Multivariate Multifractal Models: Estimation of Parameters and Applications to Risk Management

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

vorgelegt von

MBA Ruipeng Liu
aus V.R. China, geb. 13 Juni 1977

Melbourne, September 2008

Gedruckt 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: 29. September 2008
Tag der mündlichen Prüfung: 18. November 2008

ACKNOWLEDGEMENTS

First of all I would like to express my sincere gratitude to Professor Dr. Thomas Lux, who brought me into the world of fractals, and shared with me his expertise and research insight, which have been invaluable to me. I also wish to express my appreciation to Professor Dr. Roman Liesenfeld, who made many valuable suggestions and gave constructive advice, which helped improve of this thesis. I particularly acknowledge the encouragement from Professor Dr. Stefan Mittnik during my starting stage of my Ph.D in the University of Kiel.

I am tempted to individually thank all of my friends which, have joined me in the discovery of what is life about and how to make the best of it. However, because the list might be too long and by fear of leaving someone out, I will simply say *thank you very much to you all.*

I cannot finish without saying how grateful I am with my entire extended family, particular thanks, of course, to my parents, who bore me, raised me, supported me, and loved me. Last but not least, to my wife and my son. To them I dedicate this thesis.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Numerous empirical studies have found non-normal moments in the (unconditional) distributions of returns in financial markets. One early example is Fama (1965), who reports evidence of excess kurtosis in the unconditional distribution of daily returns of Dow Jones listed stocks, there are also more recent studies such as those by Ding et al. (1993), Pagan (1996), Guillaume et al. (2000), Cont (2001), Lo and MacKinlay (2001) and so on.

More traces have revealed the so-called abnormal phenomena in the distribution of financial time series (though some of them are viewed as stylized facts of financial markets):

(1) Fat tail – the distribution is 'fatter' at the extremes ('tails') than a truly Normal one. The fat tail property implies that the unconditional distribution of price changes (returns) has more probability mass in the tails and the center than the standard Normal distribution. This also means that extreme changes occur more often than would be expected under the assumption of Normality of relative price changes.

(2) Volatility clustering, which means that periods of quiescence and turbulence tend to cluster together. Hence, the volatility (conditional variance) of the financial market is not homogeneous over time, but is itself subject to temporal variation.

(3) Long-term dependence (long-range dependence or long memory), which describes the correlation structure of a time series at long lags. If a series exhibits long memory,

there is persistent temporal dependence even between distant observations. Such series are characterized by distinct but nonperiodic cyclical patterns. The presence of long memory in returns would provide evidence against the efficient markets hypotheses (EMH) since long memory implies nonlinear dependence and hence a potentially predictable component in the series dynamics. It would also raise issues regarding linear modeling, statistical testing and forecasting of pricing models based on traditional standard statistical methods, such as the capital asset pricing model (CAPM). However, long memory is typically found in higher moments only, e.g. squared return, which is often viewed as a proxy of financial volatility, therefore, the ability of replicating long-term dependence in return volatility becomes one of the important criteria for financial modelling assessments.

In addition, other features in financial time series are also recognized, e.g., (multi)scaling, leverage effect and volume and volatility correlation, see Weron et al. (1999), Lux and Marchesi (1999), Lux and Marchesi (2000), Black (1976), Bouchaud and Potters (2001), and Lobato and Velasco (2000). Consequently, various approaches based on a wide range of theoretical analyses, such as non-linear dynamic theory, behavioural finance and agent-based models have provided further insights into the mechanisms of the financial markets. See, Kirman (1993), Lux (1995), Brock and Hommes (1997), Lux (1998), Chiarella and He (2001) and Iori (2002).

An important contribution was made by Benoit Mandelbrot, the famous father of fractals who proposed a multifractal model of asset returns (MMAR), a theory which inherits all the hallmarks of Mandelbrot's earlier work that has emerged since the 1960s. A fractal is an irregular or fragmented geometric shape that can be subdivided into parts, each of which is, at least approximately and qualitatively, a reduced copy of the whole. Natural fractals are characterized by their so called fractal dimension depending on the scale, and a multifractal is composed of many fractals with different fractal dimensions.

The major innovation of the MMAR model is the use of a multifractal cascade as a transformation of chronological time into 'business time', with the multifractal component interpreted as a process governing local instantaneous volatility. Adapting multifractal (MF) processes to financial economics presents major challenges for researchers working in this area. The original MF processes in statistical physics are combinatorial operations on probability measures, which suffer from non-stationarity due to being limited to a bounded interval and to non-convergent moments in the continuous time limit. A more convenient model for financial data would have to be cast in the format of a stochastic process. This task has been achieved by introducing iterative versions of multifractal models, which can be viewed as new types of stochastic models for the volatility dynamics of asset prices (commodity or stock prices, foreign exchange rates and so on). These preserve the hierarchical multiplicative structure of volatility components with different characteristic time scales, and their time series are in harmony with the stylized facts of financial time series, such as fat/thin tails, excess kurtosis and volatility clustering. Particularly, the essential new feature of MF models is their ability to generate different degrees of long-term dependence in various powers of returns – a feature that pervades empirical financial data, and as a result, they have been introduced as a competitive alternative to long memory models; cf. Calvet and Fisher (2001, 2004), Lux (2007).

Current MF studies are mainly confined to univariate models; however, for many important questions, a multivariate setting would be preferable. Some of the reasons include:

(1) It has been well accepted that financial markets or assets are correlated, and their volatilities move together over time across markets and assets. It is this correlation that gives rise to the financial uncertainties that can subsequently result in worldwide financial disasters. So, this is particularly important when considering asset allocation, Value-at-Risk and portfolio hedging strategies.

(2) The literature has been relatively silent on the source of long-range dependence

within volatility processes; multivariate (bivariate) models may provide additional information about the factors responsible for long memory. For example, by taking account of the observed strong relationship between stock price volatility and trading volume, bivariate mixture models assume volatility and trading volume are jointly directed by the latent number of information arrivals based on the mixture distribution hypothesis, see Andersen (1996), Liesenfeld (1998) and Liesenfeld (2001). This implies that the dynamics of both variables are restricted to depend only on the behavior of the information arrival process. Therefore, trading volume may provide information about the factors which generate the volatility persistence.

In this thesis, we concentrate on developing multivariate multifractal models, and implement their estimation via different approaches. We then apply our new multivariate MF processes to empirically relevant topics, such as portfolio risk measurement and management.

The rest of the thesis is organized as follows:

Chapter 2 of this thesis provides a background and introduction to fractals and multifractals, including the definition of fractals and fractal geometry. Since fractals imply long memory, some traditional methods of fractal analysis and well-known long memory models are reviewed. They include, the Rescaled Range analysis (R/S), Fractional Brownian motion, the Fractional Integrated Autoregressive Moving Average (ARFIMA) model, the Fractional Integrated Autoregressive Conditional Heteroskedasticity (FIGARCH) model, and the long memory stochastic volatility (LMSV) model. In the last section of the chapter, the multifractal model of asset returns (MMAR) by Mandelbrot et al. (1997) is revisited, together with earlier estimation attempts and empirical studies.

The first section of Chapter 3 gives a review of one iterative MF version – the Markov switching multifractal (MSM) model. By discarding the combinatorial nature, the MSM

model preserves the multifractal and stochastic properties, and it conceives the local instantaneous volatility as the product of volatility components. As a simple alternative to the recently proposed bivariate MF model by Calvet et al. (2006), an alternative bivariate multifractal model with a more parsimonious setting is introduced. We estimate both bivariate models with a maximum likelihood approach and simulation-based inference via particle filter algorithm, which are discussed in Calvet et al. (2006). To overcome the computational limit of the likelihood approach (particularly in the case of models with continuous distribution of volatility components), we adopt the GMM (Generalized Method of Moments) approach formalized by Hansen (1982), which for its implementation requires analytical moment conditions. Monte Carlo experiments are undertaken to examine the performance of our GMM estimator. These results demonstrate the capability of GMM and its advantage when increasing the number of cascade levels, particularly for the continuous (Lognormal) bivariate MF model. In the last section of this chapter, we present empirical studies of bivariate multifractal models for different types of financial data, including stock exchange indices, foreign exchange rates and government bond maturity rates. By using three estimation approaches (ML, simulation based ML and GMM), empirical estimates of multifractal parameters are reported with various cascade levels. We also provide a heuristic method for specifying the number of cascade levels, by matching the simulated long memory index with the empirical one.

Chapter 4 focuses on higher-dimensional multifractal models. In this chapter we present Monte Carlo results and empirical estimates via the three different approaches introduced in Chapter 3; GMM does not suffer from the very high dimensionality because it allows each pair of time series to be treated as a bivariate case. The maximum likelihood approach again encounters the restriction of an upper limit to the number of cascade levels; a simulation based inference is implemented. We demonstrate the advantage of GMM by comparing estimators in terms of the efficiency and applicability when increasing the dimensionality

of the multifractal model. The last section of Chapter 4 is an empirical study of tri-variate MF models. We also report ML, SML and GMM estimates by considering a portfolio of three foreign exchange rates.

Chapter 5 is an empirical study of the multivariate multifractal models on financial risk measurement and management. Based on the number of cascade levels specified in the last section of Chapter 3, we re-estimate our bivariate MF model as well as the model by Calvet et al. (2006) using in-sample data, we then conduct out-sample forecast assessments by employing two popular tools, namely, Value-at-Risk and Expected Shortfall. To demonstrate the applicability of multivariate multifractal processes, we also compare their forecast performances with a benchmark multivariate GARCH model.

Chapter 6 contains additional summaries of results and conclusions from our study, and some suggestions for worthwhile future work in the field of multifractal models.

The Appendix at the end of this thesis provides the analytical moment conditions for our Binomial and Lognormal bivariate MF models, followed by the moment conditions for the bivariate MF Model of Calvet et al. (2006). In addition, some numerical solutions of the moments of the log of absolute (bivariate) Normal variates which are used within GMM estimation are presented.

Chapter 2

Fractal and Multifractal models

2.1 Introduction to fractals

The Gaussian hypothesis was questioned and seriously criticized by the famous mathematician Benoit Mandelbrot (e.g. Mandelbrot (1963)), who advanced the hypothesis that price change distributions are stable Paretian, or what is called Lévy stable. The hypothesis was subsequently supported by empirical evidence in Mandelbrot (1967b) with time series of commodity markets (cotton and wheat price variations), railroad stock prices and some financial rates (interest rate on call money, Dollar-Sterling exchange rate), see Cootner (1964) and Shiller (1989) for further discussions. There are more empirical studies on the stable Paretian Hypothesis of asset returns and detailed description of stable Paretian models in finance, cf. Upton and Shannon (1979), Lux (1996b), and Rachev and Mittnik (2000).

The definition of stable Paretians implies that the distribution of sums of independent, identically-distributed stable Paretical variables is itself stable Paretian. Fractals somehow share this 'stable' property. The term *fractal* first appeared in Mandelbrot's papers, and it was defined together with his arguments on traditional geometry, in "The idea of the fractal dimension" as part of the November 1975 issue of Scientific American, and then republished in "The Fractal Geometry of Nature, Chapter 1":

Why is geometry often described as cold and dry? One reason lies in its inability to

describe the shape of a cloud, a mountain, a coastline, a tree. Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line. Nature exhibits not simply a higher degree but an altogether different level of complexity. The number of distinct scales of length of patterns is for all purposes infinite. The existence of these patterns challenges us to study these forms that Euclid leaves aside as being 'formless', to investigate the morphology of the 'amorphous'. Mathematicians have disdained this challenge, however, and have increasingly chosen to flee from nature by devising theories unrelated to anything we can see or feel.

2.1.1 Self-similarity and fractals

A random process $X(t)$ that satisfies:

$$X(ct_1), \ldots, X(ct_n) \simeq c^H X(t_1), \ldots, c^H X(t_n), \tag{2.1.1}$$

(where $\simeq$ represents 'equal in terms of distribution') for $c, n, t_1, \ldots, t_n \geq 0$, is called self-similar with the self-similarity parameter H, which will often be called the scaling exponent or Hurst exponent in the following sections.

Therefore, the term 'self-similarity' tells us that the partial processes look qualitatively the same irrespective of their size (strictly, it does not mean exactly the same, but similar in terms of the distribution's properties), where c^H denotes the scaling property (power law). In this way, self-similarity implies a scaling that describes a specific relationship between data samples of different time scales. Mandelbrot (1963) presents some earlier empirical evidence graphically, by examining cotton price changes in terms of different time periods (1880-1958) and different time frequencies, which are revisited in Mandelbrot (2001).

Other plausible definition of self-similarity focuses on the increment of $X(t)$:

$$X(t + c\triangle t) - X(t) \simeq M(c)[X(t + \triangle t) - X(t)],$$

$M(c)$ is a random variable whose distribution does not depend on t but c, and it is called

the scaling factor. Nevertheless, self-similarity is the key characteristic property of fractals, and since a fractal is an object in which individual parts are similar to the whole, it is a particularly interesting class of self-similar objects.

2.1.2 Principles of fractal geometry

Fractal theory has generated a great deal of interest, which has been fuelled by the intuitive observation that the world around us is filled with fractal shapes showing the property of self-similarity.

A fractal can be constructed by taking a generating rule and iterating it over and over again: one example is the branching network in a tree; each branching is different, as are the successive smaller ones, but all are qualitatively similar to the structure of the whole tree. Other examples commonly used to illustrate this idea include the triadic Koch curve, Sierpinski Triangle, Cantor Dust, and so on.

One way to describe a fractal shape is to calculate its fractal dimension, a number that quantitatively describes how an object fills its space. Traditional Euclidean geometry applies to objects which are solid and continuous, have no holes or gaps, and have a dimensionality that is an integer. According to Euclidean geometry, we have a clear understanding of dimensionality; that is, a line has dimension one , and a plane and a space have dimensions two and three respectively.

Fractals are rough and often discontinuous, like the Sierpinski Triangle, and so have fractional, or fractal dimensions. One way of calculating fractal dimension uses rulers of varying lengths. For the example of a coast line, one can count the number of circles with a certain diameter that will cover the coastline, then decrease (or increase) the diameter for another round, and again count the number of circles, and so on $\cdots$. The fractal dimension D can be revealed by the relationship between the radius $1/r$ of the circles or other rulers,

and the number of units (length, area, or volume) N according to:

$$N = r^D, \tag{2.1.2}$$

which can be transformed to:

$$D = \frac{\log N}{\log r}. \tag{2.1.3}$$

So, the fractal dimension can be calculated by taking the limit of the quotient of the log change in object size and the log change in measurement scale, as the measurement scale approaches zero. We can take some simple and straightforward examples, the Koch Curve (Figure 2.1) and the Sierpinski Triangle (Figure 2.2).

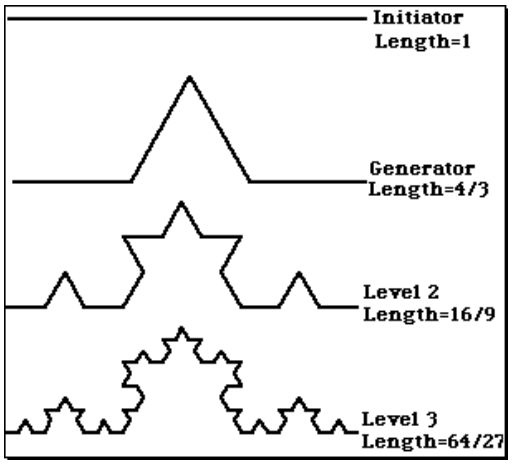

Figure 2.1: Koch curve.

Figure 2.1 shows the realizations of the Koch curve discovered by Helge von Koch (1870 - 1924). Starting from a straight line of length 1, we then divide it into three subsets of equal length, and replace the middle part of the line with two lines of the same length (1/3) like the remaining lines on each side, which can be seen in the second panel of Figure 2.1. Thus, we have four sub-lines of equal length and the original line of length 1 becomes a jagged line of length 4/3. Applying this rule again for each of these four sub-lines, we reach

the third level of Figure 2.1; repeating the same rule once again leads to the the fourth level of Figure 2.1. One can see that each level is 4/3 times longer than the previous one, therefore, the fractal dimension obtained from $D = \frac{\log 4}{\log 3} = 1.26$. Similarly, the Sierpinski Triangle is constructed by iterating equilateral triangles inwardly (Figure 2.2); consequently doubling the length of the sides gives us three times the number of triangles, and it has dimension $\frac{\log 3}{\log 2} = 1.585$.

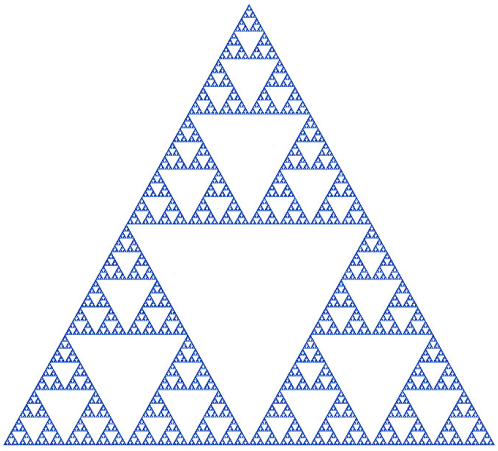

Figure 2.2: Sierpinski Triangle.

Eq. (2.1.3) can also be applied to Euclidean objects by reducing their linear sizes by 1/r in each spatial direction; so that, for a straight line, for example, scaling it by a factor of two, so that it is now twice as long as before, gives $\frac{\log 2}{\log 2} = 1$, a dimensionality of 1; for a square figure, scale up the square by a factor of two, and the square is now 4 times as large (i.e. 4 original squares can be placed on the original square) and $\frac{\log 4}{\log 2} = 2$, gives a dimension of 2; likewise, for a cube $\frac{\log 8}{\log 2} = 3$, (see Figure 2.3).

The determination of fractal dimensions above belongs to the category of Hausdorff dimensions,[1] which are more general algorithms used by mathematicians and physicists.

[1]Some mathematics sources refer to them as Minkowski-Bouligand dimensions.

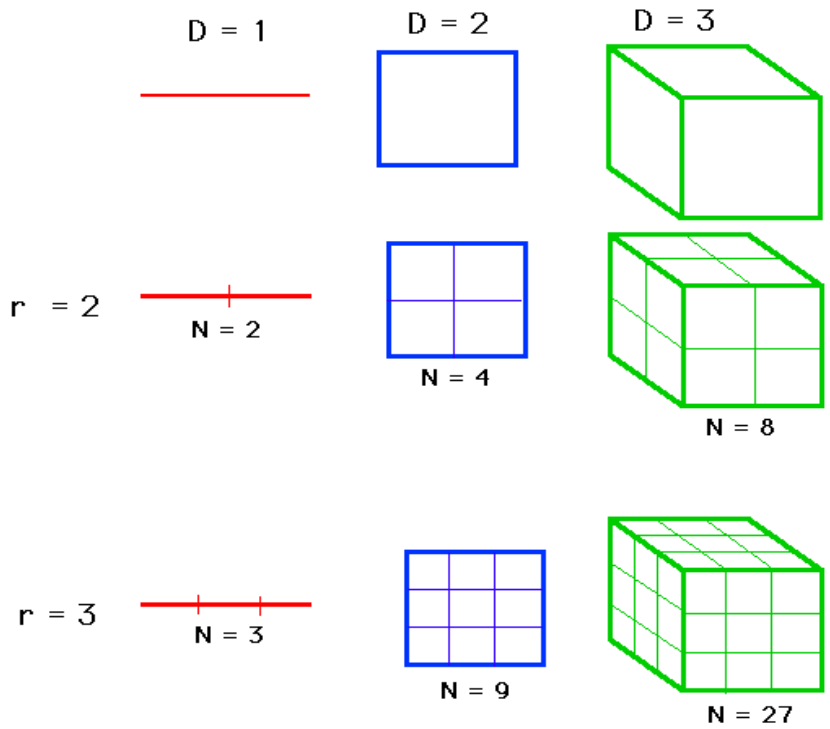

Figure 2.3: Fractal dimension for Euclidean objects.

Though various definitions somehow generate different results, the differences are due to what exactly is meant by 'object size', 'measurement scale' and how to get an average number out of many different parts of a geometrical object.[2]

Mandelbrot (1997) proposed a further tentative approach to computing the fractal dimension, by using the Hurst or Hölder exponent (which as a continuum of local scaling factors replaces the unique Hurst exponent of uni-fractal processes such as fractional Brownian motion). As an alternative measurement, the Hurst exponent H and the fractal dimension are related as:

$$D = 2 - H. \tag{2.1.4}$$

In the next sections, we focus on the Hurst exponent H by introducing various methods

[2]Mandelbrot (1967a) pointed out that it is actually difficult to measure the length of a coast of Britain explicitly since it is jagged, and it depends on the length of the rulers that one uses.

which are used to calculating H.

2.1.3 Rescaled range analysis

The Hurst exponent is named after the well-known hydrologist, Harold E. Hurst(1900 - 1978), who worked on projects involving the design of reservoirs along the Nile River water storage using a statistical method – Rescaled range analysis (R/S analysis) in 1950s. It was firstly introduced in Hurst (1951), later refined in Mandelbrot and Wallis (1969b), as well as in Mandelbrot and Taqqu (1979). These became popular in finance due to the clear exposition of the methods in Feder (1988) and the empirical work of Peters (1994).

The widely used R/S analysis is based on a heuristic graphical approach, whereby the basic idea is to compare the minimum and maximum values of running sums of deviations from the sample mean, then re-normalized by the sample standard deviation. The following is a summary of Hurst's R/S statistic:

(1) Divide the N-length of time series Y_t into $A \times n$ sub periods with length n; each sub-period is denoted N_a, $a = 1, \cdots A$, for each observation in N_a is labeled $y_{k,a}$, $k = 1, \cdots, n;$, then calculate the mean value for each N_a of length n, that is

$$M_a = \frac{1}{n} \sum_{k=1}^{n} y_{k,a}. \tag{2.1.5}$$

(2) The time series of accumulated departures ($y_{k,a}$) from the mean value for each sub-period N_a is given by:

$$X_{k,a} = \sum_{i=1}^{k} (y_{i,a} - M_a), \qquad k = 1, 2, \cdots, n; \tag{2.1.6}$$

(3) The range (R) is defined as the maximum minus the minimum value of $X_{k,a}$ within each sub-period N_a:

$$R_{N_a} = \max_{1 \le k \le n} (X_{k,a}) - \min_{1 \le k \le n} (X_{k,a}); \qquad k = 1, 2, \cdots, n; \tag{2.1.7}$$

(4) The sample standard deviation is calculated for each sub-period:

$$S_{N_a} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_{k,a} - M_a)^2}.$$

Each range R_{N_a} is then normalized by S_{N_a}, so we have the rescaled range for each N_a sub-period, which is equal to R_{N_a}/S_{N_a}. The calculations from the second step must be repeated for different time horizons, and we have A sub-periods with equal length n. The average R/S value for each n is defined as:

$$(R/S)_n = \frac{1}{A}\sum_{a=1}^{A}\frac{R_{N_a}}{S_{N_a}}. \tag{2.1.8}$$

Applying Hurst's Empirical Law for different values of n:

$$\log(R/S)_n = C + H \cdot \log(n), \tag{2.1.9}$$

and running an ordinary least square (OLS) regression we get the Hurst exponent H; the intercept is the estimator for the constant C.

Figure 2.4 shows the graphical result for returns of U.S. Dollar to British Pound exchange rate (March 1973 to February 2004), with a Hurst exponent estimator of $H = 0.59$, by applying Hurst's empirical law (Eq. 2.1.9) via the OLS regression.

Using the classical R/S approach, Peters (1994) finds that stock and bond markets follow a biased random walk, indicating that the information ripples forward in time, instead of being reflected immediately in prices. This would be interpreted as an evidence against the efficient market hypothesis (EMH). However, some later studies indicate that financial asset returns do not significantly deviate from $H = 0.5$, see Mills (1993), Crato and De Lima (1994) and Lux (1996a). Other earlier applications of R/S analysis include Mandelbrot and Wallis (1969a), Mandelbrot and Wallis (1969b), as well as some empirical studies: Greene and Fielitz (1977) (stock returns data for securities listed on the New York Stock Exchange); Booth and Kaen (1979) (Gold prices); Booth et al. (1982) (foreign exchange rates); Helms et al. (1995) (commodity futures contracts).

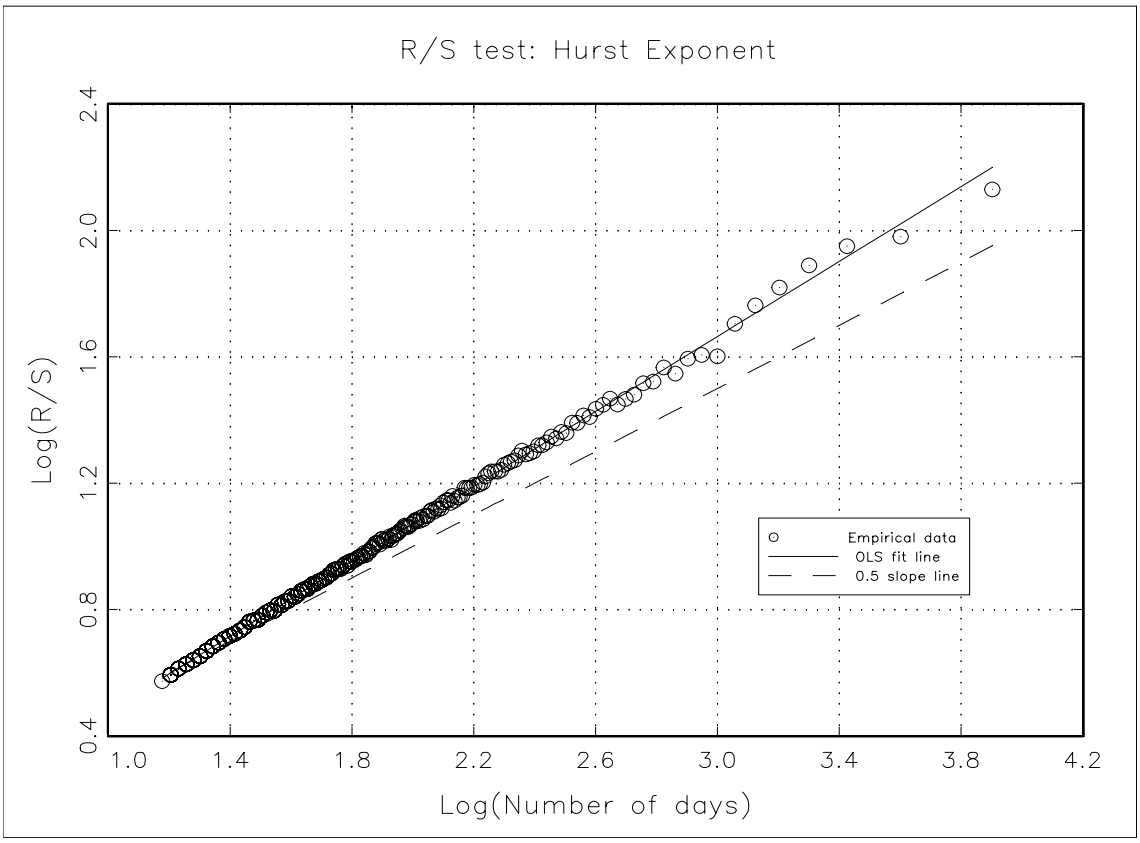

Figure 2.4: Hurst's R/S analysis for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004).

Modified R/S test

The less attractive features of the classical R/S analysis, as discussed in a number of empirical studies, are its sensitivity to the presence of short-range autocorrelation, which exists especially in the high frequency data analysis; more specifically, the classical R/S test tends to indicate that a time series has long memory when it actually does not. In fact, Lo (1991) has shown that, even for a short-memory process, such as a simple AR(1) process, the classical R/S test does not reject the null hypothesis of short-term dependence; similar results also have been presented by Davies and Harte (1987). Another weakness is the lack of a distribution theory for the underlying statistics of Eq. (2.1.8). Therefore, the Hurst

exponent derived by the classical R/S analysis must be interpreted with some caution.

These facts motivated further rigorous tests to detect long memory properties exhibited in financial time series. So, Lo (1991) proposed a modified R/S statistic that is obtained by replacing the denominator S_{N_a} in Eq. (2.1.8), (the sample standard deviation), by a consistent estimator of the square root of the variance of the partial sum. The motivation for this modification is that, in the case of dependent random variables, the variance of the partial sum is not simply the sum of the variances of the individual $y_{k,a}$, but also includes their auto-covariances up to some lag q. The modified R/S by Lo (1991) uses the Newey-West style variance estimator:

$$S(q) = \left\{ \frac{1}{n} \sum_{k=1}^{n} (y_{k,a} - M_a)^2 + \frac{2}{n} \sum_{j=1}^{q} \omega(q) \sum_{k=1}^{n} (y_{k,a} - M_a)(y_{k-j,a} - M_a) \right\}^{1/2}, \qquad (2.1.10)$$

with weights defined as

$$\omega_j(q) = 1 - \frac{j}{q+1}.$$

This autocovariance part of the denominator $S_{N_a}(q)$ is non-zero for data exhibiting short-term dependence and this makes the statistics robust to heteroscedasticity.

Lo (1991) standardizes the statistic $R/S(q)$ by introducing a modified R/S statistic $V_q(n)$:[3]

$$V_q(n) = \frac{1}{\sqrt{n}} \cdot [(R/S(q)]_n. \qquad (2.1.11)$$

The distribution function of $V_q(n)$ is explicitly given as

$$F_V = 1 + 2 \sum_{\alpha=1}^{\infty} (1 - 4\alpha^2 V^2) e^{-2\alpha^2 V^2}. \qquad (2.1.12)$$

The fractiles of the distribution of $V_q(n)$ are given in Lo (1991):

$$\lim_{n \to \infty} Prob \{V_q(n) \in [0.809, \quad 1.862]\} = 0.95,$$

[3]For $q = 0$, it is the classical R/S statistic.

which gives the 95 percent confidence interval (more can be found in the Table II of Lo (1991)), when testing the null hypothesis that there is no long memory in the given time series; specifically, if $V_q(n)$ lies within this interval of [0.809, 1.862], one infers that the data are not characterized by long-term dependence.

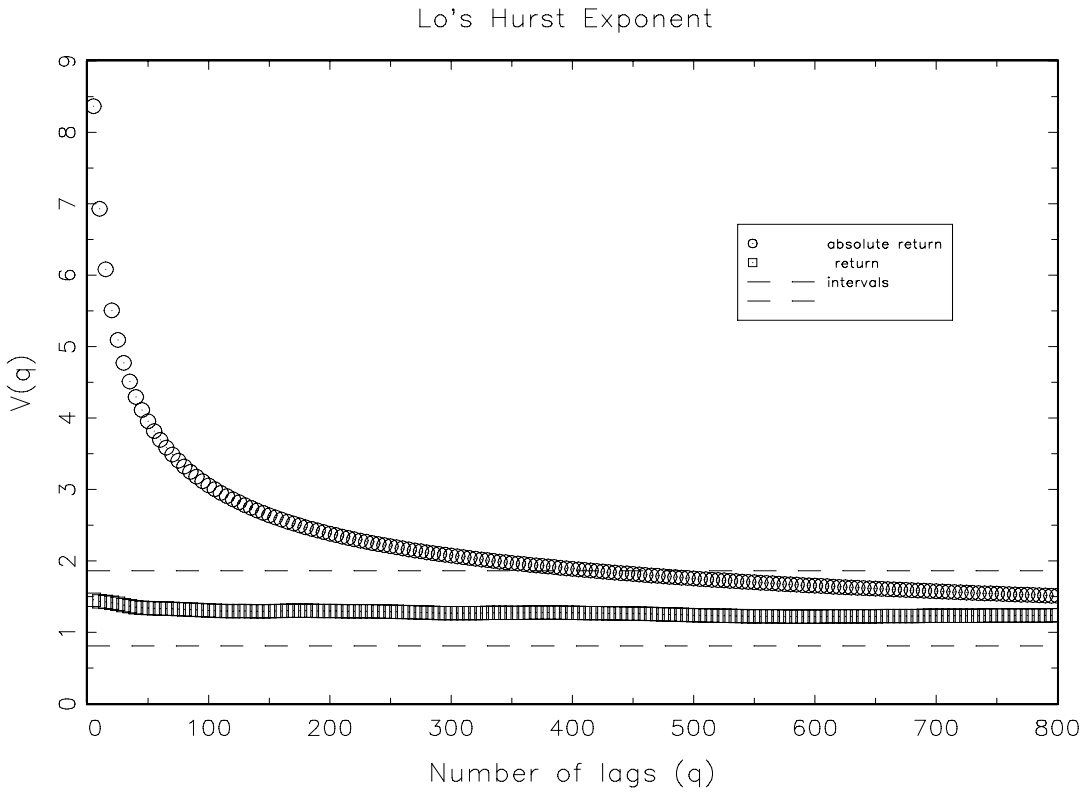

Figure 2.5: Lo's R/S for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004).

Using the empirical data of U.S Dollar to British Pound exchange rate, Figure 2.5 shows empirical results (for returns and the absolute value of returns of U.S. Dollar to British Pound exchange rate) of Lo's modified R/S statistic $V_q(n)$ plotted for different values of the truncation lag q; the two dashed horizontal lines in the figure represent the interval $[0.809, 1.862]$, revealing that there are some values of the truncation lag (approximately below 320) for which the statistic lies outside the region (for the absolute value of return

data). According to Lo's rule, the test accepts the hypothesis of long memory up to q, but the opposite inference would be made for a larger value of q. While, for the raw return time series, it fails to reject the null hypothesis of no long memory for all truncation lags q.

Lo's method represents a theoretical improvement over the classic Rescaled range statistic, but its practical application requires care. There have been long-standing debates on this modified R/S version. Moody and Wu (1996) further propose the so-called robust R/S estimates, by defining the denominator as ($\sigma(n)$ is the standard deviation):

$$S(q) = \left\{ \left[1 + 2\sum_{j=1}^{q} \omega_j(q) \frac{n-j}{n^2} \right] \sigma^2(n) + \frac{2}{n} \sum_{j=1}^{q} \omega(q) \sum_{k=1}^{n} (y_{k,a} -, M_a)(y_{k-j,a} - M_a) \right\}^{1/2}. \tag{2.1.13}$$

Willinger et al. (1999) argue that Lo's R/S statistic shows strong preference for accepting the null hypothesis: no long-term dependence, irrespective of whether long memory is present in the empirical data or not, which implies the evidence is not absolutely conclusive; Teverovsky et al. (1999) perform Monte Carlo simulations with long-range and short-range dependent time series, and find the selection of q is critical. These studies strongly advise against its use as the sole technique for testing long-term dependence in empirical data.

KPSS test

Kwiatkowski et al. (1992) proposed a so-called Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test which is based on the second moment of X_k (defined as Eq. (2.1.6) in the second step of Hurst's R/S analysis). the KPSS test is then defined as

$$KPSS_q(n) = \frac{1}{n^2 \cdot S(q)^2} \sum_{k=1}^{n} (X_k)^2. \tag{2.1.14}$$

Originally, KPSS proposed a test of the null hypothesis of stationarity. However, the distribution theory under the null hypothesis assumes that the series in question has short

memory; that is, its partial sum satisfies an invariance principle. Lee and Schmidt (1996) shows that the KPSS test is consistent against stationary long memory alternatives, such as $I(d)$ processes for $d \in (-0.5, 0.5)$, $d \neq 0$. It has been shown that under the null null hypothesis of $I(0)$, this statistic asymptotically converges to a well-defined random variable

$$KPSS_q(n) \to \int_0^1 V(t)^2,$$

where $V(t)$ is the so-called Brownian bridge (see Lee and Schmidt (1996) for details).[4] It can therefore be used to distinguish short memory and long memory stationary processes. The power of the KPSS test in finite samples is found to be comparable to that of Lo's modified rescaled range test. Similar to Hurst's empirical law of Eq. (2.1.9), Giraitis et al. (2000) extend the "pox-plot" analysis to the KPSS statistics for the estimation of the Hurst exponent by claiming the following regression:

$$\log KPSS_q(n) = \alpha + \beta \cdot \log(n), \tag{2.1.15}$$

and the Hurst exponent is then obtained via $H = 0.5 + 0.5 \times \beta$. An empirical study of U.S. Dollar to British Pound exchange rate (March 1973 to February 2004) is presented in Table 2.1. Base on the KPSS statistic provided by Kwiatkowski et al. (1992), one may conclude that long-term dependence exhibits given the empirical statistic above the critical value, that is, rejecting the null hypothesis of short memory at certain confidence levels. For instance, at 95% confidence level, the critical value is 0.463 according to Kwiatkowski et al. (1992), therefore, we can conclude that long memory exists for the truncation lag $q = 200$ at the 95% confidence level based on the results in Table 2.1 ($KPSS_{200} = 0.804$), but vanishes when considering more time lags, e.g. $q = 500$ ($KPSS_{500} = 0.452$); other applications can be found in Kirman and Teyssière (2001).

[4]Marmol (1997) shows that the KPSS test is also consistent against $I(d)$-processes with $d > 0.5$; and Sibbertsen and Kramer (2006) focus on $I(1 + d)$ processes.

Table 2.1: KPSS and Hurst exponent H values for absolute value of returns

$Data$	$q=0$	$q=5$	$q=10$	$q=20$	$q=50$	$q=100$	$q=200$	$q=500$
$KPSS_q$	7.131	5.358	4.361	2.899	1.960	1.264	0.804	0.452
H	0.609	0.593	0.581	0.559	0.537	0.513	0.488	0.456

Note: $KPSS_q$ and H are the empirical $KPSS$ statistic and Hurst exponent for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004). The critical value of 95% confidence level is 0.463 for the KPSS test.

2.2 Modelling long memory in financial time series

2.2.1 Introduction

Traditionally, long term dependence has been specified in the time domain in terms of long-lag autocorrelation, or in the frequency domains in terms of the explosion of low frequency spectra, as defined below.

A stationary time series y_t exhibits long memory when the autocorrelation function $\rho(\tau)$ behaves as:

$$\lim_{\tau\to\infty} \frac{\rho(\tau)}{C\cdot\tau^{2d-1}} = 1, \tag{2.2.1}$$

where C is a constant term, and d is the memory parameter. This definition clearly shows that long memory processes decay at a hyperbolic rate, which means that observations far distant from each other are still strongly correlated.

For the representation in the frequency domain, a Fourier transformation of the autocorrelation function $\rho(\tau)$ is required:

$$f(\lambda) = \int_{-\infty}^{+\infty} \rho(\tau) e^{-2\pi i\lambda\tau} d\tau, \tag{2.2.2}$$

and long term dependence is therefore equivalently defined by the spectral density of the time series $f(\lambda)$, for which there exists a real number $\alpha \in (0,1)$ and a constant $C' > 0$

such that:

$$\lim_{\lambda \to 0} \frac{f(\lambda)}{C'|\lambda|^{-\alpha}} = 1. \tag{2.2.3}$$

Long memory processes have been pervasively observed in hydrology, climatology and other natural phenomena. In addition, Leland et al. (1994) analyzed the long memory property of the network traffic flows. It also has been applied in social science, such as, Byers et al. (1997) modelled long-term dependence in opinion poll series; Dolado et al. (2003) investigated Spanish political polls, as well as in some studies on partisanship measures.

In particular, the presence of long-term dependence has important implications for many of the paradigms used in modern financial economics. For example, optimal consumption/savings and portfolio decisions may become extremely sensitive to the investment horizon if stock returns would exhibit long memory. Problems also arise in the pricing of derivative securities (such as options and futures) with martingale methods, since the class of continuous time stochastic processes most commonly employed are inconsistent with long-term dependence. Traditional tests of the capital asset pricing model and the arbitrage pricing theory are no longer valid since the usual forms of statistical inference do not apply to time series exhibiting such persistence. The conclusions of 'the efficient markets hypotheses' or stock market rationality also hang precariously on the presence or absence of long memory in raw returns.

Among the first to have considered the possibility and implications of persistent statistical dependence in asset returns was Mandelbrot (1971). Since then, several empirical studies have lent further support to Mandelbrot's findings. Baillie et al. (1996) found long memory in the volatility of the DM-USD exchange rate; long-term dependence in the German DAX was found by Lux (1996a); as well as stock market trading volume by Lobato and Velasco (2000), examining 30 stocks in the Dow Jones Industrial Average index. In addition, Chung (2002) and Zumbach (2004) also provide convincing evidence in favour of long memory models. Besides research on developed financial markets, there are also dozens

of papers on smaller and less developed markets: the stock market in Finland is analyzed by Tolvi (2003); Madhusoodanan (1998) provides evidence on the individual stocks in Indian Stock Exchange and Nath (2002) also gives significant indication of long-term dependence for all time lags in the Indian market; similar evidence on the Greek financial market is given by Barkoulas and Baum (2000); Cavalcante and Assaf (2002) demonstrate long memory in the Brazilian stock market.

2.2.2 Fractality and long memory

In contrast to the inference made by most hydrologists that the Nile's water inflow was a random process that had no underlying order and revealed no patterns between observations, Hurst (1951) found the existence of persistence – that is, large overflows tend to be followed by further larger overflows – by studying recorded data covering almost a millennium. Equipped with the Rescaled-range analysis and Hurst exponent value H, a time series can be classified into three categories:

(1) $H = 0.5$ indicates a random process.

(2) $0 < H < 0.5$ indicates an anti-persistent time series.

(3) $0.5 < H < 1$ indicates a persistent (long memory) time series.

An anti-persistent series has a characteristic of mean-reverting, which means an up value is more likely to be followed by a down value, and vice versa. The strength of "mean-reversion" increases as the Hurst exponent value approaches 0. A persistent time series indicates long memory, sometimes it is roughly called trend reinforcing, which means the direction (up or down compared to the last observation) of the next one is more likely the same as current value. The strength of trend increases as H approaches 1.0.

Beside the traditional R/S analysis, some alternative methods for estimating the Hurst exponent have also been developed, for instance Detrended Fluctuation Analysis (see, Peng

et al. (1994) and Peng et al. (1994)) and Wavelet Spectral Density inference (see, Jensen (1999)). The fact that fractals are characterized by long memory processes is widely accepted, and dozens of empirical works are using statistical inference methods to detect the persistence property in financial time series, such as Greene and Fielitz (1977), Booth and Kaen (1979), Booth et al. (1982), Helms et al. (1995).

There has been considerable recent interest in the estimation of long memory in frequency domain. One widely used method for this purpose was introduced by Geweke and Porter-Hudak (1983), hence it is known in the literature as the GPH estimator. The GPH estimator of persistence in volatility is based on an ordinary linear regression of the log periodigram of a time series x_t, which serves as a proxy for volatility, such as absolute returns, squared returns, or log squared returns of a financial asset. The single explanatory variable in the regression is log frequency for small Fourier frequencies in a neighbourhood, which degenerates towards zero frequency as the sample size T increases.

This procedure, based on least squares regression in the spectral domain, exploits the simple form of the pole of the spectral density (recall Eq. 2.2.3) at the origin:

$$f(\lambda) \sim |\lambda|^{-\alpha}, \lambda \to \infty,$$

with the jth periodigram $I(\lambda_j)$ such that

$$I(\lambda_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} x_t e^{i\lambda_j t} \right|^2 \tag{2.2.4}$$

where $\lambda_j = 2\pi j/T$ represents the pertinent Fourier frequency, T is the number of observations, and $j = 1, ..., m$; $m \ll T$ is the number of considered Fourier frequencies, that is the number of periodogram ordinates. The long memory parameter is estimated via the following regression:

$$\log[I(\lambda_j)] = \alpha_0 + \alpha \log\left[4sin^2(\lambda_j/2)\right] + \epsilon_t, \tag{2.2.5}$$

Under some conditions summarized by the band condition $\frac{1}{m} + \frac{m}{T} \to 0$ as $T \to +\infty$, the estimator of α can be obtained via OLS regression and $-\hat{\alpha}$ provides the long memory estimator (denoted as d in most of literatures). Figure 2.6 shows an empirical graph of the log periodogram for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004). An important practical problem in implementing the GPH estimator is the choice of m. Clearly this choice entails a bias-variance tradeoff. On the one hand, m should be sufficiently small in order to consider only frequencies near zero. On the other hand, m should be sufficiently large to ensure convergence of OLS estimation.[5]

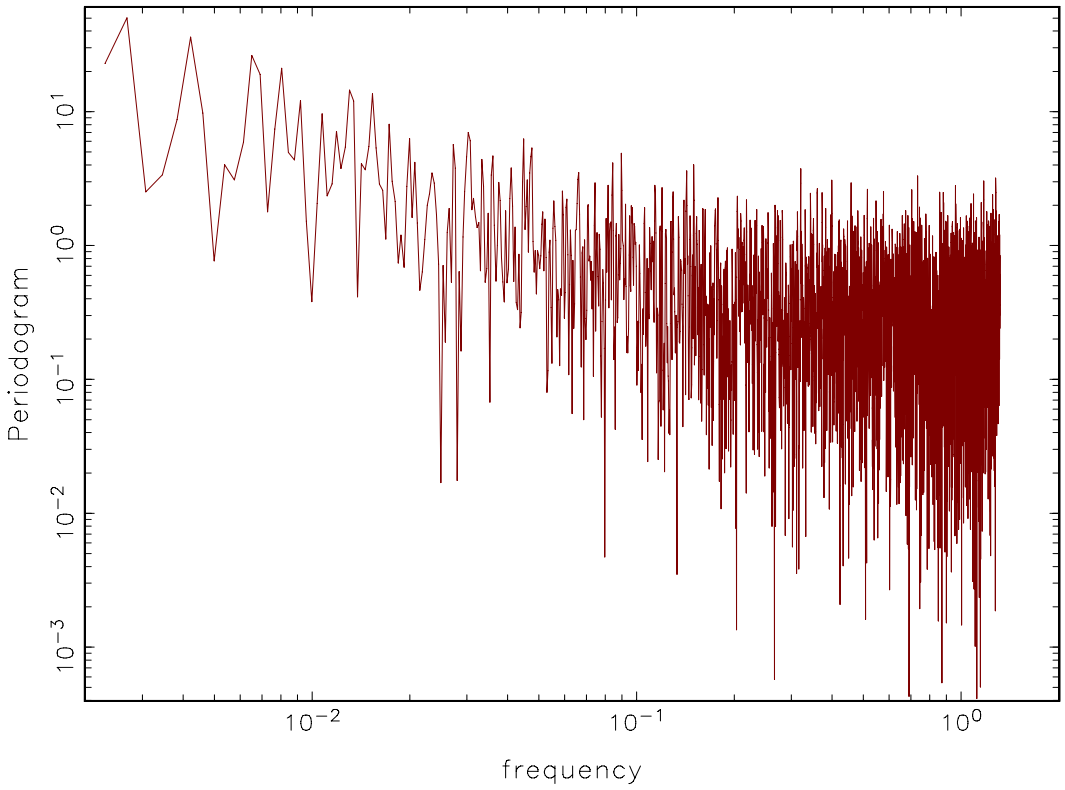

Figure 2.6: The periodogram of U.S. Dollar to British Pound exchange rate (March 1973 to February 2004).

[5]Traditionally, the number of periodogram ordinates is chosen from the interval $T^{0.45} \leq m \leq T^{0.55}$; Crato and De Lima (1994) and Boutahar et al. (2007) showed that m should be selected among $T^{0.5}$, $T^{0.55}$ and $T^{0.6}$. However, Hurvich et al. (1998) claimed that the optimal m is of order $T^{0.8}$.

Geweke and Porter-Hudak (1983) also provided the asymptotic distribution of the estimator:

$$m^{1/2}(\hat{\alpha} - \alpha) \rightarrow N\left(0,\ \frac{\pi^2}{6\sum_{j=1}^{m}(Z_j - \bar{Z})}\right) \tag{2.2.6}$$

with $Z_j = \log\left[4sin^2(\lambda_j/2)\right]$, $\bar{Z}$ being the mean of Z_j, with $j = 1, \ldots m$. Applications of GPH in the context of financial volatility have been presented in, for example, Andersen and Bollerslev (1996), Ray and Tsay (2001).

We recognize that the empirical determination of the long-term dependence property of a time series is a difficult problem. The reason is that the strong autocorrelation of long memory processes makes statistical fluctuations very large. Thus tests for long memory tend to require large quantities of data and can often give inconclusive results. Furthermore, different methods of statistical analysis often give contradictory results.

2.2.3 Long memory models

The material dealt with in the last section has inspired much debate as to the existence of long memory in financial time series. In this section, some popular long memory models in econometrics and their estimation issues are briefly reviewed; they are fractional Brownian motion, Fractional Integrated Autoregressive Moving Average (ARFIMA) model, Fractional Integrated Autoregressive Conditional Heteroskedasticity (FIGARCH) model, and the long memory stochastic volatility model.

Fractional Brownian motion

Fractional Brownian motion[6] (FBM) is presented by Granger and Joyeux (1980), Hosking (1981), where the process is defined as:

$$(1 - L)^d y_t = \epsilon_t, \tag{2.2.7}$$

[6]Given here is the discrete time version; a more detailed summary of the continuous version can be found in Baillie (1996).

for simplicity a zero mean is assumed, and ϵ_t is white noise. L is an operator that shifts the time index of a time series variable backwards by one unit of time. The process y_t is stationary if the fractional parameter $d < 0.5$ and invertible if $d > -0.5$. Granger and Joyeux (1980), Hosking (1981) also demonstrate its long memory property by transforming FBM to an infinite autoregressive (AR) process (Eq. 2.2.8) or infinite moving average (MA) representation (Eq. 2.2.9) as follows:

$$y_t = \sum_{\tau=0}^{\infty} \psi_\tau y_{t-\tau} + \epsilon_t, \tag{2.2.8}$$

with $\psi_\tau = \frac{\Gamma(\tau-d)}{\Gamma(\tau+1)\Gamma(-d)} \sim \frac{1}{\Gamma(-d)}\tau^{-d-1}$ as $\tau \to \infty$, $\Gamma(\cdot)$ denoting the gamma function. Alternatively,

$$y_t = \sum_{\tau=0}^{\infty} \tilde{\psi}_\tau \epsilon_{t-\tau}, \tag{2.2.9}$$

with $\tilde{\psi_\tau} = \frac{\Gamma(\tau+d)}{\Gamma(\tau+1)\Gamma(d)} \sim \frac{1}{\Gamma(d)}\tau^{d-1}$, as $\tau \to \infty$.

ARFIMA model

Let us start with ARMA models. An autoregressive process of order p, AR(p), is defined as:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \epsilon_t, \tag{2.2.10}$$

where $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. AR($p$) models a time series y_t, such that the level of its current observations depends on the level of its lagged observations.

A moving average model of order q, MA(q), can be written as:

$$y_t = \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + \cdots + b_q \epsilon_{t-q}. \tag{2.2.11}$$

In an MA(q) process, the observations of a random variable at time t are not only affected by the shock at time t, but also by the shocks that have taken place before time t. We

combine these two models and get a general Autoregressive Moving Average, ARMA(p, q) model:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_p y_{t-p} + \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + \cdots + b_q \epsilon_{t-q}. \tag{2.2.12}$$

With the appropriate modification, a non-stationary series is handled by the Autoregressive Integrated Moving Average model, ARIMA$(p, 1, q)$ which essentially begins with first order differencing to convert the original non-stationary series to a stationary series:

$$A(L)(1 - L)y_t = B(L)\epsilon_t, \tag{2.2.13}$$

L^p is an operator that shifts the time index of a time series variable backwards by p unit of time, that is $L^p y_t = y_{t-p}$, and AR lag polynomials $A(L) = 1 - a_1 L - \cdots - a_p L^p$; MA lag polynomials $B(L)) = 1 + b_1(L) + \cdots + b_q L^q$.

One flexible and popular econometric long memory time series model is the Fractionally Integrated Autoregressive Moving Average model ARFIMA (p, d, q), with a fractional differencing parameter d. It is introduced in the papers of Granger and Joyeux (1980), and Hosking (1981). Considering a time series observation y_t,

$$A(L)(1 - L)^d y_t = B(L)\epsilon_t, \tag{2.2.14}$$

$(1 - L)^d$ is the dth fractional difference operator defined by:

$$(1 - L)^d = 1 - dL - \frac{d(1-d)}{2!}L^2 - \frac{d(1-d)(2-d)}{3!}L^3 - \ldots = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}. \tag{2.2.15}$$

For a stationary process with $d \in (-0.5, 0.5)$ and $d \neq 0$, Hosking (1981) proves that the autocorrelations of an ARFIMA process decay hyperbolically to zero, in contrast to the faster geometric decay of a stationary ARMA process.[7] Nonzero values of the fractional parameter imply strong dependence between distant observations.

[7] Corresponding to $d = 0$.

Several estimation techniques have been proposed for ARFIMA(p, d, q) processes (see e.g. Sowell (1992)). Recalling the GPH, it is helpful to construct a two-step procedure: in the first step, the fractional differencing parameter d is found; and the AR and MA parameters are then estimated in the second step.

The parameters of ARFIMA model can also be jointly estimated through a maximum likelihood approach, under the assumption of Normality of the disturbances ϵ_t with variance σ_ϵ^2. The log-likelihood then can be written:

$$LogL(\Theta; \{y_t\}_{t=1,...T}) = -\frac{T}{2}log(2\pi) - \frac{1}{2}log(|\Xi|) - \frac{1}{2}y'\Xi^{-1}y, \tag{2.2.16}$$

Θ is the vector of parameters to be estimated, and $\Xi = \Xi(\Theta, \sigma_\epsilon^2)$ is the covariance matrix depending on Θ and σ_ϵ^2.

FIGARCH model

Analogous to the way that ARFIMA (p, d, q) constitutes an intermediary between ARMA and ARIMA processes, Baillie (1996) and Baillie et al. (1996) have presented a Fractional Integrated Autoregressive Conditional Heteroskedasticity model - FIGARCH (p, d, q) by combining the fractionally integrated processes of the regular GARCH process for the conditional variance.

We begin with the ARCH (Autoregressive Conditional Heteroskedasticity) process by Engle (1982):

$$\varepsilon_t = \sigma_t \cdot \eta_t, \tag{2.2.17}$$

where η_t is white noise, and σ_t is conditional variance, postulated to be a linear function of the lagged squared innovations:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2. \tag{2.2.18}$$

The ARCH model can be generalized by extending it with autoregressive terms of the volatility. Bollerslev (1986) provides a more flexible lag structure in the GARCH model,

the GARCH(p,q) model, defined by

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2. \tag{2.2.19}$$

Alternatively, it can also be expressed as an ARMA process in ε_t^2:[8]

$$[1-\alpha(L)-\beta(L)]\varepsilon_t^2 = \omega + [1-\beta(L)](\varepsilon_t^2 - \sigma_t^2), \tag{2.2.20}$$

$\varepsilon_t^2 - \sigma_t^2$ can be interpreted as the innovations of σ_t^2. Then the corresponding Integrated GARCH model is given as:[9]

$$\phi(L)(1-L)\varepsilon_t^2 = \omega + [1-\beta(L)](\varepsilon_t^2 - \sigma_t^2), \tag{2.2.21}$$

with $\phi(L) = [1-\alpha(L)-\beta(L)](1-L)^{-1}$.

