a given scale. This performs the decomposition stage of Mallat (1989)'s pyramid algorithm. Second, we average the squared wavelet coefficients of this transformation.[17] Third, a regression of the logarithm of the average on the logarithm of the scale parameter of the transform give us the estimator of $H = (\hat{\beta} + 1)/2$:

$$\log(\hat{\sigma}_j^2(X(t)) = a + \beta \log(\lambda_j) + \epsilon_j,$$

where $\hat{\sigma}^2_{X(t)}$ denotes the averages of the squared coefficients of wavelet transformation and the scale parameter is used as $\lambda_j = 2^j, j = 2, \ldots, 8$.

2.3.1 Empirical Detection of LRD

In this section we investigate the long range dependence (LRD) of financial time series using the above ten different methods. We use the data from five different financial markets[18]: the New York Stock Exchange Composite Index (hereafter NYCI), the German share price index DAX, the Korean Composite Stock Price Index (KOSPI), the U.S. $-Deutsche Mark exchange rate (USD) and the price of gold (GOLD).

The daily returns are calculated as follows: $x(t) = 100 \cdot (\log P(t) - \log P(t-1))$. Besides the raw returns, we focus on the various power transformation of the absolute returns. This is motivated by the long-memory study of Ding, Granger, and Engle (1993). The following power transformations of the absolute returns are used : $|X(t)|^q, q = 0.25, 0.5, 0.75, 1, 1.25, 1.50, 1, 75, 2$. Each of these transformations can be considered as a measure of volatility for the five data sets. For these time series of volatility, we estimate the Hurst exponent H as the long-memory parameter. This long-memory parameter is often estimated in the frequency domain using the spectral density. Note that Ding et al. (1993) studied the long-memory using the temporal effect of autocorrelations in the time domain. Performing the Hurst exponent (H) estimates by means of the various methods, we are able to understand the behaviour of each method's estimates as by-product. All the methods have been implemented in the open source software, *R* by R Development Core Team (2006) and its package *fSeries* by Wuertz, many others, and see the SOURCE file (2006).

The results are given in the table (2.1), (2.2), (2.3), (2.4), (2.5). We can confirm that the raw returns contain only weak dependence or temporal correlation ($H \simeq 0.5$) in the second column of each table. The small positive deviation from $H = 0.5$ is in accordance with a very small positive first order autocorrelation in Ding et al. (1993). Therefore, there is a small portion of memory that could be used for prediction. This is evidence that the efficient market theory does not hold strictly.

Across all the methods, the Hurst exponent estimates obtained from the various power transformation of absolute returns differ significantly from the Hurst exponent estimate from the raw returns. In other words these methods allows to detect the long range dependence of the underlying data.

For example the data of NYCI (see the table (2.1)), the estimated value of the Hurst exponent increases, reaches its maximum and decreases for many methods as the power q increases from 0.25 to 2. This $\cap$-shaped behaviour of the estimates are shared across all the methods except for the Higuchi method (IV) and the method of aggregated absolute value (III). The underlying returns have long-memory across all the powers and the intensity of long range dependence reaches its maximum when the power of absolute returns is $q = 0.75$, 1.0, and 1.25. This result confirms

[17] Jensen (1999) explained the role of these variance of wavelet coefficient: the wavelet coefficients' variance decomposes the variance of the series across different scales. Those scales which contribute most to the data's variance are associated with the those wavelet coefficients with the largest variance. Hence, the wavelet coefficients' sample variance provides a consistent estimator of its population variance.

[18] see the subsection 4.5 for the detailed information of the data

the findings of Ding et al. (1993) that the absolute returns, $|x_t|^1$ has more long-memory than any power transformation of absolute returns. The autocorrelation function among the various power transformations $\rho_q(\Delta t) = Corr(|x_t|^q, |x_{t+\Delta t}|^q)$ has a unique maximum point when q is around 1.

We also remark that the estimates of Higuchi method (IV) are very stable and do not vary with the various power transformation of absolute returns. The Higuchi method is originally introduced to efficiently calculate the fractal dimension which is an index for describing the characteristics of the time series (see Higuchi (1988, 1990) for details). However, it seems to be not adequate for describing the non-linear scaling or multi-scaling of time series.

According to the method of aggregated absolute values (III), it is also expected that the estimated value H of the aggregated absolute values (III) increases with the power of absolute returns. Using the absolute return $|x(t)|^1$, we also get relatively large estimated value of the Hurst exponent by the method of GPH(VII) and the method of modified periodogram(VIII).

Another interesting finding is that the data of KOSPI has the highest estimate of the Hurst exponent for ($|x_t|^2$) in not all, but most of the methods.[19] As a typical emerging market in the eighties and nineties, the index of Korean stock market has more long-memory than the other stock returns. Mandelbrot (1997) argues that the random walk hypothesis or the market efficiency hypothesis is not valid any more if the asset returns exhibit long range dependence. In particular long-memory implies strong forecastibility. Thus, we can say that the Korean stock market is far from the efficient market in the investigated period. We can measure the degree of inefficiency in a financial market as the degree of long-memory (H). On the other side, as the power increases from 0.25 to 2, the Hurst exponent estimate does not vary too much within the methods, but the estimate is strongly varying between the methods (for instance KOSPI $\hat{H}_{min} = 0.675(I), \hat{H}_{max} = 1.21(III)$). The estimate $\hat{H}_{max} = 1.21(0.122)$ is obtained by the method (II) of aggregated variance with differencing. Note that the method II yields not only the H estimates larger than 1, but also two or three times the standard deviation of those of the other methods.

The returns of the exchange rate USD also exhibit strong long-range dependence as indicated by the Hurst exponent H. The power transformation of absolute returns (m = 0.75) among the others yields the highest estimates over the different methods. The Higuchi method produces a robust estimation of the fractal dimension (H = 0.976) across the different transformations. The estimated values of the exchange rate by this methods are larger than those of the stock market returns. But we get the highest estimated value of Higuchi's fractal dimension (H = 0.980 for $|x_t|^2$) for the data GOLD.

2.4 Summary

The fractal market hypothesis (FMH)[20] is based on the idea that price movements show certain patterns, i.e. self-similarity instead of a pure random walk. A sign of self-similarity is the clustering and the bursting of time series in raw returns or volatility. Long-memory is characterised by hyperbolically decaying autocorrelations: the large (small) volatility are more likely to be followed by large (small) volatility than small (large) volatility.

The result of the long-memory studies is that the long-memory effects is also found to be significant in the various power transformation of different asset returns. The price increments of the underlying processes are self-similar with each fractal dimension instead of i.i.d.

The single scaling parameter H is estimated mainly using variance and autocovariance of time series. Then, the natural question arises whether the self-similarity parameter H can fully describe the higher moments of the financial series including time-varying variances. What is to be done

[19] The methods like V, VI, X give smaller estimated value than the estimate H for NYCI and DAX.
[20] see Peters (1994) for details.

if the moments cannot be characterised appropriately by a single Hurst exponent H, i.e the series don't obey mono-scaling laws?

The answer to this question requires to analyse multi-scaling or multi-fractal behaviour whose goal is to find the multi-fractal spectrum. By the multi-fractal analysis, financial time series is well described by a non-linear function of scaling parameters. The multi-fractal analysis enables us to describe the whole dynamics of the data using a hierarchy of exponents.

In summary, self-similar processes with stationary increments are characterised by a scaling function which is linear, whereas multi-fractal processes instead are characterised by a non-linear and convex scaling function.[21]

Estimation of Long-Memory Parameter: NYCI

Method	x_t	$\lvert x_t\rvert^{0.25}$	$\lvert x_t\rvert^{0.5}$	$\lvert x_t\rvert^{0.75}$	$\lvert x_t\rvert^{1}$	$\lvert x_t\rvert^{1.25}$	$\lvert x_t\rvert^{1.5}$	$\lvert x_t\rvert^{1.75}$	$\lvert x_t\rvert^{2}$
I	0.502	0.836	0.842*	0.835	0.815	0.780	0.730	0.672	0.618
	(0.019)	(0.048)	(0.049)	(0.055)	(0.064)	(0.093)	(0.059)	(0.037)	(0.024)
II	0.586	0.726	0.796	0.798	0.749	0.798*	0.765	0.726	0.696
	(0.103)	(0.122)	(0.097)	(0.089)	(0.103)	(0.095)	(0.073)	(0.067)	(0.088)
III	0.533	0.859	0.872	0.879	0.882	0.884	0.885	0.889	0.894*
	(0.023)	(0.044)	(0.047)	(0.052)	(0.062)	(0.092)	(0.058)	(0.038)	(0.025)
IV	0.515	0.966*	0.965	0.965	0.964	0.963	0.962	0.961	0.960
	(0.030)	(0.030)	(0.030)	(0.031)	(0.031)	(0.031)	(0.031)	(0.031)	(0.033)
V	0.511	0.644	0.681	0.711	0.732	0.743*	0.741	0.731	0.723
	(0.015)	(0.066)	(0.062)	(0.054)	(0.045)	(0.038)	(0.039)	(0.050)	(0.068)
VI	0.685	0.756	0.791	0.806	0.813	0.814*	0.811	0.807	0.800
	(0.035)	(0.032)	(0.031)	(0.031)	(0.031)	(0.030)	(0.029)	(0.029)	(0.028)
VII	0.494	0.758	0.790	0.821	0.849*	0.838	0.781	0.722	0.674
	(0.000)	(0.023)	(0.024)	(0.022)	(0.021)	(0.020)	(0.017)	(0.011)	(0.008)
VIII	0.498	0.748	0.793	0.811	0.819*	0.782	0.729	0.675	0.643
	(0.034)	(0.029)	(0.031)	(0.028)	(0.025)	(0.024)	(0.021)	(0.014)	(0.009)
IX	0.591	0.673	0.698	0.721	0.745	0.767	0.772*	0.751	0.711
	(0.009)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X	0.535	0.704	0.741	0.767	0.783	0.788*	0.784	0.776	0.766
	(0.012)	(0.058)	(0.066)	(0.070)	(0.072)	(0.075)	(0.081)	(0.091)	(0.102)

Table 2.1: For Whittle we use ARFIMA(0,d,0) for the data of stock markets (NYCI, DAX, KOSPI) and ARFIMA (1,d,0) for the data USD and GOLD. For Wavelet we use the wavelets from the Dabechies compactly supported family. We mark the largest estimates in each method with $*$.

[21] Although Bouchaud, Potters, and Meyer (2000) introduce an exactly solvable model that mimics the long range volatility correlations, their model is mono-fractal by construction. It shows apparent multi-scaling, in good agreement with empirical data, as a result of very long transient effects, induced by the long range nature of the volatility correlations [Iori (2000)].

Estimation of Long-Memory Parameter: DAX

Method	x_t	$\lvert x_t\rvert^{0.25}$	$\lvert x_t\rvert^{0.5}$	$\lvert x_t\rvert^{0.75}$	$\lvert x_t\rvert^{1}$	$\lvert x_t\rvert^{1.25}$	$\lvert x_t\rvert^{1.5}$	$\lvert x_t\rvert^{1.75}$	$\lvert x_t\rvert^{2}$
I	0.534	0.811	0.818*	0.816	0.808	0.794	0.775	0.751	0.725
	(0.036)	(0.045)	(0.047)	(0.046)	(0.044)	(0.041)	(0.037)	(0.034)	(0.030)
II	0.555	0.834	0.872	0.902*	0.889	0.864	0.831	0.753	0.722
	(0.064)	(0.111)	(0.111)	(0.096)	(0.093)	(0.094)	(0.096)	(0.112)	(0.111)
III	0.563	0.822	0.836	0.844	0.849	0.855	0.860	0.866	0.873*
	(0.032)	(0.041)	(0.044)	(0.044)	(0.042)	(0.039)	(0.036)	(0.034)	(0.032)
IV	0.529	0.966*	0.965	0.964	0.963	0.961	0.960	0.959	0.958
	(0.045)	(0.031)	(0.031)	(0.030)	(0.030)	(0.030)	(0.029)	(0.030)	(0.030)
V	0.509	0.695	0.738	0.762	0.773*	0.772	0.760	0.741	0.716
	(0.013)	(0.040)	(0.033)	(0.027)	(0.023)	(0.020)	(0.019)	(0.019)	(0.020)
VI	0.594	0.735	0.828	0.854	0.862*	0.859	0.849	0.835	0.816
	(0.031)	(0.032)	(0.039)	(0.045)	(0.047)	(0.049)	(0.050)	(0.051)	(0.052)
VII	0.540	0.737	0.781	0.800	0.810*	0.804	0.777	0.743	0.706
	(0.001)	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)
VIII	0.523	0.748	0.795	0.817	0.823*	0.807	0.777	0.743	0.705
	(0.028)	(0.029)	(0.027)	(0.026)	(0.027)	(0.026)	(0.027)	(0.026)	(0.026)
IX	0.531	0.722	0.757	0.800	0.794	0.798*	0.789	0.767	0.737
	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X	0.524	0.700	0.716	0.747	0.769	0.782	0.786*	0.783	0.774
	(0.017)	(0.016)	(0.014)	(0.014)	(0.018)	(0.024)	(0.031)	(0.037)	(0.043)

Table 2.2: For Whittle we use ARFIMA(0,d,0) for the data of stock markets (NYCI, DAX, KOSPI) and ARFIMA (1,d,0) for the data USD and GOLD. For Wavelet we use the wavelets from the Dabechies compactly supported family. We mark the largest estimates in each method with $*$.

Estimation of Long-Memory Parameter: KOSPI

Method	x_t	$\lvert x_t\rvert^{0.25}$	$\lvert x_t\rvert^{0.5}$	$\lvert x_t\rvert^{0.75}$	$\lvert x_t\rvert^{1}$	$\lvert x_t\rvert^{1.25}$	$\lvert x_t\rvert^{1.5}$	$\lvert x_t\rvert^{1.75}$	$\lvert x_t\rvert^{2}$
I	0.565	0.889	0.896	0.900	0.902*	0.900	0.897	0.892	0.885
	(0.022)	(0.034)	(0.036)	(0.039)	(0.041)	(0.043)	(0.045)	(0.046)	(0.047)
II	0.590	0.875	0.949	0.994	1.010	1.021	1.021*	1.019	0.946
	(0.096)	(0.136)	(0.133)	(0.120)	(0.119)	(0.113)	(0.122)	(0.138)	(0.133)
III	0.566	0.858	0.868	0.882	0.896	0.908	0.919	0.928	0.936*
	(0.017)	(0.027)	(0.028)	(0.030)	(0.032)	(0.034)	(0.036)	(0.037)	(0.038)
IV	0.532	0.967*	0.967	0.966	0.965	0.964	0.964	0.963	0.962
	(0.038)	(0.031)	(0.030)	(0.031)	(0.032)	(0.033)	(0.029)	(0.039)	(0.044)
V	0.511	0.679	0.697	0.705	0.706*	0.702	0.693	0.681	0.666
	(0.011)	(0.053)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)	(0.051)
VI	0.523	0.690	0.704	0.710*	0.710	0.702	0.692	0.683	0.675
	(0.033)	(0.050)	(0.047)	(0.046)	(0.046)	(0.045)	(0.044)	(0.044)	(0.044)
VII	0.535	0.801	0.822	0.842	0.844	0.844*	0.838	0.831	0.820
	(0.001)	(0.025)	(0.026)	(0.027)	(0.027)	(0.026)	(0.027)	(0.027)	(0.028)
VIII	0.512	0.743	0.759	0.772	0.778	0.783*	0.781	0.774	0.762
	(0.036)	(0.032)	(0.033)	(0.032)	(0.034)	(0.032)	(0.033)	(0.032)	(0.032)
IX	0.576	0.756	0.776	0.788	0.792*	0.791	0.784	0.774	0.761
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X	0.513	0.735	0.744*	0.744	0.738	0.728	0.716	0.701	0.685
	(0.023)	(0.029)	(0.027)	(0.026)	(0.027)	(0.028)	(0.030)	(0.032)	(0.034)

Table 2.3: For Whittle we use ARFIMA(0,d,0) for the data of stock markets (NYCI, DAX, KOSPI) and ARFIMA (1,d,0) for the data USD and GOLD. For Wavelet we use the wavelets from the Dabechies compactly supported family. We mark the largest estimates in each method with $*$.

Estimation of Long-Memory Parameter: USD

Method	x_t	$\lvert x_t\rvert^{0.25}$	$\lvert x_t\rvert^{0.5}$	$\lvert x_t\rvert^{0.75}$	$\lvert x_t\rvert^{1}$	$\lvert x_t\rvert^{1.25}$	$\lvert x_t\rvert^{1.5}$	$\lvert x_t\rvert^{1.75}$	$\lvert x_t\rvert^{2}$
I	0.549	0.825*	0.825	0.817	0.806	0.791	0.773	0.752	0.729
	(0.046)	(0.022)	(0.020)	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)
II	0.606	0.828	0.868	0.870*	0.866	0.853	0.830	0.779	0.822
	(0.094)	(0.095)	(0.106)	(0.081)	(0.081)	(0.078)	(0.082)	(0.087)	(0.073)
III	0.578	0.825	0.826*	0.823	0.820	0.817	0.814	0.812	0.811
	(0.041)	(0.029)	(0.025)	(0.022)	(0.020)	(0.017)	(0.016)	(0.015)	(0.016)
IV	0.546	0.967*	0.967	0.967	0.967	0.967	0.967	0.967	0.967
	(0.032)	(0.031)	(0.032)	(0.032)	(0.032)	(0.032)	(0.033)	(0.033)	(0.033)
V	0.502	0.679	0.704	0.711*	0.709	0.699	0.683	0.662	0.638
	(0.018)	(0.036)	(0.033)	(0.030)	(0.027)	(0.026)	(0.024)	(0.024)	(0.023)
VI	0.653	0.652	0.675	0.684	0.688	0.692	0.693*	0.690	0.683
	(0.034)	(0.064)	(0.060)	(0.056)	(0.051)	(0.045)	(0.039)	(0.033)	(0.028)
VII	0.545	0.769	0.787	0.798*	0.792	0.762	0.739	0.712	0.691
	(0.027)	(0.027)	(0.026)	(0.028)	(0.028)	(0.027)	(0.026)	(0.027)	(0.028)
VIII	0.501	0.729	0.749	0.756*	0.745	0.726	0.700	0.670	0.643
	(0.036)	(0.032)	(0.032)	(0.033)	(0.032)	(0.034)	(0.031)	(0.034)	(0.032)
IX	0.527	0.704	0.721	0.725*	0.722	0.713	0.699	0.681	0.659
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X	0.521	0.759	0.788	0.794*	0.789	0.775	0.752	0.724	0.691
	(0.008)	(0.033)	(0.033)	(0.033)	(0.032)	(0.031)	(0.030)	(0.031)	(0.034)

Table 2.4: For Whittle we use ARFIMA(0,d,0) for the data of stock markets (NYCI, DAX, KOSPI) and ARFIMA (1,d,0) for the data USD and GOLD. For Wavelet we use the wavelets from the Dabechies compactly supported family. We mark the largest estimates in each method with $*$.

Estimation of Long-Memory Parameter: GOLD

Method	x_t	$\lvert x_t\rvert^{0.25}$	$\lvert x_t\rvert^{0.5}$	$\lvert x_t\rvert^{0.75}$	$\lvert x_t\rvert^{1}$	$\lvert x_t\rvert^{1.25}$	$\lvert x_t\rvert^{1.5}$	$\lvert x_t\rvert^{1.75}$	$\lvert x_t\rvert^{2}$
I	0.550	0.925	0.926*	0.918	0.906	0.889	0.871	0.850	0.830
	(0.031)	(0.025)	(0.021)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)
II	0.688	0.856	0.903	0.913	0.916	0.925*	0.903	0.812	1.030
	(0.086)	(0.098)	(0.101)	(0.135)	(0.111)	(0.139)	(0.115)	(0.135)	(0.155)
III	0.583	0.935	0.941	0.944*	0.943	0.941	0.939	0.936	0.934
	(0.028)	(0.033)	(0.032)	(0.031)	(0.030)	(0.028)	(0.026)	(0.025)	(0.023)
IV	0.527	0.967	0.968	0.970	0.972	0.974	0.976	0.978	0.980*
	(0.035)	(0.030)	(0.031)	(0.032)	(0.032)	(0.033)	(0.033)	(0.034)	(0.035)
V	0.496	0.691	0.729	0.745	0.752	0.753*	0.749	0.741	0.730
	(0.033)	(0.037)	(0.036)	(0.034)	(0.032)	(0.030)	(0.028)	(0.026)	(0.025)
VI	0.706	0.792	0.818*	0.817	0.807	0.780	0.768	0.748	0.727
	(0.042)	(0.042)	(0.040)	(0.037)	(0.035)	(0.034)	(0.033)	(0.033)	(0.034)
VII	0.541	0.815	0.851	0.864	0.885	0.887*	0.880	0.884	0.875
	(0.002)	(0.029)	(0.027)	(0.029)	(0.032)	(0.030)	(0.030)	(0.030)	(0.029)
VIII	0.521	0.771	0.806	0.807	0.816	0.820	0.821	0.823*	0.818
	(0.043)	(0.035)	(0.033)	(0.034)	(0.034)	(0.032)	(0.034)	(0.033)	(0.032)
IX	0.519	0.751	0.772	0.780	0.784*	0.783	0.779	0.772	0.761
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
X	0.521	0.738	0.774	0.788	0.794	0.794*	0.790	0.782	0.771
	(0.040)	(0.041)	(0.045)	(0.048)	(0.051)	(0.054)	(0.056)	(0.057)	(0.057)

Table 2.5: For Whittle we use ARFIMA(0,d,0) for the data of stock markets (NYCI, DAX, KOSPI) and ARFIMA (1,d,0) for the data USD and GOLD. For Wavelet we use the wavelets from the Dabechies compactly supported family. We mark the largest estimates in each method with $*$.

Chapter 3

The Multi-fractal Model of Asset Returns

In the previous chapter, the fractal analysis is limited to study a single scaling law for financial time series, see Peters (1994). The long-term behaviour of financial time series is too complex to be described just by means of a single fractal dimension. Thus, we will turn our point of view from fractal to multi-fractal behaviour. The fractal formalism corresponds to the invariant probability distribution, while the multi-fractal formalism studies a non-linear relation between the invariant probability densities on a generic support [Chhabra and Jensen (1989)]. In the framework of multi-fractals, the moments of distributions can be characterised by a non-linear exponent function.

As reported in Calvet, Fisher, and Mandelbrot (1997) and Lux (2000), there is some evidence of multi-fractality in financial time series. What is then the source of such multi-scaling in financial markets? According to Müller, Dacorogna, Dave, Pictet, Olsen, and Ward (1995), heterogeneous traders are responsible on the multi-fractality of the market who act and react with different strategies in various time horizons of investment. On a long time scale, the volatility of the market is affected by information shocks on macroeconomic performance. On a short time scale, it may be mainly changed by transient information shocks. One can model such information shocks in terms of a multiplicative cascade, where the information at the different levels of the cascade reflects the range of coherent influences. The multi-scaling comes from such hierarchical structure of a cascade.

At first, we will show the generating process of a multi-fractal in terms of the trinomial multiplicative cascade, and then investigate the dynamics of a multi-fractal in the context of local and global moments. Furthermore, we look into the multi-scaling using the multi-fractal formalism[1], and illustrate the scaling function of a trinomial multi-fractal.

3.1 Multiplicative cascade

Here, we will construct the multi-fractal by a multiplicative process, i.e. multiplicative cascade on a generic support (be it fractal or not). In physics, this phenomenological cascade models are employed to study turbulence: a large structure multiplicatively modulates the various information on short time scales [Schertzer and Lovejoy (1987)]. In finance, we use to modulate the time-varying

[1]See Murthy, Kehr, and Giacometti (1996) for simple examples. In the physics literature, the multi-fractal formalism has received much attention as a popular framework to describe and to analyse multi-fractal processes which cover and connect both local and global scaling in terms of various sample moments [Chainais, Riede, and Abry (2005)].

volatility by means of the multiplication of volatility components of which lifespans have different scales. We generally need multiple exponents for adequate characterisation of the generated multi-fractal.

Construction of multi-fractal process

Mandelbrot (1974) introduced the canonical binomial cascade as the prototype of multi-fractal random measures. Let's here consider the trinomial multi-fractal cascade in which the measure M_t is given by

$$M_t = \prod_{k=1}^{\bar{k}} m_{t,k} \,. \tag{3.1}$$

Here, $m_{t,k}$ denotes the components of the measure, and takes one of the trinomial masses m_0, m_1 and m_2 which is constrained by $m_0 + m_1 + m_2 = 1$ because of the normalisation. Note that for each cascade the total measure $M = M_1 + \ldots + M_t$ is conserved as one. k refers to the number of cascade steps. $t = 1, 2, \ldots$ is the discrete count of the subintervals of the unit interval $[0, 1]$ with size $(1/3)^{\bar{k}}$.

We will more explicitly describe the process of trinomial multi-fractal construction on the basis of $m_0 = 0.1$, $m_1 = 0.3$ and $m_2 = 0.6$. And we restrict the whole time region to $[0, 1]$. In figure (3.1), we illustratively present the process of construction.

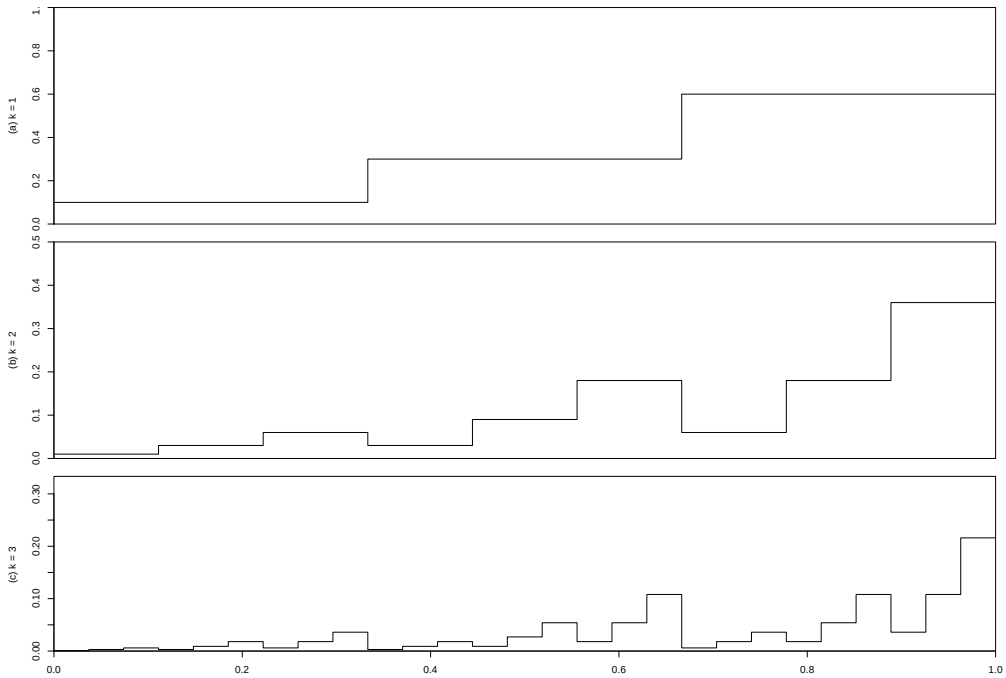

Figure 3.1: Combinatorial construction of trinomial multi-fractal based on $m_0 = 0.1$, $m_1 = 0.3$ and $m_2 = 0.6$ for $\bar{k} = 1, 2, 3$. In each cascade, the sum of discrete measures M_t equals to 1, e.g. it is exactly conservative measure.

As shown in the figure, there are, at the first cascade, three time-intervals of $[0, 1/3]$, $[1/3, 2/3]$ and $[2/3, 1]$. In each interval, M_t is distributed by the uniformly spreading masses of m_0, m_1 and m_2, respectively. At the next cascade, the interval $[0, 1/3]$ is again split into $[0, 1/9]$, $[1/9, 2/9]$ and $[2/9, 3/9]$. In each interval, the measures M_t are, due to equation (3.1), given as $m_0 m_0$, $m_0 m_1$ and $m_0 m_2$, respectively. For the other two intervals of $[1/3, 2/3]$ and $[2/3, 1]$ which are also divided into three subintervals at $\bar{k} = 2$, we determine M_t by the same way so that

$$M_{[3/9,4/9]} = m_1 m_0 \quad M_{[4/9,5/9]} = m_1 m_1 \quad M_{[5/9,6/9]} = m_1 m_2$$
$$M_{[6/9,7/9]} = m_2 m_0 \quad M_{[7/9,8/9]} = m_2 m_1 \quad M_{[8/9,1]} = m_2 m_2$$

For the match with the figure (3.1), we presented measures in an interval, instead of M_t. The cascade that is built by the above procedure, is called as *multiplicative cascade* and we refer to the generated M_t to as *multi-fractal measure*.

In an arbitrary *triadic* interval of $[t, t+3^{-\bar{k}}]$, the measures are specified by $M_{[t,t+3^{-\bar{k}}]} = m_0^a m_1^b m_2^c$ where a, b and c are the frequencies of m_0, m_1 and m_2, respectively and $a+b+c = \bar{k}$. This means that the construction creates large and increasing heterogeneity in the allocation of mass. As a result, the trinomial multi-fractal, like many other multi-fractals, has a *continuous* but *singular* probability measure, and is therefore not differentiable.

In the figure, the multi-fractal measures build a generic clustering, and then have wide range of scales. Hence, we will define them as volatility in the multi-fractal model of asset return and the Markov switching multi-fractal model[2] which we are going to treat in the next chapter.

Extension of the multi-fractal measure

The construction of trinomial multiplicative measure can be extended in several ways. First, at each cascade, the time intervals can be divided not in three but two or four spans with equal size. This defines the class of *multinomial multi-fractals*: binomial or quadronomial multi-fractal, respectively. The multi-fractal can be also generated on the support of a Cantor set and then we get a Cantor multi-fractal which also belongs to the class of multinomial multi-fractal.

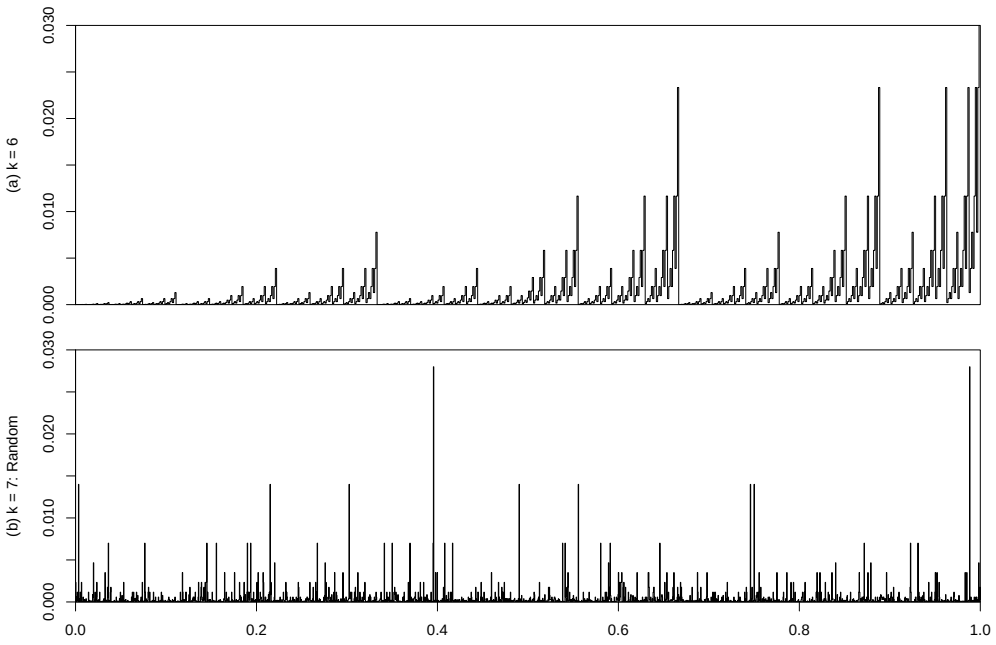

Figure 3.2: (a) Trinomial multi-fractal from the combinatorial cascade $\bar{k} = 6$ (upper). The trinomial measures are obtained according to equation (3.1) using $m_0 = 0.1$, $m_1 = 0.3$ and $m_2 = 0.6$. (b) Trinomial multi-fractal from the random cascade. We randomly allocate the obtained measures from the combinatorial cascade at $\bar{k} = 7$. We zoom out the resulting trinomial multi-fractal measure in the y-axes to see the self-similarity clearly.

[2]See, the original paper of Mandelbrot et al. (1997) that models the log-price as multi-fractal process.

Second, the allocation of mass in subintervals can be randomised. Through the reshuffling of these measures, we obtain a *random multi-fractal* instead of deterministic multi-fractal. Even though the measures are reshuffled, the multi-fractality does not vanish and the non-linear scaling function remains. Figure (3.2) shows one example of this random multi-fractal at $\bar{k} = 7$ (bottom plot) compared with the combinatorial multi-fractal at $\bar{k} = 6$ (upper plot). The mass is randomly allocated which is determined according to equation (3.1) using $m_0 = 0.1$, $m_1 = 0.3$ and $m_2 = 0.6$. Third, we can randomly draw positive masses m_i from a log-normal or a gamma distribution in each stage. It is called the *lognormal* or the *gamma* multi-fractal.[3]

Scaling function and multi-fractal dimension

Let's investigate some properties of the multi-fractal measures to understand a distribution of multiplied quantities on a geometric support, see Muzy, Bacry, and Arneodo (1994), Feder (1988) and Evertz and Mandelbrot (1992) for further details. We image that a multi-fractal is covered with N number of grid boxes of width ϵ. The measure $M_j(\epsilon)$ to find some parts of fractal within jth box scales as

$$M_j(\epsilon) \sim \epsilon^{\alpha_j}, \tag{3.2}$$

where α_j is the coarse Lipschitz-Hölder exponent or singularity strength, characterising scaling in the jth box [Chhabra, Meneveau, Jensen, and Sreenivasan (1989)]. Similar α_j values can be found at different positions within the multi-fractal distribution. The number of boxes $N(\alpha)$ where M_j has singularity strength α is found to scale as

$$N(\alpha) \sim \epsilon^{-f(\alpha)}, \tag{3.3}$$

where $f(\alpha)$ is defined as the fractal dimension of the set of boxes with singularity α. Sometimes, one refers to $f(\alpha)$ as the singularity spectrum. We use this singularity or multi-fractal spectrum to describe the local behaviour of a multi-fractal process. The exponent α can take on values from the intervals $[\alpha_{-\infty}, \alpha_{\infty}]$.

A multi-fractal set can be also characterised on the basis of the generalised dimensions D_q of the qth moment orders of the distribution, where D_q is defined as

$$D_q = \frac{1}{(q-1)} \lim_{\epsilon \to 0} \frac{S(q,\epsilon)}{\log \epsilon} \tag{3.4}$$

Here, $S(q,\epsilon)$ is the partition function[4] that is given by use

$$S(q,\epsilon) = \sum_{j=1}^{N} |M_j(\epsilon)|^q \tag{3.6}$$

The generalised dimension D_q is monotone decreasing function for all real q values within $[-\infty, \infty]$. When $q < 0$, $S(q,\epsilon)$ emphasises regions in the distribution with less concentration of a measure, where the opposite is true for $q > 0$.

[3] See Mandelbrot (1989) for details.

[4] This partition function can be used in more general fluctuation function [Davis, Marshak, Wiscombe, and Cahalan (1994)].

$$\left\{ \frac{1}{N_\epsilon} S(q,\epsilon) \right\}^{1/q} \sim \epsilon^{h(q)}, \tag{3.5}$$

where $h(q) = [\tau(q) + 1]/q$. The exponent $h(q)$ may depend on q in general. Note that $h(1)$ is identical to the Hurst exponent.

The partition function scales as

$$S(q,\epsilon) \sim \epsilon^{\tau(q)}, \tag{3.7}$$

where the scaling exponent $\tau(q)$ of the qth order is defined

$$\tau(q) = (q-1)D_q \tag{3.8}$$

The scaling function $\tau(q)$ describes the global change of the behaviour of $S(q,\epsilon)$. The connection between power exponents $f(\alpha)$ in equation (3.3) and $\tau(q)$ is made via the Legendre transformation:

$$\begin{aligned} f(\alpha(q)) &= q\alpha(q) - \tau(q) \\ \alpha(q) &= \frac{d\tau(q)}{dq} \end{aligned} \tag{3.9}$$

The multi-fractal spectrum $f(\alpha)$ and the generalised dimension D_q contain the same information, both characterising interwoven ensembles of fractals of dimension $f(\alpha_j)$. In each of the jth fractals, the measure M_j scales with the Lipschitz-Hölder exponent α_j.

The generalised dimensions for $q = 0$, $q = 1$ and $q = 2$ are known as the capacity dimension D_0, the information dimension D_1, and the correlation dimension D_2. The capacity dimension D_0 is independent of q, and provides global (or average) information of the system [Muzy et al. (1994)]. D_1 is related to the information. For a measure $M \in [0,1]$, the value of D_1 is in the range of $0 < D_1 < 1$. D_1 value close to one present a system uniformly distributed throughout all scales, whereas D_1 close to zero reflects a subset of scale in which the irregularities are concentrated. Since the information dimension is defined in terms of the relative frequency of visitation of a typical stochastic process or trajectory, it utilises information about the time behaviour of the dynamical system [Parker and Chua (1989), see equation (3.14)]. D_2 is associated to the correlation function[5], and computes the correlation of measures contained in intervals of size ϵ. The relationship between D_0, D_1 and D_2 is

$$D_2 \leq D_1 \leq D_0, \tag{3.10}$$

where the equality $D_0 = D_1 = D_2$ occurs only if the fractal is statistically or exact self similar and homogeneous.

Exercise: Analysis of Trinomial Multi-fractal Measures

As a sample sum over states measured on the trinomial components, the partition function[6] of the trinomial multi-fractal measure can be written for any order $\bar{k}$ as follows:

$$\begin{aligned} S(q,\epsilon) &= \sum_{j}^{N_\epsilon} M_j^q \\ &= \sum_{s_1,s_2,s_3}^{N_\epsilon} N_s m_0^{qs_1} m_1^{qs_2} m_2^{qs_3}, \quad N_s = \frac{N_\epsilon!}{s_1!s_2!s_3!} \\ &= \left(m_0^q + m_1^q + m_2^q\right)^{\bar{k}}, \end{aligned} \tag{3.11}$$

[5] Grassberger and Procaccia (1983) developed the concept and the method of estimation. The correlation dimension can be used to distinguish deterministic chaos from random noise.

[6] In statistical mechanics, the statistical properties of a thermodynamic system can be described by the partition function [Landau and Lifshitz (1996)].

where $\epsilon = (1/3)^{\bar{k}}$, i.e. $\bar{k} = -\log_3 \epsilon$. Substituting the scaling function

$$\tau(q) = -\log_3(m_0^q + m_1^q + m_2^q) \tag{3.12}$$

resulting from the equations (3.6), (3.1) into the equation (3.8), we obtain the generalised dimension[7] or so-called Renyi dimension D_q of the trinomial multi-fractal

$$D_q = \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{\log S(q,\epsilon)}{\log \epsilon} = -\frac{\log_3(m_0^q + m_1^q + m_2^q)}{q-1} \tag{3.13}$$

for $q \neq 1$. Like other multi-fractals, the spectrum of trinomial multi-fractal D_q decreases from $D_{-\infty} = -\log_3\{min\,(m_0,\, m_1,\, m_2)\}$ to $D_{\infty} = -\log_3\{max\,(m_0,\, m_1,\, m_2)\}$.

For $q = 0$, we obtain $D_0 = 1$ for any measures, while the information dimension or entropy dimension D_1 is given by $D_1 = -(m_0 \log m_0 + m_1 \log m_1 + m_2 \log m_2)/\log 3$ for a trinomial multi-fractal process.[8] The information dimension weights three dimensional subsets according to the frequency of visitation of a typical process [Parker and Chua (1989)]. And the correlation dimension D_2 is here represented by $-\log_3(m_0^2 + m_1^2 + m_2^2)$. The intermittency or sparseness in a stochastic process is characterised by the codimension $C_1 = 1 - D_1$ [Marshak, Davis, Cahalan, and Wiscombe (1994)]. Due to $0 < D_1 < 1$ for a measure $M \in [0, 1]$, the C_1 has positive value. If a codimension is zero and $m_0 = m_1 = m_2$, the process is just a standard Brownian motion.

In figure (3.3), we illustrate the generalised dimension D_q of equation (3.13) versus q, where I) $(m_0,\, m_1,\, m_2) = (0.1,\, 0.3,\, 0.6)$ (left) and II) $(m_0,\, m_1,\, m_2) = (0.36,\, 0.3,\, 0.34)$ (right) are used. Note that the dotted lines $D_q = 1$ owe to $m_0 = m_1 = m_2 = 1/3$. That means, the stochastic process has the mono fractal dimension in this case. The information dimension D_1, the corresponding codimension C_1 and the correlation dimension D_2 evaluated using above set of m_0, m_1 and m_2 result in

$$\begin{aligned} &\text{I)}\; D_1 = 0.8173454 \qquad C_1 = 0.1826446 \qquad D_2 = 0.706827 \\ &\text{II)}\; D_1 = 0.9974232 \qquad C_1 = 0.0025768; \qquad D_2 = 0.9949169 \end{aligned}$$

As results indicate, the second set of m_0, m_1 and m_2 in which the values are only slightly different, lead to lower codimension than that from the first set. In other words, the larger a codimension the sparser the multi-fractal process and vice versa. The multi-fractal process constructed by the second set, varies smoothly around the average value with a very small standard deviation. The multi-fractal measure from almost equal masses has very strong singularity within some specific intervals. Otherwise, the multi-fractal process has very small measures. Thus, the codimension can be used as a barometer of extreme events.[9]

[7] Muzy et al. (1994) remark that the concept of the generalised dimensions is used to characterise an ergodic measure associated with a given dynamical system.

[8] See Turcotte (1997) for the modified equation of information dimension. The information dimension can be derived using Shannon's entropy [Shannon (1948)]:

$$\begin{aligned} D_1 &= \lim_{q\to 1} \frac{1}{q-1} \lim_{\epsilon \to 0} \frac{\log S(q,\epsilon)}{\log \epsilon} = \lim_{\epsilon \to 0} \lim_{q \to 1} \frac{\log \sum_i P_i^q}{\log \epsilon (q-1)} = \lim_{\epsilon \to 0} \lim_{q \to 1} \frac{\frac{d}{dq} \log \sum_i P_i^q}{\frac{d}{dq} \log \epsilon (q-1)} \\ &= \lim_{\epsilon \to 0} \lim_{q \to 1} \frac{\frac{\sum_i \log(P_i) P_i^q}{\sum_i P_i^q}}{\log \epsilon} = \lim_{\epsilon \to 0} \frac{-\sum_i \log(P_i) P_i}{-\log \epsilon} = \lim_{\epsilon \to 0} \frac{H(\epsilon)}{\log(1/\epsilon)} \\ &= -\frac{m_0 \log m_0 + m_1 \log m_1 + m_2 \log m_2}{\log 3}, \end{aligned} \tag{3.14}$$

where $\sum_i P_i = 1$ and the entropy is defined by $H(\epsilon) = -\sum_i \log(Pi) P_i$.

[9] Of particular interest would be an investigation of these codimension with respect to the fat tail behaviour

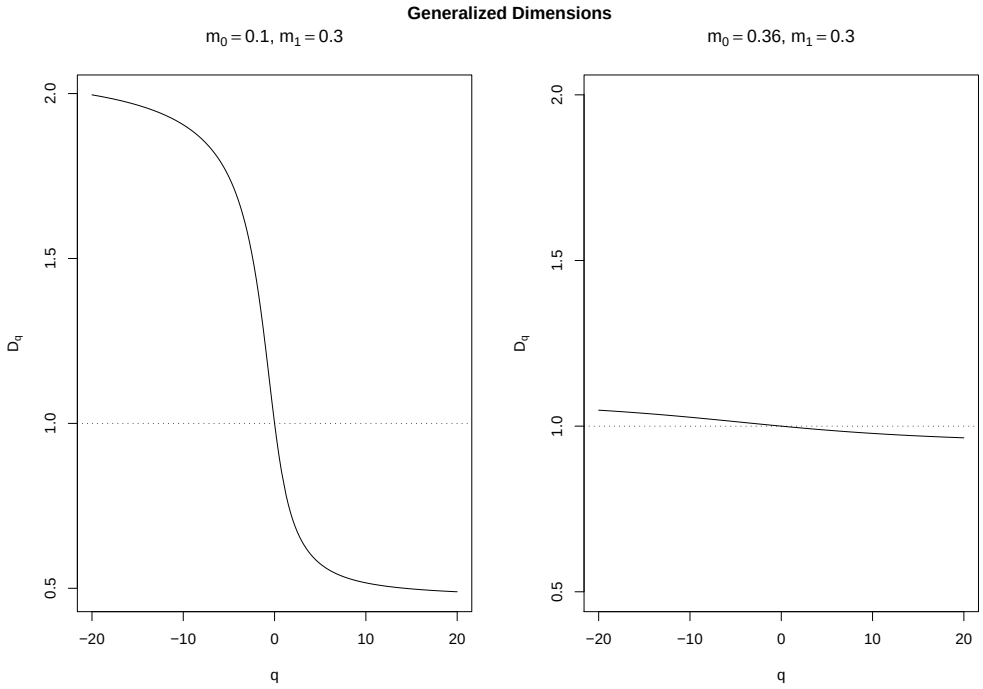

Figure 3.3: The generalised dimension D_q using $(m_0, m_1, m_2) = (0.1, 0.3, 0.6)$ (left plot) as well as $(m_0, m_1, m_2) = (0.36, 0.3, 0.34)$ (right plot). The dotted lines represent the $D_q = 1$ using $m_0 = m_1 = m_2 = 1/3$

In the right plot that is generated by almost equal m_0, m_1 and m_2, the dimension function D_q seems all but linear, and change little even though q varies over the wide range, i.e. $-20 < q < 20$. For the trinomial cascade from the first non-homogeneous masses, the spectrum D_q is a non-linear, but rapidly falling function, see the left plot. The non-linear changing of D_q is closely related to fat-tails and leptokurtic distribution, whereas the effect of clustering results mainly from the fractal structure assigned the measures. As $q \to \infty$, the generalised dimension asymptotically goes closer to $D_\infty = -\log_3 0.6 \approx 0.4649735$.

Figure (3.4) displays the spectrum $\tau(q)$ for trinomial (left plot) and binomial cascade (right plot). Following $D_q = 1$ for $m_0 = m_1 = m_2$, see figure (3.3), the $\tau(q)$ is reduced to a linear function of $q - 1$. For a heterogeneous stochastic measure of $m_0 \neq m_1 \neq m_2$, the spectrum $\tau(q)$ has the concavity of $\tau^{''}(q) < 0$ that depends only on the difference between $max\,(m_0, m_1, m_2)$ and $min\,(m_0, m_1, m_2)$. The $E[M_t^q]$ is then generally very different from $E[M_t]^q$ since even a small difference in the exponents causes a large change of moments due to the small-time interval of $(1/3)^{\bar{k}}$ [Marshak et al. (1994)].

From equation (3.9) the coarse Lipschitz-Hölder exponent α_q is followed as

$$\alpha(q) = -\frac{1}{\log 3}\left(\frac{m_0^q \log m_0 + m_1^q \log m_1 + m_2^q \log m_1}{m_0^q + m_1^q + m_2^q}\right) \tag{3.15}$$

The obtained $\alpha(q)$ and the scaling function in equation (3.12) lead to the multi-fractal spectrum for the trinomial measure in the form of

of financial return distribution. Recently, Muzy, Bacry, and Kozhemyak (2006) find that there is no discrepancy between the coefficient of intermittency and the estimated power-law behaviour of probability density function.

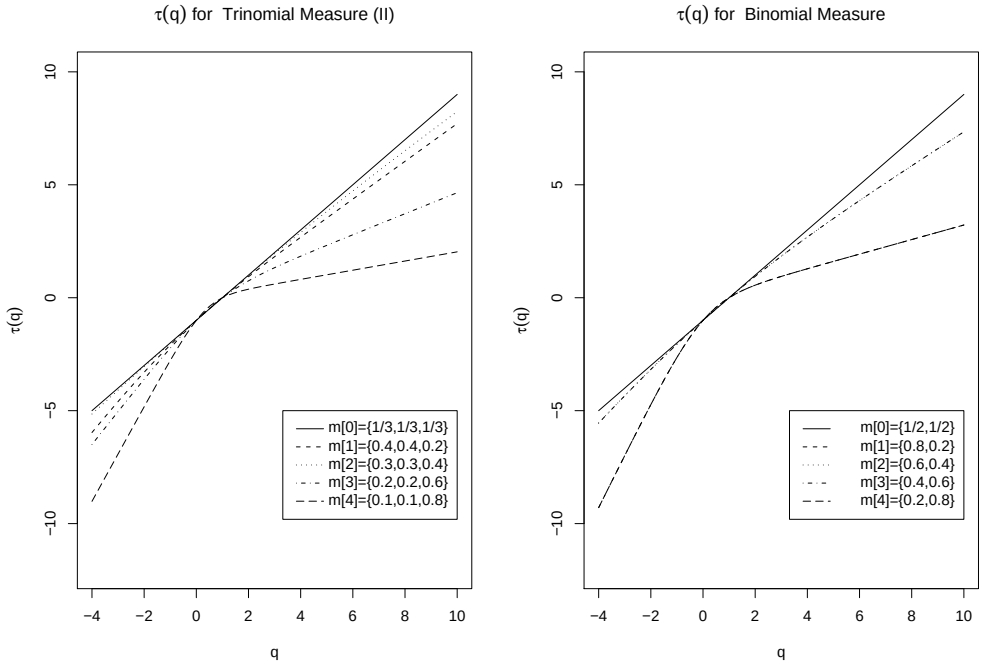

Figure 3.4: $\tau(q)$ for binomial and trinomial multi-fractal processes: The solid line represents the combination of homogeneous measure: $m[0] = \{m_0, m_1\} = \{0.5, 0.5\}$ for binomial multi-fractal in the right panel and $m[0] = \{m_0, m_1, m_2\} = \{1/3, 1/3, 1/3\}$for trinomial multi-fractal in the left panel. The scaling function $\tau(q)$ is more concave when the difference between multi-fractal parameters m_i is large as like the combination $m[4]$ in both panel

$$f\big(\alpha(q)\big) = -\frac{q}{\log 3}\left(\frac{m_0^q \log m_0 + m_1^q \log m_1 + m_2^q \log m_1}{m_0^q + m_1^q + m_2^q}\right) + \frac{\log(m_0^q + m_1^q + m_2^q)}{\log 3} \tag{3.16}$$

The multi-fractal spectrum $f(\alpha(q))$ describes the convergence rate of the probability for trinomial distribution, see Harte (2001) for the detailed discussion. The theoretical multi-fractal spectra of binomial and trinomial multi-fractal are illustrated in figure (3.5). The spectrum is bounded. The minimum and maximum exponent of binomial multi-fractal is given by $\alpha_{min} = -\log_2 0.6 = 0.737$ and $\alpha_{max} = -\log_2 0.4 = 1.322$, respectively. For the trinomial multi-fractal, $\alpha_{min} = -\log_3 0.6 = 0.455$ and $\alpha_{max} = -\log_3 0.1 = 2.096$. Note that the fractal dimension of these measures corresponds to the maximum of the multi-fractal spectrum $f(\alpha)$. It is obvious for two multi-fractal spectra that $D_1 < D_0$ when the multi-fractal measures are not equal. The singularity exponents of trinomial multi-fractal process is stronger then those of binomial multi-fractal process. If $q = \infty$, we get $\alpha_{min} = D_\infty = 0.455$ that is the strongest singularity exponents.

3.2 MMAR

Here, we explicitly look into the structure of the multi-fractal model of asset returns (hereafter, MMAR) that has been introduced by Mandelbrot et al. (1997), Calvet et al. (1997), Lux (2000) and Calvet and Fisher (2002). In the framework of MMAR, the return of a financial asset is modelled as a compound process, i.e. the subordination of standard Brownian motion to a multi-fractal process. The asset price results from a multi-fractal process with long memory and fat-tails. The

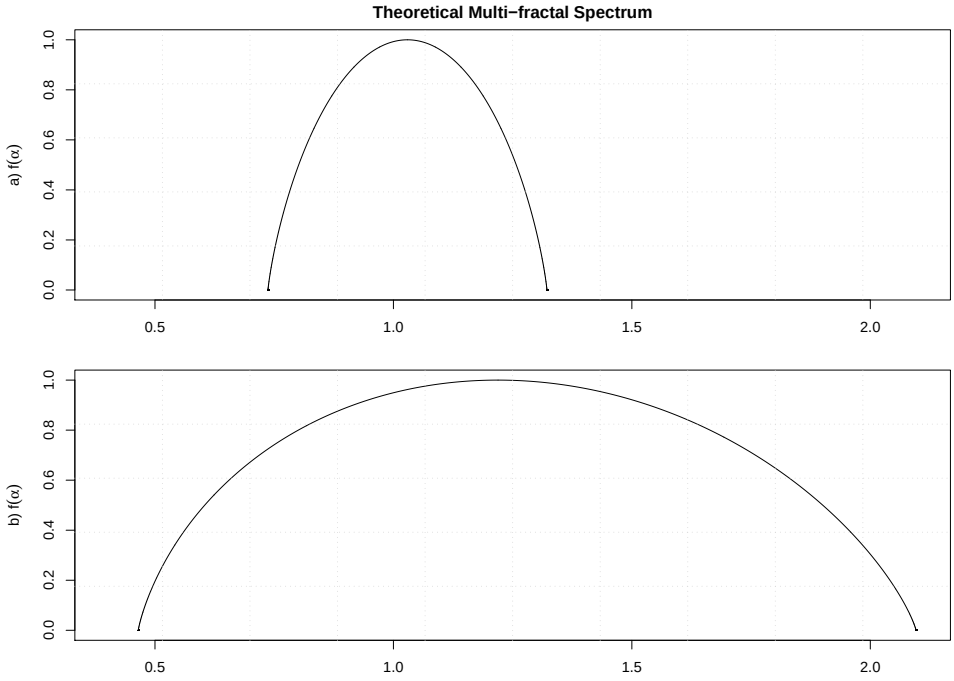

Figure 3.5: **Theoretical multi-fractal Spectrum** The multi-fractal spectra are obtained using a) the binomial mass $m_0 = 0.6$ and b) the trinomial masses of $m_0 = 0.1$, $m_1 = 0.3$, respectively. They are calculated in a region of $-100 \leq q \leq 100$ with a 0.01 step increments.

Brownian motion guarantees that the asset returns is a martingale.

3.2.1 Structure of the Model

Let us introduce the notation X_t defined as

$$X_t \equiv \ln P_t - \ln P_0, \tag{3.17}$$

where P_t and P_0 denote the financial asset price at any time t and $t_0 = 0$, respectively. The time interval is bounded as $[0, T]$. It is assumed that [Mandelbrot et al. (1997)] that

- X_t is a compound process in the form of

$$X_t = B_H\left[\vartheta(t)\right], \tag{3.18}$$

  where $B_H(t)$ is a fractional Brownian motion with self-similarity index H.[10] And $\vartheta(t)$ denotes the stochastic trading time.

- A stochastic trading time of $\vartheta(t)$ is calculated by the cumulative density function of a random multi-fractal measure defined on $[0, T]$ [Mandelbrot et al. (1997)]. In other words, the $\vartheta(t)$ is a multi-fractal process with continuous, non-decreasing paths and stationary increments.

- $\{B_H(t)\}$ and $\{\vartheta(t)\}$ are independent.

[10]The fractional Brownian motion in random trading time is a special form of subordination. Note that subordination requires independent increments before its natural filtration. The trading time $\vartheta(t)$ fulfils this condition as a cumulative density function of random multi-fractal measure. See Clark (1973) for different kinds of subordinations that have been extensively used in the financial literature. Note that in general the distributions of subordinated stochastic processes do not possess scaling properties.

The process X_t is then multi-fractal and it satisfies[11] that

$$E[|X_t|^q] = E[\vartheta(t)^{Hq}]E[|B_H(1)|^q]. \tag{3.20}$$

The trading time that serves as dynamics of volatility in MMAR, is transmitted to the process X_t through compounding. That is, this trading time is multiplied by a fractional Brownian motion.

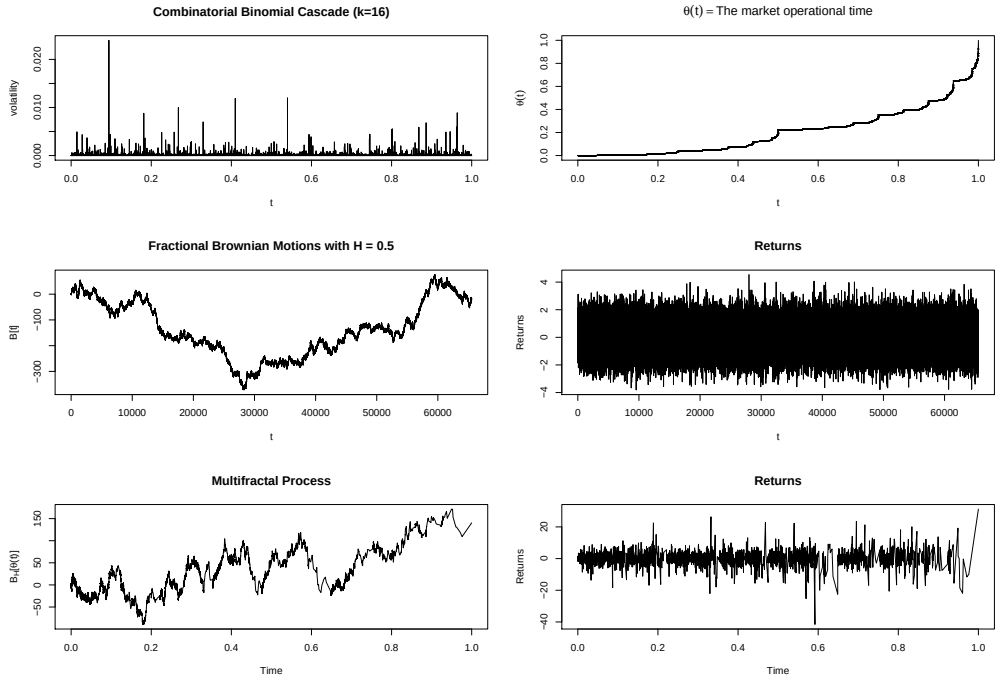

Figure 3.6: Construction Principle of MMAR After generating binomial multi-fractal process with $\bar{k} = 16$ and the binomial parameter $m_0 = 0.6$, we build the cumulative distribution of this multi-fractal that is used as the operation time $\vartheta(t)$ of financial markets. A fractional Brownian motion $B_{0.5}(t)$ is generated in the left-middle panel and its difference are build just to show a simulated return. In the last step, we compound the operational time by the fractional Brownian motion: $B_H[\vartheta(t)]$.

Figure (3.6) shows the generating procedure of a multi-fractal process. In the upper panels, we produce a randomised binomial multi-fractal process with the binomial parameter $m_0 = 0.6$ and the maximum number of cascade $\bar{k} = 16$ (left panel).[12] The measure M_t in equation (3.1) is used as a discrete approximation to the quadratic variation of a multi-fractal process. We get the trading time $\vartheta(t)$ from aggregating the increments of M_t. A simulated path of this trading time is shown as the cumulative density function in the right panel. The smoothed increase of the cumulative density function reflects the clustered volatility, while the non-smoothed increases or vertical increases of this function mean abrupt changes of trading time that reflect occasional bursts of volatility. Fat-tails in the compound process stem from these abrupt changes of trading

[11] Calvet and Fisher (2002).

Because of the independence between the trading time process and the self-similarity process, the conditional expectation yields

$$E[|X_t|^q] = E[|B_H[u]||^q\,|\vartheta(t) = u] = \vartheta(t)^{Hq}E\left[|B_H[1]|^q\right] \tag{3.19}$$

The process X_t fulfils the multi-scaling law of equation (3.22) with $\tau(q) \equiv \tau_\vartheta(Hq)$ and $c_q \equiv c_\vartheta(Hq)E[|B_H[1]|^q]$.

[12] The principle of construction is the same as the trinomial multiplicative cascade. Now, a unit interval $[0, 1]$ is subdivided into two equal segments and two binomial masses m_0 and m_1 are used that are constrained as $m_0 + m_1 = 1$. Each of these two intervals is subdivided as the same manner. At $\bar{k} = 1$, the former interval receives m_0 and the next interval m_1. At $\bar{k} = 2$, The measures of m_0m_0, m_0m_1, m_1m_0 and m_1m_1 are allocated to each interval, respectively. And so on.

time. The trading time is not only highly variable, but also contains long-memory at the same time.

In the middle panel, the left plot presents a sample path of standard Brownian motion of $B_{0.5}(t)$ and the right plot exhibits its difference that is equal to white noise. In the bottom panel, a multi-fractal process of price is simulated using $P_t = \exp(B_H[\vartheta(t)])$. For the property of martingale in MMAR, we multiply the trading time $\sqrt{M_t}$ by a standard Gaussian distribution. This price process contains small crashes and up-crashes, some persistent growth and rapid declines. Its return process in the right panel is strongly similar to that of real financial time series. We observe various degrees of temporal heterogeneity at all time scales and non-Gaussian behaviour of multi-fractal return process that have tremendous implications for behaviour of risk-averse investors and the pricing of financial derivatives [Mandelbrot et al. (1997)].

3.2.2 Properties of MMAR

We let the process $X_{t,\Delta t}$ denote as

$$X_{t,\Delta t} \equiv X_{t+\Delta t} - X_t \tag{3.21}$$

Following the definition of multi-fractality in a stochastic process [Mandelbrot et al. (1997)], a stationary increment $X_{t,\Delta t}$ of multi-fractal process satisfies a scaling law of

$$E\left[|X_{t,\Delta t}|^q\right] = c(q)\Delta t^{\tau(q)+1} \tag{3.22}$$

for all $t \in \mathcal{T}$ and $q \in \mathcal{Q}$. The scaling function $\tau(q)$ contains all information about the rate of growth of the moments, and takes into account the effects of time horizons on the absolute moment of order q. The $c(q)$ is the time-independent prefactor. The multi-scaling function (3.22) imposes that the moments of multi-fractal process are finite for all t. In particular, the qth moment of x_t exists if and only if the trading time process $\vartheta(t)$ has a moment of order Hq.