The FIGARCH model is obtained by replacing the first difference operator by the dth fractional difference operator, by rearranging the conditional variance of ε_t:

$$[1-\beta(L)]\sigma_t^2 = \omega + \left[1-\beta(L)-\phi(L)(1-L)^d\right]\epsilon_t^2. \tag{2.2.22}$$

Baillie et al (1996) showed that the FIGARCH process also implies a slow hyperbolic rate of decay for lagged squared innovations and persistent impulse response weights. One remark that may be necessary is that here d is constraint within $d \in (0,1)$, which is different from the case of ARFIMA model; its log-likelihood has the form:

$$LogL(\Theta; \{\varepsilon_t\}_{t=1,\dots T}) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T} log(\sigma_t^2) - \frac{1}{2}\sum_{t=1}^{T}(\varepsilon_t^2 \cdot \sigma_t^{-2}). \tag{2.2.23}$$

To have the proper order, ie. $\{p,q\}$, information criteria (IC) are required for model identification and specification, such as Akaike IC, Hannan-Quinn IC, Schwarz IC, Shibata IC.[10]

[8] It is an ARMA(q',p) process with $q' = \max(p,q)$.

[9] Here it is the GARCH$(q'-1,p)$ process.

[10] They are constructed with different approaches but their structures are very similar in the end, with the only difference being the penalization term.

Long memory stochastic volatility model

During the last decade, there has been an increasing interest in modelling the dynamic evolution of the volatility of financial time series using stochastic volatility (SV) models. The main distinction between ARCH type models and SV model relies on whether volatility is observable. In the standard stochastic volatility model, the volatility is modelled as an unobserved latent variable, and this characteristic feature coincides with other theoretical models in finance which build on the concept of unobservable latent factors generating asset returns, for example information flow interpretations of the mixture of distribution hypothesis, see Andersen (1996), Liesenfeld (1998).

The standard stochastic volatility model introduced by Taylor (1982) assumes a stochastic process y_t, such that

$$\begin{aligned} y_t &= \sigma \cdot \exp(h_t/2) \cdot \epsilon_t \\ h_t &= \mu_0 + \mu_1 h_{t-1} + \sigma_\eta \cdot \eta_t, \end{aligned} \tag{2.2.24}$$

and the logarithm of volatility h_t is an AR(1) process. σ is a scale parameter, and μ_0 is a constant term.[11] ϵ_t and η_t are mutually independent standard Normals (although it is rather common to assume that the error components have a Gaussian distribution, several authors have also considered heavy-tailed distributions, e.g. Liesenfeld and Jung (2000).). The volatility of the log-volatility process, σ_η measures the uncertainty about future volatility.

By adopting the ARFIMA process, the traditional standard Stochastic Volatility (SV) model can also be extended to a long memory SV model (LMSV) replacing the AR process in the state equation with a fractional differencing operator, cf. Harvey (1993) and Breidt

[11] Some simplified SV version assumes $\sigma = 1$ and the constant term μ_0 being set to zero without loss of generality.

et al. (1998). LMSV is then defined as:

$$\begin{aligned} y_t &= \sigma \cdot \exp(h_t/2) \cdot \epsilon_t \\ (1-L)^d h_t &= \mu_0 + \sigma_\eta \cdot \eta_t, \end{aligned} \tag{2.2.25}$$

the log volatility h_t at time t follows a stationary fractionally integrated process with the long-memory parameter, $d < 1/2$, and L is the lag operator as introduced in the previous fractional integrated models.

SV models are attractive because they are close to the models often used in financial theory (e.g. Black-Scholes option pricing model) to represent the behaviour of financial prices whose statistical properties are easy to derive using well-known results on the Log-normal distribution. However, until recently, their empirical applications have been very limited mainly because the exact likelihood function is difficult to evaluate and Maximum Likelihood (ML) estimation of the parameters is not straightforward. Nevertheless, several estimation methods have been proposed and the literature on SV models has grown substantially. These include simple moment matching by Taylor (1986); simulated method of moment by Gourieroux et al. (1993); Generalized Method of Moments by Melino and Turnbull (1990) and Andersen and Sorensen (1996); quasi-maximum likelihood by Harvey et al. (1994); simulation based maximum likelihood by Danielsson and Richard (1993), Liesenfeld and Richard (2003); the Markov Chain Monte Carlo (MCMC) approach by Shephard (1993) and Jacquier et al. (1994). By means of an extensive Monte Carlo study, Jacquier et al. (1994) show that MCMC is more efficient than both quasi-maximum likelihood (QML) and GMM estimators.

In addition, some important studies have been made on the estimation of the long memory SV model, which cover the quasi-maximum likelihood estimator of Breidt et al. (1998) and GMM estimation by Wright (1999). Kim et al. (1998) also provide a simulation-based method on the estimation and diagnostics of the LMSV model employing MCMC, and So

(1999) develops a new algorithm based on the state space formulation of Gaussian time series models with additive noise where full Bayesian inference is implemented through MCMC techniques. The MCMC algorithm creates a Markov chain on the blocks of the unknown parameters, latent volatilities, and state mixing variables, whereby, on repeated sampling from the distribution of each block conditioned on the current values of the remaining blocks, the chain geometrically converges to the desired multivariate posterior density. The main attraction of MCMC procedures is that they permit to obtain simultaneously sample inference about the parameters, smooth estimates of the unobserved variances and predictive distributions of multi-step forecasts of volatility. In any case, noticed that an important advantage of the MCMC estimators is that the inference is based on finite sample distributions and, consequently, the asymptotic approximation is not needed.

2.3 Multifractal models

2.3.1 Introduction

Financial markets display some properties in common with fluid turbulence. For example, both fluid turbulence and financial fluctuations display intermittency at all scales. A cascade of energy flux is known to occur from the large scale of injection to the small scale of dissipation, which is typically modelled by a multiplicative cascade, which then leads to a multifractal concept; see Lux (2001) 'Turbulence in Financial Markets: The Surprising Explanatory Power of Simple Cascade Models'.

A stochastic process $X(t)$ is called multifractal if it has stationary increments and satisfies (cf. Mandelbrot et al. (1997), Calvet and Fisher (2002)):

$$E[|X(t)|^q] = c(q)t^{\tau(q)+1}, \tag{2.3.1}$$

where $\tau(q)$ is called the scaling function, and $c(q)$ is some deterministic function of q.

2.3.2 Multifractal model of asset returns

The development of the multifractal approach goes back to Benoit Mandelbrot's work on turbulent processes in statistical physics (see Mandelbrot (1974)). One widespread influential contribution is Mandelbrot et al. (1997), which, by introducing the multifractal model of asset returns (MMAR), translates the approach from physics into finance. In the MMAR model, returns $r(t)$ are assumed to follow a compound process:

$$r(t) = B_H\left[\theta(t)\right], \tag{2.3.2}$$

in which, $B_H[\cdot]$ represents an incremental fractional Brownian motion with Hurst exponent index H (we know already that it is ordinary Brownian motion for $H = 0.5$); $\theta(t)$ is an increasing function of chronological time t. Both $B_H[\cdot]$ and $\theta(t)$ are independent.

The MMAR provides a fundamentally new class of stochastic processes to financial economists. It is able to generate fat tails in the unconditional distribution of financial returns, and it also has long memory in the absolute value of returns (FIGARCH model and the Long Memory Stochastic Volatility (LMSV) model characterize long memory in squared returns). In addition, the multifractal process has appealing temporal aggregation properties and is parsimoniously consistent with the moment scaling of financial data - in the sense that a well-defined rule relates returns over different sampling intervals. Mandelbrot (1963) suggested that the shape of the distribution of returns should be the same when the time scale is changed (as defined in Eq. (2.1.1)). Finally, one essential feature of MMAR is the compounding of $\theta(t)$ as trading time or time deformation process, and it is characterized by the cumulative distribution function of a multifractal measure, first introduced and used in Mandelbrot (1974), when modelling the distribution of energy in turbulent dissipation.

The simplest way to create a multifractal measure is as a "binomial multifractal" constructed on a unit interval [0; 1] with uniform density. One proceeds as follows: divide the interval into two parts of equal length, and let m_0 and m_1 be two positive numbers adding

up to 1. In step 1, when $k = 1$, the two subintervals are equal, and the measure uniformly spreads mass equal to m_0 on the subinterval $[0;\ 0.5]$ and mass equal to m_1 on $[0.5;\ 1]$, in step 2, the sub-set $[0;\ 0.5]$ is split into two subintervals, $[0;\ 0.25]$ and $[0.25;\ 0.5]$; which respectively receive fractional measures m_0 and m_1 of the total mass $[0;\ 0.5]$; We apply the same procedure to the dyadic set $[0.5;\ 1]$, $\cdots$, the above procedure is then repeated *ad infinitum*, and this iteration generates an infinite sequence of measures. Using this simple mechanism, Figure 2.7 illustrates the density function of the binomial multifractal measure with $m_0 = 0.4$ and $m_1 = 0.6$. From the top to the bottom of Figure 2.7 are the realizations for $k = 3$, $k = 5$ and $k = 7$ respectively, and one may easily recognize the fractal property with increasing the number of k.

The major innovation of MMAR model is the use of a multifractal cascade as a transformation of chronological time into 'business time', and the multifractal component is interpreted as a process governing instantaneous volatility. For a minor extension of the original binomial measure, one could simply dispense the rule of always assigning m_0 to the left, and m_1 in the right with randomizing the assignment; or, one may uniformly split the interval into an arbitrary number b larger than 2 at each stage of the cascade, and receive the fractions $m_0, m_1 \ldots m_{b-1}$, which leads to a so-called multinomial measure. Furthermore, one can also randomize the allocations between the subintervals, taking $m_0, m_1 \ldots m_{b-1}$ with certain probabilities, or using random numbers for m_0 instead of a constant value, such as drawing from Lognormal distribution, see Mandelbrot (1974, 1997).

2.3.3 Scaling estimators

The scaling estimation is an early entry into the issue of estimating multifractal processes. A multifractal time series $X(t)$ on the time interval $[0, T]$, is divided into N intervals of length Δt, and its partition function is defined as:

$$S(q; \Delta t) = \sum_{i=0}^{N-1} |X(i\Delta t + \Delta t) - X(i\Delta t)|^q. \tag{2.3.3}$$

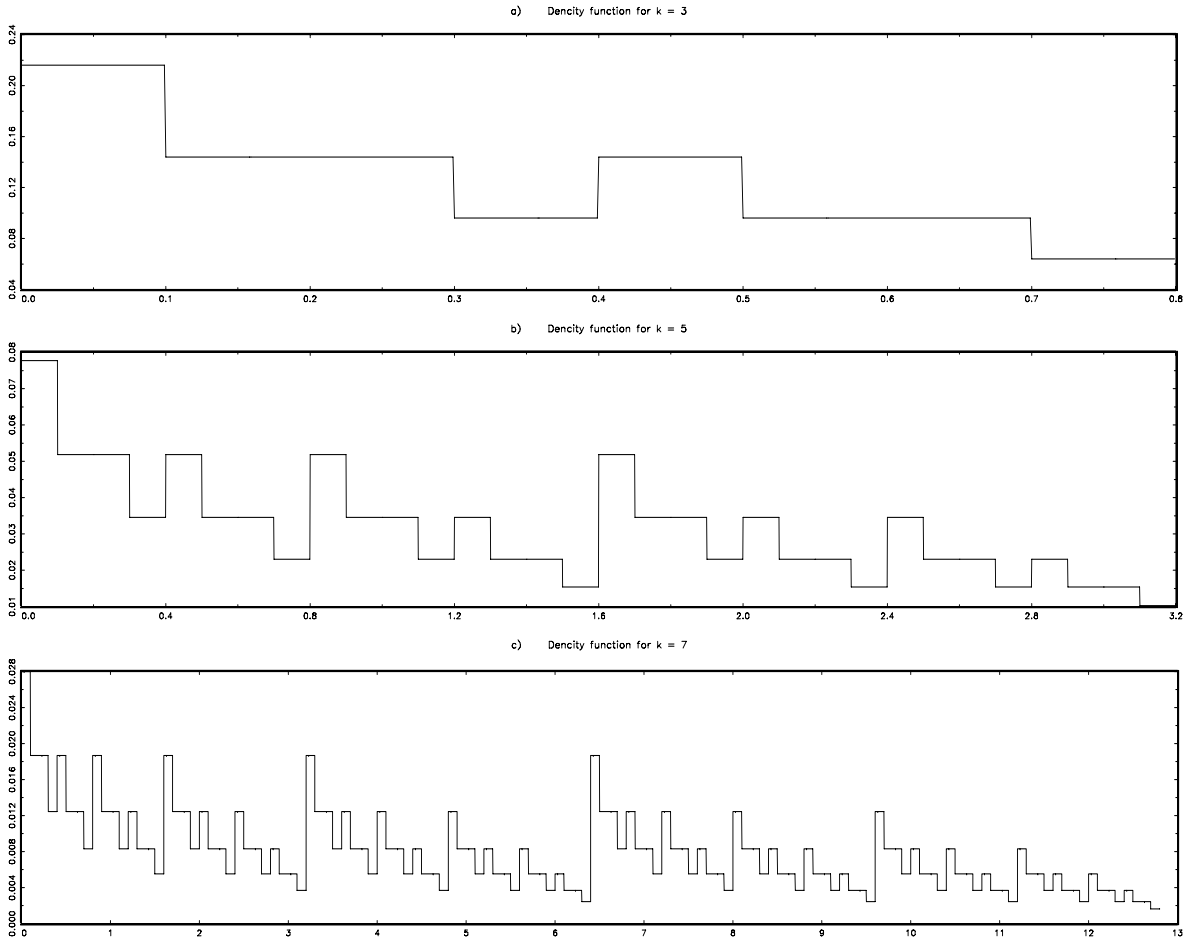

Figure 2.7: Density function of the binomial multifractal measure with $m_0 = 0.4$ and $m_1 = 0.6$.

The multifractal measures are characterized by a non-linear moment scaling function (scaling law):

$$E[S(q;\Delta t)] = C(q)(\Delta t)^{\tau_X(q)+1}, \tag{2.3.4}$$

$\tau_X(q)$ is a non-linear moment scaling function depending on the particular variant of the multifractal process.

The last essential component of the scaling estimator is the multifractal spectrum $f(\alpha)$. In the pertinent literature, the parameters of multifractal cascades are usually not estimated

directly from the scaling function $\tau(q)$, but rather from its Legendre transformation:

$$f_\theta(\alpha) = Inf[\alpha q - \tau_X(q)], \tag{2.3.5}$$

that allows an estimation of the MF process by matching the empirical and hypothetical spectra of the Hölder exponents (a continuum of local scaling factors replaces the unique Hurst exponent of uni-fractal processes such as fractional Brownian motion). One may notice the subscript of θ in Eq. (2.3.5), which refers to the spectrum of $\theta(t)$, and the shape of the spectrum carries over from the multifractal time transformation to returns in the compound process via the equations:[12]

$$\tau_X(q) = \tau_\theta(Hq). \tag{2.3.6}$$

$$f_X(\alpha) = f_\theta(\alpha/H). \tag{2.3.7}$$

Eq. (2.3.6) allows the estimation of the Hurst Index H in empirical work by using the relationship:[13]

$$\tau_X(1/H) = \tau_\theta(1) = 0. \tag{2.3.8}$$

However, most studies restrict the price process assuming that the logs of prices follow a Brownian motion with arbitrary $H = 0.5$ instead of fractal Brownian motion. The reason is that empirical evidence of long-term dependence (which give an estimator of $H > 0.5$) is confined to various powers of returns, but is almost absent in the raw data, and statistical tests can usually not reject the null hypothesis of $H = 0.5$ for raw returns. Hence, one does not need to assume a fractional Brownian motion of returns.

Mandelbrot et al. (1997) derived the analytical solution of the scaling function and multifractal spectrum with respect to the binomial and Lognormal MF; the pertinent $\tau(q)$ and $f(\alpha)$ functions are obtained as below:

[12]More technical details about Legendre transforms are provided by Harte (2001): *Multifractal: Theory and Applications.* and Mandelbrot et al. (1997).

[13]The proof of this result can be found in Mandelbrot et al. (1997).

For the Binomial distribution with $m_0 \geq 0.5$, the scaling function has the form:

$$\tau(q) = -\log_2(m_0^q + (1-m_0)^q), \tag{2.3.9}$$

and the spectrum $f(\alpha)$ is:

$$f_\theta(\alpha) = -\frac{\alpha_{max}-\alpha}{\alpha_{max}-\alpha_{min}}\log_2\left(\frac{\alpha_{max}-\alpha}{\alpha_{max}-\alpha_{min}}\right) - \frac{\alpha-\alpha_{min}}{\alpha_{max}-\alpha_{min}}\log_2\left(\frac{\alpha-\alpha_{min}}{\alpha_{max}-\alpha_{min}}\right), \tag{2.3.10}$$

with $\alpha_{min} = -\log_2(m_0)$; $\alpha_{max} = -\log_2(1-m_0)$.

For the Lognormal (LN) distribution MF, that is

$$M_t^{(i)} \sim LN(-\lambda, \sigma_m^2), \tag{2.3.11}$$

and conservation of mass imposes that $E[M_t] = 1/b$, or equivalently $\sigma_m^2 = 2\ln b(\lambda - 1)$, which leaves us only one parameter to estimate. For b = 2, Mandelbrot et al. (1997) presented the scaling function is

$$\tau(q) = q\lambda - q^2(\lambda - 1) - 1, \tag{2.3.12}$$

and the pertinent multifractal spectrum has the form of:

$$f_\theta(\alpha) = 1 - \frac{(\alpha-\lambda)^2}{4(\lambda-1)}. \tag{2.3.13}$$

Figure 2.8 to 2.10 illustrate the traditional method of estimation of the multifractal process with Lognormal cascades. One starts with the empirical partition functions $S(q;\Delta t)$ in Eq. (2.3.3) given a set of positive moments q and time scales Δt of the data, and the partition functions are then plotted against Δt in logarithmic scales. Regression estimates of the slopes then provide the corresponding scaling function $\tau(q)$ from Eq. (2.3.4). Figure 2.8 shows a selection of partition functions for some low (up) and higher moments (down) for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004). As can be observed, the empirical behavior is very close to the presumed linear shape for moments of

small order which reveals striking visual evidence of moment scaling, while the fluctuations around the regression line become more pronounced for higher powers. This is, however, to be expected as the influence of chance fluctuations is magnified with higher powers q. The resulting scaling function for moments in the range [-10, 10] is exhibited in Figure 2.9. A broken line gives a clear deviation from pure Brownian motion, which is scaling according to $q/2 - 1$.

Finally, the last step consists in computing the multifractal $f(\alpha)$ spectrum. Figure 2.10 is a visualization of the Legendre transformation. The spectrum is obtained by drawing lines of slope q and intercept $\tau(q)$ for various q. If the underlying data indeed exhibits multifractal properties, these lines would turn out to constitute the envelope of the distribution $f(\alpha)$. As can be seen, a convex envelope emerges from our scaling functions. It seems worthwhile to emphasize that this outcome is shared by all other studies available hitherto, which may suggest that such a shape of the spectrum is a robust feature of financial data.

For fitting the empirical spectrum by its theoretical counterpart, the inverted parabolic shape of the Lognormal cascade, we have to keep in mind, that the cascade model is used for the volatility or time deformation $\theta(t)$ and that the returns themselves result from the compound process $B_{0.5}[\theta(t)]$. We, therefore, have to take into account the shift in the spectrum as detailed in Eq. (2.3.7). In order to arrive at parameter estimates for λ, the common approach pursued in physical applications is to compute the best fit to Eq. (2.3.7) for the empirical spectrum using a least square criterion. One particularly uses the positively sloped (left-hand) part of the spectrum, because the right-hand is computed from partition functions with negative powers and is, therefore, strongly affected by fluctuations due to the Brownian process.

However, there are no results on the consistency and asymptotic distribution of the $f(\alpha)$ estimates available in the relevant literature, and this approach also does not provide us with estimates of the standard deviation of the incremental Brownian motion nor of

the number of steps k to be used in the cascade. The later omission is somewhat natural since the underlying physical models assume an infinite progression of the cascade which is also the reason for their initially scale free approach. Besides that, the $\tau(q)$ and $f(\alpha)$ fits also require judgmental selection of the number and location of steps used for the scaling estimates of the moments and the non-linear least square fit of the spectrum.

2.3.4 Uni-fractals and multifractals

The issue of distinguishing between uni-fractal and multifractal models lies in the linearity or non-linearity of the scaling function $\tau(q)$. A uni-fractal process is characterized by a linear scaling function, which can be derived by recalling the self-similar fractal process in Eq. (2.1.1), implying that $X(ct) \simeq c^H X(t_1)$. We have:

$$E[|X(t)|^q] = t^{Hq} E[|X(1)|^q]. \tag{2.3.14}$$

Recalling the scaling law in Eq. (2.3.4), we then end up with $\tau(q) + 1 = Hq$. Therefore an uni-fractal process is characterized by its scaling function such that:

$$\tau(q) = Hq - 1. \tag{2.3.15}$$

One may also reach the well-known uni-fractal relationship for the case of Brownian motion:

$$\tau(q) = \frac{q}{2} - 1, \tag{2.3.16}$$

(see Parisi and Frisch (1985), Fisher et al. (1997), Schmitt et al. (2000)).[14]

There are uni-fractal limiting cases for both Binomial and Lognormal multifractal processes. In the former case, the limit is obtained when $m_0 = 0.5$, and it is easy to see that this means a split of mass with probabilities 0.5. In the Lognormal case, this limit is given by $\lambda = 1$ due to the vanishing variance.

[14]Note that, in practice, simulations of uni-fractal processes would hardly give this perfectly linear relationship.

In contrast, a multifractal process has a nonlinear scaling function; other descriptions state that a uni-fractal process is characterized by a single exponent H, whereas multifractal processes have varying H. By this way, the traditional long memory time series models, for instance fractional Brownian motion (FBM) model, as well as the ARFIMA, FIGARCH, LMSV models, are all in the category of uni-fractal models.

Empirical evidences for (multi)fractal in financial economics

Dozens of financial markets have been examined for their fractal properties: Näslund (1990) analyzes the fractal structure of capital markets; Fang et al. (1994) studies fractal structure in currency futures price dynamics; Batten and Ellis (1996), Gallucio et al. (1997), and Mulligan (2000) present evidence of fractals for various foreign exchange rates. Similar research has been conducted by Matsushita et al. (2003), which shows the fractal structure in the Chinese Yuan/US Dollar exchange rate; Mulligan (2004) focuses on technology stocks as representatives of highly volatile markets.

There are also a number of works on multifractal properties, they include: Demos and Vassilicos (1994), who show the multifractal structure of high-frequency foreign exchange rate fluctuations; Fisher et al. (1997) provide evidence for the multifractality of the German Mark (DM) - U.S. Dollar exchange rate. By means of Monte Carlo simulations with Kolmogorov-Smirnov criterion, Lux (2001) demonstrates the success of the multifractal model in comparison with the GARCH model in matching various empirical financial market data; Fillol (2003) also shows the multifractal properties of the French stock exchange (CAC40) and its Monte Carlo simulations prove that the multifractal model of asset returns (MMAR) is a better model to replicate the scaling properties observed in the CAC40 series than traditional models like GARCH and FIGARCH.

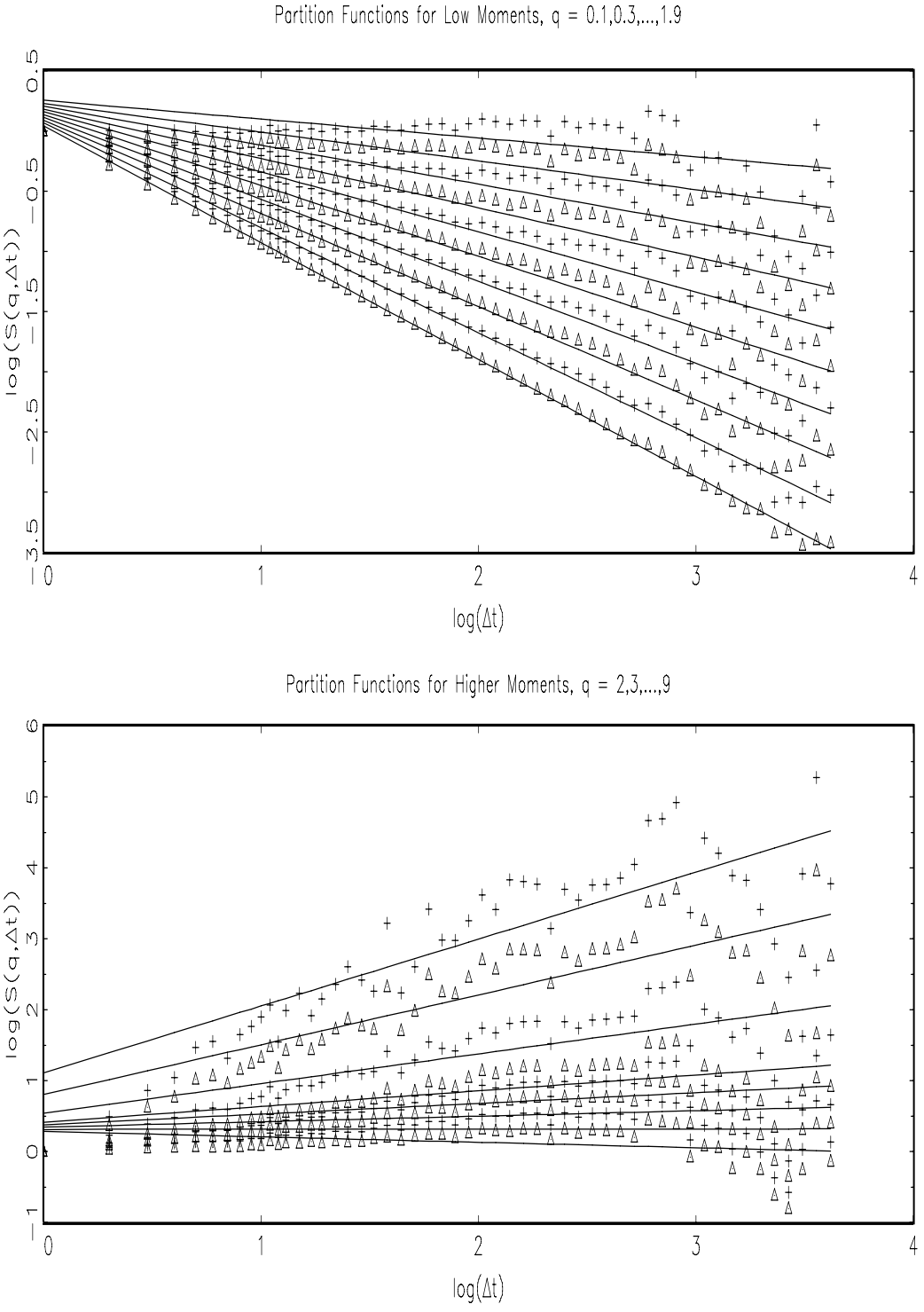

Figure 2.8: Partition functions of U.S. Dollar to British Pound exchange rate (March 1973 to February 2004) for different moments.

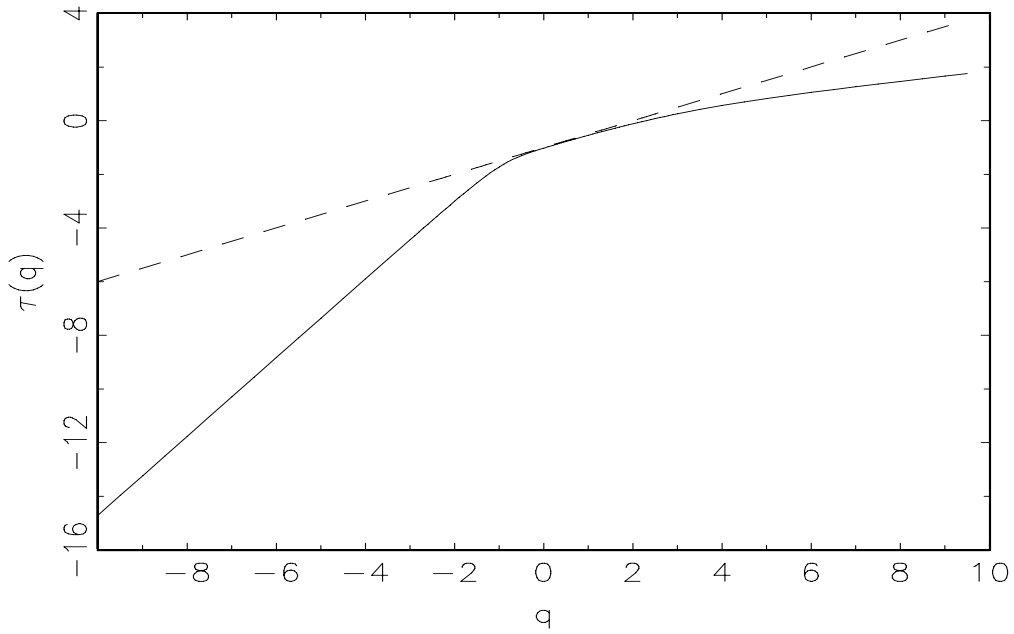

Figure 2.9: The scaling function for U.S. Dollar to British Pound exchange rate (March 1973 to February 2004).

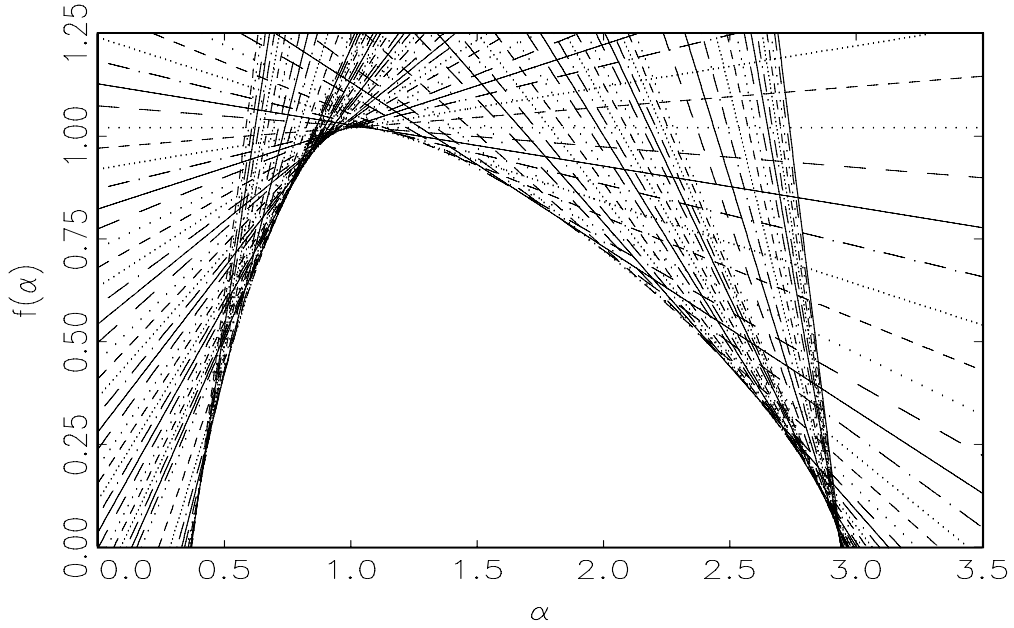

Figure 2.10: The $f(\alpha)$ spectrum of U.S. Dollar to British Pound exchange rate (March 1973 to February 2004).

Chapter 3

The Bivariate Markov Switching Multifractal Models

3.1 Introduction

MMAR provides us with a new model for financial time series with attractive stochastic properties, which take into account stylized facts of the financial market, such as fat tail, volatility clustering, long-term dependence and multi-scaling.

However, the practical applicability of MMAR suffers from its combinatorial nature and from its non-stationarity due to the restriction to a bounded interval. Taking a binomial example with an underlying binary cascade extending over k steps, there are exactly 2^k realizations of different subintervals at our disposal; consequently we are limited to 'time series' no longer than 2^k; in addition, the model suffers from a dearth of applicable statistical methods, see Mandelbrot et al. (1997).

These limitations have been overcome by introducing iterative versions of MF models. Calvet and Fisher (2001) present a multifractal model with random times for the changes of the multipliers (Poisson multifractal), demonstrating weak convergence of a discretized version of this process to its continuous-time limit. Iterative MF models preserve the multifractal and stochastic properties, in particular they make econometric analysis applicable.

In the Markov-switching multifractal model (cf. Calvet and Fisher (2004), Lux (2008)),

financial asset returns are modelled as:

$$r_t = \sigma \left(\prod_{i=1}^{k} M_t^{(i)} \right)^{1/2} \cdot u_t, \tag{3.1.1}$$

with a constant scale parameter σ, and increments u_t drawn from a standard Normal distribution $N(0,1)$. The local instantaneous volatility is then determined by the product of k volatility components or multipliers $M_t^{(1)}$, $M_t^{(2)}$ ..., $M_t^{(k)}$; various restrictions can be imposed with respect to the choices of M_t, moreover, $E[M_t]$ or $E[\sum M_t]$ being equal to some arbitrary value is a requirement for normalizing the time-varying components of volatility; for instance, Calvet and Fisher (2004) assume a Binomial distribution with parameters $m_0 \in (1,2)$ and $2 - m_0$, guaranteeing an expectation of unity for all $M_{i,t}$ that is, $E[M] = 1$. In addition, Lux (2008) introduces a Lognormal multifractal process.

Furthermore, a hierarchical structure is introduced to regulate the multifractal dynamics. To begin with, either each volatility component is renewed at time t with probability γ_i, depending on its rank within the hierarchy of multipliers, or else it remains unchanged with probability $1 - \gamma_i$. The transition probabilities are specified as:

$$\gamma_i = 1 - (1 - \gamma_k)^{(b^{i-k})} \qquad\qquad i = 1, \ldots k, \tag{3.1.2}$$

with parameters $\gamma_k \in [0,1]$ and $b \in (1, \infty)$. Estimation using this specification, then, involves the parameters γ_k and b, as well as those characterizing the distribution of the volatility components $M_{i,t}$. Lux (2008) uses a parsimonious setting by fixing $b = 2$ and $\gamma_k = 0.5$; Similarly, other specifications are used in its previous versions, e.g.

$$\gamma_i = b^{-(k-i)} \qquad\qquad i = 1, \ldots k, \tag{3.1.3}$$

which leads to a very close approximation to that of Eq. (3.1.2), with parameters that are more parsimonious but sufficient to capture the hierarchical structure within the multifractal dynamics. This says that the volatility component in a higher frequency state will be renewed twice ($b = 2$) as often as its next lower neighbour, and it happens with

certainty for the highest frequency component ($i = k$). Simulations with both Eq. (3.1.2) and Eq. (3.1.3) show similar patterns, and our study will focus on the parsimonious setting by fixing the transition parameters. One may also notice the relatively high standard errors of these two estimates (γ_k and b) in Calvet and Fisher (2004), and our empirical studies will further assess its ability to replicate the stylized facts.

Using the iterative version of the multifractal model instead of its combinatorial predecessor and confining attention to unit time intervals, the resulting dynamics of Eq. (3.1.1) can also be seen as a particular version of a stochastic volatility model. With this rather parsimonious approach, this pertinent MF model preserves the hierarchical structure of MMAR while dispensing with its restriction to a bounded interval, similarly, the model captures some properties of financial time series, namely, outliers, volatility clustering and the power-law behaviour of the autocovariance function:

$$Cov(|r_t|^q, |r_{t+\tau}|^q) \propto \tau^{2d(q)-1}. \tag{3.1.4}$$

As has also been pointed out by Calvet and Fisher (2001), although models of this class are partially motivated by empirical findings of long-term dependence of volatility, they do not obey the traditional definition of long-memory, i.e. asymptotic power-law behavior of autocovariance functions in the limit $t \to \infty$ or divergence of the spectral density at zero, see Beran (1994). The iterative MF model is rather 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. In the case of Markov-Switching multifractal process, therefore, approximately hyperbolic decline along the line of Eq. (3.1.4) holds only over an interval $1 \ll \tau \ll b^k$ with b and k defined as in Eq. (3.1.2).

Eq. (3.1.4) also implies that different powers of the measure have different decay rates in their autocovariances, one essence of multifractality, agreeing with Calvet and Fisher (2002), which shows that this feature carries over to the absolute moments of returns in

MMAR. One should note that it is this characteristic that distinguishes MF models from other long memory processes, such as FIGARCH and ARFIMA models, which belong to the category of uni-fractal models, i.e. they have the same decay rate for all moments.

Various approaches have been employed to estimate multifractal models. The parameters of the combinatorial MMAR have been estimated via an adaption of the scaling estimator and $f(\alpha)$ approach of statistical physics (cf. Calvet and Fisher (2002)). However, this approach has been shown to yield very unreliable results (cf. Lux (2004)). A broad range of more rigorous estimation methods have been developed for the iterative MF model. Calvet and Fisher (2001) propose maximum likelihood (ML) estimation due to its Markov structure. Together with ML estimation, Bayesian forecasting of volatility has also been successfully applied to forecast foreign exchange rate volatilities. However, with the computing capability of current personal computers, the applicability of ML estimation encounters a upper bound for the number of cascade levels; and it is also restricted to cases that have only a discrete distribution of volatility components.

Lux (2008) adopts the Generalized Method of Moments (GMM) approach of Hansen (1982), which can be applied not only to the discrete but also to continuous distributions of the volatility components. Likewise, the best linear forecast based on the Levinson-Durbin algorithm has been applied successfully. In empirical studies, Calvet and Fisher (2004), Lux (2008), and Lux and Kaizoji (2007) report the potential advantages of the multifractal process compared to GARCH and FIGARCH models for various financial time series in terms of their volatility forecasts.

In this chapter, we focus on constructing a bivariate multifractal model and implementing its estimation using various approaches. The motivation for this work is actually quite straightforward; univariate models limit financial applications, particularly in portfolio analysis, thus motivating research into multivariate settings, while there are only few studies

along these lines with multifractal models. In the following sections, we start with a bivariate multifractal (MF) process of Calvet et al. (2006), then introduce our bivariate MF model as a simple alternative one. We implement both models' estimation by using a maximum likelihood approach as well as simulation based inference (particle filter). Furthermore, we also apply the GMM approach to estimate bivariate multifractal models, which not only relaxes the upper bound of the number of cascade levels, but also applies to models with a continuous distribution of the volatility components. For both models, we conduct Monte Carlo studies to compare the performance of each estimation method.

3.2 Bivariate multifractal (BMF) models

It is now a well-established fact that financial markets and their respective assets are correlated, and this has received much attention from the finance profession. Multivariate settings provide relatively more information for portfolio management, and there have been increasing numbers of studies along these lines. For example, Bollerslev (1990) studies the changing variance structure of the exchange rate regime in the European Monetary System. Baillie and Myers (1996) further apply a bivariate GARCH model to derive optimal hedge ratios for commodity futures; Engle and Susmel (1993) propose multivariate models with common factors; other models with multivariate (bivariate) settings, include Harvey et al. (1994) on multivariate stochastic volatility models and Liesenfeld (1998) modelling return volatility and trade volume based on the mixture of distribution hypothesis. Accordingly, these previous works motivate the construction of multivariate multifractal models.

3.2.1 Calvet/Fisher/Thompson model

One recently appearing bivariate MF model in Calvet et al. (2006)[1] considers two financial time series returns $r_{q,t}$ for $q = 1, 2$, and assumes volatility is composed of heterogenous

[1] Called the Calvet/Fisher/Thompson model henceforth.

frequencies. For each frequency i, the local volatility components M_t are:

$$M_t = \begin{bmatrix} M_{1,t}^{(i)} \\ M_{2,t}^{(i)} \end{bmatrix} \tag{3.2.1}$$

The column vectors M_t are stacked into a $2 \times n$ matrix, and each column contains a particular volatility component at the corresponding cascade level. $M_t^{(i)}$ denotes the volatility component M at the cascade level i at time t. By defining

$$g(M_t) = \prod_{i=1}^{n} M_t^{(i)}, \tag{3.2.2}$$

Calvet/Fisher/Thompson's approach assumes that each time series follows a univariate MF process in Eq. (3.1.1), and specifies the bivariate time series $r_{q,t}$ (2×1 vector) as:

$$r_{q,t} = \sigma_q \otimes [g(M_t)]^{1/2} \otimes u_{q,t}. \tag{3.2.3}$$

$\otimes$ denotes element by element multiplication, σ_q is a 2×1 vector of scale parameters, $u_{q,t}$ is a 2×1 vector whose elements follow a bivariate standard Normal distribution with the correlation parameter ρ. M_t is drawn from a bivariate binomial distribution $M = (M_1, \ M_2)'$. M_1 takes value $m_1 \in (1, 2)$ and $2 - m_1$, and $P(M_1 = m_1) = 1/2$; M_2 takes value $m_2 \in (1, 2)$ and $2 - m_2$, and $P(M_2 = m_2) = 1/2$. Thus the random vector M has four possible values, whose probabilities are determined by a 2×2 matrix:

$$\begin{bmatrix} \frac{1+\rho_m}{4} & \frac{1-\rho_m}{4} \\ \frac{1-\rho_m}{4} & \frac{1+\rho_m}{4} \end{bmatrix}$$

with ρ_m being the correlation between M_1 and M_2 under the distribution of M, and $\rho_m \in [-1, 1]$. The model focuses on the specification $\rho_m = 1$ for simplicity, cf. Calvet et al. (2006).

In addition, whether or not certain volatility components (new arrivals) are updated is governed by the transition probabilities γ_i, which is specified as in the univariate version:

$$\gamma_i = 1 - (1 - \gamma_n)^{(b^{i-n})}, \qquad i = 1, \ldots n, \tag{3.2.4}$$

with parameters $\gamma_n \in (0,1)$ and $b \in (1,\infty)$. It defines the probability of a new arrival happening at the cascade level i, i.e., whether a volatility component $M_t^{(i)}$ is updated by a new arrival or not. Furthermore, arrivals across two series are characterized by a correlation parameter $\lambda_m \in [0,1]$. New arrivals are independent if $\lambda_m = 0$ and simultaneous if $\lambda_m = 1$. Therefore, estimating of Calvet/Fisher/Thompson model involves overall eight parameters, which are $\sigma_1, \sigma_2, m_1, m_2, \rho, b, \gamma_n, \lambda_m$.

3.2.2 Liu/Lux model

In this section, a bivariate multifractal model[2] is introduced as a simple alternative of Calvet/Fisher/Thompson model with a more parsimonious setting. It assumes that two time series have a certain number of joint cascade levels in common. The economic intuition is that the observed correlation between different markets/assets can either be due to common news processes, or to common factors, such as the business cycle or technology shocks. We model the bivariate asset returns $r_{q,t}$ as

$$r_{q,t} = \sigma_q \otimes \left[g'(M_t)\right]^{1/2} \otimes u_{q,t}. \tag{3.2.5}$$

$q = 1, 2$ refers to the two time series respectively, having n levels of their volatility cascades. $\otimes$ denotes element by element multiplication, σ_q is the vector of scale parameters (unconditional standard deviations); $u_{q,t}$ is a 2×1 vector whose elements follow a bivariate standard Normal distribution, with an unknown correlation parameter ρ. In our model, we assume for the column vector M_t that

$$g'(M_t) = \prod_{i=1}^{k} M_t^{(i)} \otimes \prod_{j=k+1}^{n} M_{q,t}^{(j)}, \tag{3.2.6}$$

$\prod_{j=k+1}^{n} M_{q,t}^{(j)}$ stands for $M_t^{(k+1)} \otimes M_t^{(k+2)} \cdots$, that means, both time series share k number of joint cascades that govern the strength of their volatility correlation. Consequently, the

[2]Called the Liu/Lux model henceforth.

larger the k, the higher the correlation between them. After k joint multiplications, each series has additional separate multifractal components.

Furthermore, the restriction for the specification of the transition probabilities is arbitrarily imposed. In our model, we allow two starting cascades within each time series (the multifractal process starts again after the joint cascade level k), that is

$$\begin{aligned} \gamma_i &= 2^{-(k-i)}, \qquad for \quad i = 1, \ldots k; \\ \gamma_i &= 2^{-(n-i)}, \qquad for \quad i = k+1, \ldots n. \end{aligned} \tag{3.2.7}$$

Each component is either renewed at time t with probability γ_i, depending on its rank i within the hierarchy of multipliers, or remains unchanged with probability $1 - \gamma_i$, following the hierarchical structure, which implies that the one at a higher cascade level has a higher probability of being updated. With regard to the heterogeneous multipliers, we follow the published routine in specifying them to be random draws from either a Binomial or Lognormal distribution. In the Binomial case, it is assumed that each volatility component within the column vector M_t is drawn from $M_t \sim \{m_0, \ m_1\}$ for $m_0 \in (0, 2)$ and alternatively $m_1 = 2 - m_0$; For the latter, we assume $-log_2 M_t \sim N(\lambda, \ \sigma_m^2)$.

Simulations of our bivariate MF model (k = 4, n = 15) are depicted in Figure 3.1 to 3.3. Figure 3.1 is the local instantaneous volatility of the simulated bivariate MF time series; Figure 3.2 shows simulations of the corresponding returns, together with their autocorrelation functions of returns and absolute returns in Figure 3.3. The dot dashed lines are roughly fitted by power functions of $0.5\tau^{-\beta}$ (time lags τ) with β approximately being 0.371 and 0.413 respectively. The simulations are seen to have some of the stylized facts of financial time series, namely: outliers, volatility clustering and hyperbolical decay of the autocorrelation function (long-term dependence). It is impressive to consider the remarkable 'long memory' depicted in Figure 3.3, and one also easily recognizes the correlation of the volatilities between each of the simulated time series.

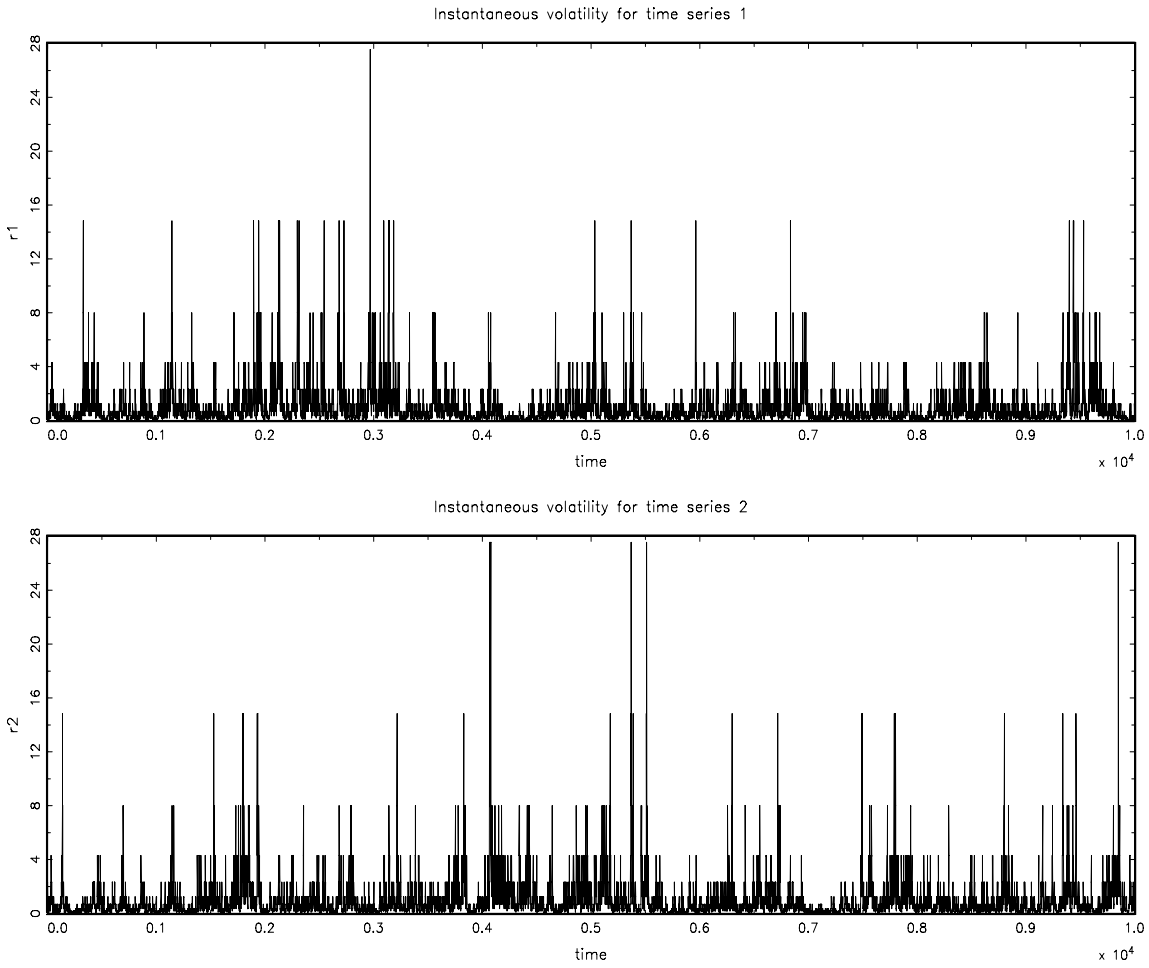

Figure 3.1: Local instantaneous volatility of the simulated bivariate MF (Binomial) time series.

3.3 Exact maximum likelihood estimation

The extension from univariate multifractal process to bivariate one trivially guarantee the positive semi-definiteness of the covariance matrix of the bivariate time series, and the likelihood function can be written in closed-form which allows maximum likelihood estimation to be implemented for a certain size of state spaces, see Calvet et al. (2006).

The dynamics of MF can be taken as a special case of a Markov-switching process which makes ML estimation feasible. Since the state spaces are finite when the multipliers

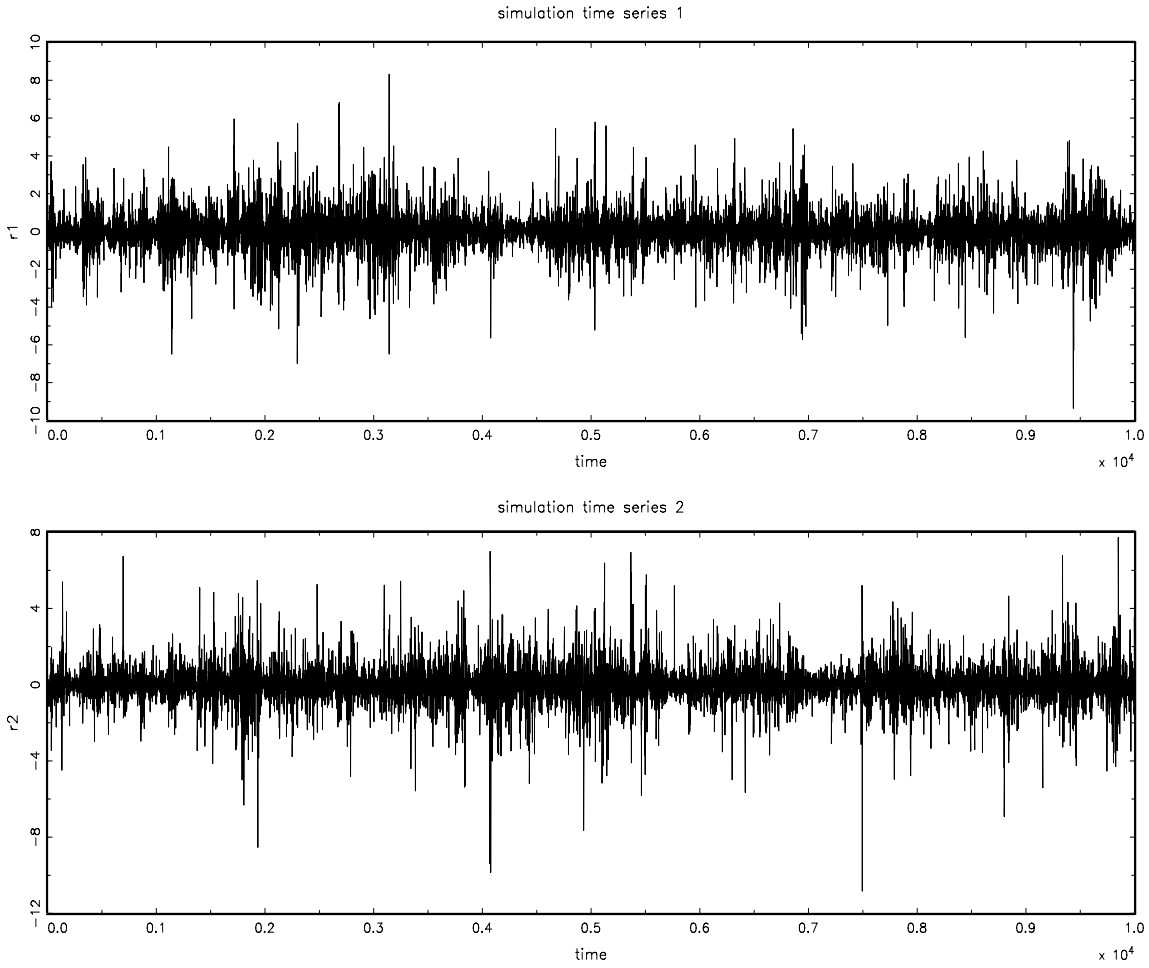

Figure 3.2: Simulation of the bivariate multifractal (Binomial) model.

follow a discrete distribution (e.g. a Binomial distribution), the likelihood function can be derived by determining the exact form of each possible component in the transition matrix, as developed by Calvet and Fisher (2004), Calvet et al. (2006). Let r_t be the set of joint return observations $\{r_{q,t}\}$ for $q = 1, 2$, and $t = 1, 2 \ldots T$. The explicit likelihood function is

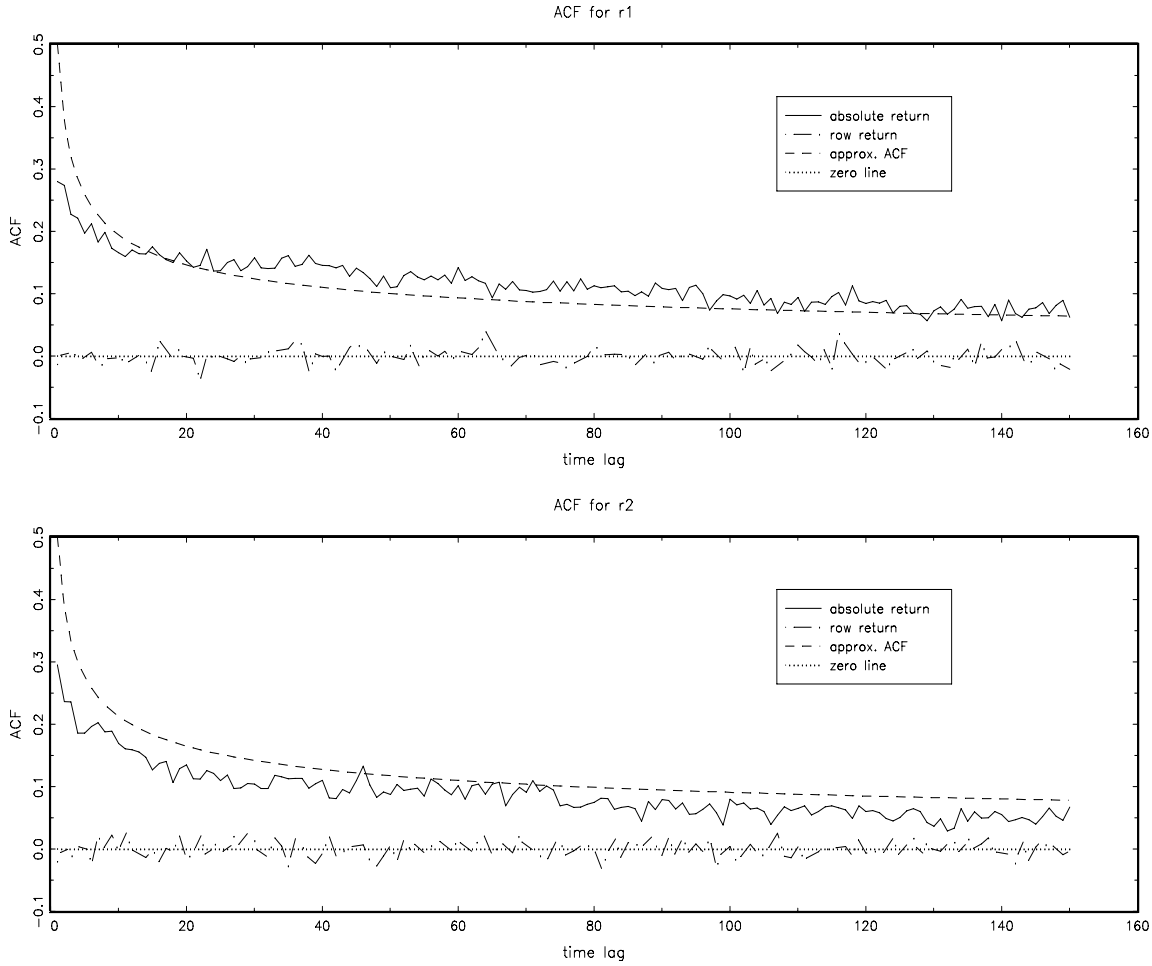

Figure 3.3: ACF for the simulations of the BMF (Binomial) Model above.

below:

$$
\begin{aligned}
f(r_1, \cdots, r_T; \Theta) &= \prod_{t=1}^{T} f(r_t | r_1, \cdots, r_{t-1}, \Theta) \qquad (3.3.1) \\
&= \prod_{t=1}^{T} \left[\sum_{i=1}^{4^n} f(r_t | M_t = m^i) \cdot P(M_t = m^i | r_1, \cdots, r_{t-1}) \right] \\
&= \prod_{t=1}^{T} f(r_t | M_t = m^i) \cdot (\pi_{t-1} A) \,.
\end{aligned}
$$

Θ is a set of parameters to be estimated. There are three elements within the likelihood function above, namely, the transition matrix A, $f(r_t | M_t = m^i)$, and a vector of conditional

probability of $\pi_{t-1} = (\pi_{t-1}^1, \ldots, \pi_{t-1}^{4^n})$. We interpret each element in our objective function, one by one, as follows: the transition matrix A contains components A_{ij} which are equal to

$$P(M_{t+1} = m^j | M_t = m^i). \tag{3.3.2}$$

Note that $i, j = \{\ 1, 2 \ldots 4^n\}$ and it indicates the transition matrix A has the dimension of $4^n \times 4^n$.