The equation (3.22) also holds for mono-fractal processes whose scaling function $\tau(q)$ is linear. A linear scaling function $\tau(q)$ means that the underling processes should be generated by an additive process. In this process, we expect the implied invariance conditions of

$$E[|X_{t,\Delta t}|^q] = t^{Hq}E[|X_1|^q] \tag{3.23}$$

The scaling law of equation (3.22) then holds with

$$c(q) = E\left[|X_1|^q\right] \quad \text{and} \quad \tau(q) = Hq - 1 \tag{3.24}$$

Note that a linear scaling function $\tau(q)$ is determined by its slope. As a typical example, standard Brownian motion for $H = 0.5$ has $\tau(q) + 1 = q/2$. In particular, fractional Brownian motion can be constructed by the fractional integration of order d of a white noise. Hence, we get the scaling exponent function with the degree of fractional integration:

$$\tau(q) + 1 = Hq = q\left(d - \frac{1}{2}\right) \tag{3.25}$$

The standard Brownian motion is here just an ordinary integral of white noise: $d = 1$, which gives the Hurst exponent $H = 1/2$.[13]

[13] Davis et al. (1994). In the case of the additive processes with Levy noise the behaviour of $\tau(q)+1$ is still linear. Note that there is a Levy index $\alpha(0 \leq \alpha \leq 2)$, which characterises the divergence of the moments of the Levy noise. In this case we have: $\tau(q) + 1 = q/\alpha$ for finite samples only if $q < \alpha$. The process diverges for $\alpha < q$.

The correlation function[14] is then given by

$$\gamma_{X_t} = E\left[X_{a,\Delta t} X_{a+t,\Delta t}\right] \tag{3.26}$$

This correlation function is not dependent on a because the process X_t has stationary increments. When $H = 1/2$, the MMAR generates a price process that has a white spectrum. If $E[\vartheta^{2H}]$ is finite, the autocovariance function of the price process satisfies for a fixed $t \geq \Delta t$

$$\gamma_{X_t} = K[(t+\Delta)^m + (t-\Delta)^m - 2t^m], \tag{3.27}$$

where $m = \tau_\vartheta(2H+1)+1$ and $K = c_\vartheta(2H)Var[B_H(1)]/2$. When $H > 0.5$, the fractional Brownian motion $B_H(t)$ has long-memory, and the price increments are positively correlated. When $H < 0.5$, the price increments are negatively correlated. However, the long-memory of multi-fractal price process is defined only on a bounded time range $[0, T]$. The limit of long-memory has at most the same length as the interval. Within this interval, the cumulative density function of multiplicative cascade has long-memory in the sense of hyperbolic decay of autocovariances.

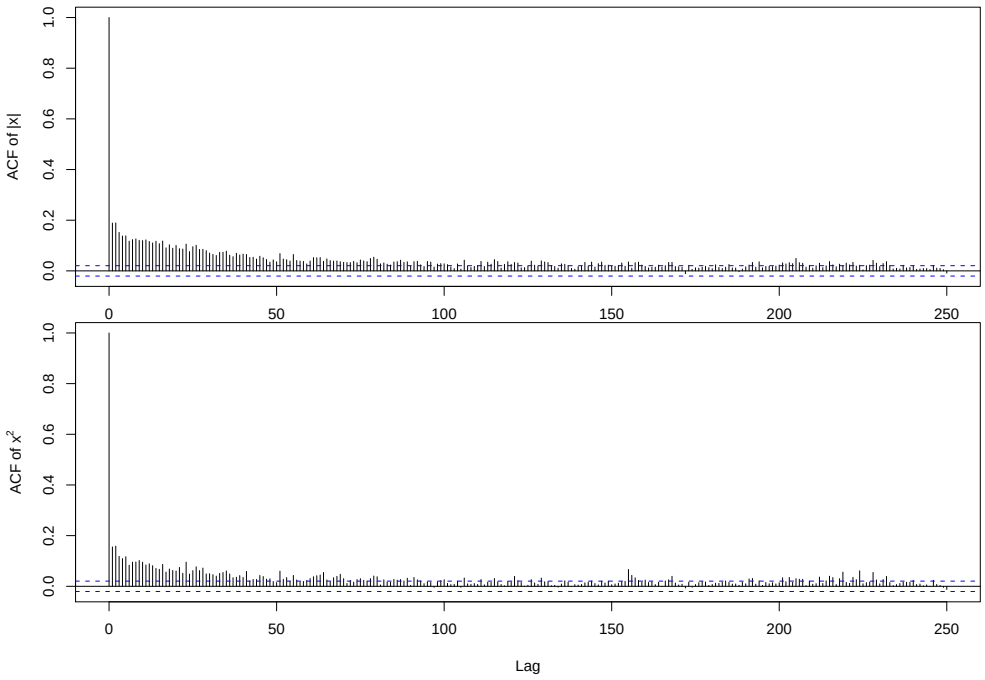

Figure 3.7: ACF of Binomial MMAR A sample path is generated by the binomial MMAR with $m_0 = 0.6$. The upper panel shows the empirical autocorrelation (ACF) of the absolute values of compound process. The bottom panel is the ACF of the squared compound process.

Next, the autocovariance of the absolute values of returns,

$$\delta_{X_{t,q}} = E\left[|X_{a,\Delta t}|^q |X_{a+t,\Delta t}|^q\right], \tag{3.28}$$

can be written by

$$\delta_{X_{t,q}} \geq \delta_{\vartheta_{t,Hq}} E[|B_H(1)|^q]^2 \tag{3.29}$$

when $H \geq 1/2$ and $E[\vartheta^{Hq}]$ is finite and all q is non-negative and $t \geq \Delta t$. The equation (3.29)

[14] See Mandelbrot et al. (1997) and Calvet and Fisher (2002) for the detailed proofs.

is hyperbolic in t when $0.5 \leq H$ and $t/\Delta t \to \infty$. That is, the increments of multi-fractal process have long-memory if $E[|X_{a,\Delta t}, X_{a+t,\Delta t}|^{2q}] < \infty$. The strong correlation of volatility in various time horizons are generated due to the hierarchical construction of cascades. Thus, the MMAR can generate a wide range of autocorrelation structures. Figure 3.7 shows the hyperbolic decay of the empirical autocorrelation for a simulated path generated by the binomial MMAR. It is most obvious that the autocorrelation of $|x_t|$ decays more slowly than the autocorrelation of x_t^2.

Apart from the bounded process and its theoretical limit of long-memory, MMAR accommodates long-memory in volatility and a variety of tail behaviours. With the property of martingale, the returns of MMAR are unpredictable. In particular, MMAR can generate volatility persistence at all time frequencies when the level $\bar{k}$ of cascade is adjusted [Mandelbrot et al. (1997)].

3.3 Empirical investigation of Multi-fractality

MMAR shows that multiplicative measures generate a stochastic process which has martingale, long memory in volatility and fat tails [Calvet and Fisher (2002)]. In this section we will show how one can test the multi-scaling function with real data in the framework of MMAR. To present the empirical evidence of the multi-fractality by means of the Korean Composite Stock Price Index (KOSPI)[15], we employ two concepts, $\tau(q)$ and $f(\alpha)$. The scaling function $\tau(q)$ uses a global average of qth powers of the underlying processes, whereas the multi-fractal spectrum $f(\alpha)$ describes the distribution of local Hölder exponents. Once, we estimate the scaling function $\tau(q)$, which is transformed into an estimate of the multi-fractal spectrum $f(\alpha)$, we can recover the individual components of the MMAR using the multi-fractal spectrum and simulate the return process.

3.3.1 Test of Multi-fractality

The methodology of Fisher, Calvet, and Mandelbort (1997) is used to test multi-fractality of financial time series.[16] The major consequence of multi-fractal process lies in the fact that the partition function presents temporal scaling of different power-law behaviours. Assume that the price X_t is defined within an interval $[0, T]$ of positive length. Dividing $[0, T]$ into N intervals of length Δt, we define the sample sum as the partition function:

$$S(q, \Delta t) \equiv \sum_{j=1}^{N-1} |X_{j+1,\Delta t} - X_{j,\Delta t}|^q.$$

Note that although the absolute differences of prices in various time intervals may have arbitrary correlation, these are identically distributed by the stationary increments property of multi-fractal processes. If X_t is multi-fractal, the scaling law yields

$$\log E[S(q, \Delta t)] = \tau(q) \log(\Delta t) + const \tag{3.30}$$

The test of multi-fractality is to check the linear relationship in equation (3.30). The testing procedure of multi-fractality is as follows:

[15]The Korean Composite Stock Price Index (KOSPI) is the main index in Seoul, South Korea. The KOSPI Index contains 200 of the largest and most liquid issues traded on the Korean Stock Exchange. The index is market capitalisation weighted, meaning that firms with the largest market value have the greatest influence on the KOSPI.

[16]Fisher et al. (1997) mention theoretical difficulties of justification of this test including problematic consistency of inference methods. but they justify the methodology on two counts. First, while inversion of logarithms and expectations will produce small sample biases in estimation, these need not prevent transmission of linearity in the equation of regression to approximate linearity in the sample partition function. Secondly, estimation biases for $\tau(q)$ can be studied and understood for particular cases of interest.

1. We plot $logS(q, \Delta t)$ versus $log\Delta t$ for various values of q. The linearity of these plots for given values of q is proposed as a test of MMAR. We check linearity by visual inspection.

2. Using OLS regression, we estimate the slope of the lines, which gives an estimate of the scaling function $\tau(q)$.[17]

3. Then, we transform the estimated scaling function into an estimated multi-fractal spectrum.

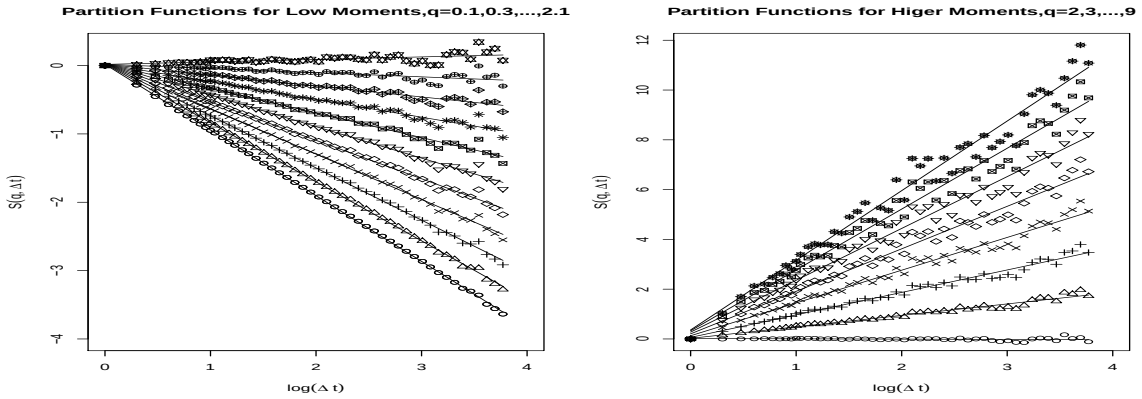

Figure 3.8: **Partition function of binomial Multi-Fractal** The simulated returns are generated by binomial multi-fractal Model $m_0 = 0.6$. The partition functions are obtained for a variety of positive moments ranging from $q = 0.1, 0.3, \ldots, 2.1$ in the left panel and $q = 2, 3, \ldots, 9$ in the right panel.

Let's compare the partition functions of a simulated multi-fractal with those of real data KOSPI. First of all, we generate a binomial multi-fractal process with the parameter $m_0 = 0.6$. Due to the multi-fractality, we expect that the partition functions of the simulated multi-fractal process is linear in $\log(\Delta t)$ with a different slope for each moment. Figure (3.8) presents the estimated partition functions for some low moments and higher moments. We observe a series of perfectly linear relationships for the lower moments and some higher moments. For the higher moments of $q = 7$, 8, 9, some bias and increasing standard deviations are observed.

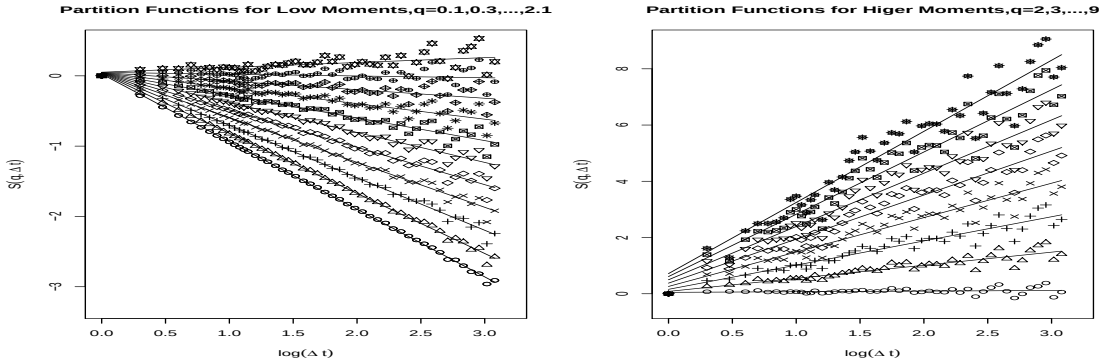

Figure 3.9: **Partition functions of KOSPI Returns**. The returns of Korean Composite Stock Price Index, or KOSPI is used for scaling. The daily data ranges from 1.1.1980 to 12.28.2001(Number of observations:6031) The panel shows the partition functions obtained for a variety of positive moments ranging from q = 0.1 to q = 9.

Figure (3.9) shows the empirical partition functions of KOSPI returns for some low moments and higher moments. The left panel presents the partition functions obtained for a variety of positive

[17]It is necessary to note that there is a problem of theoretical justification with respect to the consistency of estimation. See Fisher et al. (1997) for details.

moments ranging from q = 0.1 to 2.1. We also confirm a series of perfectly linear relationships for the lower moments. In the right panel, the higher positive moments ranging from 2 to 9. are depicted for the corresponding partition functions. The multi-scaling behaviour is consistently found in the data KOSPI. We remark that there are relatively larger bias on the regression lines in the scaling of higher moments.

Apart from the strong linearity of partition functions, it is of particular interest to see the slope zero. The value of q with slope zero in the figure (3.9) is about 1.95. Using the condition $\tau(1/H) = 0$, $\hat{H} \approx 0.513$. This confirms that there is almost no persistence in the raw returns of the data KOSPI. Note that the long-range dependence are obtained by using various power of returns, especially absolute returns (see Chapter 1).

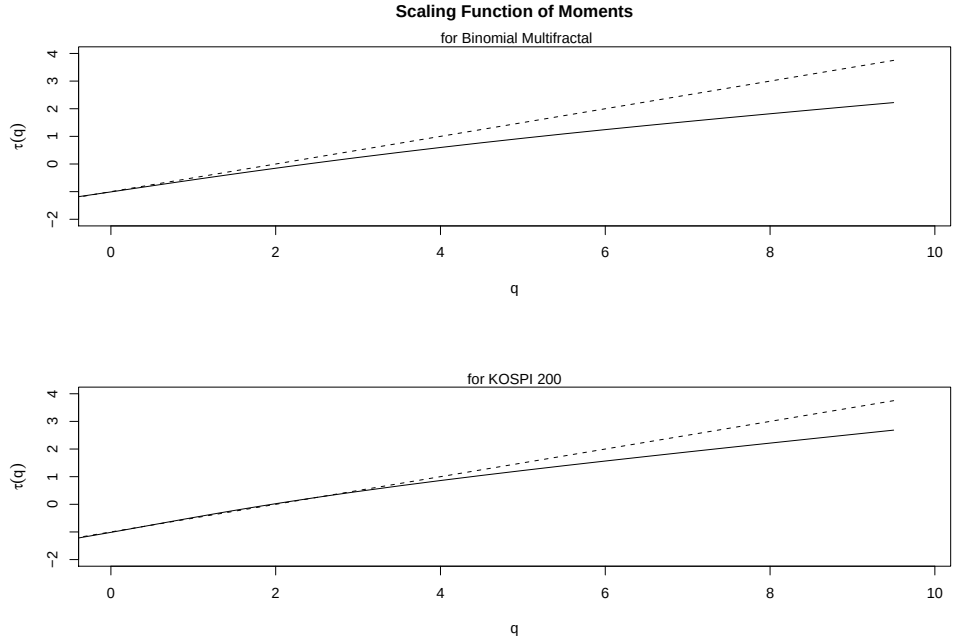

Figure 3.10: **Scaling Function of Moments**. The dotted line presents a mono-scaling like the scaling of fractional Brownian motion. The scaling function of binomial multi-fractal process is calculated using the parameter $m_0 = 0.6$. The returns of Korean Composite Stock Price Index, or KOSPI shows multi-scaling. The daily data ranges from 1.1.1980 to 12.28.2001(Number of observations: 6031) The bottom panel shows the scaling function of moments obtained for a variety of positive moments ranging from q = 0 to q = 9.5

We obtain the estimated scaling function $\hat{\tau}(q)$ from the estimating slopes of the partition functions for many q. Using (3.8) and (3.9), the scaling function are estimated for the simulated multi-fractal process and the data KOSPI at 88 different q in a region of $-10 < q < 10$. In a whole region, we take the increment of 0.5. For the interval of $-3 < q < 3$, the increment of 0.1 is additionally used. As we have shown the theoretical scaling functions of binomial and trinomial multi-fractals in the figure (3.4), extreme negative (positive) exponents of $|q| > 3$ tend to enhance small (large) fluctuations in this empirical scaling function. For comparison, the dotted line shows the scaling behaviour of standard Brownian motion: $\tau(q) = q/2 - 1$.

Figure (3.10) presents the estimated scaling function in solid line. In the upper panel the scaling function of simulated binomial multi-fractals illustrates its expected non-linearity (concavity) of $\hat{\tau}(q)$. The bottom panel of (3.10) also shows strong signs of curvature in the behaviour of some empirical exponent function for the data KOSPI. Furthermore, we observe the non-decreasing scaling function like the empirical scaling function of Deutschemark/US dollar estimated by Fisher et al. (1997).

The multi-fractal spectrum $f(\alpha)$[18] of the KOSPI data is showed in the figure (3.11)[19] It is just a visualisation of the Legendre transformation[20]. Using the estimated $\tau\hat{(q)}$, we get the multi-fractal spectrum (α) through a Legendre transformation:

$$f(\alpha) = [q\alpha - \tau(q)]_{q_{argmin}}.$$

Therefore, this spectrum is obtained by drawing lines of slope q and intercept $-\tau(q)$ for various q. The data KOSPI exhibits multi-scaling properties. This multi-fractal spectrum (3.11) constitutes an hump-shape envelop that comes from the estimated scaling functions. Considering that the length scale affects the calculation of the multi-fractal spectrum and there are not sufficient data available to determine such a clear spectrum like the theoretical one (see figure (3.5)), it is remarkable to find such non-linear function of singularity in the positive area of moments. Binomial or other multinominal multiplicative cascade may be proposed as a proxy to the return distribution of the data KOSPI. The higher the positive moments (q) the lower the singularity spectrum $f(\alpha)$. By the way, we find that the estimated multi-fractal spectrum is too sensitive for $q < 0$ and thus the right part of this spectrum is not reliable. For such unreliable results in the area of negative moments (q) in physics, Chhabra et al. (1989) give two reasons: 1) corruption by noise in the data 2) occasional non-turbulent portions of fluid. The latter leads to the effects that the spuriously low values of dissipation are emphasised with a negative moments (q). In financial markets we often find such *non-turbulent* time of very low volatility that seems just to be noise. To consider these tranquil periods in the dynamics of volatility, it will be of interest to study multi-fractal generated on a Cantor set as a general support instead of Binomial or other multinomial cascade.

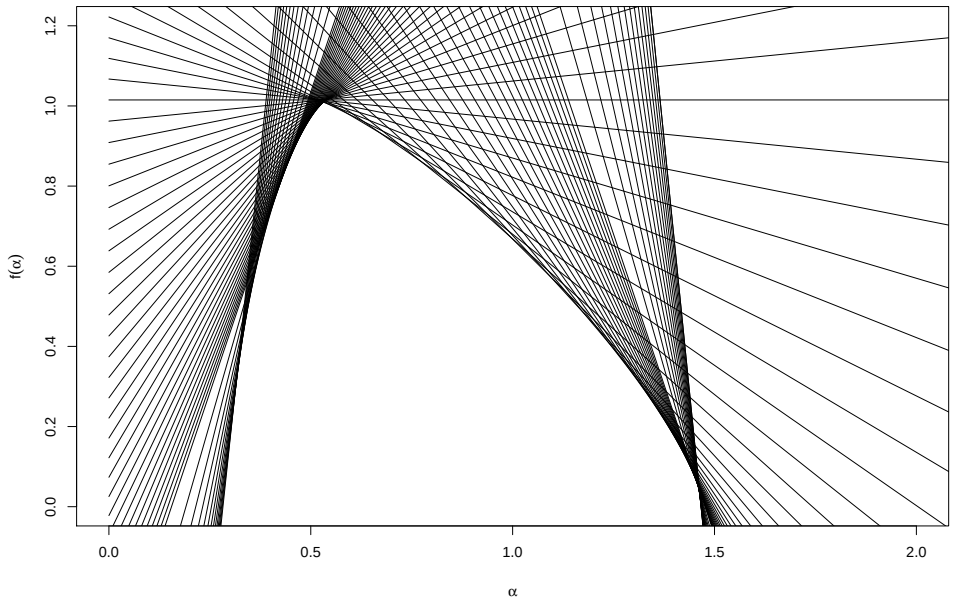

Figure 3.11: **KOSPI Multi-fractal Spectrum**. The returns of Korean Composite Stock Price Index, or KOSPI shows multi -scaling. The daily data ranges from 1.1.1980 to 12.28.2001 with 6031 observations. The empirical multi-fractal spectrum is obtained from the scaling function of moments obtained for a variety of moments ranging from $-10 \leq q \leq 10$

[18]For the calculation of $f(\alpha)$, I use the program written in GAUSS by Lux (1999).
[19]Because the strong similarity of $f(\alpha)$, we leave out the corresponding figure of binomial multi-fractal.
[20]see Evertz and Mandelbrot (1992) for details.

Chapter 4

Trinomial Markov switching multi-fractal model

Calvet and Fisher (2004) first proposed a discrete-time Markov switching multi-fractal model (hereafter MSMF model) as an alternative to the stochastic volatility model. Differently from the MMAR in which the multi-fractal volatility state is changed randomly, in the MSMF model it is assumed that the latent volatility states follow a first-order Markov chain process. They specify the switching mechanism with restricted transition probabilities using a binomial multi-fractal process. In this chapter, we will extend the binomial MSMF model to the case of a trinomial multi-fractal process. After the construction of the model, we will estimate the multifractal parameters m_0, m_1 and σ using two different methods: the generalised method of moments (GMM) and the maximum likelihood method (ML). With the empirical results using the real financial time series, we will discuss the problem of the model selection with respect to the number of multipliers with some suggested tests. Finally, we will propose the quadronomial MSMF process and derive the required moments to estimate the quadronomial multi-fractal parameters. We will briefly investigate the property of estimates using Monte Carlo experiments.

4.1 The model construction

The Makov switching multi-fractal model is represented by

$$\begin{aligned} x_t &\equiv 100 \cdot (\ln P_t - \ln P_{t-1}) \\ &= \sigma_t \, u_t, \end{aligned} \tag{4.1}$$

where

$$\begin{aligned} \sigma_t &= \sigma \sqrt{M_t} \\ &= \sigma \sqrt{\prod_{k=1}^{\bar{k}} m_{t,k}} \end{aligned} \tag{4.2}$$

x_t refers to the asset return. P_t and P_{t-1} denote an asset price at time t and $t-1$, respectively. The white noise u_t is here given by Gaussian distribution with zero mean value and unit variance $N(0,1)$. σ is a positive constant. As shown in equation (4.2), the MSMF process σ_t has a finite

number of hidden volatility states M_t generated by products of multipliers or volatility components $m_{t,k}$. Assumed that the dynamics of states M_t follows a first-order Markov chain. That is, a volatility state M_t is only affected by a state of M_{t-1}:

$$P[M_t|M_{t-1}, M_{t-2}, \dots] = P[M_t|M_{t-1}] \tag{4.3}$$

Due to equation (4.2) and the assumption of first-order Markov chain M_t, equation (4.3) can be written by

$$\begin{aligned} & P\left[\prod_{k=1}^{\bar{k}} m_{t,k} \middle| \prod_{k=1}^{\bar{k}} m_{t-1,k}, \prod_{k=1}^{\bar{k}} m_{t-2,k}, \dots\right] \\ &= P\left[\prod_{k=1}^{\bar{k}} m_{t,k} \middle| \prod_{k=1}^{\bar{k}} m_{t-1,k}\right] \\ &= P\Big[m_{t,1}|m_{t-1,1}\Big] P\Big[m_{t,2}|m_{t-1,2}\Big] \dots P\Big[m_{t,\bar{k}}|m_{t-1,\bar{k}}\Big]. \end{aligned} \tag{4.4}$$

We obtain equation (4.4) by the assumption that volatility components in each level of cascade are indepenent. A switching of $m_{t-1,k}$ into $m_{t,k}$ is governed by a transition probability[1] γ_k in form

$$\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}, \tag{4.5}$$

where $\gamma_1 \in (0, 1)$ and $b \in (1, \infty)$. How often a volatility component $m_{t,k}$ changes is determined by k if the maximal number of multipliers $\bar{k}$ is fixed. In this context k is refered to as frequency of volatility component.

4.1.1 Properties of transition probabilities

Assuming that $m_{t,k}$ takes one of three values m_0, m_1 and m_2 from the trinomial distribution:

$$m_{t,k} \in \{m_0, m_1, m_2\}. \tag{4.6}$$

Constrained $m_0 + m_1 + m_2 = 3$ so that the normalization condition of $E[M_t] = 1$ is satisfied. Using $b = 3$ and $\gamma_{\bar{k}} = 1/3$, the transition probability in equation (4.5) is specifed such that

$$\gamma_k = 1 - \left(1 - \frac{1}{3}\right)^{3^{k-\bar{k}}}. \tag{4.7}$$

The probability of staying in one state $m_{t,k} = m_{t-1,k}$ is then determined by $1 - \gamma_k$. It is assumed that the switching events and new draws of $m_{t,k}$ from the trinomial distribution are not governed by the number of multipliers k and time t. Figure (4.1) diplays the transition probability at three different values of $\bar{k} = 2$ (left), $\bar{k} = 4$ (center) and $\bar{k} = 6$ (right). The solid line describes the specifed probability in equation (4.7), while the parameterization of $\gamma_{\bar{k}} = 1/20$ and $b = 2.5$ in Calvet and Fisher (2005) is drawn by dashed line. We can see that the switching probabilty grows dramatically with increasing number of $\bar{k}$. The moderate increase of the dashed lines is due to the small values assigned to $\gamma_{\bar{k}}$ and b in the transition probability. The specification by Calvet and

[1] In Calvet and Fisher (2001), Calvet and Fisher (2004) and Calvet and Fisher (2005) they used $b \in (1, \infty)$ and $\gamma_{\bar{k}}$ as a parameter to be estimated. They introduced this specification in connection with the discretization of a Poisson arrival process. With constant rate of $b = 3$, we let the transition probability of low frequency components grow approximately at geometric rate 3. Lux (2006) used for the binomial case $b = 2$ and $\gamma_{\bar{k}} = 1/2$.

Fisher (2005) has the advantage to analyse financial time series with a relatively strongly persistent volatility. It requires, however, a large area of frequency to achieve the corresponding long-range dependence. With this specified probability of equation (4.7) and relatively small $\bar{k}$, the trinomial MSMF model can also accomodate high frequency shocks with large price changes and various degrees of long range dependence.

4.1.2 Dynamics of volatility components

The behaviour of the probability γ_k is directly related to the persistence of $m_{t,k}$.[2] As said before, the switching probability grows with increasing k. That means, the persistency of a component $m_{t,k}$ is apperently reduced as $k \longrightarrow \bar{k}$. In each region of low, intermediate and high k, a hidden switching of volatility components $m_{t,k}$ shows the following properties [Calvet and Fisher (2004)]:

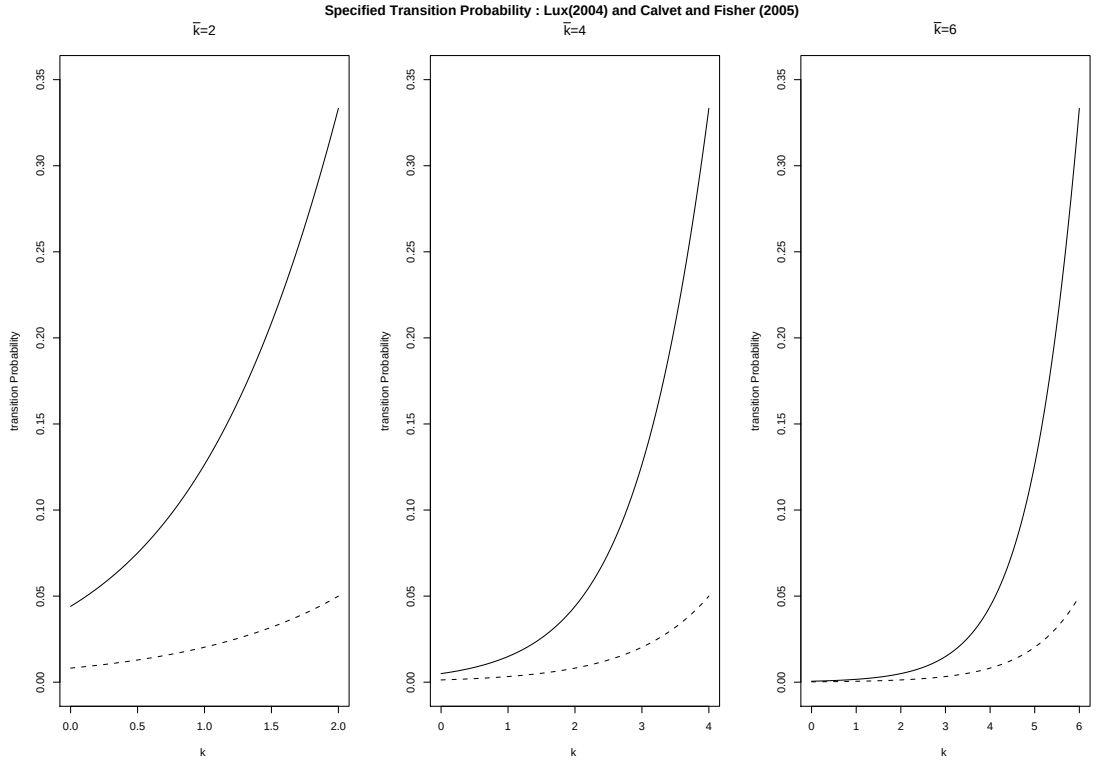

Figure 4.1: The specified transition probability $\gamma_k = 1 - (1 - \gamma_{\bar{k}})^{b^{k-\bar{k}}}$ at three different values of $\bar{k} = 2$ (left), $\bar{k} = 4$ (center) and $\bar{k} = 6$ (right). The solid line stands for the parametrization of $\gamma_{\bar{k}} = 1/3$ and $b = 3$, while the dashe lines describe the specification $\gamma_{\bar{k}} = 1/20$ and $b = 2.5$ by Calvet and Fisher (2005).

- Low-frequency multipliers $m_{t,k}$ build a fundamental block for a persistent volatility. Changing of the low-level multipliers means that there may be a strong tendency for movement in multiplied volatility. This switching of multiplied volatility may often be observed when

[2]The persistence of $m_{t,k}$ is affected by the duration of the most persistent volatility component, $1/\gamma_{\bar{k}}$ and maximum value $\bar{k}$. According to Calvet and Fisher (2004) the duration should be typically of the same order as the length of the data. That means, the duration of the most frequent-varying component just depends on the data frequency in time (hourly, daily, weekly, etc). But it may lead to a tremendous limitation of application areas as a general tool of time series analysis [Lux (2004)]. In general, a discrete MSMF model can be applied for any length of data. The choice of optimal $\bar{k}$ must be considered in the context of a model selection.

negative returns lead to a large increase of the volatility[3]. For instance, Schwert (1989b) observed the change of volatilitiy regimes during the stock market crash of October 19, 1987. According to his observations in Schwert (1989a) the average level of volatility is much higher during recessions. We refer to an asymmetry in the return-volatility relation. Volatility can quickly return to a low or an average level after highly volatile periods. It is an obvious evidence for regime switching in volatility. This discrete change of regimes was firstly formalised by Hamilton's regime-switching model, see Hamilton (1989).

- Intermediate-frequency multipliers $m_{t,k}$ care for a smoothed changing of volatility states and play a role of moving average in the volatility. The resulted volatility may be clustered mainly due to the highly correlated volatility components on the different time units.
- High-frequency multipliers $m_{t,k}$ give additional outliers through their direct effects on the tails of MSMF process [Calvet and Fisher (2005)]. These volatility components are very short-lived and in extreme case behave like noise. Therefore, it is not easy to catch the effect of this component over a certain threshold of $\bar{k}$.

We can confirm that this dynamics of the volatility component in Figure (4.2) is produced by the simulation of equation (4.2) using the trinomial multi-fractal parameter $m_0 = 1.2$, $m_1 = 0.5$, the scaling parameter $\sigma = 1$ and the number of multipliers $\bar{k} = 5$. The multiplied volatility $M_t = m_{t,1}m_{t,2}\dots m_{t,5}$ is also visualised on the bottom panel of the right side. Here, x-axis of time runs to 1000. The switching of multipliers $m_{t,k}$ occurs frequently while k is getting to vary from 1 to 5. In other words, the expected duration of each volatility component deceases on average as the number of multipliers k increases.

By means of the same combination of parameters, we illustrate the trinomial Markov-switching multi-fractal process in Figure (4.3). The upper panel of this figure shows the latent volatility process M_t with running time 10000, whereas the bottom panel presents the dynamics of the asset returns $x_t = M_t^{1/2}u_t$. The sequence of the return is similar to that of Calvet and Fisher (2004) and Lux (2006) generated by using binomial MSMF model. The evolution of a multiplied volatility $M_t = m_{t,1}m_{t,2}\dots m_{t,\bar{k}}$ directly results from the time-dependency of components $m_{t,k}$ in each k.

When M_{t+1} has lots of shared components with M_t, it does not change abruptly. The clustered volatility may come from these shared components. This generating mechanism makes a discrete MSMF process attractive for modelling financial volatility which has a wide range of variability and strong persistency. High volatility states turn up whenever each constitutent multiplier takes the largest values among m_0, m_1 and m_2 in each cascade. In Figure (4.2) the high volatility states are observed most frequently in the region of $400 < t < 1000$, just where the constitutent component takes frequently the value of either $m_0 = 1.2$ or $m_2 = 1.3$ in each cascade.

Based on the generated volatility in the upper panel of figure (4.3), we get a simulated sample path of the trinomial MSMF process through the multiplication of the squared volatility process and white noise. The bottome panel of the figure (4.3) is strongly similar to the return process of the real financial time series. The figure presents a variety of stylized facts facts such as volatility clustering at all time scale and intermittent fluctuations.

4.2 Moments of the trinomial MSMF process

In this section we consider some property of trinomial MSMF process. The unconditional and conditonal moments of multi-fractal process M_t and the compound process x_t are examined separately. We calculate these statistical moments to estimate the multi-fractal parameters m_0, m_1

[3]Schwert (1989b) mentioned that it is an increased chance of large returns of indeterminate sign.

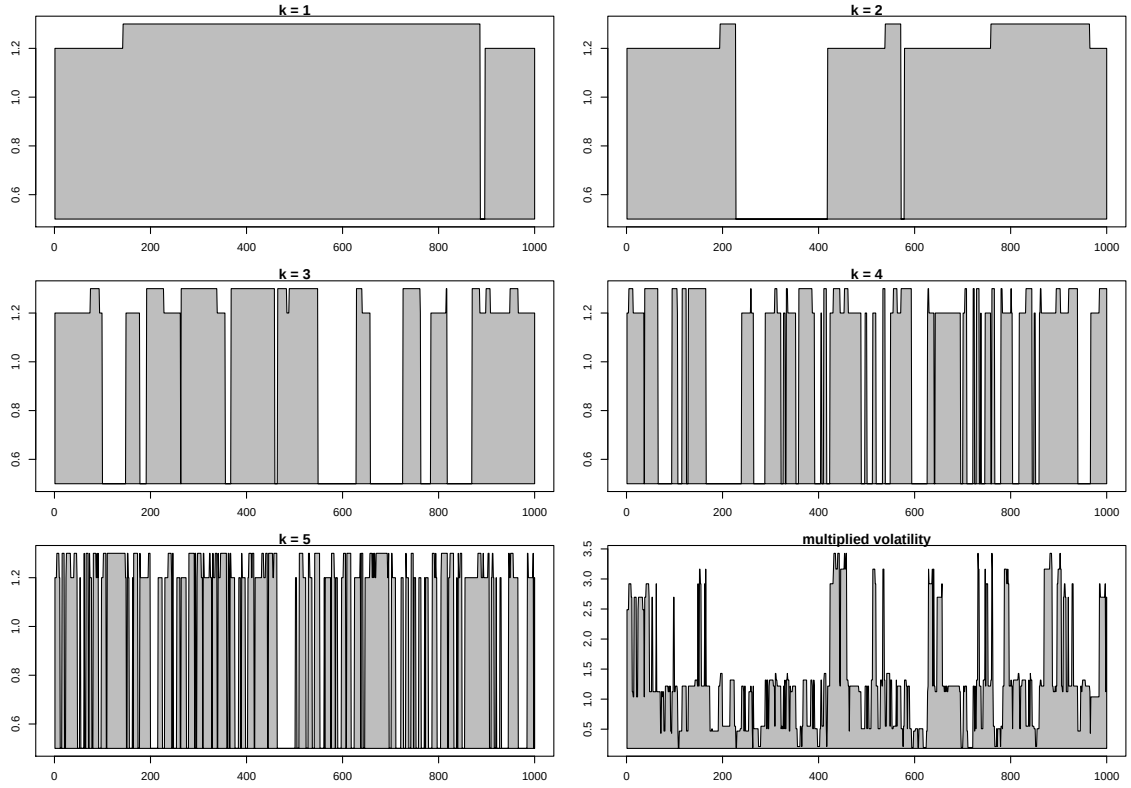

Figure 4.2: The expected duration of volatility components depends on the number of k. In each k the expected duration varies almost 1/3 times from $k = 1$ to $k = 5$ on average. The multiplied volatility M_t is presented on the botton panel of right side.

as well as the scaling parameter σ when applying the generalized method of moments (GMM). First, we derive the unconditional moments and autocovariance of the multi-fractal process. Due to the instability or huge variabilty of these moments, we introduce as next the log increments of volatility process $\eta_{t,\Delta t} = \ln M_t - \ln M_{t-\Delta t}$ and calculate not only the second unconditional moment of this process, but also the first and second correlated moments. According to the assumption of independent compounding, the moment of the MSMF process $E[|x_t|^q]$ is obtained simply by the multiplication of two moments: the moments of the multi-fractal process and the moments of Gaussian noise. We point out that the distribution of compound process x_t has heavy tails. At last, we get the moments of log increments of transformed compound process $E[\xi_{t+T,T}\xi_{t,T}]$ and $E[\xi^2_{t+T,T}\xi^2_{t,T}]$ on the basis of the previous results.

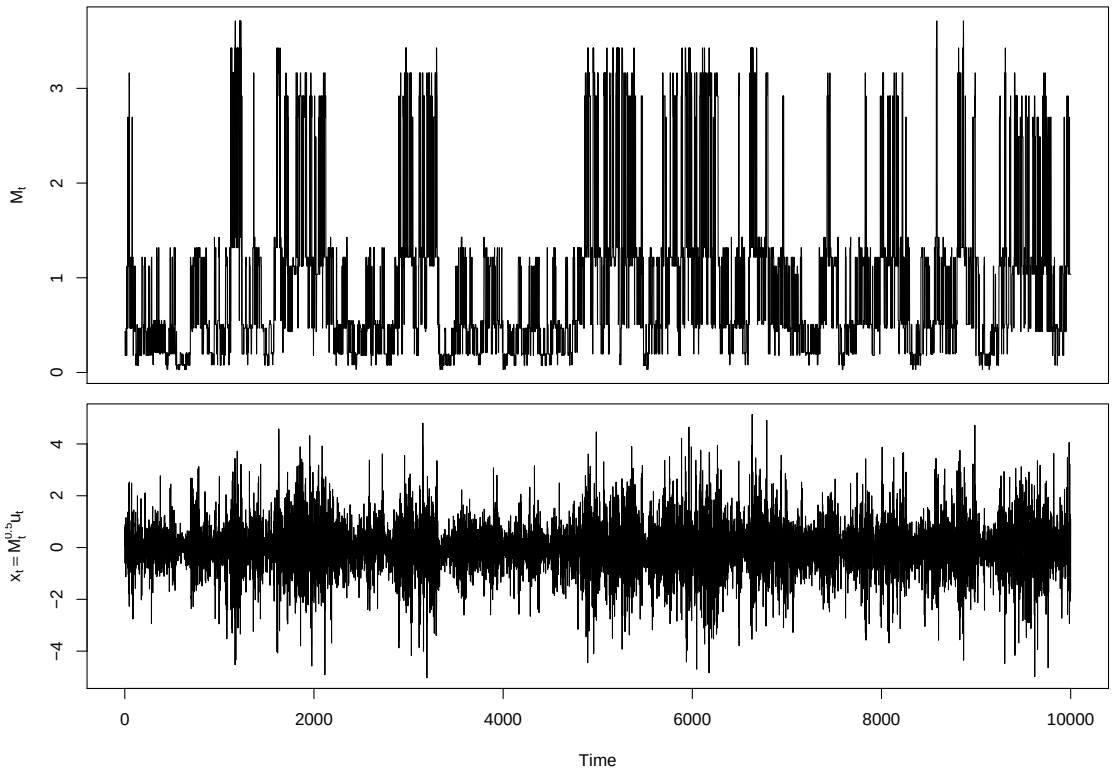

Figure 4.3: The upper panel shows the simulated volatility process $M_t = m_{t,1} m_{t,2} \dots m_{t,5}$. Here, each component $m_{t,k}$ is drawn from $m_0 = 1.2$, $m_1 = 0.5$, $m_2 = 1.3$ and $\sigma = 1$. Note that $m_2 = 3 - m_0 - m_1$. The switching probability γ_k is given in equation (4.7). The bottom panel shows simulated return $x_t = M_t^{1/2} u_t$ based on the volatility in the upper panel.

4.2.1 Moments of the trinomial multi-fractal process

The unconditional moments of the multi-fractal process[4] $E[M_t^q]$ result in

$$\begin{aligned} E[M_t^q] &= E\left[\left(\prod_{k=1}^{\bar{k}} m_{t,k}\right)^q\right] \\ &= \left(\frac{1}{3}m_0^q + \frac{1}{3}m_1^q + \frac{1}{3}m_2^q\right)^{\bar{k}}. \end{aligned} \tag{4.8}$$

Here, q is the order of moment. The first moment $E[M_t]$ and the second moment $E[M_t^2]$ have

[4]See Mandelbrot et al. (1997) for details. Using combinatorial multi-fractal process, Mandelbrot, Calvet and Fisher derive the following unconditional moment:

$$E[M_{\delta t}^q] = c(q) \cdot \delta t^{\tau(q)+1},$$

where $c(q)$ is a normalization factor. In MMAR, the binomial multipliers m_0 and m_1 are constrained by $0 < m_0 < 1$ and $m_1 = 2 - m_0$. Hence, the average of each cascade results in $E[M_t^1] = \frac{1}{2^k},\ k = 1, \dots, \bar{k}$. It should be normalized by multiplication of $c(q) = 2^{\frac{k}{2}}$. In the multipliers of binomial MSMF model(Calvet and Fisher (2004) and Lux (2004)) the parameters are constrained such as $0 < m_0 < 1$ and $m_1 = 2 - m_0$ so that $E[M_t^1] = 1$ for each cascade. The normalising factor is now $c(q) = 2^k$ over all moments.

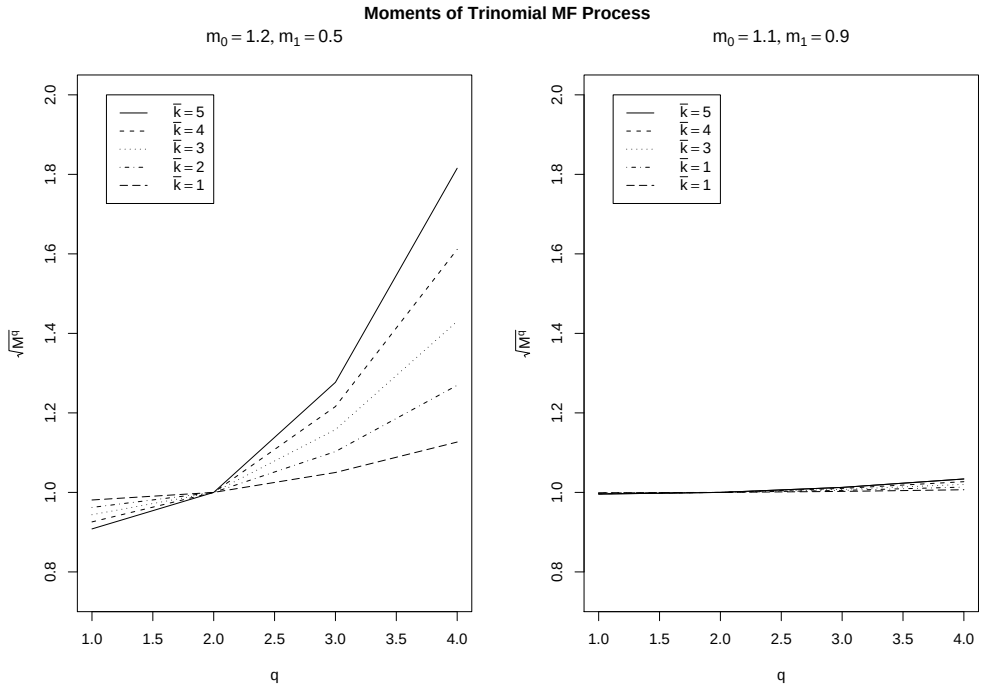

Figure 4.4: Unconditional moments $E[M_t^{q/2}]$ of trinomial multi-fractal process to the order of moment $q = 1 \ldots 4$ at fixed $\bar{k} = 5$ (solid), $\bar{k} = 4$ (dashed), $\bar{k} = 3$(dotted), $\bar{k} = 2$ (dotted-dashed) and $\bar{k} = 1$ (long dashed). In case of $m_0 = 1.2$ and $m_1 = 0.5$ (left panel) the moments $E[M_t^{q/2}]$ increase obviously with growing k, while the $E[M_t^{q/2}]$ shows no apparent change in value against increasing k if using $m_0 = 1.1$ and $m_1 = 0.9$ (right panel). The unconditional moments are strongly dependent on $max\,|m_i - m_j|$, $i = j = 0, 1, 2$ at fixed $\bar{k}$

particulaly simple forms:

$$E[M_t] = \left(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\right)^{\bar{k}} = 1$$
$$E\left[M_t^2\right] = \left(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\right)^{\bar{k}}. \tag{4.9}$$

Following to equation (4.9) we get the following variance of MF process:

$$Var[M_t] = E[M_t^2] - \Big(E[M_t]\Big)^2$$
$$= \left(\frac{1}{3}\,m_0^2 + \frac{1}{3}\,m_1^2 + \frac{1}{3}\,m_2^2\right)^{\bar{k}} - 1. \tag{4.10}$$

Figure (4.4) shows some moments $E[M_t^{q/2}]$ of the trinomial multi-fractal process. The x-axis displays the order of moment. The left plot is generated by means of $m_0 = 1.2$ and $m_1 = 0.5$, while the right one is produced by $m_0 = 1.1$ and $m_1 = 0.9$. The maximal number of multiplier $\bar{k}$ varies from 1 to 5. In the left plot the moment is growing rapidly when $\bar{k}$ increases. Depdenent on the number of multipliers $\bar{k}$, this combination of mulit-fractal parameters can generate huge range of variability which reflects multi-scaling. But the moment in the right plot shows no apparent change in values even though k is turning from 1 to 5. It seems to be mono-scaling apparently. Hence, it might be deduced that the unconditional moments are strongly dependent on $max\,|m_i - m_j|$ $\;i, j = 0,\, 1,\, 2$. The $E[M_t^q]$ has a monoscaling at $m_0 = m_1 = m_2$.

Autocorrelation function

Under consideration of the model equations (4.1) and (4.2), we get the autocorvariance $E[M_{t+1}M_t]$ at time interval $\Delta t = 1$

$$\begin{aligned}
E[M_{t+1}M_t] = \prod_{k=1}^{\bar{k}} \Bigg[& \frac{1}{3}m_0\Big\{(1-\gamma_k)m_0 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\} \\
& + \frac{1}{3}m_1\Big\{(1-\gamma_k)m_1 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\} \\
& + \frac{1}{3}m_2\Big\{(1-\gamma_k)m_2 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\} \Bigg] \\
= \prod_{k=1}^{\bar{k}} \Bigg[& \frac{2}{3}\Big(1-(1-\gamma_k)\Big)\Big(\frac{1}{3}m_0m_1 + \frac{1}{3}m_1m_2 + \frac{1}{3}m_2m_0\Big) \\
& + \Big\{(1-\gamma_k) + \frac{1}{3}\Big(1-(1-\gamma_k)\Big)\Big\}\Big(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\Big) \Bigg]. \qquad (4.11)
\end{aligned}$$

For an arbitrary Δt, the $E[M_{t+\Delta t}M_t]$ has a form of

$$\begin{aligned}
E[M_{t+\Delta t}M_t] = \prod_{k=1}^{\bar{k}} \Bigg[& \frac{2}{3}\Big\{1-(1-\gamma_k)^{\Delta t}\Big\}\Big(\frac{1}{3}m_0m_1 + \frac{1}{3}m_1m_2 + \frac{1}{3}m_2m_0\Big) \\
& + \Big\{(1-\gamma_k)^{\Delta t} + \frac{1}{3}\Big(1-(1-\gamma_k)^{\Delta t}\Big)\Big\}\Big(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\Big)\Bigg]. \qquad (4.12)
\end{aligned}$$

Let's consider the first *analytical* autocorrelation function[5] $\rho_{\Delta t}$ defined by

$$\begin{aligned}
\rho_{\Delta t} &= \rho[M_{t+\Delta t},\, M_t] \\
&= \frac{E[M_{t+\Delta t}M_t] - E[M_t]^2}{Var[M_t]}. \qquad (4.13)
\end{aligned}$$

Plugging equation (4.9)and equation (4.10) in equation (4.13), we get then

$$\rho_{\Delta t} = \frac{E[M_{t+\Delta t}M_t] - 1}{\Big(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\Big)^{\bar{k}} - 1} \qquad (4.14)$$

In Figure (4.5) the simulated autocorrelation of trinomial MSMF process is illustrated to time interval Δt. Hereby, the two different sets of $m_0 = 1.2, m_1 = 0.5$ (upper panel) and $m_0 = 1.7, m_1 = 0.1$ (bottom panel) are employed. The number of multipliers $\bar{k}$ varies from 2 to 10. The figure shows that the $\rho_\Delta t$ decays hyperbolically. The autocorrelation function $\rho_\Delta t$ is affected mainly by absolute difference between the m_i and m_j as well as $\bar{k}$. At fixed $\bar{k}$, the autocorrelation function $\rho_\Delta t$ with $m_0 - m_1 = 0.7$ (panel I) has larger values than the one with $m_0 - m_1 = 1.6$ (panel II). At fixed multi-fractal parameters m_i, $i = 0, 1$ the simulated values increase with growing k. This means, the degree of long range dependence can be controlled by the combinations of the parameters m_i and k. These results can be extended to other discrete MSMF process such as quadronomial MSMF

[5]See the proof of Calvet and Fisher (2004) about the hyperbolic decline of autocovariance for binomial MSMF process: $\ln \rho^q(n) \sim \delta(q) \ln n$. We note that the hyperbolic decline of autocovariance is valid only for a finite time interval and the autocovarance decays exponentially beyond the time interval.

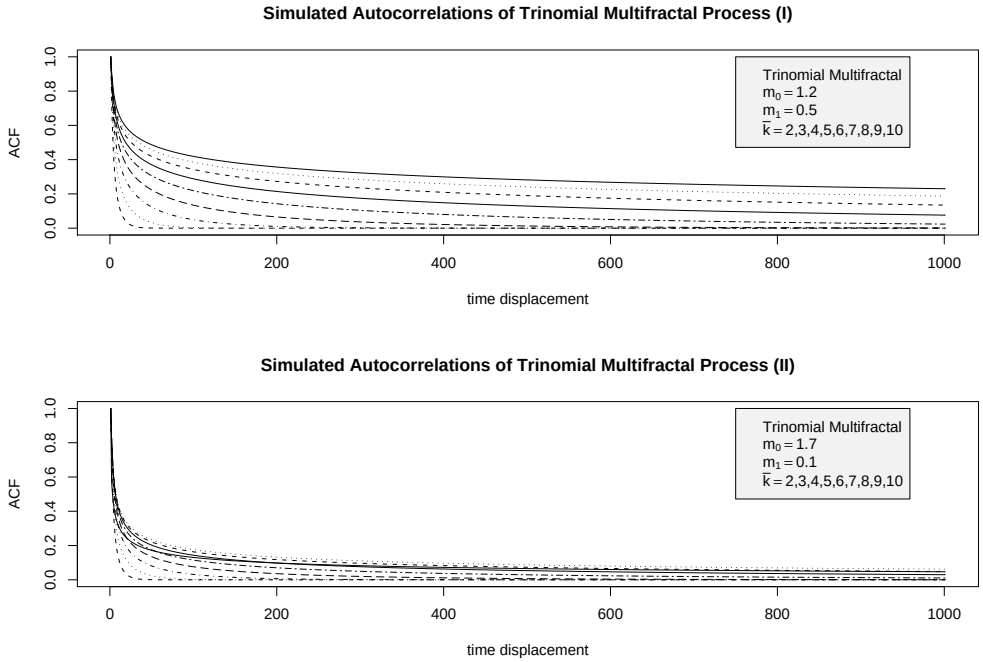

Figure 4.5: Simulated autocorrelation function(ACF) of trinomial multi-fractal process. The parameters values are assigned such that $m_0 = 1.2$, $m_1 = 0.5$ in upper panel(I) and $m_0 = 1.7$, $m_1 = 0.1$ in bottom panel(II).

process, see the section (4.7). It is obivous that the MSMF process can mimic arbitrary degrees of long memory. Compared with the binomial process illustrated in Figure (4.6), the trinomial MSMF is able to generate a wider range of long memory. Note that the autocorrelation function is not defined for $m_0 = m_1 = m_2 = 1$, because the multiplied volatility M_t in this case is a constant. Therefore, there is no variance.[6]

4.2.2 Moment of log-transformed volatility process

As previously said, we are going to estimate multi-fractal parameters $(m_i, i = 0, 1)$ and scaling factor (σ). Thereby, we use two methods: GMM and ML. Note that the 'long range dependence over a finite interval' of MSMF process does still guarantee existence of all conditional and unconditional moments of the Markov switching multi-fractal volatility process. Although applicability of the regulartiy conditions are hampered by this type of 'long memory on a bounded interval', the proximity to 'true' long memory might rise practical concerns [Lux (2006)]. For example, note that with $b = 3$ and $\bar{k} = 10$, the range of 'apparent' long memory might exceed the size of most available data for daily financial prices. In finite samples, application of GMM to Markov switching multi-fractals could yield poor results since usual estimates of the covariance matrix of the moment conditions might show large pre-asymptotic variation [Lux (2006)]. For a practical solution, Lux (2006) suggested log differences of absolute returns firstly. We introduce the new variable $\eta_{t,T}$ of

[6]For this speical case the compound process x_t follows just the uncorrelated white noise and the autocovariance ist the autocovariance of fractional Gaussian noise with the Hurst exponent $H = 0.5$.

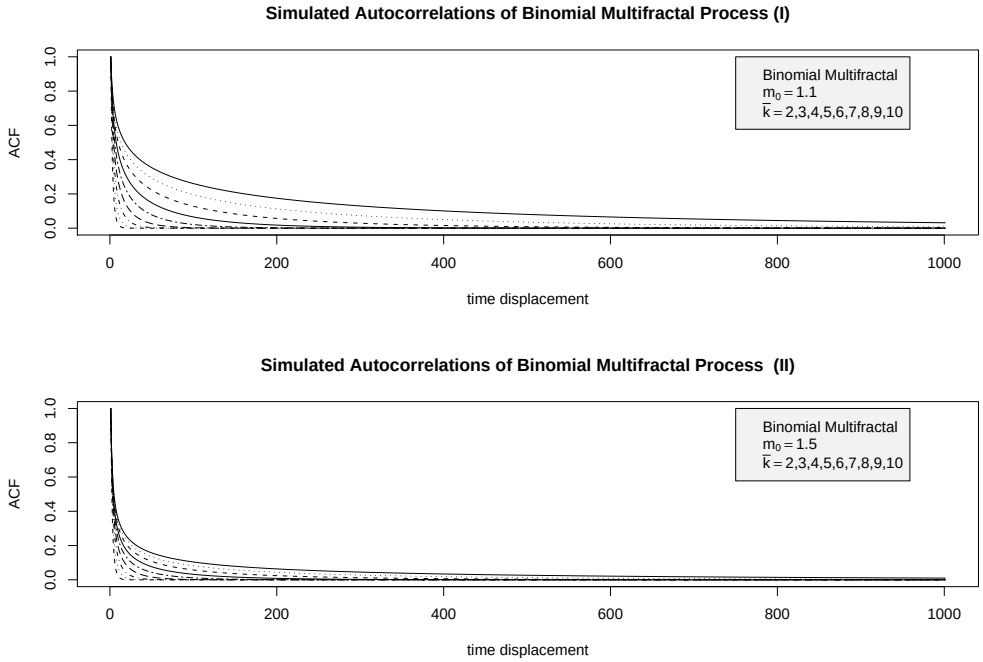

Figure 4.6: Simulated autocorrelation function(ACF) of binomial multi-fractal process. In the upper panel $m_0 = 1.1$, whereas in the lower panel $m_0 = 1.5$. The autocorrelation function depends not only on the number of volatility component, but also on the parameter m_0. The transition probability is here specified as $\gamma_k = 1 - (1 - 1/2)^{2^{(k-\bar{k})}}$ for $k = 2, \ldots, \bar{k}$

log increments

$$\begin{aligned}\eta_{t,\Delta t} &= \ln M_t - \ln M_{t-\Delta t} \\ &= \sum_{k=1}^{\bar{k}} \ln m_{t,k} - \sum_{k=1}^{\bar{k}} \ln m_{t-\Delta t,k} \\ &= \sum_{k=1}^{\bar{k}} \varepsilon_{t,k} - \sum_{k=1}^{\bar{k}} \varepsilon_{t-\Delta t,k}, \end{aligned} \tag{4.15}$$

where $\varepsilon_{t,k} = \ln m_{t,k}$. It is obvious that the first moment of log increments $E[\eta_{t,\Delta t}] = 0$ for all Δt. We will represent briefly the first and second autocovariance of log-transformed volatility process for $\Delta t > 1$. The explicit derivations of the equations are found in the appendix A.

- The second unconditional moment of the log-transformed volatility process:

$$\begin{aligned}&E[\eta^2_{t+\Delta t,\Delta t}] = \\ &\quad \frac{4}{3}\Big(\ln^2 m_0 + \ln^2 m_1 + \ln^2 m_2 - \ln m_0 \ln m_1 - \ln m_1 \ln m_2 - \ln m_2 \ln m_0\Big) \\ &\qquad \sum_{k=1}^{\bar{k}} \left\{ \frac{1}{3}\Big(1 - (1 - \gamma_k)^{\Delta t}\Big) \right\} \end{aligned} \tag{4.16}$$

- The autocovariance of the log-transformed volatility process:

$$E[\eta_{t+\Delta t,\Delta t}\,\eta_{t,\Delta t}] = 2\Big(\ln m_0 \ln m_1 + \ln m_1 \ln m_2 + \ln m_2 \ln m_0 - \ln^2 m_0 - \ln^2 m_1 - \ln^2 m_2\Big) \sum_{k=1}^{\bar{k}} \left\{ \frac{1}{3}\Big(1-(1-\gamma_k)^{\Delta t}\Big)\right\}^2. \quad (4.17)$$

- The squared autocovariance of the log-transformed volatility process:

$$\begin{aligned} &E[\eta^2_{t+\Delta t,\Delta t}\eta^2_{t,\Delta t}] \\ &= 2\sum_{k=1}^{\bar{k}} \left\{ \frac{1}{9}\Big(1-\Big(1-\gamma_k\Big)^{\Delta t}\Big)^2\right\}\xi \\ &\quad + \frac{16}{9}\sum_{k=1}^{\bar{k}} \left\{ \frac{1}{3}\Big(1-\Big(1-\gamma_k\Big)^{\Delta t}\Big) \sum_{k'=1,k'\neq k}^{\bar{k}} \frac{1}{3}\Big(1-\Big(1-\gamma_{k'}\Big)^{\Delta t}\Big)\right\}\xi \\ &\quad + 8\sum_{k=1}^{\bar{k}} \left\{ \frac{1}{9}\Big(1-\Big(1-\gamma_k\Big)^{\Delta t}\Big)^2 \sum_{k'=1,k'\neq k}^{\bar{k}} \frac{1}{9}\Big(1-\Big(1-\gamma_{k'}\Big)^{\Delta t}\Big)^2\right\}\xi, \end{aligned} \quad (4.18)$$

with

$$\xi = \Big(\ln m_0 \ln m_1 + \ln m_1 \ln m_2 + \ln m_2 \ln m_0 - \ln^2 m_0 - \ln^2 m_1 - \ln^2 m_2\Big)^2. \quad (4.19)$$

4.2.3 Moments of compound process

From equation (4.8) and the absolute moments of Gaussian distribution, we have the moment of absolute returns as a compound process $E[|x_t|^q] = E[|\sigma_t|^q]E[|u_t|^q]$, $q = 1, \dots 4$ as followed:

$$\begin{aligned} E[|x_t|] &= \frac{2\sigma}{\sqrt{2\pi}}\Big(\frac{1}{3}m_0^{1/2} + \frac{1}{3}m_1^{1/2} + \frac{1}{3}m_2^{1/2}\Big)^{\bar{k}} \\ E[|x_t|^2] &= \sigma^2 \\ E[|x_t|^3] &= \frac{4\sigma^3}{\sqrt{2\pi}}\Big(\frac{1}{3}m_0^{3/2} + \frac{1}{3}m_1^{3/2} + \frac{1}{3}m_2^{3/2}\Big)^{\bar{k}} \\ E[|x_t|^4] &= 3\sigma^4\Big(\frac{1}{3}m_0^{2} + \frac{1}{3}m_1^{2} + \frac{1}{3}m_2^{2}\Big)^{\bar{k}}. \end{aligned} \quad (4.20)$$

When estimating the parameters of MSMF model, one of the main tasks is to filter this integrated component as standard deviation of Gaussian distribution out from the compound process. Due to equation (4.20) the variance of the compound process $|x_t|$ results in

$$\begin{aligned} Var[|x_t|] &= E\big[|x_t|^2\big] - \Big(E\big[|x_t|\big]\Big)^2 \\ &= \sigma^2\left[1 - \frac{2}{\pi}\left(\frac{1}{3}m_0^{1/2} + \frac{1}{3}m_1^{1/2} + \frac{1}{3}m_2^{1/2}\right)^{2\bar{k}}\right]. \end{aligned} \quad (4.21)$$

As equation (4.20) shows, the volatility of multi-fractal process and the scaling factor are integrated into the variance σ^2. It means that the volatility generated by the multiplication of discrete components can be replaced by the one that is integrated over a continuous measure.

Hence, the discrete MSMF model can be considered an approximation of the continuous log-normal MSMF model [Lux (2006)] We note that a discrete MSMF process can generate arbitrary volatility cycles with the switching mechanism that also give abrupt volatility changes. Implementing a Markov-switching mechanism in the multiplicative structure of a multi-fractal process, the MSMF process combines 'apparent' long memory behavior with sudden volatility movements [Calvet and Fisher (2004)].

We point out that it should be distinguished between 'true' long memory and 'apparent' long memory. The MSMF processes are characterized by only 'apparent' long memory with an asymptotic hyperbolic decline of the autocorrelation of absolute powers over a finite horizon and exponential decline thereafter [Lux (2006)]. Therefore, this do not obey the traditional definition of long memory, i.e. asymptotic power-law behaviour of autocovariance functions in the limit $t \to \infty$ or divergence of the spectral density at zero (cf. Beran (1994)).

4.3 Generalized method of moments

After we have derived the analytical moments of trinomial MSMF process, we will employ the generalized method of moments (hereafter, GMM[7]) to estimate the parameters of trinomial MSMF model. GMM has a number of advantages over other estimation methods. First, GMM does not require the distribution of the underlying process to be the Gaussian distribution. Note that the distribution of financial returns and MSMF processes are leptokurtic. The asymptotic justification for the GMM procedure only needs that the distribution of the process is covariance stationary and ergodic. In other words, relevant moments must exit. Second, the GMM estimates and their standard errors are consistent even if the disturbances are conditionally heteroskedastistic. Since the volatility of a multiplicative process can not be easily distinguished from the distributions of disturbances, GMM should remove the impact of this approximation error on the parameter estimates as possible.[8]

In the following subsections we briefly introduce GMM[9] and we are going to discuss the applicability of this method to trinomial MSMF processes, too. Using the pertinent moments, we excute Monte Carlo experiments to assess the perperty of the GMM estimates for the parameters of trinomial MSMF model and finish with an evaluation of the Monte Carlo results.

4.3.1 Brief introduction to GMM

GMM can be considered of just as a generalization of the classical method of moments. A central statement of GMM is to use a set of population moment conditions that are drived from the assumption of the underling econometric model. Given data on the obserable variables the GMM finds estimates for the model parameters such that corresponding sample moment conditons are satisfied as closely as possible. We introduce GMM briefly.

Let x_t be an s x 1 vector of variables that are observed at date t, let θ denote the m x 1 unknown parameter vector, and let $g_t = g(x_t, \theta)$ be an r x 1 covariance stationary vector valued function, such that for true parameter value θ_0

$$E[g_t] = E[g(x_t; \theta_0)] = 0 \tag{4.22}$$

[7] see Hansen (1982)

[8] See Mittelhammer, Judge, and Miller (2000) for details.

[9] See Mittelhammer et al. (2000) and Campbell, Lo, and A.C. (1997) for the theoretical explanation. We follow the approach of Lux (2006) with repect to the MSMF model

In GMM function $g(x_t, \theta)$ define the moment conditions of the model. A sample conterpart of the expected value in (4.22) is the sample average

$$f_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} g(x_t; \theta_0). \tag{4.23}$$

Note that for the true parameter value θ_0, $f_T(\theta_0)$ measures the average sampling error with $E[f_T(\theta_0)] = 0$. An estimator of θ is selected such that $f_T(\theta)$ becomes as close as possible to zero. In general, there are more moment conditions than estimated parameters the estimator for θ_0 is a compromise that makes (4.23) close to zero. We get this in the GMM by selecting the estimator for θ_0 such that the sampling error with respect to the estimated value is as small as possible in the generlized least squares sense.

Objective function

The optimal value $\hat{\theta}$ of parameters is obtained from the minimization of the quadratic function

$$Q_T(\theta) = f_T(\theta)'\, W_T(\theta)\, f_T(\theta) \tag{4.24}$$

where $f_T(\theta)'$ stands for the transpose of $f_T(\theta)$ and $W_T(\theta)$ is positive definite and possibly random weighting matrix.

First-order condition

Since the objective function Q_T is non-linear, the minimization must be performed numerically. The first-order condition for the minimization is

$$D_T(\hat{\theta})' W_T(\hat{\theta}) f_T(\hat{\theta}) = 0, \tag{4.25}$$

where $D_T(\theta)$ is the matrix of a partial derivative:

$$D_T(\theta) = \frac{\partial f_T(\theta)}{\partial \theta}. \tag{4.26}$$

For the unrestricted model, the parameters are just identified and the object function attains zero for any choice of $W_T(\theta)$. In the trinomial MSMF model, the count of the parameters are restricted and the number of derived moments are more than the number of parameters to be estimated. Hence, the GMM estimation of the over-identified θ does depend on the choice of the matrix $W_T(\theta)$. The weighting matrix determines how each moment condition is weighted in the estimation. As a matter of principle more accurate moment conditions should be weighted nore than less accurate ones. The accuracy of the moment conditions can be measured by the variance covariance matrix

$$\begin{aligned} Cov[f_T(\theta)] &= E[f_T(\theta) f_T(\theta)'] \\ &= \frac{1}{T^2} E\left[\Big(\sum_{t=1}^{T} g_t \Big) \Big(\sum_{t=1}^{T} g_t \Big)' \right] \\ &= \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} E[g_t g_t'] \end{aligned} \tag{4.27}$$

Under the assumption of stationarity of g_t we let $j = |t-s|$ and then get $E[g_t g_t'] = S_j(\theta) = S'_{-j}(\theta)$ for all t. Thus in (4.27) we can write

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} E[g_t g_t'] \quad = \quad \sum_{j=-T}^{T}(T-|j|)S_j(\theta), \tag{4.28}$$

or by using $E[g_t g_t'] = S_j(\theta) = S'_{-j}(\theta)$

$$\frac{1}{T^2}\sum_{s=1}^{T}\sum_{t=1}^{T} E[g_t g_t'] \quad = \quad TS_0(\theta) + \sum_{j=1}^{T}(T-j)(S_j(\theta) + S_j'(\theta)). \tag{4.29}$$

Under the stationarity and some technical assumptions(see Hansen (1982)), it can be shown that

$$\lim_{k\to\infty}\sum_{j=-k}^{k} S_j(\theta) = S(\theta), \tag{4.30}$$

where $S(\theta)$ is a positive definite matrix, called the long run covariance matrix of g_t. Thus, because $(T-|j|)/T \to 1$ as $T \to \infty$

$$Cov[\sqrt{T}g_T] \to \sum_{j=-\infty}^{\infty} S_j(\theta) = S(\theta), \tag{4.31}$$

as $T \to \infty$. To estimate $S(\theta)$ it is practical to replace the true autocovariance matrices $S_j(\theta)$ with sample autocovariances

$$\hat{S}_j(\hat{\theta}) = \frac{1}{T}\sum_{t=j+1}^{T} \hat{g}_t \hat{g}'_{t-j}, \tag{4.32}$$

$j = 0, 1, \ldots, l$, where l is the selected maximum lag length. The long-run covariance matrix is estimated by

$$\hat{S}_j(\hat{\theta}) = S_0(\hat{\theta}) + \sum_{j=1}^{l}\omega_j(\hat{S}_j(\hat{\theta}) + \hat{S}_j(\hat{\theta})', \tag{4.33}$$

where ω_js are weights. If $\omega_j \equiv 1$ then all lag lengths are equally weighted. But in practice the more near lags are weighted more and vice versa. One popular weighting scheme is suggested by Newey and West (1987):

$$\omega_j = 1 - \frac{j}{l+1}. \tag{4.34}$$

To obtain the estimate θ, we use two step procedure:

Step 1: Set the vector of starting values, θ_0
W is set equal to the identity matrix and then estimate the parameters θ

$$\hat{\theta}_i := arg\ min_{\theta\in\Theta} f(\theta)'\, W_I(\theta)\, f(\theta), \tag{4.35}$$

where Θ denotes the parameter space. Step 2: Compute $\hat{g}_t = g(x_t; \hat{\theta}_i$ and estimate $\hat{S}_j$ as

$$\hat{S}_j = \frac{1}{T}\sum_{t=j+1}^{T} \hat{g}_t \hat{g}_{t+j}, \tag{4.36}$$

$j = 0, 1, \ldots, l.$ and estimate S by

$$\hat{S}_j = \hat{S}_0(\hat{\theta}) + \sum_{j=1}^{l} \omega_j (\hat{S}_j + \hat{S}_j'), \tag{4.37}$$

We select $W = \hat{S}^{-1}$ and obtain the second step estimate

$$\hat{\theta}_i := arg\ min_{\theta\in\Theta} f(\theta)'\, W(\theta)\, f(\theta), \tag{4.38}$$

These steps are repeated until convergence is obtained: $max|\hat{\theta}_i - \hat{\theta}_{i-1}| \leq \epsilon$.

For the MSM model we consider the following moment conditions

$$E[\xi_{t+\Delta t,\Delta t}\xi_{t,\Delta t}] \tag{4.39}$$
$$E[\xi^2_{t+\Delta t,\Delta t}\xi^2_{t,\Delta t}], \tag{4.40}$$

where $\Delta t = 1, 5, 10, 20$. Here $\xi_{t,\Delta t}$ is represented either by the log difference of the absolute returns or by the squared log-difference:

$$\xi_{t,\Delta t} = \ln|x_t| - \ln|x_{t-\Delta t}| \tag{4.41}$$

or

$$\xi_{t,\Delta t} = \ln x_t^2 - \ln x_{t-\Delta t}^2 \tag{4.42}$$

The $\xi_{t,\Delta t}$ of equation (4.41) and the analytical moment conditions are used to fulfil the requirement of stationarity.[10] θ is the parameter vector with elements m_0, m_1, and σ for the trinomial MSMF model:

$$\theta = \{m_0,\, m_1,\, \sigma\} \tag{4.43}$$

Since the moment conditions for the trinomial MSMF model parameters m_0, m_1 are independent of σ, the covariance matrix of the parameters should be block-diagonal and estimated values of σ should be essentially identical to the sample standard deviation.

[10] Some regularity conditions are, for example, discussed, in Harris and Matyas (2000).

Asymptotic distribution

Designating the covariance matrix $S_0(\theta)$, the asymptotic variance-covariance matrix Ω for the estimation of θ is then given by

$$\Omega = (D_0'(\theta)S_0^{-1}(\theta)\, D_0(\theta))^{-1}. \tag{4.44}$$

Here, $D_0(\theta)$ is the Jacobian matrix that is evaluated by using the estimated values of the parameters. This variance-covariance matrix is used to test the significance of the individual parameters. The GMM estimates $\hat{\theta}$ are consistent and asymptotically normal distributed:

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega). \tag{4.45}$$

The values satisfying equation (4.24) follow the chi-square distribution. The degree of freedom of this distribution is determined by the number of orthogonality conditions minus the number of estimated parameters. The χ^2 measure provides a goodness-of-fit test for the model. A high value of this statistic means that the model is misspecified.

4.3.2 Applicability of the GMM to MSMF

The applicability of the GMM to the trinomial MSMF dependens on whether the GMM estimates meet the conditions of consistency and asymptotic normality. The weak consistency in Harris and Matyas (2000) can be observed if and only if

1) $E[f(\theta)]$ exists for all θ and is finite.

2) There is a θ_0 such that $E[f(\theta)] = 0$ if and only if $\theta = \theta_0$.

3) The difference between average sample moment and population moment in probability converges uniformly to zero. That means, $E[f(\theta)]$ satisfies a weak law of large numbers.

4) The sequences of weighting matrices $W(\theta)$ converge in probability to a constant covariance matrix.

If, instead of *in probability*, we use *almost surely* in conditions 3) and 4), we arrive at the condition for strong consistency. Except for the above first and second conditions, the following conditions are required for asymtotic normality:

3-1) The difference between the average sample and population moments *almost surely* converges uniformly to zero. i.e., $E[f(\theta)]$ satisfies a strong law of large numbers.

4-1) The sequences of weighting matrices converges *almost surely* to a constant covariance matrix.

5) $E[f(\theta)]$ needs to be continuously differentiable.

6) The matrix of the first derivatives $\partial f(\theta)/\partial\theta$ should converge to a constant matrix $\partial \bar{f}(\theta)/\partial\theta_0$ for $\theta \overset{p}{\sim} \theta_0$.

7) $f(\theta)$ now needs to satisfy a central limit theorem.

There are two diffculties to overcome when we will apply GMM to multi-fractal process: The first one is non-stationarity. The other is the property of long range dependence. For example,

the fourth $(4-1)$ and last (7) conditions have caused problems by GMM application to MMAR [Mandelbrot et al. (1997), Calvet and Fisher (2002) and Lux (2006)]. Note that the process of MMAR is non-stationary. This problem of non-stationarity is easily removed by implementing the first order Markov chain in the multi-fractal process. The long range dependence may violate the necessary rapidness of convergence particularly with respect to the autocorrelation. We discuss the suitable solution for these difficulties.

Non-stationary

To resolve the problem of non-stationarity, Bacry, Delour, and Muzy (2001) suggest a multi-fractal random walk. Hereby it is supposed that a number of underlying volatility components is fixed.[11] Above the bounded number of cascade, the process behaves as a standard Brownian motion. It is a mixture of a multi-fractal process with finite cascade and a simple Brownian motion. The compound process contains stationary increments, but a bounded range of autocorrelation as soon as $\bar{k}$ is fixed.

Inspired by Breymann, Ghashgaie, and Talckner (2000), Lux (2003) suggests to replace the non-causal construction, i.e. the combinatorial and bounded cascade by an iterative cascade: the volatility components $m_{t,k}$ are replaced over time by new multipliers $m_{t+1,k}$ with the predetermined probabilities. This approach also conserves the hierarchical nature of the volatility process. It allows for iterative (stochastic) changes of individual components over time. It is implicitly assumed that the switching takes place randomly. This process is hence named the random-switching multi-fractal process.

Instead of the prearranged probabilities between levels of cascade, Calvet and Fisher (2002) have used a continuous-time Poisson process as a regulator of the replacement of multipliers and call it as Poisson multi-fractal process.

As an alternative to Poisson multi-fractal process, Calvet and Fisher (2004) propose a stationary Markov switching multi-fractal process using ergodic Markov chains. A MSMF process converges asymptotically to a finite distribution with $E[x_t] = 0$ and $Var[x_t] = E[x_t^2] < \infty$. As we will see later, the MSMF model is a novel approach that preserves the hierarchical nature of the volatility process and the long range dependence. It enables us to estimate the parameters by means of the maximum likelihood method in the framework of regime-switching model.

Long range dependence

Let's turn to the problem of long range dependence. In general, the autocorrelation function (ACF) for every moment q is given by $Cor\left(|x_t|^q, |x_{t+\delta t}|^q\right) \sim \delta t^{-\delta(q)}$ for $1 < t < b^k$ (Calvet and Fisher (2004)). Here, $\delta(q) = log_b(E[M_t^q/[E(M_t^{q/2})]^2)$ and M_t is multi-fractal measure. $b = 3$ for trinomial MSMF. As in Figure (4.5) and Figure (4.6) exhibited, the MSMF autocorrelation function declines hyperbolically that has various degree of long range dependence. This property is, however, valid only in a finite time range $1 < t < b^{\bar{k}}$. Beyond this time horizon, the autocorrelation of MSMF decays exponentially.[12]

Given the long memory features of MSMF process, Lux (2006) notes that poor results might be obtained according to the large asymptotic variation of the covariance matrix when one applies GMM to MSMF model with finite samples. As a practical solution to this problem, Lux (2006) suggests to use log differences of absolute returns. Alternatively, the log differences of squared

[11] Note that the compound process of multi-fractal random walk has fatter tails than Gaussian.

[12] It will be interesting to compare the autocorrelation of a MSMF model with that of multi-fractal random walk process in Bacry et al. (2001). It might be not easy to identify the zero-crossing point because of the noisiness of the autocorrelation function at long time lags even if there are a bounded correlation length [Lux (2006)]. As we can confirm by two figures (4.5) and (4.6), it is hardly possible to identify such point with empirical data.

returns can also be applied. As shown in the Appendix A, log-differences of the multi-fractal process yield a stationary stochastic process. Hence, we are now able to apply GMM for such transformed process without violating necessary conditions.

4.3.3 GMM estimation

In this section, we present two kinds of moments of the transformed process[13] that are implemented as the moment condition in GMM: log-differences of absolute returns and log differences of squared returns.

Log differences of absolute returns

Let's denote log differences of absolute returns[14] by $\xi_{t,\Delta t}$. Due to the assumption of independence between Markov switching process and fractional Gaussian noise process, we can rewrite $\xi_{t,\Delta t}$:

$$\begin{aligned}\xi_{t,\Delta t} &= \ln|x_t| - \ln|x_{t-\Delta t}| \\ &= \ln\left(\left|\sqrt{\prod_{k=1}^{\bar{k}} m_{t,k}}\,\sigma\, u_t\right|\right) - \ln\left(\left|\sqrt{\prod_{k=1}^{\bar{k}} m_{t-\Delta t,k}}\,\sigma\, u_{t-\Delta t}\right|\right) \\ &= \frac{1}{2}\sum_{k=1}^{\bar{k}}\left(\ln m_{t,k} - \ln m_{t-\Delta t,k}\right) + \ln|u_t| - \ln|u_{t-\Delta t}| \end{aligned} \tag{4.47}$$

As shown in Lux (2004), this process has non-zero autocovariances at the first time lag $\Delta t = 1$. With all the entries on the right hand side stemming from Gaussian noise u_t and $u_{t-\Delta t}$, the regularity conditions are satisfied so that GMM is applied to estimate the parameters of MSMF model.

Derivation of Moments

The moments of the compound process are made from the part of the log (absolute) multi-fractal process and that of the log absolute innovations. Following to equation (A.8) and equation (4.47), we have the first and second autocovariances of the compound process:

[13] To obtain a stationary process from a non-stationary one, we usually build the differences of the process. In some cases, we transform a process to a logarithm. It is also possible to combine logarithm with difference.

[14] This transformation reminds us the volatility of volatility (VOV) proposed by Engle (2002). He defines the standard deviation of the volatility of volatility (VOV):

$$VOV = \sqrt{Var[(\ln v_{T,t} - \ln v_{T,t-1})]} \approx \sqrt{Var[E(\ln \xi_{t,\Delta t}) - E(\ln \xi_{t,\Delta t-1})]}, \tag{4.46}$$

where v_t is an instaneous volatility and it is calculated unconditionally over all returns. He argues that this measure can be used for implied volatilities which can be interpreted as the square root of a forecast of variance from today to the expiration of the option at Δt. The concept of the VOV is related to our squared ξ_t. Like implied volatility in option theory, the variabilty of volatility in our model is composed of two part: VOV of original multi-fractal volatility and VOV of Gaussian innovation. By filtering the white noise VOV from the multi-fractal VOV, we can easily calculate the variability of volatility.

$$
\begin{aligned}
E[\xi_{t+\Delta t,\Delta t}\xi_{t,\Delta t}] &= E\Bigg[\bigg(\frac{1}{2}\sum_{k=1}^{\bar{k}}\big(\ln m_{t+\Delta t,k}-\ln m_{t,k}\big)+\ln|u_{t+\Delta t}|-\ln|u_t|\bigg)\cdot \\
&\qquad \bigg(\frac{1}{2}\sum_{k'=1}^{\bar{k}}\big(\ln m_{t,k'}-\ln m_{t-\Delta t,k'}\big)+\ln|u_t|-\ln|u_{t-\Delta t}|\bigg)\Bigg] \\
&= \frac{1}{4}E[\eta_{t+\Delta t}\eta_{t,\Delta t}]+\Big(E[\ln|u_t|]\Big)^2-E\Big[(\ln|u_t|)^2\Big]. \qquad (4.48)
\end{aligned}
$$

and

$$
\begin{aligned}
E[\xi^2_{t+\Delta t,\Delta t}\xi^2_{t,\Delta t}] &= E\Bigg[\bigg(\frac{1}{2}\sum_{k=1}^{\bar{k}}\big(\ln m_{t+\Delta t,k}-\ln m_{t,k}\big)+\ln|u_{t+\Delta t}|-\ln|u_t|\bigg)^2\cdot \\
&\qquad \bigg(\frac{1}{2}\sum_{k'=1}^{\bar{k}}\big(\ln m_{t,k'}-\ln m_{t-\Delta t,k'}\big)+\ln|u_t|-\ln|u_{t-\Delta t}|\bigg)^2\Bigg] \\
&= \frac{1}{16}E[\eta^2_{t+\Delta t}\eta^2_{t,\Delta t}]-\Big\{E[\eta^2_{t,\Delta t}]-E[\eta_{t+\Delta t,\Delta t}\eta_{t,\Delta t}]\Big\}\cdot \\
&\quad \Big\{\big(E[\ln|u_t|]\big)^2-E\big[(\ln|u_t|)^2\big]\Big\}+3E\big[(\ln|u_t|)^2\big]^2 \\
&\quad -4E[\ln|u_t|]\big[E\big[(\ln|u_t|)^3\big]+E\big[(\ln|u_t|)^4\big]. \qquad (4.49)
\end{aligned}
$$

Explicit derivation of equation (4.48) and equation (4.49) is found in the appendix A. The moments of log-transformed volatility $E[\eta^2_{t,\Delta t}]$, $E[\eta_{t+\Delta t}\eta_{t,\Delta t}]$ and $E[\eta^2_{t+\Delta t,\Delta t}\eta^2_{t,\Delta t}]$ are given in equation (4.16), equation (A.13) and equation (A.36). The scaling parameter σ that is excluded from equation (4.48) and equation (4.49) can be estimated by means of the unconditional moments of returns in equation (4.20) The determination of the optimal $\bar{k}$ will be subject-matter of model selection, see the section 4.6. For simulations as well as estimations, we assign $\bar{k}=3,4,5$.

Log differences of squared returns

As an alternative transform, we suggest to use log differences of squared returns. Since squared returns are often used as a proxy for volatility, it is also interesting to compare estimates with different transforms: log differences of absolute returns and log differences of squared returns. Let us define ξ by the squared log-difference. As an alternative proxy for volatility, we Substituting the squared and then log-transformed x_t of

$$
\begin{aligned}
\ln x_t^2 &= \ln\sigma_t^2+\ln u_t^2 \\
&= \ln\sigma^2+\sum_{k=1}^{\bar{k}}\ln m_{t,k}+\ln u_t^2 \qquad (4.50)
\end{aligned}
$$

into equation (4.42), we obtain

$$
\xi_{t,\Delta t}=\sum_{k=1}^{\bar{k}}\Big\{\ln m_{t,k}-\ln m_{t-\Delta t,k}\Big\}+\ln u_t^2-\ln u_{t-\Delta t}^2 \qquad (4.51)
$$

The transformed volatility model of equation (4.50) consists of the log volatility components $\ln m_{t,k}$ as the latent regime variables and the essential ingredient term $\ln\sigma^2$. Due to the assumption

of independence between the multi-fractal process and the standard Gaussian noise, we can easily separate the moments of transformed process into the moments of the log squared innovations and that of log standard Gaussian. As stated in the subsection 4.3.2, the process described by equation (4.51) is stationary and has no long memory so that the GMM method for the parameter estimation can be applied. The autocavariances of this transformed process and of the squared process are straightforward. Note that the moment conditions of the log differences of squared returns are in principle the same as the log differences of absolute returns.

4.3.4 Monte Carlo performance of GMM estimates

This section shows the results of Monte Carlo simulations which are designed to assess small sample properties of the GMM estimates. We estimate the values of the multipliers m_0, m_1 and the unconditional standard deviation σ. We set the maximal number of cascade $\bar{k} = 5$ for simplicity. The switching probability of multipliers is specified by $\gamma_k = 1 - (1 - 1/3)^{3^{k-\bar{k}}}$.

The aim of our experiments is to examine the efficiency of the GMM estimates. Moreover, we compare our GMM results with the Monte Carlo result of ML approach developed by Calvet and Fisher (2004). Therefore, we use the same setting of Monte Carlo simulations reported in the paper of Calvet and Fisher (2004). Note that Calvet and Fisher used the binomial specification in their work. The simulations are designed as follows: we choose three samples of sizes $n =$ 2500, 5000, 10000. As a whole, we conduct the simulations with six different combinations of parameters which are combined into two groups as follows:

- $(m_0,\, m_1,\, \sigma) \in \{1.2,\, 0.5,\, 1\}, \{1.3,\, 0.6,\, 1\}, \{1.4,\, 0.7,\, 1\}$
- $(m_0,\, m_1,\, \sigma) \in \{1.1,\, 0.9,\, 1\}, \{1.2,\, 0.8,\, 1\}, \{1.3,\, 0.7,\, 1\}$

We assume that the selected input values lead to results that can depict the main property of trinomial MSMF process.[15] In the first set of parameters, the implied parameter $m_2 = 3 - m_0 - m_1$ takes one value among $\{1.3,\, 1.1,\, 0.9\}$, while in the second one m_2 is fixed as one. The differenes between $max\,(m_0, m_1, m_2)$ and $min\,(m_0, m_1, m_2)$ are much larger in the former case than in the latter one. The value of parameters are confined in such a way that $1 < m_0 < 3$ and $0 < m_1 < 1$. The choice of bounds can also be viewed as model selection. But, it does not limit the performance of GMM to trinomial MSMF. As to every set of m_0, m_1 and σ with sample length n, we simulate 400 independent sample paths. GMM estimation provides the estimates $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$ and we calculate finite sample standard error (FSSE) and root mean squared error (RMSE)[16]. The time lag is assigned by $\Delta t = 1,\, 5,\, 10,\, 20$, where the large time intervals such as $\Delta t = 10$ and $\Delta t = 20$ serve to exploit the long-term dependent properties of the multi-fractal process.

The results using the autocovariances of equation (4.48) and equation (4.49) is shown in the tables (4.1) and (4.2), whereas the tables of (4.3) and (4.4) are obtained by employing the moment conditions of the log differences of squared returns. The applied autocovariance gives informations over the temporal dependency that are fully used by GMM. The tables shows the simulated values of parameters together with the finite sample standard error (FSSE) as well as the root mean squared error (RMSE). Note that the RMSE will always be greater than or equal to the FSSE. The root mean squared error is equal to the FSSE only if the average point estimate is identical to the true parameter value.

An quick inspection of the tables (4.1) and (4.2) reveal that the GMM estimates are unbiased. Not only the two parameters m_0 and m_1 are mainly determining the variability of volatility dynamics, but also σ has low standard errors relative to its size. The FSSE and the RMSE are as

[15] Remember that the trinomial MSMF model is reduced to the binomial one if $m_0 = m_1 = 0.5$. And we get a Gaussian noise for $m_0 = m_1 = 1$.

[16] We start the optimization at the true parameter values and iterate to convergence once.

expected to become smaller with growing sample size. As a whole, GMM estimation produces reliable results for given sample sizes. The results prove that the derived moments are correct.

Comparing the estimates $\hat{m}_0$ and $\hat{m}_1$ with the log difference of absolute returns, the estimated m_0 and m_1 in tables (4.3) and (4.4) have higher standard errors in every sample size. The standard errors are roughly 10% of the true parameter values. The increase of FSSE can be caused by relatively high uncertainty about the first and the second moments of log-squared innovation $\ln u_t^2$. The average of log-squared Gaussian variate, $E[\ln u_t^2]$ amounts to -1.27036, which is exactly two times than the log absolute Gaussian innovation $E[\ln |u_t|] = -0.635181$. Furthermore, the second monent of $E[(\ln u_t^2)^2] = 6.54862$ is four times larger than the $E[(\ln |u_t|)^2] = 1.63716$. That means, GMM exploits more efficient for the transformation of log increments of absolute returns than the other transformation. The point is that the moment conditions are in principle the same, but the GMM estimation using the log increments of squared returns produces more sampling variability. Insofar it is interesting to repeat these results. In contrast to m_0 and m_1, the standard error of σ becomes in this case smaller than the results using absolute returns and is getting to be negligible as the sample size increases. As a result, the parameters can be better identified if we use the log increments of absolute returns.

Monte Carlo GMM results with trinomial MSMF model (Ia)

Estimation of parameters m_0, m_1 and σ with $k = 5$									
m_0	1.2			1.3			1.4		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_0$	1.219	1.216	1.210	1.277	1.276	1.284	1.365	1.382	1.385
FSSE	0.087	0.079	0.082	0.085	0.071	0.060	0.079	0.059	0.038
RMSE	0.090	0.080	0.083	0.088	0.075	0.062	0.086	0.062	0.041
m_1	0.5			0.6			0.7		
$\hat{m}_1$	0.501	0.499	0.502	0.610	0.606	0.604	0.696	0.696	0.697
FSSE	0.052	0.045	0.029	0.038	0.028	0.027	0.051	0.045	0.035
RMSE	0.052	0.045	0.029	0.039	0.028	0.027	0.051	0.046	0.035
σ	1								
$\hat{\sigma}$	0.985	0.995	0.997	0.987	0.994	0.997	0.995	0.994	0.994
FSSE	0.081	0.056	0.039	0.069	0.050	0.037	0.074	0.056	0.039
RMSE	0.083	0.056	0.039	0.070	0.050	0.037	0.074	0.057	0.039

Table 4.1: Simulated Biases, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for GMM estimates. The value of parameters are assigned by $m_0 = 1.2$, 1.3, 1.4 and $m_1 = 0.5$, 0.6, 0.7. The scaling factor $\sigma = 1$. For estimation of σ we use the absolute moments $E[|x_t|^q]$, $q = 1$, 2, 3, 4.

Monte Carlo GMM results with trinomial MSMF model (Ib)

Estimation of parameters m_0, m_1 and σ with $k = 5$									
m_0	1.1			1.2			1.3		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_0$	1.085	1.088	1.088	1.166	1.176	1.182	1.259	1.270	1.277
FSSE	0.043	0.040	0.032	0.065	0.052	0.043	0.072	0.060	0.048
RMSE	0.045	0.042	0.034	0.073	0.057	0.046	0.083	0.067	0.054
m_1	0.9			0.8			0.7		
$\hat{m}_1$	0.899	0.898	0.900	0.807	0.803	0.802	0.703	0.703	0.701
FSSE	0.043	0.035	0.027	0.039	0.033	0.026	0.043	0.035	0.027
RMSE	0.043	0.035	0.027	0.040	0.033	0.027	0.043	0.035	0.027
σ	1								
$\hat{\sigma}$	0.997	0.998	0.999	0.998	0.996	0.997	0.989	0.993	0.997
FSSE	0.026	0.018	0.012	0.043	0.030	0.021	0.061	0.044	0.029
RMSE	0.026	0.018	0.013	0.043	0.031	0.022	0.062	0.044	0.030

Table 4.2: Simulated Biases, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for GMM estimators. The value of parameters are assigned by $m_0 = 1.1$, 1.2, 1.3 and $m_1 = 0.9$, 0.8, 0.7. The scaling factor $\sigma = 1$. For estimation of σ we use the absolute moments $E[|x_t|^q]$, $q = 1$, 2, 3, 4.

Monte Carlo GMM results with trinomial MSMF model (IIa)

Estimation of parameters m_0, m_1 and σ with $k = 5$									
m_0	1.2			1.3			1.4		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_0$	1.220	1.223	1.222	1.273	1.279	1.270	1.313	1.316	1.336
FSSE	0.119	0.122	0.125	0.218	0.204	0.191	0.236	0.211	0.198
RMSE	0.120	0.124	0.127	0.220	0.206	0.193	0.251	0.227	0.208
m_1	0.5			0.6			0.7		
$\hat{m}_1$	0.563	0.545	0.538	0.653	0.650	0.652	0.747	0.752	0.736
FSSE	0.168	0.138	0.133	0.161	0.143	0.134	0.161	0.144	0.123
RMSE	0.179	0.145	0.138	0.169	0.151	0.144	0.167	0.153	0.129
σ	1								
$\hat{\sigma}$	0.987	0.988	0.987	0.977	0.974	0.976	0.976	0.978	0.979
FSSE	0.085	0.063	0.059	0.087	0.077	0.075	0.098	0.084	0.074
RMSE	0.086	0.064	0.061	0.090	0.081	0.078	0.101	0.087	0.077

Table 4.3: Simulated Biases, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for GMM estimates. The value of parameters are assigned by $m_0 = 1.2, 1.3, 1.4$ and $m_1 = 0.5, 0.6, 0.7$, respectively. The scaling factor $\sigma = 1$. For estimation of scaling parameter, we use the absolute moments $|x_t|^q|$, $q = 1, 2, 3, 4$.

Monte Carlo GMM results with trinomial MSMF model (IIb)

Estimation of parameters m_0, m_1 and σ with $k = 5$									
m_0	1.1			1.2			1.3		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_0$	1.113	1.112	1.103	1.193	1.172	1.179	1.253	1.255	1.277
FSSE	0.163	0.149	0.133	0.181	0.165	0.136	0.197	0.185	0.157
RMSE	0.163	0.150	0.133	0.182	0.168	0.137	0.202	0.191	0.159
m_1	0.9			0.8			0.7		
$\hat{m}_1$	0.889	0.892	0.904	0.794	0.816	0.819	0.725	0.722	0.706
FSSE	0.141	0.118	0.107	0.155	0.126	0.097	0.147	0.123	0.096
RMSE	0.141	0.118	0.107	0.155	0.127	0.099	0.150	0.125	0.096
σ	1								
$\hat{\sigma}$	0.999	0.998	0.999	0.997	1.000	1.000	0.995	0.993	0.997
FSSE	0.028	0.016	0.014	0.044	0.029	0.022	0.054	0.040	0.030
RMSE	0.028	0.017	0.014	0.044	0.029	0.022	0.055	0.041	0.030

Table 4.4: Simulated Biases, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for GMM estimates. The value of parameters are assigned by $m_0 = 1.1, 1.2, 1.3$ and $m_1 = 0.9, 0.8, 0.7$, respectively. The scaling factor$\sigma = 1$. For estimation of scaling parameter we use the absolute moments $|x_t|^q$, $q = 1, 2, 3, 4$.

4.4 Maximum Likehood

Next, we are going to apply the maximum likehood method (ML) to estimate the parameters of trinomial MSMF. The Maximum likelihood method takes serial dependence of the hidden volatility states whose dynamics are regulated by the first-order Markov-chain process into account. Remember that the hidden hierarchical structure of states causes the long memory of trinomial MSMF.

Short history of regime switching processes

Originally, Baum, Petrie, Soules, and Weiss (1970) have proved mathematically that the likelihood always increases according to the Kullback-Leibler information criterion. Furthermore, they have established the convergence property of maximum likelihood estimator using the Kullback-Leibler divergence of two distributions. The Baum-Welch algorithm was applied to estimate regression model with Markov regime switching process by Lindgren (1978). Cosslet and Lee (1985) derived an algorithm of the non-linear filtering that gives an inference of the volatility state by means of a recurrence relation.

In the context of time series analysis, Hamilton (1989) introduced the iterative maximum likelihood procedures to estimate the model of autoregressive regime switching process.[17] Hamilton (1989) presented the so-called Hamilton filter that gives not only an inference of the (hidden) state at time t, but also enables a simple calculation of log likelihood function for maximum likelihood estimate. Calvet and Fisher (2004) suggest that the maximum likelihood method (ML) with the filtering methods can be employed to estimate the parameters of MSMF.

In this section, we first describe the maximum likelihood method with standard filtering and calculate the expected duration of the volatility states generated by the trinomial MSMF. We investigate small sample properties of the ML estimation based on the results of Monte Carlo simulation.

4.4.1 Quasi state-space model

The MSMF model can be rewritten in the *quasi* state space[18]

$$\begin{aligned} x_t &= \psi_t u_t \\ \psi_t &= A\psi_{t-1}, \end{aligned} \tag{4.52}$$

where innovations in measurement equation, u_t, are drawn from a standard Normal distribution N(0,1). Here, $\psi_{t-1} = \left(\psi_{t-1}^{(1)}, \psi_{t-1}^{(2)}, \dots \psi_{t-1}^{(d)}\right)$ and $\psi_t = \left(\psi_t^{(1)}, \psi_t^{(2)}, \dots \psi_t^{(d)}\right)$ stand for the volatility state vector at time $t-1$ and t, respectively. There is no additional innovation term for the transition equation $\psi_t = A\psi_{t-1}$. The maximal number of states is given by $d = 3^{\bar{k}}$ in trinomial MSMF.[19] We denote by u_t the unpredictable noise of the observed time series x_t and A denotes the transition matrix from state ψ_t into ψ_{t+1}. Formulating the MSMF in *quasi* state space model,

[17] See chapter 22 in Hamilton (1994) that offers a comprehensive discussion of the theory of Markov chains with application to Markov switching model.

[18] In the state space model, we formulate a process as a linear combination of a set of variables. A noise process is usually addded to the state space equation which consists of measurement equation and transition equation generally. However, we use the phrase "quasi state space model" because the trinomial MSMF process is taken to be a multiplicative combination of a set of state variables and a white noise.

[19] In general, we suppose that the volatility state ψ_t has a finite number $b_m^{\bar{k}}$ of values from each distribution. For instance, we have $2^{\bar{k}}_{m=\{m_0,m_1\}}$ for binomial MSMF, $3^{\bar{k}}_{m=\{m_0,m_1,m_2\}}$ for trinomial MSMF.

it is easy to see how we can visually separate the latent state ψ_t from the whole process x_t which is contaminated multiplicatively by noise.

The finite support of the volatility state in MSMF allows to derive the complete conditional distribution of the unobservable state variable. And the Hamilton filtering algorithm can be then engaged to calculate the optimal ψ_{t+1} on the basis of information set $I_t = (x_1, x_2, \ldots, x_t)'$. Here, $'$ denotes the transposed matrix. We present the inference algorithm to determine the filtered state probability $\psi_{t|t}$ [Calvet and Fisher (2004) and Krolzig (1997)]. In particular, we follow the paper of Krolzig (1997) and describe the procedure of filtering.

In the Markov switching model, the transition matrix contains too many coefficients to be estimated; the number of parameter n increases with $(d^2 - d)$. For example, the numbers of state d at $\bar{k} = 5$ are $3^5 = 243$ in trinomial MSMF. In this case, n amounts to 58806. We can not hope to estimate all these parameters efficiently from a finite set of observations. Therefore, we assume the constrained transition matrix A which is a $3^{\bar{k}} \times 3^{\bar{k}}$ matrix in trinomial MSMF. Specially, we will use equation (4.7) having the form of scaling properties in order to calculate the transition matrix.[20] By assuming that all parameters of the model are known, the conditional probability distribution of the volatility state can be written down as

$$\hat{\psi}_{t+1|t} \equiv E[\psi_{t+1}|I_t] = \begin{bmatrix} P(\psi_{t+1}^{(1)}|I_t) \\ P(\psi_{t+1}^{(2)}|I_t) \\ \vdots \\ P(\psi_{t+1}^{(d)}|I_t) \end{bmatrix}. \tag{4.53}$$

The equation (4.53) says that $\hat{\psi}_{t+1|t}$ can be interpreted as the conditional mean $E[\psi_{t+1}|I_t]$, which is the best prediction of ψ_{t+1} given I_t. Analogously, the filter inference $\hat{\psi}_{t|t}$ on the state vector ψ_t based only on currently available data is defined as:

$$\hat{\psi}_{t|t} = E[\psi_t|I_t] = \begin{bmatrix} P(\psi_t^{(1)}|I_t) \\ P(\psi_t^{(2)}|I_t) \\ \vdots \\ P(\psi_t^{(d)}|I_t) \end{bmatrix}. \tag{4.54}$$

The standard filtering algorithm calculates $\hat{\psi}_{t|t}$ by deriving the joint probability density of ψ_t and the x_t conditioned on available data I_t. By applying Bayes' law, the posterior probability $P(\psi_t|x_t, I_{t-1})$ are given by

$$P(\psi_t|x_t, I_{t-1}) \equiv P(\psi_t|I_t) = \frac{f(x_t|\psi_t, I_{t-1})P(\psi_t|I_{t-1})}{f(x_t|I_{t-1})} \tag{4.55}$$

[20] Calvet and Fisher (2004) specified this as followed: $A = \prod_{k=1}^{\bar{k}} [(1-\gamma_k)\mathbf{1}_{\{m_k^i = m_k^j\}} + \gamma_k P(m = m_k^j)]$, where m_k^i denotes the ith component of vector m. The dummy variable $\mathbf{1}_{m_k^i = m_k^j}$ is equal to one if the $m_k^i = m_k^j$. Otherwise it is zero. The constrained matrix is expressed as a known, or assumed function of only two parameters!

with the prior probability

$$P(\psi_t|I_{t-1}) = \sum_{\psi_{t-1}} P(\psi_t|\psi_{t-1})P(\psi_{t-1}|I_{t-1}) \tag{4.56}$$

and the density

$$f(x_t|I_{t-1}) = \sum_{\psi_t} f(x_t, \psi_t|I_{t-1}) = \sum_{\psi_t} P(\psi_t|I_{t-1})f(x_t|\psi_t, I_{t-1}). \tag{4.57}$$

Let f_t be the density vector of x_t conditional on ψ_t and I_{t-1}

$$f_t = \begin{bmatrix} f(x_t|\theta_1, I_{t-1}) \\ f(x_t|\theta_2, I_{t-1}) \\ \vdots \\ f(x_t|\theta_d, I_{t-1}) \end{bmatrix} = \begin{bmatrix} f(x_t|\theta_1, \psi_t^{(1)}, I_{t-1}) \\ f(x_t|\theta_2, \psi_t^{(2)}, I_{t-1}) \\ \vdots \\ f(x_t|\theta_d, \psi_t^{(d)}, I_{t-1}) \end{bmatrix}. \tag{4.58}$$

A vector of parameters θ is given in the trinomial MSMF by $\theta \in \{m_0, m_1, \sigma\}$. The conditional density $f(x_t|I_{t-1})$ is determined by

$$f(x_t|I_{t-1}) = f_t'\hat{\psi}_{t|t-1} = 1_d'(f_t \odot \hat{\psi}_{t|t-1}). \tag{4.59}$$

Then, the filter inference $\hat{\psi}_{t|t}$ is written in matrix notation by

$$\hat{\psi}_{t|t} = \frac{f_t \odot \hat{\psi}_{t|t-1}}{\mathbf{1}_d' f_t \odot \hat{\psi}_{t|t-1}}, \tag{4.60}$$

where $\odot$ denotes the element-wise matrix multiplication and $\mathbf{1}_d^{'} = (1, \ldots, 1)'$ is a unit vector. The filter $\hat{\psi}_{t|t}$ is a weighted probability that results from the ratio of the conditional density of the return x_t with the predicted probability for each regime to the sum of these conditional density of return multiplied by the predicted probability for each regime.[21] Thus, equation (4.60) describes the filtered regime probabilities as the updated estimate $\hat{\psi}_{t|t}$ of $\hat{\psi}_{t|t-1}$ given new information x_t.

On the other side, we can obtain the predicted filter probabilities $\hat{\psi}_{t+1|t}$ by means of the constrained transition matrix A and the contemporaneous filter inference since $\hat{\psi}_{t+1|t}$ is a linear function of the filtered probabilities $\hat{\psi}_{t|t}$, see Krolzig (1997) for details:

$$\hat{\psi}_{t+1|t} = A \cdot \hat{\psi}_{t|t}. \tag{4.61}$$

Equation (4.61) says that the vector of marginal probabilities at time t multiplied by the transition matrix A will give the vector for the next time $t+1$. Thus, the matrix A represents the pattern of change, whereas the expected marginal probability $\hat{\psi}_{t|t}$ refers the underlying trend of the process. Note that our constrained transition matrix is homogeneous because it has the same form in all periods.[22] The stationary marginal density is satisfied the steady-state condition: $\psi^\star = \psi^\star A$. The steady state arises after the process has run for sufficiently long time. Then, the

[21] Given the parameter vector θ of volatility state $\psi_t = \left\{\psi_t^{(1)}, \psi_t^{(2)} \ldots \psi_t^{(i)}\right\}$, $i = 1, 2 \ldots d$, we are able to calculate the filter with the predicted probability of being in state $\psi_t^{(i)}$ at time t given the information set I_{t-1}.

[22] The transition matrix of conditional probability of each state may be time-dependent and then non-homogeneous.

Markov chain process of the trinomial MSMF is also stationary. Finally, iterating the inference results in the filter sequence $\hat{\psi}_{t|t-1}$, $t = 1, 2 \dots T$:

$$\hat{\psi}_{t+1|t} = \frac{[f_t \odot A]\, \hat{\psi}_{t|t-1}}{1_d'[f_t \odot A]\, \hat{\psi}_{t|t-1}}. \tag{4.62}$$

In the Bayesian context, $\hat{\psi}_{t|t-1}$ is the prior distribution of ψ_t. The posterior distribution $\hat{\psi}_{t|t}$ is calculated by linking the new information x_t with the prior one through Bayes' law and becomes the prior distribution for the next state ψ_{t+1} and so on. Equipped with the above results we can derive the log-likelihood function as a by-product of the filter:

$$\ln L(\theta|X) = \sum_{t=1}^{T} \ln \hat{\psi}_{t|t-1}[f_t \odot A]1_d'. \tag{4.63}$$

Bayes' rule implies

$$\begin{aligned} \hat{\psi}_{t|t-1}[f_t \odot A]1_d' &= \sum_{i=1}^{d}\sum_{j=1}^{d} P(\psi_{t-1} = m^i, \psi_t = m^j | x_1, \dots, x_{t-1}) f(x_t | \psi_{t-1} = m^i, \psi_t = m^j) \\ &= \sum_{i=1}^{d}\sum_{j=1}^{d} \hat{\psi}^{(i)}_{t-1|t-1} a_{ij} f^{ij}(x_t), \end{aligned} \tag{4.64}$$

where we have the matrix $f^{ij}(x_t) \equiv f(x_t|\psi_{t-1} = m^i, \psi_t = m^j)$ for any return.

It should be noted that the maximisation of the log likelihood requires the solution of a non-linear optimisation problem. However, as we see in equation (4.63), it is not possible to find analytic solutions. Therefore, we employ the so-called numerical maximisation methods that integrate equation (4.63) to accomplish the numerical maximisation.[23]

For a simple Markov switching model, Lindgren (1978) derived the consistency and asymptotic distribution of the maximum likelihood estimates.[24] Leroux (1992) proved the consistency of the maximum likelihood estimates for general Markov switching models. Finally, Bickel, Ritov, and Ryden (1998) showed that the maximum likelihood estimation is consistent and asymptotically normal under suitable conditions. We can transfer the results to the trinomial MSMF since the constrained transition matrix enforces the parameter vector to fall almost surely into the allowable space. Especially, the ergodic probability of volatility states is not equal to zero so that the parameter will be very likely to converge in spite of huge number of states.

4.4.2 Expected duration of a state in MSMF model

How long does a volatility state $\psi_t^{(i)}$ lasts on the average? To obtain the required information on the expected duration of a volatility state, we need to calculate the diagonal elements of the transition matrix. Note that the conditional expected durations forms a well-known geometric

[23] For the numerical methods we use the module *constrained optimisation* in GAUSS.
[24] Because he did not estimate the transition probabilities.

Expected duration of volatility state in binomial and trinomial MSMF

k	S_{tri}	$E[D]$	# of data	S_{bi}	$E[D]$	# of data
1	3	9	27	2	4	8
2	9	6.731	60.579	4	2.779	11.116
3	27	6.208	167.616	8	2.435	19.48
4	81	6.052	490.212	16	2.298	36.768
5	243	6.001	1458.243	32	2.236	71.552
6	729	5.989	4363.065	64	2.207	141.248

Table 4.5: The expected duration gives the lasting time unit per a volatility state $\psi_t^{(i)}$. S_{bi} and S_{tri} denote the number of volatility states with binomial and trinomial measures, respectively. $E[D]$ refers the expected duration. # of a data presents minimal expected number of data points to be required.

distribution with mean[25]:

$$E[D] = \frac{1}{1 - P(\psi_{t|t-1})}. \tag{4.66}$$

The expected duration of each volatility state reflects the switching time units at each frequency k. We summerize the expected duration $E[D]$ in the trinomial MSMF in table (4.5), comparing it with the case of the binomial MSMF. Here, the minimal data points are given by the product of number of volatility states and the expected duration. According to equation (4.2) there are three volatility states m_0, m_1 and m_2 wenn $\bar{k} = 1$ which $E[D]$ is on average 9 time units. For $\bar{k} = 2$, each nine possible volatility state keeps 6.371 time units on average and so on.

4.4.3 Monte Carlo performance of ML estimates

Again, let's deal with the small-sample properties of the maximum likelihood estimation in the context of MSMF. For that we carry out the Monte Carlo simulation under the same setting that was used in GMM (see the section 4.3.4). The ML estimates $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$ were obtained employing GAUSS 7.0 and its module of constrained maximum likelihood as well as constrained optimization. We put the assigned values of parameters into the starting points.

The results are shown in the tables (4.6), (4.7), (4.8) and (4.9). The structure of the tables is the same as for the GMM. As shown in all tables, ML yields almost unbiased estimates $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$ for all cases. That is, ML estimates are not sensitive for any kind of parameter combinations including $\bar{k}$. The bias is not only negligible in every sample, but is also decreasing with growing sample size. In particular, the scaling parameter remains unbiased despite changing values of m_0 and m_1. Additionally, the bias of σ is much smaller than that from binomial MSMF of Calvet and Fisher (2004) and Lux (2004). As a result, the ML estimates are consistent and asymptotically efficient. The reported FSSE and RMSE are also reasonably small. For a low $\bar{k}$, the standard errors for σ fall significantly short in comparision with FSSE of m_0 and m_1.

A trinomial MSMF with low $\bar{k}$ behaves similarly to the increments of standard Brownian process if the multi-fractal parameters ($m_0 = 1.1$, $m_1 = 0.9$) take the value of approximated 1. Then, the esimates $\hat{m}_0$, $\hat{m}_1$ and the estimated σ can not be differentiated from each other. Note that this trinomial MSMF process is similar to the fractional Gaussian noise. It leads to higher FSSE and RMSE of the Monte Carlo results. However, the dynamics of this trinomial MSMF process becomes

[25]Kim and Nelson (1999) derived the expected duration of state such that:

$$E[D] = \frac{1 - P(\psi_{t|t-1})}{P(\psi_{t|t-1})} \sum_{h=0}^{\infty} hP(\psi_{t|t-1})^h = \frac{1}{1 - P(\psi_{t|t-1})} \tag{4.65}$$

obviously with increasing frequency. Then, it is not so difficult to distinguish between m_0 and m_1. For $\bar{k} = 5$, the bias, FSSE and RMSE are much smaller than for $\bar{k} = 2$. See the combinations in tables (4.7) and (4.9). In this case, the parameters m_0 and m_1 are well estimated with small FSSE and RMSE.

Smoothing probability

The smoothing probability is often used to identify a state or regime in the literature of Markov switching model. Given parameter estimates of the model, we can use full-sample information to make an inference about the unobservable states. Let us look at the smoothing probability in the trinomial MSMF in details. Related to the amount of available Information $I_\tau = (x_1, x_2, \ldots, x_\tau)$, we can rewrite equation (4.54) in a generalized form[26]:

$$\hat{\psi}_{t|\tau} = E[\psi_t|I_\tau] = \begin{bmatrix} P[\psi_t^{(1)}, |I_\tau] \\ P[\psi_t^{(2)}, |I_\tau] \\ \vdots \\ P[\psi_t^{(d)}, |I_\tau] \end{bmatrix}. \tag{4.67}$$

According to the bound of τ, the meaning of this probability is changed as follows:

- If $\tau = t$, $\hat{\psi}_{t|\tau}$ is *a filtering probability* derived from estimates for the underlying latent process with the given data $I_\tau = (x_1, \ldots, x_t\}$.
- When $\tau > t$ we obtain the *smoothing probability* which plays a central role in model selection based on Markov switching criteria. We focus on this smoothing probability. See Kim (1994) for details. Given parameter estimates of the model, we can get an inference about ψ_t and x_t based on all the information in the sample, $P[\psi_t^{(j)}|I_T], t = 1, 2, \ldots, T$. The smoothed joint probability, $P[\psi_t^{(j)}, \psi_{t+1}^{(i)}|I_T]$, based on full information is given by

$$P[\psi_t^{(j)}, \psi_{t+1}^{(i)}|I_T] = \frac{P[\psi_{t+1}^{(i)}|I_T]\ P[\psi_t^{(j)}|I_t]\ P[\psi_{t+1}^{(i)}|\psi_t^{(j)}]}{P[\psi_{t+1}^{(i)}|I_t]} \tag{4.68}$$

$$P[\psi_t^{(j)}|I_T] = \sum_{i=1}^{d} P[\psi_t^{(j)}, \psi_{t+1}^{(i)}|I_T]. \tag{4.69}$$

  Given $P[\psi_T|I_T]$ at the last iteration of the filter, (4.68) and (4.69) can be iterated for $t = T-1, T-2, \ldots, 1$ to get the smoothed probabilities, $P[\psi_t|I_T], t = T-1, T-2, \ldots, 1$. This smoothing algorithm can be interpreted as a backward filter that start at the end point $t = T$ of the applied filter after estimation. We note that the compuation of this approch is less demanding than that of the Hamilton filter.[27]

- For $\tau < t$ a *forecasting probability* $\hat{\psi}_{t|\tau}$ gives inference probability over future periods on the basis of estimates of states probabilities. Owing to this probability we can construct the forecasting value which is a weighted fitting value.

At fixed $\bar{k} = 3$, there are 27 different states $\psi_t^{(i)}$. As to these states we illustrate the smoothing probabilities in figure (4.7). The plots are generated by applying $m_0 = 1.3$,

[26] see Hamilton (1994) for details.
[27] see Hamilton (1989) for the Hamilton filter.

Monte Carlo ML results with Trinomial MSMF (Ia)

Parameters m_0, m_1 and σ estimation with $k = 2$									
m_0	1.2			1.3			1.4		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_{0,sim}$	1.180	1.177	1.189	1.287	1.285	1.294	1.366	1.381	1.391
FSSE	0.095	0.088	0.070	0.102	0.082	0.068	0.088	0.060	0.044
RMSE	0.097	0.091	0.071	0.103	0.083	0.068	0.095	0.063	0.045
m_1	0.5			0.6			0.7		
$\hat{m}_{1,sim}$	0.505	0.506	0.502	0.615	0.608	0.603	0.706	0.709	0.706
FSSE	0.032	0.025	0.017	0.049	0.032	0.021	0.070	0.055	0.046
RMSE	0.033	0.025	0.017	0.051	0.033	0.021	0.070	0.056	0.047
σ	1								
$\hat{\sigma}$	0.998	0.999	0.999	0.999	0.999	1.000	1.000	1.001	1.000
FSSE	0.021	0.016	0.011	0.022	0.014	0.010	0.021	0.015	0.010
RMSE	0.021	0.016	0.011	0.022	0.014	0.010	0.021	0.015	0.011

Table 4.6: Simulated biases $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for ML estimates. The value of parameters are assigned as that $m_0 = 1.2, 1.3, 1.4$, $m_1 = 0.5, 0.6, 0.7$ and $\sigma = 1$, respectively.

Monte Carlo ML results with Trinomial MSMF (Ib)

Parameters m_0, m_1 and σ estimation with $k = 2$									
m_0	1.1			1.2			1.3		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_{0,sim}$	1.077	1.084	1.086	1.171	1.177	1.186	1.268	1.275	1.285
FSSE	0.064	0.056	0.044	0.076	0.062	0.057	0.092	0.077	0.061
RMSE	0.068	0.058	0.046	0.081	0.066	0.059	0.097	0.081	0.063
m_1	0.9			0.8			0.7		
$\hat{m}_{1,sim}$	0.913	0.913	0.912	0.816	0.813	0.808	0.718	0.713	0.706
FSSE	0.071	0.060	0.048	0.070	0.053	0.045	0.065	0.050	0.040
RMSE	0.072	0.062	0.049	0.072	0.055	0.046	0.067	0.052	0.040
σ	1								
$\hat{\sigma}$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.001	1.001
FSSE	0.014	0.011	0.007	0.017	0.011	0.008	0.019	0.014	0.010
RMSE	0.014	0.011	0.007	0.017	0.011	0.008	0.019	0.014	0.010

Table 4.7: Simulated biases $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for ML estimates. The value of parameters are assigned as that $m_0 = 1.1, 1.2, 1.3$, $m_1 = 0.9, 0.8, 0.7$ and $\sigma = 1$, respectively.

Monte Carlo ML results with Trinomial MSMF(IIa)

Parameters m_0, m_1 and σ estimation with $k = 5$									
m_0	1.2			1.3			1.4		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_{0,sim}$	1.191	1.197	1.194	1.277	1.281	1.287	1.367	1.385	1.393
FSSE	0.090	0.075	0.066	0.095	0.076	0.060	0.085	0.060	0.036
RMSE	0.091	0.075	0.066	0.097	0.079	0.061	0.091	0.062	0.036
m_1	0.5			0.6			0.7		
$\hat{m}_{1,sim}$	0.508	0.503	0.503	0.615	0.606	0.602	0.717	0.712	0.707
FSSE	0.029	0.019	0.013	0.046	0.029	0.017	0.062	0.054	0.040
RMSE	0.031	0.019	0.014	0.048	0.029	0.017	0.064	0.055	0.040
σ	1								
$\hat{\sigma}$	0.991	0.994	0.996	0.997	0.996	0.998	1.003	1.003	1.003
FSSE	0.045	0.030	0.022	0.042	0.030	0.020	0.046	0.031	0.023
RMSE	0.046	0.031	0.022	0.042	0.030	0.020	0.046	0.031	0.023

Table 4.8: Simulated biases $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for ML estimates. The value of parameters are assigned as that $m_0 = 1.2, 1.3, 1.4$, $m_1 = 0.5, 0.6, 0.7$ and $\sigma = 1$, respectively.

Monte Carlo ML results with Trinomial MSMF(IIb)

Parameters m_0, m_1 and σ estimation with $k = 5$									
m_0	1.1			1.2			1.3		
n	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
$\hat{m}_{0,sim}$	1.080	1.083	1.088	1.168	1.178	1.188	1.261	1.284	1.287
FSSE	0.046	0.035	0.028	0.059	0.052	0.045	0.089	0.070	0.054
RMSE	0.050	0.039	0.030	0.067	0.057	0.046	0.097	0.072	0.055
m_1	0.9			0.8			0.7		
$\hat{m}_{1,sim}$	0.918	0.913	0.907	0.820	0.813	0.808	0.717	0.711	0.703
FSSE	0.040	0.033	0.027	0.049	0.039	0.031	0.056	0.043	0.029
RMSE	0.044	0.035	0.028	0.053	0.042	0.032	0.058	0.044	0.030
σ	1								
$\hat{\sigma}$	1.002	0.999	1.000	0.998	1.000	1.000	1.002	1.001	1.002
FSSE	0.025	0.017	0.011	0.043	0.029	0.022	0.065	0.044	0.032
RMSE	0.025	0.017	0.011	0.043	0.029	0.022	0.065	0.044	0.032

Table 4.9: Simulated biases $\hat{m}_0$, $\hat{m}_1$ and $\hat{\sigma}$, finite sample standard errors(FSSE) and root mean squared errors(RMSE) for ML estimates. The value of parameters are assigned as that $m_0 = 1.1, 1.2, 1.3$, $m_1 = 0.9, 0.8, 0.7$ and $\sigma = 1$, respectively.