To have the intuition of A, let us begin with one univariate version, with cascade level of 2; there are four values for the Binomial distribution of volatility components as in Calvet and Fisher (2004) and Lux (2008), namely:

$$\{m_1 m_1, \quad m_1 m_0, \quad m_0 m_1, \quad m_0 m_0\}.$$

Given one current volatility state, it is quite obvious that there are four possible volatility components for the next time step; therefore we have the following possible combinations of new volatility component arrivals ($M_{t+1} = m^j | M_t = m^i$):

$$M_{uni} = \begin{bmatrix} m_1m_1|m_1m_1 & m_1m_1|m_1m_0 & m_1m_1|m_0m_1 & m_1m_1|m_0m_0 \\ m_1m_0|m_1m_1 & m_1m_0|m_1m_0 & m_1m_0|m_0m_1 & m_1m_0|m_0m_0 \\ m_0m_1|m_1m_1 & m_0m_1|m_1m_0 & m_0m_1|m_0m_1 & m_0m_1|m_0m_0 \\ m_0m_0|m_1m_1 & m_0m_0|m_1m_0 & m_0m_0|m_0m_1 & m_0m_0|m_0m_0 \end{bmatrix} \tag{3.3.3}$$

which is associated with probabilities of $P(M_{t+1} = m^j | M_t = m^i)$ in the transition matrix. In the univariate MF model, the volatility state vector $M_t = (M_t^1, \cdots, M_t^n)$, M_t^n denotes one volatility component at the cascade level n. In the univariate binomial case, there are 2^n possible volatility components combinations, i.e. $m^1, \ldots, m^{(2^n)}$. Therefore, we have the transition matrix A_{uni} for the number of cascade level $n = 2$:

$$A_{uni} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

More precisely, in the univariate MF process of Eq. (3.1.1) with the transition probability defined as Eq. (3.1.3) with $b = 2$, we get

$$A_{uni} = \begin{bmatrix} 0.375 & 0.375 & 0.125 & 0.125 \\ 0.375 & 0.375 & 0.125 & 0.125 \\ 0.125 & 0.125 & 0.375 & 0.375 \\ 0.125 & 0.125 & 0.375 & 0.375 \end{bmatrix} \tag{3.3.4}$$

By extending to bivariate models, the transition matrix becomes remarkably sophisticated with expanding cascade levels, since we need to take into consideration the new arrivals of possible volatility components for the other time series given each state of the first one.

Let us take one simplest example of the Calvet/Fisher/Thompson model. The transition probability is defined as in the univariate case of Eq. (3.1.3) with $b = 2$, and the volatility arrival correlation parameters are set as $\lambda_m = 1$ and $\rho_m = 1$. The transition matrix for the bivariate model can be directly derived from the one for the univariate version above but additionally considering all possibilities of states for another time series, given a current state of a volatility component in one time series, and it has the dimension of 16×16. With our assumption on the correlation parameters for new volatility components between both time series, it would be also not too difficult to imagine there are zero values within the transition matrix, since our assumption of the correlation parameters implies updating of one multiplier in one multifractal process simultaneously leads to updating in the other one at the same cascade level. Therefore, those events that the volatility components for both time series at the same cascade level do not update simultaneously give zero entries. The 16×16 dimension of A (Eq. 3.3.1) for the Calvet/Fisher/Thompson model with cascade level $n = 2$ is given as:

0.375	0	0	0	0	0.375	0	0	0	0	0.125	0	0	0	0	0.125
0	0.375	0	0	0.375	0	0	0	0	0	0	0.125	0	0	0.125	0
0	0	0.375	0	0	0	0	0.375	0.125	0	0	0	0	0.125	0	0
0	0	0	0.375	0	0	0.375	0	0	0.125	0	0	0.125	0	0	0
0	0.375	0	0	0.375	0	0	0	0	0	0	0.125	0	0	0.125	0
0.375	0	0	0	0	0.375	0	0	0	0	0.125	0	0	0	0	0.125
0	0	0	0.375	0	0	0.375	0	0	0.125	0	0	0.125	0	0	0
0	0	0.375	0	0	0	0	0.375	0.125	0	0	0	0	0.125	0	0
0	0	0.125	0	0	0	0	0.125	0.375	0	0	0	0	0.375	0	0
0	0	0	0.125	0	0	0.125	0	0	0.375	0	0	0.375	0	0	0
0.125	0	0	0	0	0.125	0	0	0	0	0.375	0	0	0	0	0.375
0	0.125	0	0	0.125	0	0	0	0	0	0	0.375	0	0	0.375	0
0	0	0	0.125	0	0	0.125	0	0	0.375	0	0	0.375	0	0	0
0	0	0.125	0	0	0	0	0.125	0.375	0	0	0	0	0.375	0	0
0	0.125	0	0	0.125	0	0	0	0	0	0	0.375	0	0	0.375	0
0.125	0	0	0	0	0.125	0	0	0	0	0.375	0	0	0	0	0.375

Note: The transition matrix of the Calvet/Fisher/Thompson model with the number of cascade levels $n = 2$.

From the simple example of the Calvet/Fisher/Thompson model, each element within the transition matrix is associated with the probabilities of $(m^a m^a | m^a m^a)$, given $(m^b m^b | m^b m^b)$ (denoted $\langle m^a m^a | m^a m^a \rangle \langle m^b m^b | m^b m^b \rangle$ henceforth); a and b correspond the bivariate time series; m^a and m^b are drawn from the Binomial distributions. More specifically, let's refer to the univariate case in the matrix 3.3.3, i.e. M_{uni}, therefore the matrix for the bivariate case can be implicitly viewed as Kronecker product of M_{uni} with another one for the second time series, and we arrive the dimension of 16×16 transition matrix. For instance, the first element of 0.375, which is calculated as the probability for $\langle m_1^a m_1^a | m_1^a m_1^a \rangle \langle m_1^b m_1^b | m_1^b m_1^b \rangle$, and it implies 0.375×1 due to the assumption of simultaneous updating of volatility components. Let's then take the second element (the first row and second column), which corresponds to the probability for the occurrence of $\langle m_1^a m_1^a | m_1^a m_1^a \rangle \langle m_1^b m_1^b | m_1^b m_0^b \rangle$, and it means the volatility components for time series 1 remain the same at both cascade levels, but the volatility component for time series 2 is updated at the second cascade level, which implies the zeros entry due to violating the assumption of the simultaneous updating. In the same way, we calculate the third element (the first row and third column) of $\langle m_1^a m_1^a | m_1^a m_1^a \rangle \langle m_1^b m_1^b | m_0^b m_1^b \rangle$ with zero value, as well as the fourth component $\langle m_1^a m_1^a | m_1^a m_1^a \rangle \langle m_1^b m_1^b | m_0^b m_0^b \rangle$, and one may also easily obtain all other entries whin this transition matrix.

Turning to the Liu/Lux model with $n = 2$ and joint cascade level $k = 1$. In this parsimonious version of bivariate models, we actually have two combined multifractal processes with cascade level of 1, since it starts again after the first cascade level as defined in our model. Likewise, the transition matrix for the bivariate model can be obtained by considering all possibilities of states for one time series, given a current state of a volatility component in another time series. Apart from the Calvet/Fisher/Thompson model, the zero entries in the transition matrix are corresponding to those events of non-identical volatility components within the joint cascades of both time series, by recalling our assumption of joint cascades. For example, the probability for the occurrence of $\langle m_1 m_1 | m_1 m_1 \rangle \langle m_1 m_1 | m_1 m_1 \rangle$

can be obtained by the probability of 0.5×0.5 for the first $\langle m_1 m_1 | m_1 m_1 \rangle$ and 0.5×0.5 for the second one, which implies $0.5 \times 0.5 \times 0.5 = 0.125$.[3] The same case for the scenario of $\langle m_1 m_1 | m_1 m_1 \rangle \langle m_1 m_1 | m_1 m_0 \rangle$. But the probability for occurrence of $\langle m_1 m_1 | m_1 m_1 \rangle$ $\langle m_1 m_1 | m_0 m_1 \rangle$ gives zero entry due to violating the assumption of the joint cascade level; analogically, the case of of $\langle m_1 m_1 | m_1 m_1 \rangle \langle m_1 m_1 | m_0 m_0 \rangle$ has zero probability. These are the first four entries (the first to the fourth column of the first row) of the transition matrix of our simple example of Liu/Lux model:

$$
\begin{vmatrix}
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 \\
0.125 & 0.125 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0 & 0 & 0.125 & 0.125 & 0 & 0 & 0.125 & 0.125
\end{vmatrix}
$$

[3]Note that there is only one common Binomial distribution for two volatility component arrivals, that means, volatility components for both time series are drawn from $\{m_0, m_1\}$.

We easily recognize the regular appearance of zero entries due to our assumption of joint cascade levels. Practicably, we can remove all these zeros by excluding the scenarios that violate the model's assumption, which reduces the dimension of the transition matrix from $4^n \times 4^n$ to $2^{2n-k} \times 2^{2n-k}$. For example, in the case of $n = 3$ and joint cascade level $k = 2$, we only need to evaluate the 16×16 matrix instead of 64×64 matrix (without excluding the zero entries). As illustrated in previous pages, all possible combinations of volatility components can be viewed as the Kronecker product of two univariate ones (we skip it for the case of three cascade levels due to the page size limit). Let us start with the scenario of $\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle$, we know the probability of $\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle$ for the first time series is $(0.5+0.5\times0.5)\times0.5\times0.5 = 0.1875$.[4] Since we have the joint cascade levels of two, which implies the probability of $\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle$ is $0.1875 \times 0.5 = 0.094$; it is the same for the case of $\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_1 m_1 m_0 \rangle$. But for cases of

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_1 m_0 m_1 \rangle$ and

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_1 m_0 m_0 \rangle$, we obtain a probability of zero. The same applies to those scenarios where the volatility components at the first cascade level for the time series one remain unchange, while the ones for the second time series would change:

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_0 m_1 m_1 \rangle$,

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_0 m_1 m_0 \rangle$,

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_0 m_0 m_1 \rangle$,

$\langle m_1 m_1 m_1 | m_1 m_1 m_1 \rangle \langle m_1 m_1 m_1 | m_0 m_0 m_0 \rangle$, etc.

By removing all those zeros entries, we arrive at the 16×16 transition matrix:

[4] It is calculated via $\Pi_{k=1}^{n}[(1-\gamma_k)1_{m_k^i = m_k^j} + \gamma_k P(M = m_k^j)]$. Note that, here, m_k^i is the m^{th} component of vector m^i, and $1_{m_k^i = m_k^j}$ is the dummy variable equal to 1 if $m_k^i = m_k^j$ and zero otherwise, c.f. Calvet and Fisher (2002). In our case, we have $\gamma_1 = 1/2$, $\gamma_2 = 1$; $\gamma_3 = 1$ by recalling Eq. 3.1.3.

0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094
0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094	0.031	0.031	0.031	0.031	0.094	0.094	0.094	0.094

The density of the innovation r_t conditional on M_t is:

$$f(r_t|M_t = m^i) = \frac{F_N\left\{r_t \div \left[\sigma_q \otimes \eta^{1/2}\right]\right\}}{\sigma_q \otimes \eta^{1/2}}. \tag{3.3.5}$$

$F_N\{\cdot\}$ denotes the bivariate standard Normal density function and $\div$ represents element-by-element division. $\eta = g(M_t)$ for the Calvet/Fisher/Thompson model, and $\eta = g'(M_t)$ for the Liu/Lux model.

The last unknown element in the likelihood function is π_t, which is the conditional probability, defined as

$$\pi_t^i = P(M_t = m^i|r_1, \cdots, r_t), \tag{3.3.6}$$

and due to $\sum_{i=1}^{4^n} \pi_t^i = 1$; by Bayesian updating, we get

$$\pi_{t+1} = \frac{f(r_{t+1}|M_{t+1} = m^i) \otimes (\pi_t A)}{\sum f(r_{t+1}|M_{t+1} = m^i) \otimes (\pi_t A)}. \tag{3.3.7}$$

We easily recognize that the computational demands of this Bayesian updating highly depend on the dimensionality of the transition matrix A.

We implement the ML estimation for both the Calvet/Fisher/Thompson model and the Liu/Lux model with $n = 5$, which is the limit of computational feasibility for the bivariate ML approach. For the Calvet/Fisher/Thompson model, we keep the same parsimonious setting as in the example before, namely, the the specification of transition probability for each time series following the univariate case of Eq. (3.1.3) with $b = 2$; the volatility arrival correlation parameter being $\lambda_m = 1$ (simultaneous updating for the same cascade level). Therefore, estimation of the Calvet/Fisher/Thompson model with simple specifications involves five parameters, which are σ_1, σ_2, m_1, m_2, ρ.

We conducted Monte Carlo studies to explore the performance of ML estimators. 400 Monte Carlo simulations and estimations were carried out and there were 100,000 observation generated in each simulation. Then, three different sizes of sub-samples ($N_1 = 2000$,

$N_2 = 5000$, and $N_3 = 10000$) were randomly selected for estimation. Initial parameter values are fixed such that $m_1 = 1.2$; $m_2 = 1.4$; $\sigma_1 = 1$; $\sigma_2 = 1$; $\rho = 0.5$. Table 3.1 presents the statistical results of the Monte Carlo experiments for the Calvet/Fisher/Thompson model.

Table 3.1: ML estimation for the Calvet/Fisher/Thompson model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_1}$	N_1	-0.009	0.025	0.025
	N_2	0.002	0.012	0.013
	N_3	0.002	0.006	0.008
$\hat{m_2}$	N_1	0.005	0.027	0.027
	N_2	0.006	0.014	0.014
	N_3	0.001	0.007	0.007
$\hat{\sigma_1}$	N_1	-0.009	0.033	0.030
	N_2	0.0011	0.025	0.026
	N_3	0.005	0.013	0.014
$\hat{\sigma_2}$	N_1	0.011	0.039	0.041
	N_2	0.006	0.022	0.023
	N_3	-0.007	0.011	0.011
$\hat{\rho}$	N_1	0.010	0.034	0.033
	N_2	0.010	0.023	0.023
	N_3	0.004	0.011	0.010

Note: Simulations are based on the Calvet/Fisher/Thompson model with the number of cascade levels $n = 5$, and other parameters are $m_1 = 1.2$, $m_2 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

We also performed a similar Monte Carlo study for the Liu/Lux model. The initial parameter values of $m_0 = 1.4$; $\rho = 0.5$; $\sigma_1 = 1$; $\sigma_2 = 1$ are fixed; sub-sample sizes again are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$, respectively. Table 3.2 gives the ML estimates for the Liu/Lux model (binomial model) in the case of $n = 5$, $k = 2$, which is almost the limit of computational feasibility for the ML process (studies on other randomized parameter values have also been pursued, we omit them here due to their very similar results).

Both Table 3.1 and Table 3.2 report convincing performance of ML estimators. The

Table 3.2: ML estimation for the Liu/Lux model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	-0.008	0.030	0.032
	N_2	-0.012	0.017	0.017
	N_3	-0.008	0.009	0.010
$\hat{\sigma_1}$	N_1	0.010	0.031	0.032
	N_2	0.003	0.019	0.020
	N_3	-0.007	0.010	0.010
$\hat{\sigma_2}$	N_1	0.0011	0.033	0.032
	N_2	-0.003	0.022	0.022
	N_3	-0.001	0.010	0.010
$\hat{\rho}$	N_1	0.008	0.029	0.029
	N_2	0.002	0.018	0.017
	N_3	0.001	0.008	0.009

Note: Simulations are based on the Liu/Lux model with the number of cascade levels $n = 5$, and other parameters are $m_0 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

average bias of the Monte Carlo estimates is close to zero throughout different sub-sample sizes; SD (standard deviation) and RMSE (root mean squared error) are all quite small even in the small sample sizes $N = 2000$, and they are decreasing with increasing sub-sample sizes. However, applicability of the ML approach is constrained by its computational demands: first, it is not applicable to models with an infinite state space, i.e. continuous distributions of the volatility component, such as the Lognormal distribution that has been introduced in a few multifractal papers. Secondly, even for discrete distributions, say the Binomial case, current computational limitations make choices of cascades with a number of steps n beyond 5 unfeasible because of the implied evaluation of a $4^n \times 4^n$ transition matrix in each iteration. Thirdly, it is worthwhile remarking that the implementation of ML for the Binomial model is quite time-consuming, we report that it takes 87 hours for Table 3.1

and 116.5 hours for Table 3.2 which involve 400 Monte Carlo simulations and estimations.[5]

3.4 Simulation based ML estimation

As pointed out in the last section, there is an upper limit of $n = 5$ in ML estimation of bivariate MF models; to overcome this restriction, a simulation based inference is proposed. Recalling the Bayesian updating in Eq. (3.3.7) above, we can think of $\pi_t A$, which is

$$\sum_{j=1}^{4^n} P(M_{t+1} = m^i | M_t = m^j) P(M_t = m^j | r_t) \tag{3.4.1}$$

as a prior probability of

$$P(M_{t+1} = m^i | r_t), \tag{3.4.2}$$

then combine it with the conditional density

$$f(r_{t+1} | M_{t+1} = m^i) \tag{3.4.3}$$

to generate a posterior, that it is

$$\pi_{t+1} = P(M_{t+1} = m^i | r_{t+1}). \tag{3.4.4}$$

This procedure can also be expressed as a repeated application of a two-stage procedure:

(1) considering the conditional probabilities of current states π_t as an input, passing through a system of dynamic transformations, the transition probability matrix A here, to propagate the prediction density Eq. (3.4.2);

(2) one then uses the Bayesian updating to produce the conditional probabilities of future states π_{t+1} as output.

This procedure is called filtering. This is straightforward if $P(M_{t+1} = m^i | M_t = m^j)$ has a reasonable size of finite discrete elements as the previous calculation can be computed explicitly.

[5]Note: All simulation and estimations are pursued by using a PC with Pentium® IV Processor.

As one may realize, the bivariate multifractal process has the dimension of the transition (filtering) probability matrix increasing exponentially with increasing number of cascade levels n. Furthermore, there is an infinite number of states if the distribution of the volatility components is continuous. This implies that it would be difficult (or impossible for the continuous version) to evaluate the procedure above exactly, and some numerical methods must be used. Numerous attempts have been made to provide algorithms that approximate these filtering probabilities, cf. Gordon et al. (1993), Jacquier et al. (1994), Berzuini et al. (1997), Kim et al. (1998). We use a so-called particle filter, which is a class of simulation-based filters that recursively approximate the filtering of random variable by a certain finite number of particles, which are discrete points viewed as approximated samples from the prior. In our case, we evaluate Eq. (3.3.7) by combining the conditional density with Eq. (3.4.1) up to proportionality (for $R = 4^n$):

$$\pi_{t+1}^i \propto f(r_{t+1}|M_{t+1} = m^i) \sum_{j=1}^{R} P(M_{t+1} = m^i|M_t = m^j)\pi_t^j. \tag{3.4.5}$$

As particle filters treat the discrete support generated by the particles as the 'true' filtering density, this allows us to produce an approximation to the prediction probability density $P(M_{t+1} = m_t^i|r_t)$, by using the discrete support of the number B of particles, and then the one-step-ahead conditional probability is

$$\pi_{t+1}^i \propto f(r_{t+1}|M_{t+1} = m^i)\frac{1}{B}\sum_{b=1}^{B} P(M_{t+1} = m^i|M_t = m^{(b)}). \tag{3.4.6}$$

This leaves only one issue – how to design the finite number of B draws? We adopt Sampling/Importance Resampling (SIR) introduced by Rubin (1987); Pitt and Shephard (1999). This algorithm generates $\{M_t^{(b)}\}_{b=1}^B$ recursively with updating information:

Starting from $t = 0$ with the initial condition π_0, and $M_0^{(1)}, \ldots, M_0^{(B)}$ are drawn. For $t = 1$, we simulate each $\{\hat{M}_t^{(b)}\}_{b=1}^B$ independently and reweighting to obtain the importance

sampler $\{M_t^{(b)}\}_{b=1}^B$ via drawing a random number q from 1 to B with the probability of:

$$P(q=b) = \frac{f(r_t|M_t = m^{(b)})}{\sum_{i=1}^B f(r_t|M_t = m^{(i)})}. \tag{3.4.7}$$

$M_t^{(1)} = \hat{M}_t^{(q)}$ is then selected, repeat it B times and obtain B draws with $M_t^{(1)}, \dots M_t^{(B)}$. Therefore, for any $t \geq 1$, we then can simulate the Markov chain one-step-ahead to obtain $\{M_{t+1}^{(b)}\}_{b=1}^B$ based on Eq. (3.4.7), which is adjusted to account for the new information. Instead of evaluating each exact component of R numbers of A associated with $\pi_t^{(\cdot)}$, SIR produces B draws ('particles') from the prior $P(M_{t+1} = m^i|r_t)$ that are used to generate an approximation of the corresponding one-step-ahead conditional probability, as Eq. (4.3.3); this will converge with increasing B (cf. Pitt and Shephard (1999)). This procedure avoids an extremely high dimensional state space evaluation.

By using the particle filter, $\hat{M}_t^{(b)}$ is simulated from $M_t|I_{t-1}$, and the recursive approximation of Eq. (4.3.3) hence becomes

$$\pi_{t+1}^i \propto f(r_{t+1}|M_{t+1} = m^i)\frac{1}{B}\sum_{b=1}^B P(M_{t+1} = m^i|M_t = m^{(b)}), \tag{3.4.8}$$

Therefore, we have the one-step-ahead density of:

$$\begin{aligned} f(r_t|r_1, \cdots, r_{t-1}) &= \sum_{i=1}^R f(r_t|M_t = m^i)P(M_t = m^i|I_{t-1}) \\ &\approx \frac{1}{B}\sum_{b=1}^B f(r_t|M_t = \hat{M}_t^{(b)}), \end{aligned} \tag{3.4.9}$$

then the approximate likelihood function is given below:

$$\begin{aligned} g(r_1, \cdots, r_T; \Theta) &= \prod_{t=1}^T f(r_t|r_1, \cdots, r_{t-1}) \\ &\approx \prod_{t=1}^T \left[\frac{1}{B}\sum_{b=1}^B f(r_t|M_t = \hat{M}_t^{(b)})\right]. \end{aligned} \tag{3.4.10}$$

Table 3.3: Simulation based ML estimation for the Calvet/Fisher/Thompson model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	0.014	0.041	0.042
	N_2	-0.010	0.032	0.028
	N_3	0.011	0.019	0.022
$\hat{m_2}$	N_1	0.015	0.047	0.048
	N_2	-0.011	0.031	0.032
	N_3	0.009	0.019	0.022
$\hat{\sigma_1}$	N_1	0.011	0.042	0.042
	N_2	0.009	0.029	0.030
	N_3	0.010	0.020	0.022
$\hat{\sigma_2}$	N_1	-0.012	0.046	0.047
	N_2	0.012	0.035	0.036
	N_3	0.010	0.024	0.024
$\hat{\rho}$	N_1	-0.014	0.050	0.052
	N_2	0.009	0.039	0.039
	N_3	-0.010	0.026	0.028

Note: Simulations are based on the Calvet/Fisher/Thompson model with the number of cascade levels $n = 5$, and other parameters are $m_1 = 1.2$, $m_2 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

We implement the simulation based inference with the aid of particle filter for both the Calvet/Fisher/Thompson model and the Liu/Lux model. Table 3.3 and Table 3.5 are based on the bivariate multifractal processes with cascade levels of $n = 5$; Table 3.4 and Table 3.6 are based on the bivariate multifractal processes with cascade levels of $n = 6$, for which it is almost not possible to conduct exact ML estimation due to the extremely large state space. Similar Monte Carlo experiments were constructed as the ones in the section of ML estimation. 400 Monte Carlo simulations and estimations were carried out and there were 100,000 observation generated in each simulation. Then, three different sizes of sub-sample ($N_1 = 2000$, $N_2 = 5000$, and $N_3 = 10000$) were randomly selected for estimation, and number of particles $B = 500$ was used for both models. For the Calvet/Fisher/Thompson

Table 3.4: Simulation based ML estimation for the Calvet/Fisher/Thompson model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	-0.010	0.043	0.043
	N_2	-0.011	0.032	0.031
	N_3	-0.008	0.021	0.022
$\hat{m_2}$	N_1	0.014	0.048	0.048
	N_2	-0.012	0.033	0.034
	N_3	0.009	0.023	0.023
$\hat{\sigma_1}$	N_1	-0.011	0.044	0.042
	N_2	-0.009	0.030	0.030
	N_3	0.010	0.021	0.022
$\hat{\sigma_2}$	N_1	0.009	0.047	0.048
	N_2	0.015	0.036	0.036
	N_3	-0.008	0.023	0.024
$\hat{\rho}$	N_1	0.013	0.051	0.052
	N_2	0.012	0.039	0.040
	N_3	0.009	0.027	0.028

Note: Simulations are based on the Calvet/Fisher/Thompson model with the number of cascade levels $n = 6$, and other parameters are $m_1 = 1.2$, $m_2 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

model, initial parameter values are fixed such that $m_1 = 1.2$; $m_2 = 1.4$; $\sigma_1 = 1$; $\sigma_2 = 1$; $\rho = 0.5$; For the Liu/Lux model, joint cascade level $k = 2$ and initial parameter values of $m_0 = 1.4$; $\rho = 0.5$; $\sigma_1 = 1$; $\sigma_2 = 1$ are fixed.

Examining the statistics reported from Table 3.3 to Table 3.6, we observe that the bias is minor throughout different sub-sample sizes; SD (standard deviation) and RMSE (root mean squared error) are relatively moderate and are decreasing with increasing sub-sample sizes. We also notice that there is a considerable deterioration of efficiency compared with the performance of ML estimator. One may take into account that exact ML estimation extracts the full information of the data, while, simulation based estimation uses only the information of a limited number of particles ($B = 500$ in our studies). In addition, Table

3.4 and Table 3.6 involve a higher cascade level of $n = 6$ (Table 3.1 and Table 3.2 using $n = 5$). One extra point of the empirical issue is that, simulation based maximum likelihood estimation (particle filter) is still not very economic in terms of its computation time; for example, it costs 112 hours to execute the Monte Carlo experiment in Table 3.6.

Table 3.5: Simulation based ML estimation for the Liu/Lux model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_1}$	N_1	0.013	0.039	0.041
	N_2	-0.010	0.031	0.030
	N_3	-0.009	0.020	0.019
$\hat{\sigma_1}$	N_1	-0.010	0.041	0.041
	N_2	0.019	0.028	0.029
	N_3	0.005	0.015	0.015
$\hat{\sigma_2}$	N_1	0.014	0.041	0.041
	N_2	0.010	0.028	0.028
	N_3	0.010	0.019	0.019
$\hat{\rho}$	N_1	0.013	0.040	0.039
	N_2	-0.011	0.027	0.028
	N_3	0.010	0.018	0.019

Note: Simulations are based on the Liu/Lux model with the number of cascade levels $n = 5$ and $k = 2$, other parameters are $m_0 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

Table 3.6: Simulation based ML estimation for the Liu/Lux model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_1}$	N_1	-0.011	0.038	0.041
	N_2	0.011	0.030	0.029
	N_3	0.010	0.021	0.019
$\hat{\sigma_1}$	N_1	0.015	0.040	0.042
	N_2	-0.009	0.029	0.029
	N_3	0.012	0.014	0.015
$\hat{\sigma_2}$	N_1	0.014	0.042	0.043
	N_2	-0.011	0.029	0.029
	N_3	0.008	0.019	0.020
$\hat{\rho}$	N_1	0.013	0.041	0.039
	N_2	-0.008	0.028	0.028
	N_3	0.009	0.019	0.021

Note: Simulations are based on the Liu/Lux model with the number of cascade levels $n = 6$ and $k = 2$, other parameters are $m_0 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. 400 Monte Carlo simulations have been carried out.

3.5 Generalized method of moments

In this section, we adopt the GMM (Generalized Method of Moments) approach formalized by Hansen (1982), which has become one of the most widely used methods of estimation for models in economics and finance. With analytical solutions of a set of appropriate moment conditions provided, the vector of parameters, say β, can be obtained through minimizing the differences between analytical moments and empirical moments:

$$\widehat{\beta}_T = \arg\min_{\beta \in \Theta} \bar{M}_T(\beta)' W_T \bar{M}_T(\beta). \tag{3.5.1}$$

Θ is the parameter space, $\bar{M}_T(\beta)$ stands for the vector of differences between sample moments and analytical moments, and W_T is a positive definite weighting matrix, which controls over-identification when applying GMM. Implementing Eq. (3.5.1), one typically starts with the identity matrix; then the inverse of the covariance matrix obtained from the

first round estimation is used as the weighting matrix in the next step; and this procedure continues until the estimates converge. Advocated by Newey and West (1987), the weighting matrix is estimated by the Bartlett kernel using a fixed lag length of $12\times(N/100)^{0.25}$ with N being the sample size (we found the arbitrary choice of the lag length is not much sensitive in our studies).

As is well-known, $\hat{\beta}_T$ is consistent and asymptotically Normal if suitable 'regularity conditions' are fulfilled (sets of which are detailed, for example, in Harris and Mátyás (1999)). $\hat{\beta}_T$ then converges to

$$T^{1/2}(\hat{\beta}_T - \beta_0) \sim N(0, \Xi), \tag{3.5.2}$$

with covariance matrix $\Xi = (\bar{F}_T' \bar{V}_T^{-1} \bar{F}_T)^{-1}$ in which β_0 is the true parameter vector, $\hat{V}_T^{-1} = Tvar\bar{M}_T(\beta)$ is the covariance matrix of the moment conditions, $\hat{F}_T(\beta) = \frac{\partial \bar{M}_T(\beta)}{\partial \beta}$ is the matrix of first derivatives of the moment conditions, and $\bar{V}_T$ and $\bar{F}_T$ are the constant limiting matrices to which $\hat{V}_T$ and $\hat{F}_T$ converge.

Hansen (1982) also develops the J-statistic, which refers to the value of the GMM objective function evaluated using an efficient GMM estimator $J = J(\widehat{\beta}(W^{-1}), W^{-1})$, aiming to verify whether over-identification exists. Let K, L denote the number of moment conditions used and the number of parameters to be estimated, in a well specified over-identified model with valid moment conditions, the J-statistic behaves like a chi-square random variable with degrees of freedom equal to the number of over-identifying restrictions (for $K > L$):

$$J \sim \chi^2(K - L). \tag{3.5.3}$$

Hence, if the model is mis-specified, the J-statistic will be large relative to a chi-square random variable with $K - L$ degree of freedom.

The applicability of GMM to multifractal models and the detailed regularity conditions have been discussed by Lux (2008). Markov-Switching multifractal models do not obey the

traditional definition of long memory, i.e. asymptotic power-law behavior of autocovariance functions in the limit $t \to \infty$ or divergence of the spectral density at zero, see Beran (1994), and they are rather characterized by only 'apparent' long-memory with an asymptotic hyperbolic decline of the auto-correlation of absolute powers over a finite horizon and exponential decline thereafter. Although the applicability of GMM is not hampered by this type of "long memory on a bounded interval", the proximity to 'true' long memory might rise practical concerns. In particular, if a larger cascade level is employed, say $k = 15$, the extent of the power law might exceed the size of most available daily financial data. In finite samples, application of GMM to the multifractal models could yield inferior results since usual estimates of the covariance matrix V_T might show large pre-asymptotic variation.

We follow the practical solution in Lux (2008) by using the log differences of absolute observations, together with the pertinent analytical moment conditions, i.e. transforming the observed return data r_t into τth log differences:

$$\begin{aligned} X_{t,\tau} &= \ln|r_{1,t}| - \ln|r_{1,t-\tau}| \\ &= \left(0.5\sum_{i=1}^{k}\varepsilon_{1,t}^{(i)} + 0.5\sum_{j=k+1}^{n}\varepsilon_{1,t}^{(j)} + ln|u_{1,t}|\right) - \left(0.5\sum_{i=1}^{k}\varepsilon_{1,t-\tau}^{(i)} + 0.5\sum_{j=k+1}^{n}\varepsilon_{1,t-\tau}^{(j)} + ln|u_{1,t-\tau}|\right) \\ &= 0.5\sum_{i=1}^{k}(\varepsilon_{1,t}^{(i)} - \varepsilon_{1,t-\tau}^{(i)}) + 0.5\sum_{j=k+1}^{n}(\varepsilon_{1,t}^{(j)} - \varepsilon_{1,t-\tau}^{(j)}) + (ln|u_{1,t}| - ln|u_{1,t-\tau}|) \end{aligned} \tag{3.5.4}$$

with $\varepsilon_t^{(i)} = \ln\left(M_t^{(i)}\right)$, and in the same way defining the second time series, say $Y_{t,\tau}$:

$$\begin{aligned} Y_{t,\tau} &= \ln|r_{2,t}| - \ln|r_{2,t-\tau}| \\ &= 0.5\sum_{i=1}^{k}(\varepsilon_{1,t}^{(i)} - \varepsilon_{2,t-\tau}^{(i)}) + 0.5\sum_{j=k+1}^{n}(\varepsilon_{2,t}^{(j)} - \varepsilon_{2,t-\tau}^{(j)}) + (ln|u_{2,t}| - ln|u_{2,t-\tau}|) \end{aligned} \tag{3.5.5}$$

We recognize the transformation above excluding the scale parameters σ_1 and σ_2, while estimating the scale parameters can be pursued by adding additional moment conditions (e.g. unconditional moments of r_t). In our practice, we use the second moment of empirical

data by considering each observation's contribution to the standard deviation of the sample returns.

Unlike maximum likelihood (ML) estimation, GMM does not require complete knowledge of the distribution of the data, instead, only specified moment conditions derived from an underlying model are needed. In order to exploit information from the bivariate multifractal models as much as possible, the moment conditions to be considered include two groups: the first set of conditions is obtained by deriving the moments as in the univariate time series (autocovariance), the second set considers the moment conditions for covariances of X_t and Y_t. In particular, we select moment conditions for the powers of $X_{t,\tau}$ and $Y_{t,\tau}$, i.e. moments of the raw transformed observations and squared transformed observations; these two groups of moments are described as below:

$$Cov[X_{t+\tau,\tau}, X_{t,\tau}], \quad Cov[Y_{t+\tau,\tau}, Y_{t,\tau}], \quad Cov[X^2_{t+\tau,\tau}, X^2_{t,\tau}], \quad Cov[Y^2_{t+\tau,\tau}, Y^2_{t,\tau}];$$

and

$$Cov[X_{t+\tau,\tau}, Y_{t,\tau}], \quad Cov[X_{t+\tau,\tau}, Y_{t,\tau}], \quad Cov[X^2_{t+\tau,\tau}, Y^2_{t,\tau}], \quad Cov[X^2_{t+\tau,\tau}, Y^2_{t,\tau}].$$

We focus on the time lag of $\tau = 1$, and the detailed eight analytical moments solutions of our bivariate MF model are given in the Appendix 7.1.[6] We consider two particular variants of multifractal processes, namely, one discrete version (Binomial model) and one continuous version (Lognormal model).

3.5.1 Binomial model

Our bivariate binomial MF model is characterized by binomial random draws taking the values of m_0 $(1 \leq m_0 \leq 2)$ and $2 - m_0$ with equal probability. As in the combinatorial setting, it is necessary to preserve the average mass of the interval [0, 1] during the evolution of the cascade, thus preventing non-stationarity.

[6] Moment conditions for $\tau > 1$ (say, $\tau = 5, 10, 20$) are cumbersome, and we leave them for future work.

We proceed by first applying our GMM approach to relatively small numbers of cascade levels, and eight moment conditions as in Appendix 7.1 and two additional moment conditions for the scale parameters (by considering each observation's contribution to the standard deviation). Monte Carlo studies were undertaken to compare the estimation performances between GMM and the other two approaches we applied for bivariate MF models. For the Liu/Lux model, by recalling Table 3.2 for maximum likelihood estimation, and Table 3.6 for simulation based inference (particle filter), we use the same numbers of cascade levels $n = 5$ and $n = 6$, and implement the estimations with initial parameters of $m_0 = 1.4$; $\sigma_1 = 1$; $\sigma_2 = 1$ and $\rho = 0.5$. For each scenario, 400 Monte Carlo simulations and estimations were carried out, and there were 100,000 observations generated in each simulation; subsequently three different sizes of sub-samples ($N_1 = 2000$, $N_2 = 5000$, and $N_3 = 10000$) were randomly selected for estimation. The performance is reported in Table 3.7.[7]

Monte Carlo studies were also carried out to demonstrate the performance of GMM estimator for the Calvet/Fisher/Thompson model. We simulated population sizes of $100,000$ with initial parameter values $m_1 = 1.2$; $m_2 = 1.4$; $\sigma_1 = 1$; $\sigma_2 = 1$; $\rho = 0.5$, and estimations were conducted with respect to three different sub-sample sizes ($N_1 = 2000$, $N_2 = 5000$, and $N_3 = 10000$) randomly selected; eight moment conditions (detailed analytical solutions for the Calvet/Fisher/Thompson model are given in the second section of the Appendix 7.2) and two additional moment conditions for the scale parameters are used. Table 3.8 presents the statistics of 400 simulations and estimations for the number of cascade level $n = 5$ and $n = 6$, for the purpose of comparing with the performances of ML and simulated ML with particle filter algorithm (in Table 3.1 and 3.4).

Looking through Table 3.7 and Table 3.8 (part $n = 5$), biases are moderate and approaching zero with increasing sub-sample sizes. In addition, we observe that there are considerable distortions for small sample size (N_1), but the performances are significantly

[7]Studies on other randomized parameter values have also been pursued, we omit them here due to their very similar results but keep the same parameter values as ones in the ML and SML estimations.

improving with increased sample size (N_3). This again witnesses our GMM estimator is efficient. Comparing with Table 3.1 (Calvet/Fisher/Thompson model) and Table 3.2 (Liu/Lux model), it is apparent that ML provides relatively smaller SD and RMSE than GMM owing to the fact that ML extracts all the information in the data, while the latter uses only a few moment conditions. By comparing the performances of GMM and simulation based ML approach (part $n = 6$ in Table 3.4 and Table 3.6), there is no dominating method over all scenarios. One additional note is that, maximum likelihood approach involves very intensive computation, while GMM is much faster while relatively less efficient, for example, it takes 11 hours to implement 400 Monte Carlo simulations and estimation for part $n = 5$ in Table 3.8 (10.5 hours for part $n = 5$ in Table 3.7).

Monte Carlo experiments were undertaken to further examine the performance of the GMM estimation using large numbers of cascade levels.[8] Starting with the Binomial model with the number of cascade level n = 12, joint multipliers being $k = 3$ and $k = 6$, the correlation parameter was fixed at $\rho = 0.5$, and different multipliers chosen from $m_0 = 1.2$ to 1.5 by 0.1 increment; $\sigma_1 = 1$, $\sigma_2 = 1$. For each scenario, 400 Monte Carlo simulations and estimations were carried out, and there were 100,000 observations generated in each simulation; subsequently three different sizes of sub-sample ($N_1 = 2000$, $N_2 = 5000$, and $N_3 = 10000$) were randomly selected for estimation.

Table 3.9 and Table 3.10 provide the performance of our GMM estimator across a wide set of parameters for the Binomial bivariate MF model: for $\hat{m_0}$, not only the bias but also the standard deviation (SD) and root mean squared error (RMSE) show quite encouraging behaviour. Even in the small sample sizes $N = 2000$ and $N = 5000$, the average bias of the Monte Carlo estimates is moderate throughout and practically zero for the larger sample sizes $N = 10000$. We also observe similar results for other parameters: the almost zero-bias and relatively small SD and RMSE. It is also interesting to note that our estimates are in

[8]Similar experiments can be found in Andersen and Sorensen (1996) who study the GMM estimator of standard Stochastic Volatility (SV) models.

Table 3.7: GMM estimations of the Liu/Lux (Binomial) model

		n = 5			n = 6		
$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	0.031	0.102	0.102	-0.067	0.102	0.107
	N_2	0.017	0.063	0.065	-0.024	0.063	0.067
	N_3	-0.004	0.025	0.030	-0.008	0.025	0.031
$\hat{\sigma_1}$	N_1	-0.013	0.027	0.028	-0.001	0.028	0.028
	N_2	0.009	0.018	0.019	-0.001	0.018	0.018
	N_3	0.001	0.010	0.013	-0.001	0.013	0.013
$\hat{\sigma_2}$	N_1	-0.007	0.028	0.028	-0.002	0.029	0.029
	N_2	-0.004	0.017	0.018	-0.002	0.018	0.018
	N_3	0.002	0.010	0.011	-0.001	0.013	0.013
$\hat{\rho}$	N_1	0.010	0.059	0.062	-0.001	0.061	0.066
	N_2	-0.005	0.040	0.040	-0.003	0.040	0.041
	N_3	0.004	0.025	0.028	-0.004	0.025	0.030

Note: This table shows the comparisons with ML and SML estimations for the Liu/Lux model. Simulations are based on the same parameters used for ML and SML: the number of cascade levels $n = 5$ and $n = 6$, the joint cascade level $k = 2$; $m_0 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. For each scenario, 400 Monte Carlo simulations have been carried out.

harmony with $T^{\frac{1}{2}}$ consistency. Table 3.10 also shows reductions in SD and RMSE when proceeding from $m_0 = 1.2$ to $m_0 = 1.5$.

Although the performances between GMM and maximum likelihood approach have been demonstrated, a rough comparison with ML estimates in Table 3.2 for the case $m_0 = 1.4$ (the same parameters as in ML) shows that the variability of the GMM estimator is higher than ML, and ML generates relatively smaller SD and RMSE (except for $\hat{\sigma_1}$ and $\hat{\sigma_2}$). However, one may recognize that the GMM approach only collects information of certain moments instead of full information of the data like ML; and a much larger number of cascade levels of $n = 12$ was employed in GMM; while only $n = 5$ was used by ML.

In addition, we also have witnessed each simulation and estimation over 400 iterations, and they all ended up with convergence in all scenarios, (note that, there are certain cases of non-convergence that occur in smaller sample size cases in Andersen and Sorensen (1996) when studying the GMM estimator of the SV model). We recognize there is a limited number of parameters estimated within the multifractal processes, and fixing of constant parameters in the transition probability (e.g. $b = 2$) that might be responsible for a relatively decent performance in the small sample ($N = 2000$) exercises. Nevertheless, all these results can be viewed as a positive indication of the log transformation in practice, furthermore, we can also see that there is almost no significant difference between $k = 3$ (Table 3.9) and $k = 6$ (Table 3.10); the very slight sensitivity of the estimates of m_0 with respect to the number of joint cascades might even be viewed as a very welcome phenomenon since it implies that the estimation is barely affected by the potential mis-specification of joint cascade level k.

We also check the performance of Hansen's J test to examine the issue of over-identification. Our Monte Carlo simulation setting is ideal for an investigation of the standard χ^2 test for goodness of fit of the over-identification problem. We calculate the χ^2 test statistic and evaluate the associate p value in the appropriate $\chi^2(K - L)$ distribution. The findings are qualitatively similar across different scenarios, so, for the sake of brevity, we present only one case of $m_0 = 1.3$. Figure 3.4 displays the fraction of p values that fall within the 2% interval fractiles for different sample sizes. The three graphs from top to bottom correspond to three different sample sizes: $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. We observe that increasing the sample size leads to a leftward shift in the entire distribution. Indeed, the leftward shifting suggests that the size distortion of the χ^2 test statistic is growing as the sample size expands. In particular, the fractiles from 1% to 10% shed light on the size of these goodness-of-fit test from the asymptotic 1% to 10% levels, and the mass located in these fractiles increases.

Table 3.8: GMM estimation of the Calvet/Fisher/Thompson model

		n = 5			n = 6		
$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$
$\hat{m_1}$	N_1	0.037	0.110	0.118	0.017	0.126	0.134
	N_2	-0.015	0.085	0.091	-0.010	0.090	0.096
	N_3	0.006	0.062	0.067	-0.007	0.060	0.068
$\hat{m_2}$	N_1	0.054	0.130	0.137	-0.050	0.136	0.144
	N_2	-0.025	0.084	0.089	0.033	0.104	0.108
	N_3	-0.012	0.058	0.062	0.004	0.058	0.062
$\hat{\sigma_1}$	N_1	0.021	0.108	0.117	-0.040	0.116	0.124
	N_2	-0.029	0.083	0.088	-0.036	0.080	0.084
	N_3	0.016	0.050	0.053	-0.010	0.051	0.053
$\hat{\sigma_2}$	N_1	-0.033	0.093	0.098	0.031	0.100	0.106
	N_2	-0.015	0.063	0.068	-0.017	0.066	0.070
	N_3	0.009	0.025	0.028	-0.009	0.024	0.027
$\hat{\rho}$	N_1	0.024	0.105	0.117	-0.038	0.116	0.123
	N_2	0.010	0.074	0.079	0.030	0.081	0.086
	N_3	-0.004	0.046	0.049	0.011	0.050	0.054

Note: This table shows the comparisons with ML and SML estimation for the Calvet/Fisher/Thompson model. Simulations are based on the same parameters used for ML and SML: the number of cascade levels $n = 5$ and $n = 6$; $m_1 = 1.2$, $m_2 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 3.9: GMM estimation of the bivariate MF (Binomial) model

		$\hat{m}_0$			$\hat{\rho}$			$\hat{\sigma_1}$			$\hat{\sigma_2}$		
		Bias	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
	N_1	-0.081	0.125	0.153	0.007	0.065	0.072	-0.005	0.073	0.073	0.001	0.074	0.075
$m_0 = 1.20$	N_2	0.066	0.126	0.140	0.002	0.046	0.047	0.002	0.040	0.046	0.002	0.045	0.047
	N_3	0.042	0.098	0.108	-0.001	0.026	0.028	0.001	0.030	0.031	0.001	0.034	0.034
	N_1	0.110	0.128	0.175	0.019	0.070	0.071	0.017	0.108	0.110	-0.002	0.112	0.112
$m_0 = 1.30$	N_2	0.043	0.101	0.137	-0.019	0.050	0.051	-0.011	0.070	0.073	0.002	0.069	0.069
	N_3	-0.022	0.062	0.068	0.007	0.031	0.033	0.000	0.050	0.054	0.003	0.050	0.053
	N_1	-0.091	0.133	0.150	0.012	0.075	0.078	0.020	0.154	0.155	-0.010	0.159	0.161
$m_0 = 1.40$	N_2	0.035	0.066	0.071	0.005	0.052	0.051	-0.003	0.098	0.098	0.001	0.093	0.094
	N_3	-0.004	0.035	0.038	-0.008	0.033	0.035	0.003	0.071	0.074	-0.001	0.066	0.070
	N_1	-0.064	0.095	0.097	-0.017	0.080	0.085	-0.025	0.204	0.205	-0.014	0.191	0.192
$m_0 = 1.50$	N_2	-0.017	0.040	0.032	-0.010	0.056	0.057	0.016	0.130	0.135	-0.004	0.118	0.120
	N_3	-0.003	0.022	0.026	-0.010	0.034	0.035	-0.004	0.093	0.097	0.003	0.087	0.089

Note: All simulations are based on the bivariate multifractal process with the number of cascade levels equal to 12, $k = 3$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$; eight moment conditions as in Appendix are used. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 3.10: GMM estimation of the bivariate MF (Binomial) model

		$\hat{m}_0$			$\hat{\rho}$			$\hat{\sigma_1}$			$\hat{\sigma_2}$		
		Bias	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
$m_0 = 1.20$	N_1	0.089	0.134	0.155	0.021	0.072	0.073	-0.015	0.047	0.046	0.001	0.043	0.043
	N_2	0.060	0.124	0.143	0.009	0.045	0.045	0.009	0.030	0.031	0.000	0.027	0.028
	N_3	-0.029	0.100	0.112	0.000	0.030	0.032	-0.003	0.017	0.019	0.000	0.019	0.020
$m_0 = 1.30$	N_1	0.080	0.154	0.169	0.013	0.084	0.085	-0.025	0.066	0.067	0.010	0.064	0.065
	N_2	0.049	0.110	0.124	0.005	0.051	0.053	0.007	0.045	0.045	-0.008	0.040	0.041
	N_3	-0.011	0.069	0.071	-0.002	0.034	0.035	-0.001	0.028	0.029	0.001	0.030	0.030
$m_0 = 1.40$	N_1	0.080	0.130	0.139	-0.025	0.085	0.088	-0.020	0.095	0.096	-0.009	0.079	0.078
	N_2	-0.013	0.066	0.069	-0.020	0.048	0.050	0.011	0.061	0.063	0.010	0.055	0.056
	N_3	-0.001	0.034	0.035	-0.011	0.034	0.036	-0.001	0.040	0.041	0.003	0.038	0.041
$m_0 = 1.50$	N_1	-0.054	0.081	0.087	0.008	0.090	0.093	0.014	0.126	0.128	-0.022	0.116	0.117
	N_2	0.021	0.041	0.043	-0.016	0.055	0.058	-0.007	0.079	0.080	0.009	0.080	0.081
	N_3	0.003	0.024	0.025	-0.019	0.040	0.043	-0.001	0.057	0.057	0.003	0.060	0.062

Note: All simulations are based on the bivariate multifractal process with the number of cascade levels equal to 12, $k = 6$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, and eight moment conditions as in the Appendix are used. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

3.5.2 Lognormal model

To accommodate our bivariate multifractal model with continuous distribution of volatility components, we next turn to the Lognormal model. This means that when a new multiplier M_t is needed at any cascade level, it will be determined via a random draw from a Lognormal distribution with parameters λ and σ_m, i.e.,

$$-log_2 M_t \sim N(\lambda, \sigma_m^2). \tag{3.5.6}$$

We adopt the one used in Mandelbrot et al. (1997) by assigning $E[M_t^{(i)}] = 0.5$ on the Lognormal model ($E[M_t^{(i)}] = 1$ is used in Lux (2008)).[9] It implies:

$$exp[-\lambda \ln 2 + 0.5\sigma_m^2 (\ln 2)^2] = 0.5, \tag{3.5.7}$$

which leads to

$$\sigma_m^2 = 2(\lambda - 1)/ln2, \tag{3.5.8}$$

Hence, we end up with a one-parameter family of multifractal models as in the Binomial case. Unlike the Binomial model, multifractal processes with continuous distribution of volatility components imply an infinite dimension of the transitional matrix, and the exact form of likelihood function can not be identified explicitly. Therefore, the maximum likelihood approach is not applicable to the Lognormal model.[10] GMM provides a solution for estimating multifractal processes with continuous state spaces. Moment conditions for the Lognormal model are given in Appendix. Note that the admissible parameter space for the location parameter λ is $\lambda \in [1, \infty)$ where in the borderline case $\lambda = 1$ the volatility process collapses to a constant (as $m_0 = 1$ in Binomial model).

Analogously, the Calvet/Fisher/Thompson model can also be extended to the continuous case by allowing two Lognormal distributions for the volatility components of the bivariate

[9] To avoid the non-stationarity in the data generating process, a factor of 2^n is multiplied for compensating the mean of multipliers being 0.5.

[10] Simulation based maximum likelihood could be applicable, and we leave it for future work.

multifractal processes, namely, $-log_2 M_1 \sim N(\lambda_1, \sigma^2_{m1})$, and $-log_2 M_2 \sim N(\lambda_2, \sigma^2_{m2})$. We have also derived the pertinent moment conditions in the Appendix.

The Monte Carlo study reported in Table 3.11 and Table 3.12, covers parameter values $\lambda = 1.10$, 1.20, 1.30 and 1.40; $\sigma_1 = 1$, $\sigma_2 = 1$ and $\rho = 0.5$. The numbers of joint multiplier levels, the population and sub-sample sizes are set as in the Binomial case above. As can be seen, results are not too different from those obtained with the binomial BMF model: biases are very close to zero again; SD and RMSE are moderate and decrease with increase in the sub-sample size. We also observe the similar behaviour in the Calvet/Fisher/Thompson (Lognormal) model, cf. Table 3.13. All in all, these results from both the Binomial and Lognormal Monte Carlo simulations and estimations show that GMM works quite well for the bivariate multifractal process, with both discrete and continuous state spaces.