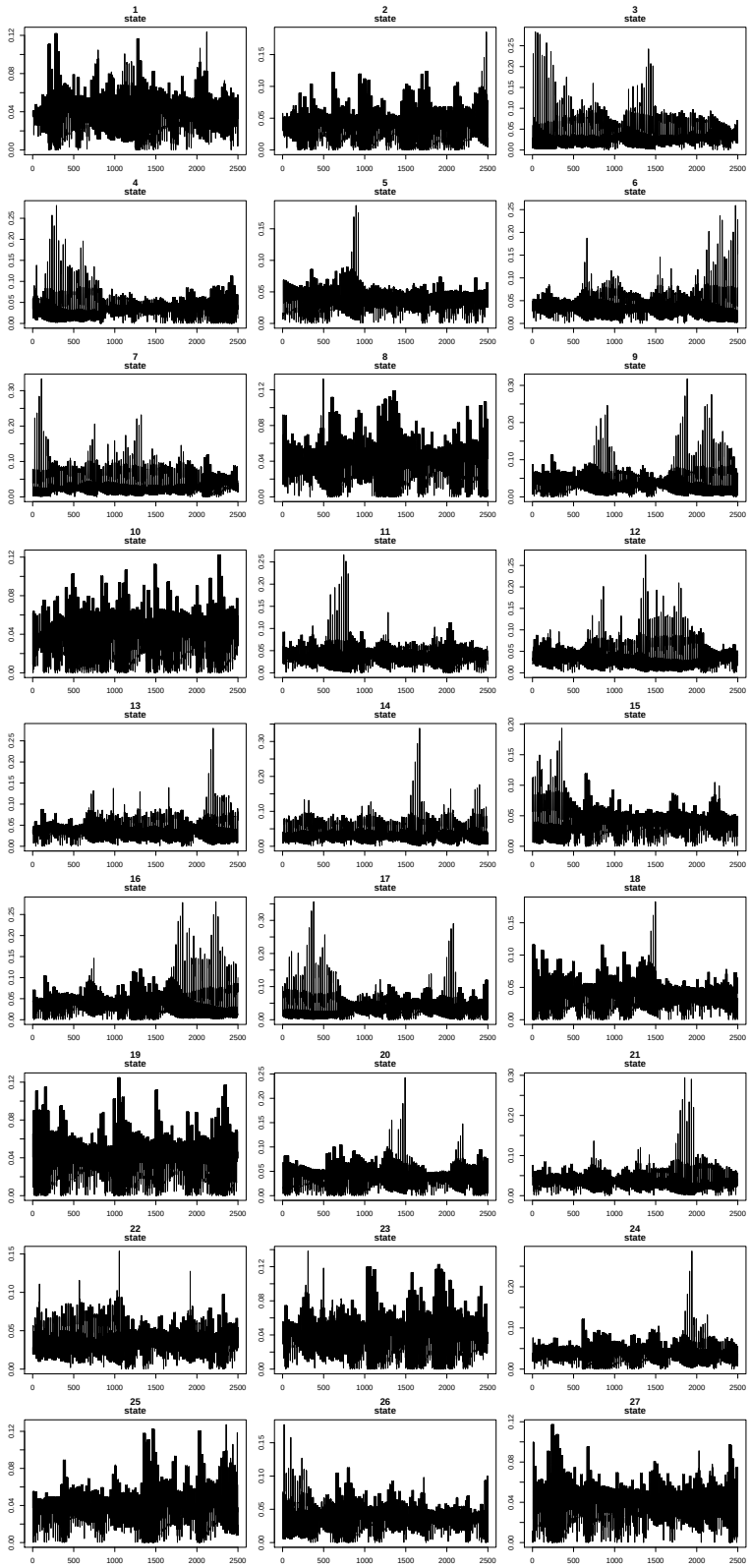

Figure 4.7: The plots represent some variations of smoothing probability that is gained as by-product of the ML estimation. We select randomly a sample with size of 2500 among the 10^5 data that are produced by $m_0 = 1.3$, $m_1 = 0.6$ and $\bar{k} = 3$. The smoothing probability depends on the states $\psi_t^{(d)}$. We zoom the smoothing probability: y-axes is not between 0 and 1.

$m_1 = 0.6$ and a sample size $n = 2500$ to the Monte Carlo simulation. Some plots such as 1, 2, 5, 8, 10, 18, 19, 22, 23, 25, 26, 27 show less variability than the rest. That is, the change of probability $\hat{\psi}_{t|\tau}$ is relatively few in a considered period. The highly variable probability means that the corresponding states are evolving dynamically and are clustered tightly around the average smoothing probability. Such states can be seen as an origin of a clustered volatility.

In the Markov switching model, it is typically to identify a position of each state with the aid of the smoothing probability. However, as the figure (4.7) shows, there is no absolute identifiable state in discrete MSMF model.[28] At most, we can infer possible states using weighted smoothing probabilities. The weighted probability in each time plays the important role in the dynamics of volatility, although there is large variability among the smoothed probabilities.

4.5 Empirical analysis using real data

In this section we estimate using some real financial data the parameters m_0, m_1 and σ based on GMM as well as ML. Because of the currently computational capacity and the size of real data we will change the maximum number of multipliers $\bar{k}$ from 1 up to 7.

As input of observation x_t we use five daily recorded financial data such as three stock market indices i.e. the German DAX and the New York Stock Exchange Composite Index(NYCI) as well as the Korean stock market(KOSPI), the exchange rate of Deutsch Mark/US$(USD-DM) and the daily price of gold from the London Precious Metal Exchange(GOLD). Stock market series were obtained from the New York, Frankfurt and Korea Stock Exchanges, while the exchange rate and precious metal series were provided by the financial database at the University of Bonn. The period and whole number T of observations of these empirical data are as followed:

Gold	$01/1978 - 12/1998\,(T = 5140)$
USD-DM	$01/1974 - 12/1998\,(T = 6140)$
NYCI	$01/1966 - 12/1998\,(T = 8308)$
DAX	$10/1959 - 12/1998\,(T = 9818)$
KOSPI	$01/1980 - 12/2001\,(T = 6301)$

At any day t, the daily return is defined as $x_t = 100 \ln\{P_t/P_{t-1}\}$, where P_t denotes the exchange rate at every trading day. Aassumed that the daily returns are dependent, we correct the returns by means of ordinary least square. First, let's analyse the results from GMM and then turn to the one from ML.

GMM estimation results

To obtain the GMM estimates, we use the conditional moments $f_1(T)$ and $f_2(T)$ as well as the unconditional moments $f_{i,3,4,5,6}$, where

$$\begin{aligned} f_1(T) &= E[\xi_{t+T,T}\xi_{t,T}] - \xi_{t+T,T}\xi_{t,T} \\ f_2(T) &= E[\xi^2_{t+T,T}\xi^2_{t,T}] - \xi^2_{t+T,T}\xi^2_{t,T} \end{aligned} \tag{4.70}$$

[28] From this observation, we can introduce instead of discrete hidden Markov chain a continuous Markov chain. See the Log-normal MSMF model of Lux (2006).

with $T = \{1, 5, 10, 20\}$ and

$$\begin{aligned} f_3(q=1) &= E[|x_t|] - |x_t| \\ f_4(q=2) &= E[(x_t - E[x_t])^2] - (x_t - \bar{x}_t)^2 \\ f_5(q=3) &= E[|x_t|^3] - |x_t|^3 \\ f_6(q=4) &= E[|x_t|^4] - |x_t|^4 \end{aligned} \tag{4.71}$$

The variable $\xi_{t,T}$ is defined by log difference of the absolute returns. The conditional moments $f_1(T)$ and $f_2(T)$ serve to exploit the temporal structure of the transformed process, while the unconditional moments $f_{3,4,5,6}(q = 1, 2, 3, 4)$ are needed to identify the shared fundamental volatility σ which is independent of $\bar{k}$. We filter out linear dependence of in-sample data (NYCI, DAX, KOSPI, USD, GOLD) using AR(1) before we carry out the absolute and squared transformation of log-differences. Table (4.10) presents the GMM estimation results for the trinomial MSMF model together with the Hansen's J-statistic and P-value of this statistic in parenthesis.

The J-statistic is obtained by the multiplication of the lengh of the considered data set and the objective log likelihood function in equation (4.24) evaluated at estimates $\hat{\theta} = (\hat{m}_0, \hat{m}_1, \hat{\sigma})$. For the trinomial MSMF model with correctly specified $\bar{k}$, the J-statistic has asymptotically a χ-distribution with degrees of freedom equal to the number of overidentifying conditions. Thus, Hansen's statistic is employed to test the overidentifying conditions of moments.

In our estimation, the whole number of applied moments is 12 and we have to estimate three parameters. So the number of overidentifying conditions is $12-3 = 9$. And the 5(1)% critical value of chi-square with degree of freedom 9 is 16.919(21.666). In data of NYCI, DAX and GOLD the J-statistic decreases after $\bar{k} = 2$. The value of J-statistic from KOSPI and USD[29] shows monotonic decline from $\bar{k} = 1$ to $\bar{k} = 5$ and then increases again. Most of the J-values lies under 21.666. Hence, the J-statistics are significant and we can't reject the validity of the estimating conditions except for the data of NYCI, KOSPI and USD at $\bar{k} = 1, 2$.

As next, we turn to the results of multi-fractal parameters. In the data sets of NYCI, KOSPI, USD and GOLD, the estimated value of the first multi-fractal parameter $\hat{m}_0$ decreases with growing $\bar{k}$, while the estimated value of the second one increases with riging $\bar{k}$. The directions of convergence of the two parameters are opposite! Only in a case of DAX, the multi-fractal parameters seem to be oscillated around 1.4 and 0.8, respectively.

The estimated scaling parameter $\hat{\sigma}$ shows mainly two different developments as $\bar{k}$ moves from 1 to 7. In cases of KOSPI, USD and GOLD the fundamental volatility remains stable across $\bar{k}$ as $\hat{\sigma}_{KOSPI} \approx 1.45$, $\hat{\sigma}_{USD} \approx 0.68$, $\hat{\sigma}_{GOLD} \approx 0.73$, respectively. On the other side, there are monotonical increasing of $\hat{\sigma}_{DAX}$ and $\hat{\sigma}_{NYCI}$. Note that the standard errors of $\hat{\sigma}$ are tiny compared with that of two multi-fractal parameters. As an independent component of volatility, each market shares this scaling parameter. We guess that the smaller the scaling parameter the more efficient will be the market and vice versa. Therefore, we interpret it as a basic measure of market efficiency.

ML estimation

We report the results of ML estimation in table(4.11). The first column of $\bar{k} = 1$ represents a Markov switching model with only three possible states for volatility. As $\bar{k}$ is rising, the number of states growes at the rate $3^{\bar{k}}$. There are 2187 volatility states when $\bar{k} = 7$.

In most case, the values of parameter $\hat{m}_0$ tend to decline monotonically with growing $\bar{k}$. There are two exceptional cases. That is, $\hat{m}_0$ moves again up at $\bar{k} = 7$ in KOSPI and after $\bar{k} = 5$ in USD. In contrast to $\hat{m}_0$, the second estimated parameter $\hat{m}_1$ displays different developments for each

[29] For $\bar{k} = 2$ the value of J-statistic increases exceptionally.

GMM Estimates of Trinomial MSMF(k) model

k	1	2	3	4	5	6	7
NYCI							
$\hat{m}_0$	2.061	1.819	1.710	1.633	1.573	1.526	1.488
	(0.078)	(0.051)	(0.040)	(0.034)	(0.030)	(0.027)	(0.025)
$\hat{m}_1$	0.501	0.597	0.644	0.682	0.712	0.735	0.754
	(0.087)	(0.040)	(0.027)	(0.021)	(0.018)	(0.016)	(0.014)
$\hat{\sigma}$	0.750	0.772	0.792	0.804	0.809	0.813	0.815
	(0.021)	(0.022)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)
J-Statistic	28.804	22.834	17.423	14.214	12.669	11.907	11.511
(P-value)	(0.001)	(0.007)	(0.043)	(0.115)	(0.178)	(0.219)	(0.242)
DAX							
$\hat{m}_0$	1.435	1.339	1.412	1.438	1.432	1.415	1.397
	(0.356)	(0.254)	(0.157)	(0.118)	(0.097)	(0.085)	(0.077)
$\hat{m}_1$	0.888	0.879	0.807	0.781	0.779	0.786	0.796
	(0.235)	(0.162)	(0.102)	(0.074)	(0.059)	(0.050)	(0.044)
$\hat{\sigma}$	0.885	0.886	0.913	0.940	0.958	0.969	0.977
	(0.034)	(0.036)	(0.043)	(0.047)	(0.049)	(0.051)	(0.052)
J-Statistic	13.816	13.458	12.674	11.635	10.796	10.192	9.751
(P-value)	(0.129)	(0.143)	(0.178)	(0.235)	(0.290)	(0.335)	(0.371)
KOSPI							
$\hat{m}_0$	2.160	2.009	1.919	1.801	1.712	1.646	1.594
	(0.092)	(0.061)	(0.041)	(0.035)	(0.032)	(0.029)	(0.027)
$\hat{m}_1$	0.446	0.499	0.542	0.599	0.643	0.676	0.702
	(0.072)	(0.034)	(0.021)	(0.019)	(0.017)	(0.016)	(0.014)
$\hat{\sigma}$	1.210	1.359	1.474	1.464	1.452	1.442	1.436
	(0.049)	(0.067)	(0.063)	(0.057)	(0.054)	(0.053)	(0.052)
J-Statistic	28.200	20.915	11.151	7.183	6.403	6.575	6.993
(P-value)	(0.001)	(0.013)	(0.266)	(0.618)	(0.699)	(0.681)	(0.638)
USD							
$\hat{m}_0$	2.121	1.944	1.707	1.622	1.560	1.513	1.475
	(0.048)	(0.036)	(0.034)	(0.031)	(0.028)	(0.026)	(0.024)
$\hat{m}_1$	0.450	0.539	0.650	0.691	0.721	0.745	0.763
	(0.026)	(0.018)	(0.017)	(0.015)	(0.014)	(0.013)	(0.012)
$\hat{\sigma}$	0.669	0.694	0.681	0.684	0.686	0.687	0.687
	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)
J-Statistic	22.566	27.153	19.307	18.669	18.613	18.764	18.975
(P-value)	(0.007)	(0.001)	(0.023)	(0.028)	(0.029)	(0.027)	(0.025)
GOLD							
$\hat{m}_0$	2.017	1.747	1.622	1.546	1.491	1.450	1.418
	(0.059)	(0.047)	(0.040)	(0.036)	(0.032)	(0.030)	(0.028)
$\hat{m}_1$	0.496	0.629	0.690	0.728	0.755	0.775	0.791
	(0.031)	(0.024)	(0.020)	(0.018)	(0.016)	(0.015)	(0.014)
$\hat{\sigma}$	0.720	0.726	0.730	0.733	0.735	0.736	0.736
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.021)	(0.021)
J-Statistic	17.703	17.801	17.520	17.187	17.028	16.975	16.973
(P-value)	(0.039)	(0.038)	(0.041)	(0.046)	(0.048)	(0.049)	(0.049)

Table 4.10: The table shows results of estimation for the trinomial MSMF that are obtained with five real data using generalized method of moments. Hansen's statistic is denoted by J-statistic. The corresponding P-value are represented in parenthesis. Note that the 5(1)% critical value for a chi-square with degree of freedom 9 is 16.919(21.666).

data set. The estimate $\hat{m}_1$ in NYCI shows little fluctuations, while the one from DAX is increasing till $\bar{k} = 7$. For the cases of KOSPI and USD, the estimates are increasing monotonically at first and then they fall suddenly to their minimum value at $\bar{k} = 6$ and $\bar{k} = 5$, respectively. The standard errors of $\hat{m}_1$ by USD are as a whole very large compared with the results from other data. For GOLD the value of the second parameter is increasing with rising $\bar{k}$ and reaches a maximum at $\bar{k} = 5$ and decreases afterward. Note that the standard error of $\hat{m}_1$ is mostly larger than those of the first multi-fractal parameter $\hat{m}_0$. We observe that the estimate of all three parameters using GMM are converging to the estimates using ML with increasing $\bar{k}$ and there are some differences between the GMM estimates and the ML estimates in smaller $\bar{k} = 1, 2, 3$.

At fixed $\bar{k}$, the variability of MSMF process depends mainly on $max\,|m_i - m_j|$, $i = j = 0, 1, 2$. As the number of multipliers is decreasing, then the value of $max\,|m_i - m_j|$ should be larger to absorb the volatility possessed by the data and vice versa. In case of binomial MSMF($\bar{k}$) model, the multiplier parameter $\hat{m}_0$ alone takes this role: with a larger number of multipliers, lower value of $\hat{m}_0$ is required in each $m_{t,k}$ to match the fluctuations in volatility exhibited by the data [Calvet and Fisher (2004)].

In general, the change in value of scaling parameter is moderate in every data. The fundamental volatility $\hat{\sigma}$ gives basically a sharing term of volatility in a financial market. The fundamental volatility is persistent and can not be distinguished from volatility components of very low frequency $\bar{k}$. As a result, the fundamental volatility should not and does not change so much although the number of multipliers $\bar{k}$ is increasing from 1 up to 7.

In each data set, $\hat{\sigma}$ and its standard error behavior are as follows: The estimated scaling parameter $\hat{\sigma}_{NYCI}$ is moving down marginally until $\bar{k} = 6$ and increases slightly again at $\bar{k} = 7$. The estimated value of $\hat{\sigma}_{DAX}$ is decreasing up to $\bar{k} = 4$ and increaes again at $\bar{k} = 5, 6$ except for $\hat{\sigma} = 0.940$ at $\bar{k} = 7$. In a comparision to NYCI, the scaling factor of DAX is larger.

The estimated value of $\hat{\sigma}_{KOSPI}$ is approximately around 1.5. In other words, the fundamental volatility is much larger than those of the NYCI and DAX data. As one among the typical emerging markets, we can confirm that the Korean stock market tends to be more volatile than the other stock markets. Let's consider the scaling function $\tau(q) = -1 - log_3\left((\hat{m}_0/3)^q + (\hat{m}_1/3)^q + (\hat{m}_2/3)^q\right)$ with the estimated values at $\bar{k} = 7$.

We can then comfirm that the scaling function of the KOSPI data is much more concave than that of the NYCI and DAX data. It means that volatility components with short or intermediate range of frequencies for the KOSPI data are also varying faster across $\bar{k}$ than for the other stock markets. For the USD data, $\hat{\sigma}_{USD}$ have the smallest values of around 0.65 among the five underlying data. Since the fundamental volatility is very small, the main fluctuation in volatility should be matched by the variation of multi-fractal parameters together with $\bar{k}$. The fundamental volatility $\hat{\sigma}_{GOLD}$ is the second highest after $\hat{\sigma}_{KOSPI}$. The highly dynamic volatility of the GOLD data can be absorbed by combining the variable multi-fractal parameters and the relatively large $\hat{\sigma}_{GOLD}$.[30] In all data, we find that the standard errors of $\hat{\sigma}$ is increasing with $\bar{k}$. If $\bar{k} = 7$, the first and second multiplier remain almost unchanged over long volatility cycles [Calvet and Fisher (2004)].

Thus, it is not easy to distinguish the long-run averages of volatility $\hat{\sigma}$ from the compound process $\hat{\sigma} \prod_{k=1}^{\bar{k}} m_{t,k}$. This leads to the increase of standard errors. The log-likelihood increases rapidly at low $\bar{k}$ and slowly with high $\bar{k}$. A better fit comes along with a larger number of multipliers that are used to reflect the heterogeneous degrees of persistence in the volatility of the data. In all cases, the log-likelihood increases with $\bar{k}$.

[30] The exception is $\hat{\sigma} = 0.975$ at $\bar{k} = 7$.

ML Estimates of Trinomial MSMF($\bar{k}$) model

k	1	2	3	4	5	6	7
NYCI							
$\hat{m}_0$	2.130	1.859	1.677	1.578	1.507	1.482	1.457
	(0.019)	(0.018)	(0.017)	(0.016)	(0.013)	(0.012)	(0.012)
$\hat{m}_1$	0.603	0.699	0.775	0.630	0.682	0.759	0.771
	(0.102)	(0.284)	(0.525)	(0.107)	(0.172)	(0.452)	(0.411)
$\hat{\sigma}$	0.849	0.849	0.836	0.820	0.819	0.768	0.796
	(0.007)	(0.008)	(0.010)	(0.012)	(0.016)	(0.018)	(0.020)
ln L	-9603.954	-9463.042	-9369.383	-9309.192	-9279.218	-9266.963	-9263.071
DAX							
$\hat{m}_0$	2.235	1.910	1.755	1.626	1.546	1.499	1.499
	(0.023)	(0.022)	(0.017)	(0.015)	(0.014)	(0.016)	(0.019)
$\hat{m}_1$	0.461	0.592	0.623	0.687	0.727	0.750	0.750
	(0.055)	(0.216)	(0.219)	(0.277)	(0.312)	(0.402)	(0.420)
$\hat{\sigma}$	1.145	1.108	1.099	1.077	1.091	1.140	0.940
	(0.012)	(0.014)	(0.014)	(0.016)	(0.018)	(0.032)	(0.029)
ln L	-6252.206	-6149.112	-6079.933	-6043.127	-6025.764	-6023.129	-6024.803
KOSPI							
$\hat{m}_0$	2.427	2.045	1.832	1.686	1.603	1.569	1.591
	(0.016)	(0.019)	(0.021)	(0.021)	(0.023)	(0.026)	(0.022)
$\hat{m}_1$	0.444	0.656	0.768	0.833	0.866	0.576	0.638
	(0.039)	(0.178)	(0.500)	(0.895)	(1.395)	(0.059)	(0.152)
$\hat{\sigma}$	1.680	1.552	1.512	1.508	1.507	1.493	1.440
	(0.012)	(0.014)	(0.017)	(0.022)	(0.030)	(0.042)	(0.045)
ln L	-10618.620	-10448.347	-10347.249	-10286.695	-10254.025	-10239.475	-10234.675
USD							
$\hat{m}_0$	2.119	1.786	1.620	1.528	1.498	1.521	1.581
	(0.024)	(0.025)	(0.024)	(0.025)	(0.024)	(0.032)	(0.016)
$\hat{m}_1$	0.685	0.870	0.938	0.969	0.547	0.589	0.710
	(0.244)	(1.376)	(3.799)	(10.657)	(0.031)	(0.087)	(1.549)
$\hat{\sigma}$	0.706	0.694	0.677	0.658	0.633	0.636	0.515
	(0.007)	(0.008)	(0.011)	(0.014)	(0.020)	(0.032)	(0.016)
ln L	-5997.645	-5884.403	-5815.427	-5779.356	-5764.106	-5759.595	-5758.694
GOLD							
$\hat{m}_0$	2.606	2.259	2.017	1.887	1.811	1.810	1.809
	(0.015)	(0.024)	(0.027)	(0.027)	(0.026)	(0.037)	(0.019)
$\hat{m}_1$	0.336	0.552	0.682	0.722	0.743	0.667	0.595
	(0.122)	(0.061)	(0.312)	(0.451)	(0.481)	(0.382)	(0.260)
$\hat{\sigma}$	1.558	1.422	1.342	1.297	1.327	1.221	0.975
	(0.016)	(0.019)	(0.023)	(0.030)	(0.044)	(0.055)	(0.025)
ln L	-6453.163	-6286.309	-6206.365	-6165.633	-6145.716	-6137.910	-6137.427

Table 4.11: This table shows the results by maximum likelihood estimation for the trinomial MSMF($\bar{k}$) model, where the real data of NYCI, DAX, KOSPI, USD, and GOLD are employed. Each columns correspond to the maximum number of multipliers ($\bar{k}$) in the estimated model. Asymptotic standard errors are given in parenthesis.

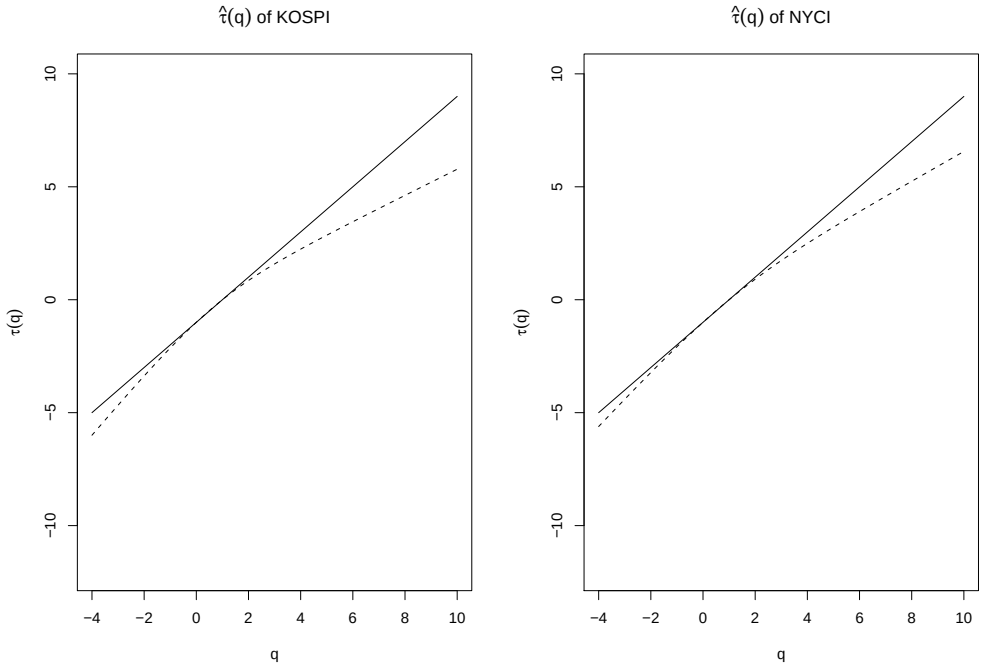

Figure 4.8: Using the estimates of GMM and ML, the scaling functions for the data set, KOSPI and NYCI, are plotted. Since the estimates of multi-fractal parameters using GMM is close to the estimates of ML, we use only the estimates of ML with $\bar{k} = 7 : \hat{m}_0 = 0.530$, $\hat{m}_1 = 0.213$ for KOSPI and $\hat{m}_0 = 0.486$, $\hat{m}_1 = 0.257$ for NYCI. The scaling functing of the KOSPI data is more concave than that of the NYCI data in the right panel.

4.6 Problem of model selection

When estimating the parameters of MSMF model with real data, we encounter the question of the appropriate frequency, i.e. the optimal number of multipliers. There are only few papers related to the selection of the optimal frequency for the estimation. Hansen (1992) develops the likelihood ratio test that can account not only for the presence of unidentified nuisance parameters, but also for the case of an ill-behaved likelihood function.[31] He treats the likelihood as an empirical process indexed by all the parameters. And his test relies mainly on taking the supremum of likelihood ratio over the nuisance parameters.

However, this test has two shortcomings. First, the computational burden is too high. In particular, the application of this test to the trinomial MSMF model with high frequency does not appear promising in practice. Second, this method provides a bound for the likelihood ratio statistics but no critical value. As an alternative, Garcia (1998) introduce the asymptotic distribution of the likelihood ratio test in supremum for a two-state Markov switching model and also provides asymptotic critical values for the likelihood ratio test.

Due to the complexity of the procedures and heavy burden of computation, we do apply neither the likelihood ratio test of Hansen nor the modification by Garcia to determine the frequency $\bar{k}$. Instead, we try to find answers in three other ways: the heuristic choice, the Vuong test and the Markov-switching criteria (MSC).

[31] Let's suppose that we will test the null hypothesis of a binomial MSMF($\bar{k}$) model against the alternative hypothesis of a trinomial MSMF($\bar{k}$) model. At the constrained maximum likelihood estimates, the score vector i.e. the numerical derivative of the log-likelihood function with respect to parameters m_0, m_1 and σ is identically zero. Under this condition, the likelihood ratio does not have a standard asymptotic distribution, see Garcia (1998).

4.6.1 The heuristic choice: GMM

Lux (2006) finds that Hansen's statistics can not distinguish the misspecified $\bar{k}$ from the most optimal $\bar{k}$ if we estimate the parameters using relatively large data sets. He tries then to find an indication of the most appropriated $\bar{k}$ by heuristic choice. He lets the maximal numer of frequencies vary from 1 to 20. The estimates show relatively large variation up to a critical frequency $\bar{k}'$. When $\bar{k}$ exceeds the critical number $\bar{k}'$, the outcome of the estimation remains practically unchanged despite increases of $\bar{k}$.

He also consider the case that the process has much larger number of volatility multipliers and the size of the available time series is limited. In this circumstance, most volatility states are hardly changed in a region of low frequencies so that the increase of $\bar{k}$ remains invisible: additional high-level multipliers make a marginal contribution to the moments so that their numerical values would stay almost constant.

However, he finds in his emprical analysis that higher choices of $\bar{k}$ improve volatility forecasts for the data from five different foreign exchange markets. In particular, mean squared errors (MSE) and mean absolute errors (MAE) decline at all forecasting horizons for increasing $\bar{k}$ while the estimated value of the multi-fractal parameters remains essentially unchanged[32]. Remember that the autocorrelation function of multi-fractal process depends not only on the constellation of multi-fractal parameters, but also on the number of multipliers $\bar{k}$. As the hyperbolic decline of the simulated autocorrelations have been shown in figure (4.5) and (4.6), we note that the long memory has been found at extremly long time lags without any apparent cut-off.

This heuristic argument is based mainly on the GMM method which is flexble enough to deal with large numbers of multipliers. However, due to the computational limitation, this heuristic choice can not be applied to the maximum likelihood estimation for trinomial or quadronomial MSMF.

4.6.2 The Vuong Test

Vuong (1989) introduced a directional and symmetric test for choosing a model among the competing models when the observations are independent and identically distributed (i.i.d.) and when the alternative models are non-nested and estimated by maximum likelihood.[33] The Vuong test requires that the likelihood ratio statistic is asymptotically normal distributed under general conditions. The likelihood ratio can be written

$$\begin{aligned} LR_T(\hat{\theta}_1, \hat{\theta}_2) &= \sqrt{T}\Big(L_t^f(\hat{\theta}_1) - L_t^g(\hat{\theta}_2)\Big) \\ &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \ln\left\{\frac{f(x_t, \hat{\theta}_1)}{g(x_t, \hat{\theta}_2)}\right\}, \end{aligned} \tag{4.73}$$

[32] Once the number of multipliers $\bar{k}$ increases beyound a certain threshold, the GMM estimates donot change any more [Lux (2006)].

[33] The Vuong test employs the Kullback and Leibler (1951) information criterion which is defined by

$$KL\Big(f(x_t, \theta_1), g(x_t, \theta_2)\Big) \equiv \sum_{x_t} f(x_t, \theta_1) \ln\left[\frac{f(x_t, \theta_1)}{g(x_t, \theta_2)}\right] = E\left[\ln\left\{\frac{L(x_t, \theta_1)}{L(x_t, \theta_2)}\right\}\right], \tag{4.72}$$

where θ_1 and θ_2 are the estimated parameters and x_t denotes the observations. L is the abbreviation of likelihood.

where

$$L_t^f(\theta_1) = \sum_{t=1}^{T} \ln f(x_t, \theta_1)$$
$$L_t^g(\theta_2) = \sum_{t=1}^{T} \ln g(x_t, \theta_2). \tag{4.74}$$

The ML estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ maximise the functions $L_t^f(\theta_1)$ and $L_t^g(\theta_2)$. Vuong proves that the expected value can be consistently estimated by n^{-1} times the likelihood ratio statistic that satisfies the null hypothesis H_0 of

$$H_0 : E\left[\ln\left\{\frac{f(x_t, \hat{\theta}_1)}{g(x_t, \hat{\theta}_2)}\right\}\right] = 0. \tag{4.75}$$

In the i.i.d case of the observation x_t, the Central Limit Theorem (CLT) implies

$$\frac{1}{\sqrt{T}} LR_T(\hat{\theta}_1, \hat{\theta}_2) \xrightarrow{d} N(0, \hat{\sigma}^2). \tag{4.76}$$

Here, $\hat{\sigma}^2$ is consistently estimated by the sample variance. The Vuong test determines whether or not the average log-likelihood ratio is statistically different from zero. If the models are close to the true specification, the mean of log-likelihood ratios should be zero. If the model $f(x_t, \theta_1)$ is better than the model $g(x_t, \theta_2)$ the mean of this ratios should be positive and vice versa.

In the non-i.i.d case, Calvet and Fisher (2004) suggest to adjust the correlation in the likelihoods ratio. Following Newey and West (1987), they estimate the adjusted variance σ_T by

$$\hat{\sigma}_T^2 = \frac{1}{T}\sum_{t=1}^{T} \ln\left[\frac{f(x_t, \hat{\theta}_{1,t})}{g(x_t, \hat{\theta}_{2,t})}\right]^2 + 2\sum_{j=1}^{m_T} \omega(j, m) \frac{1}{T} \sum_{t=j+1}^{T} \ln\left[\frac{f(x_t, \hat{\theta}_{1,t})}{g(x_t, \hat{\theta}_{2,t})}\right] \ln\left[\frac{f(x_{t-j}, \hat{\theta}_{1,t-j})}{g(x_{t-j}, \hat{\theta}_{2,t-j})}\right], \tag{4.77}$$

where $\omega(j, m) = 1 - j/(m+1)$ is the Bartlett weight. The automatic lag m_T is selected by the method of Newey and West (1994). In general, the Vuong statistic is sensitive to the number of parameters to be estimated in each model. Using a penalty term that corresponds to Schwert (1978)'s Bayesian Information Criteria (BIC), we obtain the adjusted statistic of

$$\widetilde{LR}_T(\hat{\theta}_1, \hat{\theta}_2) = LR_T\left(\hat{\theta}_1, \hat{\theta}_2\right) - \left[\left(\frac{p}{2}\right)\ln T - \left(\frac{q}{2}\right)\ln T\right], \tag{4.78}$$

where p and q are the number of estimated parameters in competing models $f(x_t, \hat{\theta}_{1,t})$ and $g(x_t, \hat{\theta}_{2,t})$, respectively.

4.6.3 Clarke test (2003)

We propose Clarke's test (Clarke (2003)) as an alternative to the Vuong test. We can select the model via the Akaike information criterion (AIC) or the Schwarz information criterion (BIC) which are based on the optimization of the objective function. Clarke suggests a distribution-free test that can be applied to a modified paired sign test for the distinction between the individual log-likelihoods of two non-nested models. The proposed test is interested in the median log-likelihood ratio, whereas the concept of Vuong test is related to the mean of log-likelihood ratio.

According to Clarke's test, model $f(x_t, \theta_1)$ is better than model $g(x_t, \theta_2)$ if more than half of the

individual log-likelihood ratios should be greater than zero and vice versa. It can be summarised in the null hypothesis of Clarke's distribution-free test:

$$H_0 : Pr\left[\ln \left\{ \frac{f(x_t, \theta_1)}{g(x_t, \theta_2)} \right\} > 0 \right] = 0.5. \tag{4.79}$$

This test assumes that the individual log-likelihood ratios of $\ln \Big\{ f(x_t, \theta_1)/g(x_t, \theta_2) \Big\}$ are mutually independent. It does not exclude some possible correlations between individual log-likelihoods themselves. We refer to the work of Clarke (2005) for the detailed proofs of consistency and unbiasedness for the distribution-free test.

The test statistic B is given by

$$B = \sum_{t=1}^{T} I_{(0,+\infty)}(d_t) \tag{4.80}$$

with $d_t = \ln f(x_t, \theta_1) - \ln g(x_t, \theta_2)$. It is intuitive that $f(x_t, \theta_1)$ is the better model than $g(x_t, \theta_2)$ only if B is significantly larger than its expected value under the null hypothesis $(n/2)$. For an upper tail test, we reject the null hypothesis of equivalence when $B \leq c_\alpha$, where c_α is chosen to be the smallest integer such that

$$\sum_{c=c_\alpha}^{T} \binom{T}{c} 0.5^T \leq \alpha. \tag{4.81}$$

For a lower tail test, the inequality is reversed, and the limits of the sum go from $c = 0$ to $c = c_\alpha$. One of the great strengths of this procedure is that its implementation is remarkably simple:

1) Compute the individual log-likelihoods of the rival models $f(x_t, \theta_1)$ and $g(x_t, \theta_2)$: $\ln f(x_t, \theta_1)$ and $\ln g(x_t, \theta_1)$.

2) Compute the differences d_t and count the number of positive values.

3) The numbers of positive difference B should be distributed by Binomial $(T, 0.5)$.

This test is also sensitive to the dimensionality of the competing models. Hence, we need a correction for the degrees of freedom. Clarke (2005) proposed to use the Schwarz correction given by $(p/2T) \ln T - (q/2T) \ln T$, where p and q are the number of estimated coefficients in models $f(x_t, \theta_1)$ and $g(x_t, \theta_2)$, respectively. It is worthwhile to note that Clarke's test applies the average correction to the individual log-likelihood ratios.[34] This correction is justified by Vuong's argument: As long as the correction factor divided by the square root of T has a stochastic order of one,

$$\frac{1}{\sqrt{T}} K_T \Big(f(x_t, \theta_1), g(x_t, \theta_2) \Big) = o_p(1), \tag{4.82}$$

the adjusted statistic has the same asymptotic properties as the unadjusted one. Clarke investigates the asymptotic relative efficiency of the Vuong test versus the distribution-free test. He notes that the distribution-free test reaches $2/\pi = 0.637$ or 64% of the efficiency of Vuong test if the individual log-likelihood ratio is normally distributed.

[34] As the individual log-likelihood ratios are used, Clarke' test cannot apply the original Schwarz correction, $(p/2) \ln T - (q/2) \ln T$ as Vuong test in his correction to the summed log-likelihood ratio.

The distribution-free test would be even more inefficient for platykurtic distribution (with kurtosis smaller than 3 in Gaussian case). Under such conditions, he suggests to use the Vuong test because it provides greater power than the distribution-free test. Using double exponential distributed data that is heavy-tailed and high peaked, Clarke finds that the Vuong test is only 50% as efficient as the distribution-free test. If log likelihood ratios are more leptokurtic, then the Vuong test is less efficient. For the leptokurtic distributed data, the Clarke test provides greater power than the Vuong test.

4.6.4 Model selection based on MSC (2005)

We also employ the Markov switching criterion (MSC) that has been newly suggested by Smith, Naik, and Tsai (2006) to find the optimal number of variables θ and $\bar{k}$ simultaneously. They applied the MSC to models which are nested in the form of an autoregressive Markov switching model to capture business cycles in real GNP. The MSC based on the Kullback-Leibler divergence is a kind of penalised likelihood criteria and has the form of

$$\mathrm{MSC} = -2\ln L + \sum_{i=1}^{N} \frac{T_i(T_i + \lambda_i \kappa)}{\delta_i T_i - \lambda_i \kappa - 2}. \tag{4.83}$$

Here, N and κ stand for the number of states as well as the number of multi-fractal parameters in a model, respectively. T_i stands for the number of observations in state i. For example let us look at data with $T = 100$ and using a Markov switching model with two states. Using smoothing probability, we can find that 30 observations belong to the state 1, i.e. $T_1 = 30$, the rest to state 2 ($T_2 = 70$). T_i is calculated as the sum of smoothing probabilities of each state i. It means that the number of observations in each state i are identifiable. Furthermore, the total number T of observations in an estimation is obtained just by $\sum_{i=1}^{N} T_i$. The MSC does not contain the scaling parameter σ. Note that increasing κ hardly leads to a change in value of MSC when we apply this criterion to the MSMF($\bar{k}$) model with a relatively large number of observations. We can select the optimal number $\bar{k}$ and the number of parameters by minimisation of equation (4.83).

As exhibited in equation (4.83), the MSC is composed of the term of the log-likelihood, $2\ln L$, and the penalty term that penalises the use of more states and more parameters in the MSC. The log-likelihoods can be affected if the number of parameters is different in the competing models, for example binomial MSMF model vs. trinomial MSMF model. To ensure positive penalty, the denominator ($\delta_i T_i - \lambda_i \kappa - 2$) should be larger than one for each state. When it is actually less than one, this denominator is considered as 1.

Figure (4.9) shows different penalties in terms of various values of δ_i and λ_i, where $N = 2$, $\kappa = 4$ and $T = 100$. As represented in the upper three plots, the penalty term is minimised at $T_1 = T_2 = T/2$ if $\delta_1 = \delta_2$ and $\lambda_1 = \lambda_2$. If there is no a priori reason to put into different δ_i and λ_i for different states, the penalty term in the MSC is symmetric in $T/2$. The three plots in bottom exhibit the asymmetric property of MSC for $\delta_1 \neq \delta_2$. We also point out that increasing λ_i leads to a higher penalty. The high value of λ_i will be useful if the competing models can not be easily distinguished from each other. Confirming simulation results are available in Smith et al. (2006). Fig.(4.9) also shows that the penalty terms of the MSC are higher at extreme values of hidden states T_1 or T_2. It indicates that MSC prefers models with 'similar' distributions of states across different regimes, in the sense that the number of T_1 and T_2 are not much different from each other. In addition, the penalty terms appear to be quite flat in the middle regions. The flat regions might be problematic if MSC can not distingush one from the other model and especially if κ does not change with rising $\bar{k}$ as in the discrete MSMF model.

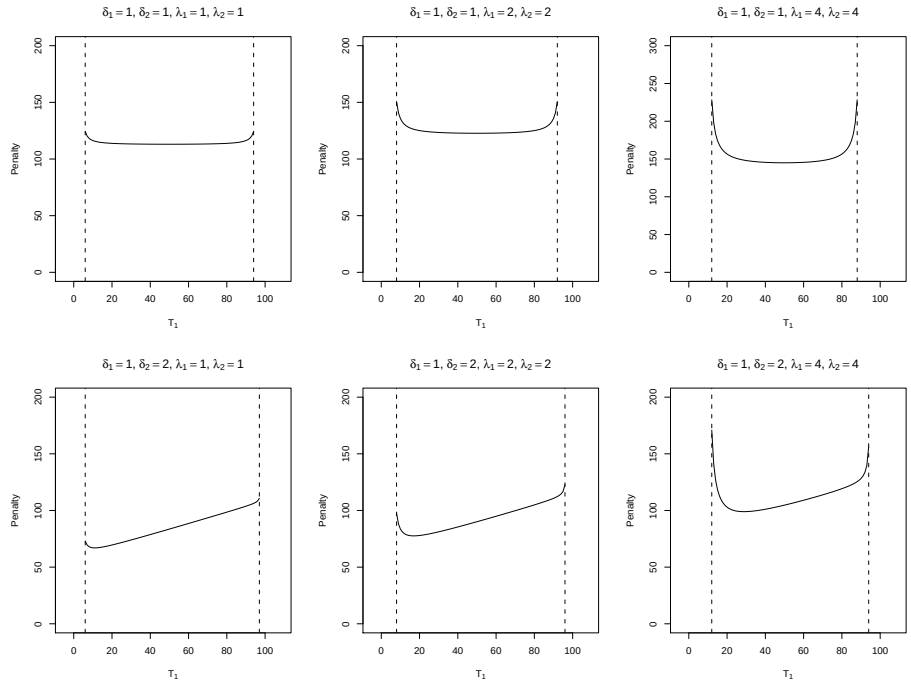

Figure 4.9: Markov switching criterion(MSC):The plots show the variations of penalty which is affected by the values of parameters $\delta \in (1, 2)$ and $\lambda \in (1, 2)$. For the simulation of penalty we use $T = 100$, $N = 2$ and $\kappa = 2$. The vertical dotted line marks the region where MSC is valid.

4.6.5 Empirical Test

We will here treat the problem of model selection by means of the five financial data set that have been used in section 4.5. First, we perform the Vuong test of equation (4.73) and equation (4.78) as well as the Clarke test of equation (4.80) for the binomial and trinomial MSMF($\bar{k}$). We consider MSMF(10) and MSMF(7) as the *true* model $f(x_t, \theta)$ for the binomial and trinomial MSMF process, respectively. For the binomial case, the maximum frequency varies from 1 to 9 , whereas $\bar{k}$ changes from 1 and to 6 for the trinomial model. The usability or validity of the competing model can be measured by one-sided values P.

The table (4.13) displays the results of the log-likelihood difference and the P-value in parentheses for the trinomial MSMF process, while table (4.12) exhibits the results of tests for the binomial MSMF. In cases of $\bar{k} = 1$, 2, 3 the log-likelihood difference is statistically significant at 1% level in non-adjusted HAC and at 5% level in HAC. This is strong evidence that MSMF(7) significantly outperforms MSMF process with $\bar{k} = 1$, 2, 3. The $\ln L$ at $\bar{k} = 4$ is significant only at 5% level in the Vuong test and the HAC-adjusted test, while the Clarke test remains weighty at 1% level. We can reject the null at 10% level in the Vuong test and 8% level in the adjusted HAC case if $\bar{k} = 5$.

But the Clarke test rejects the null at 1% level. The Clarke test has more power to differentiate the trinomial MSMF(5) from the trinomial MSMF(7). In other words, the MSMF(7) outperforms the model with $\bar{k} = 5$. For $\bar{k} = 6$, we observe lower significance levels in all tests. The null can not be rejected at the conventional level. We conclude that the trinomial MSMF model works better when $\bar{k} = 5$, 6. In order to maintain the consistency in the remaining analysis we will henceforth focuse on the trinomial MSMF(7) process for all empirical time series.

Next, we calculate the Markov switching criterion for the cases of the weighting factor to the number of parameters $\lambda_i = 1$, N, N^2 and the weighting factor to the state $\delta_i = 1$ and summerize the results in tables (4.14) and (4.15). The tables also present the values which are obtained by

other criteria such as AIC, state-adjusted AIC(AIC$_{snt}$) and BIC. Smith et al. (2006) show that MSC is the proper analogue of AIC in the sense that it balances the tradeoff between the fitting and the parsimony of the model, i.e. the number of parameters and states. AIC and BIC are useful if we are interested in forecasting of dynamic volatility in a financial time series. Each criterion is discerned by its penalty: the penalty of AIC reads as 2κ. AIC must be corrected for the degree of freedom if the competing models have different numbers of state-parameters. According to Smith et al. (2006) the adjusted AIC has the the penalty term of $2(N\kappa + N^2)$, while the penalty in BIC looks like $\kappa \ln T$.

The penalty of MSC is typically much larger than that of the AIC$_{snt}$ that requests a significant improvement in fit for additional states as MSC. For instance, the penalty of the AIC$_{snt}$ with $N = 2$ and $\kappa = 1$ is 8 for binomial MSMF(1). All the values of the penalty of the MSC, as shown in the figure(4.9), are much higher than 8. Hence, the MSC in general favours the models with smaller states than the state-adjusted AIC does.

As long as $N > 1$, the AIC$_{snt}$ is much more conservative than AIC and BIC. Unless $\ln T$ is smaller than 2, BIC is more conservative than AIC. We obtain the best fitting MSMF($\bar{k}$) if the model fits the data with the lowest value of criterion.

Based on AIC and BIC the trinomial MSMF(6) or MSMF(7) is best appropriate for the fitting of the given data. More explicitly, the MSMF(7) shall be chosen for the data sets of NYCI, KOSPI, USD and GOLD. The data of DAX is rather better described by the trinomial MSMF(6). Not only log-likelihood but also the values of both criterion are decreasing monotonically as $\bar{k}$ moves upwards. In contrast to AIC and BIC, the AIC$_{snt}$ tends to favour small $\bar{k}$ so that the binomial MSMF(3) and trinomial MSMF(2) fit all data best. The term of penalty enforces us to choose the MSMF with small $\bar{k}$ that is appreantly increased with rising number of multipliers. In fact, it is difficult to use MSMF(7) for ML estimation due to too long computing time: GMM can be employed more easily for such large $\bar{k}$ and relatively large samples.

Let's examine the MSCs of the binomial and trinomial MSMF model. Remember that the MSC has a minimised penalty if the states are equally distributed. According to the $MSC_{\lambda=1}$ statistic we have to select the trinomial MSMF(3) for all data except USD that is best fitted by the binomial MSMF(5). The criterion of $MSC_{\lambda=N}$ favours three MSMF models: the trinomial MSMF(1) model for NYCI, DAX, USD and trinomial MSMF(2) for the data KOSPI and GOLD as well as the binomial MSMF(3) model. The third criterion of $MSC_{\lambda=N^2}$ tends to choose the trinomial MSMF(1) and the binomial MSMF(2). Besides selection of the best fitting model we can put the models in order. For instance of NYCI, trinomial MSMF($\bar{k}$) is so arranged that MSMF(2) follows MSMF(3) as next best model. And MSMF(4), MSMF(1),MSMF(5), MSMF(6) and MSMF(7) is the sequence of the preferred model for this data after this criterion.

Our conclusion is inconsistent with the results of the Vuong (1989), Clarke (2003) and Clarke (2005) tests which come to the conclusion that the more volatility states the better the MSMF($\bar{k}$) model. The analysis of high-frequency financial time series might need the MSMF model with a large number of volatility components. On the basis of the results we can recommend $\bar{k} = 3$ for trinomial MSMF and $\bar{k} = 5$ for binomial MSMF as the optimal frequency, respectively.

Apparently, the MSCs may be applied explicitly only for the Markov switching model with standard transition probability. It seems rational to penalise the model with more parameters than necessary. For instance, there are twelve parameters to be estimated in addition to the transition matrix when we estimate parameters of a Markov switching model with four states. Hence, the MSC gives more penalties for such a model than our MSMF($\bar{k}$) model with the constrained transition matrix. The MSMF($\bar{k}$) models are penalty-free or penalty-resistant due to the following reasons: they have only a small number of parameters. In particular, increasing the number of volatility states does not change the number of parameters to be estimated. Note that the number

of states is implicit, i.e. it needs not to be estimated.

Binomial MSMF Model Selection using Vuong and Clarke test

data	$k=1$	2	3	4	5	6	7	8	9
A. Vuong(1989) Test									
Gold	-7.319	-5.753	-4.578	-3.745	-2.958	-2.317	-1.685	-1.224	-0.712
	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.021)	(0.092)	(0.221)	(0.477)
USD-DM	-6.052	-5.711	-4.820	-3.902	-3.055	-2.333	-1.702	-1.188	-0.728
	(0.000)	(0.000)	(0.000)	(0.000)	(0.002)	(0.020)	(0.089)	(0.235)	(0.466)
NYCI	-5.739	-5.589	-5.285	-4.706	-3.976	-3.213	-2.456	-1.788	-1.185
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.014)	(0.074)	(0.236)
DAX	-4.997	-5.302	-4.971	-4.369	-3.444	-2.586	-1.848	-1.238	-0.846
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.010)	(0.065)	(0.216)	(0.397)
KOSPI	-9.179	-7.948	-6.812	-5.745	-4.829	-4.001	-3.258	-2.622	-2.073
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.009)	(0.038)
B. HAC Adjusted Vuong Test									
Gold	-4.735	-4.172	-3.565	-3.088	-2.602	-2.117	-1.589	-1.188	-0.705
	(0.000)	(0.000)	(0.000)	(0.002)	(0.009)	(0.034)	(0.112)	(0.235)	(0.481)
USD-DM	-3.101	-2.913	-2.606	-2.262	-1.909	-1.580	-1.270	-0.991	-0.699
	(0.002)	(0.004)	(0.009)	(0.024)	(0.056)	(0.114)	(0.204)	(0.322)	(0.485)
NYCI	-3.643	-3.902	-4.030	-3.880	-3.496	-2.966	-2.346	-1.749	-1.216
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.019)	(0.080)	(0.224)
DAX	-2.883	-3.263	-3.368	-3.327	-2.816	-2.218	-1.639	-1.099	-0.747
	(0.004)	(0.001)	(0.001)	(0.001)	(0.005)	(0.027)	(0.101)	(0.272)	(0.455)
KOSPI	-6.111	-5.983	-5.572	-5.036	-4.454	-3.842	-3.210	-2.618	-2.092
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.009)	(0.036)
C. Clarke(2003, 2005) Test									
Gold	3204	3074	2982	2913	2860	2809	2797	2669	2889
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
USD-DM	3954	3912	3834	3719	3627	3489	3324	3185	3109
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.003)	(0.326)
NYCI	5438	5397	5292	5217	5153	5110	5058	5021	5099
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
DAX	3173	3156	3130	3095	3069	3051	3037	3017	2954
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
KOSPI	4677	4423	4340	4244	4210	4144	4083	4042	4022
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 4.12: This table reports test statistics and the one-sided p-values in bracket for the binomial MSMF($\bar{k}$) model. As input, the five empirical data are used. Panel A presents the results of Vuong test of equation (4.73) (Vuong (1989)). Panel B gives heteroskedasticity and autocorrelation of equation (4.78) by using automatic lag selection by Newey and West (1987) and Newey and West (1994). Panel C results from Clarke test of equation (4.80). A low p-value indicates that the corresponding model would be rejected in favour of the multi-fractal with nine different frequencies.

Trinomial MSMF Model Selection using Vuong and Clarke test

data	$k=1$	2	3	4	5	6
A. Vuong(1989) Test						
Gold	-5.612	-3.625	-2.409	-1.592	-0.882	-0.249
	(0.000)	(0.000)	(0.016)	(0.111)	(0.378)	(0.803)
USD-DM	-5.124	-3.804	-2.482	-1.482	-0.804	-0.403
	(0.000)	(0.000)	(0.013)	(0.138)	(0.422)	(0.687)
NYCI	-5.359	-4.508	-3.446	-2.328	-1.460	-0.801
	(0.000)	(0.000)	(0.001)	(0.020)	(0.144)	(0.423)
DAX	-4.934	-4.121	-2.723	-1.510	-0.391	0.131
	(0.000)	(0.000)	(0.006)	(0.131)	(0.696)	(0.896)
KOSPI	-7.234	-5.301	-3.888	-2.794	-1.928	-1.260
	(0.000)	(0.000)	(0.000)	(0.005)	(0.054)	(0.208)
B. HAC Adjusted Vuong Test						
Gold	-4.264	-3.105	-2.255	-1.637	-0.948	-0.283
	(0.000)	(0.002)	(0.024)	(0.102)	(0.343)	(0.777)
USD-DM	-4.333	-3.399	-2.316	-1.453	-0.807	-0.382
	(0.000)	(0.001)	(0.021)	(0.146)	(0.420)	(0.702)
NYCI	-3.737	-3.699	-3.126	-2.246	-1.451	-0.866
	(0.000)	(0.000)	(0.002)	(0.025)	(0.147)	(0.387)
DAX	-3.206	-3.192	-2.488	-1.424	-0.371	0.135
	(0.001)	(0.001)	(0.013)	(0.155)	(0.711)	(0.893)
KOSPI	-5.866	-4.908	-3.902	-2.934	-2.081	-1.362
	(0.000)	(0.000)	(0.000)	(0.003)	(0.037)	(0.173)
C. Clarke(2003,2005) Test						
Gold	3098	2928	2821	2736	2677	2424
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
USD-DM	3963	3850	3741	3646	3512	3728
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
NYCI	5350	5226	5082	4881	4741	4473
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.369)
DAX	2992	2942	2895	2864	2917	2774
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
KOSPI	4532	4258	4151	4078	4066	3954
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 4.13: This table reports test statistics and the one-sided p-values in bracket for trinomial MSMF($\bar{k}$) model. As input, the five empirical data are used. Panel A presents the results of Vuong test of equation (4.73). Panel B gives heteroskedasticity and autocorrelation of equation (4.78) by using automatic lag selection by Newey and West (1987) and Newey and West (1994). Panel C results from Clarke test of equation (4.80). A low p-value indicates that the corresponding model would be rejected in favour of the multi-fractal with six different frequencies.

Trinomial MSMF ($\bar{k}$) model: AICs and Markov-Switching Criterien I

k	1	2	3	4	5	6	7
NYCI							
$\ln L$	-9603.954	-9463.042	-9369.383	-9309.192	-9279.218	-9266.963	-9263.071
AIC_{snt}	19237.908	19124.084†	20304.765	32064.385	137628.436	1084331.927	9593212.142
AIC	19211.908	18930.084	18742.765	18622.385	18562.436	18537.927	18530.142†
BIC	19207.908	18926.084	18738.765	18618.385	18558.436	18533.927	18526.142†
MSC_1	27532.934	27287.320	27029.911†	27431.379	28521.807	34007.925	61877.085
MSC_N	27557.030†	27582.684	30681.886	$2.2 \cdot 10^6$	$4.3 \cdot 10^6$	$1.2 \cdot 10^7$	$3.6 \cdot 10^7$
MSC_{N^2}	27629.740†	30804.006	$1.5 \cdot 10^7$	$1.1 \cdot 10^8$	$9.8 \cdot 10^8$	$8.8 \cdot 10^9$	$7.9 \cdot 10^{10}$
DAX							
$\ln L$	-6252.206	-6149.112	-6079.933	-6043.127	-6025.764	-6023.129	-6024.803
AIC_{snt}	12534.412	12496.225†	13725.866	25532.254	131121.527	1077844.258	9586735.607
AIC	12508.412	12302.225	12163.866	12090.254	12055.527	12050.258†	12053.607
BIC	12504.412	12298.225	12159.866	12086.254	12051.527	12046.258†	12049.607
MSC_1	17018.461	16848.665	16821.896†	17106.710	18432.394	29752.837	33372.534
MSC_N	17042.641†	17150.638	21166.952	$9.9 \cdot 10^5$	$2.2 \cdot 10^6$	$6.6 \cdot 10^6$	$2.0 \cdot 10^7$
MSC_{N^2}	17115.977†	21194.123	$7.3 \cdot 10^6$	$5.9 \cdot 10^7$	$5.3 \cdot 10^8$	$4.8 \cdot 10^9$	$4.3 \cdot 10^{10}$
KOSPI							
$\ln L$	-10618.620	-10448.347	-10347.249	-10286.695	-10254.025	-10239.475	-10234.675
AIC_{snt}	21267.240	21094.695†	22260.498	34019.389	139578.050	1086276.950	9595155.351
AIC	21241.240	20900.695	20698.498	20577.389	20512.050	20482.950	20473.351†
BIC	21237.240	20896.695	20694.498	20573.389	20508.050	20478.950	20469.351†
MSC_1	27555.275	27251.007	27159.370†	27387.012	28570.238	36960.301	49450.049
MSC_N	27579.403	27548.801†	30927.784	$1.5 \cdot 10^6$	$3.2 \cdot 10^6$	$9.2 \cdot 10^6$	$2.7 \cdot 10^7$
MSC_{N^2}	27652.351†	31035.273	$1.0 \cdot 10^7$	$8.3 \cdot 10^7$	$7.4 \cdot 10^8$	$6.6 \cdot 10^9$	$6.0 \cdot 10^{10}$
USD							
$\ln L$	-5997.645	-5884.403	-5815.427	-5779.356	-5764.106	-5759.595	-5758.694
AIC_{snt}	12025.290	11966.807†	13196.853	25004.711	13058.213	1077317.191	9586203.387
AIC	11999.290	11772.807	11643.853	11562.711	11532.213	11523.191	11521.387†
BIC	11995.290	11768.807	11630.853	11558.711	11528.213	11519.191	11517.387†
MSC_1	18153.326	17963.125	17935.776†	18212.823	19485.442	28948.875	29676.975
MSC_N	18177.455†	18261.143	21396.881	$1.5 \cdot 10^6$	$3.2 \cdot 10^6$	$9.0 \cdot 10^6$	$2.6 \cdot 10^7$
MSC_{N^2}	18250.417†	21772.342	$1.0 \cdot 10^7$	$8.1 \cdot 10^7$	$7.3 \cdot 10^8$	$6.5 \cdot 10^9$	$5.8 \cdot 10^{10}$
GOLD							
$\ln L$	-6453.163	-6286.309	-6206.365	-6165.633	-6145.716	-6137.910	-6137.427
AIC_{snt}	12936.327	12770.619†	13978.730	25777.266	131361.432	1078073.819	9586960.855
AIC	12910.327	12576.619	12416.730	12335.266	12295.432	12279.819	12278.855†
BIC	12906.327	12572.619	12412.730	12331.266	12291.432	12275.819	12274.855†
MSC_1	17384.377	17051.064	17002.840†	17281.023	18618.828	28418.295	27451.591
MSC_N	17372.561	17353.211†	21396.881	$9.7 \cdot 10^5$	$2.2 \cdot 10^6$	$6.5 \cdot 10^6$	$1.9 \cdot 10^7$
MSC_{N^2}	17445.934†	21417.825	$7.2 \cdot 10^6$	$5.8 \cdot 10^7$	$5.2 \cdot 10^8$	$4.7 \cdot 10^9$	$4.2 \cdot 10^{10}$

Table 4.14: The Markov-Switching-Criterion (MSC) are calculated with smoothing probabilities of each state. MSC_{λ_i} stands for $\lambda_i = 1, N, N^2$. We denote the optimal frequency by † in each MSC's and AICs. Note that $\mathrm{AIC}_{snt} = -2 \ln L + 2(N \cdot \kappa + N^2)$.

Binomial MSMF ($\bar{k}$) model: AIC and Markov-Switching Criterien II

k	1	2	3	4	5	6	7	8	9	10
NYCI										
$\ln L$	-9729.837	-9605.454	-9520.256	-9449.760	-9391.240	-9346.197	-9314.146	-9294.004	-9282.764	-9278.027
AIC_{snt}	19471.675	19259.908	19148.512†	19433.519	20849.479	27012.394	51652.292	150172.008	543877.528	217756.054
MSC_1	27774.680	27533.931	27379.605	27270.893	27218.994†	27261.591	27473.100	28029.806	29435.498	33754.306
MSC_N	27778.686	27558.024	27492.917†	27770.183	29524.725	44692.910	$1.6 \cdot 10^8$	$2.4 \cdot 10^8$	$4.4 \cdot 10^8$	$8.5 \cdot 10^6$
MSC_{N^2}	27786.704	27655.101†	28458.368	43744.340	$1.0 \cdot 10^7$	$3.5 \cdot 10^7$	$1.3 \cdot 10^8$	$5.4 \cdot 10^8$	$2.1 \cdot 10^9$	$8.7 \cdot 10^9$
DAX										
$\ln L$	-6369.276	-6271.509	-6205.472	-6149.915	-6107.560	-6077.287	-6057.254	-6045.350	-6039.745	-6037.539
AIC_{snt}	12750.552	12583.018	12554.944†	12843.829	14327.120	20474.574	45138.508	$1.4 \cdot 10^5$	$5.4 \cdot 10^5$	$2.1 \cdot 10^6$
MSC_1	17242.563	17055.061	16939.118	16860.536	16842.026†	16918.806	17176.502	17869.299	20252.978	26040.946
MSC_N	17246.574	17079.234	17053.586†	17387.043	19543.119	$3.1 \cdot 10^5$	$7.7 \cdot 10^5$	$1.3 \cdot 10^6$	$2.4 \cdot 10^6$	$4.6 \cdot 10^6$
MSC_{N^2}	17254.606	17177.250†	18087.506	$9.0 \cdot 10^5$	$5.3 \cdot 10^6$	$1.9 \cdot 10^7$	$7.4 \cdot 10^7$	$2.9 \cdot 10^8$	$1.2 \cdot 10^9$	$4.7 \cdot 10^9$
KOSPI										
$\ln L$	-10846.729	-10645.798	-10534.445	-10446.817	-10379.941	-10329.154	-10292.662	-10268.087	-10252.714	-10244.046
AIC_{snt}	21705.458	21331.596	21212.890†	21437.634	22871.882	28978.308	53609.323	$1.5 \cdot 10^5$	$5.5 \cdot 10^5$	$2.1 \cdot 10^6$
MSC_1	28001.466	27607.627	27401.013	27258.131	27189.910†	27222.704	27433.032	28022.094	29729.994	35333.479
MSC_N	28005.474	27631.750	27514.752†	27764.009	29627.642	60919.103	$1.2 \cdot 10^6$	$1.8 \cdot 10^6$	$3.3 \cdot 10^6$	$6.5 \cdot 10^6$
MSC_{N^2}	28013.497	27729.176†	28504.627	52514.870	$7.7 \cdot 10^6$	$2.6 \cdot 10^7$	$1.0 \cdot 10^8$	$4.1 \cdot 10^8$	$1.7 \cdot 10^9$	$6.6 \cdot 10^9$
USD										
$\ln L$	-6100.101	-6000.895	-5930.191	-5873.710	-5830.445	-5800.230	-5780.482	-5769.029	-5763.033	-5760.771
AIC_{snt}	12212.202	12041.790	12004.382†	12291.420	13772.890	19920.460	44584.964	$1.4 \cdot 10^5$	$5.4 \cdot 10^5$	$2.1 \cdot 10^6$
MSC_1	18348.209	18157.822	18032.507	17951.926	17930.947†	18004.976	18249.621	18871.919	20663.099	26771.750
MSC_N	18350.211	18169.845	18088.727†	18193.823	18938.882	22171.099	35669.463	94980.507	$4.0 \cdot 10^5$	$2.6 \cdot 10^6$
MSC_{N^2}	18354.215	18217.939†	18538.486	22064.175	51192.802	28802.957	$2.3 \cdot 10^6$	$2.0 \cdot 10^7$	$2.0 \cdot 10^8$	$2.6 \cdot 10^9$
GOLD										
$\ln L$	-6681.947	-6453.002	-6347.446	-6279.172	-6227.660	-6192.337	-6168.217	-6153.910	-6146.334	-6143.090
AIC_{snt}	13375.894	12946.005	12838.893†	13102.344	14567.319	20704.673	45360.434	$1.4 \cdot 10^5$	$5.4 \cdot 10^5$	$2.1 \cdot 10^6$
MSC_1	17795.905	17346.046	17151.068	17047.652	17010.214†	17076.655	17323.578	17985.982	20073.317	28243.257
MSC_N	17799.917	17370.225	17265.562†	17564.629	19720.519	$3.3 \cdot 10^5$	$7.4 \cdot 10^5$	$1.2 \cdot 10^6$	$2.3 \cdot 10^6$	$4.6 \cdot 10^6$
MSC_{N^2}	17807.950	17468.279†	18300.996	770212.404	$5.2 \cdot 10^6$	$1.8 \cdot 10^7$	$7.3 \cdot 10^7$	$2.9 \cdot 10^8$	$1.2 \cdot 10^9$	$4.6 \cdot 10^9$

Table 4.15: MSC_{λ_i} stands for $\lambda_i = \{1,\, N,\, N^2\}$. We denote the optimal frequency by † in each MSC's and $\mathrm{AIC}_{snt} = -2\ln L + 2(N \cdot \kappa + N^2)$.

4.7 Quadronomial MSMF model

In this section we calculate the required moments of quadronmial MSMF process for the GMM estimation and estimate the parameters using GMM and ML. Monte Carlo results are also provided and we discuss the problem of ambiguity with respect to the quadronomial parameters.

4.7.1 Moments of quadronomial multi-fractal process

Let's consider the quadronomial MSMF process in which the volatility components $m_{t,k}$ are drawn from the quadronomial distribution:

$$m_{t,k} \in \{m_0, m_1, m_2, m_3\}, \tag{4.84}$$

Constrained by $m_0 + m_1 + m_2 + m_3 = 4$ so that $E[M_t] = 1$. We are going to call this process as quadronomial MSMF. The switching probability γ_k is at this time specified by

$$\gamma_k = 1 - \left(1 - \frac{1}{4}\right)^{4^{k-\bar{k}}}. \tag{4.85}$$

The structure of the model remains unchanged, see equation (4.1). The quadronomial MSMF process has the unconditional $E[M_t^q]$ and conditional $E[M_{t+T}M_T]$ moments:

$$\begin{aligned}
E\big[M_t^q\big] &= \left(\frac{1}{4}\,m_0^q + \frac{1}{4}\,m_1^q + \frac{1}{4}\,m_2^q + \frac{1}{4}\,m_3^q\right)^{\bar{k}} \\
E\big[M_{t+T}M_t\big] &= \prod_{k=1}^{\bar{k}} \left[\left\{\left(1-\gamma_k\right)^T + \frac{1}{4}\left(1-\left(1-\gamma_k\right)^T\right)\right\}\left(\frac{1}{4}m_0^2 + \frac{1}{4}m_1^2 + \frac{1}{4}m_2^2 + \frac{1}{4}m_3^2\right)\right. \\
&\left. + \frac{2}{4}\left(1-\left(1-\gamma_k\right)^T\right)\left(\frac{1}{4}m_0m_1 + \frac{1}{4}m_0m_2 + \frac{1}{4}m_0m_3 + \frac{1}{4}m_1m_2 + \frac{1}{4}m_1m_3 + \frac{1}{4}m_2m_3\right)\right].
\end{aligned} \tag{4.86}$$

4.7.2 Moments of the compound and log-transformed volatility process

Let's present the absolute moments of quadronomial MSMF $E[|x_t|^q] = E[|\sigma_t|^q]E[|u_t|^q]$, $q = 1, \ldots 4$ and the variance $Var[|x_t|]$ as well as the moments of the log-transformed volatility process which serve to estimate the volatility parameters:

$$\begin{aligned}
E[|x_t|] &= \frac{2\sigma}{\sqrt{2\pi}}\left(\frac{1}{4}m_0^{1/2} + \frac{1}{4}m_1^{1/2} + \frac{1}{4}m_2^{1/2} + \frac{1}{4}m_3^{1/2}\right)^{\bar{k}} \\
E[|x_t|^2] &= \sigma^2 \\
E[|x_t|^3] &= \frac{4\sigma^3}{\sqrt{2\pi}}\left(\frac{1}{4}m_0^{3/2} + \frac{1}{4}m_1^{3/2} + \frac{1}{4}m_2^{3/2} + \frac{1}{4}m_3^{3/2}\right)^{\bar{k}} \\
E[|x_t|^4] &= 3\sigma^4\left(\frac{1}{4}m_0^2 + \frac{1}{4}m_1^2 + \frac{1}{4}m_2^2 + \frac{1}{4}m_3^2\right)^{\bar{k}}
\end{aligned} \tag{4.87}$$

and

$$Var[|x_t|] = \sigma^2\left\{1 - \frac{2}{\pi}\left(\frac{1}{4}m_0^{1/2} + \frac{1}{4}m_1^{1/2} + \frac{1}{4}m_2^{1/2} + \frac{1}{4}m_3^{1/2}\right)^{2\bar{k}}\right\}. \tag{4.88}$$

Related to log-transformed moments, we will give only the generalized forms of $E[\eta^2_{t+T,T}]$, $E[\eta_{t+T,T}\eta_{t,T}]$ and $E[\eta^2_{t+T,T}\eta^2_{t,T}]$ for any time step $T > 1$. The explicit procedure of derivation is presented in the appendix B. The log increments of the volatility process $\eta_{t,T}$ are defined by equation (A.8). The expectation of squared log incements of volatility reads

$$E[\eta^2_{t+T,T}] = \frac{1}{2}\Big\{(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 + (\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 + (\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\} \sum_{k=1}^{\bar{k}} \left\{\frac{1}{4}\left(1 - \left(1-\gamma_k\right)^T\right)\right\}. \tag{4.89}$$

And the first and second autocovariance of log-transformed volatility $E[\eta_{t+T,T}\eta_{t,T}]$ result in

$$\begin{aligned} E[\eta_{t+T,T}\eta_{t,T}] = &-\Big\{(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 + (\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 \\ &+(\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\} \sum_{k=1}^{\bar{k}} \left\{\frac{1}{16}\left(1 - \left(1-\gamma_k\right)^T\right)^2\right\}, \\ E[\eta^2_{t+T,T}\eta^2_{t,T}] = &\sum_{k=1}^{\bar{k}} \left\{\frac{1}{16}\left(1 - \left(1-\gamma_k\right)^T\right)\right\}^2 \Omega \\ &+\frac{1}{4}\sum_{k=1}^{\bar{k}} \left\{\frac{1}{4}\left(1 - \left(1-\gamma_k\right)^T\right) \sum_{k'=1,k'\neq k}^{\bar{k}} \frac{1}{4}\left(1 - \left(1-\gamma_{k'}\right)^T\right)\right\}\Psi \\ &+2\sum_{k=1}^{\bar{k}} \left\{\frac{1}{16}\left(1 - \left(1-\gamma_k\right)^T\right)^2 \sum_{k'=1,k'\neq k}^{\bar{k}} \frac{1}{16}\left(1 - \left(1-\gamma_{k'}\right)^T\right)^2\right\}\Psi \end{aligned} \tag{4.90}$$

where

$$\begin{aligned} \Omega \;= &\Big\{(\ln m_0 - \ln m_1)^4 + (\ln m_0 - \ln m_2)^4 + (\ln m_0 - \ln m_3)^4 \\ &+(\ln m_1 - \ln m_2)^4 + (\ln m_1 - \ln m_3)^4 + (\ln m_2 - \ln m_3)^4\Big\} \end{aligned} \tag{4.91}$$

and

$$\begin{aligned} \Psi \;= &\Big\{(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 + (\ln m_0 - \ln m_3)^2 \\ &+(\ln m_1 - \ln m_2)^2 + (\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\}^2. \end{aligned} \tag{4.92}$$

4.7.3 Monte Carlo performance of GMM and ML estimates

The designs of simulation, the procedure of sample generation and the employed sample size $n \in \{2500, 5000, 10000\}$ as well as the used time lag $T \in \{1, 5, 10, 20\}$ in equation (4.89) and (4.90) are the same as in the subsection 4.3.4. We assign to the paramters $\theta = (m_0, m_1, m_2, \sigma)$ the following input values:

$$\begin{aligned} &\text{Panel A}: \;\; (m_0, m_1, m_2, \sigma) \in \{1.3, 0.4, 1.1, 1\}, \{1.4, 0.5, 1.2, 1\}, \{1.5, 0.6, 1.3, 1\} \\ &\text{Panel B}: \;\; (m_0, m_1, m_2, \sigma) \in \{1.3, 0.5, 1.0, 1\}, \{1.4, 0.6, 1.1, 1\}, \{1.5, 0.7, 1.2, 1\} \end{aligned}$$

Without loss of generality, we set the scaling factor $\sigma = 1$.

The GMM and ML results of Monte Carlo simulations for quadronomial MSMF($\bar{k} = 3$) is reported in table (4.17). Together with the estimated values $\hat{m}_0$, $\hat{m}_1$, $\hat{m}_2$ and $\hat{\sigma}$ we find also two different standard errors (FSSE and RMSE) that indicate accuracy of estimated values. In table (4.17) the estimated multi-fractal parameters do not show direct connection with the sample size. For instance let us look at the panel A: $\hat{m}_1$ and $\hat{m}_2$ are getting closer to the true value of m_1 with the increasing sample size, but $\hat{m}_0$ is running away form the true value of m_0 with the increasing sample size. The biases of m_0, m_1 and m_2 are relatively large and do not become negligible as sample size increases. Furthermore, their standard errors are also large with respect to their sizes, whereas that of σ amounts to roughly ten percent of the true parameter value and gets smaller with sample length.

Compared the GMM results, the biases and standard errors of the ML estimates are relatively small. The three quadronomial multi-fractal parameters are not perfectly identified like those of binomial MSMF. For instance, the third multi-fractal parameter is well identified with reasonable standard error in the combination of multi-fractal parareters ($(m_0, m_1, m_2, \sigma) \in \{1.4, 0.5, 1.2, 1\}$) in panel A. But the convergence of $\hat{m}_0$ and $\hat{m}_2$ is extremly slow. It is of interest to see the case of $(m_0, m_1, m_2, \sigma) \in \{1.5, 0.7, 1.2, 1\}$ in panel B. The quadronomial multi-fractal parameters $(\hat{m}_0, \hat{m}_1, \hat{m}_2)$ converge slowly to their true values whereas there is almost no bias between the estimated value of the scaling parameter $\hat{\sigma}$ and its true value. We remark that the scaling parameter is well idenified with small standard errors in all case. Lux (2006) notes that the estimated values of the scaling parameter σ should be essentially identical to the sample standard deviation[35].

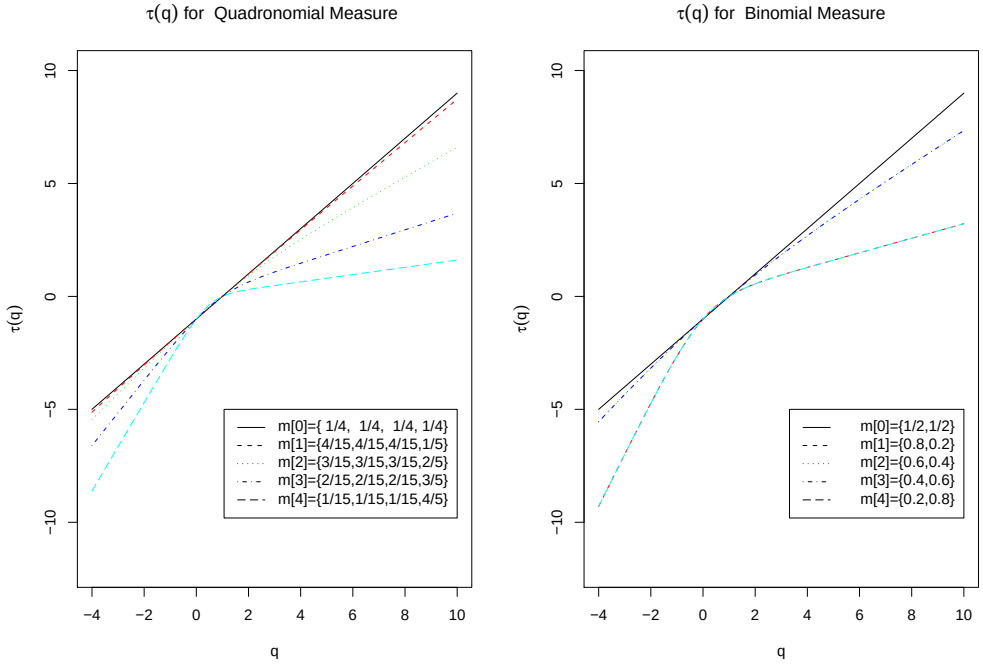

Figure 4.10: The scaling function $\tau(q)$ for quadronomial and binomial measure: The solid line represents the binomial specification of $m_0 = m_1 = 0.5$, while $m_0 = m_1 = m_2 = m_3 = 1/4$ in case of trinomial specification. Each measure is normalized so that their sum is one.

As a result, the qudronomial multi-fractal parameters are not well identified even though we constrain the valid range of quadronomial multi-fractal parameters : $1 < \hat{m}_0 < 3$, $0 < \hat{m}_1 < 1$, $1 <$

[35] We use the unconditional moments for estimating σ and therefore the covariance matrix of the parameters should be block-diagonal since the conditional moments for m_0, m_1, m_2 are not affected by the scaling parameter (σ) [Lux (2006)].

$\hat{m}_2 < 2$. This misidentification is clearly related with its scaling function $\tau(q)$. Gilbert, Willinger, and Feldmann (1999) also observed that the empirical scaling function of a trinomial multi-fractal process does not match accurately with the theoretical one $\tau(q) = -1 - \log_3\{(\frac{m_0}{3})^q + (\frac{m_1}{3})^q + (\frac{m_2}{3})^q\}$. In case of quadronomial multi-fractal process, we find that there is surely some ambiguity between the empirical scaling function and theoretical one of $\tau(q) = -1 - \log_4\{(\frac{m_0}{4})^q + (\frac{m_1}{4})^q + (\frac{m_2}{4})^q + (\frac{m_3}{4})^q\}$.

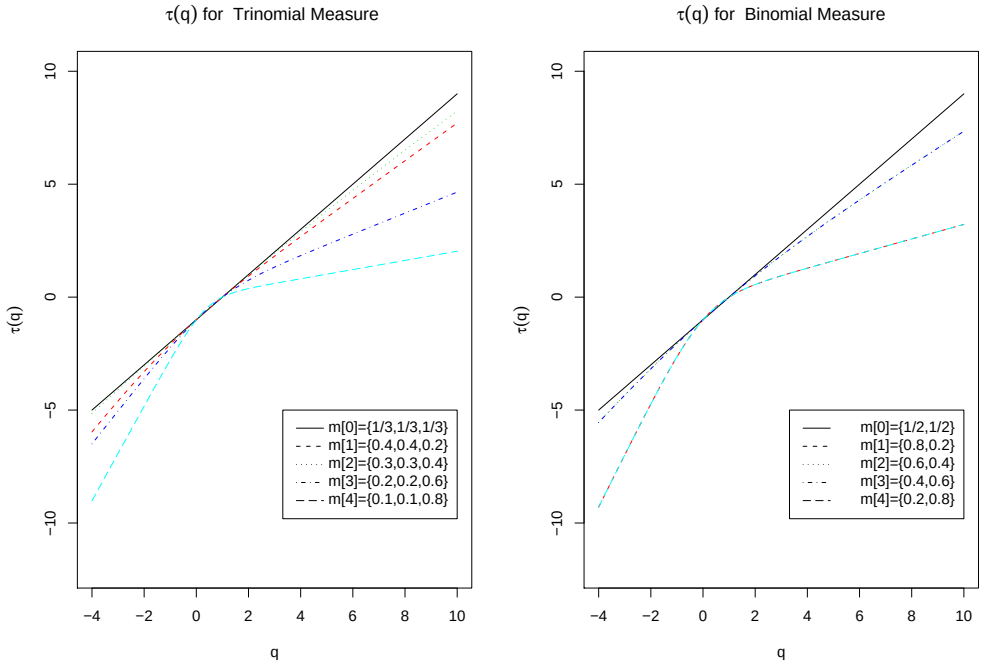

Figure 4.11: $\tau(q)$ for Trinomial and Binomial Measure: The solid line represents the binomial specification of $m_0 = m_1 = 0.5$, while $m_0 = m_1 = m_2 = 1/3$ in case of trinomial specification. Each measure is normalized so that the their sum is one.

In the right panel of figure (4.10) and (4.11) we see that the binomial scaling function takes differentiated values for some combination of parameters. Note that the scaling function is the same for the combinations of binomial parmeters: $m[1]$ and $m[4]$. However, the scaling function of the other combinations $m[2]$ and $m[3]$ is clearly different from $m[1]$ and $m[4]$. It is not difficult for the binomial multi-fractal model to identify the parameter m_0 on the base of the scaling fuction $\tau(q)$. The empirical scaling function agrees with the theoretical one: $\tau(q) = -1 - \log_2\{(\frac{m_0}{2})^q + (\frac{m_1}{2})^q\}$. This unique identification of one parameter is attractive and the binomial multi-fractal process, therefore, is often used in the physics literatur.

But the left panels show that the scaling functions of trinomial and quadronomial multi-fractal process are very close for the first two combinations of parameters (see $m[1]$ and $m[2]$ in each figure). For some moments they are nearly similar to mono-scaling. Note that the combination of parameters $m[4]$ of figure (4.10) and (4.11) gives a clear divergence from mono-scaling.

Estimating for example the combination of quadronomial multi-fractal parameters ($m[3]$), we can not identify exactly for these parameters which one belongs to m_0 or m_1 or m_2 since the parameters are identical. Even though there are some differences between quadronomial parameters, the identification is not easy for GMM (and passibly ML) because the multiplied moments are as input used. We point out that the whole variation of volatility could be well traced by the multiplication of estimated parameters with unconditional variance.

We summarize the multi-fractal parameters, the scaling parameter, the number of volatility

states and the predetermined transition probability for the three different discrete MSMF model in table (4.16).

Model	Parameter	No. of volatility state	transition probability
Binomial MSMF	m_0, σ	$2^{\bar{k}}$	$\gamma_k = 1 - (1 - \frac{1}{2})^{2^{k-\bar{k}}}$
Trinomial MSMF	m_0, m_1, σ	$3^{\bar{k}}$	$\gamma_k = 1 - (1 - \frac{1}{3})^{3^{k-\bar{k}}}$
Quadronomial MSMF	m_0, m_1, m_2, σ	$4^{\bar{k}}$	$\gamma_k = 1 - (1 - \frac{1}{4})^{4^{k-\bar{k}}}$

Table 4.16: The estimated parameters, the number of volatility states and the transition probability of three different multinomial MSMF model. Note that there is a hidden parameter in each model: $m_1 = 2 - m_0$ for binomial case, $m_2 = 3 - m_0 - m_1$ for trinomial model, and $m_3 = 4 - m_0 - m_1 - m_2$ for quadronomial process, respectively.

The considered discrete MSMF models, i.e. binomial, trinomial and quadronomial MSMF models, incorporate multi-scaling, volatility clustering and long range dependence. We show that with respect to the efficent and consistent estimation, binomial specification will be adquate for some financial time series, whereas trinomial or quadronomial MSMF model can be used for distinguishing fine multi-scaling effects in the underlying series.

In paricular, the latters can be applied for various degree of long range dependence through fine tuning of the multinomial parameters. This is one issue we have to encounter in practice: there is a dilemma between fine tuning of multi-scaling and some inconsistency of estimation when adding one more parameter to the underlying MSMF model. The estimation of trinomial parameters is not as stable as that of binomial one. For quadronomial MSMF it leads to more unstable results. The biased estimation and its relatively large standard errors are the price for complexity. But the following question gives a way to solve this difficulty. How much does adding or skipping of one paramter in a MSMF model change the performance of volatility forecasting? We can justify this kind of model selection by comparison of performances of volatility forecasting based on different MSMF specifications.

Monte Carlo GMM/ML results with Quadronomial MSMF($\bar{k} = 3$)

A Panel: GMM									
T	2500	5000	10.000	2500	5000	10.000	2500	5000	10.000
m_0	1.3			1.4			1.5		
$\hat{m}_0$	1.224	1.229	1.238	1.272	1.273	1.279	1.331	1.341	1.387
FSSE	0.156	0.156	0.135	0.232	0.209	0.196	0.232	0.222	0.204
RMSE	0.174	0.171	0.148	0.265	0.245	0.230	0.287	0.273	0.233
m_1	0.5			0.6			0.7		
$\hat{m}_1$	0.605	0.616	0.573	0.598	0.689	0.700	0.727	0.698	0.687
FSSE	0.278	0.261	0.235	0.174	0.277	0.257	0.169	0.154	0.131
RMSE	0.297	0.285	0.246	0.200	0.291	0.276	0.212	0.182	0.157
m_2	1.0			1.1			1.2		
$\hat{m}_2$	1.005	1.017	1.039	1.145	1.072	1.068	1.238	1.248	1.257
FSSE	0.148	0.143	0.132	0.231	0.181	0.161	0.290	0.293	0.279
RMSE	0.148	0.144	0.138	0.238	0.184	0.164	0.296	0.297	0.283
σ	1								
$\hat{\sigma}$	0.997	1.000	1.000	0.989	0.996	0.996	0.993	0.996	0.997
FSSE	0.065	0.055	0.044	0.071	0.058	0.023	0.055	0.043	0.026
RMSE	0.065	0.055	0.044	0.072	0.058	0.023	0.055	0.043	0.026
B Panel: ML									
m_0	1.3			1.4			1.5		
$\hat{m}_0$	1.217	1.204	1.186	1.303	1.318	1.330	1.449	1.470	1.475
FSSE	0.142	0.135	0.138	0.172	0.159	0.142	0.171	0.151	0.125
RMSE	0.164	0.166	0.179	0.198	0.179	0.158	0.178	0.154	0.128
m_1	0.5			0.6			0.7		
$\hat{m}_1$	0.517	0.506	0.500	0.641	0.631	0.613	0.767	0.751	0.722
FSSE	0.056	0.029	0.016	0.098	0.085	0.058	0.145	0.129	0.097
RMSE	0.059	0.110	0.016	0.106	0.156	0.059	0.159	0.199	0.100
m_2	1.0			1.1			1.2		
$\hat{m}_2$	1.080	1.101	1.110	1.117	1.129	1.141	1.147	1.152	1.191
FSSE	0.182	0.149	0.129	0.175	0.158	0.148	0.228	0.204	0.157
RMSE	0.199	0.149	0.170	0.176	0.174	0.153	0.234	0.252	0.158
σ	1								
$\hat{\sigma}$	0.996	1.002	1.003	0.995	0.999	0.999	0.994	0.998	0.997
FSSE	0.038	0.026	0.018	0.039	0.028	0.020	0.047	0.035	0.024
RMSE	0.038	0.026	0.018	0.039	0.028	0.020	0.048	0.035	0.025

Table 4.17: The Monte Carlo results are based on 400 simulated paths for each column in panels (A) and (B).The columns are distinguished by the number of observations of $T = 2500, 5000, 10000$ and the combinations of parameters. In each set of three rows, the first one presents the average GMM/ML value. The second one is finite sample standard error (FSSE) and the third one gives root mean squared error (RMSE) for GMM/ML estimates.

Chapter 5

Extensions of Markov Switching Multifractal

It is widely accepted that the unconditional return x_t of equation (4.1) has fatter tails than the normal distribution. In the framework of MSMF, this large kurtosis of the return distribution is mainly caused by the multi-fractal term σ_t. Can an additional thick tails provide a better fit if we replace the Gaussian noise u_t in equation (4.1) by the leptokurtic distribution ϵ_t? Which effects does an alternative noise have on the estimation of multi-fractal parameters? These are questions we will answer in the next sections.

As reported by Bollerslev, Chou, and Kroner (1992) and Pagan (1996), the use of the Student's-t distribution is widespread in the financial econometrics. First of all, Bollerslev (1987), Baillie and Bollerslev (1989) and Palm and Vlaar (1997) among others show that this distribution better captures the observed kurtosis.[1] As Nelson (1990) notes in the discussion of the continous time conditionally normal GARCH $(1, 1)$ model, the innovations are approximately t-distributed if one observes them over short time intervals. We will therefore reconstruct the MSMF process by using Student's t-distributed errors (hereafter, t-MSMF model). After the model construction, we derive some moments of the process that are necessary to estimate the parameters based on GMM as well as ML. We will discuss about the Monte Carlo results. We will present the empirical results of the ML estimation using real data, too.

5.1 The MSMF Model with Student's-t innovation

Using the noise ϵ_t from a Student's-t distribution with mean zero, variance $\sigma_s = 1$ and the degree of freedom ν, the Markov switching multi-fractal model of equation (4.1) can be written as

$$\begin{aligned} x_t &= \sigma_t \, \epsilon_t \\ &= \sigma \sqrt{\prod_{k=1}^{\bar{k}} m_{t,k}} \, \epsilon_t \end{aligned} \tag{5.1}$$

[1] Bollerslev (1987) notes that the GARCH with a conditional Student's-t distribution permits a distinction between condidtional heteroskdasticity and a conditional leptokurtic distribution, either of which could acount for the observed unconditional kurtosis in the data.

The density function of Student's-t distribution $f_s(\nu)$ is given by

$$f_s(\nu) = \frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\Gamma\left(\frac{1}{2}\nu\right)} \left\{ \pi(\nu-2)\left(1+\frac{x^2}{\nu-2}\right)^{\nu+1} \right\}^{-1/2}, \quad \nu > 2 \tag{5.2}$$

Γ stands for the gamma function. The Student's t-distribution is symmetric about zero. The parameter ν serves to capture leptokurtosis. If $\nu < 2$, no variance is defined. For $\nu \to \infty$ the Student's-t distribution converges to a normal distribution as well known. In other words, the large values of ν may be interpreted as an absence of kurtosis. The implied kurtosis of the Student's-t distribution is $6/(\nu-4)+3$ for all $4 < \nu < \infty$. Later, the value of ν is estimated together with parameters m_0, m_1 (in the binomial case only m_0) and σ. This assures that both the multiplicative volatility process in the root of equation (5.1) and the error process $\sigma\epsilon_t$ are statistically independent on each other.

5.2 Moments of t-MSMF model

Since the part of σ_t in equation (5.1) is defined in the same as the MSMF model with gaussian noise, there is no change in the unconditional moments and the autocorelation function of the multi-fractal process. See the section 4.2. We must renew only the moments of the absolute asset returns. With the aid of the abosolute moments $E[|\epsilon_t|^q]$ of the Stundent's t, we get the moments of absolute returns $E[|x_t|^q] = E[|\sigma_t|^q]E[|\epsilon_t|^q]$ as follows:

$$\begin{aligned}
E[|x_t|] &= \sigma\sqrt{\frac{\nu}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-1)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{3}m_0^{1/2}+\frac{1}{3}m_1^{1/2}+\frac{1}{3}m_2^{1/2}\right)^{\bar{k}} && \text{for } \nu > 1 \\
E[|x_t|^{1.5}] &= \sigma^{1.5}\; 0.511383\;\nu^{0.75}\;\frac{\Gamma\left[\frac{\nu}{2}-0.75\right]}{\Gamma\left[\frac{\nu}{2}\right]} && \text{for } \nu > 1.5 \\
E[|x_t|^2] &= \sigma^2\left(\frac{\nu}{\nu-2}\right) && \text{for } \nu > 2 \\
E[|x_t|^3] &= \sigma^3\sqrt{\frac{\nu^3}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-3)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{3}m_0^{3/2}+\frac{1}{3}m_1^{3/2}+\frac{1}{3}m_2^{3/2}\right)^{\bar{k}} && \text{for } \nu > 3
\end{aligned} \tag{5.3}$$

These moments are used for the trinomial t-MSMF process. Furthermore, the first and third moments of absolute returns are represented by

$$\begin{aligned}
E[|x_t|] &= \sigma\sqrt{\frac{\nu}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-1)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{2}m_0^{1/2}+\frac{1}{2}m_1^{1/2}\right)^{\bar{k}} && \text{for } \nu > 1 \\
E[|x_t|^3] &= \sigma^3\sqrt{\frac{\nu^3}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-3)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{2}m_0^{3/2}+\frac{1}{2}m_1^{3/2}\right)^{\bar{k}} && \text{for } \nu > 3
\end{aligned} \tag{5.4}$$

for the binomial t-MSMF process and by

$$\begin{aligned}
E[|x_t|] &= \sigma\sqrt{\frac{\nu}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-1)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{4}m_0^{1/2}+\frac{1}{4}m_1^{1/2}+\frac{1}{4}m_2^{1/2}+\frac{1}{4}m_3^{1/2}\right)^{\bar{k}} && \text{for } \nu > 1 \\
E[|x_t|^3] &= \sigma^3\sqrt{\frac{\nu^3}{\pi}}\,\frac{\Gamma\left[\frac{1}{2}(\nu-3)\right]}{\Gamma\left[\frac{1}{2}\nu\right]} \cdot \left(\frac{1}{4}m_0^{3/2}+\frac{1}{4}m_1^{3/2}+\frac{1}{4}m_2^{3/2}+\frac{1}{4}m_3^{3/2}\right)^{\bar{k}} && \text{for } \nu > 3
\end{aligned} \tag{5.5}$$

for the quadronomial t-MSMF process, respectively. The second and fourth moments have the same form as for the trinomial t-MSMF. Due to equation (5.3) the variance $Var[|x_t|]$ of trinomial t-MSMF results in

$$\begin{aligned} Var[|x_t|] &= E\left[\left|\sigma_t\,\epsilon_t\right|^2\right] - \left(E\left[\left|\sigma_t\,\epsilon_t\right|\right]\right)^2 \\ &= \sigma^2\left\{\left(\frac{\nu}{\nu-2}\right)^2 - \frac{\nu}{\pi}\left(\frac{\Gamma\left[\frac{1}{2}(\nu-1)\right]}{\Gamma\left[\frac{1}{2}\nu\right]}\right)^2\left(\frac{1}{3}m_0^{1/2}+\frac{1}{3}m_1^{1/2}+\frac{1}{3}m_2^{1/2}\right)^{2\bar{k}}\right\} \end{aligned} \tag{5.6}$$

The autocovariance and the squared autocovariance of the compound t-MSMF process are calculated by means of equation (4.48) and equation (4.49), where the $E[|\ln u_t|^q]\ q = 1, \ldots, 4$ are simply substituted by the log absolute moments of the Student's-t distribution $E[|\ln \epsilon_t|^q]$. The explicit forms of $E[|\ln \epsilon_t|^q]$ are found in the appendix C.[2] Like MSMF with Gaussian innovation, the log differences of the compound t-MSMF process yields a stationary stochastic process that has no 'true' long-range dependency.

5.3 Monte Carlo results: ML vs. GMM vs. MM

We perform the Monte Carlo simulations for the trinomial and binomial t-MSMF model to estimate the parameters m_0, m_1, σ and ν. Thereby, we apply again the GMM and the ML as well as the simple method of moments(MM). In MM, we fit the empirical unconditional moments to the theoretical values of $E[|x_t|^q]\ q = 1,\ 1.5,\ 2,\ 3$ using the GAUSS module *nonlinear equations*[3].

GMM I and GMM II

The estimation function $f(x_t, \theta)$ of equation (4.39) is changed in the following two different way. The one is

$$f(x_t,\theta) = \begin{bmatrix} E[\xi_{t+T,T}\xi_{t,T}] - \xi_{t+T,T}\xi_{t,T} \\ E[\xi^2_{t+T,T}\xi^2_{t,T}] - \xi^2_{t+T,T}\xi^2_{t,T} \\ E[|x_t|] - x_t \end{bmatrix}. \tag{5.7}$$

The other reads

$$f(x_t,\theta) = \begin{bmatrix} E[\xi_{t+T,T}\xi_{t,T}] - \xi_{t+T,T}\xi_{t,T} \\ E[\xi^2_{t+T,T}\xi^2_{t,T}] - \xi^2_{t+T,T}\xi^2_{t,T} \\ E[|x_t|] - |x_t| \\ E[(x_t - E[x_t])^2] - (x_t - \bar{x}_t)^2 \\ E[|x_t|^3] - |x_t|^3 \end{bmatrix}. \tag{5.8}$$

The time lags T in equation (5.7) and equation (5.8) take the values of 1, 5, 10 and 20, respectively. We call the former GMM I and the latter GMM II. Note that the parameter set to be estimated is given by $\theta \in \{m_0,\ m_1,\ \sigma,\ \nu\}$ for the trinomial t-MSMF. So, we will then see whether this changed forms of $f(x_t, \theta)$ can lead to the better results.

[2] We use Mathematica 4.2 for the calculation of the log absolute moments of the Student's-t distribuion.

[3] The detailed solution methods are described in the manual: nonlinear equations for $GAUSS^{TM}$ from Aptech Sytems, Inc.

Panel	Binomial t-MSMF model ($k=6$)	Trinomial t-MSMF model ($k=3$)
A	$\{m_0,\nu,\sigma\}\in\{1.3,4,1\}$	$\{m_{0,1},\nu,\sigma\}\in\{1.2,0.5,4,1\}$
B	$\{m_0,\nu,\sigma\}\in\{1.3,5,1\}$	$\{m_{0,1},\nu,\sigma\}\in\{1.2,0.5,5,1\}$
C	$\{m_0,\nu,\sigma\}\in\{1.5,4,1\}$	$\{m_{0,1},\nu,\sigma\}\in\{1.3,0.6,4,1\}$
D	$\{m_0,\nu,\sigma\}\in\{1.5,5,1\}$	$\{m_{0,1},\nu,\sigma\}\in\{1.3,0.6,5,1\}$

Table 5.1: Parameter combinations of t-MSMF model for Monte Carlo Simulation

ML

We have already explained in the section 4.4 how to obtain the ML estimates of trinomial MSMF. Hence, we will here give only the modified log-likelihood function $\ln L(\theta|X)$ that is induced by the introduction of the Student's-t innovation:

$$\begin{aligned}\ln L(\theta|X) &= \sum_{t=1}^{T}\ln\sum_{i=1}^{d}\sum_{j=1}^{d} P(\psi_{t-1}^{(i)}|I_{t-1},\theta)\; f(x_t|\psi_t^{(j)},X_{t-1},\theta)\\ &= \sum_{t=1}^{T}\ln\sum_{i=1}^{d}\sum_{j=1}^{d}\psi_{i,t-1|t-1}^{(i)}\\ &\quad\frac{\Gamma\left(\frac{1}{2}(\nu+1)\right)}{\sqrt{\pi(\nu-2)}\,\Gamma\left(\frac{1}{2}\nu\right)\;\hat\sigma\sqrt{\psi_t^{(j)}|\hat m_0,\hat m_1}}\left(\frac{x_t^2}{\left(1+(\nu-2)\hat\sigma\sqrt{\psi_t^{(j)}|\hat m_0,\hat m_1}\right)^{(\nu+1)/2}}\right)\end{aligned}\tag{5.9}$$

In general, the simulation of t-MSMF model requires more numerical effort than that of MSMF with Gaussian innovations, because not only this t-MSMF process involves one more parameter ν, but also the moments of t-MSMF include various special function like polynomial gamma. However, such computational burden can be overcome since the absolute moments of the Student's t-distribution possess a closed form expression. It is also possibe to distingush the effects of the scaling parameter σ from that of the tailness ν due to the plausible assumption of independence between the multi-fractal part and the innovation process.

Monte Carlo results of the binomial and trinomial $t-$MSMF

We simulate both the binomial and the trinomial t-MSMF model by using four different sets of parameters which is summarized in table (5.1). The simulation generates data of size 10^5. From these data, we choose two samples which have lengths of 5000 and 10000, respectively. This procedure is repeated 400 times. Based on the estimation of the optimal $\bar k$ with respect to MSC, we set $\bar k=3$ for the trinomial t-MSMF model, whereas $\bar k=6$ for the binomial model. The results are arrayed in tables (5.2), (5.3) and the rest is listed in the appendix C.3. With respect to the asymptotic consistency the results are similar. Hence, we only provide analyses about the case of panel A in table (5.1).

In table (5.2) and (5.3), we find not only the estimates $\hat m_i$, $\hat\sigma$ and $\hat\nu$ but also the root mean squared error (RMSE) as well as the finite sample standard error (FSSE). With increasing sample size, the estimates converges to the true values without bias. Overall, ML provides efficient and good estimates in both t-MSMF models. The FSSE and RMSE are here numerically very low and close to each other.

It is difficult to make a general statement about the Monte Carlo results estimated by GMM. In case of binomial t-M SMF, the GMM I estimate of $\hat m_0$ is downward biased. It is the other way

round if GMM II is utilized. The degree of freedom $\hat{\nu}$ by GMM I is upward biased. Furthermore, the estimate $\hat{\nu}$ using GMM II has the largest bias and RMSE among the four methods. This means that the modified GMM of equation (5.7) and (5.8) could not lead to a reduction of the bias. The scaling parameter σ is almost unbiased with relative small FSSE and RMSE in both case of GMM.

As to the trinomial t-MSMF process, GMM I is also consistent for the estimate of the multi-fractal parameter m_0, m_1 and scaling parameter σ. Only the degree of freedom, $\hat{\nu}$, is upward biased. The standard error and RMSE of this parameter are three or four times larger than those by ML. If GMM II is employed, the parameters m_0 and ν are more upward biased compared with the results from GMM I. In paricular, the FSSE and the RMSE of $\hat{\nu}$ are apparently enhanced. It seems that this upward bias of $\hat{\nu}$ is compensated by a litte downward biased $\hat{\sigma}$.

The MM for the multi-fractal paramter m_0 results in the highest upward bias among the four methods of estimation, while $\hat{m}_1$ in trinomial t-MSMF exhibits substantial downward bias. The degree of freedom and the scaling parameter are well indentified in both t-MSMF processes. Note that there are only 146 valid MM estimations for the trinomial process.

5.4 Empirical estimation

As we have applied the MSC to the trinomial MSMF using empirical data (see the table (4.14)), we have to determine the optimal number of multipliers $\bar{k}$. To do that, it is necessary to calculate the smoothing probability of equation (4.67) which comes as a by-product from ML. Hence, we will focus here on the method of maximum likelihood. Together with the estimation of parameters, we will also conduct the likelihood ratio test $LR_{1/\nu=0}$ to see whether using a Student's-t distribution improves estimation. This test is defined by the ratio of log-likelihood of the MSMF to that of the t-MSMF under the null hypothesis $H_0 : 1/\nu = 0$. The usual LR test statistic is concentrated towards the origin more than a χ^2_1 distribution. The likelihood ratio test on the χ^2_1 distribution leads to a conservative test, implying rejection at even higher levels of significance α, because the value of $1/\nu$ lies on the boundary of the admissible parameter space. To take this boundary effects into the consideration, Harvey (1989) suggests the adequate modified $LR_{1/\nu=0}$ in a form of

$$LR \sim \frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1 \tag{5.10}$$

Here, the additional term χ^2_0 is a degenerate distribution with all its mass at the origin, see Ruiz (1994) and Liesefeld and Jung (2000). For the significance level at 5% the values of $LR_{1/\nu=0}$ is 2.71, while the value of 5.41 relates at the significance level 1%.

The estimation results using the data of GOLD and USD-DM are summarized in table (5.4) and table (5.5), respectively. The other tables (5.6), (5.7) and (5.8) show the results of data from three stock markets NYCI, DAX as well as KOSPI. Compared with the results from MSMF model in the table (4.14) there are no remarkable changes in the estimates $\hat{m}_0$, $\hat{m}_1$, $\hat{\sigma}$ and their standard errors for all data, except for the lower area of frequency. The quantity of $max\,|m_i - m_j|$ for $i = j = 0, 1, 2$ is decreasing with growing $\bar{k}$ which is main indicator for the dynamic volatility. Remember that not the individual one, the combination of all multi-fractal parameters accounts for the dynamic volatility.

For the most part, the estimated values of $1/\nu$ are small: the degree of freedom is large in every frequency! That indicates a strong rejection of the t-MSMF model against the MSMF model. The values of the likelihood ratio test vary between 294.03 ($\bar{k} = 1$, GOLD) and 14.49 ($\bar{k} = 5$, DAX). That is higly significant at any level in the corresponding asymptotic distribution. The only exception is the LR value of 4.16 in the case of DAX with $\bar{k} = 6$, which is significant at the 5% level, but not significant at the 1% level. After MSC, the optimal number of $\bar{k}$ is 2 if DAX is

employed as input.

In individual data, the degree of freedom looks as follows. For the GOLD data, ν varies between $4.7\,(\bar{k}=2)$ and $8.2\,(\bar{k}=6)$ and the implied kurtosis is 4.43 and 11.92, respectively. The standard errors are relatively small in comparision to the other data. For the exchange market data USD the value of ν lies between $12.05\,(\bar{k}=1)$ and $18.18\,(\bar{k}=2)$. For the stock market data NYCI the maximum of ν is $9.17\,(\bar{k}=1)$, while the minimum of that is $12.82\,(\bar{k}=2)$. For DAX the estimated value of ν changes between $11.23\,(\bar{k}=1)$ and $3668.01\,(\bar{k}=4)$. The estimated values of $\hat{\nu}$ for KOSPI are all more than 30. Too high degree of freedom signals an absence of fat-tails in the conditional distribution. Therefore, the Gaussian innovation might be appropriate for DAX and KOSPI even though the $LR_{1/\nu=0}$ statistics for these data are significant. The standard errors on $\hat{\nu}$ for USD, NYCI, DAX, KOSPI are too large for significance of this parameter.

When the Student's-t noise is introduced, the estimated value of the scaling parameter σ is smaller than that of the MSMF model with Gaussian innovations for almost every frequency.[4] The results using the data NYCI, USD and GOLD exhibit a significant effect of this fat tailed noise on σ. With respect to DAX and KOSPI, there are no noticeable differences between the estimates m_i by gaussian MSMF and that by t-MSMF due to the small value of $1/\hat{\nu}$.

A remaining problem that also arises in this application is how to determine the optimal number of multipliers. We will again use the Markov switching criterion (MSC) that combines a measure of goodness of fit with a penalty for model complexity. We select the t-MSMF model with lowest MSC_1 and mark it by † in the tables. According to MSC_1 we choose the optimal frequency as $\bar{k}=2$ for GOLD, USD, DAX and $\bar{k}=3$ for NYCI and KOSPI. Remember that the optimal $\bar{k}$ is 3 after MSC_1 for all the data when using Gaussian innovations. In other words, the t-MSMF model with low number of $\bar{k}$, i.e. small number of states is enough to account for the dynamics of volatility in the time series.

Our result is consistent with those from other models like t-GARCH and stochastic volatility models. Bollerslev (1987) shows that the GARCH with Student's-t distribution is superior to the normal GARCH. Ruiz (1994) extends the stochastic volatility model with heavey-tailed distribution. Sandman and Koopman (1998) and Chib, Nardari, and Shepard (2002) develop Markov chain Monte Carlo based Bayesian inference in the stochastic volatility model with Student's-t distribution. It is worth to note that Liesefeld and Jung (2000) find a strong evidence in favour of the stochastic volatility model with Student's-t distribution for the US Dollar/Deutsche Mark, the US Dollar/Japanese Yen, the S&P500 stock price index, the IBM stock prices for NYSE, and the Siemens and Daimler-Benz stock prices.

Suggestions

The ML estimates of MSMF model can be easily adjusted to incorporate other heavy tailed distributions. For example, the MSMF model can be generalized to allow the disturbance, u_t in $x_t = \sigma_t u_t$ to follow the generalized error distribution (hereafter, GED), see Liesefeld and Jung (2000) for the stochastic volatility model with this distribution. This specification can be compared with the t-MSMF model and the normal MSMF model concerning its ability to capture the observed distributional and dynamic patterns of the return series. Since the t-MSMF model and the GED-MSMF model are non-nested, the Vuong test can be applied to compare these specifications.

Until now, we have considered only simpler version: MSMF model with symmetric innovations. Our range of MSMF models can be extended to a more general model allowing for skewed distribution of empirical financial time series. We propose the MSMF with the skewed Gaussian or skewed Student's innovations as a unified way of dealing with skewness, long memory, and fat tails.

[4] Liesefeld and Jung (2000) also find that the estimates of the variance σ under the stochastic volatility model with Student's-t innovation are smaller than those under the volatility model with Gaussian distribution.

Binomial MSMF($\bar{k} = 6$) with Student's-t innovation(1)

Panel A: Parameters $m_0 = 1.3$, $\nu = 4$ and $\sigma = 1$								
method	ML		GMM I		GMM II		MM*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m	1.3							
$\hat{m}_0$	1.297	1.298	1.263	1.271	1.346	1.339	1.396	1.397
FSSE	0.018	0.014	0.100	0.077	0.033	0.027	0.032	0.025
RMSE	0.018	0.014	0.107	0.082	0.056	0.048	0.101	0.100
ν	4							
$\hat{\nu}$	4.029	4.014	4.512	4.258	5.049	4.718	4.226	4.182
FSSE	0.321	0.222	0.908	0.720	0.668	0.536	0.559	0.331
RMSE	0.322	0.222	1.043	0.765	1.244	0.896	0.603	0.377
σ	1							
$\hat{\sigma}$	1.001	1.001	1.014	1.002	1.072	1.058	1.083	1.081
FSSE	0.041	0.029	0.082	0.068	0.060	0.047	0.053	0.038
RMSE	0.041	0.029	0.083	0.068	0.093	0.074	0.098	0.090

Table 5.2: Simulated average, finite sample standand error (FSSE) and root mean squared error (RMSE) based on ML, GMM I, GMM II and MM. The parameters are so assigned that $m_0 = 1.3$, $\nu = 4$ and $\sigma = 1$. MM*: there are 398 valid estimations for MM.

Trinomial MSMF($\bar{k} = 3$) with Student's-t innovation(1)

Panel A: Parameter $m_0 = 1.2$, $m_1 = 0.5$, $\nu = 4$ and $\sigma = 1$								
method	ML		GMM I		GMM II		MM*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m_0	1.2							
$\hat{m}_0$	1.178	1.186	1.227	1.228	1.327	1.327	1.774	1.612
FSSE	0.097	0.083	0.165	0.150	0.136	0.130	0.464	0.350
RMSE	0.099	0.084	0.168	0.152	0.186	0.182	0.738	0.540
m_1	0.5							
$\hat{m}_1$	0.506	0.504	0.497	0.492	0.542	0.543	0.362	0.369
FSSE	0.023	0.017	0.088	0.068	0.100	0.092	0.037	0.027
RMSE	0.024	0.017	0.088	0.069	0.108	0.101	0.143	0.134
ν	4							
$\hat{\nu}$	4.051	4.027	4.453	4.310	4.920	5.118	4.087	4.077
FSSE	0.283	0.199	0.750	0.521	1.777	2.002	0.215	0.032
RMSE	0.288	0.201	0.838	0.606	2.001	2.293	0.232	0.084
σ	1							
$\hat{\sigma}$	0.997	0.996	1.129	1.098	0.911	0.899	1.059	1.056
FSSE	0.040	0.027	0.162	0.149	0.370	0.348	0.047	0.030
RMSE	0.040	0.027	0.207	0.179	0.380	0.362	0.075	0.064

Table 5.3: Simulated average, finite sample standand error (FSSE) and root mean squared error (RMSE) based on ML, GMM I, GMM II and MM. The value of parameters are given by $m_0 = 1.2$, $m_1 = 0.5$, $\nu = 4$ and $\sigma = 1$, respectively. MM*: there are only 146 valid estimations for MM.

ML Estimates of Trinomial t-MSMF($\bar{k}$) model and Markov-Switching Criterion(Ia)

GOLD						
k	1	2	3	4	5	6
$\hat{m}_0$	2.649	2.149	1.936	1.762	1.667	1.571
	(0.019)	(0.035)	(0.035)	(0.031)	(0.047)	(0.032)
$\hat{m}_1$	0.441	0.631	0.704	0.787	0.803	0.571
	(0.062)	(0.225)	(0.465)	(0.779)	(1.611)	(0.060)
$1/\hat{\nu}$	0.179	0.122	0.160	0.187	0.203	0.214
	(0.163)	(0.263)	(0.197)	(0.166)	(0.151)	(0.136)
$\hat{\sigma}$	1.143	1.261	1.098	1.046	0.959	0.953
	(0.046)	(0.069)	(0.061)	(0.062)	(0.090)	(0.088)
ln L	-6306.146	-6216.872	-6163.049	-6128.165	-6106.307	-6091.136
AIC_{snt}	12642.291	$12631.744^{\dagger}$	13892.099	25702.330	131282.614	1077980.272
AIC	12616.291	12437.744	12330.099	12260.330	12216.614	$12186.272^{\dagger}$
SIC	12612.291	12433.744	12326.099	12256.330	12212.614	$12182.272^{\dagger}$
MSC_1	17054.341	$16912.191^{\dagger}$	16916.256	17207.550	18596.939	28638.302
MSC_N	$17078.523^{\dagger}$	17241.407	21350.503	990641.539	$2.2 \cdot 10^6$	$6.5 \cdot 10^6$
MSC_{N^2}	$17151.877^{\dagger}$	21294.248	$7.2 \cdot 10^6$	$5.8 \cdot 10^7$	$5.2 \cdot 10^8$	$4.7 \cdot 10^9$
$LR_{1/\nu=0}$	294.03	138.87	86.63	74.94	78.82	93.55

Table 5.4: The results of ML estimation are reported in each row, i.e. the number of frequencies $\bar{k}$ with standard errors in the parentheses. The Markov-Switching-Criterion(MSC) are calculated with smoothing probability of each state. MSC_{λ_i} stands for $\lambda_i = \{1, N, N^2\}$. We denote the optimal frequency by † in each MSC's and AIC. Note that $\text{AIC}_{snt} = -2\ln L + 2(N \cdot \kappa + N^2)$.

ML Estimates of Trinomial t-MSMF($\bar{k}$) model and Markov-Switching Criterion(IIa)

USD						
k	1	2	3	4	5	6
$\hat{m}_0$	2.005	1.685	1.622	1.480	1.475	1.482
	(0.030)	(0.033)	(0.031)	(0.035)	(0.041)	(0.038)
$\hat{m}_1$	0.758	0.948	0.930	0.514	0.559	0.582
	(0.490)	(5.864)	(4.064)	(0.009)	(0.047)	(0.069)
$1/\hat{\nu}$	0.083	0.055	0.000	0.058	0.063	0.065
	(0.194)	(0.317)	(0.073)	(0.311)	(0.287)	(0.284)
$\hat{\sigma}$	0.628	0.642	0.672	0.608	0.594	0.600
	(0.016)	(0.019)	(0.023)	(0.028)	(0.046)	(0.064)
ln L	-5892.069	-5817.921	-5783.694	-5756.299	-5750.262	-5749.232
AIC_{snt}	$11814.137^{\dagger}$	11833.843	13133.389	24958.598	130570.525	1077296.464
AIC	11788.137	11639.843	11571.389	11516.598	11504.525	$11502.464^{\dagger}$
SIC	11784.137	11635.843	11567.389	11512.598	11500.525	$11498.464^{\dagger}$
MSC_1	17942.172	$17830.162^{\dagger}$	17872.346	18167.783	19456.735	27943.098
MSC_N	$17966.302^{\dagger}$	18218.207	21679.853	1501181.691	$3.1 \cdot 10^6$	$9.0 \cdot 10^6$
MSC_{N^2}	$18039.264^{\dagger}$	21643.468	$1.0 \cdot 10^7$	$8.1 \cdot 10^7$	$7.2 \cdot 10^8$	$4.7 \cdot 10^9$
$LR_{1/\nu=0}$	211.15	132.96	63.56	46.11	27.69	20.73

Table 5.5: The results of ML estimation are reported in each row, i.e. the number of frequencies $\bar{k}$ with standard errors in the parentheses. The Markov-Switching-Criterion(MSC) are calculated with smoothing probability of each state. MSC_{λ_i} stands for $\lambda_i = \{1, N, N^2\}$. We denote the optimal frequency by † in each MSC's and AIC. Note that $\text{AIC}_{snt} = -2\ln L + 2(N \cdot \kappa + N^2)$.

ML Estimates of Trinomial t-MSMF($\bar{k}$) model and Markov-Switching Criterion(III)

NYCI						
k	1	2	3	4	5	6
$\hat{m}_0$	1.999	1.739	1.601	1.515	1.460	1.446
	(0.024)	(0.021)	(0.020)	(0.019)	(0.017)	(0.015)
$\hat{m}_1$	0.671	0.769	0.597	0.663	0.702	0.765
	(0.218)	(0.495)	(0.073)	(0.132)	(0.219)	(0.375)
$1/\hat{\nu}$	0.109	0.078	0.079	0.082	0.084	0.084
	(0.166)	(0.261)	(0.250)	(0.180)	(0.171)	(0.171)
$\hat{\sigma}$	0.753	0.768	0.754	0.755	0.760	0.716
	(0.018)	(0.022)	(0.023)	(0.026)	(0.047)	(0.028)
ln L	-9470.330	-9377.698	-9311.258	-9272.352	-9254.950	-9245.529
AIC_{snt}	18970.660†	18953.395	20188.517	31990.704	137579.901	1084289.058
AIC	18944.660	18759.395	18626.517	18548.704	18513.901	18495.058†
SIC	18940.660	18755.395	18622.517	18544.704	18509.901	18491.058†
MSC_1	27265.686	27116.631	27093.675†	27357.954	28474.631	34040.693
MSC_N	27289.782†	27411.989	30571.224	$2.2 \cdot 10^6$	$4.3 \cdot 10^6$	$1.2 \cdot 10^7$
MSC_{N^2}	27362.491†	30632.663	$1.4 \cdot 10^7$	$1.0 \cdot 10^8$	$9.8 \cdot 10^8$	$8.8 \cdot 10^9$
$LR_{1/\nu=0}$	267.25	170.69	116.25	73.68	48.54	42.89

Table 5.6: The results of ML estimation are reported in each row, i.e. the number of frequencies $\bar{k}$ with standard errors in the parentheses. The Markov-Switching-Criterion (MSC) are calculated with smoothing probability of each state. MSC_{λ_i} stands for $\lambda_i = \{1, N, N^2\}$. We denote the optimal frequency by † in each MSC's and AIC. Note that $\text{AIC}_{snt} = -2\ln L + 2(N \cdot \kappa + N^2)$.

ML Estimates of Trinomial t-MSMF($\bar{k}$) model and Markov-Switching Criterion(IV)

DAX						
k	1	2	3	4	5	6
$\hat{m}_0$	2.079	1.825	1.748	1.635	1.560	1.557
	(0.032)	(0.029)	(0.022)	(0.018)	(0.020)	(0.016)
$\hat{m}_1$	0.594	0.485	0.614	0.669	0.708	0.708
	(0.128)	(0.016)	(0.170)	(0.211)	(0.275)	(0.270)
$1/\hat{\nu}$	0.089	0.049	0.010	0.000	0.000	0.000
	(0.189)	(0.353)	(1.194)	(1.310)	(0.117)	(0.223)
$\hat{\sigma}$	1.010	1.064	1.101	1.127	1.157	0.931
	(0.036)	(0.055)	(0.041)	(0.044)	(0.055)	(0.036)
ln L	-6154.513	-6093.293	-6046.808	-6025.189	-6018.619	-6021.048
AIC_{snt}	12339.025†	12384.585	13659.617	25496.038	131107.238	1077840.097
AIC	12313.025	12190.585	12097.617	12054.038	12041.238†	12046.097
SIC	12309.025	12186.585	12093.617	12050.038	12037.238†	12042.097
MSC_1	16823.074	16737.032†	16755.674	17071.422	18451.684	24939.267
MSC_N	16847.254†	17039.218	21126.155	998114.692	$2.2 \cdot 10^6$	$6.6 \cdot 10^6$
MSC_{N^2}	16920.587†	21127.915	$7.3 \cdot 10^6$	$5.9 \cdot 10^7$	$5.3 \cdot 10^8$	$4.7 \cdot 10^9$
$LR_{1/\nu=0}$	215.39	111.64	66.25	35.88	14.29	4.16

Table 5.7: The results of ML estimation are reported in each row, i.e. the number of frequencies$\bar{k}$ with standard errors in the parentheses. The Markov-Switching-Criterion (MSC) are calculated with smoothing probability of each state. MSC_{λ_i} stands for $\lambda_i = \{1, N, N^2\}$. We denote the optimal frequency by † in each MSC's and AIC. Note that $\text{AIC}_{snt} = -2\ln L + 2(N \cdot \kappa + N^2)$.

ML Estimates of Trinomial t-MSMF($\bar{k}$) model and Markov-Switching Criterion (V)

KOSPI						
k	1	2	3	4	5	6
$\hat{m}_0$	2.401	2.027	1.814	1.691	1.628	1.670
	(0.019)	(0.020)	(0.023)	(0.024)	(0.033)	(0.024)
$\hat{m}_1$	0.467	0.675	0.774	0.485	0.535	0.620
	(0.025)	(0.194)	(0.575)	(0.006)	(0.029)	(0.113)
$1/\hat{\nu}$	0.032	0.000	0.000	0.000	0.008	0.009
	(0.444)	(0.127)	(0.105)	(0.096)	(4.470)	(2.611)
$\hat{\sigma}$	1.667	1.596	1.524	1.507	1.510	1.447
	(0.049)	(0.041)	(0.052)	(0.064)	(0.131)	(0.093)
ln L	-10466.860	-10345.692	-10284.444	-10252.284	-10238.721	-10232.168
AIC_{snt}	20963.720	20889.385†	22134.888	33950.567	139547.443	1086262.336
AIC	20937.720	20695.385	20572.888	20508.567	20481.443	20468.336†
SIC	20933.720	20691.385	20585.888	20504.567	20477.443	20464.336†
MSC_1	27251.755	27045.698	27033.774†	27318.463	28547.078	37856.404
MSC_N	27275.885†	27343.552	30811.184	$1.5 \cdot 10^6$	$3.2 \cdot 10^6$	$9.2 \cdot 10^6$
MSC_{N^2}	27348.847†	30838.824	$1.0 \cdot 10^7$	$8.3 \cdot 10^7$	$7.4 \cdot 10^8$	$6.6 \cdot 10^9$
$LR_{1/\nu=0}$	303.52	205.31	125.61	68.82	30.61	14.61

Table 5.8: The results of ML estimation are reported in each row, i.e. the number of frequencies $\bar{k}$ with standard errors in the parentheses. The Markov-Switching-Criterion (MSC) are calculated with smoothing probability of each state. MSC_{λ_i} stands for $\lambda_i = \{1, N, N^2\}$. We denote the optimal frequency by † in each MSC's and AIC. Note that $\text{AIC}_{snt} = -2 \ln L + 2(N \cdot \kappa + N^2)$.

Chapter 6

Forecasting of Volatility using MSMF

We will look into the forecasting processes of the volatility in the binomial and trinomial MSMF($\bar{k}$) as well as in benchmark processes such as GARCH and FIGARCH. The basic steps are as follows: We estimate the parameters based on GMM and ML, and then form out-of-sample forecasts using the results of estimation.

At first, we will perform the volatility forecasting for trinomial MSMF(3) by means of the Bayesian updating as well as the linear forecasting. It is particularly interesting to see whether the estimates based on different method of GMM and ML deliver any change in power of linear forecasting. After that, we will investigate the volatility performance for real data presented in the section 4.5.