We then examine Hansen's J test for the bivariate Lognormal model. The case of $\lambda = 1.3$ is presented due to the qualitative similarity across different scenarios. Figure 3.5 displays the fraction of p values that fall within the 2% interval fractiles. Three graphs from top to bottom corresponding to three different sample sizes: $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. We again observe that increasing the sample size leads to a leftward shift within the entire distribution.

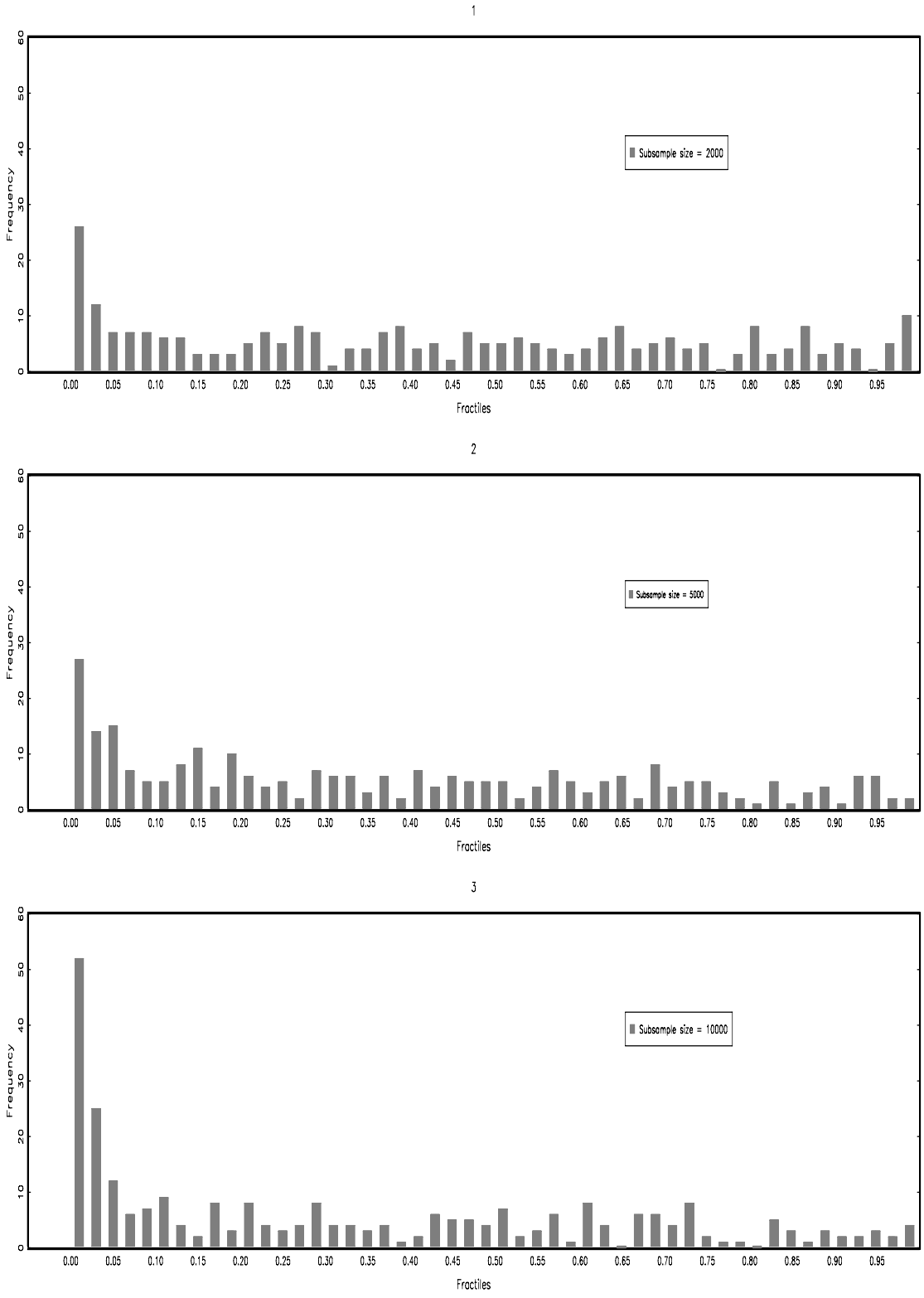

Figure 3.4: The distribution of p values for the test of over-identification restrictions for a binomial BMF model.

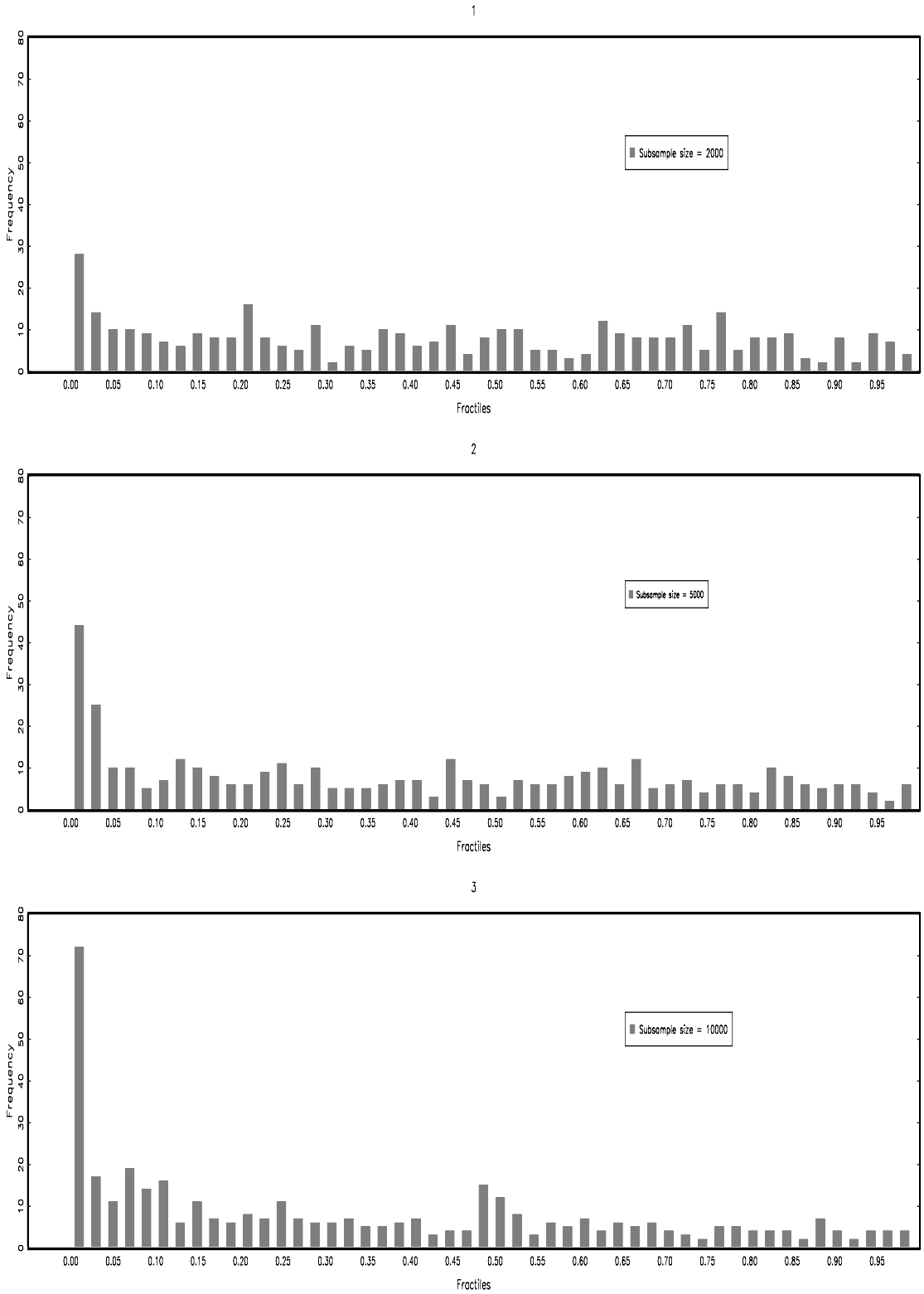

Figure 3.5: The distribution of p values for the test of over-identification restrictions for a Lognormal BMF model.

Table 3.11: GMM estimation of the bivariate MF (Lognormal) model

		$\hat{\lambda}$			$\hat{\rho}$			$\hat{\sigma_1}$			$\hat{\sigma_2}$		
		Bias	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
	N_1	0.032	0.054	0.059	0.020	0.073	0.077	0.023	0.105	0.111	-0.014	0.110	0.119
$\lambda = 1.10$	N_2	0.019	0.035	0.037	0.011	0.044	0.048	-0.010	0.070	0.076	-0.016	0.073	0.076
	N_3	-0.003	0.025	0.026	-0.006	0.039	0.039	-0.005	0.053	0.057	-0.005	0.047	0.053
	N_1	0.031	0.058	0.061	0.002	0.080	0.084	0.028	0.161	0.170	0.024	0.180	0.180
$\lambda = 1.20$	N_2	-0.018	0.036	0.039	-0.009	0.054	0.054	-0.015	0.109	0.117	-0.017	0.122	0.131
	N_3	-0.05	0.025	0.027	-0.015	0.035	0.040	-0.009	0.089	0.094	-0.009	0.087	0.091
	N_1	0.035	0.065	0.070	0.017	0.092	0.096	-0.040	0.238	0.249	-0.041	0.243	0.252
$\lambda = 1.30$	N_2	0.011	0.035	0.039	-0.024	0.056	0.061	0.027	0.173	0.181	-0.029	0.164	0.170
	N_3	-0.005	0.028	0.030	0.021	0.039	0.051	0.014	0.129	0.133	0.019	0.111	0.120
	N_1	0.037	0.064	0.068	0.038	0.092	0.097	-0.043	0.299	0.314	0.051	0.294	0.305
$\lambda = 1.40$	N_2	0.014	0.040	0.042	0.040	0.067	0.079	0.031	0.203	0.217	-0.030	0.212	0.227
	N_3	-0.009	0.028	0.032	-0.033	0.042	0.055	-0.021	0.163	0.176	0.026	0.153	0.159

Note: All simulations are based on the bivariate multifractal process with the whole number of cascade levels equal to 12, $k = 3$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, and eight moment conditions as in the Appendix are used. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 3.12: GMM estimation of the bivariate MF (Lognormal) Model

		$\hat{\lambda}$			$\hat{\rho}$			$\hat{\sigma_1}$			$\hat{\sigma_2}$		
		Bias	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*	*Bias*	*SD*	*RMSE*
$\lambda = 1.10$	N_1	0.037	0.054	0.060	-0.012	0.087	0.087	0.019	0.082	0.085	-0.010	0.085	0.091
	N_2	0.014	0.035	0.037	-0.009	0.047	0.052	0.018	0.050	0.054	-0.002	0.057	0.061
	N_3	-0.003	0.023	0.022	0.009	0.036	0.036	-0.004	0.037	0.037	-0.005	0.030	0.037
$\lambda = 1.20$	N_1	-0.021	0.057	0.064	0.017	0.089	0.089	0.028	0.150	0.157	-0.022	0.140	0.148
	N_2	0.015	0.036	0.038	-0.012	0.050	0.058	-0.013	0.086	0.095	-0.009	0.080	0.088
	N_3	-0.007	0.025	0.025	0.015	0.037	0.042	-0.005	0.071	0.071	-0.002	0.062	0.069
$\lambda = 1.30$	N_1	-0.018	0.061	0.066	0.037	0.104	0.107	-0.014	0.301	0.301	-0.031	0.205	0.222
	N_2	0.010	0.035	0.040	-0.030	0.055	0.064	-0.009	0.212	0.212	-0.023	0.135	0.145
	N_3	0.011	0.027	0.029	-0.022	0.044	0.055	-0.008	0.154	0.154	-0.015	0.089	0.100
$\lambda = 1.40$	N_1	-0.026	0.067	0.071	-0.051	0.109	0.115	0.034	0.299	0.302	-0.046	0.291	0.307
	N_2	-0.004	0.041	0.043	0.040	0.067	0.078	-0.022	0.216	0.217	-0.024	0.218	0.230
	N_3	0.015	0.028	0.033	-0.033	0.042	0.062	0.019	0.160	0.161	-0.013	0.176	0.185

Note: All simulations are based on the bivariate multifractal process with the whole number of cascade levels equal to 12, $k = 6$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$, and eight moment conditions as in the Appendix are used. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 3.13: GMM estimation of the Calvet/Fisher/Thompson (Lognormal) model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{\lambda_1}$	N_1	-0.031	0.110	0.111
	N_2	0.013	0.067	0.067
	N_3	0.008	0.032	0.032
$\hat{\lambda_2}$	N_1	-0.043	0.120	0.117
	N_2	-0.020	0.056	0.059
	N_3	0.011	0.031	0.033
$\hat{\sigma_1}$	N_1	0.027	0.108	0.110
	N_2	-0.020	0.080	0.082
	N_3	-0.019	0.061	0.063
$\hat{\sigma_2}$	N_1	0.031	0.097	0.099
	N_2	-0.017	0.077	0.078
	N_3	0.011	0.065	0.068
$\hat{\rho}$	N_1	0.020	0.095	0.102
	N_2	-0.018	0.065	0.069
	N_3	0.007	0.046	0.047

Note: This table shows the GMM estimation for the Calvet/Fisher/Thompson model with Lognormal distribution of the volatility components. Simulations are based on the same parameters of: the number of cascade levels $n = 12$; $\lambda_1 = 1.2$, $\lambda_2 = 1.4$, $\rho = 0.5$, $\sigma_1 = 1$, $\sigma_2 = 1$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. For each scenario, 400 Monte Carlo simulations have been carried out.

3.6 Empirical estimates

In this section, we present empirical applications of bivariate MF models. We consider daily data for a collection of stock exchange indices: Dow Jones Composite 65 Average Index and $NIKKEI$ 225 Average Index (DOW/NIK, January 1969 - October 2004); two foreign exchange rates, U.S. Dollar to British Pound, German Mark to British Pound (US/DM, March 1973 - February 2004); and U.S. 1-year and 2-year treasury constant maturity bond rates ($TB1/TB2$, June 1976 - October 2004), where the first symbol inside these parentheses designates the short notation for the corresponding time series, followed by the starting and ending dates for the sample at hand.[11]

Figure 3.6 to 3.8 provide the plots of the six empirical daily time series (stock exchange indices, foreign exchange rates and U.S. treasury constant maturity bond rates) p_t and return r_t calculated as

$$r_t = 100 \times [\ln(p_t) - \ln(p_{t-1})].$$

Our study in this section covers the empirical results of bivariate multifractal models (the Calvet/Fisher/Thompson model and the Liu/Lux model) obtained by using maximum likelihood (ML), simulation based ML and GMM approaches. Table 3.14 reports empirical estimates of the Calvet/Fisher/Thompson model with the number of cascade levels n from 1 to 5 via ML approach. We observe that the maximized log-likelihood values dramatically increase (by more than 1400 for DOW/NIK, about 2000 for US/DM and about 1700 for $TB1/TB2$) when the number of cascades n increases from 1 to 5, and it signals that it might be preferable to use a larger number of n ($n = 5$ is the computation limit for ML). Therefore, for the Liu/Lux (binomial) model, Table 3.15 presents the empirical estimation results via ML for fixed $n = 5$ but various joint cascade levels k from 1 to 4. We find that the maximized log-likelihood values among different k are more flat, but there are slightly

[11] For foreign exchange rate US/DM after January 2000 is transformed from Euro; The U.S. one and two-year treasury constant maturity rates have been converted to bond prices before calculating returns.

higher values when $k = 2$ for DOW/NIK and $TB1/TB2$, $k = 1$ for US/DM.

Next, we move to the simulation based maximum likelihood estimation using particle filter (with the number of particles of $B = 500$). Table 3.16 gives the empirical estimates of the Calvet/Fisher/Thompson model with a choice of $n = 8$ as in Calvet et al. (2006). For the Liu/Lux (binomial) model, Table 3.17 gives the empirical results for a range of joint cascade levels k from 1 to 7 ($n = 8$). To specify an optimal choice of cascade levels for different assets, we also provide a heuristic method as below:

(1) For each bivariate time series, we take its equal-weight portfolio, for an example of DOW/NIK, the stock exchange portfolio is $0.5 \times DOW + 0.5 \times NIK$.

(2) By using the GPH approach by Geweke and Porter-Hudak (1983) (see details in Section 2.2.2), the empirical long memory estimator $\hat{d}$ for the absolute returns of equally-weighted portfolio is calculated.

(3) Based on the empirical estimates with different numbers of cascade levels, 100 simulations are conducted for each asset; and long memory parameter d is calculated for each simulated equally-weighted portfolio.

(4) We then select the case of the cascade level whose mean value of $\hat{d}$ is close to the empirical GPH estimator d.

In Table 3.17, we find the case of $k = 3$ whose arithmetic mean of $\hat{d}$ is relatively close to the empirical one for DOW/NIK, also $k = 2$ for US/DM and $k = 5$ for $TB1/TB2$ respectively.

Table 3.18 to Table 3.20 are GMM estimates of the Calvet/Fisher/Thompson model. By using the specification method introduced above, we observe these cases which have relatively close $\hat{d}$ to empirical ones: $n = 10$ for DOW/NIK, $n = 12$ for US/DM and $n = 10$ for $TB1/TB2$.

For the GMM application to Liu/Lux models, we take the choices for the number of cascade levers n obtained from the Calvet/Fisher/Thompson model as references. Let us

begin with the Binomial model, by fixing $n = 10$, Table 3.21 presents the empirical estimates for the stock exchange indices (DOW/NIK) with different numbers of joint cascade levels k from 1 to 9, as well as the corresponding mean values of simulated GPH estimator $\hat{d}$. We observe that the case of $k = 4$ ($\hat{d} = 0.249$) has the relatively closest value of $\hat{d}$ to the empirical one. Since it rather deviates from the empirical value of $d = 0.295$, we then try large numbers of n in order to minimize this distance. Table 3.22 provides the empirical results for a range of n from 5 to 20 with fixing $k = 4$, and we find that the case of $n = 14$ gives the closest value of $\hat{d}$ to the empirical GPH estimator. Table 3.23 reports the results for the exchange rates US/DM for a range of joint cascade level of k from 1 to 11 ($n = 12$). We find that the deviations between the simulated $\hat{d}$ and empirical one are within the reasonable size and the case having close values of $\hat{d}$ to empirical one is $k = 6$. Table 3.24 gives GMM estimates of U.S. bond rates $TB1/TB2$ for various joint cascade levels k from 1 to 9 ($n = 10$), and we find that the cases having close values of $\hat{d}$ to empirical one are $k = 3$ and $k = 4$.

Since GMM approach allows multifractal processes with continuous state space applicable, we also have pursued the empirical studies for the Lognormal model introduced in Section 3.5.2. Table 3.25 presents the empirical results for DOW/NIK with different numbers of joint cascade levels k from 1 to 9 ($n = 10$); like in the Binomial case, $\hat{d}$ at $k = 4$ is comparatively close to the empirical one but with considerable deviation. Again, we also report the empirical estimates for different numbers of cascade levels n from 5 to 20 but fixing $k = 4$ in Table 3.26 and it suggests us an optimal choice of cascade levels for the stock indices is $n = 15$. Table 3.27 gives the empirical results for the exchange rates US/DM with different numbers of joint cascade levels k from 1 to 11 ($n = 12$); we observe the case of $k = 6$ has a close value to empirical one. Table 3.28 reports the empirical estimates for U.S. bond rates $TB1/TB2$ by fixing $n = 10$ but various numbers of joint cascade k from 1 to 9, and it provides that the case having close values of $\hat{d}$ to empirical ones is $k = 3$.

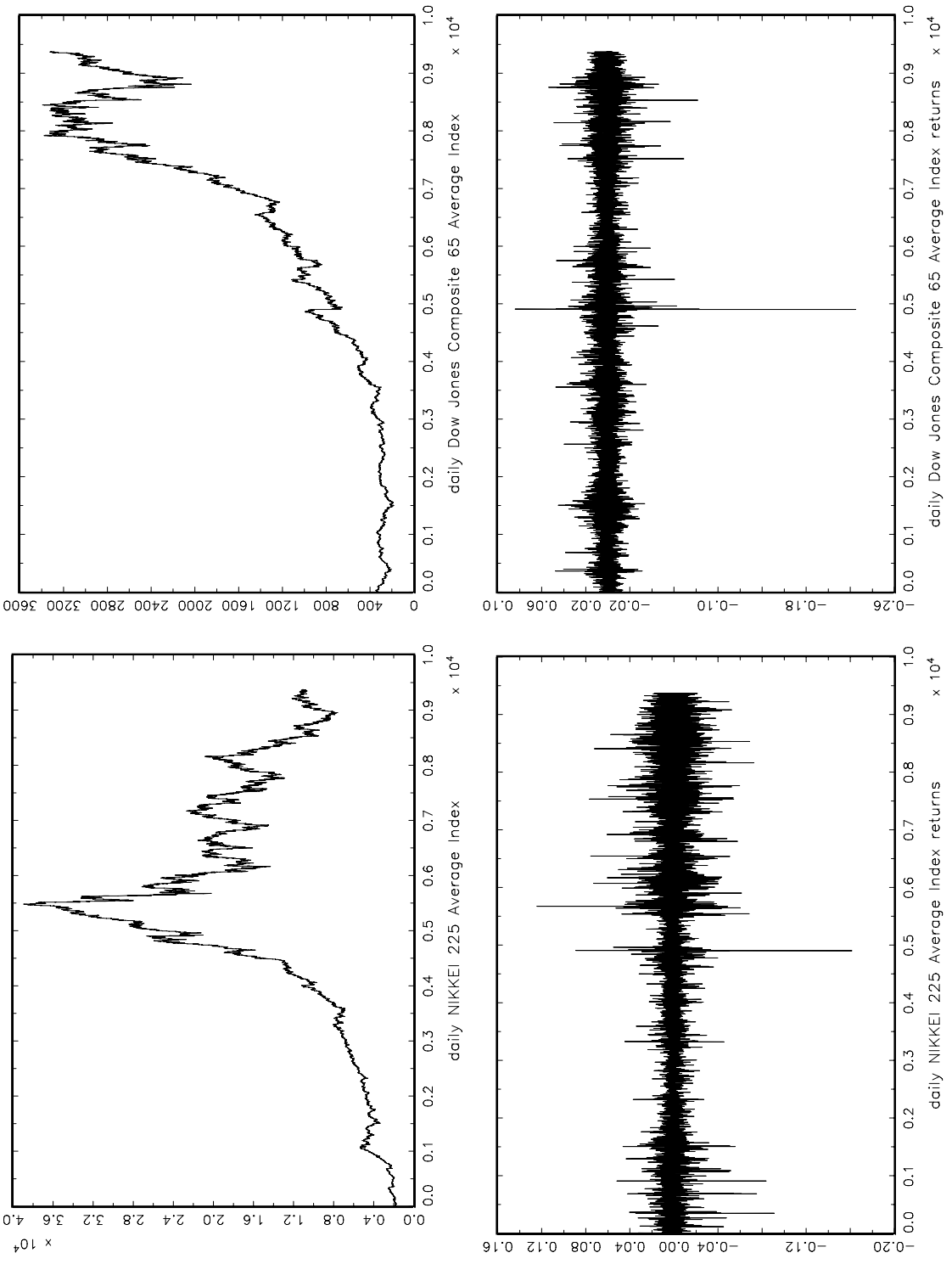

Figure 3.6: Empirical time series: *Dow* and *Nik*.

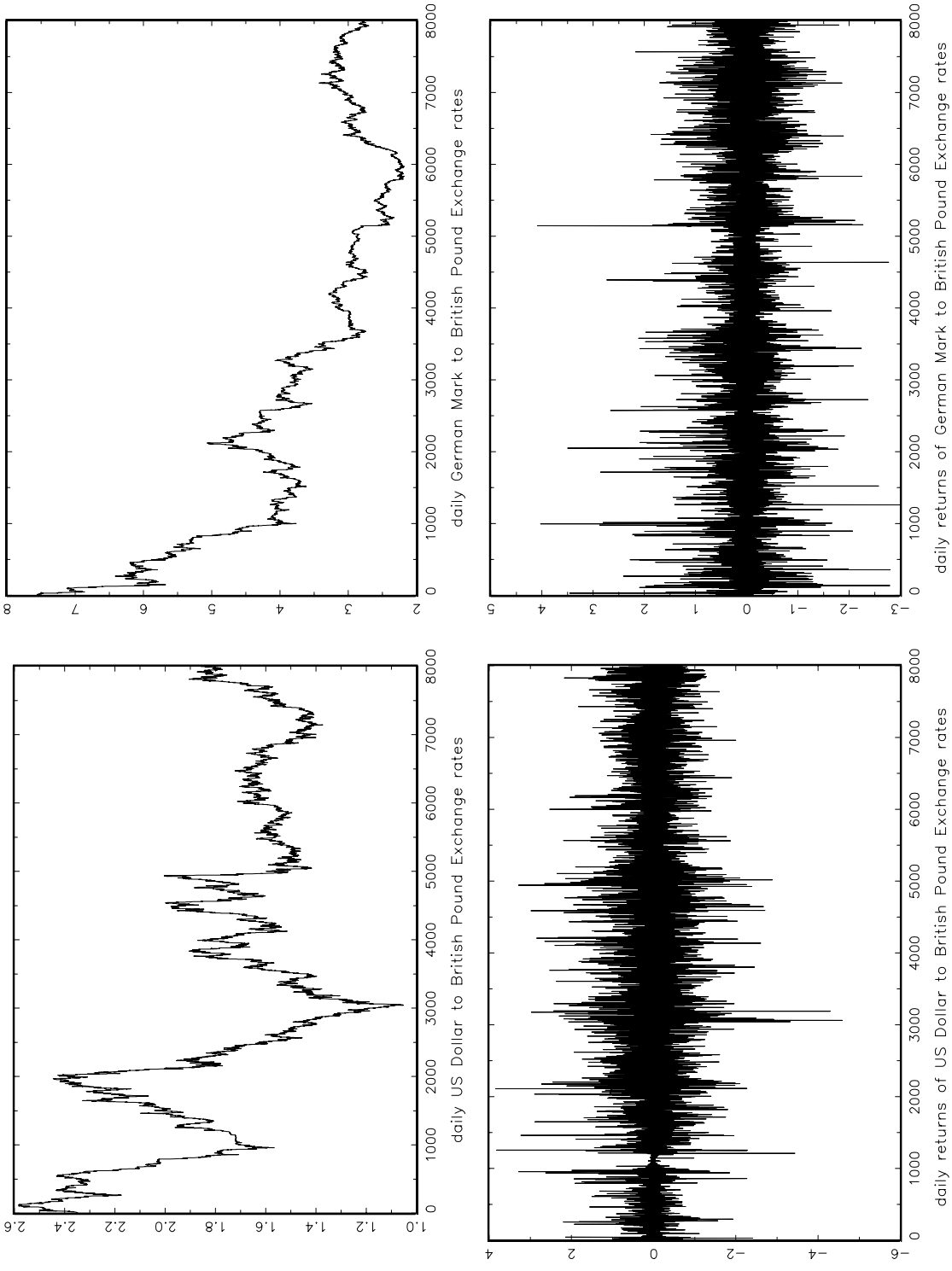

Figure 3.7: Empirical time series: US and DM.

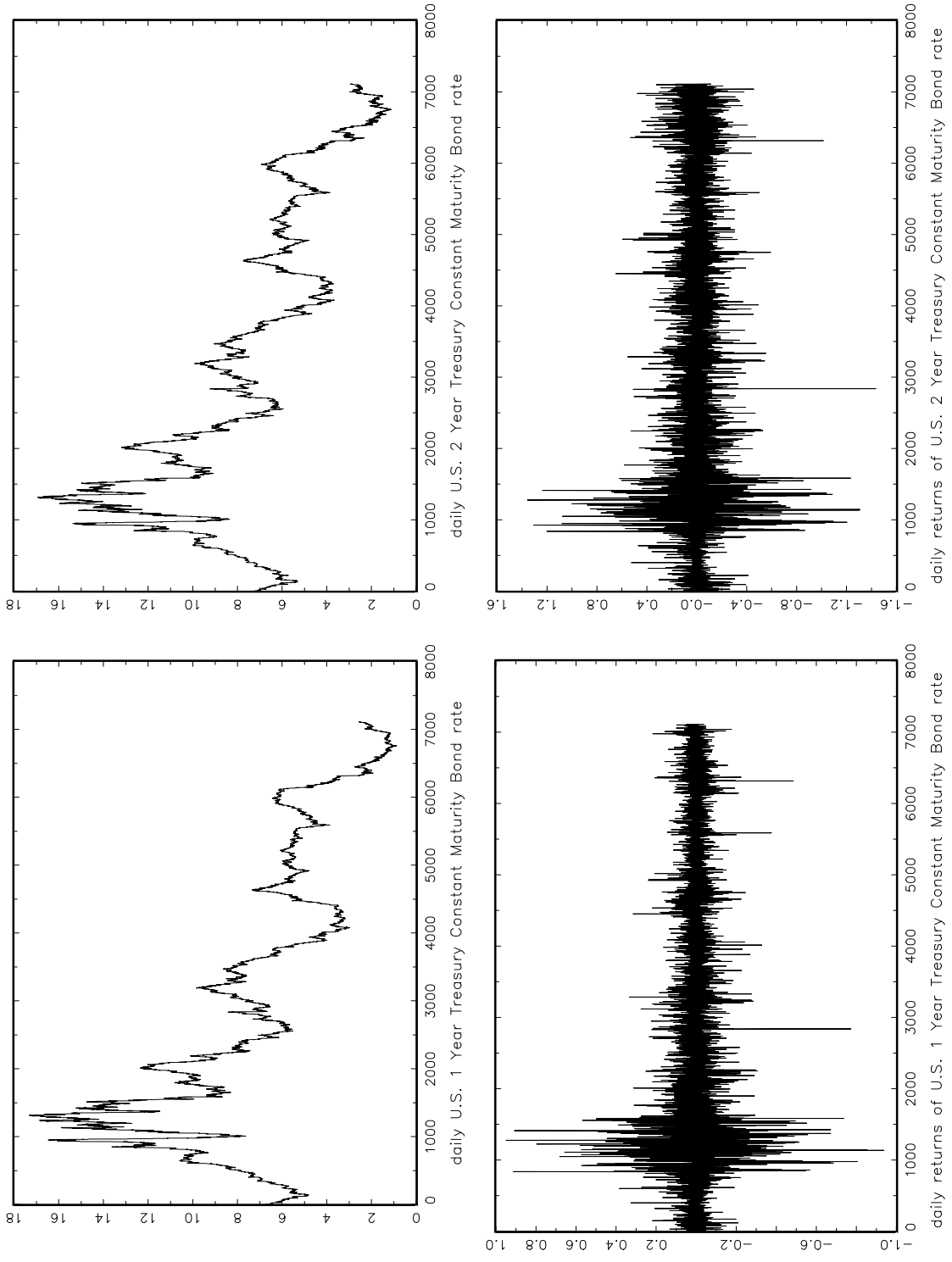

Figure 3.8: Empirical time series: $TB1$ and $TB2$.

Table 3.14: ML estimates of the Calvet/Fisher/Thompson model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$
Dow/Nik					
$\hat{m_1}$	1.777	1.723	1.632	1.550	1.521
	(0.029)	(0.026)	(0.028)	(0.029)	(0.028)
$\hat{m_2}$	1.561	1.507	1.422	1.349	1.301
	(0.024)	(0.024)	(0.023)	(0.028)	(0.022)
$\hat{\sigma_1}$	1.141	1.133	1.145	1.149	1.152
	(0.026)	(0.028)	(0.027)	(0.026)	(0.028)
$\hat{\sigma_2}$	0.928	0.945	0.916	0.920	0.987
	(0.035)	(0.032)	(0.031)	(0.030)	(0.030)
$\hat{\rho}$	0.079	0.084	0.085	0.084	0.089
	(0.025)	(0.027)	(0.022)	(0.024)	(0.018)
$\ln L$	-19007.669	-18897.703	-18397.651	-18037.670	-17559.334
US/DM ~ £					
$\hat{m_1}$	1.643	1.584	1.519	1.459	1.430
	(0.012)	(0.009)	(0.014)	(0.014)	(0.016)
$\hat{m_2}$	1.630	1.584	1.468	1.417	1.335
	(0.015)	(0.008)	(0.014)	(0.016)	(0.011)
$\hat{\sigma_1}$	0.673	0.644	0.656	0.642	0.613
	(0.022)	(0.022)	(0.023)	(0.017)	(0.024)
$\hat{\sigma_2}$	0.540	0.522	0.513	0.534	0.521
	(0.018)	(0.017)	(0.018)	(0.015)	(0.018)
$\hat{\rho}$	0.014	0.015	0.016	0.015	0.014
	(0.010)	(0.009)	(0.007)	(0.013)	(0.012)
$\ln L$	-18754.571	-18371.033	-17573.378	-17175.534	-16826.873
TB2/TB1					
$\hat{m_1}$	1.811	1.793	1.783	1.728	1.644
	(0.030)	(0.032)	(0.031)	(0.029)	(0.030)
$\hat{m_2}$	1.802	1.776	1.704	1.641	1.604
	(0.031)	(0.031)	(0.029)	(0.029)	(0.028)
$\hat{\sigma_1}$	1.321	1.310	1.339	1.293	1.289
	(0.041)	(0.040)	(0.039)	(0.044)	(0.043)
$\hat{\sigma_2}$	1.308	1.299	1.292	1.298	1.312
	(0.047)	(0.050)	(0.049)	(0.054)	(0.052)
$\hat{\rho}$	0.729	0.719	0.711	0.712	0.716
	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)
$\ln L$	-16457.126	-15925.302	-15649.534	-15120.155	-14773.443

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n; $\ln L$ represents log likelihood value. Numbers in parenthesis are standard errors.

Table 3.15: ML estimates of the Liu/Lux (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$
Dow/Nik				
$\hat{m_0}$	1.552	1.508	1.587	1.555
	(0.016)	(0.013)	(0.015)	(0.017)
$\hat{\sigma_1}$	1.094	1.121	1.120	1.112
	(0.022)	(0.019)	(0.016)	(0.020)
$\hat{\sigma_2}$	0.934	0.906	0.905	0.893
	(0.018)	(0.018)	(0.018)	(0.020)
$\hat{\rho}$	0.212	0.202	0.194	0.180
	(0.020)	(0.020)	(0.021)	(0.019)
$\ln L$	-18301.900	-18291.380	-18313.255	-18330.861
$US/DM \sim £$				
$\hat{m_0}$	1.458	1.451	1.442	1.420
	(0.018)	(0.016)	(0.016)	(0.017)
$\hat{\sigma_1}$	0.877	0.898	0.0801	0.789
	(0.006)	(0.006)	(0.005)	(0.007)
$\hat{\sigma_2}$	0.681	0.700	0.672	0.693
	(0.007)	(0.005)	(0.008)	(0.008)
$\hat{\rho}$	0.018	0.014	0.011	0.007
	(0.017)	(0.019)	(0.019)	(0.020)
$\ln L$	-16246.194	-16301.766	-16328.546	-16316.936
$TB2/TB1$				
$\hat{m_0}$	1.678	1.671	1.667	1.671
	(0.035)	(0.033)	(0.035)	(0.032)
$\hat{\sigma_1}$	1.199	1.206	1.202	1.194
	(0.023)	(0.025)	(0.022)	(0.023)
$\hat{\sigma_2}$	1.193	1.190	1.187	1.196
	(0.021)	(0.020)	(0.020)	(0.018)
$\hat{\rho}$	0.715	0.711	0.714	0.713
	(0.015)	(0.018)	(0.015)	(0.016)
$\ln L$	-16606.330	-16497.429	-16588.104	-16535.378

Note: Each column corresponds to the empirical estimate with different numbers of joint cascade level k (n =5); $\ln L$ represents log likelihood value. Numbers in parenthesis are standard errors.

Table 3.16: SML estimates of the Calvet/Fisher/Thompson model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
Dow/Nik								
$\hat{m_1}$	1.726	1.703	1.670	1.657	1.622	1.605	1.554	1.538
	(0.104)	(0.099)	(0.063)	(0.093)	(0.087)	(0.129)	(0.082)	(0.102)
$\hat{m_2}$	1.592	1.565	1.534	1.492	1.471	1.421	1.404	1.375
	(0.117)	(0.078)	(0.082)	(0.067)	(0.064)	(0.113)	(0.099)	(0.079)
$\hat{\sigma_1}$	1.110	1.104	1.121	1.132	1.147	1.154	1.137	1.121
	(0.037)	(0.039)	(0.038)	(0.038)	(0.038)	(0.035)	(0.043)	(0.035)
$\hat{\sigma_2}$	0.987	0.994	1.113	0.971	0.993	0.894	0.998	1.097
	(0.042)	(0.036)	(0.043)	(0.044)	(0.039)	(0.040)	(0.031)	(0.050)
$\hat{\rho}$	0.164	0.182	0.140	0.173	0.166	0.156	0.150	0.143
	(0.018)	(0.027)	(0.017)	(0.024)	(0.016)	(0.018)	(0.019)	(0.015)
$\ln L$	-22772.84	-22525.29	-21623.89	-20761.96	-19590.49	-19032.88	-18735.87	-18818.66
$US/DM \sim £$								
$\hat{m_1}$	1.607	1.602	1.589	1.577	1.550	1.421	1.409	1.400
	(0.056)	(0.062)	(0.058)	(0.063)	(0.062)	(0.060)	(0.061)	(0.061)
$\hat{m_2}$	1.589	1.560	1.549	1.537	1.521	1.515	1.485	1.439
	(0.038)	(0.042)	(0.042)	(0.041)	(0.039)	(0.037)	(0.042)	(0.042)
$\hat{\sigma_1}$	0.688	0.690	0.645	0.671	0.680	0.644	0.639	0.685
	(0.089)	(0.089)	(0.092)	(0.095)	(0.090)	(0.085)	(0.091)	(0.091)
$\hat{\sigma_2}$	0.732	0.894	0.793	0.741	0.693	0.698	0.708	0.787
	(0.087)	(0.089)	(0.089)	(0.088)	(0.092)	(0.091)	(0.087)	(0.092)
$\hat{\rho}$	0.026	0.022	0.023	0.021	0.024	0.029	0.020	0.021
	(0.011)	(0.020)	(0.017)	(0.016)	(0.020)	(0.018)	(0.013)	(0.017)
$\ln L$	-22103.75	-21784.11	-20112.25	-19882.73	-18972.05	-18445.38	-17216.59	-17119.23
$TB2/TB1$								
$\hat{m_1}$	1.815	1.790	1.771	1.752	1.728	1.695	1.650	1.607
	(0.073)	(0.077)	(0.081)	(0.083)	(0.083)	(0.080)	(0.084)	(0.081)
$\hat{m_2}$	1.790	1.761	1.733	1.710	1.672	1.534	1.430	1.485
	(0.102)	(0.104)	(0.101)	(0.105)	(0.105)	(0.105)	(0.107)	(0.101)
$\hat{\sigma_1}$	1.534	1.430	1.485	1.426	1.342	1.790	1.761	1.733
	(0.075)	(0.079)	(0.071)	(0.071)	(0.072)	(0.072)	(0.071)	(0.070)
$\hat{\sigma_2}$	1.610	1.545	1.558	1.563	1.556	0.695	0.680	0.669
	(0.052)	(0.051)	(0.053)	(0.057)	(0.058)	(0.055)	(0.059)	(0.051)
$\hat{\rho}$	0.695	0.680	0.669	0.658	0.644	1.610	1.545	1.558
	(0.038)	(0.039)	(0.033)	(0.033)	(0.032)	(0.035)	(0.036)	(0.036)
$\ln L$	-17834.38	-17510.80	-17037.89	-16819.24	-16321.50	-15923.89	-15761.96	-15590.44

Note: Each column corresponds to the empirical estimate with different numbers of cascade levels n. Numbers in parenthesis are standard errors.

Table 3.17: SML estimates of the Liu/Lux (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
Dow/Nik							
$\hat{m_0}$	1.386	1.387	1.386	1.397	1.381	1.373	1.346
	(0.032)	(0.029)	(0.029)	(0.030)	(0.029)	(0.031)	(0.031)
$\hat{\sigma_1}$	1.042	1.032	1.052	1.109	1.084	1.030	1.068
	(0.025)	(0.026)	(0.024)	(0.026)	(0.023)	(0.024)	(0.023)
$\hat{\sigma_2}$	1.092	1.076	1.078	1.077	1.090	1.043	1.068
	(0.042)	(0.041)	(0.042)	(0.043)	(0.042)	(0.043)	(0.042)
$\hat{\rho}$	0.051	0.064	0.096	0.058	0.076	0.089	0.047
	(0.034)	(0.040)	(0.041)	(0.033)	(0.024)	(0.031)	(0.027)
$\ln L$	-17701.673	-17139.486	-17008.225	-17046.905	-17088.234	-17100.648	-17396.042
$\hat{d}$	0.210	0.224	0.247	0.237	0.203	0.215	0.199
	(0.029)	(0.024)	(0.024)	(0.031)	(0.033)	(0.034)	(0.025)
$US/DM \sim £$							
$\hat{m_0}$	1.477	1.442	1.345	1.470	1.376	1.372	1.390
	(0.047)	(0.049)	(0.048)	(0.049)	(0.048)	(0.049)	(0.047)
$\hat{\sigma_1}$	0.695	0.692	0.613	0.745	0.618	0.605	0.617
	(0.027)	(0.021)	(0.042)	(0.029)	(0.044)	(0.025)	(0.030)
$\hat{\sigma_2}$	0.611	0.581	0.499	0.607	0.510	0.507	0.510
	(0.016)	(0.022)	(0.029)	(0.013)	(0.022)	(0.025)	(0.021)
$\hat{\rho}$	0.023	0.022	0.019	0.022	0.016	0.015	0.015
	(0.016)	(0.012)	(0.015)	(0.010)	(0.013)	(0.015)	(0.013)
$\ln L$	-18592.383	-18396.088	-18390.851	-18648.198	-18625.808	-18672.593	-18588.641
$\hat{d}$	0.238	0.201	0.175	0.177	0.159	0.160	0.168
	(0.024)	(0.023)	(0.033)	(0.023)	(0.020)	(0.017)	(0.028)
$TB2/TB1$							
$\hat{m_0}$	1.579	1.568	1.525	1.503	1.478	1.461	1.437
	(0.035)	(0.036)	(0.035)	(0.034)	(0.035)	(0.035)	(0.035)
$\hat{\sigma_1}$	1.142	1.141	1.150	1.142	1.142	1.142	1.142
	(0.043)	(0.043)	(0.046)	(0.046)	(0.047)	(0.044)	(0.047)
$\hat{\sigma_2}$	1.167	1.167	1.134	1.166	1.165	1.166	1.166
	(0.037)	(0.039)	(0.037)	(0.038)	(0.038)	(0.038)	(0.039)
$\hat{\rho}$	0.668	0.658	0.661	0.662	0.660	0.661	0.663
	(0.041)	(0.041)	(0.042)	(0.041)	(0.040)	(0.041)	(0.041)
$\ln L$	-17128.033	-17075.728	-17104.141	-17026.014	-16932.184	-16989.836	-16977.686
$\hat{d}$	0.162	0.173	0.186	0.210	0.229	0.239	0.266
	(0.023)	(0.024)	(0.024)	(0.028)	(0.026)	(0.026)	(0.026)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k (n = 8); $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of Dow/Nik is 0.295; the empirical GPH estimator d of US/DM is 0.192; the empirical GPH estimator d of $TB2/TB1$ is 0.226.

Table 3.18: GMM estimates of the Calvet/Fisher/Thompson model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
Dow/Nik								
$\hat{m_1}$	1.720	1.681	1.651	1.604	1.588	1.570	1.656	1.554
	(0.049)	(0.048)	(0.047)	(0.048)	(0.048)	(0.047)	(0.048)	(0.047)
$\hat{m_2}$	1.677	1.621	1.581	1.560	1.549	1.523	1.501	1.479
	(0.039)	(0.038)	(0.037)	(0.037)	(0.037)	(0.039)	(0.036)	(0.037)
$\hat{\sigma_1}$	1.110	1.113	1.119	1.117	1.110	1.111	1.112	1.112
	(0.050)	(0.053)	(0.051)	(0.051)	(0.053)	(0.051)	(0.049)	(0.048)
$\hat{\sigma_2}$	0.951	0.945	0.958	0.963	0.965	0.966	10.967	0.966
	(0.058)	(0.062)	(0.061)	(0.060)	(0.061)	(0.061)	(0.060)	(0.059)
$\hat{\rho}$	0.216	0.204	0.208	0.199	0.214	0.217	0.218	0.217
	(0.037)	(0.035)	(0.034)	(0.034)	(0.033)	(0.034)	(0.031)	(0.032)
J_{Prob}	0.343	0.348	0.350	0.353	0.351	0.355	0.358	0.362
$\hat{d}$	0.0725	0.0113	0.167	0.212	0.240	0.252	0.266	0.278
	(0.011)	(0.015)	(0.017)	(0.018)	(0.018)	(0.020)	(0.019)	(0.019)

	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=20$
Dow/Nik								
$\hat{m_1}$	1.552	1.552	1.615	1.552	1.552	1.552	1.551	1.551
	(0.046)	(0.046)	(0.046)	(0.046)	(0.045)	(0.046)	(0.045)	(0.047)
$\hat{m_2}$	1.478	1.477	1.478	1.478	1.473	1.478	1.478	1.472
	(0.036)	(0.035)	(0.036)	(0.035)	(0.036)	(0.035)	(0.036)	(0.035)
$\hat{\sigma_1}$	1.112	1.112	1.111	1.110	1.112	1.112	1.112	1.112
	(0.051)	(0.049)	(0.046)	(0.048)	(0.046)	(0.047)	(0.048)	(0.047)
$\hat{\sigma_2}$	0.967	0.967	0.967	0.934	0.966	0.965	0.966	0.966
	(0.056)	(0.057)	(0.056)	(0.057)	(0.058)	(0.059)	(0.060)	(0.059)
$\hat{\rho}$	0.219	0.218	0.218	0.221	0.202	0.210	0.221	0.220
	(0.330)	(0.030)	(0.031)	(0.032)	(0.031)	(0.030)	(0.031)	(0.031)
J_{Prob}	0.367	0.367	0.367	0.367	0.367	0.367	0.367	0.367
$\hat{d}$	0.285	0.299	0.311	0.316	0.314	0.315	0.314	0.314
	(0.020)	(0.022)	(0.024)	(0.025)	(0.024)	(0.026)	(0.026)	(0.027)

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n; J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of *Dow/Nik* is 0.295.

Table 3.19: GMM estimates of the Calvet/Fisher/Thompson model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
$US/DM \sim £$								
$\hat{m_1}$	1.644	1.609	1.593	1.585	1.556	1.562	1.542	1.507
	(0.111)	(0.108)	(0.107)	(0.107)	(0.104)	(0.102)	(0.103)	(0.102)
$\hat{m_2}$	1.700	1.692	1.663	1.672	1.668	1.664	1.665	1.664
	(0.055)	(0.051)	(0.050)	(0.048)	(0.047)	(0.047)	(0.045)	(0.043)
$\hat{\sigma_1}$	0.619	0.613	0.609	0.612	0.612	0.611	0.611	0.611
	(0.028)	(0.025)	(0.021)	(0.021)	(0.020)	(0.018)	(0.014)	(0.015)
$\hat{\sigma_2}$	0.550	0.486	0.488	0.486	0.487	0.487	0.487	0.487
	(0.027)	(0.024)	(0.020)	(0.021)	(0.018)	(0.015)	(0.016)	(0.015)
$\hat{\rho}$	0.027	0.029	0.027	0.027	0.025	0.026	0.026	0.022
	(0.011)	(0.010)	(0.018)	(0.016)	(0.017)	(0.016)	(0.014)	(0.013)
J_{Prob}	0.253	0.250	0.250	0.247	0.246	0.246	0.245	0.246
$\hat{d}$	0.083	0.090	0.110	0.121	0.138	0.157	0.164	0.170
	(0.012)	(0.015)	(0.013)	(0.016)	(0.016)	(0.019)	(0.020)	(0.023)

	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=20$
$US/DM \sim £$								
$\hat{m_1}$	1.466	1.433	1.426	1.407	1.382	1.575	1.353	1.339
	(0.102)	(0.103)	(0.103)	(0.102)	(0.101)	(0.101)	(0.104)	(0.103)
$\hat{m_2}$	1.662	1.665	1.664	1.662	1.661	1.661	1.664	1.665
	(0.042)	(0.042)	(0.041)	(0.040)	(0.041)	(0.041)	(0.040)	(0.044)
$\hat{\sigma_1}$	0.611	0.611	0.611	0.612	0.611	0.611	0.611	0.611
	(0.014)	(0.014)	(0.013)	(0.014)	(0.015)	(0.014)	(0.017)	(0.017)
$\hat{\sigma_2}$	0.487	0.487	0.486	0.487	0.486	0.487	0.487	0.487
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)
$\hat{\rho}$	0.024	0.022	0.022	0.023	0.022	0.026	0.022	0.023
	(0.016)	(0.011)	(0.010)	(0.012)	(0.011)	(0.012)	(0.013)	(0.011)
J_{Prob}	0.246	0.246	0.246	0.246	0.246	0.246	0.246	0.246
$\hat{d}$	0.175	0.178	0.181	0.190	0.199	0.198	0.200	0.198
	(0.026)	(0.030)	(0.032)	(0.035)	(0.041)	(0.044)	(0.052)	(0.069)

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n; J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of US/DM is 0.192.

Table 3.20: GMM estimates of the Calvet/Fisher/Thompson model

	$n=1$	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=8$
$TB2/TB1$								
$\hat{m_1}$	1.822	1.795	1.780	1.774	1.768	1.750	1.746	1.744
	(0.047)	(0.048)	(0.047)	(0.048)	(0.048)	(0.047)	(0.046)	(0.046)
$\hat{m_2}$	1.799	1.780	1.766	1.742	1.723	1.708	1.706	1.704
	(0.035)	(0.036)	(0.032)	(0.034)	(0.034)	(0.033)	(0.035)	(0.031)
$\hat{\sigma_1}$	1.301	1.300	1.296	1.292	1.298	1.296	1.072	1.296
	(0.040)	(0.040)	(0.042)	(0.041)	(0.043)	(0.044)	(0.041)	(0.040)
$\hat{\sigma_2}$	1.214	1.204	1.208	1.209	1.201	1.206	1.202	1.201
	(0.063)	(0.063)	(0.063)	(0.064)	(0.067)	(0.062)	(0.063)	(0.064)
$\hat{\rho}$	0.805	0.790	0.769	0.732	0.722	0.727	0.726	0.726
	(0.030)	(0.030)	(0.031)	(0.031)	(0.030)	(0.034)	(0.030)	(0.029)
J_{Prob}	0.020	0.018	0.019	0.019	0.017	0.017	0.017	0.017
$\hat{d}$	0.090	0.107	0.133	0.154	0.162	0.181	0.190	0.199
	(0.017)	(0.017)	(0.017)	(0.017)	(0.016)	(0.019)	(0.021)	(0.020)

	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=20$
$TB2/TB1$								
$\hat{m_1}$	1.742	1.742	1.742	1.742	1.741	1.742	1.742	1.742
	(0.046)	(0.046)	(0.046)	(0.046)	(0.045)	(0.046)	(0.045)	(0.047)
$\hat{m_2}$	1.703	1.704	1.704	1.704	1.704	1.704	1.704	1.704
	(0.033)	(0.033)	(0.033)	(0.032)	(0.037)	(0.033)	(0.033)	(0.032)
$\hat{\sigma_1}$	1.297	1.296	1.297	1.297	1.297	1.296	1.297	1.297
	(0.048)	(0.048)	(0.042)	(0.045)	(0.048)	(0.049)	(0.056)	(0.058)
$\hat{\sigma_2}$	1.201	1.201	1.201	1.201	1.201	1.201	1.201	1.201
	(0.045)	(0.045)	(0.044)	(0.042)	(0.044)	(0.044)	(0.045)	(0.046)
$\hat{\rho}$	0.726	0.726	0.726	0.726	0.726	0.726	0.726	0.726
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)
J_{Prob}	0.017	0.017	0.017	0.017	0.017	0.017	0.017	0.017
$\hat{d}$	0.215	0.230	0.239	0.248	0.252	0.355	0.358	0.358
	(0.028)	(0.026)	(0.026)	(0.028)	(0.024)	(0.027)	(0.028)	(0.027)

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n; J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of $TB2/TB1$ is 0.226.

Table 3.21: GMM estimates of the Liu/Lux (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
Dow/Nik							
$\hat{m_0}$	1.446	1.437	1.428	1.379	1.381	1.385	1.422
	(0.020)	(0.021)	(0.022)	(0.022)	(0.024)	(0.025)	(0.031)
$\hat{\sigma_1}$	1.144	1.126	1.123	1.121	1.120	1.120	1.119
	(0.034)	(0.031)	(0.031)	(0.031)	(0.030)	(0.031)	(0.029)
$\hat{\sigma_2}$	0.922	0.919	0.914	0.907	0.904	0.902	0.901
	(0.034)	(0.030)	(0.027)	(0.024)	(0.024)	(0.022)	(0.020)
$\hat{\rho}$	0.204	0.193	0.202	0.196	0.199	0.207	0.202
	(0.015)	(0.015)	(0.013)	(0.013)	(0.012)	(0.012)	(0.011)
J_{Prob}	0.272	0.266	0.266	0.269	0.268	0.268	0.269
$\hat{d}$	0.241	0.244	0.247	0.251	0.246	0.240	0.239
	(0.018)	(0.021)	(0.020)	(0.018)	(0.019)	(0.018)	(0.020)

	$k=8$	$k=9$
Dow/Nik		
$\hat{m_0}$	1.411	1.420
	(0.028)	(0.028)
$\hat{\sigma_1}$	1.119	1.119
	(0.033)	(0.034)
$\hat{\sigma_2}$	0.901	0.901
	(0.021)	(0.021)
$\hat{\rho}$	0.201	0.202
	(0.010)	(0.010)
J_{Prob}	0.269	0.269
$\hat{d}$	0.236	0.232
	(0.019)	(0.019)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 10$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of Dow/Nik is 0.295.