6.1 Linear Forecasting: Durbin-Levinson procedure

Durbin and Levinson proposed an iterative procedure of linear forecasting for the determination of the weighting function.[1] This forecasting method known as Durbin-Levinson algorithm (DL) can be applied when we know the covariance function of a considered model. Owing to the autocovariance derived in the section 4.3.3, we can employ this method to the trinomial MSMF model, even though we do not have any information on volatility components, i.e. each multiplier.

One-step and multi-step prediction

Consider the zero-mean stationary processes $\{X_t\}$ with covariance $E[X_{t+u}X_t] = \gamma(u)$, $t, u \in \{0, \pm 1, \ldots\}$. Suppose that we will predict X_{n+1} from $X_n, \ldots X_{n-i+1}$. The best linear predictor of X_{n+1} is then written as

$$\hat{X}_{n+1} = \sum_{i=1}^{n} c_{ni} X_{n-i+1} \tag{6.1}$$

and the corresponding mean squared forecasting error ν_n is obtained by

$$\nu_n = E\left[\left(X_{n+1} - \hat{X}_{n+1}\right)^2\right] \tag{6.2}$$

[1] See Morettin (1984), Brockwell and Davis (1991) as well as Brockwll and Dahlhaus (2004) for details.

Then the coefficients c_{ni}'s and ν_n satisfy the following recursions:

$$\begin{aligned} c_{n+1n+1} &= \left\{ \gamma(n+1) - \sum_{i=1}^{n} c_{ni}\gamma(n-i+1) \right\} \nu_n^{-1} \\ c_{n+1i} &= c_{ni} - c_{n+1n+1}c_{nn-i+1} \end{aligned} \tag{6.3}$$

and

$$\nu_{n+1} = (1 - c_{n+1n+1}^2)\nu_n \tag{6.4}$$

where the generalised inverse is $\nu_n^{-1} = 0$ if $\nu_n = 0$ and we use the starting values of $\nu_0 = \gamma(0)$, $c_{00} = 0$ and $c_{11} = \gamma(1)/\gamma(0)$. It is important for MSMF($\bar{k}$) model because this recursive solution does not require the information about each latent volatility component. It depends only on the variance-autocovariance function.

For h-step ahead forecasting, the best linear predictor of X_{n+h} given $X_n, \dots X_{n-i+1}$ can be expressed as

$$\hat{X}_{n+1}^{(h)} = \sum_{i=1}^{n+1} c_{n+1i}^{(h)} X_{n-i+1} \tag{6.5}$$

The Durbin-Levinson algorithm is then generalised as followed:

$$\begin{aligned} c_{nn}^{(h+1)} &= \left[\gamma(n+h) - \sum_{i=1}^{n-1} c_{n-1i}^{(h)}\gamma(n-i+h) \right] \nu_{n-1}^{-1} \\ c_{ni}^{(h+1)} &= c_{ni+1}^{(h)} + c_{n1}^{(h)}c_{n-1i} - c_{nn}^{(h+1)}c_{n-1n-i} \end{aligned} \tag{6.6}$$

and

$$\nu_n^{(h+1)} = \nu_n^{(h)} + \left[\left(c_{n1}^{(h)}\right)^2 - \left(c_{nn}^{(h+1)}\right)^2 \right] \nu_{n-1} \tag{6.7}$$

Using c_{ni} and ν_n from equation (6.3) and equation (6.4) we set the starting values as $c_{ni}^{(1)} = c_{ni}$ and $\nu_n^{(1)} = \nu_n$. The generalised Durbin-Levinson algorithm does not require any inverse matrix so that the computational burden is enormously reduced.

In our application of the Durbin-Levinson algorithm to trinomial MSMF, we consider $x_t^2 - E[x_t^2] = x_t^2 - \hat{\sigma}_t^2$ as a zero-mean stationary process X_t, where x_t is asset returns of equation (4.1) and $\hat{\sigma}_t$ is the estimated standard deviation of the innovations. Therefore, we try to predict squared returns x_t^2 as a usual proxy of volatility.

How much information do we need for prediction when using the DL algorithm? For a long-memory process, it is useful to use as much information as available. As Lux (2004) mentioned, we should use all available data to forecast the binomial and trinomial MSMF($\bar{k}$) process even though the resulting predictions might not change too much beyond some very long lags .

6.1.1 Criteria for Evaluation of Forecasting Performance

The forecasting power of any model can be measured in terms of the mean squared error (MSE) and the mean absolute error (MAE) which are presented by

$$\begin{aligned}\text{MSE} &= \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\sigma}^2_{n+i|n+i-1} - \sigma^2_{n+i}\right)^2 \\ \text{MAE} &= \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\sigma}^2_{n+i|n+i-1} - \sigma^2_{n+i}\right|,\end{aligned} \tag{6.8}$$

where the estimated volatility is denoted as $\hat{\sigma}^2_t$, i is the forecasting horizon and the realized volatility is calculated as the squared returns $\sigma^2 = x^2_t$. The best linear predictor is obtained only if the mean squared error is minimised. Hence, MSE as a loss function is a natural choice. The MAE is also commonly used.

When we will compare the forecasting power of two models, the relative error measures lead to more reliable results, see Armstrong and Fildes (1995). For instance, let's consider the trinomial MSMF($\bar{k}$) and a benchmark model of the historical volatility that has the constant in-sample standard deviation $\bar{\sigma}$. The forecasting accuracy of the former relative to the latter can be evaluated by

$$\begin{aligned}\text{MSPE} &= \frac{\sum_{i=1}^{m}\left(\hat{\sigma}^2_{n+i|n+i-1} - \sigma^2_{n+i}\right)^2}{\sum_{i=1}^{m}\left(\bar{\sigma}^2 - \sigma^2_{n+i}\right)^2} \\ \text{MAPE} &= \frac{\sum_{i=1}^{m}\left|\hat{\sigma}^2_{n+i|n+i-1} - \sigma^2_{n+i}\right|}{\sum_{i=1}^{m}\left|\bar{\sigma}^2 - \sigma^2_{n+i}\right|},\end{aligned} \tag{6.9}$$

where the denominators of the first and second equations represent the MSE and MAE of the historical volatility, respectively. MSPE and MAPE then indicate an improvement or deterioration of the MSMF($\bar{k}$) model relative to the constant variance prediction. There is no difference in the forecasting power of both models when the relative error assumes the value one. If the MSPE and MAPE are below one, we can better forecast the volatility by the competing model and vice versa.

6.1.2 Diebold-Mariano test

The Diebold-Mariano test (DM) is another way to check the relative accuracy of forecasting by two competing models A and B. This test is based on the null hypothesis that there is no difference in the forecasting performance of two competing models, see Diebold and Mariano (1995) for details. Let's introduce the *loss differential* d_j, $j = 1, 2, \ldots, m$ given by

$$d_j = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\sigma}^2_{n+i|n+i-1,A} - \sigma^2_{n+i}\right|^s - \frac{1}{m}\sum_{i=1}^{m}\left|\hat{\sigma}^2_{n+i|n+i-1,B} - \sigma^2_{n+i}\right|^s, \tag{6.10}$$

where $\hat{\sigma}^2_{n+i|n+i-1,A}$ and $\hat{\sigma}^2_{n+i|n+i-1,B}$ means the forecasting of the volatility from the alternative models A and B at time $n+i$. If the goal is to compare the MSEs, s is equal to 2. For the comparison of MAEs we must take $s = 1$. The Diebold-Mariano test is constructed by the average loss differential of $\bar{d} = (d_1 + \ldots + d_m)/m$ as followed:

$$DM = \frac{\bar{d}}{\sqrt{\omega}} \overset{d}{\sim} N(0,1). \tag{6.11}$$

Here, $\sqrt{\omega}$ is the asymptotic standard deviation of the mean loss differential. In practical applications, a sum of the unweighted autocovariances of d_j, denoted $\hat{\gamma}_i(d)$, is used to estimate a consistent ω:

$$\hat{\omega} = \sum_{i=-(h-1)}^{h-1} \hat{\gamma}_i(d), \tag{6.12}$$

where h is the forecast horizon for which the prediction errors are compared.

6.1.3 Monte Carlo simulation

Here, we will show the forecasting power from the trinomial MSMF ($\bar{k} = 3$) relative to historical volatility by means of MSPE and MAPE of equation (6.9). To do that, we firstly generate a data sample with 10.000 observation points using trinomial MSMF ($\bar{k} = 3$). Hereby, we impose the parameter $\theta = (m_0, m_1, \sigma)$ on the same values presented in the subsection 4.3.4. We apply then half of the data to estimate the parameters based on GMM and ML over the in-sample period. The estimates which are held constant over the out-of-sample period, are needed for the forecasting of volatility. And we employ the in-sample standard deviation as historical volatility.

The remainder of 5000 data points will be used to evaluate the forecasting performance out-of-sample. We execute 400 Monte Carlo replications. We have performed not only Bayesian updating, but also linear forecasting on the base of the GMM estimates as well as the ML estimates. We denote the performances by BU, BL1 and BL2, respectively. As mentioned at the beginning of the chapter, we are interested in the potential loss of forecasting efficiency due to these different estimates in a performance of linear forecasting [Lux (2006)]. The results of predictions are displayed in table (6.1). Note that the forecast horizon h varies from one day over 1, 5, 10, 20 days.

In the table, we find the following results: the Bayesian updating and the best linear forecasting are better than the historical average. Furthermore, the MSPE statistics indicates that Bayesian updating provides more accurate forecasts than the linear predictions, see table (6.1). The MAPE statistics also favours Bayesian updating. This is why we call this Bayesian updating the optimal forecast.

On the other side, there is hardly any difference in the MSPE and the MAPE of BL1 / BL2, respectively. It seems that BL1 is, however, marginally better than BL2. Furthermore, we observe that the efficiency of the forecasting performance is related with the values imposed on m_0 and m_1.

For $m_0 = 1.2$, $m_1 = 0.5$ for which $max|m_i - m_j|$, $i = j = 0, 1, 2$ takes the largest value among the combinations, we get the least MSPE as well as the least MAPE. In other words, this combination leads to the best results of forecasting performance. It is due to that the generated MSMF process using these parameters is highly variable and intermittent. Note that the scaling function $\tau(q) = -1 - \log_3\{(m_0/3)^q + (m_1/3)^q + (m_2/3)^q\}$ for $m_0 = 1.1$, $m_1 = 0.9$ is not significantly different than mono-scaling. On the other hand, the performance of volatility forecasting using trinomial MSMF ($\bar{k} = 3$) is almost identical to that of the historical volatility if the parameters $m_0 = 1.1$, $m_1 = 0.9$ are employed and then the generated MSMF is similar to a Gaussian noise, see the first three columns of the table (6.1).

Hence, we can conclude that the more variable and intermittent the underlying stochastic processes the more efficient the performance of volatility forecasting from the trinomial MSMF.

Bayesian vs. Best linear Forecasts for trinomial MSMF(3) model

		$m_0 = 1.2,\ m_1 = 0.5$			$m_0 = 1.3,\ m_1 = 0.6$			$m_0 = 1.4,\ m_1 = 0.7$		
	h	BU	BL1	BL2	BU	BL1	BL2	BU	BL1	BL2
MSPE	1	0.967	0.976	0.976	0.980	0.985	0.984	0.981	0.985	0.984
		(0.007)	(0.004)	(0.004)	(0.006)	(0.003)	(0.003)	(0.006)	(0.003)	(0.004)
	5	0.986	0.992	0.992	0.991	0.995	0.995	0.992	0.995	0.995
		(0.005)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)	(0.004)	(0.002)	(0.002)
	10	0.993	0.997	0.997	0.996	0.998	0.998	0.996	0.998	0.998
		(0.004)	(0.002)	(0.001)	(0.003)	(0.001)	(0.001)	(0.003)	(0.001)	(0.001)
	20	0.998	1.000	0.999	0.999	1.000	1.000	0.999	1.000	1.000
		(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.002)	(0.001)	(0.000)
MAPE	1	0.970	0.974	0.980	0.984	0.985	0.989	0.988	0.987	0.991
		(0.019)	(0.011)	(0.010)	(0.013)	(0.009)	(0.006)	(0.012)	(0.007)	(0.006)
	5	0.987	0.986	0.993	0.993	0.991	0.996	0.995	0.992	0.997
		(0.014)	(0.010)	(0.008)	(0.009)	(0.008)	(0.005)	(0.009)	(0.006)	(0.004)
	10	0.994	0.990	0.997	0.996	0.993	0.998	0.998	0.994	0.999
		(0.011)	(0.010)	(0.007)	(0.007)	(0.008)	(0.004)	(0.007)	(0.006)	(0.003)
	20	0.998	0.991	0.999	0.999	0.994	1.000	1.000	0.995	1.000
		(0.008)	(0.010)	(0.007)	(0.005)	(0.008)	(0.004)	(0.005)	(0.006)	(0.003)

		$m_0 = 1.1,\ m_1 = 0.9$			$m_0 = 1.2,\ m_1 = 0.8$			$m_0 = 1.3,\ m_1 = 0.7$		
	h	BU	BL1	BL2	BU	BL1	BL2	BU	BL1	BL2
MSPE	1	1.000	1.000	1.000	0.996	0.998	0.997	0.988	0.988	0.988
		(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.005)	(0.004)	(0.004)
	5	1.000	1.000	1.000	0.998	0.999	0.999	0.995	0.995	0.995
		(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)
	10	1.000	1.000	1.000	0.999	1.000	1.000	0.997	0.997	0.997
		(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)
	20	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.999
		(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
MAPE	1	1.000	0.997	1.000	0.998	0.996	0.999	0.992	0.989	0.992
		(0.002)	(0.005)	(0.001)	(0.005)	(0.004)	(0.002)	(0.009)	(0.008)	(0.008)
	5	1.000	0.997	1.000	0.999	0.996	1.000	0.996	0.993	0.997
		(0.001)	(0.006)	(0.001)	(0.003)	(0.004)	(0.001)	(0.006)	(0.007)	(0.006)
	10	1.000	0.997	1.000	0.999	0.997	1.000	0.998	0.995	0.998
		(0.001)	(0.006)	(0.001)	(0.002)	(0.004)	(0.001)	(0.005)	(0.006)	(0.004)
	20	1.000	0.997	1.000	1.000	0.997	1.000	1.000	0.995	1.000
		(0.001)	(0.006)	(0.001)	(0.002)	(0.004)	(0.001)	(0.003)	(0.005)	(0.003)

Table 6.1: The table reports the forecasting accuracy of the trinomial MSMF ($\bar{k} = 3$) relative to historical volatility by means of relative mean squared error (MSPE) and the relative mean absolute error (MAPE), where we employ in-sample standard deviation to the constant variance of the naive forecasting of equation (6.9). BU denotes the Bayesian updating using ML estimates. The linear forecasting based on GMM estimates as well as ML estimates is presented by BL1 and BL2, respectively.

6.2 GARCH and FIGARCH

Instead of historical volatility, we introduce GARCH and FIGARCH model as a benchmark model so that we will compare the forecasting performance from these model with that from the MSMF($\bar{k}$) model.

GARCH process

Engle (1982) has introduced the Autoregressive Conditionally Heteroscedastistic model (ARCH) and subsequently Bollerslev (1986) proposed the Generalised ARCH model (GARCH). Refer to Bollerslev et al. (1992) and Bollerslev, Engle, and Nelson (1994) for a comprehensive survey on GARCH processes. In this model, the volatility is represented as a time-varying function of the current information. Therefore, the GARCH process generates realistic mean-reverting volatility and can be used to forecast volatility. Following Bollerslev (1986), Bollerslev and Mikkelsen (1996) suggest the stochastic GARCH process of the orders p and q, or the GARCH(p,q) process written in the form of

$$\begin{aligned} x_t &= \mu + a x_{t-1} + \sqrt{\sigma_t} u_t; \quad u_t \sim N(0,1) \\ \sigma_t^2 &= \omega + \alpha(L) u_t^2 + \beta(L) \sigma_t^2 \end{aligned} \tag{6.13}$$

with $\alpha(L) = \alpha_1 L^1 + \ldots + \alpha_p L^p$ and $\beta(L) = \beta_1 L^1 + \ldots + \alpha_q L^q$. The coefficient a refers to a regression and L denotes the time-lag operator. As equation (6.13) shows, the asset returns x_t in GARCH (p,q) is constituted of the unconditional mean μ, the AR(1) term and the conditional variance σ_t. To ensure that the recursive relationship of equation (6.13) remains positive for all realization of the noise process, the parameters ω, α and β are so assumed that $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$. For stationarity of u_t^2, we suppose additionally that all the roots of the polynomials $1 - \alpha(L) - \beta(L)$ and $1 - \beta(L)$ lie outside the unit circle.[2] This assumption means that the effect of the past squared shocks on the current conditional variance decays exponentially with the length of time-lag.

GARCH forecasting

We use a GARCH(1,1) to model return volatility. For the GARCH(1,1) model, the h-step ahead forecast $\sigma_{t+h|t}^2$ is given by

$$\sigma_{t+h|t}^2 = \sigma^2 + (\alpha + \beta)^{n-1} (\sigma_{t+1|t}^2 - \sigma^2), \tag{6.14}$$

where $\sigma_{t+1|t}^2$ stands for the one-step ahead forecast in the GARCH model. If the conditional mean equals zero and $\alpha + \beta < 1$, the unconditional variance σ may be expressed by

$$\sigma^2 = (1 - \alpha - \beta)^{-1} \omega \tag{6.15}$$

The multi-period forecasts of volatility reverts to the long run unconditional variance at rate $(\alpha+\beta)$, see Campbell et al. (1997) for details. Since the autocorrelation of GARCH processes decays exponentially over time, the GARCH model is not able to explain adequately the long-memory

[2]The GARCH (p,q) process can be formulated as the ARCH process with infinite order, namely ARCH (∞):

$$h_t = \{1 - \beta(L)\}^{-1} \omega + \{1 - \beta(L)\}^{-1} \alpha(L) u_t^2$$

or persistence in the volatility of financial returns [Caporin (2003)]. In a general investigation of fitting of conditional volatility models, the GARCH $(1,1)$ specification is often preferred and we will then use this as the benchmark model.

FIGARCH

As another benchmark, we choose the Fractionally Integrated Generalised Auto Regressive Conditionally Heteroscedastistic (FIGARCH) model to take long-memory of financial time series into consideration. FIGARCH implies a slow, hyperbolic decay for the lagged squared innovations in the conditional variance function. It is more flexible to capture the observed temporal dependence of the volatility in various financial markets. Firstly, the FIGARCH (p,d,q) process is introduced as a generalisation of the Integrated GARCH (IGARCH). Under the assumption that the autoregressive lag polynomial $1-\alpha(L)-\beta(L)$ has a unit root, Baillie et al. (1996) proposed the GARCH (p,q) to be integrated in variance, see Engle and Bollerslev (1986). The IGARCH (p,q) process is then represented by

$$\phi(L)(1-L)u_t^2 = \omega + [1-\beta(L)]v_t, \tag{6.16}$$

where $v_t = \epsilon_t^2 - \sigma_t^2$ refers to the innovation at time t. And $\phi(L) \equiv \{1-\alpha(L)-\beta(L)\}(1-L)^{-1}$ is of order $m-1$ with $m = max(p,q)$. The shocks on the conditional variance persist indefinitely in the IGARCH model.

Replacing $(1-L)$ of equation (6.16) by the fractional difference operator[3]

$$\begin{aligned}(1-L)^d &= 1 - d\sum_{k=1}^{\infty}\Gamma(k-d)\Gamma(1-d)^{-1}\Gamma(k+1)^{-1}L^k \\ &\equiv 1-\delta_d(L),\end{aligned} \tag{6.17}$$

we obtain the FIGARCH(p,d,q) model of

$$\phi(L)(1-L)^d u_t^2 = \omega + [1-\beta(L)]v_t, \tag{6.18}$$

where the fractional difference parameter d takes a value between zero and one. All the roots of $1-\beta(L)$ and $\phi(L)$ must be constrained to lie outside the unit circle. Equation (6.18) reduces to the GARCH (p,q) for $d=0$. We substitute v_t by $(u_t^2-\sigma_t^2)$[4] and then the FIGARCH (p,d,q) can be rewritten in terms of infinite ARCH representation [Robinson (1991)]:

[3]The fractional difference operator can be expressed in terms of the Gaussian hyper-geometric function $F(a,b,c;z)$:

$$(1-L)^d = \sum_{k=0}^{\infty}\frac{\Gamma(k-d)L^k}{\Gamma(k+1)\Gamma(-d)} = F(-d,1,1;L),$$

[4]Analogously to the FIGARCH (p,d,q) process for the fractional integrated variance, the Auto Regressive Fractional Integrated Moving Average, or ARFIMA (p,d,q) process is defined by

$$a(L)(1-L)^{d_0}(y_t-\mu) = b(L)\epsilon_t,$$

where $-0.5 < d_0 < 0.5$ and all the roots of $a(L)$ and $b(L)$ lie outside the unit circle. The fractional difference operator $(1-L)^{d_0}$ is applied to μ in the ARFIMA model, while $(1-L)^d$ is not applied to ω in the FIGARCH model. Note that the concept of fractional integrated process for the mean is somewhat differently applied to models of the fractionally integrated process in variance.

$$\begin{aligned}\sigma_t^2 &= [1-\beta(1)]^{-1}\omega + \left\{1-[1-\beta(L)]^{-1}\phi(L)(1-L)^d\right\}u_t^2 \\ &= [1-\beta(1)]^{-1}\omega + \lambda(L)u_t^2 \end{aligned} \tag{6.19}$$

We truncate the response coefficient $\lambda(L)$ at 1000 lags, i.e. $\lambda(L) = \lambda_1 L^1 + \ldots + \lambda_{1000}L^{1000}$ to save computational burden. Since $\delta_d(1) = 1$ for all $d > 0$, the FIGARCH model of equation (6.19) is not covariance stationary in general: the unconditional variance of FIGARCH models is infinite. To ensure the stationarity of covariances, we have to impose restrictions on the parameters. Baillie et al. (1996) derived the following conditions[5] that the parameters of the FIGARCH $(1, d, 1)$ model should satisfy:

$$\beta_1 - d \leq \phi_1 \leq (2-d)/3 \quad \text{and} \quad d\{\phi_1 - (1-d)/2\} \leq \beta_1(d - \beta_1 + \phi_1).$$

The memory property of the GARCH process depends on the value of d. In case of $d = 0$, the shock on the conditional variance of GARCH process dies out at an exponential rate, while for $d = 1$, the shocks to the variance of IGARCH process remain persistently. Davidson (2004) points out that the conditional variance of IGARCH is not well-defined. The FIGARCH model with $0 < d < 1$ allows long memory in variance, shocks on the conditional variance dies out at an hyperbolic rate. Thus, Davidson (2004) calls it hyperbolic memory. The memory increases as d goes to zero.[6]

FIGARCH forecasting

One- and h-step ahead variance forecasts can be expressed by:

$$\begin{aligned}\hat{\sigma}_{t+1}^2 &= [1-\beta_1]^{-1}\omega + \left[1-\{1-\beta_1\}^{-1}(1-\phi_1 L)(1-L)^d\right]u_t^2 \\ \hat{\sigma}_{t+h}^2 &= [1-\beta_1]^{-1}\omega + \left[1-\{1-\beta_1\}^{-1}(1-\phi_1 L)(1-L)^d\right]u_{t+h-1}^2 \end{aligned} \tag{6.20}$$

Note that the fractional difference parameter provides important information on the pattern of volatility and the rate with which the shocks on the volatility process are propagated. As shown in most practical application, relatively low order GARCH models provide a good approximation of the conditional variance process. We use hence also the low order FIGARCH model, i.e. the FIGARCH $(1, d, 1)$.

[5] Chung (2002) also gives a unique restriction for the parameters :

$$0 \leq \phi \leq \beta \leq d \leq 1.$$

As noted by Chung (2002), this condition is not exactly equivalent with 6.2 but admissible. There may exist a set of parameters values that satisfy the one condition and not the other. see Chung (2002) for details. Even though there is no general methods to ensure the non-negativity of the conditional variances, the GAUSS module "FANPAC" let us to constrain the conditional variances to be nonnegative. See the manual "FANPAC" for details.

[6] Consider that $\rho_k \approx c\,k^{-1+2d_0}$ for $c > 0$. The memory increases as d_0 becomes closer to 0.5 in ARFIMA(p, d_0, q) model with $-0.5 \leq d_0 \leq 0.5$. As we often discuss the long-memory in terms of the Hurst exponent H, which is related to the exponent of time lags in autocorrelation function, the d of the FIGARCH process can be given by $d = 1 - H$, $0 \leq d < 1$. The larger H, i.e. the larger the length of memory, the smaller d. Therefore, Davidson (2004) warns of the wrong interpretation that one characterises the FIGARCH model as an intermediate case between the GARCH and the IGARCH, just as the $I(d)$ process in levels is intermediate between $I(0)$ and $I(1)$.

6.3 Estimation and Volatility Forecasting: GARCH, FIGARCH and MSMF

In this section, we will compare the results of parameter estimations and the relative accuracy of forecasting from the trinomial MSMF ($\bar{k}$), the binomial MSMF ($\bar{k}$), the AR (1) - GARCH $(1, 1)$ and the AR (1) - FIGARCH $(1, d, 1)$ model against historical volatility. For the binomial process, we take the MSMF (10) and MSMF (20) into the consideration, where we apply ML to the former and GMM to the latter for the estimation of the parameters $\theta = (m_0, \sigma)$.[7] In the trinomial model, we estimate $\theta = (m_0, m_1, \sigma)$ of MSMF($\bar{k}$), $\bar{k} = 3, 4, 5$ based on ML as well as on GMM.

As sample data we use the empirical data described in section 4.5. That is, the New York Stock Exchange Composite Index (NYCI), the German Stock Price Index (DAX), the Korean Composite Stock Price Index (KOSPI), the US$-Deutsche Mark Exchange rate (USD) and the price of gold (GOLD). The chosen periods of observations are as followed: the samples of NYCI,DAX,USD and GOLD cover twenty years starting from 1 January 1979 and ending on 31 December 1998. The data of KOSPI covers eighteen years starting from 4 January 1980 and ending on 28 December 1998. We use the data of the year 1979 (1980 for KOSPI) to 1996 for in-sample estimation and the data from the remaining two years for out-of-sample evaluation of volatility forecasts. This gives about $4,400$ in-sample observations and 500 out-of-sample entries. There are slight variations of the data points across markets due to the differences in the number of active days. We consider again the forecast horizon of $h = 1, 5, 20, 50, 100$ days.

6.3.1 Estimation

Before we perform prediction of volatility, we have to estimate the parameters of the considered models. Let's begin with AR (1) - GARCH $(1, 1)$ and AR (1) - FIGARCH $(1, d, 1)$ model. The results are given in table (6.2). For all data, the estimates β of GARCH $(1, 1)$ are relatively high, but are still away from unity. This is not definitely indicative of IGARCH. As a whole, the estimates of μ and a take small values, and are especially significant in the data of NYCI, DAX, KOSPI. According to AIC and BIC, the AR (1) - FIGARCH $(1, 1)$ model describes the data better than the AR (1) - GARCH $(1, 1)$, except for the data of USD.

The estimated values of the fractional difference parameter d lies between 0 and 0.5. This confirms that the dynamics of long-range dependence are better modelled by FIGARCH. In other words, the impact of shocks on the conditional volatility exhibits a hyperbolic decay. It is interesting that the estimated d for the stock markets NYCI, DAX, KOSPI are quite similar in magnitude. The $d = 0.344(0.004)$ for KOSPI is the smallest, while $d = 0.378(0.119)$ for DAX is the largest. It can be said that the behaviour of d reflects the degree of efficiency in stock markets: We may conclude that the German stock market is more efficient than the Korean stock market. The estimates d for the exchange market and that for the gold market are somewhat different. In the case of USD, the estimated value $d = 0.467(0.161)$ for AR (1) - FIGARCH $(1, d, 1)$ is smaller than $d = 0.652(0.160)$ for FIGARCH $(1, d, 1)$ and $d = 0.660(0.043)$ for Moving Average, shortly MA(1) - FIGARCH $(1, d, 1)$ which are reported by Baillie et al. (1996) and Beltratti and Morana (1999), respectively. That means, our specification has captured more long run components of volatility than the other specifications.

Furthermore, Beltratti and Morana (1999) report $d = 0.211(0.010)$ for MA(1)-FIGARCH$(1, d, 1)$, where the data of USD with high frequency of half-hours is employed. In general, the long-memory in conditional variance of *intra-daily* data is much more pronounced

[7] See, Lux (2006) in which he estimated the parameters m_0 of the binomial MSMF(10) based on ML as well as that of the binomial MSMF($\bar{k}$), $\bar{k} = 5, 10, 15, 20$ based on GMM.

than that in conditional variance of *daily* data.[8] As a whole, each financial market seems clearly to have own degree of long-memory in the conditional volatility process.

Table (6.3) exhibits in-sample estimation for the binomial and trinomial MSMF together with standard errors (SE). Related to GMM, we also give the values of *J-statistic*. The notation of ML10 means the results from the binomial MSMF with $\bar{k} = 10$ based on the method of ML. The others notations are given in the same way. From this table, we find the following aspects:

First, the estimates $\hat{m}_0$ tend to decrease with increasing numbers of multipliers. As a whole, the standard errors of the estimates are significantly small, only except for the parameter m_1 of trinomial MSMF model. It is due to the ambiguity in the estimation of m_0 and m_1, even though these parameters are constrained by $1 < m_0 < 3$, $0 < m_1 < 1$.

Second, each market has its unique level of time invariant volatility (see the estimated values of $\hat{\sigma}$ for five different markets). In the MSMF model described as equation (4.1), the volatility σ_t^2 is composed of a time-varying component M_t and time resistant or time invariant component σ^2. In this context, we call the σ a scaling parameter. Thus, we can interpret the different $\hat{\sigma}$ in each data as time invariant own volatility.

Let's more explicitly look into the results of $\hat{\sigma}$. Among the stock markets, the NYCI has the smallest $\hat{\sigma}$. This means that the level of time-invariant volatility is very low in the New York stock market. In particular, the estimate $\hat{\sigma}_{USD}$ is the lowest among the underlying data. With this result, we conjecture that time varying components of volatility will be dominant in this foreign exchange market. Hence, the multi-fractal parameter might play a central role in this case. The estimate $\hat{\sigma}_{GOLD}$ takes the largest value. Furthermore, the estimates of multi-fractal parameters are uniquely differentiated from those of other data. Due to the large difference in $\hat{m}_0$ and $\hat{m}_1$, we expect that the volatility maybe more extremely clustered and the alternation of volatility clustering may be enforced in the Gold market. The volatility in the market is typically *intermittent*. In the context of dynamical systems, we may call it 'intermittent volatility' when the states of volatility are periodically interspersed by erratic bursts. The estimation of the scaling parameter enables us to assess the common (fundamental) level of volatility in each market, whereas the time varying components of volatility are covered by the multi-fractal parameter(s).

Third, there is a lock-in effect. Using the binomial specification, Lux (2006) has mentioned that the estimate m_0 converge to a constant value when the number of multipliers exceeds the threshold of $5 < \bar{k} < 10$. This can be explained as followed: the contribution of multipliers to moments is tailing off above the threshold so that the values of multi-fractal parameters would stay almost constant. Because of this marginal influence of higher multipliers, it should be hard to distinguish between multi-fractal models with different numbers of multiplier, once $\bar{k}$ increases beyond a certain threshold. This is also valid for the both MSMF ($\bar{k}$) models. Nonetheless of the limited range of $\bar{k}$ in the trinomial specification, this kind of lock-in effect can be also observed in (6.3) across five data.

Last, we investigate the probability of the *J-statistic* denoted by J_{prob}. For the trinomial MSMF, we can not reject the null hypothesis of correct specification. However, we find that the statistics of Hansen's test varies marginally with growing $\bar{k}$. Based on J_{prob}, we may select the trinomial MSMF ($\bar{k} = 5$) as the best model for our data. The higher probability of *J-statistics* implies that the binomial models can better fit our selection of moments.[9] Note that it is evaluated by the GMM objective function of equation (4.24).

[8] By taking advantage of this more long-memory effect of variance, the intra-daily trader could make profits than the daily trader without specification of transaction costs. It is possible if financial market is defined as linear efficiency, and allows negligible complicated autocorrelations. See Lillo and Farmer (2004) for the discussion about the long-memory and the market inefficiency.

[9] With the data of various exchange rate, Lux (2006) finds that the log-normal MSMF model can not beat the binomial MSMF model with sufficiently high $\bar{k}$ generally.

6.3.2 Volatility Forecasting

Together with measurement of relative MSE and MAE, we also carry out the DM test of equation (6.11) for the considered models. Note that DM test is based on the null hypothesis that two model have the same forecast accuracy. But, our test establishes the findings that the forecasts from the binomial MSMF(20) and the trinomial MSMF ($\bar{k}$), $\bar{k}$ = 3, 4, 5 are more accurate than those from GARCH, FIGARCH and even binomial MSMF (10) using ML estimates. We summarise the results in tables (6.5) and (6.4) in which $*$ denotes an improvement against the GARCH benchmark which is significant at the 99% level. On the other hand, + stands for an improvement against the FIGARCH benchmark with the same significance level. And ‡ means an improvement of GMM20 relative to the binomial MSMF (10) using ML estimates which is also significant at the 99% level. We will firstly analysis the results of MSPE in table (6.5).

MSPE

For the data of NYCI and DAX, there are actually little variations of MSPEs across all models. The forecasting accuracy from the binomial MSMF (10) using Bayesian updating is slightly better than the other models. The accuracy of prediction from GARCH and that from FIGARCH are almost indifferent.

In case of KOSPI, the FIGARCH model forecasts the volatility of this data better than does GARCH. For the time horizons of h = 20, 50, 100, the forecasting performance of the FIGARCH model is comparable with those of ML10, ML5 and GMM5, while MSMF models provide more accurate forecasts than GARCH, except for the case of GMM3.

Inspecting the MSPEs of USD, the forecasting from FIGARCH is marginally better than that from GARCH. We obtain the best results from the binomial MSMF (20) model for all time horizons. But out-of-sample forecasts using trinomial MSMF ($\bar{k} = 5$) are also competing with those using binomial MSMF ($\bar{k}$), $\bar{k}$ = 10, 20. It is then concluded that the MSMF model can help to capture the underlying nonlinear dynamics of the volatility on the foreign exchange markets (cf. Lux (2006) and Calvet and Fisher (2004)). It is interesting that the predictions from both MSMF models are improved significantly at time horizons h = 50, 100. In other words, the accuracy of long-horizon predictability is increased. Whereas, the forecasting performance from GARCH and FIGARCH is getting worse with increasing time horizons.

Why does the MSMF model provide the best outcome for the exchange market? According to Kilian and Taylor (2003), there is evidence for nonlinear mean reversion of exchange rates to fundamentals. The exchange rates are then detached (attached) from (to) fundamentals by switchings in expectations about future values of the exchange rate. Hereby, the expectations are not only based on fundamentals, but also connected with a time-varying uncertainty caused by the interaction of heterogeneous agents. Such dynamic transition of expectations can be compactly absorbed by the discrete MSMF model. Hence, the risk of this market can be correctly evaluated when we regard the estimation of the scaling parameter as well as multi-fractal parameters.

For the GOLD data, the binomial MSMF ($\bar{k} = 20$) outperforms GARCH, FIGARCH and the other MSMF models. The trinomial MSMF ($\bar{k} = 4$) and MSMF ($\bar{k} = 5$) using ML have slight advantage over GARCH and FIGARCH at h = 50, 100. The forecasting accuracy of ML5 amounts to more than 55% of the historical volatility. Remember that we can illustrate the advantage of these MSMF by means of the scaling function whose shape reflects the degree of multi-scaling effects in a model and is determined by multi-fractal parameters. The scaling parameter and the multi-fractal one have the largest estimated value of $\hat{\sigma} = 1.334$ and $\hat{m}_0 = 1.391$ at ML20, respectively. The estimation of the trinomial MSMF ($\bar{k} = 5$) based on ML results in $\hat{m}_0 = 1.819$ and $\hat{m}_1 = 0.443$ and then the maximal difference between the multi-fractal parameters $max|m_i - m_j|$, i, j = 0, 1, 2

is the largest.

It is also interesting to note that the binomial MSMF ($\bar{k} = 20$) outperforms other trinomial specifications, even though there are only small differences in MSPEs in some cases. With increasing $\bar{k}$, not only the binomial MSMF, but also the trinomial MSMF model improves remarkably the out-of-sample forecasting performance. As we have shown in figure (4.5) and (4.6), the discrete (binomial and trinomial) MSMF model can absorb a larger region of long-range dependence for increasing number of multipliers. Therefore we obtain the improved forecasting performance from these models (see Lux (2006)).

As a whole, the binomial and trinomial MSMF model have an advantage over GARCH as well as FIGARCH at longer prediction horizons. We also observe that MSEs of binomial and trinomial specifications decline with increasing $\bar{k}$ for all h.

MAPE

For the stock markets data, the trinomial MSMF ($\bar{k}$) carries out the best forecasting of volatility among the competing models. The improvement of forecasting accuracy by the trinomial model is significant by means of Diebold-Mariano test. The MAPEs of the trinomial MSMF are unbeatable across all time horizons. Above all, the trinomial MSMF ($\bar{k} = 3$) provides the best results for all stock market data. This conflicts with the results in term of MSPE for which the trinomial MSMF ($\bar{k} = 5$) yields the most reliable forecasting of volatility. GARCH, FIGARCH and the binomial MSMF($\bar{k}$) have values of MAPE which are larger than one. It means that the forecast accuracy of these models are as bad as that of historical volatility.

In case of USD and GOLD, the binomial MSMF ($\bar{k} = 20$) outperforms all other models in terms of MAPE as well as MSPE. In other words, this model delivers further improvements of forecast accuracy, i.e. the lowest relative MSEs and MAEs over all time horizons.

On the other side, it should not be overlooked that the performance by some trinomial MSMF ($\bar{k}$) models is better than that of GARCH and FIGARCH with increasing time horizons. Note that the MAPEs of all models are lower than one. That means, the investigated models are better than the historical volatility. With respect to the long time horizons of $h = 20$, 50, 100, the performance of MSMF (5) is unbeatable among the competing models. Compared with the naive forecasts, the performance of MSMF (5) is improved by about 50% at $h = 20$, 48 percents at $h = 50$ and 51 percent at the time horizon of 100 days.

Let us look into the results for USD in detail. The forecasting performance of FIGARCH is better than that of GARCH. Out-of-sample forecasts from the discrete MSMF ($\bar{k}$) model are much better than the results from GARCH and FIGARCH. Especially for the long time horizons, we can confirm this aspect at 99% significance level according to the DM test.[10]

For GOLD, the relative accuracy of forecasting from MSMF ($\bar{k}$) model becomes more plausible at longer time horizons. This result is consistent with long-memory models like FIGARCH. One would not expect to see such pattern of increased long-horizon predictability. Usually, the MSEs and MAEs of traditional time series models are increasing as we take the larger time horizon for out-of-sample forecasts. Thus, it is surprising for us to observe increased long-horizon predictability. This phenomenon could be understood when considering the scaling function $\tau(q)$ in connection with the large difference between the values of multi-fractal parameters.

As a whole, there are sporadically poor performances of all competing models. It might have some thing to do with our data splitting. Unfortunately, we don't have any guideline how to partition the sample data. We conjecture that it is probably hard for any time series approach to

[10]The DM test uses an asymptotic standard error that is much larger than the Newey-West HAC adjusted standard error. This leads to a more conservative test and therefore we take this normal DM test.

catch the huge increase of volatility in the out-of-sample period 1997/98.

6.3.3 Summary

MSMF models are superior to GARCH and FIGARCH in capturing dynamic features of volatility. In this empirical comparison of forecast accuracy, the binomial or the trinomial MSMF models are beneficial for all our samples collected from various financial markets, even though the outcomes are sensitive to the optimal number of multipliers.

The forecasting performance from binomial ML10 is not better than that of GMM20. Clements and Krolzig (1998) also confirm in the Monte Carlo study of two-state Markov switching model that linear forecasting methods are relatively robust.

In case of trinomial model, the linear forecastings are overall better than Bayesian updating, especially in the stock markets, whereas for the data USD and GOLD we have the opposite situation. In any way, it looks like that the discrete MSMF model can capture multi-scaling more flexibly compared with GARCH or FIGARCH.

Our empirical results suggest that both MSMF models with small numbers of parameters are as easy to fit like other non-linear models like GARCH and FIGARCH. Using the Levison-Durbin algorithm makes it possible to compute h-step forecasts without difficulty.

Although our analysis is limited only to five financial time series, we conclude that the discrete MSMF model is useful to identify the nonlinear dynamics of financial volatility as well as to forecast the volatility. These strong evidence indicates a success of the discrete MSMF model with respect to the higher power of out-of-sample predictability. The optimal frequency after MSC is not always identical with the best MSMF model with the lowest MSPE and MAPE. But it serves as a good starting point at least.

Last but not least, if the MSMF model ($\bar{k}$) is supported by lots of other financial time series, its application should be widely developed. Moreover, we expect to find more decisive empirical results for estimates and forecasts with high-frequency data.

Estimation of GARCH and FIGARCH parameters

Data	Model	μ	a	ω	β	α	ϕ	d	Logl	AIC	BIC
NYCI	GARCH	0.050	0.114	0.013	0.913	0.072			-5225.21	10460.43	10492.54
		(0.010)	(0.016)	(0.003)	(0.010)	(0.007)					
	FIGARCH	0.051	0.112	0.024	0.664		0.442	0.350	**-5199.82**	**10411.65**	**10450.19**
		(0.010)	(0.016)	(0.012)	(0.097)		(0.100)	(0.106)			
DAX	GARCH	0.042	0.077	0.044	0.843	0.125			-6232.18	12474.36	12506.41
		(0.013)	(0.017)	(0.007)	(0.015)	(0.013)					
	FIGARCH	0.045	0.74	0.075	0.344		0.052	0.378	**-6200.93**	**12413.86**	**12452.32**
		(0.014)	(0.019)	(0.030)	(0.080)		(0.044)	(0.119)			
KOSPI	GARCH	0.032	0.146	0.112	0.747	0.172			-7340.01	14690.02	14722.59
		(0.014)	(0.015)	(0.015)	(0.021)	(0.016)					
	FIGARCH	0.034	0.147	0.114	0.223		0.008	0.344	**-7319.42**	**14650.85**	**14689.92**
		(0.001)	(0.000)	(0.007)	(0.036)		(0.034)	(0.004)			
USD-DEM	GARCH	-0.002	-0.042	0.018	0.876	0.095			**-4673.87**	**9357.74**	**9389.68**
		(0.010)	(0.016)	(0.003)	(0.011)	(0.009)					
	FIGARCH	-0.003	-0.040	0.022	0.601		0.215	0.467	-4676.51	9365.02	9403.35
		(0.011)	(0.024)	(0.011)	(0.127)		(0.063)	(0.161)			
GOLD	GARCH	-0.006	0.077	0.024	0.899	0.092			-6648.87	13307.74	13339.71
		(0.014)	(0.017)	(0.003)	(0.009)	(0.009)					
	FIGARCH	0.004	-0.095	0.067	0.600		0.382	0.408	**-6639.15**	**13290.30**	**13328.67**
		(0.023)	(0.022)	(0.075)	(0.397)		(0.260)	(0.241)			

Table 6.2: We use the quasi-maximum likelihood method to estimate the parameters of GARCH and FIGARCH model. Both models have a constant and AR(1) term in the level of returns denoted by μ and a. The remaining parameters are specified in the main text. FIGARCH estimates are based on a truncation lag $T = 1000$ together with 1000 presample values set equal to the sample variance of the time series. Logl is the maximised log-likelihood. AIC and BIC are Akaike and Bayesian information criteria.

Empirical Estimation of Binomial and Trinomial MSMF

		Binomial MSMF		Trinomial MSMF					
Data	Θ	ML10	GMM20	ML3	ML4	ML5	GMM3	GMM4	GMM5
NYCI	$\hat{m}_0$	1.278	1.247	1.693	1.592	1.529	1.569	1.633	1.720
	(SE)	(0.012)	(0.047)	(0.035)	(0.029)	(0.029)	(0.041)	(0.045)	(0.052)
	$\hat{m}_1$			0.536	0.607	0.659	0.715	0.683	0.641
	(SE)			(0.028)	(0.011)	(0.008)	(0.027)	(0.031)	(0.039)
	$\hat{\sigma}$	1.062	0.848	0.847	0.835	0.844	0.787	0.784	0.778
	(SE)	(0.063)	0.070	(0.027)	(0.030)	(0.037)	(0.027)	(0.027)	(0.027)
	J_{prob}		(0.711)				(0.031)	(0.032)	(0.031)
DAX	$\hat{m}_0$	1.252	1.196	1.754	1.626	1.546	1.412	1.439	1.432
	(SE)	(0.011)	(0.033)	(0.025)	(0.023)	(0.022)	(0.158)	(0.117)	(0.097)
	$\hat{m}_1$			0.623	0.687	0.727	0.808	0.780	0.779
	(SE)			(0.052)	(0.038)	(0.032)	(0.102)	(0.073)	(0.059)
	$\hat{\sigma}$	1.151	1.034	1.099	1.077	1.091	0.913	0.940	0.958
	(SE)	(0.090)	(0.063)	(0.029)	(0.033)	(0.038)	(0.043)	(0.047)	(0.049)
	J_{prob}		(0.448)				(0.178)	(0.235)	(0.290)
KOSPI	$\hat{m}_0$	1.302	1.179	1.599	1.521	1.544	1.654	1.561	1.497
	(SE)	(0.013)	(0.010)	(0.044)	(0.054)	(0.081)	(0.038)	(0.033)	(0.030)
	$\hat{m}_1$			0.464	0.515	0.563	0.674	0.720	0.751
	(SE)			(0.019)	(0.022)	(0.039)	(0.019)	(0.017)	(0.015)
	$\hat{\sigma}$	1.215	1.118	1.101	1.092	1.090	1.109	1.106	1.105
	(SE)	(0.130)	(0.029)	(0.032)	(0.048)	(0.082)	(0.028)	(0.028)	(0.028)
	J_{prob}		(0.470)				(0.884)	(0.856)	(0.816)
USD	$\hat{m}_0$	1.269	1.181	1.563	1.496	1.468	1.622	1.546	1.492
	(SE)	(0.013)	(0.010)	(0.041)	(0.045)	(0.049)	(0.040)	(0.036)	(0.032)
	$\hat{m}_1$			0.508	0.569	0.596	0.690	0.728	0.755
	(SE)			(0.023)	(0.027)	(0.029)	(0.020)	(0.018)	(0.016)
	$\hat{\sigma}$	0.720	0.741	0.728	0.714	0.710	0.730	0.733	0.735
	(SE)	(0.083)	(0.020)	(0.021)	(0.029)	(0.046)	(0.020)	(0.020)	(0.020)
	J_{prob}		(0.626)				(0.041)	(0.046)	(0.048)
GOLD	$\hat{m}_0$	1.391	1.312	2.028	1.898	1.819	2.032	1.928	1.829
	(SE)	(0.014)	(0.014)	(0.042)	(0.041)	(0.039)	(0.064)	(0.058)	(0.053)
	$\hat{m}_1$			0.674	0.388	0.443	0.492	0.538	0.586
	(SE)			(0.044)	(0.019)	(0.021)	(0.038)	(0.035)	(0.032)
	$\hat{\sigma}$	1.324	1.443	1.345	1.303	1.329	1.278	1.297	1.281
	(SE)	(0.098)	(0.086)	(0.054)	(0.068)	(0.102)	(0.083)	(0.079)	(0.073)
	J_{prob}		(0.171)				(0.001)	(0.001)	(0.001)

Table 6.3: We estimate $\theta = (m_0, \sigma)$ of the binomial MSMF($\bar{k}$) and $\theta = (m_0, m_1, \sigma)$ of the trinomial MSMF($\bar{k}$). The number after the name of estimation methods denotes the number of multipliers. For instance, ML3 denotes empirical estimates of Trinomial models via maximum likelihood with $\bar{k} = 3$ multipliers and GMM5 is empirical estimates via GMM with $\bar{k} = 5$ multipliers, etc. For GMM, the ten moment conditions have been used: first and second unconditional moment in equation (4.20) and conditional moments in equation (4.48) and (4.49) with $q = 1, 2$ and $T = 1, 2, 5, 20$. SE stands for the standard error of the pertinent estimates. J_{prob} gives the probability of the pertinent J statistic in the case of GMM.

Volatility Forecasts using GARCH, FIGARCH and MSMF: MSPE

		ARCH		Binomial MSMF		Trinomial MSMF					
Data	horizon	GARCH	FIGARCH	ML10	GMM20	ML3	ML4	ML5	GMM3	GMM4	GMM5
NYCI	1	0.943	0.921	0.925	0.929	0.952	0.950	0.951	0.929	0.923	0.922
	5	0.983	0.975	0.963	0.970	0.976	0.976	0.979	0.976	0.968	0.968
	20	1.011	1.004	0.989	1.003	0.992	0.994	1.002	1.008	1.000	0.999
	50	1.011	1.014	0.995^{*}	1.013	0.999	1.008	1.022	1.019	1.012	1.009
	100	1.007	1.015	0.995	1.014	1.027	1.031	1.035	1.020	1.016	1.012
DAX	1	0.779	0.788	0.823	0.849	0.917	0.849	0.842	0.920	0.847	0.830
	5	0.899	0.891	0.884	0.891	0.935	0.880	0.874	0.967	0.908	0.892
	20	0.937	0.935	0.933	0.935	0.953	0.914	0.918	1.013	0.964	0.943
	50	0.976	0.973	0.965	0.973	0.988	0.983	0.994	1.034	1.007	0.986
	100	0.983	0.976	0.969^{+}	0.979	0.992	0.981	0.982	1.037	1.021	1.000
KOSPI	1	0.761	0.735^{*}	0.740	0.748	0.887	0.834	0.773	0.781	0.755	0.744
	5	0.806	0.761^{*}	0.780	0.768	0.901	0.859	0.810	0.841	0.795	0.773
	20	0.938	0.820^{*}	0.811^{*}	0.813^{*}	0.915^{*}	0.880^{*}	0.839^{*}	0.937	0.871^{*}	0.834^{*}
	50	0.988	0.876^{*}	0.847^{*}	0.869^{*}	0.925^{*}	0.893^{*}	0.861^{*}	0.987	0.939^{*}	0.897^{*}
	100	0.993	0.915^{*}	0.913^{*}	0.909^{*}	0.953^{*}	0.935^{*}	0.918^{*}	1.000	0.978^{*}	0.942^{*}
USD	1	0.899	0.891	0.906	0.888	0.894	0.886	0.882	0.910	0.899	0.892
	5	0.899	0.888	0.884	0.878	0.917	0.910	0.905	0.913	0.894	0.885
	20	0.957	0.932	0.920	$0.909^{*\ddagger}$	0.916^{*}	0.915^{*}	0.911^{*}	0.967	0.939	0.921
	50	1.031	0.980^{*}	0.942^{*+}	0.936^{*+}	1.003	0.995	0.988	0.992	0.970^{*}	0.947^{*+}
	100	1.062	1.013^{*}	0.948^{*+}	0.940^{*+}	0.984	0.981	0.978	0.997	0.986^{*+}	0.960^{*+}
GOLD	1	0.370	0.362	0.381	$0.356^{*\ddagger}$	0.497	0.426	0.401	0.463	0.425	0.388
	5	0.386	0.375^{*}	0.385	0.366^{*+}	0.495	0.424	0.400	0.592	0.513	0.432
	20	0.484	0.429^{*}	0.429^{*}	$0.403^{*+\ddagger}$	0.556	0.480	0.452	0.797	0.700	0.549
	50	0.640	0.465^{*}	0.430^{*+}	0.414^{*+}	0.581	0.494^{*}	0.462^{*}	0.905	0.860	0.661
	100	0.941	0.549^{*}	0.475^{*+}	0.461^{*+}	0.531^{*}	0.467^{*+}	0.444^{*+}	0.940	0.987	0.795^{*}

Table 6.4: The table shows the ratio of empirical mean squared errors (MSE) of various models to the MSEs of the benchmark of historical volatility. ML10 means the binomial MSMF with $\bar{k} = 10$ using maximum likelihood. GMM3 is the trinomial MSMF model with $\bar{k} = 3$ estimated by GMM. $*$ denotes an improvement against the GARCH benchmark which is significant at the 99% level, whereas + an improvement against the FIGARCH benchmark with the same significant level. And ‡ stands for an improvement of GMM20 against the ML10 which is also significant at the 99% level. All comparisons are based on the test statistics of Diebold and Mariano (1995).

Volatility Forecasts using GARCH, FIGARCH and MSMF: MAPE

		ARCH		Binomial MSMF		Trinomial MSMF					
Data	horizon	GARCH	FIGARCH	ML10	GMM20	ML3	ML4	ML5	GMM3	GMM4	GMM5
NYCI	1	1.084	1.044^{*}	1.074	$1.058^{*\ddagger}$	1.030	1.042	1.060	0.993^{*+}	1.028^{*+}	1.048^{*}
	5	1.079	1.040^{*}	1.070	1.055^{*}	1.029	1.043	1.061	0.968^{*+}	1.004^{*+}	1.030^{*+}
	20	1.119	1.050^{*}	1.094	1.079	1.062	1.083	1.113	0.964^{*+}	1.004^{*+}	1.039^{*+}
	50	1.079	1.026^{*}	1.083	1.060^{*}	1.062	1.086	1.116	0.954^{*+}	0.983^{*+}	1.015^{*+}
	100	1.032	0.991^{*}	1.061	$1.014^{*\ddagger}$	1.059	1.071	1.082	0.953^{*+}	0.966^{*+}	0.986^{*}
DAX	1	1.051	1.047	1.025	1.029	0.969^{+}	0.999	1.020	0.961^{*+}	0.994^{*+}	1.022
	5	1.071	1.061	1.024	1.041	0.981^{+}	1.024	1.042	0.971^{*+}	0.993^{*+}	1.024^{+}
	20	1.036	1.046	1.025	1.043	0.995^{+}	1.046	1.073	0.992	0.988^{+}	1.011^{+}
	50	1.015	1.040	1.026	1.042	1.023	1.090	1.121	1.009	1.000	1.013^{+}
	100	1.002	1.014	1.003	1.017	1.012	1.056	1.069	1.017	1.007	1.005
KOSPI	1	1.029	1.063	1.016^{+}	1.058	0.971^{*+}	0.974^{*+}	0.990^{*+}	0.991^{*+}	1.017^{+}	1.039^{+}
	5	0.997	1.049	1.038	1.057	0.977^{+}	0.985^{+}	1.010^{+}	0.977^{+}	1.004^{+}	1.029^{+}
	20	0.981	1.044	1.049	1.059	0.986^{+}	0.995^{+}	1.020	0.976^{+}	0.986^{+}	1.016^{+}
	50	0.996	1.042	1.066	1.062	0.987^{+}	0.999^{+}	1.029	0.994	0.988^{+}	1.012^{+}
	100	0.998	1.034	1.119	1.058	1.012	1.034	1.074	1.000	0.994^{+}	1.001^{+}
USD	1	0.822	0.797^{*}	0.792^{*}	$0.777^{*+\ddagger}$	0.841	0.814	0.795^{*}	0.855	0.824	0.801^{*}
	5	0.850	0.815^{*}	0.802^{*+}	$0.782^{*+\ddagger}$	0.852	0.825^{*}	0.807^{*}	0.895	0.850	0.818^{*}
	20	0.946	0.886^{*}	0.841^{*+}	$0.816^{*+\ddagger}$	0.868^{*}	0.843^{*+}	0.826^{*+}	0.961	0.912^{*}	0.866^{*+}
	50	1.031	0.937^{*}	0.853^{*+}	$0.826^{*+\ddagger}$	0.890^{*+}	0.861^{*+}	0.842^{*+}	0.987^{*}	0.948^{*}	0.894^{*+}
	100	1.075	0.999^{*}	0.873^{*+}	$0.842^{*+\ddagger}$	0.862^{*+}	0.837^{*+}	0.820^{*+}	0.995^{*+}	0.976^{*+}	0.925^{*+}
GOLD	1	0.456	0.443^{*}	0.446	0.439^{*}	0.568	0.493	0.463	0.611	0.563	0.508
	5	0.483	0.473	0.465^{*}	0.457^{*+}	0.562	0.487	0.460^{*}	0.733	0.662	0.570
	20	0.599	0.548^{*}	0.525^{*}	0.510^{*+}	0.601	0.526^{*}	0.498^{*+}	0.881	0.813	0.688
	50	0.759	0.607^{*}	0.554^{*+}	0.541^{*+}	0.616^{*}	0.548^{*+}	0.517^{*+}	0.946	0.919	0.786
	100	0.967	0.686^{*}	0.606^{*+}	0.591^{*+}	$0.584^{*}+$	0.518^{*+}	0.491^{*+}	0.967	0.993	0.878^{*}

Table 6.5: The table shows the ratio of empirical mean absolute errors (MAE) of various models to the MAEs of the benchmark of historical volatility. ML10 means the binomial MSMF with $\bar{k} = 10$ using maximum likelihood. GMM3 is the trinomial MSMF model with $\bar{k} = 3$ estimated by GMM. $*$ denotes an improvement against the GARCH benchmark which is significant at the 99% level, whereas $+$ an improvement against the FIGARCH benchmark with the same significant level. And ‡ stands for an improvement of GMM20 against the ML10 which is also significant at the 99% level. All comparisons are based on the test statistics of Diebold and Mariano (1995).

Chapter 7

Conclusion and Outlook

We have studied in this thesis the discrete Markov switching multi-fractal model for modelling the returns of financial assets and for forecasting volatility with some real financial time series. In chapter two, geometric and stochastic fractals are introduced with the related concept like self-similarity, fractal dimension and information dimension that are useful to measure the scaling effect of multi-fractal process. In particular the Hurst exponent is introduced as a measure of long range dependence in stochastic fractals and is estimated over various methods with the real data. Since fractals are a generic form of clustering, we have focused on the long range dependence of financial time series.

Chapter three is devoted to the multi-fractal model of asset returns proposed by Mandelbrot et al. (1997). This model has been shown the way forward to how multi-fractal processes can be applied for financial economics. After describing the construction of multi-fractal process by means of the trinomial multiplicative cascade, we investigate the multi-scaling property of MMAR and estimate the partition function of the financial time series to detect the multi-fractality of them.

In chapter four, we have proposed the trinomial MSMF and later the quadronomial MSMF as an extension of the binomial MSMF. The properties of these MSMF are discussed in detail and the necessary moments of the MSMFs are derived for the GMM estimation. Specially, the theoretical understanding of long-range dependence in this model class of discrete MSMF is presented in the analytical form of autocorrelation functions that show how the range of long memory depends on the multi-fractal parameters and the number of multipliers. The properties of the autocorrelation function in case of the binomial MSMF model are similar to that of the trinomial and the quadronomial MSMF model. We find that the estimation of multi-fractal parameters is closely related to the empirical and theoretical autocorrelation function. Hence, the fine tuning of such autocorrelations is possible when applying the trinomial or quadronomial MSMF model. The ML procedure for the estimation of Markov switching models is introduced and applied in a modified form for the trinomial and the quadronomial MSMF. We have assessed the performance of these two estimations, GMM and the ML, with respect to the small sample property through Monte Carlo simulation. We discuss the problem of model selection based on Voung (1989), Clarke (2003, 2005), and the Markov switching criterion (2006). Using real financial time series, we use these tests to obtain the optimal number of multipliers. The results of these criteria are not uniform and differed by a small amount of frequencies in case of the binomial and the trinomial MSMF model.

On the other side, we allow the MSMF to have Student-t innovations instead of Gaussian. This extension has shown that the same techniques like GMM and ML can be used to estimate the parameters of the t-MSMF model. Chapter five includes both a theoretical and empirical investigation with the t-MSMF. We derive the log moments of Student-t distribution and implement

it for the GMM estimation. With increasing number of multipliers, there is some trade-off between the estimated values of the multi-fractal parameters and the estimated degree of freedom. But we find that the normal MSMF is a better alternative than the t-MSMF for some data.

In chapter 6, we compare the performances of forecasting volatility of various models. The Levinson-Durbin algorithm is shortly introduced and is applied to forecast volatility by means of the autocorrelation of underlying MSMF with the estimated values of multi-fractal parameters using GMM. We also compare the results of linear forecasting with those of optimal forecasting using ML.

GARCH and FIGARCH are used as the benchmark models. The MSPEs and MSPEs in the results of forecasting volatility are examined and compared with the performance of some binomial and trinomial MSMF models. Our analysis shows that the trinomial MSMF model could give acceptable results with the relatively small number of multipliers $\bar{k} = 3, 4, 5$ whereas the corresponding results can be achieved only with the relatively large number of multipliers in case of the binomial one.

In the estimation of GMM and ML, we find that there is also a trade-off between the multi-fractal parameters and the number of multipliers in the binomial and the trinomial models. Though a particular model by the binomial MSMF model ($\bar{k} = 10$) is superior in terms of forecasting performance measured relative MSE, we find that the forecasts from other models like the trinomial MSMF model ($\bar{k} = 3$) is superior in terms of the MAPE. It implies that two models (binomial and trinomial MSMF) should be evaluated simultaneously to get optimal results. It is worth to note that the lengths of the autocorrelations mainly depend on the number of multipliers even though the maximal difference between the multi-fractal parameters has some influence on the order of magnitude in the autocorrelation. We point out that there are certainly computational limitations to obtain the results of estimation if we would use the trinomial (quadronomial) model with larger than $\bar{k} = 7(4)$

Future work

Since the discrete MSMF model provides an economically appealing and computationally tractable alternative to previous GARCH and stochastic volatility models, this approach can be developed in different applications. For example, one could apply it to identify business fluctuations and to forecast the real interest rate (cf. Kim and Nelson (1999)). The discrete MSMF model may be used to examine whether it can be applied successfully to option pricing. With drift or without drift, and with displacement or without displacement, this can be easily modified with only small numbers of parameters.

Furthermore, the scope of the MSMF model can be extended to the multivariate MSMF model. In the recent literature, a bivariate MSMF model has been considered to apply value at risk (VaR) [Liu and Lux (2005)]. Calvet, Fisher, and Thompson (2006) develop a bivariate MSMF model which models bivariate shocks with heterogeneous durations in stochastic volatility and covariation in financial prices. They show that the multi-frequency approach performs well in- and out-of-sample relative to a standard benchmark.