Table 3.22: GMM estimates of the Liu/Lux (Binomial) model

	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
Dow/Nik								
$\hat{m_0}$	1.527	1.440	1.427	1.412	1.374	1.379	1.372	1.355
	(0.017)	(0.022)	(0.024)	(0.021)	(0.023)	(0.022)	(0.019)	(0.020)
$\hat{\sigma_1}$	1.112	1.119	1.120	1.120	1.120	1.121	1.112	1.094
	(0.029)	(0.032)	(0.031)	(0.032)	(0.030)	(0.031)	(0.035)	(0.033)
$\hat{\sigma_2}$	0.893	0.908	0.907	0.906	0.907	0.906	0.893	0.934
	(0.024)	(0.022)	(0.020)	(0.023)	(0.020)	(0.023)	(0.021)	(0.023)
$\hat{\rho}$	0.167	0.202	0.204	0.201	0.205	0.196	0.204	0.212
	(0.010)	(0.012)	(0.010)	(0.012)	(0.012)	(0.012)	(0.014)	(0.009)
J_{Prob}	0.317	0.291	0.280	0.283	0.274	0.270	0.267	0.267
$\hat{d}$	0.174	0.198	0.219	0.227	0.240	0.252	0.259	0.267
	(0.015)	(0.016)	(0.014)	(0.012)	(0.016)	(0.015)	(0.018)	(0.018)

	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$	$n=18$	$n=19$	$n=20$
Dow/Nik								
$\hat{m_0}$	1.358	1.371	1.368	1.371	1.370	1.366	1.361	1.360
	(0.018)	(0.018)	(0.021)	(0.018)	(0.019)	(0.020)	(0.020)	(0.023)
$\hat{\sigma_1}$	1.121	1.112	1.118	1.112	1.121	1.050	1.112	1.112
	(0.031)	(0.031)	(0.032)	(0.034)	(0.034)	(0.034)	(0.026)	(0.031)
$\hat{\sigma_2}$	0.906	0.893	0.906	0.893	0.906	0.910	0.893	0.893
	(0.021)	(0.021)	(0.022)	(0.023)	(0.021)	(0.023)	(0.021)	(0.022)
$\hat{\rho}$	0.214	0.215	0.226	0.230	0.224	0.226	0.227	0.224
	(0.013)	(0.013)	(0.015)	(0.013)	(0.012)	(0.013)	(0.013)	(0.011)
J_{Prob}	0.267	0.267	0.267	0.267	0.267	0.267	0.267	0.267
$\hat{d}$	0.278	0.291	0.302	0.309	0.315	0.319	0.323	0.327
	(0.022)	(0.023)	(0.020)	(0.025)	(0.021)	(0.018)	(0.025)	(0.027)

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n ($k=4$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of Dow/Nik is 0.295.

Table 3.23: GMM estimates of the Liu/Lux (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$US/DM \sim £$							
$\hat{m_0}$	1.343	1.340	1.264	1.250	1.250	1.260	1.234
	(0.076)	(0.070)	(0.075)	(0.088)	(0.073)	(0.072)	(0.079)
$\hat{\sigma_1}$	0.607	1.000	0.607	0.633	0.634	0.743	0.691
	(0.024)	(0.021)	(0.029)	(0.029)	(0.024)	(0.028)	(0.028)
$\hat{\sigma_2}$	0.524	0.539	0.555	0.524	0.568	0.672	0.596
	(0.018)	(0.017)	(0.018)	(0.013)	(0.018)	(0.014)	(0.013)
$\hat{\rho}$	0.019	0.022	0.021	0.021	0.021	0.021	0.021
	(0.012)	(0.017)	(0.016)	(0.015)	(0.012)	(0.010)	(0.012)
J_{Prob}	0.271	0.272	0.274	0.274	0.274	0.274	0.274
$\hat{d}$	0.209	0.181	0.174	0.178	0.189	0.198	0.198
	(0.018)	(0.018)	(0.016)	(0.018)	(0.018)	(0.018)	(0.017)

	$k=8$	$k=9$	$k=10$	$k=11$
$US/DM \sim £$				
$\hat{m_0}$	1.251	1.260	1.255	1.250
	(0.087)	(0.086)	(0.078)	(0.079)
$\hat{\sigma_1}$	0.605	0.622	0.630	0.609
	(0.021)	(0.024)	(0.025)	(0.022)
$\hat{\sigma_2}$	0.587	0.572	0.584	0.582
	(0.014)	(0.015)	(0.017)	(0.015)
$\hat{\rho}$	0.021	0.022	0.020	0.019
	(0.013)	(0.011)	(0.013)	(0.013)
J_{Prob}	0.274	0.274	0.274	0.274
$\hat{d}$	0.199	0.204	0.187	0.181
	(0.014)	(0.014)	(0.015)	(0.017)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 12$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of US/DM is 0.192.

Table 3.24: GMM estimates of the Liu/Lux (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$TB2/TB1$							
$\hat{m_0}$	1.719	1.708	1.698	1.683	1.668	1.658	1.657
	(0.035)	(0.036)	(0.035)	(0.034)	(0.035)	(0.035)	(0.035)
$\hat{\sigma_1}$	1.332	1.331	1.330	1.332	1.332	1.332	1.332
	(0.046)	(0.046)	(0.046)	(0.046)	(0.047)	(0.048)	(0.047)
$\hat{\sigma_2}$	1.367	1.367	1.364	1.366	1.365	1.366	1.366
	(0.057)	(0.058)	(0.057)	(0.057)	(0.057)	(0.060)	(0.059)
$\hat{\rho}$	0.718	0.718	0.721	0.722	0.720	0.721	0.720
	(0.041)	(0.041)	(0.042)	(0.041)	(0.041)	(0.041)	(0.041)
J_{Prob}	0.032	0.035	0.035	0.035	0.035	0.035	0.035
$\hat{d}$	0.210	0.216	0.223	0.235	0.234	0.238	0.242
	(0.021)	(0.021)	(0.019)	(0.022)	(0.022)	(0.021)	(0.020)

	$k=8$	$k=9$
$TB2/TB1$		
$\hat{m_0}$	1.655	1.656
	(0.037)	(0.038)
$\hat{\sigma_1}$	1.330	1.34
	(0.045)	(0.045)
$\hat{\sigma_2}$	1.365	1.364
	(0.055)	(0.053)
$\hat{\rho}$	0.719	0.716
	(0.040)	(0.040)
J_{Prob}	0.035	0.035
$\hat{d}$	0.241	0.240
	(0.020)	(0.022)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 10$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of $TB2/TB1$ is 0.226.

Table 3.25: GMM estimates of the Liu/Lux (Lognormal) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
Dow/Nik							
$\hat{\lambda}$	1.153	1.140	1.132	1.124	1.125	1.125	1.126
	(0.035)	(0.040)	(0.024)	(0.020)	(0.038)	(0.033)	(0.033)
$\hat{\sigma_1}$	1.125	1.122	1.121	1.117	1.121	1.121	1.112
	(0.029)	(0.025)	(0.027)	(0.030)	(0.029)	(0.034)	(0.032)
$\hat{\sigma_2}$	0.904	0.906	0.906	0.908	0.906	0.906	0.893
	(0.041)	(0.036)	(0.042)	(0.024)	(0.043)	(0.041)	(0.036)
$\hat{\rho}$	0.222	0.211	0.204	0.197	0.195	0.195	0.194
	(0.021)	(0.020)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
J_{Prob}	0.253	0.263	0.265	0.265	0.267	0.262	0.262
$\hat{d}$	0.229	0.237	0.242	0.246	0.253	0.243	0.246
	(0.021)	(0.022)	(0.020)	(0.021)	(0.019)	(0.019)	(0.020)

	$k=8$	$k=9$
Dow/Nik		
$\hat{\lambda}$	1.128	1.127
	(0.037)	(0.038)
$\hat{\sigma_1}$	1.112	1.111
	(0.029)	(0.027)
$\hat{\sigma_2}$	0.893	0.912
	(0.044)	(0.045)
$\hat{\rho}$	0.194	0.194
	(0.017)	(0.017)
J_{Prob}	0.262	0.262
$\hat{d}$	0.248	0.245
	(0.020)	(0.020)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 10$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of Dow/Nik is 0.295.

Table 3.26: GMM estimates of the Liu/Lux (Lognormal) model

	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
Dow/Nik								
$\hat{\lambda}$	1.137	1.229	1.128	1.125	1.125	1.124	1.123	1.123
	(0.021)	(0.020)	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
$\hat{\sigma_1}$	1.133	1.126	1.122	1.120	1.118	1.117	1.115	1.115
	(0.030)	(0.031)	(0.030)	(0.029)	(0.030)	(0.031)	(0.031)	(0.030)
$\hat{\sigma_2}$	0.920	0.915	0.916	0.912	0.911	0.908	0.907	0.907
	(0.026)	(0.025)	(0.023)	(0.023)	(0.022)	(0.020)	(0.022)	(0.023)
$\hat{\rho}$	0.220	0.219	0.212	0.204	0.205	0.197	0.195	0.195
	(0.020)	(0.023)	(0.022)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)
J_{Prob}	0.377	0.368	0.320	0.289	0.2710	0.265	0.265	0.265
$\hat{d}$	0.144	0.190	0.208	0.223	0.240	0.246	0.264	0.275
	(0.017)	(0.017)	(0.018)	(0.019)	(0.020)	(0.020)	(0.021)	(0.020)

	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$	$n=18$	$n=19$	$n=20$
Dow/Nik								
$\hat{\lambda}$	1.122	1.122	1.123	1.123	1.122	1.122	1.122	1.122
	(0.020)	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)	(0.021)	(0.020)
$\hat{\sigma_1}$	1.115	1.115	1.115	1.115	1.115	1.115	1.115	1.115
	(0.030)	(0.031)	(0.031)	(0.030)	(0.029)	(0.030)	(0.030)	(0.031)
$\hat{\sigma_2}$	0.907	0.907	0.907	0.907	0.907	0.907	0.907	0.907
	(0.023)	(0.022)	(0.022)	(0.020)	(0.021)	(0.024)	(0.022)	(0.022)
$\hat{\rho}$	0.195	0.195	0.195	0.195	0.195	0.195	0.195	0.195
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
J_{Prob}	0.265	0.265	0.265	0.265	0.265	0.265	0.265	0.265
$\hat{d}$	0.281	0.289	0.296	0.304	0.307	0.305	0.305	0.305
	(0.020)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)	(0.020)

Note: Each column corresponds to the empirical estimate with different numbers of cascade level n ($k=4$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of Dow/Nik is 0.295.

Table 3.27: GMM estimates of the Liu/Lux (Lognormal) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$US/DM \sim £$							
$\hat{\lambda}$	1.192	1.163	1.171	1.166	1.164	1.164	1.162
	(0.024)	(0.040)	(0.037)	(0.035)	(0.035)	(0.034)	(0.029)
$\hat{\sigma_1}$	0.608	0.616	0.608	0.610	0.607	0.607	0.602
	(0.018)	(0.019)	(0.024)	(0.016)	(0.013)	(0.013)	(0.009)
$\hat{\sigma_2}$	0.906	0.905	0.893	0.906	0.906	0.907	0.906
	(0.021)	(0.025)	(0.016)	(0.019)	(0.026)	(0.025)	(0.023)
$\hat{\rho}$	0.019	0.020	0.021	0.019	0.021	0.021	0.021
	(0.015)	(0.011)	(0.015)	(0.022)	(0.020)	(0.019)	(0.015)
J_{Prob}	0.314	0.306	0.305	0.301	0.299	0.300	0.300
$\hat{d}$	0.212	0.180	0.171	0.159	0.181	0.190	0.189
	(0.021)	(0.018)	(0.018)	(0.018)	(0.018)	(0.016)	(0.018)

	$k=8$	$k=9$	$k=10$	$k=11$
$US/DM \sim £$				
$\hat{\lambda}$	1.165	1.162	1.165	1.164
	(0.020)	(0.022)	(0.020)	(0.025)
$\hat{\sigma_1}$	0.607	0.607	0.605	0.605
	(0.015)	(0.012)	(0.016)	(0.014)
$\hat{\sigma_2}$	0.486	0.490	0.491	0.491
	(0.014)	(0.015)	(0.017)	(0.017)
$\hat{\rho}$	0.021	0.019	0.018	0.018
	(0.023)	(0.025)	(0.022)	(0.025)
J_{Prob}	0.300	0.300	0.300	0.300
$\hat{d}$	0.197	0.210	0.214	0.214
	(0.018)	(0.019)	(0.017)	(0.017)

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 12$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of US/DM is 0.192.

Table 3.28: GMM estimates of the Liu/Lux (Lognormal) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$TB2/TB1$							
$\hat{\lambda}$	1.524	1.510	1.492	1.481	1.474	1.468	1.467
	(0.044)	(0.043)	(0.042)	(0.042)	(0.043)	(0.043)	(0.043)
$\hat{\sigma_1}$	1.132	1.130	1.135	1.126	1.131	1.130	1.139
	(0.064)	(0.060)	(0.059)	(0.059)	(0.058)	(0.054)	(0.056)
$\hat{\sigma_2}$	1.257	1.248	1.240	1.244	1.243	1.240	1.242
	(0.035)	(0.036)	(0.034)	(0.033)	(0.029)	(0.030)	(0.031)
$\hat{\rho}$	0.728	0.718	0.712	0.686	0.663	0.641	0.641
	(0.031)	(0.031)	(0.030)	(0.028)	(0.026)	(0.027)	(0.030)
J_{Prob}	0.024	0.021	0.021	0.019	0.020	0.021	0.021
$\hat{d}$	0.235	0.231	0.233	0.218	0.230	0.239	0.241
	(0.023)	(0.023)	(0.025)	(0.023)	(0.025)	(0.023)	(0.022)
$TB2/TB1$							
$\hat{\lambda}$	1.460	1.461					
	(0.041)	(0.042)					
$\hat{\sigma_1}$	1.137	1.137					
	(0.055)	(0.052)					
$\hat{\sigma_2}$	1.246	1.240					
	(0.034)	(0.035)					
$\hat{\rho}$	0.641	0.639					
	(0.027)	(0.028)					
J_{Prob}	0.021	0.021					
$\hat{d}$	0.244	0.240					
	(0.020)	(0.021)					

Note: Each column corresponds to the empirical estimate with different joint numbers of cascade level k ($n = 10$); J_{Prob} gives the probability of the pertinent J test statistic; $\hat{d}$ is the mean of 100 simulated GPH estimators, and numbers in parenthesis are standard errors. The empirical GPH estimator d of $TB2/TB1$ is 0.226.

Chapter 4

Beyond the Bivariate case: Higher Dimensional Multifractal Models

4.1 Tri-variate (higher dimensional) multifractal models

The bivariate MF model deals with only two financial time series, whereas an investment portfolio could be allocated to more than two assets. Analogously to the bivariate case, a multivariate ML model assumes N assets (time series) that have the same overall number of volatility cascades containing k joint cascade levels; after the kth level, each time series independently has separate additional multifractal processes. Our parsimonious multivariate multifractal model is given by

$$r_{q,t} = \sigma_q \otimes [G(M_t)]^{1/2} \otimes u_{q,t}. \tag{4.1.1}$$

$r_{q,t}$ are asset returns with $q = 1, 2, \cdots, N$ being the number of assets. σ_q are scale parameters (unconditional standard deviations) of the return series, and both $r_{q,t}$ and σ_q are $N \times 1$ vector; $\otimes$ denotes element by element multiplication; $u_{q,t}$ is a $N \times 1$ vector, whose elements are multivariate standard Normal distributions. Considering a three-asset

portfolio, $N = 3$, it is the trivariate standard Normal with correlation matrix

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

The multifractal process denoted by $G(M_t)$ is assumed to be

$$G(M_t) = \prod_{i=1}^{k} M_t^{(i)} \otimes \prod_{j=k+1}^{n} M_{q,t}^{(j)}, \tag{4.1.2}$$

and it defines the number of joint cascade levels k within multivariate multifractal processes. In addition, we restrict the specification of the transition probabilities as the one in the bivariate setting:

$$\begin{aligned} \gamma_i &= 2^{-(k-i)}, \qquad for \quad i = 1, \ldots k; \\ \gamma_i &= 2^{-(n-i)}, \qquad for \quad i = k+1, \ldots n. \end{aligned} \tag{4.1.3}$$

This specification allows each multifractal process having two starting cascades, that is, it starts again after the joint cascade level k. Each component is either renewed at time t with probability γ_i, depending on its rank i within the hierarchy of multipliers, or remains unchanged with probability $1 - \gamma_i$.

For the distribution of the multifractal volatility components M_t, we specify the multipliers to be random draws from either a binomial or Lognormal distribution. In the binomial (discrete) case, we assume two volatility component draws, $m_0 \in (0, 2)$ and the alternative $m_1 = 2 - m_0$; for the Lognormal (continuous) version, we preserve the assumption as the one in the bivariate case by setting

$$-log_2 M \sim N(\lambda, \sigma_m^2). \tag{4.1.4}$$

We impose an additional restriction of $E[M_t^{(i)}] = 0.5$, which leads to $exp[-\lambda \ln 2 + 0.5\sigma_m^2(\ln 2)^2] = 0.5$, and we arrive at

$$\sigma_m^2 = 2(\lambda - 1)/ln2. \tag{4.1.5}$$

This transformation reduces the estimation of λ and σ_m in Eq. (4.1.4) to one parameter.

4.2 Maximum likelihood estimation

We expand the likelihood approach from the bivariate model to the more general multivariate setting. Let r_t be the set of N assets' joint return observations $\{r_{q,t}\}$ for $q = 1, 2 \dots N$ and $t = 1, 2 \dots T$; Θ be the unknown parameters to be estimated, the exact likelihood function has the form like in Eq. (3.3.1):

$$f(r_1, \cdots, r_T; \Theta) \quad = \quad \prod_{t=1}^{T} (\pi_{t-1} A) \cdot f(r_t | M_t = m^i). \tag{4.2.1}$$

A is the transition matrix, which specifies the dynamics of each possible state of

$$P(M_{t+1} = m^j | M_t = m^i), \tag{4.2.2}$$

both M_t and $m^{(\cdot)}$ are vectors, which contain the combinations of each possible volatility component. For multifractal processes with discrete distribution of volatility components, say the Binomial model, $i, j \in \{1, 2 \dots (2^N)^n\}$, for N being the number of assets allocated and n being the number of cascade levels. It indicates that the transition matrix A has the dimension of $(2^N)^n \times (2^N)^n$.

$f(r_t | M_t = m^i)$ is the density of the innovation r_t conditional on M_t, which is

$$f(r_t | M_t = m^i) = \frac{F_N \left\{ r_t \div \left[\sigma \otimes \eta^{1/2} \right] \right\}}{\sigma \otimes \eta^{1/2}}, \tag{4.2.3}$$

$F_N\{\cdot\}$ denotes the multivariate standard Normal density function (trivariate standard Normal for three assets) and $\eta = G(M_t)$; $\div$ represents element-by-element division. π_t

denotes the conditional probability defined by

$$\pi_t^i = P(M_t = m^i | r_1, \cdots, r_t). \tag{4.2.4}$$

For Binomial model, we know $\sum_{i=1}^{R} \pi_t^i = 1$ for $R = (2^N)^n$ and the one-step-ahead conditional probabilities of π_{t+1} can be obtained by the Bayesian updating:

$$\pi_{t+1} = \frac{f(r_{t+1}|M_{t+1} = m^i) \otimes (\pi_t A)}{\sum f(r_{t+1}|M_{t+1} = m^i) \otimes (\pi_t A)}. \tag{4.2.5}$$

We easily recognize that this recursive process is highly dependent on the dimensions of the transition matrix, as highlighted in the bivariate MF models. In the trivariate case, the maximization of the likelihood function implies to evaluate the size of $8^n \times 8^n$ dimensions of A in each iteration for the binomial model, and current computational limitations make choices of n beyond 3 unfeasible. Besides, it is not applicable for models with continuous distributions of the multipliers such as Lognormal distribution, because these imply an infinite number of states.

4.3 Simulation based maximum likelihood estimation

We recognize that ML estimation is constrained by its immense computational demands, in particular for higher dimensional multifractal models. The maximization process depends strongly on the dimension of the transition (filtering) probability matrix, which increases exponentially with the rate of $(2^N)^n$ for binomial multifractal processes, and one can imagine the intensive computations when evaluating a 4096×4096 transition matrix in each iteration for the trivariate model with number of cascade levels $n = 4$.

In this section, we also implement simulation based inference via particle filter, which is introduced in Chapter 3. The essence of this algorithm is to avoid formalizing and evaluating the exact form of the transition matrix, which is required to generate π_{t+1} given π_t through

the Bayesian updating. Instead, some limited number of importance samples (particles) are employed to pass the filtering process

$$\sum_{j=1}^{R} P(M_{t+1} = m^i | M_t = m^j) P(M_t = m^j | r_t), \tag{4.3.1}$$

by adopting Sampling/Importance Resampling (SIR), the method developed by Rubin (1987) and further analyzed by Pitt and Shephard (1999). The SIR algorithm assists to generate independent draws $\{M_t^{(b)}\}_{b=1}^{B}$ from the conditional distribution π_t, and given new information r_{t+1}, an approximation of Bayesian updating provides simulations of $\{M_{t+1}^{(b)}\}_{b=1}^{B}$ recursively.

As detailed in Section 3.4, we complete the approximation by simulating each $M_t^{(b)}$ one step forward and re-weighting using an importance sampler: simulate the Markov Chain one step ahead with B draws $\hat{M}_{t+1}^{(1)}, \hat{M}_{t+1}^{(2)}, \ldots, \hat{M}_{t+1}^{(B)}$, which are taken independently from the prior; this preliminary step only uses information available at time t, therefore these draws have to account for new return r_{t+1}. Under SIR, each of these draws is chosen by draw a random number q from 1 to B with the probability of:

$$P(q = b) = \frac{f(r_{t+1} | M_{t+1} = \hat{M}_{t+1}^{(b)})}{\sum_{i=1}^{B} f(r_{t+1} | M_{t+1} = m^{(i)})}, \tag{4.3.2}$$

then $M_{t+1}^{(1)} = \hat{M}_{t+1}^{(q)}$ is a draw from π_{t+1}. Repeating B times, we obtain B draws $M_{t+1}^{(1)}, \ldots M_{t+1}^{(B)}$ which have been adjusted to account for the new information.

As particle filters treat the discrete support generated by the particles as the "true" filtering density, this allows us to produce an approximation to the prediction probability density $P(M_{t+1} = m_t^i | r_t)$, by using the discrete support of the particles, and then one-step-ahead conditional probability is

$$\pi_{t+1}^i \propto f(r_{t+1} | M_{t+1} = m^i) \frac{1}{B} \sum_{b=1}^{B} P(M_{t+1} = m^i | M_t = \hat{M}_t^{(b)}) \tag{4.3.3}$$

Thus one step ahead density hence becomes:

$$f(r_t|r_1,\cdots,r_{t-1}) \approx \frac{1}{B}\sum_{b=1}^{B} f(r_t|M_t=\hat{M}_t^{(b)}), \tag{4.3.4}$$

and this implies that we maximize the approximate likelihood function:

$$g(r_1,\cdots,r_T;\Theta) \approx \prod_{t=1}^{T}\left[\frac{1}{B}\sum_{b=1}^{B} f(r_t|M_t=\hat{M}_t^{(b)})\right]. \tag{4.3.5}$$

We can use these calculations to carry out simulated likelihood estimation of higher dimensional multifractal processes, whose Monte Carlo studies are reported in the following section.

4.4 GMM Estimation

Generalized Method of Moments (GMM) approach by Hansen (1982) assists us to obtain the vector of parameters β by minimizing the distance between the analytical moments and empirical moments:

$$\widehat{\beta}_T = \arg\min_{\beta\in\Theta} \bar{M}_T(\beta)' W_T \bar{M}_T(\beta), \tag{4.4.1}$$

with Θ being the admissible parameter space. $\bar{M}_T(\beta)$ stands for the vector of differences between sample moments and analytical moments, and W_T is a positive definite weighting matrix, which controls over-identification when applying GMM. Implementing Eq. (4.4.1), one typically starts with the identity matrix; then the inverse of the covariance matrix obtained from the first round estimation is used as the weighting matrix in the next step; and this procedure continues until the estimates converge.

The applicability of GMM for multifractal models has been discussed in Chapter 3, and we follow the same practical solution by transforming the observed return data r_t into τth log differences::

$$\begin{aligned} R_{t,\tau} &= \ln|r_t| - \ln|r_{t-\tau}| \\ &= 0.5\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-\tau}^{(i)}) + 0.5\sum_{j=k+1}^{n}(\varepsilon_t^{(j)} - \varepsilon_{t-\tau}^{(j)}) + (ln|u_t| - ln|u_{t-\tau}|) \end{aligned} \tag{4.4.2}$$

One may notice that in the transformation above the scale parameter σ_1 and σ_2 drops out, while estimating the scale parameters can be pursued by adding additional moment conditions (e.g. unconditional moments of r_t). In our practice, we use the moments as in bivariate models by considering each observation's contribution to the standard deviation of the sample return time series.

GMM provides a more convenient and efficient way towards the estimation of higher dimensional MF models as it allows us to treat each pair of time series as a bivariate case.

In order to exploit as much information as possible, the moment conditions that we consider include two categories:

The first category of moment conditions is the moments for individual time series' autocovariance, which contains the moments of raw transformed observations and squared transformed observations:[1]

$$Cov[R_{t,1}, R_{t-1,1}]; \qquad Cov[R^2_{t,1}, R^2_{t-1,1}];$$

in the trivariate case, let $X_{t,\tau}$, $Y_{t,\tau}$ and $Z_{t,\tau}$ be transformed time series through Eq. (4.4.2), and the pertinent moment conditions are

$$Cov[X_{t,1}, X_{t-1,1}]; \quad Cov[Y_{t,1}, Y_{t-1,1}]; \quad Cov[Z_{t,1}, Z_{t-1,1}];$$

and

$$Cov[X^2_{t,1}, X^2_{t-1,1}]; \quad Cov[Y^2_{t,1}, Y^2_{t-1,1}]; \quad Cov[Z^2_{t,1}, Z^2_{t-1,1}];$$

The second category of conditions is obtained by considering moment conditions of covariances (moments of raw transformed observations and squared transformed observations):

$$Cov[R_{t,1}, R^-_{t,1}]; \qquad Cov[R_{t,1}, R^-_{t-1,1}];$$

and

$$Cov[R^2_{t,1}, (R^-_{t,1})^2]; \qquad Cov[R^2_{t,1}, (R^-_{t-1,1})^2];$$

$R^-_{t,1}$ represents any one other time series rather than $R_{t,1}$. In the trivariate case, we have six moment conditions corresponding to six pairs of time series for raw transformed observations, namely,

$$Cov[X_{t,1}, Y_{t,1}]; \quad Cov[X_{t,1}, Z_{t,1}]; \quad Cov[Y_{t,1}, Z_{t,1}];$$

$$Cov[X_{t,1}, Y_{t-1,1}]; \quad Cov[X_{t,1}, Z_{t-1,1}]; \quad Cov[Y_{t,1}, Z_{t-1,1}];$$

[1] As in the bivariate case, we leave moment conditions of $Cov[R_{t+\tau,\tau}, R_{t,\tau}]$ with $\tau > 1$ for future work.

plus six moment conditions for squared transformed observations, these 12 moment conditions provide additional information for the correlation parameters ρ_{ij} for $i, j = 1, 2, 3$. Altogether, we use 18 moment conditions when implementing GMM for tri-variate multifractal models, and the detailed analytical moments are given in the Appendix for both Binomial model and Lognormal model.

We firstly apply GMM approach to our trivariate MF model with relatively small numbers of cascade levels to compare their performances with the likelihood approach and simulation base inference. Table 4.1 reports the comparison of ML and GMM estimations with Monte Carlo studies (designed as in previous sections). Initial settings include $n = 3$ (the limit of computational feasibility for ML), $k = 1$, $m_0 = 1.3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. It should be not too surprising that the ML estimators are more efficient than GMM since ML extracts all the information in the data. Obviously, variability of estimates with GMM is much higher particularly in small sub-sample size, but dramatically decreases in the case of larger sub-sample size; for example, it is from around 4 times that of the ML estimator for m_0's. While we also observe that the efficiency of $\hat{\sigma}$ from GMM is pretty close to ML, which shows that there is no loss of efficiency using moment conditions of sample standard deviation for $\hat{\sigma}$ within GMM. One additional gain for GMM is that it takes only a small fraction of computation time that ML costs; they are 12.5 and 109 hours for GMM and ML, respectively.

Table 4.2 reports the comparison of simulation based ML and GMM estimation with Monte Carlo studies: initial settings includes $n = 4$, $k = 2$, $m_0 = 1.3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. We used $B = 500$ particles for SML estimation. Let us begin with the estimator of m_0, variability of estimates with GMM is higher only in small sub-sample size (SD and RMSE), but dramatically decreases in the case of larger sub-sample size; we also observe that the efficiency of the σ estimates from GMM (by considering sample standard deviation) is almost identical to that from SML; for correlation estimators,

a mixed picture shows that GMM estimator of ρ_{12} is dominated by the one from SML, but is opposite in the cases of ρ_{23} and ρ_{13}. As in the ML section, we report the computation time needed for implementing 400 Monte Carlo simulations and estimations in Table 4.2; they are 14 and 120.5 hours for GMM and SML respectively.

Note that there is no constraint to the number of cascade levels within our higher dimensional multifractal processes for GMM approach. Next, we proceed by reporting further results of Monte Carlo experiments designed to explore the performance of our GMM estimator for trivariate multifractal models with larger numbers of cascade levels. We fixed n = 12 with number of joint cascades $k = 3$, the scale parameters $\sigma_1 = \sigma_2 = \sigma_3 = 1$, and the increment correlation matrix is set as $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. As before,we begin by simulating 100,000 observations in each iteration, and randomly choose three different sub-samples with sample sizes $N_1 = 2000$, $N_2 = 5000$, and $N_3 = 10000$, which is a robust design to assess the estimation performance. Let us start with the Binomial model with parameter value $m_0 = 1.3$.[2] Table 4.3 shows the statistical result of the GMM estimator: for Binomial distribution parameter $\hat{m_0}$, not only the bias but also the standard deviation and root mean squared error show quite encouraging behavior. It is also the case for other parameters, even in the small sample size $N = 2000$ and $N = 5000$, the average bias of the Monte Carlo estimates is moderate, and particularly close to zero for the larger sample sizes $N = 10000$. One also easily observe that it is asymptotically efficient with increasing the sub-sample size and in harmony with $T^{\frac{1}{2}}$ efficiency which further underscores the good performance of the log transformation within our GMM estimation.

One advantage of GMM is that it allows to estimate multifractal processes with continuous distributions of volatility components. Following the work of Mandelbrot et al. (1997)

[2]Studies on other parameter values have also been pursued, we omit them here due to their qualitatively similar results.

and Calvet and Fisher (2002), we use the same Lognormal distribution of

$$-log_2 M \sim N(\lambda, \sigma_m^2)$$

in the tri-variate case, and Eq. (4.1.5) provides the relationship between λ and σ_m by assigning a restriction of $E[M] = 0.5$.[3] In our Monte Carlo simulations and estimations reported in Table 4.4, we cover parameter values $\lambda = 1.20$, and other initial setting as in the Binomial model, namely $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. As can be seen, the picture is not too different from that obtained in the Binomial case: Biases are moderate again, SD and RMSE significantly decrease with increasing the sub-sample size. Monte Carlo studies towards other different choice of parameter values have also been conducted, we skip them for saving some space as no significant deviation is observed comparing with the performance above.

Figure 4.1 and Figure 4.2 present the fraction of p values that fall within the 2% interval fractiles from Hansen's J tests of over-identification restrictions (three figures from up to down corresponding to three different sample sizes: $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$). There are much more significant left-shifts compared with the bivariate cases, Especially, we observe the sharp left-shift in the Lognormal model. All in all, the Monte Carlo studies again show the satisfactory behaviour of our GMM estimator for higher dimensional multifractal models with both discrete and continuous state spaces.

[3] A factor of 2^n is multiplied for compensating the mean of multipliers being 0.5.

4.5 Empirical estimates

For empirical applications of the multivariate multifractal models, we consider daily data for a equally weighted portfolio of three foreign exchange rates, U.S. Dollar to British Pound, German Mark to British Pound, and Japanese Yen to British Pound ($US/DM/JP$, March 1973 - February 2004), where the first symbol inside these parentheses designates the short notation for the corresponding time series, followed by the starting and ending dates for the sample at hand. Figure 4.3 is the plot of Japanese Yen to British Pound exchange rate p_t and return $r_t = 100 \times [\ln(p_t) - \ln(p_{t-1})]$ (Figure 3.7 gives a plot of empirical exchange rates and returns for U.S. Dollar to British Pound and German Mark to British Pound).

Table 4.5 presents the empirical estimates of the trivariate multifractal model with $n = 3$ (it is the computational limit) via the maximum likelihood estimation. There is not much different for the parameter m_0 with two different joint cascade levels; one may also observe that the correlation parameter estimates for U.S. Dollar to British Pound and German Mark to British Pound ρ_{12} is a bit deviated with the one from the bivariate case which has been reported in the last chapter. It is actually plausible by considering the fact that different total number of the cascade levels n are used. We also find the maximized log-likelihood values are quite close between $k = 1$ and $k = 2$ with only slightly higher values when $k = 1$. Indeed, one would prefer to have more comparisons among various joint cascade levels (k) in the case for larger number of cascades n.

By using particle filter, we are allowed to employ relatively more cascade levels. Table 4.6 reports empirical estimates of the trivariate MF model with the number of cascade levels of $n = 5$ via simulation based maximum likelihood approach. With the fixed total cascade levels, there are not much fluctuations for the estimates among different joint cascade levels k, but one may aware the maximized log-likelihood values increases about 1100 from the case $k = 1$ to $k = 2$. Practically, we find that the computation is still very intensive even

when using a relatively small number of particles of $B = 500$ in the trivatiate model.

In order to remove the upper limit of cascade levels of multivariate MF processes, we also present the empirical estimates for higher numbers of n via the GMM approach; and moment conditions described in Section 4.2 are used. With fixed joint cascade level $k = 4$, Table 4.7 provides the empirical results for the Binomial model with a range of numbers of cascade levels n from 5 to 20. It shows the estimates for m_0 is decreasing with the number of cascade levels n increasing, but then almost keep constant after the case $n = 12$; while other estimates are relatively stable, as well as the J-test statistic. Table 4.8 reports the empirical results with different numbers of joint cascade levels k but $n = 20$ fixed. In contrast to Table 4.7, there are only slight variations for m_0 (among $k = 1$ to $k = 5$). To demonstrate the applicability of GMM to the MF process with continuous state space, Table 4.9 and Table 4.10 give the empirical estimates for the Lognormal model with various numbers of cascade levels, and similar picture is observed comparing with the Binomial case.

Table 4.1: Comparison between ML and GMM estimators

		ML			GMM		
$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	0.006	0.026	0.026	-0.097	0.122	0.156
	N_2	0.007	0.012	0.013	-0.027	0.059	0.065
	N_3	0.003	0.007	0.008	-0.005	0.032	0.033
$\hat{\sigma_1}$	N_1	0.011	0.022	0.023	0.001	0.031	0.031
	N_2	0.007	0.014	0.014	0.004	0.019	0.019
	N_3	-0.002	0.008	0.009	-0.001	0.013	0.013
$\hat{\sigma_2}$	N_1	-0.008	0.020	0.021	-0.010	0.032	0.032
	N_2	0.003	0.013	0.015	-0.004	0.019	0.020
	N_3	0.001	0.007	0.008	-0.007	0.013	0.013
$\hat{\sigma_3}$	N_1	0.012	0.022	0.023	-0.003	0.032	0.032
	N_2	0.012	0.013	0.013	0.007	0.020	0.021
	N_3	0.004	0.010	0.011	0.004	0.014	0.014
$\hat{\rho_{12}}$	N_1	0.014	0.024	0.027	0.027	0.122	0.125
	N_2	0.006	0.012	0.012	0.003	0.091	0.090
	N_3	-0.003	0.008	0.009	-0.018	0.064	0.067
$\hat{\rho_{23}}$	N_1	0.008	0.023	0.024	0.020	0.070	0.073
	N_2	-0.005	0.013	0.013	0.002	0.048	0.048
	N_3	0.001	0.008	0.010	-0.007	0.033	0.034
$\hat{\rho_{13}}$	N_1	0.014	0.024	0.025	0.005	0.038	0.038
	N_2	-0.006	0.015	0.016	-0.003	0.030	0.030
	N_3	-0.003	0.007	0.007	-0.005	0.027	0.028

Note: Simulations are based on the trivariate binomial multifractal process with $n = 3$, $k = 1$, which is almost the limit of computational feasibility for ML, and initial value $m_0 = 1.3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 4.2: Comparison between SML and GMM estimators

		SML			GMM		
$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	-0.013	0.062	0.062	-0.103	0.129	0.166
	N_2	0.015	0.045	0.046	-0.028	0.055	0.061
	N_3	0.010	0.034	0.036	-0.007	0.032	0.033
$\hat{\sigma_1}$	N_1	-0.022	0.039	0.040	-0.011	0.027	0.028
	N_2	0.017	0.022	0.023	0.006	0.016	0.016
	N_3	-0.019	0.011	0.012	0.001	0.010	0.012
$\hat{\sigma_2}$	N_1	0.020	0.033	0.035	-0.008	0.027	0.029
	N_2	0.014	0.021	0.022	0.010	0.017	0.018
	N_3	0.018	0.011	0.013	0.003	0.010	0.011
$\hat{\sigma_3}$	N_1	0.033	0.038	0.040	0.005	0.026	0.027
	N_2	-0.026	0.017	0.019	0.001	0.015	0.016
	N_3	0.021	0.012	0.012	-0.001	0.011	0.012
$\hat{\rho_{12}}$	N_1	-0.035	0.051	0.052	0.079	0.125	0.128
	N_2	0.022	0.038	0.040	0.007	0.086	0.087
	N_3	-0.017	0.021	0.023	-0.006	0.062	0.062
$\hat{\rho_{23}}$	N_1	0.029	0.056	0.057	0.058	0.070	0.072
	N_2	-0.019	0.035	0.035	0.002	0.048	0.049
	N_3	0.013	0.027	0.028	-0.005	0.034	0.035
$\hat{\rho_{13}}$	N_1	0.031	0.046	0.046	0.004	0.045	0.047
	N_2	-0.021	0.029	0.031	-0.001	0.031	0.031
	N_3	-0.015	0.019	0.019	-0.004	0.020	0.021

Note: Simulations are based on the trivariate binomial multifractal process with $n = 4$, $k = 2$, and initial value $m_0 = 1.3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. For each scenario, 400 Monte Carlo simulations have been carried out.

Table 4.3: GMM estimation for the trivariate multifractal (Binomial) model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{m_0}$	N_1	-0.097	0.128	0.161
	N_2	-0.042	0.075	0.086
	N_3	-0.019	0.056	0.059
$\hat{\sigma_1}$	N_1	-0.011	0.078	0.079
	N_2	-0.001	0.055	0.055
	N_3	-0.001	0.038	0.038
$\hat{\sigma_2}$	N_1	0.000	0.084	0.084
	N_2	0.000	0.055	0.055
	N_3	-0.004	0.039	0.039
$\hat{\sigma_3}$	N_1	-0.002	0.086	0.086
	N_2	-0.003	0.052	0.052
	N_3	-0.002	0.040	0.040
$\hat{\rho_{12}}$	N_1	0.011	0.133	0.133
	N_2	0.000	0.102	0.102
	N_3	-0.009	0.085	0.085
$\hat{\rho_{23}}$	N_1	-0.014	0.124	0.124
	N_2	-0.017	0.109	0.110
	N_3	-0.021	0.098	0.100
$\hat{\rho_{13}}$	N_1	-0.006	0.089	0.089
	N_2	-0.011	0.073	0.074
	N_3	-0.009	0.056	0.057

Note: Simulations are based on the trivariate binomial multifractal process with $n = 12$, $k = 4$, and initial value $m_0 = 1.3$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. 400 Monte Carlo simulations have been carried out.

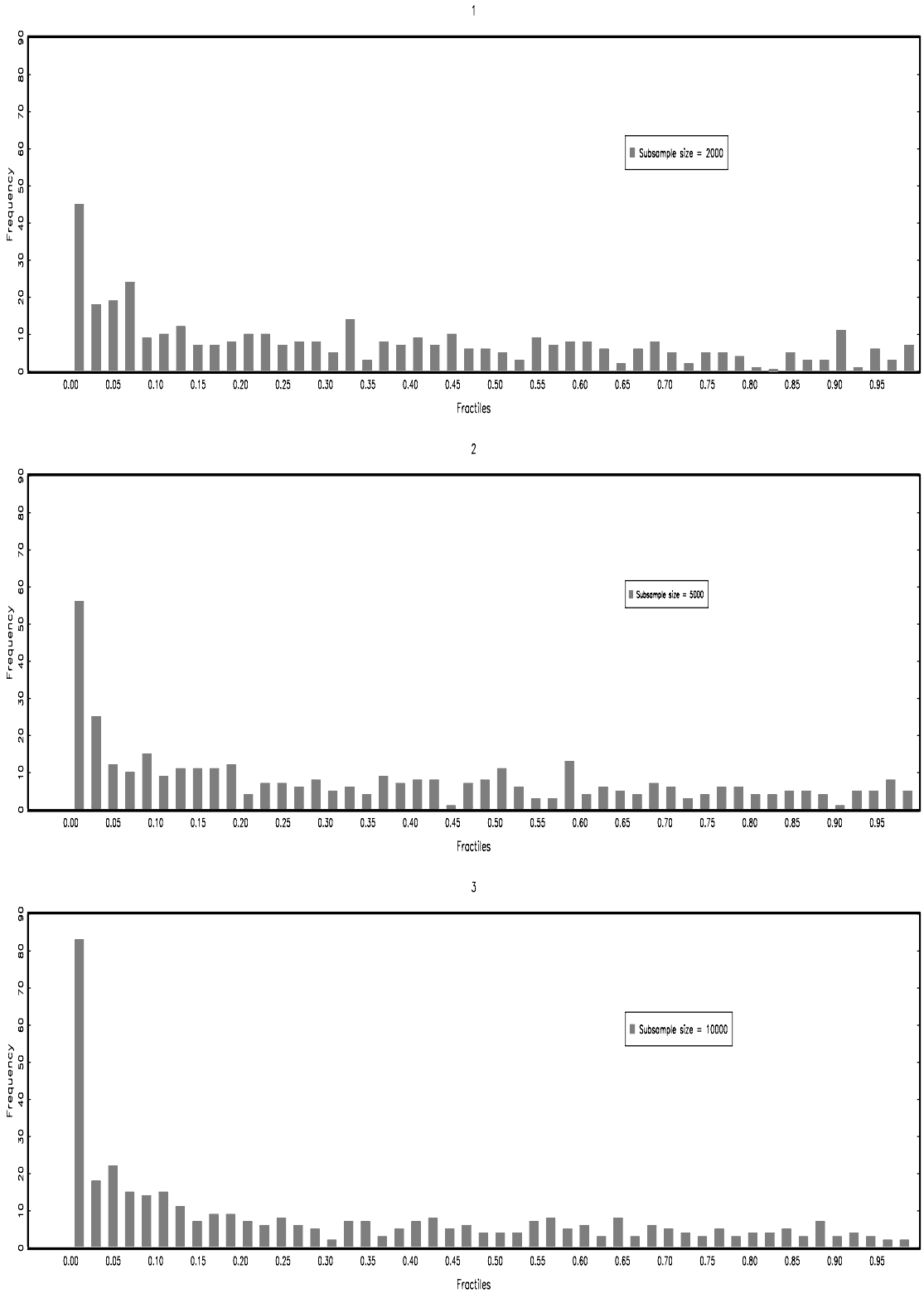

Figure 4.1: The distribution of p value for the test of over-identification restrictions (binomial trivariate MF model).

Table 4.4: GMM estimation for the trivariate multifractal (Lognormal) model

$\hat{\theta}$	Sub-sample Size	$Bias$	SD	$RMSE$
$\hat{\lambda}$	N_1	-0.047	0.051	0.068
	N_2	-0.012	0.031	0.033
	N_3	-0.003	0.021	0.021
$\hat{\sigma_1}$	N_1	-0.056	0.295	0.300
	N_2	-0.029	0.210	0.211
	N_3	-0.027	0.154	0.156
$\hat{\sigma_2}$	N_1	-0.068	0.277	0.285
	N_2	-0.033	0.213	0.215
	N_3	-0.008	0.158	0.158
$\hat{\sigma_3}$	N_1	-0.055	0.283	0.288
	N_2	-0.034	0.200	0.203
	N_3	-0.011	0.177	0.177
$\hat{\rho_{12}}$	N_1	0.014	0.142	0.142
	N_2	-0.018	0.101	0.102
	N_3	-0.029	0.073	0.078
$\hat{\rho_{23}}$	N_1	0.020	0.088	0.088
	N_2	-0.013	0.056	0.058
	N_3	-0.016	0.040	0.043
$\hat{\rho_{13}}$	N_1	-0.009	0.048	0.048
	N_2	-0.016	0.027	0.031
	N_3	-0.019	0.021	0.029

Note: Simulations are based on the trivariate Lognormal multifractal process with $n = 12$, $k = 4$, and initial value $\lambda = 1.2$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 1$, $\rho_{12} = 0.3$, $\rho_{23} = 0.5$, $\rho_{13} = 0.7$. Sample lengths are $N_1 = 2,000$, $N_2 = 5,000$ and $N_3 = 10,000$. Bias denotes the distance between the given and estimated parameter value, SD and RMSE denote the standard deviation and root mean squared error, respectively. 400 Monte Carlo simulations have been carried out.

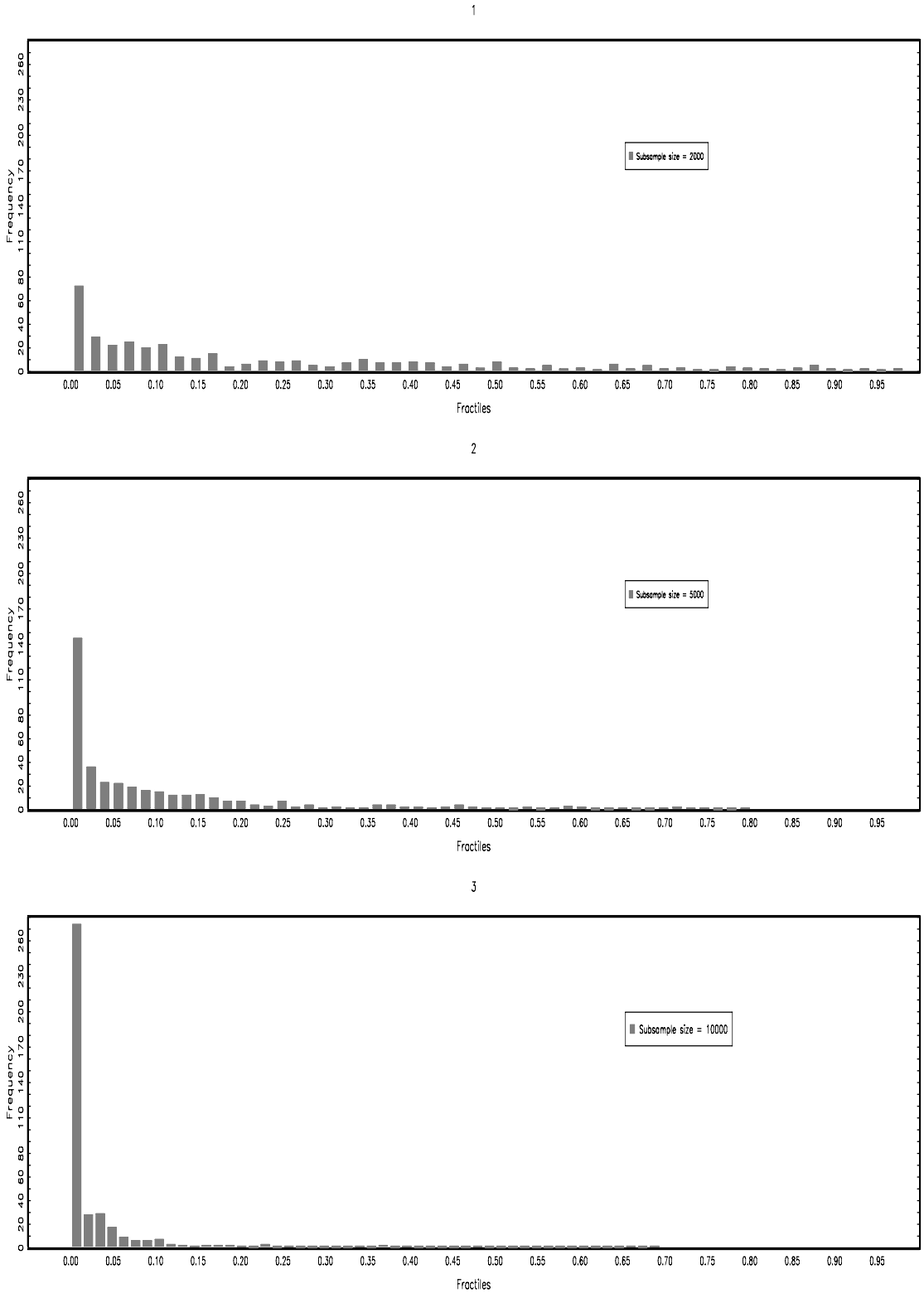

Figure 4.2: The distribution of p values for the test of over-identification restrictions for a Lognormal trivariate MF model.

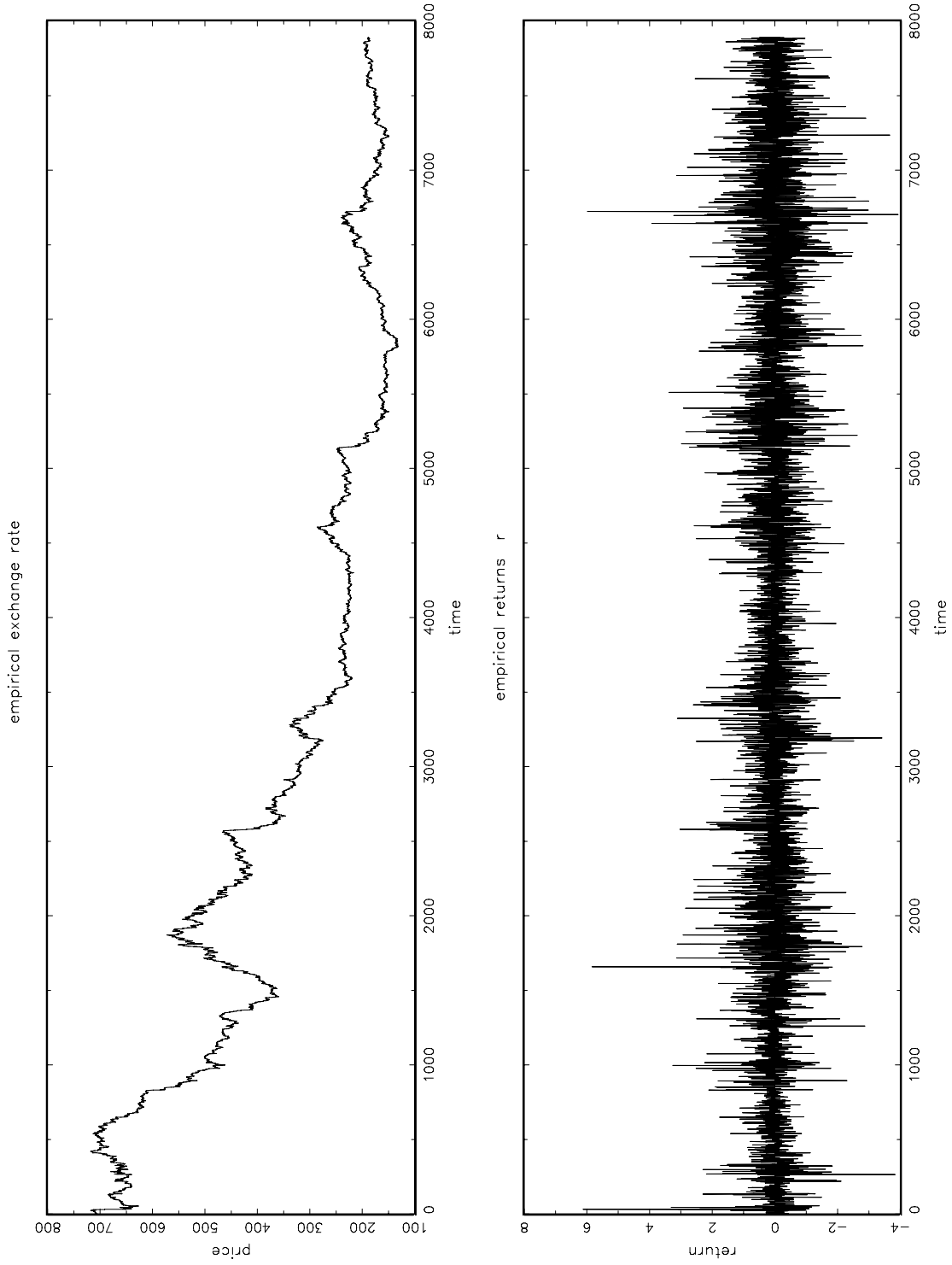

Figure 4.3: Empirical time series: Japanese Yen to British Pound exchange rate.

Table 4.5: ML estimates of the tri-variate (Binomial) MF model

	$k=1$	$k=2$
$US/DM/JP$		
$\hat{m_0}$	1.647	1.630
	(0.044)	(0.045)
$\hat{\sigma_1}$	0.870	0.874
	(0.033)	(0.032)
$\hat{\sigma_2}$	0.904	0.910
	(0.040)	(0.039)
$\hat{\sigma_3}$	0.817	0.812
	(0.031)	(0.029)
$\hat{\rho_{12}}$	0.031	0.027
	(0.019)	(0.019)
$\hat{\rho_{23}}$	0.362	0.357
	(0.030)	(0.031)
$\hat{\rho_{13}}$	0.633	0.628
	(0.042)	(0.042)
$\ln L$	-33279.440	-33251.378

Note: The number of cascade levels $n = 3$.

Table 4.6: SML estimates of the tri-variate (Binomial) MF model

	$k=1$	$k=2$	$k=3$	$k=4$
$US/DM/JP$				
$\hat{m_0}$	0.533	0.529	0.528	0.530
	(0.044)	(0.042)	(0.042)	(0.042)
$\hat{\sigma_1}$	0.811	0.824	0.822	0.823
	(0.045)	(0.044)	(0.044)	(0.044)
$\hat{\sigma_2}$	0.899	0.893	0.890	0.890
	(0.039)	(0.040)	(0.040)	(0.040)
$\hat{\sigma_3}$	0.722	0.718	0.719	0.718
	(0.047)	(0.048)	(0.048)	(0.048)
$\hat{\rho_{12}}$	0.045	0.042	0.037	0.035
	(0.019)	(0.020)	(0.019)	(0.019)
$\hat{\rho_{23}}$	0.245	0.242	0.238	0.236
	(0.022)	(0.024)	(0.022)	(0.022)
$\hat{\rho_{13}}$	0.537	0.532	0.528	0.527
(0.045)	(0.044)	(0.042)	(0.042)	
$\ln L$	-32167.252	-31076.057	-31095.723	-31103.371

Note: The number of cascade levels $n = 5$, and the number of particles of $B = 500$ is used.

Table 4.7: GMM estimates of the tri-variate MF (Binomial) model

	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
$US/DM/JP$								
$\hat{m_0}$	1.538	1.472	1.436	1.400	1.387	1.361	1.352	1.333
	(0.041)	(0.041)	(0.042)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
$\hat{\sigma_1}$	0.794	0.791	0.790	0.795	0.791	0.791	0.791	0.794
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
$\hat{\sigma_2}$	0.910	0.894	0.893	0.911	0.893	0.894	0.898	0.897
	(0.047)	(0.047)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
$\hat{\sigma_3}$	0.711	0.714	0.713	0.711	0.713	0.714	0.718	0.717
	(0.053)	(0.053)	(0.053)	(0.054)	(0.053)	(0.054)	(0.054)	(0.054)
$\hat{\rho_{12}}$	0.046	0.045	0.045	0.041	0.042	0.041	0.040	0.040
	(0.019)	(0.020)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
$\hat{\rho_{23}}$	0.223	0.225	0.223	0.221	0.221	0.221	0.222	0.223
	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
$\hat{\rho_{13}}$	0.526	0.527	0.527	0.519	0.519	0.521	0.519	0.518
	(0.045)	(0.044)	(0.044)	(0.043)	(0.044)	(0.044)	(0.043)	(0.043)
J_{Prob}	0.033	0.030	0.031	0.031	0.031	0.030	0.030	0.030
	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=20$
$\hat{m_0}$	1.330	1.329	1.329	1.329	1.329	1.328	1.329	1.328
	(0.040)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
$\hat{\sigma_1}$	0.792	0.793	0.795	0.793	0.795	0.795	0.795	0.793
	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)	(0.047)
$\hat{\sigma_2}$	0.909	0.894	0.893	0.907	0.899	0.899	0.899	0.912
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
$\hat{\sigma_3}$	0.729	0.724	0.723	0.727	0.723	0.723	0.723	0.723
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
$\hat{\rho_{12}}$	0.046	0.045	0.043	0.041	0.041	0.041	0.040	0.040
	(0.019)	(0.019)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
$\hat{\rho_{23}}$	0.226	0.225	0.223	0.221	0.221	0.221	0.220	0.220
	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)
$\hat{\rho_{13}}$	0.520	0.517	0.518	0.516	0.515	0.511	0.515	0.516
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
J_{Prob}	0.030	0.030	0.030	0.030	0.030	0.030	0.030	0.030

Note: The number of joint cascade levels $k = 4$.

Table 4.8: GMM estimates of the tri-variate MF (Binomial) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$US/DM/JP$								
$\hat{m_0}$	1.335	1.331	1.329	1.328	1.325	1.325	1.326	1.324
	(0.039)	(0.038)	(0.038)	(0.040)	(0.038)	(0.037)	(0.035)	(0.035)
$\hat{\sigma_1}$	0.797	0.795	0.792	0.780	0.789	0.783	0.781	0.779
	(0.049)	(0.048)	(0.047)	(0.047)	(0.047)	(0.045)	(0.046)	(0.044)
$\hat{\sigma_2}$	0.950	0.933	0.920	0.912	0.910	0.911	0.912	0.912
	(0.052)	(0.053)	(0.051)	(0.050)	(0.050)	(0.051)	(0.049)	(0.048)
$\hat{\sigma_3}$	0.741	0.735	0.728	0.723	0.725	0.726	0.721	0.726
	(0.063)	(0.060)	(0.055)	(0.052)	(0.051)	(0.051)	(0.050)	(0.059)
$\hat{\rho}$	0.050	0.056	0.058	0.059	0.054	0.057	0.058	0.057
	(0.047)	(0.045)	(0.044)	(0.044)	(0.043)	(0.044)	(0.041)	(0.042)
$\hat{\rho_{12}}$	0.046	0.045	0.043	0.041	0.041	0.041	0.040	0.040
	(0.018)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
$\hat{\rho_{23}}$	0.226	0.225	0.223	0.221	0.221	0.221	0.220	0.220
	(0.022)	(0.022)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.020)
$\hat{\rho_{13}}$	0.528	0.521	0.519	0.516	0.515	0.511	0.515	0.516
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
J_{Prob}	0.033	0.032	0.032	0.032	0.032	0.032	0.032	0.032
	$k=9$	$k=10$	$k=11$	$k=12$	$k=13$	$k=14$	$k=15$	$k=16$
$\hat{m_0}$	1.325	1.325	1.327	1.328	1.329	1.325	1.322	1.324
	(0.039)	(0.0340)	(0.042)	(0.042)	(0.041)	(0.041)	(0.041)	(0.041)
$\hat{\sigma_1}$	0.780	0.772	0.778	0.784	0.785	0.785	0.778	0.783
	(0.047)	(0.045)	(0.045)	(0.044)	(0.044)	(0.045)	(0.040)	(0.041)
$\hat{\sigma_2}$	0.913	0.910	0.908	0.910	0.912	0.911	0.911	0.911
	(0.055)	(0.053)	(0.052)	(0.052)	(0.052)	(0.052)	(0.0450)	(0.050)
$\hat{\sigma_3}$	0.725	0.728	0.722	0.725	0.720	0.722	0.718	0.720
	(0.051)	(0.052)	(0.052)	(0.052)	(0.052)	(0.051)	(0.051)	(0.051)
$\hat{\rho}$	0.053	0.056	0.058	0.059	0.054	0.057	0.058	0.057
	(0.040)	(0.042)	(0.041)	(0.042)	(0.043)	(0.042)	(0.042)	(0.042)
$\hat{\rho_{12}}$	0.046	0.045	0.043	0.044	0.047	0.048	0.048	0.050
	(0.015)	(0.015)	(0.016)	(0.017)	(0.017)	(0.016)	(0.016)	(0.016)
$\hat{\rho_{23}}$	0.225	0.226	0.226	0.225	0.225	0.224	0.230	0.231
	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)
$\hat{\rho_{13}}$	0.515	0.514	0.516	0.519	0.515	0.511	0.515	0.516
	(0.042)	(0.040)	(0.041)	(0.042)	(0.042)	(0.042)	(0.041)	(0.041)
J_{Prob}	0.032	0.032	0.032	0.032	0.032	0.032	0.032	0.032

Note: The number of cascade levels $n = 20$.

Table 4.9: GMM estimates of the tri-variate MF (Lognormal) model

	$n=5$	$n=6$	$n=7$	$n=8$	$n=9$	$n=10$	$n=11$	$n=12$
$US/DM/JP$								
$\hat{\lambda}$	1.305	1.262	1.246	1.237	1.215	1.191	1.169	1.123
	(0.038)	(0.037)	(0.037)	(0.037)	(0.036)	(0.037)	(0.037)	(0.038)
$\hat{\sigma_1}$	0.794	0.791	0.790	0.795	0.791	0.791	0.791	0.794
	(0.043)	(0.043)	(0.043)	(0.044)	(0.043)	(0.043)	(0.043)	(0.043)
$\hat{\sigma_2}$	0.910	0.894	0.893	0.911	0.893	0.894	0.898	0.897
	(0.050)	(0.049)	(0.050)	(0.051)	(0.051)	(0.051)	(0.051)	(0.050)
$\hat{\sigma_3}$	0.711	0.714	0.713	0.711	0.713	0.714	0.718	0.717
	(0.055)	(0.055)	(0.054)	(0.054)	(0.053)	(0.054)	(0.054)	(0.054)
$\hat{\rho_{12}}$	0.046	0.045	0.045	0.041	0.042	0.041	0.040	0.040
	(0.023)	(0.023)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
$\hat{\rho_{23}}$	0.223	0.225	0.223	0.221	0.221	0.221	0.222	0.223
	(0.037)	(0.036)	(0.036)	(0.036)	(0.035)	(0.035)	(0.035)	(0.035)
$\hat{\rho_{13}}$	0.526	0.527	0.527	0.519	0.519	0.521	0.519	0.518
	(0.042)	(0.042)	(0.043)	(0.044)	(0.044)	(0.044)	(0.043)	(0.044)
J_{Prob}	0.033	0.032	0.032	0.033	0.031	0.031	0.028	0.028
	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$	$n=18$	$n=19$	$n=20$
$\hat{\lambda}$	1.114	1.109	1.108	1.108	1.108	1.108	1.108	1.108
	(0.038)	(0.037)	(0.037)	(0.037)	(0.036)	(0.037)	(0.037)	(0.038)
$\hat{\sigma_1}$	0.792	0.793	0.795	0.793	0.795	0.795	0.795	0.793
	(0.044)	(0.044)	(0.043)	(0.044)	(0.045)	(0.043)	(0.042)	(0.043)
$\hat{\sigma_2}$	0.909	0.894	0.893	0.907	0.899	0.899	0.899	0.912
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
$\hat{\sigma_3}$	0.729	0.724	0.723	0.727	0.723	0.723	0.723	0.710
	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)	(0.054)
$\hat{\rho_{12}}$	0.046	0.045	0.043	0.041	0.041	0.041	0.040	0.040
	(0.021)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
$\hat{\rho_{23}}$	0.226	0.225	0.223	0.221	0.221	0.221	0.220	0.220
	(0.035)	(0.035)	(0.036)	(0.035)	(0.035)	(0.034)	(0.034)	(0.034)
$\hat{\rho_{13}}$	0.520	0.517	0.518	0.516	0.515	0.511	0.515	0.516
	(0.042)	(0.042)	(0.041)	(0.042)	(0.041)	(0.040)	(0.040)	(0.040)
J_{Prob}	0.028	0.028	0.028	0.028	0.028	0.028	0.028	0.028

Note: The number of joint cascade levels $k = 4$.

Table 4.10: GMM estimates of the tri-variate MF (Lognormal) model

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$
$US/DM/JP$								
$\hat{\lambda}$	1.114	1.112	1.110	1.108	1.108	1.090	1.107	1.107
	(0.040)	(0.041)	(0.041)	(0.041)	(0.042)	(0.041)	(0.043)	(0.041)
$\hat{\sigma_1}$	0.803	0.802	0.795	0.794	0.795	0.797	0.804	0.808
	(0.047)	(0.047)	(0.047)	(0.045)	(0.045)	(0.045)	(0.045)	(0.046)
$\hat{\sigma_2}$	0.925	0.914	0.912	0.912	0.913	0.918	0.925	0.917
	(0.049)	(0.047)	(0.047)	(0.047)	(0.045)	(0.043)	(0.043)	(0.043)
$\hat{\sigma_3}$	0.729	0.725	0.713	0.710	0.710	0.694	0.701	0.7178
	(0.051)	(0.051)	(0.052)	(0.053)	(0.051)	(0.050)	(0.051)	(0.051)
$\hat{\rho_{12}}$	0.044	0.040	0.044	0.040	0.039	0.040	0.042	0.042
	(0.021)	(0.020)	(0.019)	(0.017)	(0.018)	(0.018)	(0.018)	(0.018)
$\hat{\rho_{23}}$	0.238	0.217	0.219	0.220	0.218	0.213	0.199	0.210
	(0.036)	(0.036)	(0.035)	(0.034)	(0.037)	(0.037)	(0.036)	(0.036)
$\hat{\rho_{13}}$	0.545	0.536	0.527	0.516	0.520	0.529	0.527	0.530
	(0.048)	(0.045)	(0.042)	(0.041)	(0.044)	(0.043)	(0.045)	(0.048)
J_{Prob}	0.026	0.023	0.025	0.023	0.023	0.022	0.023	0.023

	$k=9$	$k=10$	$k=11$	$k=12$	$k=13$	$k=14$	$k=15$	$k=16$
$\hat{m_0}$	1.108	1.108	1.108	1.108	1.108	1.108	1.095	1.092
	(0.042)	(0.042)	(0.043)	(0.042)	(0.042)	(0.043)	(0.041)	(0.041)
$\hat{\sigma_1}$	0.788	0.786	0.786	0.808	0.783	0.783	0.793	0.787
	(0.045)	(0.047)	(0.046)	(0.043)	(0.045)	(0.047)	(0.045)	(0.044)
$\hat{\sigma_2}$	0.909	0.905	0.893	0.903	0.882	0.909	0.919	0.889
	(0.046)	(0.046)	(0.046)	(0.048)	(0.047)	(0.045)	(0.048)	(0.048)
$\hat{\sigma_3}$	0.718	0.711	0.705	0.703	0.720	0.718	0.719	0.698
	(0.049)	(0.048)	(0.045)	(0.047)	(0.046)	(0.048)	(0.046)	(0.048)
$\hat{\rho_{12}}$	0.043	0.044	0.047	0.046	0.043	0.045	0.045	0.047
	(0.020)	(0.020)	(0.021)	(0.020)	(0.020)	(0.020)	(0.021)	(0.019)
$\hat{\rho_{23}}$	0.240	0.242	0.242	0.247	0.246	0.243	0.246	0.248
	(0.036)	(0.033)	(0.037)	(0.033)	(0.034)	(0.035)	(0.035)	(0.035)
$\hat{\rho_{13}}$	0.535	0.537	0.543	0.544	0.536	0.551	0.547	0.549
	(0.045)	(0.044)	(0.046)	(0.045)	(0.048)	(0.044)	(0.043)	(0.046)
J_{Prob}	0.023	0.022	0.022	0.024	0.022	0.022	0.022	0.022

Note: The number of cascade levels $n = 20$.

Chapter 5

Risk Management Applications of Bivariate MF Models

5.1 Introduction

In general, risk management is the process of measuring or assessing risk, and then developing strategies to manage and control it. Conventional risk management focuses on risks stemming from physical or legal causes, such as natural disasters, accidents, and lawsuits.

Financial risk management, on the other hand, focuses on risks that can be managed by using practical financial instruments. With the aid of modern technology, information is instantaneously available, which means that change, and subsequent market reactions, occur very quickly. Financial markets can also be very quickly affected by changes in other stock prices, exchange rates and interest rates. In addition, the fact that financial markets and their individual assets are correlated attracts considerable attention of finance professionals; it is this correlation that accelerates financial uncertainties that can subsequently result in worldwide financial disasters. As an example, (1) take the market crash in 1987: The stock market crash of 1987 was the largest one day stock market crash in history. The

Dow Jones index dropped 22.6% on the 19th of October 1987, and 500 billion dollars have been evaporated within one day. Markets in most countries around the world collapsed in the same fashion. (2) Southeast Asian financial crises in 1990s: It started with the devaluation of Thai Baht, which took place on July 2nd 1997, a 15% to 20% devaluation that occurred two months after this currency started to suffer from a massive speculative attack. The devaluation of the Thai Baht was soon followed by that of the Philippine Peso, the Malaysian Ringgit, the Indonesian Rupiah and, to a lesser extent, the Singapore Dollar. The currency turmoil erupted in Southeast Asia, in turn triggering the market collapse on Wall Street on October 27, 1997.

In this Chapter, we present empirical applications of bivariate MF models using empirical financial time series, including stock indices, foreign exchange rates, and U.S. government bond maturity rates. Firstly the unconditional distribution of simulated data including portfolios are studied based on bivariate MF models. In particular we focus on the tail part which describes the most extreme outliers which are the interest of risk measurement and management. We then conduct assessments in terms of two widely applied risk instruments, namely Value-at-Risk and Expected Shortfall, comparing to the ones implied from traditional GARCH type model.

5.2 Data description

Recalling the six empirical financial markets data in Section 3.6, we separate each time series into two subsets (in-sample data used for estimation, out-of-sample data for forecast assessment); for the in-sample periods we use for DOW/NIK: January 1969 - February 1988; US/DM: March 1973 - January 1990; $TB1/TB2$: June 1976 - May 1991. For the out-of-sample subset, we use for DOW/NIK: March 1988 - October 2004; US/DM: February 1990 - February 2004; $TB1/TB2$: June 1991 - October 2004. The relevant descriptive

statistics reported in Table 5.1 for in-sample data indicate the substantial violation of normality (asymmetry and kurtosis).

Table 5.1: Descriptive statistics for empirical daily (in-sample) returns r_t

Data	Mean	Variance	Minimum	Maximum	Skewness	Kurtosis
Dow	-0.0280	0.7974	-8.4185	22.4746	1.7585	49.2165
Nik	-0.0294	1.2365	-12.4303	16.1354	0.4405	25.1747
US	-0.0006	0.4294	-23.2647	3.8427	-1.5304	37.2622
DM	0.0122	0.2432	-2.9999	4.0822	0.4867	7.8081
$TB1$	0.0135	2.1238	-14.3101	16.3764	0.1227	13.6966
$TB2$	0.0127	2.3705	-13.1336	16.6024	-0.1219	14.1261

5.3 Multivariate GARCH model

Besides the widely used univariate version of the ARCH model proposed by Engle (1982), GARCH model by Bollerslev (1986) and variants of modified versions, multivariate GARCH models have also been extensively applied throughout the financial economics literatures.

Bollerslev et al. (1988) provide the basic framework for a multi-variate GARCH model. For the number of assets N, returns $r_t = \{r_{1,t}, \ldots, r_{N,t}\}$, and $t = 1, \ldots, T$, a multivariate GARCH model can be defined as follows:

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t; \\ \varepsilon_t &= H_t^{1/2} \cdot z_t, \end{aligned} \tag{5.3.1}$$

$H_t^{1/2}$ is a $N \times N$ positive definite matrix. Furthermore, $N \times 1$ random vector $\{z_t\}$ is assumed to be independent and identically-distributed, with $E[z_t] = 0, Var[z_t] = I_N$, I_N refers to

the identity matrix of order N and

$$\begin{aligned} E[r_t|\Omega_{t-1}] &= \mu_t; \\ Var(r_t|\Omega_{t-1}) &= H_t^{1/2}(H_t^{1/2})', \end{aligned} \tag{5.3.2}$$

Ω_{t-1} is the information set available till time $t-1$, $H_t = h_{ij,t}$ $(i, j = 1, 2, \ldots, N$ refers to different time series). Bollerslev et al. (1988) propose a general formulation of H_t by extending the GARCH representation in the univariate case to the vectorized conditional-variance matrix (VEC model). In the general VEC model, each element of H_t is a linear function of the lagged squared errors and cross-products of errors and lagged values of the elements of H_t. The VEC(1,1) model is defined as:

$$h_t = C + A\varepsilon_{t-1}^2 + Gh_{t-1}, \tag{5.3.3}$$

where

$$h_t = vech(H_t), \tag{5.3.4}$$

$$\varepsilon_t = vech(\epsilon_t\epsilon_t)', \tag{5.3.5}$$

and $vech(\cdot)$ denotes the operator that stacks the lower triangular portion of a $N \times N$ matrix as a $N(N+1)/2 \times 1$ vector. A and G are square parameter matrices of order $N(N+1)/2$ and C is a $N(N+1)/2 \times 1$ constant parameter vector.

However this vectorized representation involves a large number of parameters which is equal to $N(N+1)(N(N+1)+1)/2$. Empirical applications require further restrictions and simplifications. One useful member of the vech-representation family is the diagonal form. Under the diagonal form, each variance-covariance term is postulated to follow a GARCH-type equation with the lagged variance-covariance term and the product of the corresponding lagged residuals as the right-hand-side variables in the conditional-(co)variance equation.

It is often difficult to verify the condition that the variance-covariance matrix of an estimated multi-variate GARCH model is positive definite. Furthermore, such conditions are often very difficult to impose during the optimization of the log-likelihood function. A constant correlation GARCH (CC-GARCH) model proposed by Bollerslev (1990) overcomes these difficulties. The CC-GARCH (1, 1) model assumes constant cross correlations,[1] such that:

$$\begin{aligned} h_{ii,t} &= \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i h_{ii,t-1} \qquad i = 1, 2, \ldots, N, \\ h_{ij,t} &= \rho_{ij} \sqrt{h_{i,t} h_{j,t}} \qquad \forall i \neq j. \end{aligned} \tag{5.3.6}$$

The normality assumption implies that the log-likelihood has the form:

$$\log L(\theta) = -\frac{T}{2} log(2\pi) - \frac{1}{2} \sum_{t=1}^{T} log(H_t) - \frac{1}{2} \sum_{t=1}^{T} (r_t - \mu_t)' H_t^{-1} (r_t - \mu_t). \tag{5.3.7}$$

With the assumption of constant correlations, the maximum likelihood (ML) estimate of the correlation matrix is equal to the sample correlation matrix, which is always positive semi-definite; the positive semi-definiteness of the conditional variance-covariance matrix can be ensured when the conditional variances are all positive. In addition, when the correlation matrix is concentrated out of the likelihood function, further simplification is achieved in the optimization by the fact that the Gaussian ML estimator is consistent, provided that the conditional mean and variance are correctly specified. Table 5.2 reports the empirical CC-GARCH(1, 1) parameters with respect to the in-sample data described in the previous section.

[1] Instead of assuming correlations to be time invariant, Engle and Kroner (1995) propose a class of multivariate GARCH model called the BEKK (named after Baba, Engle, Kraft and Kroner) model, and provide some theoretical analysis related to the vech-representation form which also ensures the condition of a positive definite conditional-variance matrix in the process of optimization. More details on different versions of multivariate GARCH models can be found in Gourieroux (1997) (Chapter 6).

Table 5.2: CC-GARCH(1, 1) model estimates (in-sample data)

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_{12}$
Dow/Nik	-0.03	0.01	0.02	0.01	0.19	0.06	0.79	0.93	0.05
	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
US/DM	0.01	-0.01	0.19	0.08	0.10	0.15	0.88	0.78	0.02
	(0.01)	(0.00)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)	(0.00)
$TB1/TB2$	-0.03	-0.02	0.13	0.09	0.02	0.03	0.87	0.89	0.74
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)	(0.03)

Note:The ML estimation of the CC-GARCH(1, 1) model is implemented via the GAUSS module 'Fanpac' provided by Aptech^{TM} Systems Inc.

5.4 Unconditional coverage

In this section, we study the unconditional distribution of simulated data based on bivariate multifractal processes. In particular we are interesting in the tail part, which describes the frequency of extreme events, implying the information of financial risk. It is now widely accepted that asset returns exhibit excess kurtosis (fat-tail), which means that the unconditional distribution of data has more probability mass in the tails than the one under the convenient assumption of Normal distribution (extreme observations occur more often than would be expected under Normality), cf . Fama (1965), Ding et al. (1993), Pagan (1996), Guillaume et al. (2000), Cont (2001), Lo and MacKinlay (2001) and so on.

We proceed by comparing the probability distribution of empirical data and the Gaussian. Figure 5.1 shows the probability density function (pdf) of empirical returns of *Dow* (data standardized), as well as the pdf of Gaussian (left panel). In addition to the descriptive statistics reported in Table 5.1, one easily recognize that empirical data are not Normal distributed but leptokurtic from the plot in Figure 5.1. To examine more details

of the tail distribution, we also report the corresponding complementary cumulative distributions $1 - F(x)$, where $F(x)$ is the cumulative distribution function (CDF) defined as $\int_{-\infty}^{x} P(t)dt$ with $P(t)$ being the probability density function. The right panel of Figure 5.1 presents the heavily-tailed stylized fact which has been pervasively found in empirical financial data, a comparison with complementary cumulative distribution of Gaussian is also reported.

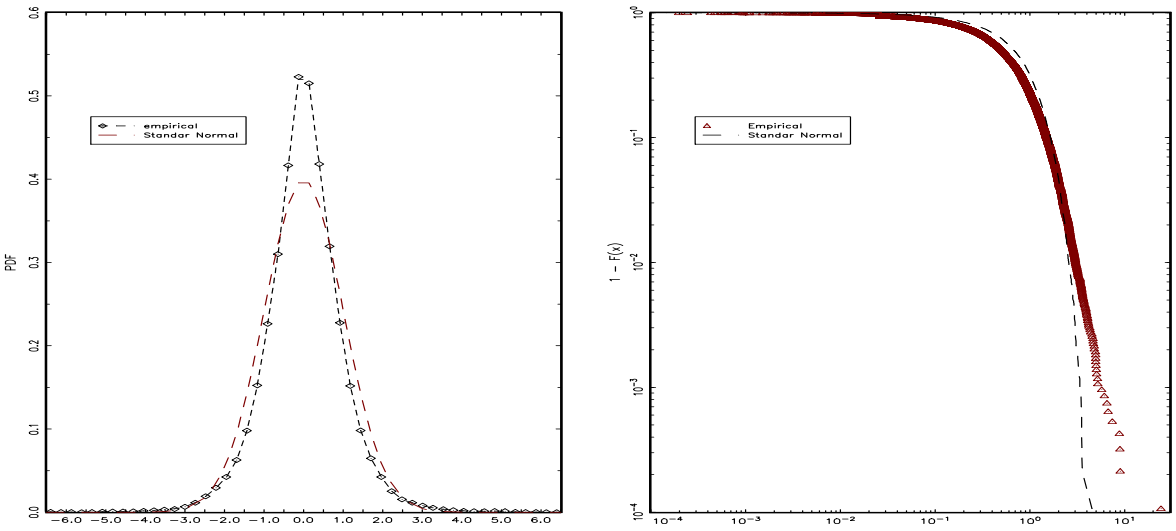

Figure 5.1: This graph shows the probability density function (pdf) of empirical *DOW* (left panel), and log-log plot of the complementary cumulative distribution of empirical DOW (right panel), for comparison, the dashed lines give the pdf and complementary of the cumulative distribution of Gaussian distribution.

In addition to the graphical illustration, we study the simulation data based on the in-sample estimates, and investigate the similarities between the empirical and simulated data regarding their tail part. Let us define $\tilde{r}_{t,t+h}$ as the forward-looking h-period return at time t:

$$\tilde{r}_{t,t+h} = \sum_{i=1}^{h} r_{t+i}, \tag{5.4.1}$$

and we also calculate the $(1-\alpha)^{th}$ quantile Q_h of the unconditional distribution of $\tilde{r}_{t:t+h}$

by:

$$Pr\left(\tilde{r}_{t:t+h} \leq Q_h^{\alpha}\right) = \alpha. \tag{5.4.2}$$

Given α level, Eq. (5.4.2) allows us to obtain Q_h from the cumulative distribution of $\tilde{r}_{t:t+h}$. Our study of unconditional distribution focuses on α equal to 0.1, 0.05 and 0.01, which the risk management is interested in, and we compare the ones based on both empirical and simulated data. Thus, the performance of the model can be assessed by computing the empirical failure rate (for both the left and right tails of the distribution of returns). The failure rate is defined as the number of times returns exceed the simulated Q_h^{α}. If the model is well specified, it is expected to be as close as possible to the pre-specified α level. We then perform Kupiec's likelihood ratio (LR) test, cf. Kupiec (1995); because the computation of the empirical failure rate is characterized as a sequence of yes/no observations, the identity of the hypothesis and empirical failure rate can be tested finally through the hypothesis:

H_0: $\alpha = \hat{\alpha}$, against

H_1: $\alpha \neq \hat{\alpha}$,

where $\hat{\alpha}$ is the empirical failure rate estimated. Let T be the out-of-sample size, then, at the 1% level, an approximate confidence interval for $\hat{\alpha}$ is given by

$$\left[\hat{\alpha} - 2.58\sqrt{\hat{\alpha}(1-\hat{\alpha})/T}, \quad \hat{\alpha} + 2.58\sqrt{\hat{\alpha}(1-\hat{\alpha})/T}\right] \tag{5.4.3}$$

We estimate the bivariate MF models via the GMM and maximum likelihood (ML) approaches using in-sample data. The number of cascade levels is selected based on our empirical studies in Section 3.6, namely: $n = 10$ for Dow/Nik, $n = 12$ for US/DM, and $n = 10$ for $TB1/TB2$ are used in the Calvet/Fisher/Thompson model; $n = 14$, $k = 4$ for Dow/Nik, $n = 12$, $k = 5$ for US/DM, and $n = 10$, $k = 4$ for $TB1/TB2$ are used in the Liu/Lux model. The empirical results are reported in Table 5.3 and Table 5.4 (the value inside the parentheses is the standard error), respectively. For the maximum likelihood

Table 5.3: GMM estimates for Liu/Lux model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_0$	$\hat{\rho}$
Dow/Nik	1.117	0.913	1.334	0.206
	(0.028)	(0.022)	(0.034)	(0.017)
US/DM	0.914	0.873	1.409	0.022
	(0.014)	(0.021)	(0.024)	(0.009)
$TB1/TB2$	1.109	0.890	1.433	0.792
	(0.038)	(0.052)	(0.027)	(0.041)

Note: The number of cascade levels are: $n = 14$, $k = 4$ for Dow/Nik, $n = 12$, $k = 5$ for US/DM, and $n = 10$, $k = 4$ for $TB1/TB2$.

(ML) estimates (Table 5.5 and Table 5.6), we use $n = 5$ for both bivariate MF models (it is almost the computational limit with ML), and the joint cascade levels used (the Liu/Lux model) are $k = 2$ for Dow/Nik, $k = 1$ for US/DM and $k = 2$ for $TB1/TB2$, respectively. Notice that these estimates are slightly different from the ones obtained in Section 3.6 due to the fact that only the in-sample data are used here. Based on these empirical estimates, simulations of the bivariate time series are conducted to calculate the cumulative returns for single time series and portfolios: Equal-weight portfolio and Hedge portfolio (an investment that is taken out specifically to reduce or cancel out the risk in another investment, e.g. long an asset and short another one). Figure 5.2 gives the plot for the two portfolios of empirical data (stock indices, foreign exchange rates and U.S bonds).[2] We then calculate the quantile for each simulated return time series by Eq. (5.4.2); empirical observations (single and portfolio return data) are used to compare with the pertinent level quantile $Q^{\alpha}_{t:t+h}$; the failure rate is calculated as the ratio of the number of empirical observations

[2]Given two assets (x and y), an equally weighted portfolio is one with portfolio fraction 0.5 of both assets; a Hedge portfolio is a zero net investment portfolio of $x - y$.

Table 5.4: GMM estimates for Calvet/Fisher/ Thompson model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_1$	$\hat{m}_2$	$\hat{\rho}$
Dow/Nik	0.764	0.932	1.488	1.263	0.252
	(0.038)	(0.041)	(0.102)	(0.088)	(0.086)
US/DM	0.820	0.764	1.557	1.433	0.024
	(0.014)	(0.032)	(0.081)	(0.043)	(0.021)
$TB1/TB2$	1.063	0.928	1.372	1.485	0.801
	(0.027)	(0.041)	(0.063)	(0.079)	(0.035)

Note: $n = 10$ for Dow/Nik, $n = 12$ for US/DM, and $n = 10$ for $TB1/TB2$.

over $Q^{\alpha}_{t:t+h}$ and the number of out-of-sample size. Failure rates including the Kupiec test statistics for the different time series are given in Table 5.7 (GMM) and Table 5.8 (ML) that are based on the Liu/Lux model; Table 5.9 (GMM) and Table 5.10 (ML) are based on the Calvet/Fisher/Thompson model. The results are presented separately below:

Beginning with the Liu/Lux model, and looking firstly at the stock indices, it shows quite positive results based on GMM and ML at a confidence level of 10% (except for one risky case in Table 5.7), while at the 5% level, we observe three cases which are too risky and one which is too conservative in Table 5.7, also three risky cases in two and five-day horizons, cf. Table 5.8. There are also six cases (two conservative and four risky ones) that lie outside the confidence interval in Table 5.7 at a confidence level of 1%; at the same confidence level of 1%, we find three risky and two conservative cases out of the twelve scenarios in Table 5.8.

For foreign exchange rates US and DM, the unconditional coverage from GMM seem entirely successful throughout each scenario in Table 5.7. In contrast, there are three unsuccessful cases in Table 5.8, namely, two risky ones for US at the 10% level in one-day

Table 5.5: ML estimates for Liu/Lux model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_0$	$\hat{\rho}$
Dow/Nik	1.014	0.913	1.512	0.207
	(0.031)	(0.024)	(0.027)	(0.018)
US/DM	0.849	0.901	1.397	0.020
	(0.021)	(0.020)	(0.017)	(0.011)
$TB1/TB2$	1.275	1.032	1.711	0.730
	(0.023)	(0.027)	(0.022)	(0.033)

Note: The number of cascade levels $n = 5$. We use the number of joint cascade level $k = 2$ for Dow/Nik, $k = 1$ for US/DM and $k = 2$ for $TB1/TB2$, respectively.

and two-day horizon, respectively; another conservative one for EW at the 1% level in five-day horizon.[3]

For U.S. treasury bond maturity rates, we find that results based on GMM and ML in both tables are successful at confidence level of 10%, except for the conservative hedge portfolios. At 5% level, there are two risky cases for $TB2$ and equal-weighted portfolio in Table 5.7, and risky EW in Table 5.8 at both two and five-day horizons, again, all HG are too conservative; of course, excessive conservativeness (as in the most cases of hedge portfolios) is also not an indication of superior risk management. At 1% level, we observe the positive results for individual $TB1$ and $TB2$ (except one risky $TB2$ at two-day horizon in Table 5.7).

We then move to the results from the Calvet/Fisher/Thompson model. For the stock indices, it shows the reasonable success for the individual Nikkei index in all scenarios and HG portfolio in two and five-day horizons in Table 5.9, but there are entirely unsuccessful

[3]One may notice there are insignificant estimates of $\hat{\rho}$ for the foreign exchange rates in Table 5.4, 5.5 and 5.6; we have conducted the study by setting $\hat{\rho} = 0$ and we omit these supplementary tables here due to very similar picture of the results.

Table 5.6: ML estimates for Calvet/Fisher/Thompson model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_1$	$\hat{m}_2$	$\hat{\rho}$
Dow/Nik	0.984	0.859	1.426	1.384	0.158
	(0.026)	(0.029)	(0.023)	(0.034)	(0.029)
US/DM	0.837	0.769	1.511	1.428	0.024
	(0.046)	(0.035)	(0.032)	(0.027)	(0.019)
$TB1/TB2$	1.202	1.038	1.491	1.527	0.780
	(0.039)	(0.032)	(0.030)	(0.043)	(0.042)

Note: The number of cascade levels $n = 5$.

for DOW in all scenarios. For results from ML, we find the similar success for NIK (except with two risky cases at at 10% and 1% levels) and DOW at the 1% level in Table 5.10; for EW and HG portfolios, we find only seven relative satisfied results which are out of total 18 cases.

In contrast, there are quite satisfactory results for foreign exchange rates. Table 5.9 reports only two conservative and one risky case at confidence levels of 10%, as well as three too conservative ones at confidence levels of 1%. In Table 5.10, we observe five conservative cases (two for equal-weighted and hedge portfolios respectively, one for DM) as well as three risky cases (only for US at the 5% and 1% levels) across all time horizons.

For U.S. treasury bond maturity rates, we find that VaRs coverage through GMM is successful for $TB1$ (except for one risky case at 1% level in Table 5.10) and $TB2$ across all confidence levels. As the results from the Liu/Lux model, we observe the failures for the too conservative hedge portfolios in all time horizons and confidence levels. In contrast, there is success for EW at confidence level of 10%, but leaves risky cases across other confidence levels.

In summary, from Table 5.7 to 5.10, the bivariate MF models show rather satisfactory performances, particularly with remarkable findings for foreign exchange rates. By comparing between Table 5.7 and Table 5.9 with a total of 108 scenarios in each, we find that the Liu/Lux model (24 failure cases) outperforms the Calvet/Fisher/Thompson model (40 failure cases), which signals that our more parsimonious design of multivariate MF model performs quite well. note that, the Liu/Lux model with only one common choice of the distribution for volatility components, that is two draws of $M_t \in \{m_0,\ \ m_1\}$, with $m_0 \in (0, 2)$ and the alternative $m_1 = 2 - m_0$; The Calvet/Fisher/Thompson approach uses two separate choices, namely two binomial distributions: $\{m_1, 2 - m_1\}$ and $\{m_2, 2 - m_2\}$ for the bivariate assets. The results also reveal that the simulated data based on GMM show better fit with empirical time series than the ones based on ML estimation. This is actually plausible: there are larger numbers of cascade levels ($n = 10$ for Nik/Dow and $TB1/TB2$, $n = 12$ for US/DM) employed within BMF models for GMM, while only $n = 5$number of cascade levels is used for ML due to its computational limitation.

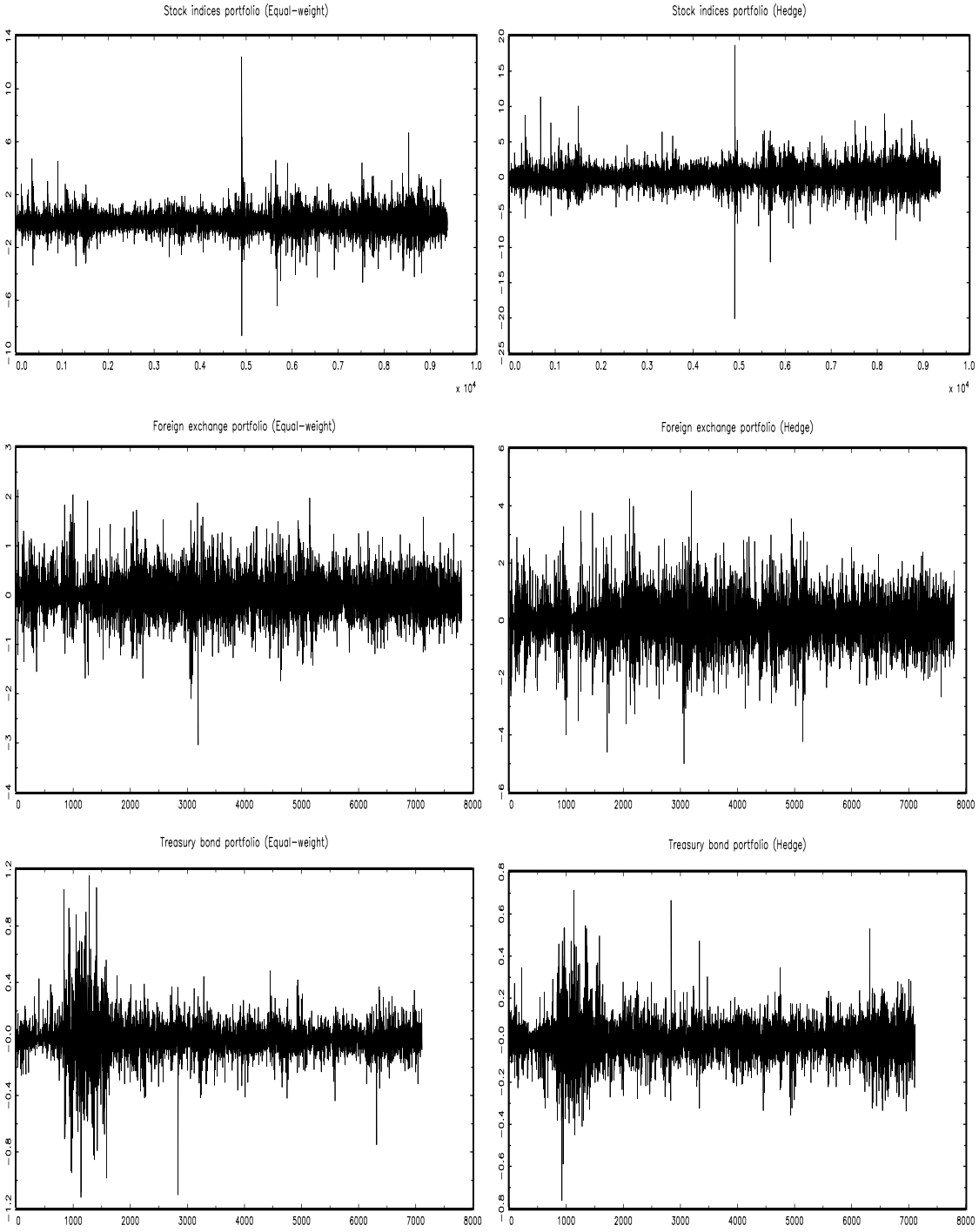

Figure 5.2: Empirical equal-weighted portfolio returns.

Table 5.7: Multi-period unconditional coverage (Liu/Lux model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	NIK	EW	HG	DOW	NIK	EW	HG	DOW	NIK	EW	HG
$Stocks$	$\alpha = 10\%$	0.1137	0.1055	0.0981	0.1104	0.1021	0.990	0.1007	0.1267^{+}	0.0962	0.1012	0.1086	0.0907
	$\alpha = 5\%$	0.0586	0.0510	0.0493	0.0533	0.0661^{+}	0.0520	0.0498	0.0472	0.0719^{+}	0.0583	0.0640^{+}	0.0377^{*}
	$\alpha = 1\%$	0.0108	0.0073	0.0112	0.031^{*}	0.0179^{+}	0.0090	0.0133	0.0061	0.0187^{+}	0.0110	0.0194^{+}	0.0034^{*}
		US	DM	EW	HG	US	DM	EW	HG	US	DM	EW	HG
FXs	$\alpha = 10\%$	0.1023	0.1023	0.1003	0.1030	0.0985	0.0950	0.1011	0.1031	0.0989	0.1014	0.0964	0.1014
	$\alpha = 5\%$	0.0533	0.0488	0.0515	0.0485	0.0520	0.0460	0.0470	0.0485	0.0438	0.0513	0.0401	0.0488
	$\alpha = 1\%$	0.0093	0.0090	0.0075	0.0063	0.0105	0.0100	0.0090	0.0105	0.0088	0.0075	0.0075	0.0075
		$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG
$Bonds$	$\alpha = 10\%$	0.1015	0.1106	0.987	0.0780^{*}	0.1056	0.1017	0.0949	0.0801^{*}	0.1013	0.1057	0.1003	0.0727^{*}
	$\alpha = 5\%$	0.0563	0.0583	0.0601	0.0327^{*}	0.0515	0.0490	0.0683^{+}	0.0373^{*}	0.0603	0.0710^{+}	0.0652^{+}	0.0373^{*}
	$\alpha = 1\%$	0.0076	0.0105	0.0210^{+}	0.0029	0.0098	0.0227^{+}	0.0244^{+}	0.0061	0.0057	0.0042	0.0260^{+}	0.0056

Note: This table shows the failure rate based on the GMM estimator of the BMF Binomial model. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). In-sample data is used for GMM estimation, using eight moments as in the Appendix. EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative scenarios, respectively.

Table 5.8: Multi-period coverage (Liu/Lux model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	NIK	EW	HG	DOW	NIK	EW	HG	DOW	NIK	EW	HG
$Stocks$	$\alpha = 10\%$	0.1044	0.0970	0.0899	0.1075	0.0985	0.0953	0.0968	0.1033	0.1064	0.0970	0.1005	0.0899
	$\alpha = 5\%$	0.0431	0.0511	0.0424	0.0401	0.0399	0.0683^{+}	0.0490	0.0501	0.0638^{+}	0.0569	0.0644^{+}	0.0478
	$\alpha = 1\%$	0.0040	0.0107	0.0049	0.0024^{*}	0.0188^{+}	0.0118	0.0171	0.0020	0.0183^{+}	0.0219^{+}	0.0166	0.0024^{*}
		US	DM	EW	HG	US	DM	EW	HG	US	DM	EW	HG
FXs	$\alpha = 10\%$	0.1295^{+}	0.1037	0.1110	0.1074	0.1223^{+}	0.0948	0.1071	0.1028	0.1058	0.0957	0.1058	0.0914
	$\alpha = 5\%$	0.0596	0.0419	0.0489	0.0459	0.0560	0.0405	0.0443	0.0463	0.0488	0.0419	0.0457	0.0426
	$\alpha = 1\%$	0.0074	0.0119	0.0048	0.0045	0.0075	0.0160	0.0058	0.0078	0.0063	0.0105	0.0027^{*}	0.0063
		$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG
$Bonds$	$\alpha = 10\%$	0.0978	0.1045	0.0966	0.0809^{*}	0.1028	0.0910	0.1002	0.0871^{*}	0.1034	0.0978	0.0970	0.0828^{*}
	$\alpha = 5\%$	0.0555	0.0461	0.0508	0.0377^{*}	0.0514	0.0534	0.0679^{+}	0.0381^{*}	0.0570	0.0599	0.0632^{+}	0.0365^{*}
	$\alpha = 1\%$	0.0097	0.0158	0.0226^{+}	0.0017^{*}	0.0094	0.0132	0.0223^{+}	0.0014^{*}	0.0073	0.0120	0.0250^{+}	0.0021^{*}

Note: This table shows the failure rate based on the ML estimator of the BMF binomial model. Stocks are Dow Jones Composite 65 Average Index and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). In-sample data is used for ML estimation. EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative scenarios, respectively.

Table 5.9: Multi-period unconditional coverage (Calvet/Fisher/Thompson model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	NIK	EW	HG	DOW	NIK	EW	HG	DOW	NIK	EW	HG
$Stocks$	$\alpha = 10\%$	0.1486^{+}	0.0990	0.1514^{+}	0.1498^{+}	0.1545^{+}	0.0760	0.1430^{+}	0.0930	0.1525^{+}	0.0850	0.1200	0.0975
	$\alpha = 5\%$	0.0833^{+}	0.0477	0.0725^{+}	0.0843^{+}	0.0805^{+}	0.0405	0.0710^{+}	0.0455	0.0887^{+}	0.0525	0.0912^{+}	0.0438
	$\alpha = 1\%$	0.0247^{+}	0.0133	0.0215^{+}	0.0115	0.0335^{+}	0.0145	0.0220^{+}	0.0197^{+}	0.0350^{+}	0.0163	0.0288^{+}	0.0078
		US	DM	EW	HG	US	DM	EW	HG	US	DM	EW	HG
FXs	$\alpha = 10\%$	0.0105	0.0113	0.092	0.1050	0.0939	0.097	0.0980	0.1284^{*}	0.0860^{*}	0.1529^{+}	0.1047	0.936
	$\alpha = 5\%$	0.0435	0.0568	0.0601	0.0566	0.0537	0.0485	0.0530	0.0531	0.0472	0.0439	0.0472	0.0549
	$\alpha = 1\%$	0.0125	0.0030^{*}	0.0061	0.0089	0.0110	0.0027^{*}	0.0046	0.0131	0.0144	0.0033^{*}	0.0065	0.0089
		$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG
$Bonds$	$\alpha = 10\%$	0.1028	0.1100	0.1094	0.0764^{*}	0.1025	0.1085	0.0954	0.0813^{*}	0.1047	0.0981	0.1293^{+}	0.0831^{*}
	$\alpha = 5\%$	0.0532	0.0474	0.0497	0.0385^{*}	0.0532	0.0608	0.0553	0.0320^{*}	0.0540	0.0520	0.0603	0.0355^{*}
	$\alpha = 1\%$	0.0047	0.0171	0.0261^{+}	0.0030^{*}	0.0061	0.0095	0.0274^{+}	0.0019^{*}	0.0057	0.0120	0.0226^{+}	0.0021^{*}

Note: This table shows the failure rate based on the GMM estimator of the Calvet/Fisher/Thompson model. Stocks are Dow Jones Composite 65 Average Index (Dow) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). In-sample data is used for GMM estimation, which is based on the Calvet/Fisher/Thompson model. EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative scenarios, respectively.

Table 5.10: Multi-period unconditional coverage (Calvet/Fisher/Thompson model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	0.1204^{+}	0.0967	0.1006	0.1246^{+}	0.1205^{+}	0.1134	0.1059	0.1172^{+}	0.0842^{*}	0.0921	0.0993	0.1220^{+}
	$\alpha = 5\%$	0.0499	0.0531	0.0598	0.0528	0.0638^{+}	0.0470	0.0395^{*}	0.0373^{+}	0.0389^{*}	0.0555	0.0620^{+}	0.0618^{+}
	$\alpha = 1\%$	0.0038	0.0124^{+}	0.0018^{*}	0.0028	0.0036	0.0042	0.0023^{*}	0.0035	0.0040	0.0121^{+}	0.0128^{+}	0.0189^{+}
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	0.1032	0.0987	0.0987	0.1010	0.0928	0.0934	0.0977	0.0917	0.1013	0.0995	0.0980	0.1096
	$\alpha = 5\%$	0.0465	0.0558	0.0383^{*}	0.0391^{*}	0.0455	0.374^{*}	0.0559	0.0408	0.0610^{+}	0.0547	0.0481	0.0392^{*}
	$\alpha = 1\%$	0.0059	0.0104	0.0017^{*}	0.0040	0.0191^{+}	0.011	0.0046	0.0062	0.0186^{+}	0.014	0.0060	0.0066
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	0.0987	0.1049	0.1046	0.0805^{*}	0.1052	0.1091	0.0959	0.0864^{*}	0.1038	0.1011	0.0982	0.0778^{*}
	$\alpha = 5\%$	0.0454	0.0456	0.0644^{+}	0.0363^{*}	0.0571	0.0500	0.0629^{+}	0.0317^{*}	0.0535	0.0474	0.0689^{+}	0.0359^{*}
	$\alpha = 1\%$	0.0061	0.0102	0.0194	0.0051	0.0260^{+}	0.0034	0.0229^{+}	0.0011^{*}	0.0113	0.0069	0.0278^{+}	0.0021^{*}

Note: This table shows the failure rate based on the ML estimator of the Calvet/Fisher/Thompson model. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). In-sample data is used for ML estimation, which is based on the Calvet/Fisher/Thompson model. EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative scenarios, respectively.

5.5 Value-at-Risk

To measure and control the potential movements of financial markets, there has been extensive research into this issue by academics and financial institutions. In ideal financial risk management, a prioritization process is followed, by which the risks with the greatest loss and the greatest probability of occurrence are handled first. For instance, according to the Basle Committee (1996), the risk capital of a bank must be sufficient to cover losses on the banks' trading portfolio over a 10-day holding period on 99% of occasions, this is termed Value-at-Risk (VaR), one of the best-known tools used to measure, gear and control financial market risks. One alternative measure of financial risk is the so-called Expected Shortfall (ES), which refers to the expected loss conditional on the losses exceeding the VaR at the target period horizon.

With the increasing demand for reliable quantitative risk measurement and management instruments, Value-at-Risk (VaR) has emerged as one of the most prominent tools of downside market risk. It was defined in Riskmetrics,[4] and was claimed to be proportional to the computed standard deviation of the pertinent portfolio (often assuming normality); it provides a quantitative and synthetic measure of financial risk. Numerous applications have appeared since VaR was introduced, some financial institutions and related organizations have recommended VaR as an alternative to the traditional mean-variance efficient frontiers for portfolio selection; other applications of VaR have been used in the risk management framework of internal supervision in order to mitigate the agency problem.

In response to the increasing interest from both academia and industry, various VaR approaches have been introduced; existing methodologies for calculating VaR differ in a number of respects, namely: non-parametric historical simulation methods which estimate VaR by using the sample quantile estimate based on historic return data;[5] fully parametric

[4] J. P Morgan. 3rd edn. 1995, Riskmetrics Technical Document.

[5] There are several varieties of this method with advantages and disadvantages, see Dowd (2002) and Christoffersen (2003).

methods based on econometric models for volatility dynamics which often impose certain distribution assumptions, such as GARCH type processes; and other methods based on extreme value theory (EVT) which is concerned with the distribution of the smallest and largest order statistics and focuses only on the tails of the return distribution.[6] Some recent works including McNeil and Frey (2000), Holton (2003), Kuester et al. (2006). Despite these variations, they share a common point which is related to the inference of the profit-and-loss distribution of given portfolios; in particular, it measures the worst loss over a specified target horizon within a given statistical confidence level, or, from a mathematical viewpoint, VaR represents a quantile of an estimated profit-loss distribution.

Recalling the studies we have conducted in the previous section, the unconditional coverage of out of sample data based on the in-sample estimates might be viewed as an implicit Value-at-Risk 'forecast', by assuming that observations are equally likely for the out of sample window (i.e. independent and identically-distributed, iid). However, this unrealistic assumption of future returns in a particular period being equally likely, fails to be satisfied, since the unconditional probabilities would not respond to the arrival of changing information and the condition of an i.i.d is clearly violated by the fact – the distribution of portfolio returns typically changing over time.

As a result, we further extrapolate value-at-risk by using multifractal process approaches, and exam their performances regarding to the forecasting ability. Instead of assuming that VaR is one particular quantile of the unconditional distribution of portfolio returns, it is more plausible to construct VaR forecasts based on current information. Let I_t be the information set until time t, the forward looking h-period return at time t being $r_{t,t+h}$ (note that, we use the cumulative return $\tilde{r}_{t,t+h}$ as in Eq. 5.4.1 in the unconditional coverage instead), and Value-at-Risk at the h-period horizon be defined by:

[6] A comprehensive overview of EVT is provided by Embrechts et al. (1997).

$$Pr\left(r_{t:t+h} \leq VaR^{\alpha}_{t:t+h} | I_t\right) = \alpha. \tag{5.5.1}$$

It places an upper bound on losses in the sense that these will exceed the VaR threshold with only a pre-assumed target probability. More specifically, conditional on the information given up to time t, the Value-at-Risk for period $t = h$ of one unit of portfolio is the $(1-\alpha)$th quantile of the conditional distribution $r_{t:t+h}$. In this section, we study the Value-at-Risk and compare the performances of our model with ones based on the Calvet/Fisher/Thompson model and bivariate CC-GARCH.

For CC-GARCH (1, 1) model, it provides a closed form solution for a one-day VaR forecast, that is,

$$VaR^{\alpha}_{t:t+1} = \mu_t + Q_{1-\alpha}\sigma_t, \tag{5.5.2}$$

and $Q_{1-\alpha}$ is the $(1-\alpha)$th quantile of the standard Normal distribution; σ_t is the square root of the conditional volatility (standard deviation) implied from CC-GARCH (via the GAUSS module 'Fanpac'); VaR forecasts for more than one day are implemented through simulations.

The algorithm particle filter in Chapter 3 provides us with a way of calculating $VaR^{\alpha}_{t:t+h}$ as of Calvet et al. (2006). We simulate each volatility component draw M_t one-step-ahead by using SIR introduced in Section 3.4:

1. After having estimated the parameters with in-sample data, we invoke once more the particle filter algorithm by starting at $t = 0$ with drawing $M_0^{(1)}, \ldots, M_0^{(B)}$ from the initial condition π_0. For $t \geq 1$, we simulate each $\{\hat{M}_t^{(b)}\}_{b=1}^{B}$ independently and reweighting to obtain the importance sampler $\{M_t^{(b)}\}_{b=1}^{B}$ via:

2. Draw a random number q from 1 to B with the probability of:

$$P(q = b) = \frac{f(r_t | M_t = m^{(b)})}{\sum_{i=1}^{B} f(r_t | M_t = m^{(i)})}.$$

3. $M_t^{(1)} = \hat{M}_t^{(q)}$ is then selected, repeat Step 2 B times and obtain B draws with $M_t^{(1)}, \ldots M_t^{(B)}$.

4. After the last iteration of the in-sample series (say time t) by reaching M_t, we simulate the Markov chain one-step-ahead to obtain $\hat{M}_{t+1}^{(1)}$ given $M_t^{(1)}$, repeat B times to generate draws $\{\hat{M}_{t+1}^{(b)}\}_{b=1}^{B}$, i.e., $\hat{M}_{t+1}^{(1)}, \ldots, \hat{M}_{t+1}^{(B)}$, which are used for one-step ahead forecast, i.e., to move from t to forecast for $t+1$.

5. For h-period forecast given information up to time t, iterating the particles obtained from importance resampling at time t h times to obtain h-period ahead volatility draws $\hat{M}_{t+h}^{(1)}, \ldots, \hat{M}_{t+h}^{(B)}$, which are used to forecast for $t+h$ from t. For all cases. we use $B = 10000$ simulated draws.

6. Simulate bivariate Normal innovations, which need to be combined with new volatility draws drawn to calculate VaR.

7. For the next one-step ahead forecast, i.e., to move from $t+1$ to forecast for $t+2$, apply the SIR via Step 2 and Step 3 to obtain importance sampler $\{M_{t+1}^{(b)}\}_{b=1}^{B}$, then simulate the Markov chain one-step-ahead to generate draws $\{\hat{M}_{t+2}^{(b)}\}_{b=1}^{B}$ given $\{M_{t+1}^{(b)}\}_{b=1}^{B}$.