They use a new inference methodology, the so-called particle filter that enables us to estimate a discrete MSMF model with very large state space, i.e. with large number of frequencies. The particle filter permits to estimate a continuous multi-fractal model like the lognormal or gamma MSMF model. Alternatively we can try to estimate the parameters of MSMF model with the method of Markov chain Monte Carlo (MCMC) using efficient important sample [Liesefeld and Richard (2006)].

Of particular interest would be an investigation of the dynamics of volatility using the gamma distribution. As a non-negative random variable the gamma distributtion is attractive to model

dynamic volatility due to its additivity. Indeed, Abraham, Balakrishna, and Sivakumar (2006) recently propose the gamma stochastic volatility model in which return volatilities evolve according to stationary gamma autoregressive specification. In the analysis with daily stock index return data they find that the gamma autoregressive stochastic volatility processes capture the leptokurtic nature of return distributions and the slowly decaying autocorrelation functions of squared returns. Futhermore they show that this stochastic volatility model with gamma autoregressive process has a superior forecasting performance compared to the GARCH and EGARCH models.

Appendix A

Moments of trinomial MSMF

We derive the selected moments of the trinomial MSMF. Let us consider the following volatility process M_t that is described as

$$M_t = \prod_{k=1}^{\bar{k}} m_{t,k}, \tag{A.1}$$

where a volatility component $m_{t,k}$ is drawn from the trinomial multipliers m_0, m_1 and m_2. To fulfil $E[M_t] = 1$ we will constrain $m_0 + m_1 + m_2 = 3$. From time t to $t+1$ a component $m_{t,k}$ changes with the specified probability γ_k of

$$\gamma_k = 1 - \left(1 - \frac{1}{3}\right)^{3^{k-\bar{k}}} \tag{A.2}$$

A.1 Moments of the multi-fractal process

The independency of the volatility components $m_{t,k}$ leads to the first $E\big[M_t\big]$ and the second moments $E\big[M_t^2\big]$ as follows:

$$\begin{aligned} E\big[M_t\big] &= E\big[m_{t,1}\, m_{t,2}\, m_{t,3} \dots\, m_{t,\bar{k}}\big] \\ &= E\big[m_{t,1}\big] E\big[m_{t,2}\big] E\big[m_{t,3}\big] \dots E\big[m_{t,\bar{k}}\big] \\ &= \left(\frac{1}{3}\, m_0 + \frac{1}{3}\, m_1 + \frac{1}{3}\, m_2\right)^{\bar{k}} = 1 \end{aligned} \tag{A.3}$$

and

$$\begin{aligned} E\big[M_t^2\big] &= E\Big[\prod_{k=1}^{\bar{k}} m_{t,k} \prod_{k'=1}^{\bar{k}} m_{t,k'}\Big] \\ &= E\Big[\Big(m_{t,1}\, m_{t,2}\, m_{t,3} \dots\, m_{t,\bar{k}}\Big)\Big(m_{t,1}\, m_{t,2}\, m_{t,3} \dots\, m_{t,\bar{k}}\Big)\Big] \\ &= E\Big[m_{t,1}^2\Big] E\Big[m_{t,2}^2\Big] E\Big[m_{t,3}^2\Big] \dots E\Big[m_{t,\bar{k}}^2\Big] \\ &= \left(\frac{1}{3}\, m_0^2 + \frac{1}{3}\, m_1^2 + \frac{1}{3}\, m_2^2\right)^{\bar{k}} \end{aligned} \tag{A.4}$$

The variance of the process results then in

$$\begin{aligned} Var[M_t] &= E[M_t^2] - \Big(E[M_t]\Big)^2 \\ &= \left(\frac{1}{3}\,m_0^2 + \frac{1}{3}\,m_1^2 + \frac{1}{3}\,m_2^2\right)^{\bar{k}} - 1 \end{aligned} \tag{A.5}$$

Covariance of the volatility process

Let us consider the covariance $E\big[M_{t+T}M_t\big]$. For one time step of $T = 1$, we have got the covariance $E\big[M_{t+1}M_t\big]$:

$$\begin{aligned} &E\big[M_{t+1}M_t\big] \\ &\quad= \prod_{k=1}^{\bar{k}} \Bigg[\frac{1}{3}m_0\Big\{(1-\gamma_k)m_0 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\} \\ &\qquad\quad + \frac{1}{3}m_1\Big\{(1-\gamma_k)m_1 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\} \\ &\qquad\quad + \frac{1}{3}m_2\Big\{(1-\gamma_k)m_2 + \gamma_k\Big(\frac{1}{3}m_0 + \frac{1}{3}m_1 + \frac{1}{3}m_2\Big)\Big\}\Bigg] \\ &\quad= \prod_{k=1}^{\bar{k}} \Bigg[\frac{2}{3}\Big(1-(1-\gamma_k)\Big)\Big(\frac{1}{3}m_0m_1 + \frac{1}{3}m_1m_2 + \frac{1}{3}m_2m_0\Big) \\ &\qquad\quad + \Big\{(1-\gamma_k) + \frac{1}{3}\Big(1-(1-\gamma_k)\Big)\Big\}\Big(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\Big)\Bigg]. \end{aligned} \tag{A.6}$$

In case of $T > 1$, we must replace the probability for renewal after one time step by the probability for T steps. The covariance $E\big[M_{t+T}M_t\big]$ is then written by

$$\begin{aligned} &E\big[M_{t+T}M_t\big] \\ &\quad= \prod_{i=1}^{\bar{k}} \Bigg[\frac{2}{3}\Big(1-\Big(1-\gamma_k\Big)^T\Big)\Big(\frac{1}{3}m_0m_1 + \frac{1}{3}m_1m_2 + \frac{1}{3}m_2m_0\Big) \\ &\qquad\quad + \Big\{\Big(1-\gamma_k\Big)^T + \frac{1}{3}\Big(1-\Big(1-\gamma_k\Big)^T\Big)\Big\}\Big(\frac{1}{3}m_0^2 + \frac{1}{3}m_1^2 + \frac{1}{3}m_2^2\Big)\Bigg] \end{aligned} \tag{A.7}$$

A.2 Moments of log-transformed volatility process

We define the log increment of volatility process[1] $\eta_{t,T}$ as

$$\begin{aligned} \eta_{t,T} &\equiv \ln M_t - \ln M_{t-T} \\ &= \sum_{k=1}^{\bar{k}} \ln m_{t,k} - \sum_{k=1}^{\bar{k}} \ln m_{t-T,k} \\ &= \sum_{k=1}^{\bar{k}} \varepsilon_{t,k} - \sum_{k=1}^{\bar{k}} \varepsilon_{t-T,k} \end{aligned} \tag{A.8}$$

[1] Priestley (1981) shows that the autocorrelation function of the derivative process is only a small scaled autocorrelation function of its original process. In particular the temporal dependence of the derivative process is hugely reduced for a stationary process. Hence, it would be suggested to analyse the autocorrelation function of the derivative (differenced) process for the multi-fractal process which is characterised by long range dependence. This is why we use the (log-transformed) differenced Markov switching multi-fractal process.

Here, $\varepsilon_{t,k} = \ln m_{t,k}$. It is obvious that the first moment of log increments $E[\eta_{t,T}] = 0$ for all T.

A.2.1 The autocovariance of log increments

The autocovariance $E[\eta_{t+T,T}\eta_{t,T}]$ of the log increments of multipliers is given at $T = 1$ as follows:

$$\begin{aligned} & E[\eta_{t+1,1}\eta_{t,1}] \\ & = E\left[\left\{\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k} - \varepsilon_{t,k}\right)\right\}\left\{\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k} - \varepsilon_{t-1,k}\right)\right\}\right] \\ & = E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}\right) - \sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t-1,k}\right) - \sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^2\right) + \sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}\varepsilon_{t-1,k}\right)\right] \end{aligned} \tag{A.9}$$

For example, we can calculate the expectation value of the first term in (A.9) which is the same as that of the forth:

$$
\begin{aligned}
& E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t,k}\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\ln m_{t+1,k}\ln m_{t,k}\Big)\right] \\
&= \sum_{k=1}^{\bar{k}}\left[\frac{1}{3}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}+\frac{1}{3}\ln m_{t,1}+\frac{1}{3}\ln m_{t,2}\big)\Big\}\right. \\
&\qquad +\frac{1}{3}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}+\frac{1}{3}\ln m_{t,1}+\frac{1}{3}\ln m_{t,2}\big)\Big\} \\
&\qquad \left.+\frac{1}{3}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}+\frac{1}{3}\ln m_{t,1}+\frac{1}{3}\ln m_{t,2}\big)\Big\}\right] \\
&= \sum_{k=1}^{\bar{k}}\Big[\frac{1}{9}\Big\{(3-2\gamma_k)(\ln m_0^2+\ln m_1^2+\ln m_2^2) \\
&\qquad\qquad +2\gamma_k(\ln m_0\ln m_1+\ln m_1\ln m_2+\ln m_2\ln m_0)\Big\}\Big] \qquad \text{(A.10)}
\end{aligned}
$$

In (A.10), we can omit the index for time, t because the multiplication of these two components, $\ln m_{t+1,0}$ and $\ln m_{t,0}$ can be simplified by $\ln m_0^2$. We also get the expectation value of the second and the third in (A.9):

$$
\begin{aligned}
&E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t-1,k}\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\ln m_{t+1,k}\ln m_{t-1,k}\Big)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t+1,0}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad +\frac{1}{3}\ln m_{t+1,1}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad +\frac{1}{3}\ln m_{t+1,2}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{9}\Big\{(\gamma_k-2)(-2\ln m_0\ln m_1-2\ln m_1\ln m_2-2\ln m_2\ln m_0) \\
&\qquad +(3-4\gamma_k+2\gamma_k^2)(\ln m_0^2+\ln m_1^2+\ln m_2^2)\Big\}\Bigg] \qquad \text{(A.11)}
\end{aligned}
$$

and

$$
E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t,k}^2\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\ln m_{t,k}^2\Big)\right] = \frac{1}{3}\big(\ln m_{t,0}^2+\ln m_{t,1}^2+\ln m_{t,2}^2\big). \qquad \text{(A.12)}
$$

Therefore, the equation (A.9) leads to

$$
\begin{aligned}
E[\eta_{t+1,1}\eta_{t,1}] = 2\Big(&\ln m_0\ln m_1+\ln m_1\ln m_2+\ln m_2\ln m_0-\ln^2 m_0-\ln^2 m_1-\ln^2 m_2\Big) \\
&\sum_{k=1}^{\bar{k}}\left(\frac{1}{3}\gamma_k\right)^2. \qquad \text{(A.13)}
\end{aligned}
$$

For time intervals $T > 1$ the autocovariance is thus

$$E[\eta_{t+T,T}\,\eta_{t,T}] = 2\Big(\ln m_0 \ln m_1 + \ln m_1 \ln m_2 + \ln m_2 \ln m_0 - \ln^2 m_0 - \ln^2 m_1 - \ln^2 m_2\Big) \sum_{k=1}^{\bar{k}} \left\{ \frac{1}{3}\Big(1 - (1-\gamma_k)^T\Big) \right\}^2 \tag{A.14}$$

Analogous to $E[\eta_{t+1,1}\eta_{t,1}]$ the expected value of squared log increments $E[\eta^2_{t+1,1}]$ is reduced to

$$\begin{aligned} E[\eta^2_{t+1,1}] &= E\left[\left\{ \sum_{k=1}^{\bar{k}} \Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big)^2 \right\} \right] \\ &= E\left[\sum_{k=1}^{\bar{k}} \Big(2\varepsilon^2_{t+1,k} - 2\varepsilon_{t+1,k}\varepsilon_{t,k}\Big) \right] \end{aligned} \tag{A.15}$$

The first term of (A.15) is to two times of (A.12) as the second one is to two times of (A.10). Therefore, we obtain

$$\begin{aligned} E[\eta^2_{t+1,1}] \;=\; & \frac{4}{3}\Big(\ln^2 m_0 + \ln^2 m_1 + \ln^2 m_2 - \ln m_0 \ln m_1 - \ln m_1 \ln m_2 - \ln m_2 \ln m_0\Big) \\ & \sum_{k=1}^{\bar{k}} \frac{1}{3}\gamma_k \end{aligned} \tag{A.16}$$

For an arbitrary time step $T > 1$ the equation (A.16) becomes

$$\begin{aligned} E[\eta^2_{t+T,T}] = & \frac{4}{3}\Big(\ln^2 m_0 + \ln^2 m_1 + \ln^2 m_2 - \ln m_0 \ln m_1 - \ln m_1 \ln m_2 - \ln m_2 \ln m_0\Big) \\ & \sum_{k=1}^{\bar{k}} \left\{ \frac{1}{3}\Big(1 - (1-\gamma_k)^T\Big) \right\} \end{aligned} \tag{A.17}$$

A.2.2 The autocovariance of squared log increments

Due to the independence of volatility components k and k', the autocovariance $E[\eta^2_{t+T,T}\eta^2_{t,T}]$ of squared log increments is given at $T = 1$ by

$$\begin{aligned} E[\eta^2_{t+1,1}\eta^2_{t,1}] &= E\left[\left\{ \sum_{k=1}^{\bar{k}} \Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big) \right\}^2 \left\{ \sum_{k'=1}^{\bar{k}} \Big(\varepsilon_{t,k'} - \varepsilon_{t-1,k'}\Big) \right\}^2 \right] \\ &= E\left[\sum_{k=k'=1}^{\bar{k}} \left\{ \Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big)^2 \Big(\varepsilon_{t,k} - \varepsilon_{t-1,k}\Big)^2 \right\} \right] \\ &\quad + E\left[\sum_{k=1}^{\bar{k}} \left\{ \sum_{k'=1,k'\neq k}^{\bar{k}} \Big(\varepsilon_{t+1,k'} - \varepsilon_{t,k'}\Big)^2 \right\} \Big(\varepsilon_{t,k} - \varepsilon_{t-1,k}\Big)^2 \right] \\ &\quad + 2E\left[\sum_{k=1}^{\bar{k}} \left\{ \sum_{k'=1,k'\neq k}^{\bar{k}} \Big(\varepsilon_{t+1,k'} - \varepsilon_{t,k'}\Big) \right\} \Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big) \right. \\ &\qquad \left. \sum_{k'=1}^{\bar{k}} \left\{ \sum_{k=1,k\neq k'}^{\bar{k}} \Big(\varepsilon_{t,k} - \varepsilon_{t-1,k}\Big) \right\} \Big(\varepsilon_{t,k'} - \varepsilon_{t-1,k'}\Big) \right] \end{aligned} \tag{A.18}$$

Firstly, let us consider the term of $(\varepsilon_{t+1,k}-\varepsilon_{t,k})^2(\varepsilon_{t,k}-\varepsilon_{t-1,k})^2$. The condition for non-zero entries is the same as equation (A.16). We calculate the first term of (A.18), and get

$$
\begin{aligned}
&E\left[\sum_{k=1}^{\bar{k}}\big(\varepsilon_{t+1,k}-\varepsilon_{t,k}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\right]\\
&=E\left[\sum_{k=1}^{\bar{k}}\big(\varepsilon_{t+1,k}^2-2\varepsilon_{t+1,k}\varepsilon_{t,k}+\varepsilon_{t,k}^2\big)\big(\varepsilon_{t,k}^2-2\varepsilon_{t,k}\varepsilon_{t-1,k}+\varepsilon_{t-1,k}^2\big)\right]\\
&=E\Bigg[\sum_{k=1}^{\bar{k}}\big(\underbrace{\varepsilon_{t+1,k}^2\varepsilon_{t,k}^2}_{A}-2\underbrace{\varepsilon_{t+1,k}^2\varepsilon_{t,k}\varepsilon_{t-1,k}}_{B}+\underbrace{\varepsilon_{t+1,k}^2\varepsilon_{t-1,k}^2}_{C}-2\underbrace{\varepsilon_{t+1,k}\varepsilon_{t,k}^3}_{D}+4\underbrace{\varepsilon_{t+1,k}\varepsilon_{t,k}^2\varepsilon_{t-1,k}}_{E}\\
&\qquad-2\underbrace{\varepsilon_{t+1,k}\varepsilon_{t,k}\varepsilon_{t-1,k}^2}_{F}+\underbrace{\varepsilon_{t,k}^4}_{G}-2\underbrace{\varepsilon_{t,k}^3\varepsilon_{t-1,k}}_{H}+\underbrace{\varepsilon_{t,k}^2\varepsilon_{t-1,k}^2}_{I}\big)\Bigg]
\end{aligned}
\tag{A.19}
$$

Let us take the expected value of each term using changing probability γ_k and staying probability $1-\gamma_k$ at each level of cascade. Using equation (A.10), (A.11) and (A.12) we calculate from the underbraced A to I step by step. The underbraced A to I are given als followed:

$$
\begin{aligned}
A&=E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}^2\varepsilon_{t,k}^2\Big)\right]\\
&=\sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t+1,0}^2\Big\{(1-\gamma_k)\ln m_{t,0}^2+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^2+\frac{1}{3}\ln m_{t,1}^2+\frac{1}{3}\ln m_{t,2}^2\big)\Big\}\\
&\qquad+\frac{1}{3}\ln m_{t+1,1}^2\Big\{(1-\gamma_k)\ln m_{t,1}^2+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^2+\frac{1}{3}\ln m_{t,1}^2+\frac{1}{3}\ln m_{t,2}^2\big)\Big\}\\
&\qquad+\frac{1}{3}\ln m_{t+1,2}^2\Big\{(1-\gamma_k)\ln m_{t,2}^2+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^2+\frac{1}{3}\ln m_{t,1}^2+\frac{1}{3}\ln m_{t,2}^2\big)\Big\}\Bigg]\\
&=\sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{9}\Big\{(\ln m_0^4+\ln m_1^4+\ln m_2^4)(3-2\gamma_k)\\
&\qquad+2\ln m_0^2\ln m_1^2\gamma_k+2\ln m_1^2\ln m_2^2\gamma_k+2\ln m_2^2\ln m_0^2\gamma_k\Big\}\Bigg]
\end{aligned}
\tag{A.20}
$$

$$
\begin{aligned}
B &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}^2\varepsilon_{t,k}\varepsilon_{t-1,k}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t+1,0}^2\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad +\frac{1}{3}\ln m_{t+1,1}^2\Big\{(1-\gamma_k)\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +(1-\gamma_k)\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad +\frac{1}{3}\ln m_{t+1,2}^2\Big\{(1-\gamma_k)\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +(1-\gamma_k)\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{k=1}^{\bar{k}} \Bigg[\frac{1}{27} \Big\{ (3-2\gamma_k)(\ln m_0^4 + \ln m_1^4 + \ln m_2^4) \\
&\quad - \ln m_0 (\ln m_1 + \ln m_2)(\ln m_1^2(\gamma_k - 3) \\
&\quad + \ln m_2^2(\gamma_k - 3) - 3 \ln m_1 \ln m_2(\gamma_k - 1))\gamma_k \\
&\quad + 2 \ln m_1^2 \ln m_2^2 (3 - 2\gamma_k)\gamma_k - \ln m_1^3 \ln m_2 (\gamma_k - 3)\gamma_k \\
&\quad - \ln m_1 \ln m_2^3 (\gamma_k - 3)\gamma_k - \ln m_0^3 (\ln m_1 + \ln m_2)(\gamma_k - 3)\gamma_k \\
&\quad + 2 \ln m_0^2 \gamma_k (\ln m_1^2 (3 - 2\gamma_k) + \ln m_2^2 (3 - 2\gamma_k) + \ln m_1 \ln m_2 \gamma_k) \Big\} \Bigg] \qquad \text{(A.21)}
\end{aligned}
$$

$$
\begin{aligned}
C &= E\left[\sum_{k=1}^{\bar{k}} \left(\varepsilon_{t+1,k}^2 \varepsilon_{t-1,k}^2 \right) \right] \\
&= E\Bigg[\sum_{k=1}^{\bar{k}} \Big(\frac{1}{3} \ln m_{t+1,0}^2 \Big\{ (1-\gamma_k)(1-\gamma_k) \ln m_{t-1,0}^2 \\
&\qquad + (1-\gamma_k)\gamma_k \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k (1-\gamma_k) \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k^2 \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \Big\} \\
&\qquad + \frac{1}{3} \ln m_{t+1,1}^2 \Big\{ (1-\gamma_k)(1-\gamma_k) \ln m_{t-1,1}^2 \\
&\qquad + (1-\gamma_k)\gamma_k \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k (1-\gamma_k) \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k^2 \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \Big\} \\
&\qquad + \frac{1}{3} \ln m_{t+1,2}^2 \Big\{ (1-\gamma_k)(1-\gamma_k) \ln m_{t-1,2}^2 \\
&\qquad + (1-\gamma_k)\gamma_k \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k (1-\gamma_k) \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \\
&\qquad + \gamma_k^2 \big(\frac{1}{3} \ln m_{t-1,0}^2 + \frac{1}{3} \ln m_{t-1,1}^2 + \frac{1}{3} \ln m_{t-1,2}^2 \big) \Big\} \Bigg] \\
&= \sum_{k=1}^{\bar{k}} \Bigg[\frac{1}{9} \Big\{ (\ln m_0^4 + \ln m_1^4 + \ln m_2^4)(3 - 4\gamma_k + 2\gamma_k^2) \\
&\qquad - (2 \ln m_0^2 \ln m_1^2 + 2 \ln m_1^2 \ln m_2^2 + 2 \ln m_2^2 \ln m_0^2)(\gamma_k - 2)\gamma_k \Big\} \Bigg] \qquad \text{(A.22)}
\end{aligned}
$$

$$
\begin{aligned}
D &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}^{3}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\left[\frac{1}{3}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}^{3}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^{3}+\frac{1}{3}\ln m_{t,1}^{3}+\frac{1}{3}\ln m_{t,2}^{3}\big)\Big\}\right. \\
&\qquad +\frac{1}{3}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}^{3}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^{3}+\frac{1}{3}\ln m_{t,1}^{3}+\frac{1}{3}\ln m_{t,2}^{3}\big)\Big\} \\
&\qquad \left.+\frac{1}{3}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}^{3}+\gamma_k\big(\frac{1}{3}\ln m_{t,0}^{3}+\frac{1}{3}\ln m_{t,1}^{3}+\frac{1}{3}\ln m_{t,2}^{3}\big)\Big\}\right] \\
&= \sum_{k=1}^{\bar{k}}\left[\frac{1}{9}\Big\{(\ln m_0^4+\ln m_1^4+\ln m_2^4)(3-2\gamma_k)+\ln m_1^3\ln m_2\gamma_k+\ln m_1\ln m_2^3\gamma_k\right. \\
&\quad \left.+\ln m_1\ln m_0^3\gamma_k+\ln m_2\ln m_0^3\gamma_k+\ln m_0\ln m_1^3\gamma_k+\ln m_0\ln m_2^3\gamma_k\Big\}\right] \qquad \text{(A.23)}
\end{aligned}
$$

$$
\begin{aligned}
E &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}^{2}\varepsilon_{t-1,k}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\left[\frac{1}{3}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}^{2}(1-\gamma_k)\ln m_{t-1,0}\right. \\
&\qquad +(1-\gamma_k)\ln m_{t,0}^{2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}^{2}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}^{2}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}^{2}(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}^{2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}^{2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad \left.+\gamma_k\frac{1}{3}\ln m_{t,2}^{2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\}\right. \\
&\quad +\frac{1}{3}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}^{2}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +(1-\gamma_k)\ln m_{t,1}^{2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big) \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,0}^{2}(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,1}^{2}(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\frac{1}{3}\ln m_{t,2}^{2}(1-\gamma_k)\ln m_{t-1,2}
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_k\frac{1}{3}\ln m_{t,0}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\}\\
+\frac{1}{3}\ln m_{t+1,2}\Big\{&(1-\gamma_k)\ln m_{t,2}^2(1-\gamma_k)\ln m_{t-1,2}\\
&+(1-\gamma_k)\ln m_{t,2}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}^2(1-\gamma_k)\ln m_{t-1,0}\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}^2(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}^2(1-\gamma_k)\ln m_{t-1,2}\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}^2\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}+\frac{1}{3}\ln m_{t-1,1}+\frac{1}{3}\ln m_{t-1,2}\big)\Big\}\\
=\sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{27}\Big\{&(3-2\gamma_k)(\ln m_0^4+\ln m_1^4+\ln m_2^4)\\
&-\ln m_0(\ln m_1+\ln m_2)(\ln m_1^2(\gamma_k-3)\\
&+\ln m_2^2(\gamma_k-3)-3\ln m_1\ln m_2(\gamma_k-1))\gamma_k\\
&+2\ln m_1^2\ln m_2^2(3-2\gamma_k)\gamma_k-\ln m_1^3\ln m_2(\gamma_k-3)\gamma_k\\
&-\ln m_1\ln m_2^3(\gamma_k-3)\gamma_k-\ln m_0^3(\ln m_1+\ln m_2)(\gamma_k-3)\gamma_k\\
&+2\ln m_0^2\gamma_k(\ln m_1^2(3-2\gamma_k)+\ln m_2^2(3-2\gamma_k)+\ln m_1\ln m_2\gamma_k)\Big\}\Bigg]
\end{aligned}
\tag{A.24}
$$

$$
\begin{aligned}
F&=E\Bigg[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t,k}\varepsilon_{t-1,k}^2\Big)\Bigg]\\
&=\sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^2\\
&\qquad+(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&\qquad+\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^2\\
&\qquad+\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^2\\
&\qquad+\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^2
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\Big\}\\
&+\frac{1}{3}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^2\\
&+(1-\gamma_k)\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\Big\}\\
&+\frac{1}{3}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^2\\
&+(1-\gamma_k)\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^2\\
&+\gamma_k\frac{1}{3}\ln m_{t,0}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,1}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^2+\frac{1}{3}\ln m_{t-1,1}^2+\frac{1}{3}\ln m_{t-1,2}^2\big)\Big\}\\
=\sum_{k=1}^{\bar{k}}\Bigg[&\frac{1}{27}\Big\{(3-2\gamma_k)(\ln m_0^4+\ln m_1^4+\ln m_2^4)\\
&-\ln m_0(\ln m_1+\ln m_2)(\ln m_1^2(\gamma_k-3)\\
&+\ln m_2^2(\gamma_k-3)-3\ln m_1\ln m_2(\gamma_k-1))\gamma_k\\
&+2\ln m_1^2\ln m_2^2(3-2\gamma_k)\gamma_k-\ln m_1^3\ln m_2(\gamma_k-3)\gamma_k\\
&-\ln m_1\ln m_2^3(\gamma_k-3)\gamma_k-\ln m_0^3(\ln m_1+\ln m_2)(\gamma_k-3)\gamma_k\\
&+2\ln m_0^2\gamma_k(\ln m_1^2(3-2\gamma_k)+\ln m_2^2(3-2\gamma_k)+\ln m_1\ln m_2\gamma_k)\Big\}\Bigg]
\end{aligned}
\tag{A.25}
$$

$$G = E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^{4}\right)\right] = \frac{1}{3}\Big\{\ln m_{t,0}^{4} + \ln m_{t,1}^{4} + \ln m_{t,2}^{4}\Big\} \tag{A.26}$$

$$\begin{aligned}
H &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^{3}\varepsilon_{t-1,k}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t,0}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,0} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0} + \frac{1}{3}\ln m_{t-1,1} + \frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad + \frac{1}{3}\ln m_{t,1}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,1} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0} + \frac{1}{3}\ln m_{t-1,1} + \frac{1}{3}\ln m_{t-1,2}\big)\Big\} \\
&\quad + \frac{1}{3}\ln m_{t,2}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,2} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0} + \frac{1}{3}\ln m_{t-1,1} + \frac{1}{3}\ln m_{t-1,2}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{9}\Big\{(\ln m_0^4 + \ln m_1^4 + \ln m_2^4)(3-2\gamma_k) \\
&\quad + \ln m_0^3 \ln m_1\gamma_k + \ln m_0^3 \ln m_2\gamma_k + \ln m_1^3 \ln m_0\gamma_k \\
&\quad + \ln m_1^3 \ln m_2\gamma_k + \ln m_2^3 \ln m_0\gamma_k + \ln m_2^3 \ln m_1\gamma_k\Big\}\Bigg]
\end{aligned} \tag{A.27}$$

$$\begin{aligned}
I &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^{2}\varepsilon_{t-1,k}^{2}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{3}\ln m_{t,0}^{2}\Big\{(1-\gamma_k)\ln m_{t-1,0}^{2} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^{2} + \frac{1}{3}\ln m_{t-1,1}^{2} + \frac{1}{3}\ln m_{t-1,2}^{2}\big)\Big\} \\
&\quad + \frac{1}{3}\ln m_{t,1}^{2}\Big\{(1-\gamma_k)\ln m_{t-1,1}^{2} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^{2} + \frac{1}{3}\ln m_{t-1,1}^{2} + \frac{1}{3}\ln m_{t-1,2}^{2}\big)\Big\} \\
&\quad + \frac{1}{3}\ln m_{t,2}^{2}\Big\{(1-\gamma_k)\ln m_{t-1,2}^{2} + \gamma_k\big(\frac{1}{3}\ln m_{t-1,0}^{2} + \frac{1}{3}\ln m_{t-1,1}^{2} + \frac{1}{3}\ln m_{t-1,2}^{2}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{9}\Big\{(\ln m_0^4 + \ln m_1^4 + \ln m_2^4)(3-2\gamma_k) \\
&\quad + 2\ln m_0^2 \ln m_1^2\gamma_k + 2\ln m_1^2 \ln m_2^2\gamma_k + 2\ln m_2^2 \ln m_0^2\gamma_k\Big\}\Bigg]
\end{aligned} \tag{A.28}$$

Using the above equations from A to I, the first term of (A.18) is given by

$$
\begin{aligned}
&E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}-\varepsilon_{t,k}\right)^2\left(\varepsilon_{t,k}-\varepsilon_{t-1,k}\right)^2\right] \\
&=\sum_{k=1}^{\bar{k}}2\Big(\frac{1}{3}\gamma_k\Big)^2\Big(\ln m_0\ln m_1+\ln m_1\ln m_2+\ln m_2\ln m_0 \\
&\qquad\qquad -\ln^2 m_0-\ln^2 m_1-\ln^2 m_2\Big)^2. \qquad \text{(A.29)}
\end{aligned}
$$

For the second term of $\left(\varepsilon_{t+1,k'}-\varepsilon_{t,k'}\right)^2\left(\varepsilon_{t,k}-\varepsilon_{t-1,k}\right)^2$ there are 36 sequences which meet the non-zero requirement of

$$
\varepsilon_{t+1,k'}\neq\varepsilon_{t,k'} \quad \text{and} \quad \varepsilon_{t,k}\neq\varepsilon_{t-1,k} \;\; (k\neq k') \qquad \text{(A.30)}
$$

That is, the sequences

$$
\begin{array}{llll}
m_0\to m_1 & m_0\to m_1 & m_0\to m_2 & m_0\to m_2 \\
m_1\to m_0 & m_1\to m_0 & m_1\to m_2 & m_1\to m_2 \\
m_2\to m_0 & m_2\to m_0 & m_2\to m_1 & m_2\to m_1 \\
\\
m_0\to m_1 & m_1\to m_0 & m_0\to m_2 & m_2\to m_0 \\
m_1\to m_0 & m_0\to m_1 & m_1\to m_2 & m_2\to m_1 \\
m_2\to m_0 & m_0\to m_2 & m_2\to m_1 & m_1\to m_2
\end{array} \qquad \text{(A.31)}
$$

as well as

$$
\begin{array}{llll}
m_0\to m_1 & m_0\to m_2 & m_0\to m_1 & m_1\to m_2 \\
m_0\to m_1 & m_2\to m_0 & m_0\to m_1 & m_2\to m_1 \\
\\
m_0\to m_2 & m_0\to m_1 & m_0\to m_2 & m_1\to m_2 \\
m_0\to m_2 & m_1\to m_0 & m_0\to m_2 & m_2\to m_1 \\
\\
m_1\to m_0 & m_0\to m_2 & m_1\to m_0 & m_1\to m_2 \\
m_1\to m_0 & m_2\to m_0 & m_1\to m_0 & m_2\to m_1 \\
\\
m_1\to m_2 & m_0\to m_1 & m_1\to m_2 & m_0\to m_2 \\
m_1\to m_2 & m_1\to m_0 & m_1\to m_2 & m_2\to m_0 \\
\\
m_2\to m_0 & m_0\to m_1 & m_2\to m_0 & m_1\to m_2 \\
m_2\to m_0 & m_1\to m_0 & m_2\to m_0 & m_2\to m_1 \\
\\
m_2\to m_1 & m_0\to m_1 & m_2\to m_1 & m_0\to m_2 \\
m_2\to m_1 & m_1\to m_0 & m_2\to m_1 & m_2\to m_0
\end{array} \qquad \text{(A.32)}
$$

comply with the conditions for the non-zero entries in equation (A.30). According to the conditions of (A.31, A.32), the change of multipliers occurs with the probability of $\left(\frac{1}{3}\gamma_k\right)^2$. Substituting

equation (A.32) into the $E\Big[\big(\varepsilon_{t+1,k'}-\varepsilon_{t,k'}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\Big]$ results in

$$\begin{aligned}
&E\Big[\big(\varepsilon_{t+1,k'}-\varepsilon_{t,k'}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\Big] \\
&\quad=\sum_{k,k'=1,k'\neq k}^{\bar{k}}\frac{16}{9}\Big(\frac{1}{3}\gamma_k\frac{1}{3}\gamma_{k'}\Big)\Big(\ln^2 m_0+\ln^2 m_1+\ln^2 m_2 \\
&\quad-\ln m_0\ln m_1-\ln m_1\ln m_2-\ln m_2\ln m_0\Big)^2
\end{aligned} \tag{A.33}$$

Now, we turn to the last term of $(\varepsilon_{t+1,k'}-\varepsilon_{t,k'})(\varepsilon_{t+1,k}-\varepsilon_{t,k})(\varepsilon_{t,k}-\varepsilon_{t-1,k})(\varepsilon_{t,k'}-\varepsilon_{t-1,k'})$. On the condition of

$$\varepsilon_{t+1,k'}\neq\varepsilon_{t,k'}\neq\varepsilon_{t-1,k'}\quad\text{and}\quad\varepsilon_{t+1,k}\neq\varepsilon_{t,k}\neq\varepsilon_{t-1,k}\qquad(k\neq k') \tag{A.34}$$

the last term contributes to $E[\eta^2_{t+1,1}\eta^2_{t,1}]$. With respect to each case of $\varepsilon_{t+1,k}=\ln m_0$, $\varepsilon_{t+1,k}=\ln m_1$ and $\varepsilon_{t+1,k}=\ln m_3$, there are 48 possible combinations of the volatility components. As a result, a total of 144 combinations with $\ln m_0$, $\ln m_1$ and $\ln m_2$ fulfil the condition in equation (A.34), where the change of multipliers occurs with the probability of $\big(\frac{1}{3}\gamma_k\big)^2\big(\frac{1}{3}\gamma_{k'}\big)^2$ Due to these entries we get the following expected value of

$$\begin{aligned}
&E\Big[(\varepsilon_{t+1,k'}-\varepsilon_{t,k'})(\varepsilon_{t+1,k}-\varepsilon_{t,k})(\varepsilon_{t,k}-\varepsilon_{t-1,k})(\varepsilon_{t,k'}-\varepsilon_{t-1,k'})\Big] \\
&=\sum_{k'=1,k'\neq k}^{\bar{k}}4\left(\frac{1}{3}\gamma_k\right)^2\left(\frac{1}{3}\gamma_{k'}\right)^2\Big(\ln m_0\ln m_1+\ln m_1\ln m_2+\ln m_2\ln m_0 \\
&\qquad\qquad-\ln^2 m_0-\ln^2 m_1-\ln^2 m_2\Big)^2
\end{aligned} \tag{A.35}$$

Putting equations (A.19), (A.33) and (A.35) together, we get the autocovariance of squared log increments:

$$\begin{aligned}
E[\eta^2_{t+1,1}\eta^2_{t,1}]&=2\sum_{k=1}^{\bar{k}}\left(\frac{1}{3}\gamma_k\right)^2\xi+\frac{16}{9}\sum_{k=1}^{\bar{k}}\left\{\left(\frac{1}{3}\gamma_k\right)\sum_{k,k'=1,k'\neq k}^{\bar{k}}\left(\frac{1}{3}\gamma_{k'}\right)\right\}\xi \\
&\quad+8\sum_{k=1}^{\bar{k}}\left\{\left(\frac{1}{3}\gamma_k\right)^2\sum_{k'=1,k'\neq k}^{\bar{k}}\left(\frac{1}{3}\gamma_{k'}\right)^2\right\}\xi,
\end{aligned} \tag{A.36}$$

where $\xi=\Big(\ln m_0\ln m_1+\ln m_1\ln m_2+\ln m_2\ln m_0-\ln^2 m_0-\ln^2 m_1-\ln^2 m_2\Big)^2$. For an arbitrary T, equation (A.36) is generalized as followed:

$$\begin{aligned}
&E[\eta^2_{t+1,1}\eta^2_{t,1}] \\
&=2\sum_{k=1}^{\bar{k}}\frac{1}{9}\Big(1-\big(1-\gamma_k\big)^T\Big)^2\xi \\
&\quad+\frac{16}{9}\sum_{k=1}^{\bar{k}}\left\{\frac{1}{3}\Big(1-\big(1-\gamma_k\big)^T\Big)\sum_{k'=1,k'\neq k}^{\bar{k}}\frac{1}{3}\Big(1-\big(1-\gamma_{k'}\big)^T\Big)\right\}\xi \\
&\quad+8\sum_{k=1}^{\bar{k}}\left\{\frac{1}{9}\Big(1-\big(1-\gamma_k\big)^T\Big)^2\sum_{k'=1,k'\neq k}^{\bar{k}}\frac{1}{9}\Big(1-\big(1-\gamma_{k'}\big)^T\Big)^2\right\}\xi
\end{aligned} \tag{A.37}$$

Appendix B

Moments of quadronomial model

The trinomial MSMF has been clearly developed by the request for more flexible stochastic multi-fractal model. Here, we will extend our approach to the quadronomial MSMF that belongs to the more general class of discrete MSMF, and can generate more flexible autocorrelation function.

B.1 The first and second moments

In quadronomial multi-fractal process M_t, the volatility component $m_{t,k}$ is drawn from the quadronomial distribution of $\{m_0, m_1, m_2, m_3\}$:

$$M_t = \prod_{k=1}^{\bar{k}} m_{t,k} \tag{B.1}$$

with $m_{t,k} \in \{m_0, m_1, m_2, m_3\}$. We assume $m_0 + m_1 + m_2 + m_3 = 4$ for the normalization of mean value, that is $E[M_t] = 1$. The $m_{k,t}$ changes from time t to $t+1$ with the probability that is specified as

$$\gamma_k = 1 - \left(1 - \frac{1}{4}\right)^{4^{k-\bar{k}}} \tag{B.2}$$

The first $E\big[M_t\big]$ and second moment $E\big[M_t^2\big]$ of the volatility process are calculated by the same consideration as the one in the trinomial model, see equations (A.3) and (A.4). It is followed that

$$E\big[M_t\big] = \left(\frac{1}{4}\, m_0 + \frac{1}{4}\, m_1 + \frac{1}{4}\, m_2 + \frac{1}{4}\, m_3\right)^{\bar{k}} = 1 \tag{B.3}$$

and

$$E\big[M_t^2\big] = \left(\frac{1}{4}\, m_0^2 + \frac{1}{4}\, m_1^2 + \frac{1}{4}\, m_2^2 + \frac{1}{4}\, m_3^2\right)^{\bar{k}}, \tag{B.4}$$

respectively. Furthermore, we have got the first covariance for one time step $T = 1$ as follows:

$$
\begin{aligned}
&E\big[M_{t+1}M_t\big] \\
&= \prod_{k=1}^{\bar{k}} \Big[\frac{1}{4}m_0\Big\{(1-\gamma_k)m_0 + \gamma_k\Big(\frac{1}{4}m_0 + \frac{1}{4}m_1 + \frac{1}{4}m_2 + \frac{1}{4}m_3\Big)\Big\} \\
&\qquad + \frac{1}{4}m_1\Big\{(1-\gamma_k)m_1 + \gamma_k\Big(\frac{1}{4}m_0 + \frac{1}{4}m_1 + \frac{1}{4}m_2 + \frac{1}{4}m_3\Big)\Big\} \\
&\qquad + \frac{1}{4}m_2\Big\{(1-\gamma_k)m_2 + \gamma_k\Big(\frac{1}{4}m_0 + \frac{1}{4}m_1 + \frac{1}{4}m_2 + \frac{1}{4}m_3\Big)\Big\} \\
&\qquad + \frac{1}{4}m_3\Big\{(1-\gamma_k)m_3 + \gamma_k\Big(\frac{1}{4}m_0 + \frac{1}{4}m_1 + \frac{1}{4}m_2 + \frac{1}{4}m_3\Big)\Big\}\Big] \\
&= \prod_{k=1}^{\bar{k}} \Big[\frac{2}{4}\Big(1-(1-\gamma_k)\Big)\Big(\frac{1}{4}m_0m_1 + \frac{1}{4}m_0m_2 + \frac{1}{4}m_0m_3 + \frac{1}{4}m_1m_2 + \frac{1}{4}m_1m_3 + \frac{1}{4}m_2m_3\Big) \\
&\qquad + \Big\{(1-\gamma_k) + \frac{1}{4}\Big(1-(1-\gamma_k)\Big)\Big\}\Big(\frac{1}{4}m_0^2 + \frac{1}{4}m_1^2 + \frac{1}{4}m_2^2 + \frac{1}{4}m_3^2\Big)\Big]
\end{aligned} \tag{B.5}
$$

The equation (B.5) is generalized for any time interval $T > 1$ as

$$
\begin{aligned}
&E\big[M_{t+T}M_t\big] \\
&= \prod_{k=1}^{\bar{k}} \Big[\frac{2}{4}\Big(1-\Big(1-\gamma_k\Big)^T\Big)\Big(\frac{1}{4}m_0m_1 + \frac{1}{4}m_0m_2 + \frac{1}{4}m_0m_3 + \frac{1}{4}m_1m_2 + \frac{1}{4}m_1m_3 + \frac{1}{4}m_2m_3\Big) \\
&\qquad + \Big\{\Big(1-\gamma_k\Big)^T + \frac{1}{4}\Big(1-\Big(1-\gamma_k\Big)^T\Big)\Big\}\Big(\frac{1}{4}m_0^2 + \frac{1}{4}m_1^2 + \frac{1}{4}m_2^2 + \frac{1}{4}m_3^2\Big)\Big]
\end{aligned} \tag{B.6}
$$

B.2 Moment of log-transformed volatility process

The autocovariance of log increments

Let us remember that the autocovariance $E[\eta_{t+1,1}\eta_{t,1}]$ of log increments is given by

$$
\begin{aligned}
&E[\eta_{t+1,1}\eta_{t,1}] \\
&= E\Big[\Big\{\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big)\Big\}\Big\{\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t,k} - \varepsilon_{t-1,k}\Big)\Big\}\Big] \\
&= E\Big[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t,k}\Big) - \sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t-1,k}\Big) - \sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t,k}^2\Big) + \sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t,k}\varepsilon_{t-1,k}\Big)\Big]
\end{aligned} \tag{B.7}
$$

By virtue of the sequences

$$
\begin{array}{lll}
m_0 \to m_1 \to m_0, & m_0 \to m_2 \to m_0, & m_0 \to m_3 \to m_0 \\
m_1 \to m_0 \to m_1, & m_1 \to m_2 \to m_1, & m_1 \to m_3 \to m_1 \\
m_2 \to m_0 \to m_2, & m_2 \to m_1 \to m_2, & m_2 \to m_3 \to m_2 \\
m_3 \to m_0 \to m_3, & m_3 \to m_1 \to m_3, & m_3 \to m_2 \to m_3
\end{array} \tag{B.8}
$$

and

$$
\begin{array}{ll}
m_0 \to m_1 \to m_2, & m_0 \to m_1 \to m_3 \\
m_0 \to m_2 \to m_1, & m_0 \to m_2 \to m_3 \\
m_0 \to m_3 \to m_1, & m_0 \to m_3 \to m_2 \\
 & \\
m_1 \to m_0 \to m_2, & m_1 \to m_0 \to m_3 \\
m_1 \to m_2 \to m_0, & m_1 \to m_2 \to m_3 \\
m_1 \to m_3 \to m_0, & m_1 \to m_3 \to m_2 \\
 & \\
m_2 \to m_0 \to m_1, & m_2 \to m_0 \to m_3 \\
m_2 \to m_1 \to m_0, & m_2 \to m_1 \to m_3 \\
m_2 \to m_0 \to m_1, & m_2 \to m_1 \to m_3 \\
 & \\
m_3 \to m_0 \to m_1, & m_3 \to m_0 \to m_2 \\
m_3 \to m_1 \to m_0, & m_3 \to m_1 \to m_2 \\
m_3 \to m_2 \to m_0, & m_3 \to m_2 \to m_1
\end{array}
\tag{B.9}
$$

that fulfill the condition for non-zero entries. For example, we can calculate the expectation value of the first term in (B.7) which is practically the same as that of the forth:

$$
\begin{aligned}
&E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t,k}\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t,k}\varepsilon_{t-1,k}\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\ln m_{t+1,k}\ln m_{t,k}\Big)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0} \\
&\qquad\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}+\frac{1}{4}\ln m_{t,1}+\frac{1}{4}\ln m_{t,2}+\frac{1}{4}\ln m_{t,3}\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1} \\
&\qquad\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}+\frac{1}{4}\ln m_{t,1}+\frac{1}{4}\ln m_{t,2}+\frac{1}{4}\ln m_{t,3}\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2} \\
&\qquad\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}+\frac{1}{4}\ln m_{t,1}+\frac{1}{4}\ln m_{t,2}+\frac{1}{4}\ln m_{t,3}\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,3}\Big\{(1-\gamma_k)\ln m_{t,3} \\
&\qquad\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}+\frac{1}{4}\ln m_{t,1}+\frac{1}{4}\ln m_{t,2}+\frac{1}{4}\ln m_{t,3}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Big[\frac{1}{16}\Big\{(4\ln m_0^2+4\ln m_1^2+4\ln m_2^2+4\ln m_3^2) \\
&\qquad -3\ln m_0^2\gamma_k-3\ln m_1^2\gamma_k-3\ln m_2^2\gamma_k-3\ln m_3^2\gamma_k \\
&\qquad +2\ln m_0\ln m_1\gamma_k+2\ln m_0\ln m_2\gamma_k+2\ln m_0\ln m_3\gamma_k \\
&\qquad +2\ln m_1\ln m_2\gamma_k+2\ln m_1\ln m_3\gamma_k+2\ln m_2\ln m_3\gamma_k)\Big\}\Big]
\end{aligned}
\tag{B.10}
$$

We also take the expectation value of the second and the third in (B.7):

$$
\begin{aligned}
&E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon_{t+1,k}\varepsilon_{t-1,k}\Big)\right] = E\left[\sum_{k=1}^{\bar{k}}\Big(\ln m_{t+1,k}\ln m_{t-1,k}\Big)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t+1,0}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,2}\big)\Big\}+ \\
&\quad \frac{1}{4}\ln m_{t+1,1}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,2}\big)\Big\}+ \\
&\quad \frac{1}{4}\ln m_{t+1,2}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,2}\big)\Big\}+ \\
&\quad \frac{1}{4}\ln m_{t+1,3}\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,3} \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,2}\big)\Big\} \\
&= \sum_{k=1}^{\bar{k}}\Big[\frac{1}{16}\Big\{(4\ln m_0^2+4\ln m_1^2+4\ln m_2^2+4\ln m_3^2) \\
&\qquad -6\ln m_0^2\gamma_k-6\ln m_1^2\gamma_k-6\ln m_2^2\gamma_k-6\ln m_3^2\gamma_k \\
&\qquad +4\ln m_0\ln m_1\gamma_k+4\ln m_0\ln m_2\gamma_k+4\ln m_0\ln m_3\gamma_k \\
&\qquad +4\ln m_1\ln m_2\gamma_k+4\ln m_1\ln m_3\gamma_k+4\ln m_2\ln m_3\gamma_k \\
&\qquad -2\ln m_0\ln m_1\gamma_k^2-2\ln m_0\ln m_2\gamma_k^2-2\ln m_0\ln m_3\gamma_k^2 \\
&\qquad -2\ln m_1\ln m_2\gamma_k^2-2\ln m_1\ln m_3\gamma_k^2-2\ln m_2\ln m_3\gamma_k^2 \\
&\qquad +4\ln m_0^2\gamma_k^2+4\ln m_1^2\gamma_k^2+4\ln m_2^2\gamma_k^2+4\ln m_3^2\gamma_k^2)\Big\}\Big] \qquad \text{(B.11)}
\end{aligned}
$$

and

$$\begin{aligned} E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^2\right)\right] &= E\left[\sum_{k=1}^{\bar{k}}\left(\ln m_{t,k}^2\right)\right] \\ &= \frac{1}{4}\big(\ln m_{t,0}^2 + \ln m_{t,1}^2 + \ln m_{t,2}^2 + \ln m_{t,3}^2\big). \end{aligned} \tag{B.12}$$

Therefore, the autocovariance of log increments, (B.7) leads to

$$\begin{aligned} E[\eta_{t+1,1}\eta_{t,1}] = -\Big\{&(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 \\ &+(\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 \\ &+(\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\}\sum_{k=1}^{\bar{k}}\left(\frac{1}{4}\gamma_k\right)^2 \end{aligned} \tag{B.13}$$

Accordingly, the moment $E[\eta_{t+T,T}\eta_{t,T}]$ for a time interval $T > 1$ is obtained in the form of

$$\begin{aligned} E[\eta_{t+T,T}\eta_{t,T}] = -\Big\{&(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 \\ &+(\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 \\ &+(\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\}\sum_{k=1}^{\bar{k}}\left\{\frac{1}{4}\Big(1 - \Big(1 - \gamma_k\Big)^T\Big)\right\}^2 \end{aligned} \tag{B.14}$$

The expected value of squared log increments $E[\eta_{t+T,T}^2]$ is represented for one time step $T = 1$ by

$$\begin{aligned} E[\eta_{t+1,1}^2] &= E\left[\sum_{k=k'=1}^{\bar{k}}\Big(\varepsilon_{t+1,k} - \varepsilon_{t,k}\Big)^2\right] \\ &= E\left[\sum_{k=1}^{\bar{k}}\Big(2\varepsilon_{t+1,k}^2 - 2\varepsilon_{t+1,k}\varepsilon_{t,k}\Big)\right]. \end{aligned} \tag{B.15}$$

The first term of (B.15) is just two times of (B.12) and the second one is two times of (B.18). Therefore we obtain the following expectation value of squared log increments:

$$\begin{aligned} E[\eta_{t+1,1}^2] = \frac{1}{2}\Big\{&(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 \\ &+(\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 \\ &+(\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\}\sum_{k=1}^{\bar{k}}\frac{1}{4}\gamma_k \end{aligned} \tag{B.16}$$

For $T > 1$ equation (B.17) is expressed by

$$\begin{aligned} E[\eta^2_{t+T,T}] &= \frac{1}{2}\Big\{(\ln m_0 - \ln m_1)^2 + (\ln m_0 - \ln m_2)^2 \\ &\quad +(\ln m_0 - \ln m_3)^2 + (\ln m_1 - \ln m_2)^2 \\ &\quad +(\ln m_1 - \ln m_3)^2 + (\ln m_2 - \ln m_3)^2\Big\} \\ &\quad \sum_{k=1}^{\bar{k}}\left\{\frac{1}{4}\Big(1 - \Big(1-\gamma_k\Big)^T\Big)\right\} \end{aligned} \tag{B.17}$$

B.2.1 The autocovariance of squared log increments

As to $E[\eta^2_{t+1,1}\eta^2_{t,1}]$ the adapting procedure of the trinomial case is straightforward. See equation (A.18) and (A.19).

Using equation (B.18), (B.11) and (B.12) we calculate from the underbraced A to I step by step. For the quadronomial case the underbraced A to I are given als followed:

$$\begin{aligned} A &= E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon^2_{t+1,k}\varepsilon^2_{t,k}\Big)\right] \\ &= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m^2_{t+1,0}\Big\{(1-\gamma_k)\ln m^2_{t,0} + \gamma_k\big(\frac{1}{4}\ln m^2_{t,0} + \frac{1}{4}\ln m^2_{t,1} + \frac{1}{4}\ln m^2_{t,2} + \frac{1}{4}\ln m^2_{t,3}\big)\Big\} \\ &\quad +\frac{1}{4}\ln m^2_{t+1,1}\Big\{(1-\gamma_k)\ln m^2_{t,1} + \gamma_k\big(\frac{1}{4}\ln m^2_{t,0} + \frac{1}{4}\ln m^2_{t,1} + \frac{1}{4}\ln m^2_{t,2} + \frac{1}{4}\ln m^2_{t,3}\big)\Big\} \\ &\quad +\frac{1}{4}\ln m^2_{t+1,2}\Big\{(1-\gamma_k)\ln m^2_{t,2} + \gamma_k\big(\frac{1}{4}\ln m^2_{t,0} + \frac{1}{4}\ln m^2_{t,1} + \frac{1}{4}\ln m^2_{t,2} + \frac{1}{4}\ln m^2_{t,3}\big)\Big\} \\ &\quad +\frac{1}{4}\ln m^2_{t+1,3}\Big\{(1-\gamma_k)\ln m^2_{t,2} + \gamma_k\big(\frac{1}{4}\ln m^2_{t,0} + \frac{1}{4}\ln m^2_{t,1} + \frac{1}{4}\ln m^2_{t,2} + \frac{1}{4}\ln m^2_{t,3}\big)\Big\}\Bigg] \\ &= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{16}\Big\{(\ln m_0^4 + \ln m_1^4 + \ln m_2^4)(4-3\gamma_k) + 2\ln m_0^2 \ln m_1^2\gamma_k + 2\ln m_0^2\ln m_2^2\gamma_k \\ &\quad +2\ln m_0^2\ln m_3^2\gamma_k + 2\ln m_1^2\ln m_2^2\gamma_k + 2\ln m_1^2\ln m_3^2\gamma_k + 2\ln m_2^2\ln m_3^2\gamma_k\Big\}\Bigg] \end{aligned} \tag{B.18}$$

$$\begin{aligned} B &= E\left[\sum_{k=1}^{\bar{k}}\Big(\varepsilon^2_{t+1,k}\varepsilon_{t,k}\varepsilon_{t-1,k}\Big)\right] \\ &= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m^2_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0} \\ &\quad +(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0} + \frac{1}{4}\ln m_{t-1,1} + \frac{1}{4}\ln m_{t-1,2} + \frac{1}{4}\ln m_{t-1,3}\big) \end{aligned}$$

$$
\begin{aligned}
&+\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,2}\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\\
&+\frac{1}{4}\ln m^2_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}\\
&+(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,2}\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\\
&+\frac{1}{4}\ln m^2_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}\\
&+(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,2}\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{4}\ln m_{t+1,3}^2\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}\\
&+(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,2}\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{3}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
=\sum_{k=1}^{\bar{k}}\Bigg[&\frac{1}{64}\Big\{16\ln m_0^4+16\ln m_1^4+16\ln m_2^4+16\ln m_3^4\\
&-24\ln m_0^4\gamma_k+4\ln m_0^3\ln m_1\gamma_k+8\ln m_0^2\ln m_1^2\gamma_k+4\ln m_0\ln m_1^3\gamma_k\\
&-24\ln m_1^4\gamma_k+4\ln m_0^3\ln m_2\gamma_k+4\ln m_1^3\ln m_2\gamma_k+8\ln m_0^2\ln m_2^2\gamma_k\\
&-24\ln m_2^4\gamma_k+4\ln m_0\ln m_2^3\gamma_k+4\ln m_1\ln m_2^3\gamma_k+8\ln m_1^2\ln m_2^2\gamma_k\\
&+4\ln m_0^3\ln m_3\gamma_k+4\ln m_1^3\ln m_3\gamma_k+4\ln m_2^3\ln m_3\gamma_k+8\ln m_0^2\ln m_3^2\gamma_k\\
&+8\ln m_1^2\ln m_3^2\gamma_k+8\ln m_2^2\ln m_3^2\gamma_k+4\ln m_0\ln m_3^3\gamma_k+4\ln m_1\ln m_3^3\gamma_k\\
&+4\ln m_2\ln m_3^3\gamma_k-24\ln m_3^4\gamma_k+9\ln m_0^4\gamma_k^2+2\ln m_0^3\ln m_1\gamma_k^2\\
&-6\ln m_0^2\ln m_1^2\gamma_k^2-2\ln m_0\ln m_1^3\gamma_k^2+9\ln m_1^4\gamma_k^2-2\ln m_0^3\ln m_2\gamma_k^2\\
&+2\ln m_0^2\ln m_1\ln m_2\gamma_k^2+2\ln m_0\ln m_1^2\ln m_2\gamma_k^2+2\ln m_1^3\ln m_2\gamma_k^2\\
&-6\ln m_0^2\ln m_2^2\gamma_k^2+2\ln m_0\ln m_1\ln m_2^2\gamma_k^2-6\ln m_1^2\ln m_2^2\gamma_k^2\\
&-2\ln m_0\ln m_2^3\gamma_k^2-2\ln m_1\ln m_2^3\gamma_k^2+9\ln m_2^4\gamma_k^2\\
&-2\ln m_0^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_1\ln m_3\gamma_k^2+2\ln m_0\ln m_1^2\ln m_3\gamma_k^2\\
&-2\ln m_1^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_2\ln m_3\gamma_k^2+2\ln m_1^2\ln m_2\ln m_3\gamma_k^2\\
&+2\ln m_0\ln m_2^2\ln m_3\gamma_k^2+2\ln m_1\ln m_2^2\ln m_3\gamma_k^2-2\ln m_2^3\ln m_3\gamma_k^2\\
&-6\ln m_0^2\ln m_3^2\gamma_k^2+2\ln m_0\ln m_1\ln m_3^2\gamma_k^2-6\ln m_1^2\ln m_3^2\gamma_k^2\\
&+2\ln m_0\ln m_2\ln m_3^2\gamma_k^2+2\ln m_1\ln m_2\ln m_3^2\gamma_k^2-6\ln m_2^2\ln m_3^2\gamma_k^2\\
&-2\ln m_0\ln m_3^3\gamma_k^2-2\ln m_1\ln m_3^3\gamma_k^2-2\ln m_2\ln m_3^3\gamma_k^2+9\ln m_3^4\gamma_k^2\Big\}\Bigg]
\end{aligned}
\tag{B.19}
$$

$$
\begin{aligned}
C &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}^2\varepsilon_{t-1,k}^2\right)\right] \\
&= E\Bigg[\sum_{k=1}^{\bar{k}}\Big(\frac{1}{4}\ln m_{t+1,0}^2\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,0}^2 \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,1}^2\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,1}^2 \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,2}^2\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,2}^2 \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\qquad +\frac{1}{4}\ln m_{t+1,3}^2\Big\{(1-\gamma_k)(1-\gamma_k)\ln m_{t-1,3}^2 \\
&\qquad +(1-\gamma_k)\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k(1-\gamma_k)\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&\qquad +\gamma_k^2\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{16}\Big\{4(\ln m_1^4+\ln m_2^4+\ln m_3^4) \\
&\qquad -2\ln m_0^2(\ln m_1^2+\ln m_2^2+\ln m_3^2)(\gamma_k-2)\gamma_k \\
&\qquad +(3\ln m_1^4+3\ln m_2^4+3\ln m_3^4-2\ln m_2^2\ln m_3^2-2\ln m_1^2\ln m_2^2 \\
&\qquad -2\ln m_1^2\ln m_3^2)(\gamma_k-2)\gamma_k)+\ln m_0^4(4+3(\gamma_k-2)\gamma_k)\Big\}\Bigg]
\end{aligned}
\tag{B.20}
$$

$$
\begin{aligned}
D &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}^{3}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}^{3} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}^{3}+\frac{1}{4}\ln m_{t,1}^{3}+\frac{1}{4}\ln m_{t,2}^{3}+\frac{1}{4}\ln m_{t,3}^{3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}^{3} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}^{3}+\frac{1}{4}\ln m_{t,1}^{3}+\frac{1}{4}\ln m_{t,2}^{3}+\frac{1}{4}\ln m_{t,3}^{3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}^{3} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}^{3}+\frac{1}{4}\ln m_{t,1}^{3}+\frac{1}{4}\ln m_{t,2}^{3}+\frac{1}{4}\ln m_{t,3}^{3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t+1,3}\Big\{(1-\gamma_k)\ln m_{t,3}^{3} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t,0}^{3}+\frac{1}{4}\ln m_{t,1}^{3}+\frac{1}{4}\ln m_{t,2}^{3}+\frac{1}{4}\ln m_{t,3}^{3}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{16}\Big\{4(\ln m_1^4+\ln m_2^4+\ln m_3^4) \\
&\qquad +\ln m_0^4(4-3\gamma_k)+\ln m_0^3(\ln m_1+\ln m_2+\ln m_3)\gamma_k \\
&\qquad +\ln m_0(\ln m_1^3+\ln m_2^3+\ln m_3^3)\gamma_k \\
&\qquad +(-3\ln m_1^4-3\ln m_2^4-3\ln m_3^4+\ln m_2^3\ln m_3+\ln m_2\ln m_3^3 \\
&\qquad +\ln m_2\ln m_1^3+\ln m_3\ln m_1^3+\ln m_1\ln m_2^3+\ln m_1\ln m_3^3)\gamma_k\Big\}\Bigg]
\end{aligned}
\tag{B.21}
$$

$$
\begin{aligned}
E &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}^{2}\varepsilon_{t-1,k}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}^{2}(1-\gamma_k)\ln m_{t-1,0} \\
&\quad +(1-\gamma_k)\ln m_{t,0}^{2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}^{2}(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}^{2}(1-\gamma_k)\ln m_{t-1,1} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}^{2}(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}^{2}(1-\gamma_k)\ln m_{t-1,3}
\end{aligned}
$$

$$
\begin{aligned}
&+\gamma_k \frac{1}{4}\ln m_{t,0}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,1}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,2}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,3}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\\
&+\frac{1}{4}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}^2 (1-\gamma_k)\ln m_{t-1,1}\\
&+(1-\gamma_k)\ln m_{t,1}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,0}^2 (1-\gamma_k)\ln m_{t-1,0}+\gamma_k \frac{1}{4}\ln m_{t,1}^2 (1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k \frac{1}{4}\ln m_{t,2}^2 (1-\gamma_k)\ln m_{t-1,2}+\gamma_k \frac{1}{4}\ln m_{t,3}^2 (1-\gamma_k)\ln m_{t-1,3}\\
&+\gamma_k \frac{1}{4}\ln m_{t,0}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,1}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,2}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,3}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\\
&+\frac{1}{4}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}^2 (1-\gamma_k)\ln m_{t-1,2}\\
&+(1-\gamma_k)\ln m_{t,2}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,0}^2 (1-\gamma_k)\ln m_{t-1,0}+\gamma_k \frac{1}{4}\ln m_{t,1}^2 (1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k \frac{1}{4}\ln m_{t,2}^2 (1-\gamma_k)\ln m_{t-1,2}+\gamma_k \frac{1}{4}\ln m_{t,3}^2 (1-\gamma_k)\ln m_{t-1,3}\\
&+\gamma_k \frac{1}{4}\ln m_{t,0}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,1}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,2}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k \frac{1}{4}\ln m_{t,3}^2 \gamma_k \big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{4}\ln m_{t+1,3}\Big\{(1-\gamma_k)\ln m_{t,3}^2(1-\gamma_k)\ln m_{t-1,3}\\
&+(1-\gamma_k)\ln m_{t,3}^2\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}^2(1-\gamma_k)\ln m_{t-1,0}+\gamma_k\frac{1}{4}\ln m_{t,1}^2(1-\gamma_k)\ln m_{t-1,1}\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}^2(1-\gamma_k)\ln m_{t-1,2}+\gamma_k\frac{1}{4}\ln m_{t,3}^2(1-\gamma_k)\ln m_{t-1,3}\\
&+\gamma_k\frac{1}{4}\ln m_{t,0}^2\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,1}^2\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,2}^2\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\\
&+\gamma_k\frac{1}{4}\ln m_{t,3}^2\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\\
=\sum_{k=1}^{\bar{k}}\Bigg[&\frac{1}{64}\Big\{16\ln m_0^4+16\ln m_1^4+16\ln m_2^4+16\ln m_3^4\\
&-24\ln m_0^4\gamma_k+4\ln m_0^3\ln m_1\gamma_k+8\ln m_0^2\ln m_1^2\gamma_k+4\ln m_0\ln m_1^3\gamma_k\\
&-24\ln m_1^4\gamma_k+4\ln m_0^3\ln m_2\gamma_k+4\ln m_1^3\ln m_2\gamma_k+8\ln m_0^2\ln m_2^2\gamma_k\\
&-24\ln m_2^4\gamma_k+4\ln m_0\ln m_2^3\gamma_k+4\ln m_1\ln m_2^3\gamma_k+8\ln m_1^2\ln m_2^2\gamma_k\\
&+4\ln m_0^3\ln m_3\gamma_k+4\ln m_1^3\ln m_3\gamma_k+4\ln m_2^3\ln m_3\gamma_k+8\ln m_0^2\ln m_3^2\gamma_k\\
&+8\ln m_1^2\ln m_3^2\gamma_k+8\ln m_2^2\ln m_3^2\gamma_k+4\ln m_0\ln m_3^3\gamma_k+4\ln m_1\ln m_3^3\gamma_k\\
&+4\ln m_2\ln m_3^3\gamma_k-24\ln m_3^4\gamma_k+9\ln m_0^4\gamma_k^2+2\ln m_0^3\ln m_1\gamma_k^2\\
&-6\ln m_0^2\ln m_1^2\gamma_k^2-2\ln m_0\ln m_1^3\gamma_k^2+9\ln m_1^4\gamma_k^2-2\ln m_0^3\ln m_2\gamma_k^2\\
&+2\ln m_0^2\ln m_1\ln m_2\gamma_k^2+2\ln m_0\ln m_1^2\ln m_2\gamma_k^2+2\ln m_1^3\ln m_2\gamma_k^2\\
&-6\ln m_0^2\ln m_2^2\gamma_k^2+2\ln m_0\ln m_1\ln m_2^2\gamma_k^2-6\ln m_1^2\ln m_2^2\gamma_k^2\\
&-2\ln m_0\ln m_2^3\gamma_k^2-2\ln m_1\ln m_2^3\gamma_k^2+9\ln m_2^4\gamma_k^2\\
&-2\ln m_0^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_1\ln m_3\gamma_k^2+2\ln m_0\ln m_1^2\ln m_3\gamma_k^2\\
&-2\ln m_1^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_2\ln m_3\gamma_k^2+2\ln m_1^2\ln m_2\ln m_3\gamma_k^2\\
&+2\ln m_0\ln m_2^2\ln m_3\gamma_k^2+2\ln m_1\ln m_2^2\ln m_3\gamma_k^2-2\ln m_2^3\ln m_3\gamma_k^2\\
&-6\ln m_0^2\ln m_3^2\gamma_k^2+2\ln m_0\ln m_1\ln m_3^2\gamma_k^2-6\ln m_1^2\ln m_3^2\gamma_k^2\\
&+2\ln m_0\ln m_2\ln m_3^2\gamma_k^2+2\ln m_1\ln m_2\ln m_3^2\gamma_k^2-6\ln m_2^2\ln m_3^2\gamma_k^2\\
&-2\ln m_0\ln m_3^3\gamma_k^2-2\ln m_1\ln m_3^3\gamma_k^2-2\ln m_2\ln m_3^3\gamma_k^2+9\ln m_3^4\gamma_k^2\Big\}\Bigg]
\end{aligned}
\tag{B.22}
$$

$$
\begin{aligned}
F &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t+1,k}\varepsilon_{t,k}\varepsilon_{t-1,k}^{2}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t+1,0}\Big\{(1-\gamma_k)\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^{2} \\
&\quad +(1-\gamma_k)\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^{2}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^{2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,3}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t+1,1}\Big\{(1-\gamma_k)\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^{2} \\
&\quad +(1-\gamma_k)\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^{2}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^{2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,3}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t+1,2}\Big\{(1-\gamma_k)\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^{2} \\
&\quad +(1-\gamma_k)\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^{2}+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^{2}+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,3}^{2} \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big) \\
&\quad +\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^{2}+\frac{1}{4}\ln m_{t-1,1}^{2}+\frac{1}{4}\ln m_{t-1,2}^{2}+\frac{1}{4}\ln m_{t-1,3}^{2}\big)\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{4}\ln m_{t+1,3}\Big\{(1-\gamma_k)\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,3}^2 \\
&+(1-\gamma_k)\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&+\gamma_k\frac{1}{4}\ln m_{t,0}(1-\gamma_k)\ln m_{t-1,0}^2+\gamma_k\frac{1}{4}\ln m_{t,1}(1-\gamma_k)\ln m_{t-1,1}^2 \\
&+\gamma_k\frac{1}{4}\ln m_{t,2}(1-\gamma_k)\ln m_{t-1,2}^2+\gamma_k\frac{1}{4}\ln m_{t,3}(1-\gamma_k)\ln m_{t-1,3}^2 \\
&+\gamma_k\frac{1}{4}\ln m_{t,0}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&+\gamma_k\frac{1}{4}\ln m_{t,1}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&+\gamma_k\frac{1}{4}\ln m_{t,2}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big) \\
&+\gamma_k\frac{1}{4}\ln m_{t,3}\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
=\sum_{k=1}^{\bar{k}}\Bigg[&\frac{1}{64}\Big\{16\ln m_0^4+16\ln m_1^4+16\ln m_2^4+16\ln m_3^4 \\
&-24\ln m_0^4\gamma_k+4\ln m_0^3\ln m_1\gamma_k+8\ln m_0^2\ln m_1^2\gamma_k+4\ln m_0\ln m_1^3\gamma_k \\
&-24\ln m_1^4\gamma_k+4\ln m_0^3\ln m_2\gamma_k+4\ln m_1^3\ln m_2\gamma_k+8\ln m_0^2\ln m_2^2\gamma_k \\
&-24\ln m_2^4\gamma_k+4\ln m_0\ln m_2^3\gamma_k+4\ln m_1\ln m_2^3\gamma_k+8\ln m_1^2\ln m_2^2\gamma_k \\
&+4\ln m_0^3\ln m_3\gamma_k+4\ln m_1^3\ln m_3\gamma_k+4\ln m_2^3\ln m_3\gamma_k+8\ln m_0^2\ln m_3^2\gamma_k \\
&+8\ln m_1^2\ln m_3^2\gamma_k+8\ln m_2^2\ln m_3^2\gamma_k+4\ln m_0\ln m_3^3\gamma_k+4\ln m_1\ln m_3^3\gamma_k \\
&+4\ln m_2\ln m_3^3\gamma_k-24\ln m_3^4\gamma_k+9\ln m_0^4\gamma_k^2+2\ln m_0^3\ln m_1\gamma_k^2 \\
&-6\ln m_0^2\ln m_1^2\gamma_k^2-2\ln m_0\ln m_1^3\gamma_k^2+9\ln m_1^4\gamma_k^2-2\ln m_0^3\ln m_2\gamma_k^2 \\
&+2\ln m_0^2\ln m_1\ln m_2\gamma_k^2+2\ln m_0\ln m_1^2\ln m_2\gamma_k^2+2\ln m_1^3\ln m_2\gamma_k^2 \\
&-6\ln m_0^2\ln m_2^2\gamma_k^2+2\ln m_0\ln m_1\ln m_2^2\gamma_k^2-6\ln m_1^2\ln m_2^2\gamma_k^2 \\
&-2\ln m_0\ln m_2^3\gamma_k^2-2\ln m_1\ln m_2^3\gamma_k^2+9\ln m_2^4\gamma_k^2 \\
&-2\ln m_0^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_1\ln m_3\gamma_k^2+2\ln m_0\ln m_1^2\ln m_3\gamma_k^2 \\
&-2\ln m_1^3\ln m_3\gamma_k^2+2\ln m_0^2\ln m_2\ln m_3\gamma_k^2+2\ln m_1^2\ln m_2\ln m_3\gamma_k^2 \\
&+2\ln m_0\ln m_2^2\ln m_3\gamma_k^2+2\ln m_1\ln m_2^2\ln m_3\gamma_k^2-2\ln m_2^3\ln m_3\gamma_k^2 \\
&-6\ln m_0^2\ln m_3^2\gamma_k^2+2\ln m_0\ln m_1\ln m_3^2\gamma_k^2-6\ln m_1^2\ln m_3^2\gamma_k^2 \\
&+2\ln m_0\ln m_2\ln m_3^2\gamma_k^2+2\ln m_1\ln m_2\ln m_3^2\gamma_k^2-6\ln m_2^2\ln m_3^2\gamma_k^2 \\
&-2\ln m_0\ln m_3^3\gamma_k^2-2\ln m_1\ln m_3^3\gamma_k^2-2\ln m_2\ln m_3^3\gamma_k^2+9\ln m_3^4\gamma_k^2\Big\}\Bigg]
\end{aligned}
\tag{B.23}
$$

$$
G=E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^4\right)\right]=\frac{1}{4}\Big\{\ln m_{t,0}^4+\ln m_{t,1}^4+\ln m_{t,2}^4+\ln m_{t,3}^4\Big\} \tag{B.24}
$$

$$
\begin{aligned}
H &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^{3}\varepsilon_{t-1,k}\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t,0}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,0} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,1}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,1} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,2}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,2} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,3}^{3}\Big\{(1-\gamma_k)\ln m_{t-1,3} \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}+\frac{1}{4}\ln m_{t-1,1}+\frac{1}{4}\ln m_{t-1,2}+\frac{1}{4}\ln m_{t-1,3}\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{16}\Big\{4(\ln m_1^4+\ln m_2^4+\ln m_3^4) \\
&\qquad +\ln m_0^4(4-3\gamma_k)+\ln m_0^3(\ln m_1+\ln m_2+\ln m_3)\gamma_k \\
&\qquad +\ln m_0(\ln m_1^3+\ln m_2^3+\ln m_3^3)\gamma_k \\
&\qquad +(-3\ln m_1^4-3\ln m_2^4-3\ln m_3^4+\ln m_2^3\ln m_3+\ln m_2\ln m_3^3 \\
&\qquad +\ln m_2\ln m_1^3+\ln m_3\ln m_1^3+\ln m_1\ln m_2^3+\ln m_1\ln m_3^3)\gamma_k\Big\}\Bigg]
\end{aligned}
\tag{B.25}
$$

$$
\begin{aligned}
I &= E\left[\sum_{k=1}^{\bar{k}}\left(\varepsilon_{t,k}^2\varepsilon_{t-1,k}^2\right)\right] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{4}\ln m_{t,0}^2\Big\{(1-\gamma_k)\ln m_{t-1,0}^2 \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,1}^2\Big\{(1-\gamma_k)\ln m_{t-1,1}^2 \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,2}^2\Big\{(1-\gamma_k)\ln m_{t-1,2}^2 \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\} \\
&\quad +\frac{1}{4}\ln m_{t,3}^2\Big\{(1-\gamma_k)\ln m_{t-1,3}^2 \\
&\qquad +\gamma_k\big(\frac{1}{4}\ln m_{t-1,0}^2+\frac{1}{4}\ln m_{t-1,1}^2+\frac{1}{4}\ln m_{t-1,2}^2+\frac{1}{4}\ln m_{t-1,3}^2\big)\Big\}\Bigg] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\frac{1}{16}\Big\{4(\ln m_1^4+\ln m_2^4+\ln m_3^4) \\
&\qquad +\ln m_0^4(4-3\gamma_k)+\ln m_0^3(\ln m_1+\ln m_2+\ln m_3)\gamma_k \\
&\qquad +\ln m_0(\ln m_1^3+\ln m_2^3+\ln m_3^3)\gamma_k \\
&\qquad +(-3\ln m_1^4-3\ln m_2^4-3\ln m_3^4+\ln m_2^3\ln m_3+\ln m_2\ln m_3^3 \\
&\qquad +\ln m_2\ln m_1^3+\ln m_3\ln m_1^3+\ln m_1\ln m_2^3+\ln m_1\ln m_3^3)\gamma_k\Big\}\Bigg]
\end{aligned}
\tag{B.26}
$$

Using the above equations from (B.18) to (B.26), the first term of autocovariance of squared log increments (A.18) is given by $E\Big[\big(\varepsilon_{t+1,k}-\varepsilon_{t,k}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\Big]$, which is determined by the sequence of equations (B.8) and (B.9) as follows:

$$
\begin{aligned}
&E\Big[\big(\varepsilon_{t+1,k}-\varepsilon_{t,k}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\Big] \\
&= \sum_{k=1}^{\bar{k}}\Bigg[\Big(\frac{1}{4}\gamma_k\Big)^2\Big\{(\ln m_0-\ln m_1)^4+(\ln m_0-\ln m_2)^4+(\ln m_0-\ln m_3)^4 \\
&+(\ln m_1-\ln m_2)^4+(\ln m_1-\ln m_3)^4+(\ln m_2-\ln m_3)^4\Big\}\Bigg]
\end{aligned}
\tag{B.27}
$$

For each pair of the second part $\big(\varepsilon_{t+1,k'}-\varepsilon_{t,k'}\big)\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)$ the sequences of equation (B.28) contribute to non-zero entries, where the switching occurs with probabilities $\big(\frac{1}{4}\gamma_k\big)\big(\frac{1}{4}\gamma_{k'}\big)$. The possible combination of the transitions of multipliers looks like

$$
\begin{array}{llll}
m_0\to m_1 & m_1\to m_0 & m_2\to m_0 & m_3\to m_0 \\
m_0\to m_2 & m_1\to m_2 & m_2\to m_1 & m_3\to m_1 \\
m_0\to m_3 & m_1\to m_3 & m_2\to m_3 & m_3\to m_2.
\end{array}
\tag{B.28}
$$

Total 144 relevant sequences lead to

$$\begin{aligned}
&E\Big[\big(\varepsilon_{t+1,k'}-\varepsilon_{t,k'}\big)^2\big(\varepsilon_{t,k}-\varepsilon_{t-1,k}\big)^2\Big] \\
&=\sum_{k=1}^{\bar{k}}\Bigg[\Big(\frac{1}{4}\gamma_k\Big)\sum_{k=1,k'\neq k}^{\bar{k}}\Big(\frac{1}{4}\gamma_k'\Big)\Big\{(\ln m_0-\ln m_1)^2+(\ln m_0-\ln m_2)^2 \\
&\qquad\qquad +(\ln m_0-\ln m_3)^2+(\ln m_1-\ln m_2)^2 \\
&\qquad\qquad +(\ln m_1-\ln m_3)^2+(\ln m_2-\ln m_3)^2\Big\}^2
\end{aligned} \tag{B.29}$$

For the last term of $(\varepsilon_{k+1,k'}-\varepsilon_{t,k'})(\varepsilon_{t+1,k}-\varepsilon_{t,k})(\varepsilon_{k,t}-\varepsilon_{t-1,k})(\varepsilon_{t,k'}-\varepsilon_{t-1,k'})$ the non- zero condition reads to $\varepsilon_{t+1,n}\neq\varepsilon_{t,n}\neq\varepsilon_{t-1,n}$ for $n=k$, k' and relevant sequences are represented in equations (B.8) and (B.9). Here, each non-zero entry occurs with the probability $\big(\frac{1}{4}\gamma_k\big)^2\big(\frac{1}{4}\gamma_{k'}\big)^2$. The expectation value of this component results in

$$\begin{aligned}
&E\Big[(\varepsilon_{t+1,k'}-\varepsilon_{t,k'})(\varepsilon_{t+1,k}-\varepsilon_{t,k})(\varepsilon_{t,k}-\varepsilon_{t-1,k})(\varepsilon_{t,k'}-\varepsilon_{t-1,k'})\Big] \\
&=\sum_{k=1}^{\bar{k}}\Bigg\{\Big(\frac{1}{4}\gamma_k\Big)^2\sum_{k'=1,k'\neq k}^{\bar{k}}\Big(\frac{1}{4}\gamma_{k'}\Big)^2\Big\{(\ln m_0-\ln m_1)^2+(\ln m_0-\ln m_2)^2 \\
&\qquad\qquad +(\ln m_0-\ln m_3)^2+(\ln m_1-\ln m_2)^2 \\
&\qquad\qquad +(\ln m_1-\ln m_3)^2+(\ln m_2-\ln m_3)^2\Big\}^2
\end{aligned} \tag{B.30}$$

Putting all three cases together we obtain

$$\begin{aligned}
E[\eta_{t+1,1}^2\eta_{t,1}^2] &= \sum_{k=k'=1}^{\bar{k}}\Big(\frac{1}{4}\gamma_k\Big)^2\Omega+\sum_{k=1}^{\bar{k}}\Big\{\Big(\frac{1}{4}\gamma_k\Big)\sum_{k'=1,k'\neq k}^{\bar{k}}\Big(\frac{1}{4}\gamma_{k'}\Big)\Big\}\Psi \\
&\quad+\sum_{k=1}^{\bar{k}}\Big\{\Big(\frac{1}{4}\gamma_k\Big)^2\sum_{k'=1,k'\neq k}^{\bar{k}}\Big(\frac{1}{4}\gamma_{k'}\Big)^2\Big\}\Psi
\end{aligned} \tag{B.31}$$

with

$$\begin{aligned}
\Omega=\Big\{&(\ln m_0-\ln m_1)^4+(\ln m_0-\ln m_2)^4+(\ln m_0-\ln m_3)^4 \\
&+(\ln m_1-\ln m_2)^4+(\ln m_1-\ln m_3)^4+(\ln m_2-\ln m_3)^4\Big\}
\end{aligned} \tag{B.32}$$

and

$$\begin{aligned}
\Psi=\Big\{&(\ln m_0-\ln m_1)^2+(\ln m_0-\ln m_2)^2+(\ln m_0-\ln m_3)^2 \\
&+(\ln m_1-\ln m_2)^2+(\ln m_1-\ln m_3)^2+(\ln m_2-\ln m_3)^2\Big\}^2
\end{aligned} \tag{B.33}$$

For $T > 1$ this covariance is formulated by

$$
\begin{aligned}
&E[\eta_{t+T,T}^2 \eta_{t,T}^2] \\
&= \sum_{k=1}^{\bar{k}} \frac{1}{16}\left(1 - \left(1-\gamma_k\right)^T\right)^2 \Omega \\
&\quad + \frac{1}{4}\sum_{k=1}^{\bar{k}} \left\{ \frac{1}{4}\left(1 - \left(1-\gamma_k\right)^T\right) \sum_{k'=1, k' \neq k}^{\bar{k}} \frac{1}{4}\left(1 - \left(1-\gamma_k'\right)^T\right) \right\} \Psi \\
&\quad + 2\sum_{k=1}^{\bar{k}} \left\{ \frac{1}{16}\left(1 - \left(1-\gamma_k\right)^T\right)^2 \sum_{k'=1, k' \neq k}^{\bar{k}} \frac{1}{16}\left(1 - \left(1-\gamma_k'\right)^T\right)^2 \right\} \Psi. \qquad \text{(B.34)}
\end{aligned}
$$

Appendix C

The log moments of the normal distribution

C.1 The log moments of the normal distribution

Using the Gamma function[1] and its derivatives, we can calculate the log moments of the normal distribution. For the first log moment we obtain

$$\begin{aligned} E[\ln|u|] &= \int_{-\infty}^{\infty} \ln|u| \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) du \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{2} \big(\ln(y) - \ln(2) \big) \exp(-y) \frac{\sqrt{2}}{2} (y)^{-\frac{1}{2}} dy \\ &= \frac{1}{\sqrt{2\pi}} \big(\frac{1}{\sqrt{2}} \int_0^{\infty} \ln(y) \exp(-y) y^{-\frac{1}{2}} dy - \frac{1}{\sqrt{2}} \ln(2) \int_0^{\infty} \exp(-y) y^{-\frac{1}{2}} dy \big) \\ &= \frac{1}{\sqrt{2\pi}} \big\{ \frac{1}{\sqrt{2}} \Gamma'(0.5) - \frac{1}{\sqrt{2}} \ln(2) \Gamma(0.5) \big\}, \end{aligned} \tag{C.1}$$

where we substitute $y = \frac{u^2}{2}$ and $\ln(\epsilon) = \frac{1}{2} \big\{ \ln(y) - \ln(2) \big\}$.

For the second log moment we also use the substitution $\ln(\epsilon)^2 = \frac{1}{4} \big\{ \ln(y)^2 - 2\ln(2) * \ln(y) + \ln(2)^2 \big\}$.

$$\begin{aligned} E[\ln|u|^2] &= \int_{-\infty}^{\infty} (\ln|u|)^2 \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) du \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{4} \big(\ln(y)^2 - 2\ln(2) * \ln(y) + \ln(2)^2 \big) \exp(-y) \frac{1}{\sqrt{2}} (y)^{-\frac{1}{2}} dy \\ &= \frac{1}{4\sqrt{\pi}} \Gamma''(0.5) - \frac{1}{2\sqrt{\pi}} \ln(2) \Gamma'(0.5) + \frac{1}{4\sqrt{\pi}} \ln(2)^2 \Gamma(0.5) \end{aligned} \tag{C.2}$$

For the third log moment we use the substitution $\ln(\epsilon)^3 = \frac{1}{8} \big\{ \ln(y)^3 - 3\ln(2)\ln(y)^2 + 3\ln(y)\ln(2)^2 - \ln(2)^3 \big\}$.

[1] For $x > 0$, the Gamma function denoted $\gamma(x)$ is defined : $\gamma(x) = \int_0^{\infty} t^{x-1} \exp(-t) dt$. Its derivatives can be deduced by differentiation under the integral sign of gamma function: $\gamma'(x) = \int_0^{\infty} t^{x-1} \exp(-t) \ln(t) dt$ and $\gamma^{(n)}(x) = \int_0^{\infty} t^{x-1} \exp(-t) \ln(t)^n dt$.

$$
\begin{aligned}
E[\ln|u|^3] &= \int_{-\infty}^{\infty} (\ln|u|)^3 \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) du \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{8} \big(\ln(y)^3 - 3\ln(2)\ln(y)^2 + 3\ln(y)\ln(2)^2 - \ln(2)^3 \big) \exp(-y) \frac{1}{\sqrt{2}} (y)^{-\frac{1}{2}} dy \\
&= \frac{1}{8\sqrt{\pi}} \Gamma'''(0.5) - \frac{3}{8\sqrt{\pi}} \ln(2)\Gamma''(0.5) + \frac{3}{8\sqrt{\pi}} \ln(2)^2 \Gamma'(0.5) - \frac{1}{8\sqrt{\pi}} \ln(2)^3 \Gamma(0.5) \quad \text{(C.3)}
\end{aligned}
$$

For the forth log moment we use the substitution $\ln(\epsilon)^4 = \frac{1}{16} \big\{ \ln(y) - \ln(2) \big\}^4$.

$$
\begin{aligned}
E[\ln|u|^4] &= \int_{-\infty}^{\infty} (\ln|u|)^4 \frac{1}{\sqrt{2\pi}} \exp(\frac{-u^2}{2}) du \\
&= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \frac{1}{16} \big(\ln(y) - \ln(2) \big)^4 \exp(-y) \frac{1}{\sqrt{2}} (y)^{-\frac{1}{2}} dy \\
&= \frac{1}{16\sqrt{\pi}} \Gamma^{(4)}(0.5) - \frac{4}{16\sqrt{\pi}} \ln(2)\Gamma'''(0.5) + \frac{6}{16\sqrt{\pi}} \ln(2)^2 \Gamma''(0.5) \\
&\quad - \frac{4}{16\sqrt{\pi}} \ln(2)^3 \Gamma'(0.5) + \frac{1}{16\sqrt{\pi}} \ln(2)^4 \Gamma(0.5) \quad \text{(C.4)}
\end{aligned}
$$

C.2 The log moments of the Student-t distribution

Again, we caculate the log absolute moments of the Student-t distribution.[2] For the first log moment we obtain:

$$
\begin{aligned}
E[\ln|u|] &= \int_{-\infty}^{\infty} \ln|u| \frac{\Gamma(\frac{1}{2}(\nu+1)}{\sqrt{\nu\pi}\Gamma(\frac{1}{2}\nu)\big(1+\frac{x^2}{\nu}\big)^{(\nu+1)/2}} du \\
&= -\frac{\big(\Gamma[1+\frac{\nu}{2}](-\Gamma'[1] + \ln(4) - \ln(\nu) + \psi_0(\frac{\nu}{2})\big)}{\nu\Gamma[\frac{\nu}{2}]}, \quad \text{(C.5)}
\end{aligned}
$$

where the logarithmic derivative of gamma function $\Gamma(x)$ is defined as $\Psi_n(x) \equiv \frac{d}{dx} \ln\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$. The Euler constant, denoted $-\Gamma'(-1)$, has the numerical value $0.57721566490\cdots$. Th nth derivative of $\Psi(x)$ is called the polygamma function, denoted $\psi_n(x) = \frac{d^n}{dx^n} \frac{\Gamma'(x)}{\Gamma(x)} = \frac{d^n}{dx^n} \psi_0(x)$.

The second log moment of the Student-t distribution becomes:

$$
\begin{aligned}
E[\ln|u|^2] &= \int_{-\infty}^{\infty} \ln|u|^2 \frac{\Gamma(\frac{1}{2}(\nu+1)}{\sqrt{\nu\pi}\Gamma(\frac{1}{2}\nu)\big(1+\frac{x^2}{\nu}\big)^{(\nu+1)/2}} du \\
&= \frac{1}{8\nu\Gamma(\frac{\nu}{2})} \Big(-8\Gamma(1+\frac{\nu}{2}) \big((-\Gamma'(1) + \ln(4))\ln(\nu) - (-\Gamma'(1) + \ln(4) \\
&\quad - \ln(\nu))\psi_0(\frac{\nu}{2}) \big) + \nu\Gamma(\frac{\nu}{2}) \big(2(-\Gamma'(1))^2 + \pi^2 + 4(-\Gamma'(1))\ln(4) + 2\ln(4)^2 + 2\ln(\nu)^2 \\
&\quad + 2\psi_0(\frac{\nu}{2})^2 + 2\psi_1(\frac{\nu}{2}) \big) \Big) \quad \text{(C.6)}
\end{aligned}
$$

The third log moment of the Student-t distribution becomes:

[2] We use for the calculation of the log absolute moments of the Student-t distribuion Mathematica 4.2.

$$
\begin{aligned}
E[\ln|u|^3] &= \int_{-\infty}^{\infty} \ln|u|^3 \frac{\Gamma(\frac{1}{2}(\nu+1)}{\sqrt{\nu\pi}\Gamma(\frac{1}{2}\nu)\big(1+\frac{x^2}{\nu}\big)^{(\nu+1)/2}} du \\
&= \frac{1}{16}\Big\{ -(-\Gamma'(1)+\ln(4)-\ln(\nu))\big(3\pi^2+2(-\Gamma'(1)+\ln(4))^2 \\
&\quad +2\ln(\nu)\big(-2(-\Gamma'(1)+\ln(4))+\ln(\nu)\big)\big) \\
&\quad +\psi_0(\frac{\nu}{2})\Big(-3(\pi^2+2(-\Gamma'(1)+\ln(4))^2)+12(-\Gamma'(1)+\ln(4))\ln(\nu)-6\ln(\nu)^2 \\
&\quad -2\psi_0(\frac{\nu}{2})\big(3(-\Gamma'(1)+\ln(4)-\ln(\nu))+\psi_0(\frac{\nu}{2})\big)\Big) \\
&\quad -6\big(-\Gamma'(1)+\ln(4)-\ln(\nu)+\psi_0(\frac{\nu}{2})\big)\psi_1(\frac{\nu}{2})-2\psi_2(\frac{\nu}{2})-28\zeta(3)\Big\}, \qquad \text{(C.7)}
\end{aligned}
$$

where $\zeta(3)$,known as Riemann zeta function is equal to 1.2020569032.[3]

The forth log moment of the Student-t distribution becomes:

$$
\begin{aligned}
E[\ln|u|^4] &= \int_{-\infty}^{\infty} \ln|u|^4 \frac{\Gamma(\frac{1}{2}(\nu+1)}{\sqrt{\nu\pi}\Gamma(\frac{1}{2}\nu)\big(1+\frac{x^2}{\nu}\big)^{(\nu+1)/2}} du \\
&= \frac{1}{64}\Big\{ -(-4\Gamma'(1)^4+12(-\Gamma'(1))^2\pi^2+7\pi^4+8(-\Gamma'(1))(2(-\Gamma'(1))^2+3\pi^2)\ln(4) \\
&\quad +12(2(-\Gamma'(1))^2+\pi^2)\ln(4)^2+16(-\Gamma'(1))\ln(4)^3+4\ln(4)^4-4(2(-\Gamma'(1)) \\
&\quad +\ln(16)-\ln(\nu))\ln(\nu)(3\pi^2+2(-\Gamma'(1)+\ln(4))^2-2(-\Gamma'(1)+\ln(4))\ln(\nu) \\
&\quad +\ln(\nu)^2)+16(-\Gamma'(1)+\ln(4)-\ln(\nu))\psi_0(\frac{\nu}{2})^3+4\psi_0(\frac{\nu}{2})^4 \\
&\quad +12\big(\pi^2+2(-\Gamma'(1)+\ln(4)-\ln(\nu))^2+2\ln(\nu)(-2(-\Gamma'(1)+\ln(4))+\ln(\nu))\big)\psi_1(\frac{\nu}{2}) \\
&\quad +12\psi_1(\frac{\nu}{2})^2+12\psi_0(\frac{\nu}{2})^2\big(\pi^2+2(-\Gamma'(1)+\ln(4))^2+\ln(\nu)(-2(-\Gamma'(1)+\ln(4))+\ln(\nu)) \\
&\quad +2\psi_1(\frac{\nu}{2})\big)+4\psi_3(\frac{\nu}{2})+16(-\Gamma'(1)+\ln(4)-\ln(\nu))(\psi_2(\frac{\nu}{2})+14\zeta(3)) \\
&\quad +8\psi_0(\frac{\nu}{2})\big((-\Gamma'(1)+\ln(4)-\ln(\nu))(3\pi^2+2(-\Gamma'(1)+\ln(4))^2+2\ln(\nu)(-2(-\Gamma'(1)+\ln(4)) \\
&\quad +\ln(\nu)))+6(-\Gamma'(1)+\ln(4)-\ln(\nu))\psi_1(\frac{\nu}{2})+2\psi_2(\frac{\nu}{2})+28\zeta(3)\big)\Big\} \qquad \text{(C.8)}
\end{aligned}
$$

C.3 Monte Carlo Esimation of t-MSMF

We present for the binomial and trinomial t-MSMF models the Monte Carlo results of the estimators: ML, GMMI, GMMII and MM. For the Monte Carlo experiments the combinations of parameters are used in table (5.1). There were no convergence problem with the exception of MM. In MM we use the GAUSS module *nonlinear equations* that fits the empirical moments to the theoretical values of $E[|x_t|^q], q = 1, 1.5, 2, 3$. In MM we evaluate only the estimates without convergence problem and note the number of the valid estimates in the footnote.

[3]The Riemann zeta function can be defined by the integral $\zeta(x) = \frac{1}{\Gamma(x)}\int_0^\infty \frac{u^{x-1}}{\exp(u)-1}du$

Binomial MSMF with Student's-t innovation (1)

Panel A: Parameter $(m_0 = 1.3, \nu = 4$ **and** $\sigma = 1)$ **estimation with** $\bar{k} = 6$								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m	1.3							
$\hat{m}_{sim}$	1.297	1.298	1.263	1.271	1.346	1.339	1.396	1.397
FSSE	0.018	0.014	0.100	0.077	0.033	0.027	0.032	0.025
RMSE	0.018	0.014	0.107	0.082	0.056	0.048	0.101	0.100
ν	4							
$\hat{\nu}_{sim}$	4.029	4.014	4.512	4.258	5.049	4.718	4.226	4.182
FSSE	0.321	0.222	0.908	0.720	0.668	0.536	0.559	0.331
RMSE	0.322	0.222	1.043	0.765	1.244	0.896	0.603	0.377
σ	1							
$\hat{\sigma}$	1.001	1.001	1.014	1.002	1.072	1.058	1.083	1.081
FSSE	0.041	0.029	0.082	0.068	0.060	0.047	0.053	0.038
RMSE	0.041	0.029	0.083	0.068	0.093	0.074	0.098	0.090
Pandel B: Parameter $(m_0 = 1.3, \nu = 5$ **and** $\sigma = 1)$ **estimation with k = 6**								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m	1.3							
$\hat{m}_{sim}$	1.299	1.299	1.217	1.235	1.311	1.307	1.366	1.363
FSSE	0.017	0.012	0.099	0.071	0.028	0.025	0.037	0.026
RMSE	0.017	0.012	0.129	0.097	0.030	0.025	0.075	0.068
ν	5							
$\hat{\nu}_{sim}$	5.034	5.025	4.617	4.426	5.384	5.156	5.229	5.189
FSSE	0.505	0.339	0.721	0.499	0.609	0.504	0.396	0.344
FSSE	0.507	0.339	0.816	0.761	0.721	0.528	0.458	0.392
σ	1							
$\hat{\sigma}$	0.998	0.999	0.952	0.948	1.007	1.002	1.053	1.048
FSSE	0.040	0.028	0.058	0.044	0.045	0.036	0.055	0.038
RMSE	0.040	0.028	0.076	0.068	0.045	0.036	0.076	0.061

Table C.1: Simulated average, finite sample standand error(FSSE) and root mean squared error(RMSE) for ML, GMM I, GMM II, and MM Estimators. In panel A, the simulated parameters are $m_0 = 1.2, m_1 = 0.5$ $\nu = 4$ and $\sigma = 1$, respectively. In panel B, the simulated parameters are $m_0 = 1.3$, $\nu = 5$ and $\sigma = 1$, respectively. MM^*: there are 398 valid estimations for MM.

Binomial MSMF with Student's-t innovation (2)

Pandel C: Parameter $(m_0 = 1.5, \nu = 4$ **and** $\sigma = 1)$ **estimation with k = 6**								
method	ML		GMM I		GMM II		MM^*	
n	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m	1.5							
$\hat{m}_{sim}$	1.498	1.499	1.493	1.493	1.522	1.519	1.546	1.548
FSSE	0.016	0.011	0.040	0.032	0.019	0.031	0.025	0.023
RMSE	0.016	0.012	0.041	0.032	0.029	0.036	0.052	0.053
ν	4							
$\hat{\nu}_{sim}$	4.048	4.035	4.400	4.248	5.133	4.879	4.094	4.080
FSSE	0.421	0.278	0.900	0.726	0.686	0.601	0.141	0.133
RMSE	0.424	0.280	0.985	0.767	1.324	1.065	0.169	0.155
σ	1							
$\hat{\sigma}$	1.000	0.999	1.010	1.003	1.059	1.050	1.072	1.069
FSSE	0.068	0.044	0.094	0.075	0.073	0.075	0.072	0.052
RMSE	0.068	0.044	0.094	0.075	0.094	0.090	0.101	0.087
Pandel D: Parameter $(m_0 = 1.5, \nu = 5$ **and** $\sigma = 1)$ **estimation with k = 6**								
method	ML		GMM I		GMM II		MM^*	
n	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m	1.5							
$\hat{m}_{sim}$	1.498	1.499	1.470	1.474	1.498	1.499	1.529	1.530
FSSE	0.016	0.011	0.043	0.029	0.020	0.015	0.030	0.027
RMSE	0.017	0.011	0.053	0.039	0.020	0.015	0.041	0.040
ν	5							
$\hat{\nu}_{sim}$	5.081	5.042	4.623	4.480	5.384	5.175	5.110	5.091
FSSE	0.735	0.443	0.735	0.528	0.670	0.563	0.448	0.403
RMSE	0.740	0.445	0.826	0.741	0.772	0.590	0.462	0.413
σ	1							
$\hat{\sigma}$	0.997	0.998	0.953	0.949	0.979	0.981	1.038	1.036
FSSE	0.066	0.045	0.072	0.053	0.063	0.049	0.071	0.053
RMSE	0.066	0.045	0.086	0.073	0.067	0.053	0.081	0.064

Table C.2: Simulated average, finite sample standand error(FSSE) and root mean squared error(RMSE) for ML, GMM I, GMM II, and MM Estimators. In panel A, the simulated parameters are $m_0 = 1.2, m_1 = 0.5$ $\nu = 4$, and $\sigma = 1$, respectively. In panel B, the simulated paramers are $m_0 = 1.5$, $\nu = 5$, and $\sigma = 1$, respectively. MM^*: there are 396 valid estimations for MM.

Trinomial MSMF (k=3) with Student's-t innovation (1)

Pandel A: Parameter ($m_0 = 1.2, m_1 = 0.5$, $\nu = 4$ **and** $\sigma = 1$)								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m_0	1.2							
$\hat{m}_0$	1.178	1.186	1.227	1.228	1.327	1.327	1.774	1.612
FSSE	0.097	0.083	0.165	0.150	0.136	0.130	0.464	0.350
RMSE	0.099	0.084	0.168	0.152	0.186	0.182	0.738	0.540
m_1	0.5							
$\hat{m}_1$	0.506	0.504	0.497	0.492	0.542	0.543	0.362	0.369
FSSE	0.023	0.017	0.088	0.068	0.100	0.092	0.037	0.027
RMSE	0.024	0.017	0.088	0.069	0.108	0.101	0.143	0.134
ν	4							
$\hat{\nu}_{sim}$	4.051	4.027	4.453	4.310	4.920	5.118	4.087	4.077
FSSE	0.283	0.199	0.750	0.521	1.777	2.002	0.215	0.032
RMSE	0.288	0.201	0.838	0.606	2.001	2.293	0.232	0.084
σ	1							
$\hat{\sigma}$	0.997	0.996	1.129	1.098	0.911	0.899	1.059	1.056
FSSE	0.040	0.027	0.162	0.149	0.370	0.348	0.047	0.030
RMSE	0.040	0.027	0.207	0.179	0.380	0.362	0.075	0.064
Pandel B: Parameter ($m_0 = 1.2, m_1 = 0.5$, $\nu = 5$ **and** $\sigma = 1$)								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m_0	1.2							
$\hat{m}_0$	1.191	1.194	1.252	1.265	1.321	1.317	1.688	1.478
FSSE	0.099	0.081	0.103	0.082	0.123	0.108	0.441	0.347
RMSE	0.100	0.082	0.115	0.105	0.172	0.159	0.658	0.444
m_1	0.5							
$\hat{m}_1$	0.507	0.504	0.475	0.463	0.509	0.495	0.415	0.416
FSSE	0.023	0.016	0.080	0.074	0.105	0.095	0.039	0.045
RMSE	0.024	0.017	0.084	0.082	0.105	0.095	0.093	0.096
ν	5							
$\hat{\nu}_{sim}$	5.075	5.042	4.867	4.787	5.745	6.046	5.000	5.009
FSSE	0.454	0.331	0.484	0.400	1.815	1.650	0.273	0.254
RMSE	0.460	0.334	0.502	0.453	1.962	1.953	0.273	0.254
σ	1							
$\hat{\sigma}$	0.999	0.998	1.064	1.047	0.691	0.699	1.034	1.032
FSSE	0.040	0.030	0.122	0.110	0.391	0.389	0.056	0.044
RMSE	0.040	0.030	0.138	0.119	0.498	0.492	0.065	0.055

Table C.3: Simulated average, finite sample standand error(FSSE) and root mean squared error(RMSE) for ML, GMM I, GMM II, and MM Estimators. In panel A, the simulated parameters are $m_0 = 1.2, m_1 = 0.5$ $\nu = 4$, and $\sigma = 1$, respectively. MM^*: there are only 146 valid estimations for MM. In pandel B, the simulated parameters are $m_0 = 1.2, m_1 = 0.5$ $\nu = 5$, and $\sigma = 1$, respectively. MM^*: there are only 119 valid estimations for MM.

Trinomial MSMF (k=3) with Student's-t innovation (2)

Pandel C: Parameter ($m_0 = 1.3, m_1 = 0.6$, $\nu = 4$ **and** $\sigma = 1$)								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m_0	1.3							
$\hat{m}_0$	1.277	1.270	1.393	1.358	1.376	1.363	1.573	1.528
FSSE	0.104	0.091	0.226	0.193	0.175	0.156	0.413	0.317
RMSE	0.106	0.095	0.244	0.202	0.191	0.168	0.495	0.391
m_1	0.6							
$\hat{m}_1$	0.615	0.605	0.578	0.580	0.602	0.623	0.441	0.439
FSSE	0.046	0.024	0.114	0.088	0.092	0.066	0.079	0.064
RMSE	0.048	0.024	0.116	0.090	0.092	0.070	0.177	0.173
ν	4							
$\hat{\nu}_{sim}$	3.988	4.002	4.312	4.239	4.459	4.570	4.110	4.097
FSSE	0.283	0.200	0.778	0.636	1.131	1.365	0.108	0.091
RMSE	0.283	0.200	0.848	0.679	1.221	1.479	0.154	0.133
σ	1							
$\hat{\sigma}$	0.996	0.998	1.108	1.077	0.981	1.012	1.084	1.081
FSSE	0.037	0.026	0.140	0.121	0.273	0.192	0.046	0.038
RMSE	0.037	0.026	0.177	0.144	0.274	0.193	0.095	0.089
Pandel D: Parameter ($m_0 = 1.3, m_1 = 0.6$, $\nu = 5$ **and** $\sigma = 1$)								
method	ML		GMM I		GMM II		MM^*	
T	5000	10.000	5000	10.000	5000	10.000	5000	10.000
m_0	1.3							
$\hat{m}_0$	1.277	1.264	1.406	1.373	1.409	1.367	1.592	1.517
FSSE	0.097	0.085	0.221	0.188	0.194	0.162	0.332	0.247
RMSE	0.100	0.092	0.245	0.202	0.222	0.176	0.442	0.329
m_1	0.6							
$\hat{m}_1$	0.613	0.604	0.546	0.552	0.590	0.587	0.531	0.523
FSSE	0.044	0.022	0.101	0.099	0.093	0.091	0.069	0.048
RMSE	0.046	0.022	0.115	0.110	0.093	0.092	0.098	0.091
ν	5							
$\hat{\nu}_{sim}$	5.062	5.028	4.751	4.698	5.237	5.485	5.184	5.149
FSSE	0.442	0.296	0.699	0.525	1.409	1.764	0.335	0.271
RMSE	0.446	0.298	0.742	0.606	1.429	1.829	0.383	0.309
σ	1							
$\hat{\sigma}$	0.999	1.000	1.060	1.053	0.868	0.876	1.050	1.047
FSSE	0.039	0.024	0.096	0.099	0.245	0.247	0.046	0.032
RMSE	0.039	0.024	0.113	0.112	0.278	0.276	0.068	0.057

Table C.4: Simulated average, finite sample standand error(FSSE) and root mean squared error(RMSE) for ML, GMM I, GMM II, and MM Estimators. In panel A, the simulated parameters are $m_0 = 1.3, m_1 = 0.6$ $\nu = 4$ and $\sigma = 1$, respectively. MM^*: there are only 283 valid estimations for MM. In pandel B, the simulated parameters are $m_0 = 1.3, m_1 = 0.6$ $\nu = 5$, and $\sigma = 1$, respectively. MM^*: there are 331 valid estimations for MM.

Bibliography

B. Abraham, N Balakrishna, and R. Sivakumar. Gamma stochastic volatility models. *Journal of Forecasting*, 25:153–171, 2006.

P. Abry, P Baraniuk, P. Fladrin, R. Riede, and D. Veitch. The multiscale nature of netwark traffic: discovery, analysis and modelling. *IEEE Signal Processing Magzine*, 19:28–46, 2002.

J. Armstrong and R Fildes. On the selection of error measures for comparisons among forecasting methods. *Journal of Forecasting*, 14:67–71, 1995.

E. Bacry, J. Delour, and J.F. Muzy. Multifractal random walk. *Physical Review E*, 026103:409–430, 2001.

R.T. Baillie and T. Bollerslev. The message in daily exchange rates: a conditional-variance tale. *Journal of Business & Economic Statistics*, 7:297–305, 1989.

R.T. Baillie, T. Bollerslev, and H.-O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30, 1996.

L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. *Annaly of Mathematical Statistics*, 41:164–171, 1970.

A. Beltratti and C. Morana. Computing value at risk with high frequency data. *Journal of Empirical Finance*, 6:431–455, 1999.

J. Beran. *Statistics for long-memory processes*. Chapman and Hall, New York, 1994.

P.J. Bickel, Y. Ritov, and T. Ryden. Asymptotic normality of the maximum-likelihood estimator for general hidden markov models. *The Annaly of Statistics*, 26:1614–1635, 1998.

T. Bollerslev. Generalized autoregressive conditional heteroscekasticity. *Journal of Econometrics*, 31:307–327, 1986.

T. Bollerslev. A contitionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, 69:542–547, 1987.

T. Bollerslev and H. O. Mikkelsen. Modeling and pricing long memory in stock market volatility. *Jounral of Econometrics*, 73:151–184, 1996.

T. Bollerslev, R. Chou, and K. Kroner. Arch modeling in finance. *Jounral of Econometrics*, 52: 5–59, 1992.

T. Bollerslev, R.F. Engle, and D.B. Nelson. Arch models. In R.F. Engle and D. McFadden, editors, *Handbook of econometrics*, pages –. Elsevier, Amsterdam, 1994.

J.-P. Bouchaud and M. Potters. *Theory of financial risks: from statistical physics to risk management*. Cambridge University Press, Cambridge, 2000.

J.-P. Bouchaud, M. Potters, and M. Meyer. Aparent multifractality in financial time series. *The European Physical Journal B*, 13:595–599, 2000.

W. Breymann, S. Ghashgaie, and P. Talckner. A stochastick cascade model for fx dynamics. *International Journal of Theoretical and Applied Finance*, 3:357–360, 2000.

P.J. Brockwell and R.A. Davis. *Time series: theory and methods*. Springer-Verlag, New York, 1991.

P.J. Brockwll and R. Dahlhaus. Generalized levinson-durbin and burg algorithms. *Journal of Econometrics*, 118:129–149, 2004.

L. Calvet and A. Fisher. Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58, 2001.

L. Calvet and A. Fisher. Multi-fractality in asset returns: Theory and evidence. *Review of Economics & Statistics*, 84:381–406, 2002.

L. Calvet and A. Fisher. How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2:49–83, 2004.

L. Calvet and A. Fisher. Multifrequency news and stock returns. *NBER Working Paper Series*, w11441:1–57, 2005.

L. Calvet, A. Fisher, and B.B. Mandelbrot. Large deviation and the distribution of price changes. *Cowles Foundation Discussion Paper*, 1165, 1997.

L. Calvet, A. Fisher, and S.B. Thompson. Volatility comovement: a multifrequency approach. *Cowles Foundation for Research in Economics*, 131:179–215, 2006.

J.Y Campbell, A.W. Lo, and MacKinlay A.C. *The Econometrics of Finacial Markets*. Princeton University Press, Princeton, 1997.

M. Caporin. Identification of long memory in garch models. *Statistical Methods and Applications*, 12(2):133–151, 2003.

P. Chainais, R. Riede, and P. Abry. On non-scale-invariant infinitely divisible cascades. *IEEE Transactions on Information Theroy*, 51:1063–1083, 2005.

A. Chhabra and R.V. Jensen. Direct determination of the $f(\alpha)$ singularity spectrum. *Physical Review Letters*, 62:1327–1330, 1989.

A. Chhabra, C. Meneveau, R.V. Jensen, and K.R. Sreenivasan. Direct determination of the $f(\alpha)$ singularity spectrum and its application to fully developed turbulence. *Physical Review A*, 40: 5284–5294, 1989.

S. Chib, F. Nardari, and N. Shepard. Markov chain monte carlo methods for stochastic volatility models. *Journal of Econometrics*, 108:291–316, 2002.

C.-F. Chung. Estimating the fractionally integrated garch model. *Academia Sinica*, 2002.

P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–55, 1973. available at http://ideas.repec.org/a/ecm/emetrp/v41y1973i1p135-55.html.

K.A. Clarke. Nonparametric model discrimination in international relations. *Journal of Conflict Resolution*, 47:72–93, 2003. available at http://www.rochester.edu/college/psc/clarke/JCRpub.pdf.

K.A. Clarke. A simple distribution-free test for nonnested hypotheses. *Working paper*, 2005. available at http://www.rochester.edu/college/psc/clarke/SDFT.pdf.

M.P. Clements and H-M. Krolzig. A comparison of the forecast performance of markov-switching and threshold autoregressive models of u.s. gnp. *Econometrics Journal*, 1:C47–C75, 1998.

S.R Cosslet and L.-F. Lee. Serial correlation in latent discret variable models. *Journal of Econometrics*, 27:79–97, 1985.

J. Davidson. Moment and memory properties of linear conditional heteroscedasticity models, and a new model. *Journal of Business & Economic Statistics*, 22:16–29, 2004.

A. Davis, A. Marshak, W. Wiscombe, and R. Cahalan. Multifractal characterizations of nonstationarity and intermittency in geophysical fields: Observed, retrieved, or simulated. *Journal of geophysical research*, 99:8055–8072, 1994.

F.X. Diebold and R.S. Mariano. Comparing predictive accuracy. *Journal of Besiness and Economic Statistics*, 13(3):253–263, 1995.

Z. Ding, C. Granger, and R Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–196, 1993.

R. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom. *Econometrica*, 50:987–1007, 1982.

R. Engle. New frontiers for arch models. *Journal of Applied Econometrics*, 17:425–446, 2002.

R.F. Engle and T Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5(3):1–50, 1986.

C.J.G. Evertz and B.B. Mandelbrot. Multifractal measures(appendix b). In H.-O. Peitgen, H. Jügens, and D.L. Saupe, editors, *Chaos and Fractals*, pages –. Springer, Berlin, 1992.

K.J. Falconer. *Fractal Geometry: Mathematical Foundation and Applications*. John wiley and Sons, New York, 2003.

J. Feder. *Fractals*. Plenum Press, New York, 1988.

A. Fisher, L. Calvet, and B.B. Mandelbort. Multifractality of deutschemark / us dollar exchanges rates. *Cowles Foundation for Research in Economics*, 1165, 1997.

R. Fox and M.S. Taqqu. Large-sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, 14:517–532, 1986.

R. Garcia. Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, 39:763–788, 1998.

R. Gençay, F. Selç, and B. Whitcher. *An introduction to wavelets and other filtering methods in finance and economics*. Academic Press, San Diego, 2002.

J. Geweke and S. Porter-Hudak. The estimation and application of long memory time seires models. *Journal of Time Seires Analysis*, 4:221–237, 1983.

A. Gilbert, W. Willinger, and Feldmann. Scaling analysis of conservative cascades, with application to network traffic. *IEEE Trans. on Information Theory*, 45:971–992, 1999.

J. Gleick. *Chaos: making a new science*. Cardinal, London, 1987.

T. Gneiting and M. Schlather. Stochastic models that separate fractal dimension and hurst effect. *SIAM Reviw*, 46:269–282, 2004.

P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica 9D*, pages 189–208, 1983.

J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 77(2):357–384, 1989.

J.D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, 1994.

B.E. Hansen. The likelihood ratio test under non-standard conditions: testing the markov trend models of gdp. *Journal of Applied Econometrics*, 7:S61–S82, 1992.

L.P. Hansen. Large sample property of generalized method of moments estimators. *Econometrica*, 50:1029–1054, 1982.

D. Harris and L. Matyas. Introduction to the generalized method of moment estimation. In L. Matyas, editor, *Generalized Method of Moment Estimation*, pages –. Cambridge Universtiy Press, Cambridge, 2000.

D. Harte. *Multifractals: Theory and applications*. Chapman and Hall/CRC, Boca Raton, 2001.

A. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, 1989.

T. Higuchi. Approach to an irregular time series on the basis of the fractal theroy. *Physica D*, 31: 277–283, 1988.

T. Higuchi. Relationship between the fractal dimension and the power law index for a time series: a numerical investigation. *Physica D*, 46:254–264, 1990.

G. Iori. Scaling and multiscaling in financial markets. *Economics Working Paper Archive at WUSTL*, 2000.

M.J. Jensen. Using wavelets to obtain a consistent ordinary least squares estimator of the long-memory parameter. *Journal of Forecasting*, 18:17–32, 1999.

L. Kilian and M. Taylor. Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60:85–107, 2003.

Chang-Jin Kim and Charles R Nelson. *State-Space Models with Regime Switching: Classicla and Gibbs-Sampling Approches with Applications*. The MIT Press, Cambridge, 1999.

C.J. Kim. Dynamic linear models with markov-switching. *Journal of Econometrics*, 60:1–22, 1994.

H. M. Krolzig. *Markov-switching vector autoregressions, Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer, Berlin, 1997.

S. Kullback and R.A. Leibler. On information and sufficiency. *Annaly of Mathematical Statistics*, 22:79–86, 1951.

L.D. Landau and E.M. Lifshitz. *Statistical Physics, 3rd Edition Part 1.* Butterworth-Heinemann, Oxford, 1996.

B.G. Leroux. Maximum-likelihood estimation for hidden markov models. *Stochastic Processes and their Applications*, 40:127–143, 1992.

R. Liesefeld and R.C. Jung. Stochatic volatility models: conditional normality versus heavey-tailed distributions. *Journal of Applied Econometrics*, 15(2):137–160, 2000.

R. Liesefeld and J.F. Richard. Improving mcmc using efficient important sampling. *Economics Working Paper 2006-05, University of Kiel*, 2006.

F. Lillo and J.D. Farmer. The long memory of the efficient market. *Studies in Nonlinear Dynamics & Econometrics*, 8(3):1126–1126, 2004.

G. Lindgren. Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*, 5:81–91, 1978.

R. Liu and T. Lux. Long memory in financial time series: estimation of bivariate multi-fractal model and its application for value-at-risk. *Economic Working paper, University of Kiel*, 2005.

T. Lux. Multi-fractal process as models for financial returns: a first assessment. *Workingpaper , University of Bonn*, B-456, 1999.

T. Lux. Turbulence in financial markets: the surprising explanatory power of simple cascade models. *Quantitative Finance*, 1:632–640, 2000.

T. Lux. The multi-fractal model of asset returns: Its estimation via gmm and its use for volatility forecasting. *Economics Working paper*, 2003-13, 2003.

T. Lux. The markov-switching multi-fractal model of asset returns: Gmm estimation and linear forecasting of volatility. *Economics Working paper*, 2004-11, 2004.

T. Lux. The markov-switching multi-fractal model of asset returns: Gmm estimation and linear forecasting of volatility. *Economics Working paper*, 2006-17, 2006.

S. Mallat. A theory for multiresolution signal decomposition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:674–693, 1989.

B.B Mandelbrot. The variantion of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

B.B Mandelbrot. Intermittent turbulence in self similar cascades : divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62:331 – 358, 1974.

B.B. Mandelbrot. *The Fractal Geometry of Nature.* W.H. Freeman and Company, San Francisco, 1983.

B.B. Mandelbrot. Multifractal measures, especially for the geophysicist. *Pure and Applied Geophysics*, 131:5–42, 1989.

B.B. Mandelbrot. *Fractaly, Scaling in Finance: Discontinuity, Concentration, Risk.* Springer, New York, 1997.

B.B. Mandelbrot and J.W. Van Ness. Fractional brownian motion, fractional noises and applications. *SIAM Review*, 10(4):422–437, oct 1968.

B.B. Mandelbrot, L. Calvet, and A. Fisher. A multifractal model of asset returns. *Cowles Foundation for Research in Economics*, 1164, 1997.

A. Marshak, A. Davis, R. Cahalan, and W. Wiscombe. Bounded cascade models as nonstationary multifractals. *Physical Review E*, 49:55–69, 1994.

R.C. Mittelhammer, G.G. Judge, and D.J. Miller. *Econometric Foundations*. Cambridge Universtiy Press, Cambridge, 2000.

P.A. Morettin. The levinon algorithms and its application in time series analysis. *International Statistical Review*, 52:83–92, 1984.

U. A. Müller, M. M. Dacorogna, R. D. Dave, O.V. Pictet, B. Olsen, and J.R. Ward. Fractals and intrinsic time - a challenge to econometricians. *Olsen and Associates preprint*, 1995.

K. P. N. Murthy, K. W. Kehr, and A. Giacometti. Multifractal scaling of moments of mean first-passage time in the presence of sinai disorder. *Phys. Rev. E*, 53(1):444–449, Jan 1996. doi: 10.1103/PhysRevE.53.444.

J.F. Muzy, E. Bacry, and A. Arneodo. The multifractal formalisim revisited with wavelets. *International Journal of Bifurcation and Chaos*, 4(2):245–302, 1994.

J.F. Muzy, E. Bacry, and A. Kozhemyak. Exterme values and fat tails of multi-fractal fluctuations. *Physical Review E*, 73(066114), 2006.

Daniel B. Nelson. Arch models as diffusion approximations. *Journal of Econometrics*, 45:7–38, 1990.

W.K. Newey and K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708, 1987.

W.K. Newey and K.D. West. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, 61(4):631–653, 1994.

A. Pagan. The econometrics of financial markets. *Journal of Empirical Finance*, 3:15–102, 1996.

F.C. Palm and P.J.G. Vlaar. Inflation differentials and excess returns in the european monetary system. *Journal of International Financial Markets, Institutions and Money*, 7:1–20, 1997.

T.S. Parker and L.O. Chua. *Practical numerical algorithms for chaotic systems*. Springer-Verlag, 1989.

C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger. Mosaic organization of dna nucleotides. *Phys. Rev. E*, 49(2):1685–1689, Feb 1994. doi: 10.1103/PhysRevE.49.1685.

E. Peters. *Fractal market analysis: applying chaos theroy to investment and economics*. John Wiley & Sons, 1994.

M.B. Priestley. *Spectral analysis and time series*. Academic press, London, 1981.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

P.M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47:67–84, 1991.

E. Ruiz. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, 63:289–306, 1994.

G. Sandman and S. Koopman. Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics*, 67:271 – 301, 1998.

D. Schertzer and S. Lovejoy. Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92:9693 – 9714, 1987.

G Schwert. Estimation the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

W.G. Schwert. Why does stock market volatility change over time ? *Journal of Finance*, 44: 1125–1153, 1989a.

W.G. Schwert. Stock volatility and the crash of '87. *Review of Financial Studies*, 3:77–102, 1989b.

C.E. Shannon. A mathematical theroy of communication. *The Bell System Technical Journal*, 27: 379–423, 623–656, 1948.

A. Smith, P.A. Naik, and C.L. Tsai. Markov-switching model selection using kullbak-leibler divergence. *Journal of Econometrics*, 134(2):553–577, 2006.

M. S. Taqqu, V. Teverovsky, and W. Willinger. Estimators for long-range dependence: an empirical study. *Fractals*, 3:785–798, 1995.

D.L. Turcotte. *Fractals and chaos in geology and geophysics*. Cambridge Universtiy Press, Cambridge, 1997.

Q.H. Vuong. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57:307–333, 1989.

Diethelm Wuertz, many others, and see the SOURCE file. *fSeries: Rmetrics - The Dynamical Process Behind Markets*, 2006. URL `http://www.rmetrics.org`. R package version 240.10068.

Eidesstättliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich meine Doktorarbeit ,,The Markov Switching Multi-Fractal model of Asset Returns: Estimation and Forecasting of Dynamic Volatility with Multinomial Specifications" selbständig und ohne fremde Hilfe angefertigt habe und dass ich alle von anderen Autoren wörtlich übernommenen Stellen, wie auch die sich an die Gedanken anderer Autoren eng anlehnenden Ausführungen meiner Arbeit, besonders gekennzeichnet und die Quellen nach den mir angegebenen Richtlinien zitiert habe.

Kiel, 25.07. 2007 (Hwa-Taek Lee)