This recursive procedure provides a discrete approximation to Bayesian updating, which assists to simulate the bivariate series forward other h-day horizons. Thus, we calculate $VaR_{t:t+h}^{\alpha}$ as the $(1-\alpha)$th simulated quantile.

Table 5.11 and Table 5.12 report the empirical estimates from the particle filter for bivariate MF models by using in-sample data. For the Calvet/Fisher/Thompson model, we fixed $n = 8$ as in Calvet et al. (2006); For Liu/Lux model, we set the number of joint cascade suggested in the empirical study of Section 3.6, namely $k = 3$ for Dow/Nik; $k = 2$ for US/DM and $k = 5$ for $TB1/TB2$. We assess the models' performances by computing the failure rate for the individual returns, equal weight portfolio and hedge portfolio. By

Table 5.11: SML estimates for Liu/Lux model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_0$	$\hat{\rho}$
Dow/Nik	1.133	0.925	1.374	0.286
	(0.022)	(0.021)	(0.023)	(0.015)
US/DM	0.739	0.642	1.411	0.063
	(0.020)	(0.022)	(0.029)	(0.016)
$TB1/TB2$	0.791	0.946	1.455	0.792
	(0.025)	(0.021)	(0.028)	(0.020)

Note: We use the number of joint cascade suggested in the empirical study of Section 3.6, namely $k = 3$ for Dow/Nik, $k = 2$ for US/DM, $k = 5$ for $TB1/TB2$, respectively.

definition, the failure rate is the number of times returns exceed the forecasted VaR, which is expected to be close to the pre-specified VaR level given that the model is well specified. A selection of the conditional VaR forecasts plots can be found from Figure 5.3 to 5.5. The results under the alternative modeling assumptions are reported in Table 5.13, 5.14, and 5.15 corresponding to the Liu/Lux model, the Calvet/Fisher/Thompson model and the bivariate CC-GARCH(1, 1) model (empirical estimates have been reported in Section 5.3). The standard errors are calculated as ones in the study of unconditional coverage (Section 5.4), cf. Eq. (5.4.3).

We first look at the stock indices: Table 5.13 presents its success for NIK, equal weight and hedge portfolios in all scenarios, however, there are three risky cases of DOW at 10% level, as well as one case at 5% level and 1% level in five-day horizon, respectively; In contrast, for the results from the Calvet/Fisher/Thompson model, Table 5.14 shows the satisfactory results for NIK, EW (except with one risky case in one-day horizon) and HG (except with one risky case in five-day horizon), but leaves one risky case for DOW at 10% level in two-day horizon, as well as two cases at 10% and 5% levels in five-day horizon; The

Table 5.12: SML estimates for Calvet/Fisher/Thompson model (in-sample data)

	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{m}_1$	$\hat{m}_2$	$\hat{\rho}$
Dow/Nik	1.167	0.970	1.433	1.375	0.173
	(0.029)	(0.032)	(0.041)	(0.038)	(0.021)
US/DM	0.758	0.630	1.452	1.419	0.047
	(0.027)	(0.022)	(0.043)	(0.037)	(0.015)
$TB1/TB2$	0.880	1.043	1.239	1.448	0.813
	(0.029)	(0.032)	(0.057)	(0.052)	(0.022)

Note: The number of cascade levels $n = 8$ as in Calvet et al. (2006).

results from the CC-GARCH model reported in Table 5.15 are moderate for cases at 10% and 5% levels, but present the failure in most cases at 1% level.

For foreign exchange rates, there are quite convincing results for both Liu/Lux model and Calvet/Fisher/Thompson model. In Table 5.14, there are one risky case for HG at 10% level in one-day horizon, one conservative case of equal-weighted portfolio at 1% level in two-day horizon, and two conservative cases in five-day horizon; In Table 5.13, there are one too risky case of HG in one-day horizon, and additional three too conservative cases for equal-weighted portfolio in two-day and five-day horizons at 5% and 1% levels; While the VaR forecast based on the CC-GARCH(1, 1) model in Table 5.15, shows its failure in most cases at 1% level (except with US and HG in one and two-day horizons), as well as EW and HG portfolios in one-day horizon; US in two-day horizon; HG in five-day horizon at 5% level.

For U.S. bond maturity rates, Table 5.13 demonstrates its success mainly for $TB2$ and EW in all scenarios, but leaves the too conservative cases of hedge portfolios, as

well as other five cases of $TB1$ across three time horizons. For the results from the Calvet/Fisher/Thompson model, Table 5.14 reports some similar results, namely, too conservative cases for HG portfolio, also conservative ones of $TB1$ in five-day horizon at all levels, as well as additional ones at 10% and 5% levels in one-day and two-day horizons; Again, there are number of failures for HG in all scenarios found in the VaR forecasts based on the bivariate CC-GARCH(1, 1) model, while all $TB2$ are successful.

The pictures from the conditional VaR forecasts with different models are quite encouraging, in particular for the stock indices and foreign exchange rates results based on multifractal models (both Table 5.13 and Table 5.14), though there are a couple of cases tending to underestimate the frequency of extreme returns, as well as certain overestimation (too conservative predictions) scenarios for U.S. treasury bond rate, e.g., hedge portfolio forecasts. A glance at these results suggests that the performances based on both multifractal models clearly dominate the one based on CC-GARCH model. Indeed, VaR can help risk managers to estimate the cost of positions, allowing them to allocate risk in a more efficient way. Also the Basle Committee on Banking Supervision (1996) requires financial institutions such as, banks and investment firms to use VaR to measure their capital operations. However, if the underlying risk is not properly estimated, these requirements may lead to overestimation (or underestimation) of market risks and consequently to maintaining excessively high (low) capital levels.

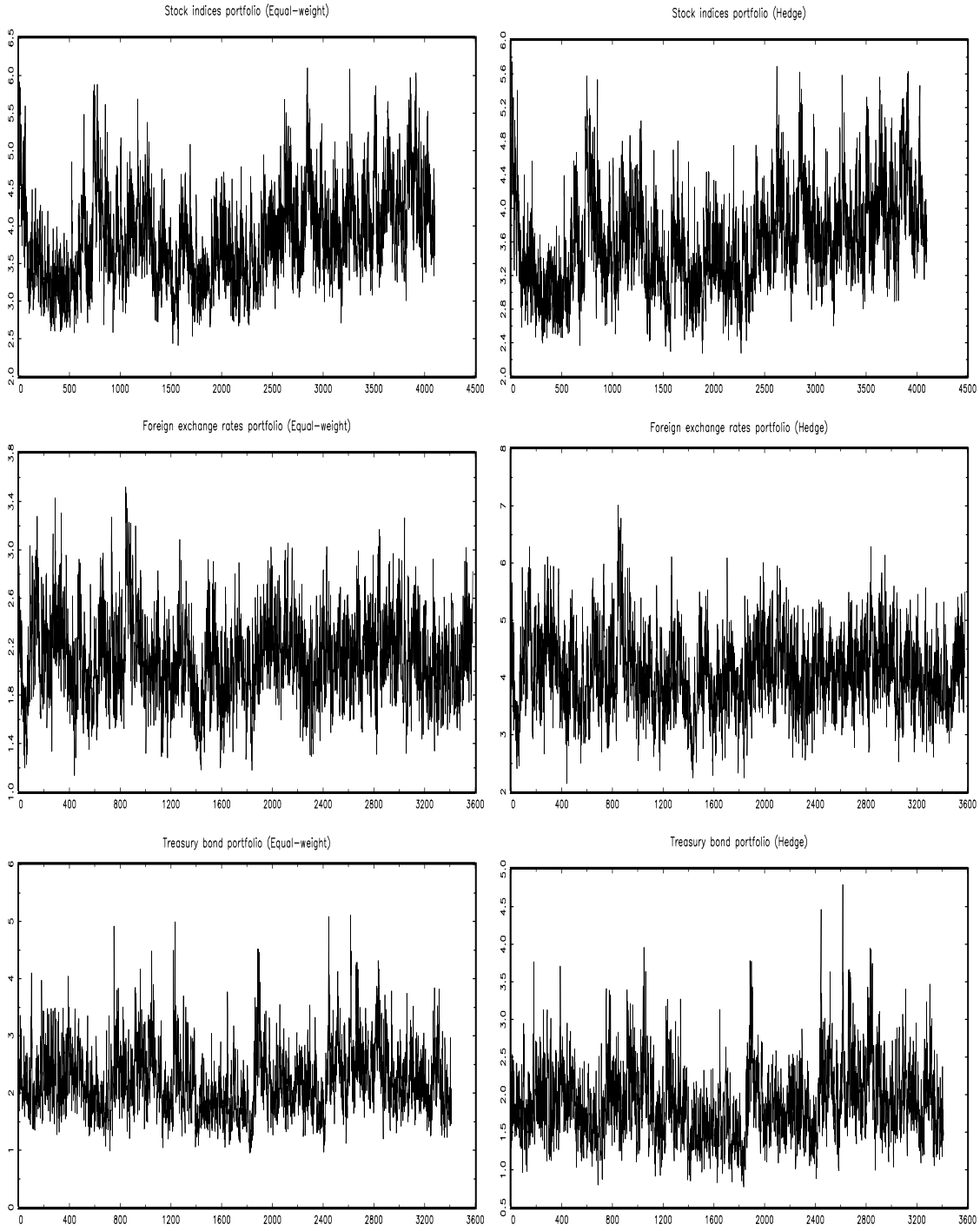

Figure 5.3: One-step ahead VaR predictions for $\alpha = 1\%$ under the Liu/Lux model.

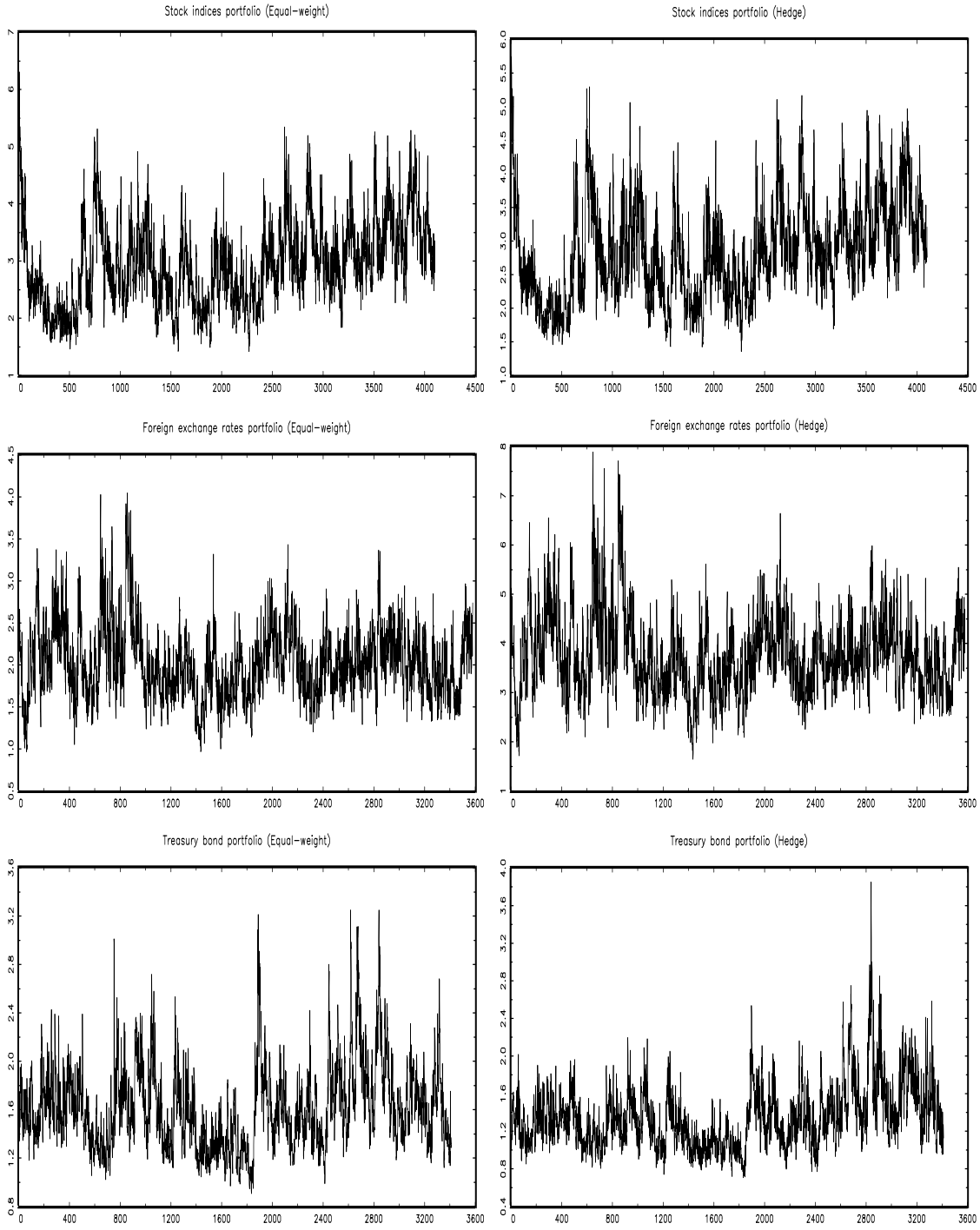

Figure 5.4: One-step ahead VaR predictions for $\alpha = 1\%$ under the Calvet/Fisher/Thompson model.

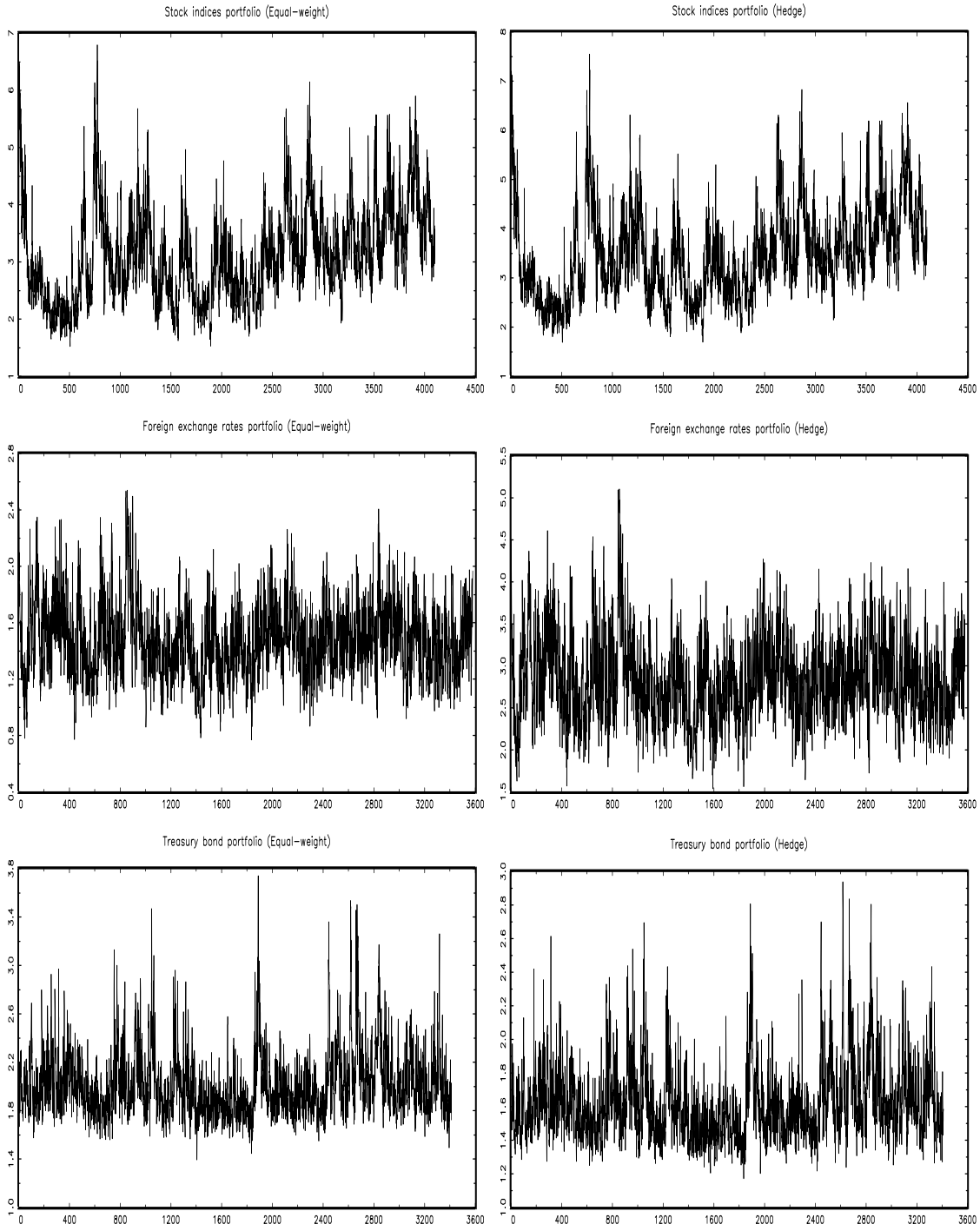

Figure 5.5: One-step ahead VaR predictions for $\alpha = 1\%$ under the CC-GARCH(1, 1) model.

Table 5.13: Failure rates for multi-period Value-at-Risk forecasts (Liu/Lux model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	NIK	EW	HG	DOW	NIK	EW	HG	DOW	NIK	EW	HG
$Stocks$	$\alpha = 10\%$	0.1166^{+}	0.1020	0.1129	0.1002	0.1173^{+}	0.1022	0.1147	0.1039	0.1204^{+}	0.1023	0.1174	0.1037
	$\alpha = 5\%$	0.0588	0.0500	0.0556	0.0529	0.0595	0.0503	0.0542	0.0517	0.0647^{+}	0.0513	0.0566	0.0568
	$\alpha = 1\%$	0.0146	0.0078	0.0127	0.0117	0.0144	0.0068	0.0102	0.0129	0.0173^{+}	0.0083	0.0137	0.0127
		US	DM	EW	HG	US	DM	EW	HG	US	DM	EW	HG
FXs	$\alpha = 10\%$	0.1076	0.0995	0.1028	0.1175^{+}	0.1126	0.0995	0.1056	0.1118	0.1091	0.1015	0.1071	0.1107
	$\alpha = 5\%$	0.0472	0.0513	0.0572	0.0474	0.0478	0.0422	0.0419	0.0447	0.0456	0.0464	0.0522	0.0450
	$\alpha = 1\%$	0.0079	0.0061	0.0053	0.0057	0.0045	0.0045	0.0030^{*}	0.0053	0.0031^{*}	0.0046	0.0025^{*}	0.0056
		$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG	$TB1$	$TB2$	EW	HG
$Bonds$	$\alpha = 10\%$	0.0899	0.1070	0.0918	0.0822^{*}	0.0893	0.1115	0.0910	0.0781^{*}	0.0817^{*}	0.1028	0.0891	0.0855^{*}
	$\alpha = 5\%$	0.0320^{*}	0.0452	0.0417	0.0339^{*}	0.0323^{*}	0.0402	0.0429	0.0323^{*}	0.0402	0.0435	0.0438	0.0385^{*}
	$\alpha = 1\%$	0.0063	0.0052	0.0046	0.0023^{*}	0.0023^{*}	0.0042	0.0049	0.0013^{*}	0.0022^{*}	0.0057	0.0043	0.0020^{*}

Note: This table shows the failure rate (proportion of observations above the VaR). Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative VaR, respectively. The standard errors are calculated as ones in the study of unconditional coverage, cf. Eq. (5.4.3).

Table 5.14: Failure rates for multi-period Value-at-Risk forecasts (Calvet/Fisher/Thompson model)

		One day horizon				Two days horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	0.1101	0.1034	0.1061	0.1022	0.1208^{+}	0.1030	0.1010	0.1054	0.1223^{+}	0.0991	0.1096	0.1044
	$\alpha = 5\%$	0.0571	0.0537	0.0493	0.0505	0.0573	0.0529	0.0510	0.0503	0.0645^{+}	0.0483	0.0525	0.0546
	$\alpha = 1\%$	0.0110	0.0076	0.0162^{+}	0.0114	0.0110	0.0083	0.0107	0.0122	0.0128	0.0073	0.0095	0.0159^{+}
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	0.1081	0.0942	0.0926	0.1065	0.1068	0.0925	0.0904	0.1059	0.1004	0.0906	0.0953	0.1034
	$\alpha = 5\%$	0.0464	0.0474	0.0560	0.0581	0.0442	0.0400	0.0347^{*}	0.0500	0.0483	0.0455	0.0322^{*}	0.0447
	$\alpha = 1\%$	0.0079	0.0083	0.0077	0.0176^{+}	0.0110	0.0068	0.0076	0.0114	0.0052	0.0049	0.0027^{*}	0.0058
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	0.0816^{*}	0.1030	0.0916	0.0770^{*}	0.0772^{*}	0.0992	0.0927	0.0746^{*}	0.0785^{*}	0.0978	0.0966	0.0793^{*}
	$\alpha = 5\%$	0.0374^{*}	0.0484	0.0481	0.0344^{*}	0.0357^{*}	0.0475	0.0443	0.0376^{*}	0.0755^{*}	0.0452	0.0405	0.0338^{*}
	$\alpha = 1\%$	0.0056	0.0103	0.0125	0.0049	0.0052	0.0065	0.0082	0.0018^{*}	0.0055	0.0068	0.0079	0.0016^{*}

Note: This table shows the failure rate (proportion of observations above the VaR). Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rates of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge, a zero investment portfolio. $+$ and $*$ denote too risky and too conservative VaR, respectively. The standard errors are calculated as ones in the study of unconditional coverage, cf. Eq. (5.4.3).

Table 5.15: Failure rates for multi-period Value-at-Risk forecasts (CC-GARCH)

		One day horizon				Two days horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	0.0968	0.0835	0.0955	0.0923	0.1065	0.0874	0.0960	0.1013	0.0928	0.0853	0.0993	0.1101
	$\alpha = 5\%$	0.0545	0.0480	0.0570	0.0505	0.0527	0.0499	0.0443	0.0517	0.0576	0.0543	0.0552	0.0721^{+}
	$\alpha = 1\%$	0.0243^{+}	0.0178^{+}	0.0218^{+}	0.0193^{+}	0.0144	0.0229^{+}	0.0205^{+}	0.0155	0.0094	0.0203^{+}	0.0150	0.0261^{+}
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	0.0968	0.1041	0.1026	0.1137	0.0878	0.0950	0.1024	0.0973	0.0931	0.1072	0.0913	0.1132
	$\alpha = 5\%$	0.0408	0.0601	0.0625^{+}	0.0629^{+}	0.0332^{*}	0.0554	0.0651^{+}	0.0609^{+}	0.0564	0.0584	0.0461	0.0704^{+}
	$\alpha = 1\%$	0.0062	0.0272^{+}	0.0290^{+}	0.0262^{+}	0.0050	0.0220^{+}	0.0229^{+}	0.0041	0.0226^{+}	0.0194^{+}	0.0176^{+}	0.0229^{+}
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	0.0707^{*}	0.0965	0.1074	0.0669^{*}	0.0686^{*}	0.0933	0.0868	0.0690^{*}	0.0614^{*}	0.0948	0.0807	0.0689^{*}
	$\alpha = 5\%$	0.0460	0.0537	0.0616^{+}	0.0367^{*}	0.0402	0.0540	0.0478	0.0349^{*}	0.0414	0.0502	0.0396	0.0366^{*}
	$\alpha = 1\%$	0.0129	0.0061	0.0235^{+}	0.0029^{*}	0.0027^{*}	0.0081	0.0041	0.0029^{*}	0.0038	0.0114	0.0094	0.0024^{*}

Note: This table shows the failure rate (proportion of observations above the VaR) based on the bivariate CC-GARCH model. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK); FXs are Foreign Exchange rates of U.S. Dollar (US) and German Mark (DM) to British Pound; Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge, as zero investment portfolio. $+$ and $*$ denote too risky and too conservative VaR, respectively. The standard errors are calculated as ones in the study of unconditional coverage, cf. Eq. (5.4.3).

5.6 Conditional Expected shortfall

Value-at-Risk seems to have become a standard tool used in financial risk management with three main attributes: it allows the potential loss associated with a decision to be quantified; it summarizes complex positions in a single figure; and it is intuitive - expressing loss in monetary terms.[7] However, there are a few shortcomings of VaR. For example, Artzner et al. (1997) argues that:

(1) VaR reports only percentiles of profit-loss distributions, thus disregarding any loss beyond the VaR level (this is called the 'tail risk' problem);

(2) VaR is not coherent, since it is not sub-additive. A risk measure is sub-additive when the risk of the total position is less than or equal to the sum of the risk of individual portfolios. It may be troublesome to base a risk-management system solely on VaR limits for individual books.

Conditional Expected Shortfall (ES) has been proposed to alleviate the arguments inherent in VaR; it is defined as the expected losses conditional on exceeding the VaR, at the h-period horizon. The conditional Expected Shortfall is hence given by:

$$ES^{\alpha}_{t:t+h} = E[\left(r_{t:t+h} | r_{t:t+h} \leq VaR^{\alpha}_{t:t+h}\right) | I_t]. \tag{5.6.1}$$

Thus, by the definition of Eq. (5.6.1), ES considers the loss beyond the VaR level, in particular, it gives information about the size of the potential losses, given that loss larger than VaR occurred. It has been shown to be sub-additive (the appendix of Acerbi et al. (2001) gives the details) which assures its coherence as a risk measure. On these grounds, a number of studies turned their attention towards Expected Shortfall. Although it has not become a standard in the financial industry, conditional ES is likely to play a major role as it currently does in the insurance industry, see Embrechts et al. (1997). Also, conditional ES is used in credit risk studies, see Bucay and Rosen (1996). Some recent works include:

[7] Despite its conceptual simplicity, the measurement of VaR is a very challenging statistical problem and none of the methodologies developed so far gives satisfactory solutions, see Engle and Manganelli (2002).

McNeil and Frey (2000), who presents both unconditional and conditional VaR and ES, and advocates conditional ES as an alternative tool with good theoretical properties by fitting different GARCH type models; Yamai and Yoshiba (2002) provides an overview of studies comparing VaR and ES by drawing implications for financial risk measurement, and also illustrates practical problems among various back-testing methods.

In this section, we assess bivariate multifractal models in terms of the performance of the multi-period ES forecasts. For comparison, conditional Expected Shortfall forecasts based on the Calvet/Fisher/Thompson model and the CC-GARCH(1, 1) model were also computed. The forecasting results are reported in Table 5.16 to 5.18, respectively. Numbers inside the parentheses are the empirical ESs; numbers without parentheses are the simulated ESs based on these three models; bold numbers show those cases for which we cannot reject identity of the empirical and simulated ES, i.e. the empirical value fall into the range between the 2.5 to 97.5 percent quantile of the simulated ones, which are corresponding to successful forecasts. We summarize those positive results as below:

Table 5.16 reports the ES forecast based on the Liu/Lux model, which shows quite positive results for stock indices in one-day horizon, except with one risky case of equal-weighted portfolio at the 1% level, in five-day horizon it shows two failure cases of DOW; For foreign exchange rates, the ES forecasts are quite successful, and there are only one risky cases for equal-weighted at 1% level and two risky hedge portfolios in five-day horizon; For U.S. bonds, it reports the too conservative results (the simulated ES above empirical one) for all HG scenarios, as well as individual $TB1$ at 10% and 5% level in five-day horizon.

The performance of the Calvet/Fisher/Thompson model in Table 5.17 shows pretty similar positive scenario with the one of Liu/Lux model for stock indices, except with one risky EW at the 1% level in one-day horizon, two risky forecasts of EW and HG portfolios at the 1% level in fiver-day horizon; The results for foreign exchange rates present the success at all cases of 10% and 5% levels; but at the 1% level, we observe the risky HG

portfolio in one-day horizon, too conservative EW and HG portfolios at the 1% level in five-day horizon; For U.S. treasury bond maturity rates, there are too conservative $TB1$ (except with one case in one-day horizon), as well as hedge portfolio at all levels in both one and five-day horizons.

For CC-GARCH(1, 1) model, Table 5.18 reports the successful forecasts for scenarios (except for HG in five-day horizon)at 10% and 5% levels, in contrast, there are too risky EW and HG portfolios at 1% level; For foreign exchange rates, we find the similar results as the stock indices, and at the 1% level, there are failures in DM and HG in one-day horizon, US and HG in five-day horizon; forecasting U.S. treasury bond maturity shows the disappointing results at 1% level, also the too conservative $TB1$ and hedge portfolio at 10% and 5% levels.

We have reported the forecast performances by using both VaR and Expected Shortfall. A rough comparison shows that there are agreements in most cases between the results from both risk management tools, though there also are some discrepancies reported with different approaches; For instance, forecasting the foreign exchange rates based on the Liu/Lux model, we observe the positive performances for hedge portfolios at 5% and 1% levels in five-day horizon in Table 5.13; but Table 5.16 repots the opposite results with Expected Shortfall forecasts; Although the debate on these two risk instruments in terms of the pro and con is continuing, there are mounting empirical studies showing the different preferences on the choice of risk measurement tools, cf. McNeil and Frey (2000) and Yamai and Yoshiba (2002).

Table 5.16: Multi-period Expected shortfall forecasts (Liu/Lux model)

		One day horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	(2.21) **1.79**	(1.80) **1.72**	(3.02) **2.86**	(2.60) **2.56**	(2.21) 1.70	(1.80) **1.71**	(3.02) **2.89**	(2.60) **2.42**
	$\alpha = 5\%$	(2.84) **2.65**	(2.39) **2.33**	(3.89) **3.73**	(3.25) **3.20**	(2.84) 2.41	(2.39) **2.48**	(3.89) **3.70**	(3.25) **3.17**
	$\alpha = 1\%$	(4.36) **4.18**	(4.40) **4.56**	(6.48) 5.93	(5.09) **4.81**	(4.36) **4.03**	(4.40) **4.61**	(6.48) **6.21**	(5.09) **4.88**
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	(1.84) **1.82**	(1.79) **1.70**	(2.51) **2.35**	(2.44) **2.09**	(1.84) **1.77**	(1.79) **1.69**	(2.51) **2.37**	(2.44) **2.27**
	$\alpha = 5\%$	(2.34) **2.55**	(2.26) **2.14**	(3.08) **3.03**	(2.99) **3.12**	(2.34) **2.50**	(2.26) **2.13**	(3.08) **3.19**	(2.99) 3.42
	$\alpha = 1\%$	(3.57) **3.79**	(3.39) **3.57**	(4.34) **4.69**	(4.22) **4.30**	(3.57) **3.73**	(3.39) **3.55**	(4.34) 4.80	(4.22) 4.70
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	(0.92) **1.22**	(1.32) **1.25**	(1.08) **1.03**	(0.62) 1.34	(0.92) 1.38	(1.32) **1.17**	(1.08) **1.15**	(0.62) 1.87
	$\alpha = 5\%$	(1.17) **1.49**	(1.67) **1.77**	(1.37) **1.23**	(0.78) 2.02	(1.17) 1.82	(1.67) **1.52**	(1.37) **1.50**	(0.78) 2.17
	$\alpha = 1\%$	(1.77) **2.03**	(2.50) **2.74**	(2.07) **2.25**	(1.14) 2.48	(3.77) **3.40**	(2.50) **2.75**	(2.07) **2.26**	(1.14) 2.91

Note: This table reports the Expected Shortfall (ES) forecast based on bivariate MF model, the numbers in parentheses are the empirical realized ES values. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK), FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound, Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge which is zero investment portfolio. Numbers inside parentheses are empirical ES, and numbers outside parentheses are corresponding ES obtained by forecast. Bold numbers show those cases for which we cannot reject identity of the empirical and simulated ES, i.e. the empirical value falls into the range between the 2.5 to 97.5 percent quantile of the simulated ones.

Table 5.17: Multi-period Expected shortfall forecasts (Calvet/Fisher/Thompson model)

		One day horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	(2.21) **1.95**	(1.80) **1.73**	(3.02) **2.90**	(2.60) **2.35**	(2.21) **1.94**	(1.80) **1.84**	(3.02) **2.95**	(2.60) **2.39**
	$\alpha = 5\%$	(2.84) **2.98**	(2.39) **2.54**	(3.89) **3.77**	(3.25) **3.13**	(2.84) **2.76**	(2.39) **2.44**	(3.89) **3.71**	(3.25) **3.22**
	$\alpha = 1\%$	(4.36) **4.20**	(4.40) **4.32**	(6.48) 5.10	(5.09) **5.27**	(4.36) **4.16**	(4.40) **4.51**	(6.48) 5.70	(5.09) 4.52
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	(1.84) **1.78**	(1.79) **1.72**	(2.51) **2.70**	(2.44) **2.20**	(1.84) **1.70**	(1.79) **1.88**	(2.51) **2.74**	(2.44) **2.30**
	$\alpha = 5\%$	(2.34) **2.25**	(2.26) **2.14**	(3.08) **3.18**	(2.99) **2.87**	(2.34) **2.44**	(2.26) **2.41**	(3.08) **3.41**	(2.99) **3.18**
	$\alpha = 1\%$	(3.57) **3.46**	(3.39) **3.09**	(4.34) **4.18**	(4.22) 3.51	(3.57) **3.66**	(3.39) **3.50**	(4.34) 5.12	(4.22) 3.87
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	(0.92) 1.29	(1.32) **1.47**	(1.08) **1.21**	(0.62) 1.42	(0.92) 1.50	(1.32) **1.48**	(1.08) **1.13**	(0.62) 1.55
	$\alpha = 5\%$	(1.17) **1.49**	(1.67) **1.82**	(1.37) **1.41**	(0.78) 1.89	(1.17) 2.25	(1.67) **1.78**	(1.37) **1.50**	(0.78) 2.02
	$\alpha = 1\%$	(1.77) 2.40	(2.50) **2.35**	(2.07) **1.88**	(1.14) 2.30	(3.77) 4.30	(2.50) **2.61**	(2.07) **2.08**	(1.14) 3.26

Note: This table reports the Expected Shortfall (ES) forecast based on Calvet/Fisher/Thompson model, the numbers in parentheses are the empirical realized ES values. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK), FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound, Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge which is zero investment portfolio. Numbers inside parentheses are empirical ES, and numbers outside parentheses are corresponding ES obtained by forecast. Bold numbers show those cases for which we cannot reject identity of the empirical and simulated ES, i.e. the empirical value falls into the range between the 2.5 to 97.5 percent quantile of the simulated ones.

Table 5.18: Multi-period Expected shortfall forecasts (CC-GARCH model)

		One day horizon				Five days horizon			
		DOW	*NIK*	*EW*	*HG*	*DOW*	*NIK*	*EW*	*HG*
Stocks	$\alpha = 10\%$	(2.21) **2.03**	(1.80) **1.71**	(3.02) **2.85**	(2.60) **2.47**	(2.21) **1.90**	(1.80) **1.95**	(3.02) **2.96**	(2.60) **2.39**
	$\alpha = 5\%$	(2.84) **2.70**	(2.39) **2.22**	(3.89) **4.01**	(3.25) **3.19**	(2.84) **2.81**	(2.39) **2.35**	(3.89) **3.80**	(3.25) 2.81
	$\alpha = 1\%$	(4.36) 3.75	(4.40) **4.51**	(6.48) 5.11	(5.09) 4.30	(4.36) **4.70**	(4.40) **4.45**	(6.48)5.39	(5.09) 4.20
		US	*DM*	*EW*	*HG*	*US*	*DM*	*EW*	*HG*
FXs	$\alpha = 10\%$	(1.84) **1.92**	(1.79) **1.77**	(2.51) **2.65**	(2.44) **2.48**	(1.84) **2.20**	(1.79) **1.91**	(2.51) **2.55**	(2.44) **2.67**
	$\alpha = 5\%$	(2.34) **2.17**	(2.26) **2.03**	(3.08) **3.22**	(2.99) **2.68**	(2.34) **2.02**	(2.26) **2.00**	(3.08) **2.71**	(2.99) 2.35
	$\alpha = 1\%$	(3.57) **3.25**	(3.39) 3.02	(4.34) **4.16**	(4.22) 3.41	(3.57) 3.19	(3.39) **3.01**	(4.34) **3.05**	(4.22) 3.10
		*TB*1	*TB*2	*EW*	*HG*	*TB*1	*TB*2	*EW*	*HG*
Bonds	$\alpha = 10\%$	(0.92) **1.10**	(1.32) **1.35**	(1.08) **1.27**	(0.62) **0.81**	(0.92) 1.51	(1.32) **1.49**	(1.08) **1.40**	(0.62) 1.24
	$\alpha = 5\%$	(1.17) 1.22	(1.67) **1.70**	(1.37) **1.39**	(0.78) 1.38	(1.17) 1.98	(1.67) **1.83**	(1.37) **1.85**	(0.78) 2.33
	$\alpha = 1\%$	(1.77) **1.63**	(2.50) 1.97	(2.07) 1.35	(1.14) 2.28	(1.77) 2.20	(2.50) 2.61	(2.07) 2.99	(1.14) 2.68

Note: This table reports the Expected Shortfall (ES) forecast based on CC-GARCH(1, 1) model, the numbers in parentheses are the empirical realized ES values. Stocks are Dow Jones Composite 65 Average Index (DOW) and NIKKEI 225 Stock Average Index (NIK), FXs are Foreign Exchange rate of U.S. Dollar (US) and German Mark (DM) to British Pound, Bonds are the U.S. 1-Year and 2-Year Treasury Constant Maturity Rate (TB1, TB2 respectively). EW denotes Equal-Weight portfolio, HG denotes Hedge which is zero investment portfolio. Numbers inside parentheses are empirical ES, and numbers outside parentheses are corresponding ES obtained by forecast. Bold numbers show those cases for which we cannot reject identity of the empirical and simulated ES, i.e. the empirical value falls into the range between the 2.5 to 97.5 percent quantile of the simulated ones.

Chapter 6

Conclusion

In this thesis, we have reviewed fractal and multifractal concepts arising from the natural science and considered their implications for financial economics. Since fractals imply long memory, we briefly covered some traditional methods of fractal analysis (various Hurst exponent methods) and well-known long memory models, e.g. fractional Brownian motion, Fractional Integrated Autoregressive Moving Average (ARFIMA) model, Fractional Integrated Autoregressive Conditional Heteroskedasticity (FIGARCH) model, and the long memory stochastic volatility (LMSV) model. To gain a better impression of the interpretation of (multi)fractals in fields ranging from physics to finance, the multifractal model of asset returns (MMAR) of Mandelbrot et al. (1997) has been revisited, together with its scaling estimator. As one variation of the MF model, the Markov-switching multifractal model was proposed by Mandelbrot's former students L. Calvet and A. Fisher, with the intention of overcoming the limitations in practical applicability of MMAR.

The main contribution of this thesis is the development of a bivariate multifractal model as an extension to the univariate Markov-switching multifractal model. Indeed, there are few contributions going beyond univariate multifractal processes, except for the model of Calvet/Fisher/Thompson (2006). We present a relatively parsimonious bivariate MF model

as a simple alternative. Since the scaling estimator for combinatorial multifractal process yields unreliable results (cf. Lux (2004)), estimating the parameters of multivariate multifractal models is a challenging task. Various approaches have been applied such as maximum likelihood, simulation based ML with particle filter algorithm and GMM approaches, which have been implemented in this thesis for the purpose of comparison. Furthermore, a higher dimensional ML model for $N > 2$ assets has been analogously introduced, and the three different approaches (GMM, ML and SML) have also been employed for the model's estimation.

To investigate the performance of different estimation approaches, Monte Carlo studies were carried out, indicating that: (1) there is no disadvantage for GMM compared to the ML estimator, although the latter had been expected to be more efficient; (2) there is no restrictions on the choice of the number of cascade levels with GMM, in contrast to the upper bound of about 5 cascade levels for ML; (3) furthermore, for the computational view, GMM is much faster compared to the very time-consuming ML process, and the particle filter algorithm when larger numbers of particles are employed. As an extension, a higher dimensional MF model with parsimonious design was defined, and three estimation approaches, as used in the bivariate model, have been implemented by taking as example a tri-variate version, which further demonstrates the advantages of GMM when increasing the dimensions of the MF model.

To demonstrate the applicability of our multivariate MF model, two well-known measures used in financial risk management, namely Value-at-Risk (VaR) and Expected Shortfall (ES) have been used to assess the performance of our model. We considered various empirical financial time series, including stock exchange indices, foreign exchange rates and U.S. treasury bond maturity rates, and compared the performance of BMF models (Liu/Lux model, Calvet/Fisher/Thompson model) and the CC-GARCH model. It has been demonstrated that bivariate multifractal models outperform the CC-GARCH model in calculating

the failure rate of VaR forecasts, particularly in the case of foreign exchange rates. For the ES forecast, bivariate MF models also provide far better results, even though, in VaR forecasting, the ES forecast performances of the CC-GARCH model is better in comparison.

With the bivariate MF models examined in this thesis, our studies have provided supportive evidences for this formalism in financial time series analysis, and opened up a broad outlook on the rich variety of volatility models available in the financial risk manager's tool box. Additionally, we suggest that further studies in this direction are likely to offer new insights:

(1) Other variant MF models can be defined, by replacing the increment with some non-Gaussian such as the t-distribution, to improve its ability to capture the stylized facts in financial markets; also different transition probabilities may be introduced into the volatility component updating mechanism.

(2) To contribute the understanding of the correlation in terms of the cascade level, we presented a heuristic method for specifying the number of joint cascade levels, by matching the simulated GPH estimator with the empirical one. More robust specification tests for the number of joint frequencies are worth developing.

(3) As in forthcoming studies on volatility forecasting of the univariate MF model, following the introduction of best linear forecast by Lux (2008) with the aid of the Levinson-Durbin algorithm by Brockwell and Dahlhaus (2004), volatility forecasts based on multivariate (bivariate) MF models can be explored by developing that algorithm with multivariate settings.

(4) Other future areas of work may include implementing a particle filter for multifractal processes with continuous distributions of volatility components; as well as applying alternative efficient algorithms for simulation based inferences, other than the SIR method with particle filter.

Chapter 7

Appendix

7.1 Moment conditions for the Liu/Lux model

Recall the model from Chapter 3. Let $\varepsilon_t^{(\cdot)} = ln(M_t^{(\cdot)})$, and we compute the first log difference:

$$
\begin{aligned}
X_{t,1} &= ln(|r_{1,t}|) - ln(|r_{1,t-1}|) \\
&= \left(\tfrac{1}{2} \sum_{i=1}^{k} \varepsilon_t^{(i)} + \tfrac{1}{2} \sum_{l=k+1}^{n} \varepsilon_t^{(l)} + ln|u_{1,t}| \right) - \left(\tfrac{1}{2} \sum_{i=1}^{k} \varepsilon_{t-1}^{(i)} + \tfrac{1}{2} \sum_{l=k+1}^{n} \varepsilon_{t-1}^{(l)} + ln|u_{1,t-1}| \right) \\
&= \tfrac{1}{2} \sum_{i=1}^{k} \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)} \right) + \tfrac{1}{2} \sum_{l=k+1}^{n} \left(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)} \right) + (ln|u_{2,t}| - ln|u_{1,t-1}|)
\end{aligned}
$$

$$
\begin{aligned}
Y_{t,1} &= ln(|r_{2,t}|) - ln(|r_{2,t-1}|) \\
&= \left(\tfrac{1}{2} \sum_{i=1}^{k} \varepsilon_t^{(i)} + \tfrac{1}{2} \sum_{h=k+1}^{n} \varepsilon_t^{(h)} + ln|u_{2,t}| \right) - \left(\tfrac{1}{2} \sum_{i=1}^{k} \varepsilon_{t-1}^{(i)} + \tfrac{1}{2} \sum_{h=k+1}^{n} \varepsilon_{t-1}^{(h)} + ln|u_{2,t-1}| \right) \\
&= \tfrac{1}{2} \sum_{i=1}^{k} \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)} \right) + \tfrac{1}{2} \sum_{h=k+1}^{n} \left(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)} \right) + (ln|u_{2,t}| - ln|u_{2,t-1}|)
\end{aligned}
$$

7.1.1 Binomial case

$$cov[X_{t,1}, Y_{t,1}]$$

$$= E\left[(X_{t,1} - E[X_{t,1}]) \cdot (Y_{t,1} - E[Y_{t,1}])\right] = E[X_{t,1} \cdot Y_{t,1}]$$

$$= E\left\{ \left[\tfrac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2} \sum_{l=k+1}^{n} (\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|) \right] \cdot \right.$$
$$\left. \left[\tfrac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2} \sum_{h=k+1}^{n} (\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \right] \right\}$$

$$= \tfrac{1}{4} E\left[\left(\sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) \right)^2 \right] + E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|) \cdot (ln|u_{2,t}| - ln|u_{2,t-1}|) \right].$$

We firstly consider $E[(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})^2]$, the only one non-zero contribution is $[ln(m_0) - ln(2 - m_0)]^2$, and it occurs when new draws take place in cascade level i between t and $t-1$, whose probability is $\frac{1}{2}\frac{1}{2^{k-i}}$ by definition. For the second component:

$$E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|) (ln|u_{2,t}| - ln|u_{2,t-1}|) \right]$$

$$= E\left[(ln|u_{1,t}|)(ln|u_{2,t}|) - (ln|u_{1,t}|)(ln|u_{2,t-1}|) - (ln|u_{2,t}|)(ln|u_{1,t-1}|) + (ln|u_{1,t-1}|)(ln|u_{2,t-1}|) \right]$$
$$= 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2.$$

Because, $E[(ln|u_{1,t}|)(ln|u_{2,t-1}|)] = E[ln|u_{1,t}|] \cdot E[ln|u_{2,t-1}|] = E[ln|u_t|]^2$. Note that, $ln|u_{1,t}|$ and $ln|u_{2,t}|$ are the log of absolute values of random variates from bivariate standard Normal distribution; $ln|u_t|$ is the log of absolute value of random variates from standard Normal distribution. Summing up we get:

$$cov[X_{t,1}, Y_{t,1}] = 0.25 \cdot [ln(m_0) - ln(2 - m_0)]^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2. \tag{A1}$$

$$cov[X_{t+1,1}, Y_{t,1}]$$

$$= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + \tfrac{1}{2}\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|)\right]\cdot \left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2}\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|)\right]\right\}$$

$$= \tfrac{1}{4}E\left[\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\cdot\sum_{i=1}^{k}\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right] + E[ln|u_t|]^2 - E[ln|u_{1,t}|\cdot ln|u_{2,t}|].$$

For $(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})$, the non-zero value only occurs in case of two changes of the multiplier from time $t+1$ to time $t-1$, the probability of this occurrence is $(\frac{1}{2}\frac{1}{2^{k-i}})^2$. So, we have the result:

$$\begin{aligned} &cov[X_{t+1,1}, Y_{t,1}] \\ &= 0.25\cdot[2ln(m_0)\cdot ln(2-m_0) - (ln(m_0))^2 - (ln(2-m_0))^2]\cdot\sum_{i=1}^{k}(\tfrac{1}{2}\tfrac{1}{2^{k-i}})^2 \\ &+E[ln|u_t|]^2 - E[ln|u_{1,t}|\cdot ln|u_{2,t}|]. \end{aligned} \tag{A2}$$

Then, we look at the moment condition for one single time series:

$$cov[X_{t+1,1}, X_{t,1}]$$

$$= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + \tfrac{1}{2}\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|)\right]\cdot \left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2}\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|)\right]\right\}$$

$$= \tfrac{1}{4}E\left[\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\cdot\sum_{i=1}^{k}\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right] + \tfrac{1}{4}E\left[\sum_{l=k+1}^{n}\left(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)}\right)\cdot\sum_{l=k+1}^{n}\left(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)}\right)\right]$$

$$+E\left[ln|u_t|\right]^2 - E\left[ln|u_t|^2\right]. \tag{A3}$$

The first component is identical to the one of the case of $cov[X_{t+1,1}, Y_{t,1}]$, and the second component can be derived in the same way. Adding together we arrive at:

$$
\begin{aligned}
&cov[X_{t+1,1}, X_{t,1}] \\
&= 0.25 \cdot \left[2ln(m_0) \cdot ln(2-m_0) - (ln(m_0))^2 - (ln(2-m_0))^2\right] \cdot \sum_{i=1}^{k} (\tfrac{1}{2}\tfrac{1}{2^{k-i}})^2 \\
&+0.25 \cdot \left[2ln(m_0) \cdot ln(2-m_0) - (ln(m_0))^2 - (ln(2-m_0))^2\right] \cdot \sum_{i=k+1}^{n} (\tfrac{1}{2}\tfrac{1}{2^{n-i}})^2 \\
&+E[ln|u_t|]^2 - E[ln|u_t|^2].
\end{aligned}
\tag{A4}
$$

By our assumption of both time series having the same number of cascade levels, the moments for the two individual time series are identical for the same length of time lags.

Then, let's turn to the moments for squared observations:

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= E\left\{ \left[\tfrac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2} \sum_{l=k+1}^{n} (\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|) \right]^2 \cdot \right. \\
&\quad \left. \left[\tfrac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \tfrac{1}{2} \sum_{h=k+1}^{n} (\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \right]^2 \right\}.
\end{aligned}
\tag{A5}
$$

Let's define

$a_1 = \frac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})$; $b_1 = \frac{1}{2} \sum_{l=k+1}^{n} (\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})$; $c_1 = ln|u_{1,t}| - ln|u_{1,t-1}|$;

and

$a_2 = \frac{1}{2} \sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) = a_1$; $b_2 = \frac{1}{2} \sum_{h=k+1}^{n} (\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})$; $c_2 = ln|u_{2,t}| - ln|u_{2,t-1}|$.

Then Eq (A5) can be replaced by the following expression:

$$E\left[\left(a_1^2+b_1^2+c_1^2+2a_1b_1+2a_1c_1+2b_1c_1\right)\left(a_2^2+b_2^2+c_2^2+2a_2b_2+2a_2c_2+2b_2c_2\right)\right]$$

$$\begin{aligned}
&= E\left[a_1^2a_2^2+a_1^2b_2^2+a_1^2c_2^2+2a_1^2a_2b_2+2a_1^2a_2c_2+2a_1^2b_2c_2+\right.\\
&b_1^2a_2^2+b_1^2b_2^2+b_1^2c_2^2+2b_1^2a_2b_2+2b_1^2a_2c_2+2b_1^2b_2c_2+\\
&c_1^2a_2^2+c_1^2b_2^2+c_1^2c_2^2+2c_1^2a_2b_2+2c_1^2a_2c_2+2c_1^2b_2c_2+\\
&2a_1b_1a_2^2+2a_1b_1b_2^2+2a_1b_1c_2^2+4a_1b_1a_2b_2+4a_1b_1a_2c_2+4a_1b_1b_2c_2+\\
&2a_1c_1a_2^2+2a_1c_1b_2^2+2a_1c_1c_2^2+4a_1c_1a_2b_2+4a_1c_1a_2c_2+4a_1c_1b_2c_2\\
&\left.2b_1c_1a_2^2+2b_1c_1b_2^2+2b_1c_1c_2^2+4b_1c_1a_2b_2+4b_1c_1a_2c_2+4b_1c_1b_2c_2\right]
\end{aligned} \tag{A6}$$

By examining each element above, we find that there are many elements which can be skipped due to their values of zero, eg. $E[2a_1^2a_2b_2] = 2E[a_1^2a_2]\cdot E[b_2] = 0$, $E[2a_1^2a_2c_2] = 2E[a_1^2a_2]\cdot E[c_2] = 0$, and so on. We therefore have the intermediate result by carefully checking each one in Eq (A6):

$$E\left[\left(a_1^2+b_1^2+c_1^2+2a_1b_1+2a_1c_1+2b_1c_1\right)\left(a_2^2+b_2^2+c_2^2+2a_2b_2+2a_2c_2+2b_2c_2\right)\right]$$

$$= E\left[a_1^2a_2^2+a_1^2b_2^2+a_1^2c_2^2+b_1^2a_2^2+b_1^2b_2^2+b_1^2c_2^2+c_1^2a_2^2+c_1^2b_2^2+c_1^2c_2^2+4a_1c_1a_2c_2\right] \tag{A7}$$

$$= E\left[a_1^4+b_1^2b_2^2+a_1^2b_2^2+b_1^2a_2^2+(2a_1^2+2b_1^2)c_1^2+c_1^2c_2^2+4a_1^2c_1c_2\right]$$

Therefore, we can study the original expression of Eq. (A6) as:

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^4\right] + \tfrac{1}{16}E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right] \\
&+\tfrac{1}{4}\left\{2E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] + 2E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right]\right\} \cdot E\left[\left(ln|u_{1,t}| - ln|u_{1,t-1}|\right)^2\right] \\
&+E\left[\left(ln|u_{1,t}| - ln|u_{1,t-1}|\right)^2\left(ln|u_{2,t}| - ln|u_{2,t-1}|\right)^2\right] \\
&+E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] \cdot E\left[\left(ln|u_{1,t}| - ln|u_{1,t-1}|\right)\left(ln|u_{2,t}| - ln|u_{2,t-1}|\right)\right].
\end{aligned}
\tag{A8}
$$

We begin by studying the first component $E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^4\right]$. The non-zero contribution comes from:

(1) $\left[ln\left(m_0\right) - ln\left(2 - m_0\right)\right]^4$ when $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$, and the possibilities for this occurrence is equal to $\sum_{i=1}^{k} \frac{1}{2}\frac{1}{2^{k-i}}$,

(2) $\left[ln\left(m_0\right) - ln\left(2 - m_0\right)\right]^2\left[ln\left(m_0\right) - ln\left(2 - m_0\right)\right]^2$ when $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$ and $\varepsilon_t^{(j)} \neq \varepsilon_{t-1}^{(j)}$, (for $i \neq j$), and the possibilities of its occurrence is $\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right)$.

therefore, the corresponding value of the first component is

$$\left[ln\left(m_0\right) - ln\left(2 - m_0\right)\right]^4\left[\sum_{i=1}^{k}\frac{1}{2}\frac{1}{2^{k-i}} + \sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right)\right].$$

Then we have a look at the second component in Eq. (A8), that is

$$E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)}-\varepsilon_{t-1}^{(l)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)}-\varepsilon_{t-1}^{(h)})\right)^2\right].$$

Because of the independence between two multifractal processes after k joint cascade levels, let us examine $E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)}-\varepsilon_{t-1}^{(l)})\right)^2\right]$ firstly, whose non-zero contribution comes from $[ln(m_0)-ln(2-m_0)]^2$ with probability of $\sum_{i=k+1}^{n}\frac{1}{2}\frac{1}{2^{n-i}}$, and the same way for $E\left[\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)}-\varepsilon_{t-1}^{(h)})\right)^2\right]$. Then we have the result:

$$[ln(m_0)-ln(2-m_0)]^4\cdot\sum_{i=k+1}^{n}\left(\frac{1}{2}\frac{1}{2^{n-i}}\sum_{j=k+1}^{n}\frac{1}{2}\frac{1}{2^{n-j}}\right).$$

For the third component in Eq. (A8), we have

$$E\left[\left(\sum_{i=k}^{n}(\varepsilon_t^{(i)}-\varepsilon_{t-1}^{(i)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)}-\varepsilon_{t-1}^{(h)})\right)^2\right]=E\left[\left(\sum_{i=k}^{n}(\varepsilon_t^{(i)}-\varepsilon_{t-1}^{(i)})\right)^2\right]\cdot E\left[\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)}-\varepsilon_{t-1}^{(h)})\right)^2\right],$$

and it equals to

$$[ln(m_0)-ln(2-m_0)]^4\cdot\sum_{i=k}^{n}\left(\frac{1}{2}\frac{1}{2^{n-i}}\sum_{j=k+1}^{n}\frac{1}{2}\frac{1}{2^{n-j}}\right).$$

Since the forth component has the same structure like the third one, we directly move to the fifth component, which is

$$\left\{2E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)}-\varepsilon_{t-1}^{(i)})\right)^2\right]+2E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)}-\varepsilon_{t-1}^{(l)})\right)^2\right]\right\}\cdot E\left[(ln|u_{1,t}|-ln|u_{1,t-1}|)^2\right].$$

We know $E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|)^2\right] = 2E[ln|u_t|^2] - 2E[ln|u_t|]^2$, and by combining with the calculations of the previous two components, it is not difficult to find the result for it, that is:

$$[ln(m_0) - ln(2 - m_0)]^2 \cdot \left(\sum_{i=1}^{k} \frac{1}{2}\frac{1}{2^{k-i}} + \sum_{i=k+1}^{n} \frac{1}{2}\frac{1}{2^{n-i}}\right) \cdot \left(E[ln|u_t|^2] - E[ln|u_t|]^2\right).$$

We also have the solution for the sixth component:

$$E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right]$$

$$= 2E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 8E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot ln|u_{2,t}|]^2 + 2E[(ln|u_t|)^2]^2.$$

For the last component:

$$E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] \cdot E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|)(ln|u_{2,t}| - ln|u_{2,t-1}|)\right],$$

by recalling the calculations of the previous moments of Eq. (A1) and A(2), we have its non-zero value which is

$$[ln(m_0) - ln(2 - m_0)]^2 \cdot \sum_{i=1}^{k} \frac{1}{2}\frac{1}{2^{k-i}} \left(2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2\right).$$

Therefore, we sum up the solutions for each component in Eq. (A8), and arrive to the final result:

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16}\left[ln\left(m_0\right) - ln\left(2-m_0\right)\right]^4 \cdot \left[\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} + \sum_{i=1}^{k}\left(\tfrac{1}{2^{k-i}} \sum_{j=1, j\neq i}^{k} \tfrac{1}{2^{k-j}}\right)\right] \\
&+ \tfrac{1}{16}\left[ln\left(m_0\right) - ln\left(2-m_0\right)\right]^4 \cdot \sum_{i=k+1}^{n}\left(\tfrac{1}{2}\tfrac{1}{2^{n-i}} \sum_{j=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-j}}\right) \\
&+ 2 \times \tfrac{1}{16}\left[ln\left(m_0\right) - ln\left(2-m_0\right)\right]^4 \cdot \sum_{i=1}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-i}} \sum_{j=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-j}}\right) \\
&+ \left[ln\left(m_0\right) - ln\left(2-m_0\right)\right]^2 \cdot \left(\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} + \sum_{i=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-i}}\right) \cdot \left(E[ln|u_t|^2] - E[ln|u_t|]^2\right) \\
&+ 2E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 8E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] \\
&+ 4E[ln|u_{1,t}| \cdot ln|u_{2,t}|]^2 + 2E[(ln|u_t|)^2]^2 \\
&+ \left[ln\left(m_0\right) - ln\left(2-m_0\right)\right]^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \left(2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2\right).
\end{aligned}
\tag{A9}
$$

$$E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$$

$$= E\left\{ \left[\frac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + \frac{1}{2}\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|) \right]^2 \cdot \right.$$
$$\left. \left[\frac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + \frac{1}{2}\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \right]^2 \right\}. \tag{A10}$$

Again, let's define

$a_1 = \frac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})$; $b_1 = \frac{1}{2}\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)})$; $c_1 = (ln|u_{1,t+1}| - ln|u_{1,t}|)$;

and

$a_2 = \frac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})$; $b_2 = \frac{1}{2}\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})$; $c_2 = (ln|u_{2,t}| - ln|u_{2,t-1}|)$.

Then Eq. (A10) reaches the following expression:

$$E\left[\left(a_1^2 + b_1^2 + c_1^2 + 2a_1b_1 + 2a_1c_1 + 2b_1c_1\right)\left(a_2^2 + b_2^2 + c_2^2 + 2a_2b_2 + 2a_2c_2 + 2b_2c_2\right)\right]$$

$$= E\left[a_1^2a_2^2 + a_1^2b_2^2 + a_1^2c_2^2 + 2a_1^2a_2b_2 + 2a_1^2a_2c_2 + 2a_1^2b_2c_2 + b_1^2a_2^2 + b_1^2b_2^2 + b_1^2c_2^2 + \right.$$
$$2b_1^2a_2b_2 + 2b_1^2a_2c_2 + 2b_1^2b_2c_2 + c_1^2a_2^2 + c_1^2b_2^2 + c_1^2c_2^2 + 2c_1^2a_2b_2 + 2c_1^2a_2c_2 + 2c_1^2b_2c_2 +$$
$$2a_1b_1a_2^2 + 2a_1b_1b_2^2 + 2a_1b_1c_2^2 + 4a_1b_1a_2b_2 + 4a_1b_1a_2c_2 + 4a_1b_1b_2c_2 +$$
$$2a_1c_1a_2^2 + 2a_1c_1b_2^2 + 2a_1c_1c_2^2 + 4a_1c_1a_2b_2 + 4a_1c_1a_2c_2 + 4a_1c_1b_2c_2 +$$
$$\left. 2b_1c_1a_2^2 + 2b_1c_1b_2^2 + 2b_1c_1c_2^2 + 4b_1c_1a_2b_2 + 4b_1c_1a_2c_2 + 4b_1c_1b_2c_2\right]$$

$$= E\left[a_1^2a_2^2 + a_1^2b_2^2 + a_1^2c_2^2 + b_1^2a_2^2 + b_1^2b_2^2 + b_1^2c_2^2 + c_1^2a_2^2 + c_1^2b_2^2 + c_1^2c_2^2 + 4a_1c_1a_2c_2\right]$$
$$= E\left[a_1^2a_2^2 + b_1^2b_2^2 + a_1^2b_2^2 + b_1^2a_2^2 + (2a_1^2 + 2b_1^2)c_1^2 + c_1^2c_2^2 + 4a_1c_1a_2c_2\right] \tag{A11}$$

Therefore, we can study the original expression of Eq. (A11) as below:

$$
\begin{aligned}
&E[X_{t+1,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] \\
&+ \tfrac{1}{16} E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)})\right)^2 \left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+ \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+ \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 \left(\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)})\right)^2\right] \\
&+ \tfrac{1}{4}\left\{2E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2\right] + 2E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right]\right\} \cdot \left(2E[ln|u_t|^2] - 2E[ln|u_t|]^2\right) \\
&+ E\left[\left(ln|u_{1,t+1}| - ln|u_{1,t}|\right)^2 \left(ln|u_{2,t}| - ln|u_{2,t-1}|\right)^2\right] \\
&+ E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)\right](E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]).
\end{aligned}
\tag{A12}
$$

After checking each component in Eq. (A12), we see the only unfamiliar one is the first term:

$E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \cdot \left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right]$, and there are three different forms to be considered:

(1) $\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2$, which have non-zero value only if $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$. and this possibility is $(\frac{1}{2}\frac{1}{2^{k-i}})^2$, combining with the non-zero expectation value, we have $\left(\sum_{i=1}^{k}(\frac{1}{2}\frac{1}{2^{k-i}})^2\right) \cdot \left[ln\left(m_0\right) - ln\left(2 - m_0\right)\right]^4$.

(2) $\left(\varepsilon_{t+1}^{(j)}-\varepsilon_{t}^{(j)}\right)^{2}\left(\varepsilon_{t}^{(i)}-\varepsilon_{t-1}^{(i)}\right)^{2}$, which are non-zero for $i\neq j$, $\varepsilon_{t+1}^{(j)}\neq\varepsilon_{t}^{(j)}$ and $\varepsilon_{t}^{(i)}\neq\varepsilon_{t-1}^{(i)}$, the probability of its occurrence is $\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right)$.

Putting together these two possible forms we get
$[ln\,(m_0)-ln\,(2-m_0)]^4\cdot\left(\sum_{i=1}^{k}\frac{1}{2}\frac{1}{2^{k-i}}\sum_{j=1}^{k}\frac{1}{2}\frac{1}{2^{k-j}}\right)$.

(3) Form $\left(\varepsilon_{t+1}^{(j)}-\varepsilon_{t}^{(j)}\right)\left(\varepsilon_{t+1}^{(i)}-\varepsilon_{t}^{(i)}\right)\left(\varepsilon_{t}^{(j)}-\varepsilon_{t-1}^{(j)}\right)\left(\varepsilon_{t}^{(i)}-\varepsilon_{t-1}^{(i)}\right)$, which for $i\neq j$ and $\varepsilon_{t+1}^{(n)}\neq\varepsilon_{t}^{(n)}\neq\varepsilon_{t-1}^{(n)}$, $n=i,j$ are non-zero, and which implies
$2\left\{\sum_{i=1}^{k}\left((\frac{1}{2^{k-i}})^2\sum_{j=1,j\neq i}^{k}(\frac{1}{2^{k-j}})^2\right)\right\}\cdot[ln\,(m_0)-ln\,(2-m_0)]^4$.

Then we have the solution for the first component in the above moment condition:

$$E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)}-\varepsilon_{t}^{(i)})\right)^{2}\cdot\left(\sum_{i=1}^{k}(\varepsilon_{t}^{(i)}-\varepsilon_{t-1}^{(i)})\right)^{2}\right]$$
$$=[ln\,(m_0)-ln\,(2-m_0)]^4\left[\sum_{i=1}^{k}\frac{1}{2}\frac{1}{2^{k-i}}\sum_{j=1}^{k}\frac{1}{2}\frac{1}{2^{k-j}}+2\sum_{i=1}^{k}(\frac{1}{2}\frac{1}{2^{k-i}})^2\sum_{j=1,j\neq i}^{k}(\frac{1}{2}\frac{1}{2^{k-j}})^2\right]$$

The other components can be solved by recalling previous calculations. All in all, we finally arrive at:

$$
\begin{aligned}
&E[X_{t+1,1}^2 \cdot Y_{t,1}^2] \\
&= [ln(m_0) - ln(2-m_0)]^4 \cdot \tfrac{1}{16}\left[\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \sum_{j=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-j}} + 2\sum_{i=1}^{k}(\tfrac{1}{2}\tfrac{1}{2^{k-i}})^2 \sum_{j=1, j\neq i}^{k}(\tfrac{1}{2}\tfrac{1}{2^{k-j}})^2\right] \\
&+ [ln(m_0) - ln(2-m_0)]^4 \cdot \tfrac{1}{16} \sum_{i=k+1}^{n}\left(\tfrac{1}{2}\tfrac{1}{2^{n-i}} \sum_{j=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-j}}\right) \\
&+ \tfrac{1}{8}[ln(m_0) - ln(2-m_0)]^4 \sum_{i=1}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-i}} \sum_{j=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-j}}\right) \\
&+ \left(E[ln|u_t|^2] - E[ln|u_t|]^2\right) \cdot [ln(m_0) - ln(2-m_0)]^2 \cdot \left(\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} + \sum_{i=k+1}^{n} \tfrac{1}{2}\tfrac{1}{2^{n-i}}\right) \\
&+ E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 4E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot ln|u_{2,t}|]E[ln|u_t|]^2 \\
&+ 3E[ln|u_t|^2]^2 - 4E[ln|u_t|^2]E[ln|u_t|]^2 \\
&+ \left[2ln(m_0) \cdot ln(2-m_0) - (ln(m_0))^2 - (ln(2-m_0))^2\right] \cdot \sum_{i=1}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-i}}\right)^2 (E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]).
\end{aligned}
\tag{A13}
$$

We also consider the moments for individual transformed observations of $E[X_{t+1,1}^2 \cdot X_{t,1}^2]$ and $E[Y_{t+1,1}^2 \cdot Y_{t,1}^2]$. The calculations have been done by Lux (2008) when applying GMM for the univariate MF model. Note that they have identical solutions since both multifractal processes share a common volatility component, and we omit it here for brevity.

7.1.2 Lognormal case

$$cov[X_{t,1}, Y_{t,1}] = E[(X_{t,1} - E[X_{t,1}]) \cdot (Y_{t,1} - E[Y_{t,1}] = E[X_{t,1} \cdot Y_{t,1}]$$

$$= \tfrac{1}{4} E\left[\left(\sum_{i=1}^{k} (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[u_t]^2$$
$$= 0.5\sigma_\varepsilon^2 \sum_{i=1}^{k} \tfrac{1}{2^{k-i}} + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2. \tag{B1}$$

Because the non-zero outcomes occur when $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$, which implies:

$$(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})^2 = 2(E[(\varepsilon_t^{(i)})^2] - E[\varepsilon_t^{(i)}]^2) = 2\sigma_\varepsilon^2$$

$$cov[X_{t+1,1}, Y_{t,1}] = \tfrac{1}{4} E\left[\sum_{i=1}^{k} \left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right) \cdot \sum_{i=1}^{k} \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right] + E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]$$

$$= -0.25\sigma_\varepsilon^2 \sum_{i=1}^{k} \left(\tfrac{1}{2^{k-i}}\right)^2 + E[u_t]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]. \tag{B2}$$

Because the non-zero outcomes occur when $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$, which implies:

$$(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) \cdot (\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) = E[\varepsilon_t^{(i)}]^2 - E[(\varepsilon_t^{(i)})^2] = -\sigma_\varepsilon^2$$

$$cov[X_{t+1,1}, X_{t,1}]$$

$$= \tfrac{1}{4} E\left[\sum_{i=1}^{k} \left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right) \cdot \sum_{i=1}^{k} \left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\right] + \tfrac{1}{4} E\left[\sum_{l=k+1}^{n} \left(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)}\right) \cdot \sum_{l=k+1}^{n} \left(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)}\right)\right]$$

$$+E\left[ln|u_t|\right]^2 - E\left[ln|u_t|^2\right]$$

$$= -0.25\sigma_\varepsilon^2 \left[\sum_{i=1}^{k} \left(\tfrac{1}{2^{k-i}}\right)^2 + \sum_{i=k+1}^{n} \left(\tfrac{1}{2^{n-i}}\right)^2\right] + E\left[ln|u_t|\right]^2 - E\left[ln|u_t|^2\right]. \tag{B3}$$

For the moments of squared observations, we recall the semi-final results of Eq. (A6) and Eq. (A11) for $E[X_{t,1}^2 \cdot Y_{t,1}^2]$ and $E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$.

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^4\right] + \tfrac{1}{16}E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right] \\
&+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right] \\
&+\tfrac{1}{4}\left\{2E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] + 2E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right]\right\} \cdot \left(2E[ln|u_t|^2] - 2E[ln|u_t|]^2\right) \\
&+2E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 8E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] \\
&+4E[ln|u_{1,t}| \cdot (ln|u_{2,t}|]^2 + 2E[(ln|u_t|)^2]^2 \\
&+E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right](2E[ln|u_{1,t}| \cdot ln|u_{2,t}|] - 2E[ln|u_t|]^2).
\end{aligned}
\tag{B4}
$$

For the first term $E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^4\right]$, let's begin with $E\left[\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^4\right]$, we know that the non-zero contributions come from:

(1) $E\left[\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^4\right] = 2E[\varepsilon_t^{(i)}]^4 + 6E[(\varepsilon_t^{(i)})^2]^2 - 8E[(\varepsilon_t^{(i)})^3]E[\varepsilon_t^{(i)}] = 12\sigma_\varepsilon^4$, and this occurs with probability $2^{\frac{1}{k-i}}$. Then we have the solution:

$$E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^4\right] = 12\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}\frac{1}{2^{k-i}}.$$

(2) $E\left[\left(\varepsilon_t^{(i)}-\varepsilon_{t-1}^{(i)}\right)^2\left(\varepsilon_t^{(j)}-\varepsilon_{t-1}^{(j)}\right)^2\right]=4\sigma_\varepsilon^4$, for $\varepsilon_t^{(i)}\neq\varepsilon_{t-1}^{(i)}$ and $\varepsilon_t^{(j)}\neq\varepsilon_{t-1}^{(j)}$, $(i\neq j)$, with the possibilities of its occurrence is $\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right)$. Therefore,

$$E\left[\left(\varepsilon_t^{(i)}-\varepsilon_{t-1}^{(i)}\right)^2\left(\varepsilon_t^{(j)}-\varepsilon_{t-1}^{(j)}\right)^2\right]=4\sigma_\varepsilon^4\cdot\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right).$$

$$\begin{aligned}
&E[X_{t,1}^2\cdot Y_{t,1}^2]\\
&=0.75\sigma_\varepsilon^4\sum_{i=1}^{k}\tfrac{1}{2^{k-i}}+0.25\sigma_\varepsilon^4\sum_{i=1}^{k}\left(\tfrac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\tfrac{1}{2^{k-j}}\right)+0.25\sigma_\varepsilon^4\sum_{l=k+1}^{n}\tfrac{1}{2^{n-l}}\sum_{h=k+1}^{n}\tfrac{1}{2^{n-h}}\\
&+0.25\sigma_\varepsilon^4\sum_{i=1}^{k}\tfrac{1}{2^{k-i}}\sum_{h=k+1}^{n}\tfrac{1}{2^{n-h}}+0.25\sigma_\varepsilon^4\sum_{i=1}^{k}\tfrac{1}{2^{k-i}}\sum_{l=k+1}^{n}\tfrac{1}{2^{n-l}}\\
&+2\sigma_\varepsilon^2\left(E[ln|u_t|^2]-E[ln|u_t|]^2\right)\cdot\left(\sum_{i=1}^{k}\tfrac{1}{2^{k-i}}+\sum_{l=k+1}^{n}\tfrac{1}{2^{n-l}}\right)\\
&+2E[(ln|u_{1,t}|)^2\cdot(ln|u_{2,t}|)^2]-8E[(ln|u_{1,t}|)^2\cdot ln|u_{2,t}|]\cdot E[ln|u_t|]\\
&+4E[ln|u_{1,t}|\cdot(ln|u_{2,t}|]^2+2E[(ln|u_t|)^2]^2\\
&+2\sigma_\varepsilon^2\cdot\sum_{i=1}^{k}\tfrac{1}{2^{k-i}}(2E[ln|u_{1,t}|\cdot ln|u_{2,t}|]-2E[ln|u_t|]^2).
\end{aligned}\tag{B5}$$

With the aid of semi-final results in Eq. (A11), $E[X_{t+1,1}^2\cdot Y_{t,1}^2]$ can be computed as follows:

$$E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$$

$$= \tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \cdot \left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right]$$
$$+\tfrac{1}{16}E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)})\right)^2 \left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right]$$
$$+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \left(\sum_{h=k+1}^{n}(\varepsilon_t^{(h)} - \varepsilon_{t-1}^{(h)})\right)^2\right]$$
$$+\tfrac{1}{16}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 \left(\sum_{l=k+1}^{n}(\varepsilon_{t+1}^{(l)} - \varepsilon_t^{(l)})\right)^2\right]$$
$$+\tfrac{1}{4}\left\{2E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right] + 2E\left[\left(\sum_{l=k+1}^{n}(\varepsilon_t^{(l)} - \varepsilon_{t-1}^{(l)})\right)^2\right]\right\} \cdot \left(2E[ln|u_t|^2] - 2E[ln|u_t|]^2\right)$$

$$+E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 4E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot (ln|u_{2,t}|]E[ln|u_t|]^2$$

$$+3E[ln|u_t|^2]^2 - 4E[ln|u_t|^2]E[ln|u_t|]^2$$

$$+E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)\right](E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]).$$

For the first term $E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \cdot \left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right]$, there are three different possible forms:

(1) $\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2$, has non-zero value only if $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$. then $E\left[\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2\right] = E[\varepsilon_t^4] + 3E[\varepsilon_t^2]^2 - 4E[\varepsilon_t^3]E[\varepsilon_t] = 6\sigma_\varepsilon^4$. ($E[\varepsilon_t^3] = 3\lambda\sigma_\varepsilon^2 + \lambda^3$ and $E[\varepsilon_t^4] = 3\sigma_\varepsilon^4 + 6\lambda^2\sigma_\varepsilon^2 + \lambda^4$,) and the probability of this occurrence is $(\frac{1}{2^{k-i}})^2$. Putting together we get $\left[\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\right)^2\right] \cdot 6\sigma_\varepsilon^4$.

(2) $\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2$, does not equal zero for $i \neq j$, $\varepsilon_{t+1}^{(j)} \neq \varepsilon_t^{(j)}$ and $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$. since $E\left[\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2\right] = 4E[(\varepsilon_t^{(i)})^2]^2 - 8E[(\varepsilon_t^{(i)})^2]E[\varepsilon_t^{(i)}]^2 + 4E[\varepsilon_t^{(i)}]^4 = 4\sigma_\varepsilon^4$, together with the probability, this overall contribution yields: $\left[\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\frac{1}{2^{k-j}}\right)\right]\cdot 4\sigma_\varepsilon^4$.

(3) $\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\left(\varepsilon_t^{(j)} - \varepsilon_{t-1}^{(j)}\right)\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)$, which for $i \neq j$ and $\varepsilon_{t+1}^{(n)} \neq \varepsilon_t^{(n)} \neq \varepsilon_{t-1}^{(n)}$, $n = i, j$ are non-zero, since $\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\left(\varepsilon_t^{(j)} - \varepsilon_{t-1}^{(j)}\right)\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right) = 4\sigma_\varepsilon^4$, we obtain a contribution $2\left[\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\right)^2\sum_{j=1,j\neq i}^{k}\left(\frac{1}{2^{k-j}}\right)^2\right]\cdot\sigma_\varepsilon^4$.

Combining those three cases, we have the result:

$$E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \cdot \left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2\right]$$

$$= 6\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}(\tfrac{1}{2^{k-i}})^2 + 4\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}\tfrac{1}{2^{k-i}}\sum_{j=1,j\neq i}^{k}\tfrac{1}{2^{k-j}} + 2\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}(\tfrac{1}{2^{k-i}})^2\sum_{j=1,j\neq i}^{k}(\tfrac{1}{2^{k-j}})^2.$$

Taking all components together, we arrive at:

$$
\begin{aligned}
&E[X_{t+1,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16}\left[6\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}(\tfrac{1}{2^{k-i}})^2 + 4\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}\tfrac{1}{2^{k-i}} \sum_{j=1,j\neq i}^{k}\tfrac{1}{2^{k-j}} + 2\sigma_\varepsilon^4 \cdot \sum_{i=1}^{k}(\tfrac{1}{2^{k-i}})^2 \sum_{j=1,j\neq i}^{k}(\tfrac{1}{2^{k-j}})^2\right] \\
&+0.25\sigma_\varepsilon^4 \sum_{l=k+1}^{n}\tfrac{1}{2^{n-l}} \sum_{h=k+1}^{n}\tfrac{1}{2^{n-h}} \\
&+0.25\sigma_\varepsilon^4 \sum_{i=1}^{k}\tfrac{1}{2^{k-i}} \sum_{h=k+1}^{n}\tfrac{1}{2^{n-h}} + 0.25\sigma_\varepsilon^4 \sum_{i=1}^{k}\tfrac{1}{2^{k-i}} \sum_{l=k+1}^{n}\tfrac{1}{2^{n-l}} \\
&+2\sigma_\varepsilon^2 \cdot \left(E[ln|u_t|^2] - E[ln|u_t|]^2\right) \cdot \left(\sum_{i=1}^{k}\tfrac{1}{2^{k-i}} + \sum_{i=k+1}^{n}\tfrac{1}{2^{n-i}}\right) \\
&+E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 4E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot (ln|u_{2,t}|]E[ln|u_t|]^2 \\
&+3E[ln|u_t|^2]^2 - 4E[ln|u_t|^2]E[ln|u_t|]^2 - \sigma_\varepsilon^2 \sum_{i=1}^{k}\left(\tfrac{1}{2^{k-i}}\right)^2 (E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|]).
\end{aligned}
\tag{B6}
$$

For the moments of individual transformed observations of $E[X_{t+1,1}^2 \cdot X_{t,1}^2]$ and $E[Y_{t+1,1}^2 \cdot Y_{t,1}^2]$, we refer to Lux (2008) when applying GMM for univariate Lognormal model.

7.2 Moment conditions for the Calvet/Fisher/Thompson model

Let $\varepsilon_t^{(\cdot)} = ln(M_{1,t}^{(\cdot)})$, and $\xi_t^{(\cdot)} = ln(M_{2,t}^{(\cdot)})$, and we compute the first log difference:

$$\begin{aligned} X_{t,1} &= ln(|r_{1,t}|) - ln(|r_{1,t-1}|) = \left(\frac{1}{2}\sum_{i=1}^{k}\varepsilon_t^{(i)} + ln|u_{1,t}|\right) - \left(\frac{1}{2}\sum_{i=1}^{k}\varepsilon_{t-1}^{(i)} + ln|u_{1,t-1}|\right) \\ &= \frac{1}{2}\sum_{i=1}^{k}\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right) + (ln|u_{1,t}| - ln|u_{1,t-1}|) \end{aligned} \tag{C1}$$

$$\begin{aligned} Y_{t,1} &= ln(|r_{2,t}|) - ln(|r_{2,t-1}|) = \left(\frac{1}{2}\sum_{i=1}^{k}\xi_t^{(i)} + ln|u_{2,t}|\right) - \left(\frac{1}{2}\sum_{i=1}^{k}\xi_{t-1}^{(i)} + ln|u_{2,t-1}|\right) \\ &= \frac{1}{2}\sum_{i=1}^{k}\left(\xi_t^{(i)} - \xi_{t-1}^{(i)}\right) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \end{aligned} \tag{C2}$$

7.2.1 Binomial case

M_1 is drawn from $\{m_1, 2-m_1\}$, M_2 is drawn from $\{m_2, 2-m_2\}$, and $m_1 \in (0,2)$, $m_2 \in (0,2)$.

$$\begin{aligned} &E[X_{t,1}, Y_{t,1}] \\ &= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|)\right]\left[\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|)\right]\right\} \\ &= \tfrac{1}{4}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)\right] - 2E[ln|u_t|]^2 + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|]. \end{aligned} \tag{C3}$$

We firstly consider $E[(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})(\xi_t^{(i)} - \xi_{t-1}^{(i)})]$, the only one non-zero contribution is $[ln(m_1) - ln(2-m_1)][ln(m_2) - ln(2-m_2)]$, and it occurs when new draws take place in cascade level i between t and $t-1$, whose probability by definition is $\frac{1}{2}\frac{1}{2^{k-i}}$. Summing up

we get:

$$
\begin{aligned}
&E[X_{t,1}, Y_{t,1}] \\
&= 0.25 \cdot [ln(m_1) - ln(2-m_1)][ln(m_2) - ln(2-m_2)] \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \\
&-2E[ln|u_t|]^2 + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|].
\end{aligned}
\tag{C4}
$$

$$
\begin{aligned}
&E[X_{t+1,1}, Y_{t,1}] \\
&= E\left\{ \left[\tfrac{1}{2} \sum_{i=1}^{k} (\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|) \right] \left[\tfrac{1}{2} \sum_{i=1}^{k} (\xi_t^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \right] \right\} \\
&= \tfrac{1}{4} E\left[\sum_{i=1}^{k} \left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)} \right) \cdot \sum_{i=1}^{k} \left(\xi_t^{(i)} - \xi_{t-1}^{(i)} \right) \right] + E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|].
\end{aligned}
\tag{C5}
$$

For $(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})(\xi_t^{(i)} - \xi_{t-1}^{(i)})$, the non-zero value requires $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)}$ and $\xi_t^{(i)} \neq \xi_{t-1}^{(i)}$), which only occur in case of two changes of the multiplier from time $t+1$ to time $t-1$, the probability of this occurrence is $(\frac{1}{2}\frac{1}{2^{k-i}})^2$. So, we have the result:

$$
\begin{aligned}
&E[X_{t+1,1}, Y_{t,1}] \\
&= -0.25 \cdot [ln(m_1) - ln(2-m_1)][ln(m_2) - ln(2-m_2)] \cdot \sum_{i=1}^{k} (\tfrac{1}{2}\tfrac{1}{2^{k-i}})^2 \\
&+E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|].
\end{aligned}
\tag{C6}
$$

We also need the moment condition for one single time series:

$$
\begin{aligned}
&E[X_{t+1,1}, X_{t,1}] \\
&= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_{t}^{(i)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|)\right]\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|)\right]\right\} \\
&= \tfrac{1}{4}E\left[\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_{t}^{(i)}\right)\cdot\sum_{i=1}^{k}\left(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right] + E\left[ln|u_t|\right]^2 - E\left[ln|u_t|^2\right] \\
&= 0.25\cdot\left[2ln(m_1)\cdot ln(2-m_1) - (ln(m_1))^2 - (ln(2-m_1))^2\right]\cdot\sum_{i=1}^{k}(\tfrac{1}{2}\tfrac{1}{2^{k-i}})^2 \\
&+E\left[ln|u_t|\right]^2 - E\left[ln|u_t|^2\right].
\end{aligned}
\tag{C7}
$$

Then we turn to moment conditions for squared transformed observations.

$$
\begin{aligned}
&E[X_{t,1}^2\cdot Y_{t,1}^2] \\
&= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|)\right]^2\cdot\left[\tfrac{1}{2}\sum_{i=1}^{k}(\xi_{t}^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|)\right]^2\right\}.
\end{aligned}
\tag{C8}
$$

By defining $a_1 = \frac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)})$; $\quad b_1 = (ln|u_{1,t}| - ln|u_{1,t-1}|)$; and

$a_2 = \frac{1}{2}\sum_{i=1}^{k}(\xi_{t}^{(i)} - \xi_{t-1}^{(i)})$; $\quad b_2 = (ln|u_{2,t}| - ln|u_{2,t-1}|)$;

Eq. (C8) is transformed to the following expression:

$$
\begin{aligned}
&E\left[\left(a_1^2 + b_1^2 + 2a_1b_1\right)\left(a_2^2 + b_2^2 + 2a_2b_2\right)\right] \\
&= E\left[a_1^2a_2^2 + a_1^2b_2^2 + 2a_1^2a_2b_2 + b_1^2a_2^2 + b_1^2b_2^2 + 2b_1^2a_2b_2 + 2a_1b_1a_2^2 + 2a_1b_1b_2^2 + 4a_1b_1a_2b_2\right] \\
&= E\left[a_1^2a_2^2 + a_1^2b_2^2 + b_1^2a_2^2 + b_1^2b_2^2 + 4a_1b_1a_2b_2\right].
\end{aligned}
\tag{C9}
$$

Therefore, we have

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 \left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2 (ln|u_{1,t}| - ln|u_{1,t-1}|)^2\right] \\
&+E\left[\left(\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)(ln|u_{1,t}| - ln|u_{1,t-1}|)\left(\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)(ln|u_{2,t}| - ln|u_{2,t-1}|)\right] \\
&+E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right].
\end{aligned}
\tag{C10}
$$

We examine each component in Eq. (C10) based on the experiences of previous calculations, and get the final solution for $E[X_{t,1}^2 \cdot Y_{t,1}^2]$ as below:

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16}\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right]^2 \cdot \left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right]^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \\
&+\tfrac{1}{4}\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right]^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] \\
&+\tfrac{1}{4}\left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right]^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] \\
&+\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right] \cdot \left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right] \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot \left(2E[ln|u_{1,t}|ln|u_{2,t}| - 2E[ln|u_t|]^2]\right) \\
&+2E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 8E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] \\
&+4E[ln|u_{1,t}| \cdot (ln|u_{2,t}|]^2 + 2E[(ln|u_t|)^2]^2.
\end{aligned}
\tag{C11}
$$

$$E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$$

$$= E\left\{ \left[\frac{1}{2} \sum_{i=1}^{k} (\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|) \right]^2 \cdot \left[\frac{1}{2} \sum_{i=1}^{k} (\xi_t^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|) \right]^2 \right\}. \tag{C12}$$

By defining $a_1 = \frac{1}{2} \sum_{i=1}^{k} (\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})$; $\quad b_1 = (ln|u_{1,t+1}| - ln|u_{1,t}|)$; and

$a_2 = \frac{1}{2} \sum_{i=1}^{k} (\xi_{t+1}^{(i)} - \xi_t^{(i)})$; $\quad b_2 = (ln|u_{2,t+1}| - ln|u_{2,t}|)$;

Eq. (C12) is then transformed to the following expression:

$$E\left[\left(a_1^2 + b_1^2 + 2a_1b_1\right) \left(a_2^2 + b_2^2 + 2a_2b_2\right) \right]$$

$$= E\left[a_1^2a_2^2 + a_1^2b_2^2 + 2a_1^2a_2b_2 + b_1^2a_2^2 + b_1^2b_2^2 + 2b_1^2a_2b_2 + 2a_1b_1a_2^2 + 2a_1b_1b_2^2 + 4a_1b_1a_2b_2 \right]$$

$$= E\left[a_1^2a_2^2 + a_1^2b_2^2 + b_1^2a_2^2 + b_1^2b_2^2 + 4a_1b_1a_2b_2 \right]. \tag{C13}$$

Eq. (C13) excludes the zero value of the moments and leaves only the non-zero contribution ones. We rewrite its original form and it is not difficult for us to have the solution of $E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$, that is

$$
\begin{aligned}
&E[X_{t+1,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2 (ln|u_{1,t+1}| - ln|u_{1,t}|)^2\right] \\
&+E\left[\left(\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)(ln|u_{1,t+1}| - ln|u_{1,t}|)\left(\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)(ln|u_{2,t}| - ln|u_{2,t-1}|)\right] \\
&+E\left[(ln|u_{1,t+1}| - ln|u_{1,t}|)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right]. \\
&= \tfrac{1}{16}\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right]^2 \cdot \left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right]^2 \\
&\quad \left[\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \sum_{j=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-j}} + 2\sum_{i=1}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-i}}\right)^2 \sum_{j=1, j\neq i}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-j}}\right)^2\right] \\
&+\tfrac{1}{4}\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right]^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot 2(E[(ln|u_t|)^2] - E[ln|u_t|]^2) \\
&+\tfrac{1}{4}\left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right]^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] \\
&+\left[ln\left(m_1\right) - ln\left(2 - m_1\right)\right] \cdot \left[ln\left(m_2\right) - ln\left(2 - m_2\right)\right] \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot \left(E[ln|u_t|]^2 - E[ln|u_{1,t}|ln|u_{2,t}|]\right) \\
&+E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 4E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot ln|u_{2,t}|]E[ln|u_t|]^2 \\
&+3E[ln|u_t|^2]^2 - 4E[ln|u_t|^2]E[ln|u_t|]^2.
\end{aligned}
\tag{C14}
$$

We also consider the moments for individual transformed observations of $E[X_{t+1,1}^2 \cdot X_{t,1}^2]$ and $E[Y_{t+1,1}^2 \cdot Y_{t,1}^2]$, which can be found in Lux (2008).

7.2.2 Lognormal case

The Lognormal case applies when a new multiplier M_t is needed at any cascade level, it will be determined via a random draw from a Lognormal distribution, i.e., $-\log_2 M_1 \sim N(\lambda_1, \sigma_{m1}^2)$ and $-\log_2 M_2 \sim N(\lambda_2, \sigma_{m2}^2)$.

$$
\begin{aligned}
&E[X_{t,1}, Y_{t,1}] \\
&= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}) + (ln|u_{1,t}| - ln|u_{1,t-1}|)\right]\left[\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|)\right]\right\} \\
&= \tfrac{1}{4}E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)\right] - 2E[ln|u_t|]^2 + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|].
\end{aligned}
\tag{D1}
$$

For the Lognormal case, $E[X_{t,1}, Y_{t,1}] = -2E[ln|u_t|]^2 + 2E[ln|u_{1,t}| \cdot ln|u_{2,t}|]$.

$$
\begin{aligned}
&E[X_{t+1,1}, Y_{t,1}] \\
&= E\left\{\left[\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}) + (ln|u_{1,t+1}| - ln|u_{1,t}|)\right]\left[\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)}) + (ln|u_{2,t}| - ln|u_{2,t-1}|)\right]\right\} \\
&= \tfrac{1}{4}E\left[\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right) \cdot \sum_{i=1}^{k}\left(\xi_t^{(i)} - \xi_{t-1}^{(i)}\right)\right] + E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|] \\
&= E[ln|u_t|]^2 - E[ln|u_{1,t}| \cdot ln|u_{2,t}|].
\end{aligned}
\tag{D2}
$$

We also need the moment condition for one single time series $E[X_{t+1,1}, X_{t,1}]$ and $E[Y_{t+1,1}, Y_{t,1}]$, which can be found in Lux (2008)

Then we turn to moment conditions for squared transformed observations. By recalling

Eq. (C10), we have

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 \left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2 (ln|u_{1,t}| - ln|u_{1,t-1}|)^2\right] \\
&+E\left[\left(\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})\right)(ln|u_{1,t}| - ln|u_{1,t-1}|)\left(\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)(ln|u_{2,t}| - ln|u_{2,t-1}|)\right] \\
&+E\left[(ln|u_{1,t}| - ln|u_{1,t-1}|)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right].
\end{aligned}
\tag{D3}
$$

We know $(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)})^2 = 2(E[(\varepsilon_t^{(i)})^2] - E[\varepsilon_t^{(i)}]^2) = 2\sigma_\varepsilon^2$, and $(\xi_t^{(i)} - \xi_{t-1}^{(i)})^2 = 2(E[(\xi_t^{(i)})^2] - E[\xi_t^{(i)}]^2) = 2\sigma_\xi^2$. By examining each component in Eq. (D3) based on the experiences of previous calculations, we arrive the final solution for $E[X_{t,1}^2 \cdot Y_{t,1}^2]$ for the Lognormal case:

$$
\begin{aligned}
&E[X_{t,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{4}\sigma_\varepsilon^2 \cdot \sigma_\xi^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} + \tfrac{1}{2}\sigma_\varepsilon^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] + \tfrac{1}{2}\sigma_\xi^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] \\
&+\sigma_\varepsilon^2 \cdot \sigma_\xi^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot \left(2E[ln|u_{1,t}|ln|u_{2,t}| - 2E[ln|u_t|]^2]\right) \\
&+2E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 8E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] \\
&+4E[ln|u_{1,t}| \cdot (ln|u_{2,t}|]^2 + 2E[(ln|u_t|)^2]^2.
\end{aligned}
\tag{D4}
$$

For the moment condition of $E[X_{t+1,1}^2 \cdot Y_{t,1}^2]$, By recalling Eq. (D5), we have

$$
\begin{aligned}
&E[X_{t+1,1}^2 \cdot Y_{t,1}^2] \\
&= \tfrac{1}{16} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 \left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right] \\
&+\tfrac{1}{4} E\left[\left(\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)^2 (ln|u_{1,t+1}| - ln|u_{1,t}|)^2\right] \\
&+E\left[\left(\tfrac{1}{2}\sum_{i=1}^{k}(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)})\right)(ln|u_{1,t+1}| - ln|u_{1,t}|)\left(\tfrac{1}{2}\sum_{i=1}^{k}(\xi_t^{(i)} - \xi_{t-1}^{(i)})\right)(ln|u_{2,t}| - ln|u_{2,t-1}|)\right] \\
&+E\left[(ln|u_{1,t+1}| - ln|u_{1,t}|)^2 (ln|u_{2,t}| - ln|u_{2,t-1}|)^2\right] \\
&= \tfrac{1}{4}\sigma_\varepsilon^2 \cdot \sigma_\xi^2 \left[\sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \sum_{j=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-j}} + 2\sum_{i=1}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-i}}\right)^2 \sum_{j=1, j\neq i}^{k}\left(\tfrac{1}{2}\tfrac{1}{2^{k-j}}\right)^2\right] \\
&+\tfrac{1}{2}\sigma_\varepsilon^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot 2(E[(ln|u_t|)^2] - E[ln|u_t|]^2) \\
&+\tfrac{1}{2}\sigma_\xi^2 \cdot \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot E[(ln|u_t| - ln|u_{t-1}|)^2] \\
&+\sigma_\varepsilon^2 \cdot \sigma_\xi^2 \sum_{i=1}^{k} \tfrac{1}{2}\tfrac{1}{2^{k-i}} \cdot \left(E[ln|u_t|]^2 - E[ln|u_{1,t}|ln|u_{2,t}|]\right) \\
&+E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2] - 4E[(ln|u_{1,t}|)^2 \cdot ln|u_{2,t}|] \cdot E[ln|u_t|] + 4E[ln|u_{1,t}| \cdot ln|u_{2,t}|]E[ln|u_t|]^2 \\
&+3E[ln|u_t|^2]^2 - 4E[ln|u_t|^2]E[ln|u_t|]^2.
\end{aligned}
\tag{D5}
$$

We also consider the moments for individual transformed observations of $E[X_{t+1,1}^2 \cdot X_{t,1}^2]$ and $E[Y_{t+1,1}^2 \cdot Y_{t,1}^2]$, which can be found in Lux (2008).

7.3 (Numerical) Solutions of the moments for the log of (bivariate) Normal variates

To implement GMM estimation for multifractal processes, we need also the solutions of the first to fourth moments for the log of absolute values of random variates from the standard Normal distribution: $E[ln|u_t|^q]$ (for $q = 1, 2, 3, 4$). It requires to analyze the following integrals:

$$\int_0^\infty [ln(u)]^q \cdot e^{-a \cdot u^2} du, \tag{D1}$$

for $a = 0.5$. Different moments of Eq. (D1) can be solved using Euler *Gamma*, which is Euler's constant γ, with numerical value of

$$\gamma \simeq 0.577216,$$

and the Riemann zeta function which is

$$Zeta(x) = \sum_{k=1}^{\infty} k^{-x}.$$

$$\int_0^\infty ln(u) \cdot e^{-a \cdot u^2} du = -\frac{\sqrt{\pi}}{4\sqrt{a}}(\gamma + ln(4) + ln(a)) \tag{D2}$$

$$\int_0^\infty [ln(u)]^2 \cdot e^{-a \cdot u^2} du \tag{D3}$$
$$= \frac{1}{16\sqrt{a}}[\sqrt{\pi}(2\gamma^2 + \pi^2 + 4\gamma ln(4) + 2(ln(4))^2 + 4(\gamma + ln(4))ln(a) + 2(ln(a))^2)]$$

$$\int_0^\infty [ln(u)]^3 \cdot e^{-a \cdot u^2} du \tag{D4}$$
$$= -\frac{1}{32\sqrt{a}}[\sqrt{\pi}(2\gamma^3 + 6\gamma^2 ln(4) + 2(ln(4))^2 + 3\gamma(\pi^2 + 2(ln(4))^2) + \pi^2 ln(64) +$$
$$3(2\gamma^2 + \pi^2 + 4\gamma ln(4) + 2(ln(4))^2)ln(a) + 6(\gamma + ln(4))(ln(a))^2) +$$
$$2(ln(a))^3 + 28 \cdot Zeta(3))]$$

$$\int_0^\infty [ln(u)]^4 \cdot e^{-a\cdot u^2} du \tag{D5}$$

$$= -\frac{1}{128\sqrt{a}}[\sqrt{\pi}(4\gamma^4 + 7\pi^4 + 16\gamma^3 ln(4) + 12\pi^2 (ln(4))^2 + 4(ln(4))^4 +$$
$$12\gamma^2(\pi^2 + 2(ln(4))^2) + 12(2\gamma^2 + \pi^2 + 4\gamma ln(4) + 2(ln(4))^2) + (ln(a))^2) +$$
$$16(\gamma + ln(4))(ln(a))^3 + 4(ln(a))^4 + 224 ln(4)\cdot Zeta(3) + 8\gamma(2(ln(4))^3) +$$
$$\pi^2 ln(64) + 28\cdot Zeta(3)) + 8ln(a)(2\gamma^3 + 6\gamma^2 ln(4) + 2(ln(4))^3 +$$
$$3\gamma(\pi^2 + 2(ln(4))^2) + \pi^2 ln(64) + 28\cdot Zeta(3))]$$

Furthermore, solutions for three forms of moments for the log of absolute value of bivariate Normal variates with the correlation parameter ρ are needed, that is:

$$E[ln|u_{1,t}| \cdot ln|u_{2,t}|],$$

$$E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)],$$

$$E[(ln|u_{1,t}|)^2 \cdot (ln|u_{2,t}|)^2],$$

which require to solve the following double integrals:

$$\int_0^\infty \int_0^\infty [ln(u_1)]^p \cdot [ln(u_2)]^q \cdot e^{-a\cdot(u_1^2 + u_2^2 - 2u_1u_2\rho)} du_1 du_2, \tag{D6}$$

(for $p = 1, 2$; $q = 1, 2$), with $a = \frac{1}{2(1-\rho^2)}$.

Let $u_1 = \sqrt{1-\rho^2}\alpha$, and $u_2 = \sqrt{1-\rho^2}\beta$, which lead to a new expression of Eq. (D6):

$$(1-\rho^2)\int_0^\infty \int_0^\infty [ln(\alpha) + ln(\sqrt{1-\rho^2})]^p \cdot [ln(\beta) + ln(\sqrt{1-\rho^2})]^q e^{\alpha\beta\rho} e^{-\frac{\alpha^2+\beta^2}{2}} d\alpha d\beta. \tag{D7}$$

We first have a look at the case of $p = q = 1$ (other cases of $p = q = 2$; p = 1, q = 2 can be calculated analogously).

By decomposition ($n \geq 50$ is required).[1]

$$e^{\alpha\beta\rho} \simeq \sum_{n\geq 0} \frac{\rho^n}{n!}\alpha^n\beta^n, \tag{D8}$$

[1]The larger n is, the more precise the decomposition is.

Eq. (D7) is transformed to

$$\begin{aligned}
& (1-\rho^2)\sum_{n\geq 0}\frac{\rho^n}{n!}\left(\int_0^\infty (ln(\alpha)+ln(\sqrt{1-\rho^2})\alpha^n e^{-\frac{\alpha^2}{2}}d\alpha\right)\left(\int_0^\infty (ln(\beta)+ln(\sqrt{1-\rho^2})\beta^n e^{-\frac{\beta^2}{2}}d\beta\right) \\
= \quad & (1-\rho^2)\sum_{n\geq 0}\frac{\rho^n}{n!}\left(\int_0^\infty (ln(\alpha)+ln(\sqrt{1-\rho^2})\cdot\alpha^n\cdot e^{-\frac{\alpha^2}{2}}d\alpha\right)^2 \\
= \quad & (1-\rho^2)\sum_{n\geq 0}\frac{\rho^n}{n!}\left(P_n+Q_n\cdot ln(\sqrt{1-\rho^2})\right)^2, \qquad \text{(D9)}
\end{aligned}$$

where, $P_n=\int_0^\infty ln(\alpha)\cdot\alpha^n\cdot e^{-\frac{\alpha^2}{2}}d\alpha$, and $Q_n=\int_0^\infty \alpha^n\cdot e^{-\frac{\alpha^2}{2}}d\alpha$, which can be solved by recurrences, cf. Gradshteyn et al. (2000):

$$P_{n+2}=(n+1)P_n+Q_n, \qquad \text{(D10)}$$

$$Q_{n+2}=(n+1)Q_n, \qquad \text{(D11)}$$

Implementing Eq. (D10) and Eq. (D11) requires the initial conditions of P_0, P_1, Q_0, and Q_1, which are given below:

$$P_0=\int_0^\infty ln(\alpha)\cdot e^{-\frac{\alpha^2}{2}}d\alpha=-0.5\sqrt{\frac{\pi}{2}}\left(\gamma+ln(2)\right), \qquad \text{(D12)}$$

$$P_1=\int_0^\infty ln(\alpha)\cdot\alpha\cdot e^{-\frac{\alpha^2}{2}}d\alpha=-0.5\left(\gamma-ln(2)\right), \qquad \text{(D13)}$$

$$Q_0=\int_0^\infty e^{-\frac{\alpha^2}{2}}d\alpha=\sqrt{\frac{\pi}{2}}, \qquad \text{(D14)}$$

$$Q_1=\int_0^\infty \alpha\cdot e^{-\frac{\alpha^2}{2}}d\alpha=1. \qquad \text{(D15)}$$

Gradshteyn et al. (2000) "Table of Integrals, Series, and Products", (Academic Press) provides the (numerical) solutions of various integrals above. And they can be verified by the fully integrated technical computing package – *Mathematica*®.

Bibliography

Acerbi, C., Nordio, C., and Sirtori, C. (2001). Expected shortfall as a tool for financial risk management. Working paper of Italian Association for financial risk management. Manuscript.

Andersen, T. G. (1996). Return volatility and trading volume: An information flow interpretation of stochastic volatility. *Journal of Finance*, 60:169–204.

Andersen, T. G. and Bollerslev, T. (1996). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4:115–158.

Andersen, T. G. and Sorensen, B. E. (1996). GMM estimation of a Stochastic Volatility Model: A Monte Carlo study. *Journal of Business and Economic Statistics*, 14:328–352.

Artzner, P., Delbean, F., Eber, M., and Heath, D. (1997). Thinking coherently. *Risk*, 10:68–71.

Baillie, R. (1996). Long memory process and fractionally integration in econometrics. *Journal of Econometrics*, 73:5–59.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30.

Baillie, R. T. and Myers, R. J. (1996). Bivariate garch estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics*, 6:109 – 124.

Barkoulas, J. and Baum, C. (2000). Long memory in the Greek stock market. *Applied Financial Economics*, 10:177–184.

Batten, J. and Ellis, C. (1996). Fractal structures in currency markets : evidence from the spot Australian/US dollar. *Advances in Pacific Basin financial markets*, 2A:173–181.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.

Berzuini, C., Best, N. B., Gilks, W. R., and Larizza, C. (1997). Dynamic conditional independence models and Markov Chain Monte Carlo methods. *Journal of the American Statistical Association*, 92:1403–1412.

Black, F. (1976). Studies of stock price volatility changes. Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section. 177.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, 72:498–505.

Bollerslev, T., Engle, R., and Wooldridge, J. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96:116–131.

Booth, G. and Kaen, F. (1979). Gold and silver spot prices and market information efficiency. *Financial Review*, 14:21–26.

Booth, G., Kaen, F., and Koveos, P. (1982). R/s analysis of foreign exchange rates under two international monetary regimes. *Journal of Monetary Economics*, 10:407–415.

Bouchaud, J. P. and Potters, M. (2001). More stylized facts of financial markets: leverage effect and downside correlations. *Physica A*, 266:60–70.

Boutahar, M., Marimoutou, V., and Nouira, L. (2007). Estimation methods of the long memory parameter: Monte carlo analysis and application. *Journal of Applied Statistics*, 34:261–301.

Breidt, F., Crato, N., and de Lima, P. (1998). On the detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325–348.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, 65:1059–1095.

Brockwell, P. J. and Dahlhaus, R. (2004). Generalized levinson- durbin and burg algorithms. *Journal of Econometrics*, 118:129–149.

Bucay, N. and Rosen, D. (1996). Credit risk of an international bond portfolio: a case study. *ALGO Research Quarterly*, 2:9–29.

Byers, D., Davidson, J., and Peel, D. (1997). Modelling political popularity: An analysis of long-range dependence in opinion poll series. *Journal of Royal Statistical Society*, 160:471–490.

Calvet, L. and Fisher, A. (2001). Forecasting multifractal volatility. *Journal of Econometrics*, 105:27–58.

Calvet, L. and Fisher, A. (2002). Multi-fractality in asset returns: Theory and evidence. *Review of Economics and Statistics*, 84:381–406.

Calvet, L. and Fisher, A. (2004). How to forecast long-run volatility: Regime switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, 2:49–83.

Calvet, L., Fisher, A., and Thompson, S. (2006). Volatility comovement: A multifrequency approach. *Journal of Econometrics*, 131:179–215.

Cavalcante, J. and Assaf, A. (2002). Long range dependence in the returns and volatility of the brazilian stock market. Rio de Janeiro. Manuscript.

Chiarella, C. and He, X. Z. (2001). Asset price and wealth dynamics under heterogeneous expectations. *Quantitative Finance*, 1:509–526.

Christoffersen, P. F. (2003). *Elements of Financial Risk Management*. Academic press, Amsterdam.

Chung, C.-F. (2002). Estimating the fractionally integrated garch model. Mimeo: Academia Sinica.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1:1–14.

Cootner, P. (1964). *The random character of stock market price*. MIT press, Cambridge, MA.

Crato, N. and De Lima, P. J. F. (1994). Long-range dependence in the conditional variance of stocks returns. *Economics Letters*, 45:281–285.

Danielsson, J. and Richard, J. F. (1993). Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, 8:153–173.

Davies, R. and Harte, D. (1987). Test for Hurst effect. *Biometrika*, 74:95–101.

Demos, A. and Vassilicos, C. (1994). The multifractal structure of high frequency foreign exchange rate fluctuations. *LSE Financial Markets Group Disscussion Paper Series No.15*.

Ding, Z., Engle, R., and Granger, C. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1:83–106.

Dolado, J., Gonzalo, J., and Mayoral, L. (2003). Long-range dependence in spanish political opinion poll series. *Journal of Applied Econometrics*, 18:137–155.

Dowd, K. (2002). *Measuring Market Risk*. John Wiley and Sons, New York.

Embrechts, P., Klueppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer, Berlin.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, 50:987–1008.

Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11:122–150.

Engle, R. F. and Manganelli, S. (2002). Caviar: Conditional autoregressive value at risk by regression quantiles. University of California, San Diego.

Engle, R. F. and Susmel, R. (1993). Common volatility in international equity markets. *Journal of Business and Economic Statistics*, 11:167–176.

Fama, E. F. (1965). The behavior of stock market prices. *Journal of Bussiness*, 38:34–105.

Fang, H., Lai, K. S., and Lai, M. (1994). Fractal structure in currency futures price dynamics. *The Journal of Futures Markets*, 14:169–181.

Feder, J. (1988). *Fractals*. Plenum Press, New York.

Fillol, J. (2003). Multifractality: Theory and evidence an application to the french stock market. *Economics Bulletin*, 3:1–12.

Fisher, A., Calvet, L., and Mandelbrot, B. (1997). Multifractality of deutschmark/US dollar exchange rate. Cowles Foundation Disscussion paper 1165. Manuscript.

Gallucio, S., Caldarelli, G., Marsili, M., and Zhang, Y. C. (1997). Scaling in currency exchange. *Physica A*, 245:423–436.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series*, 4:221–238.

Giraitis, L., Kokoszka, P., Leipus, R., and Teyssiere, G. (2000). Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity. *Mathematics and Statistics*, 3:113–128.

Gordon, N. J., Salmond, D. J., and Smith, A. (1993). Anovel approach to non-linear and non-gaussian bayesian state estimation. *IEEE Proceedings F*, 140:107–113.

Gourieroux, C. (1997). *ARCH Models and Financial Applications*. Springer, New York.

Gourieroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, 8:85–118.

Gradshteyn, I. S., Ryzhik, I. M., Jeffrey, A., and Zwillinger, D. (2000). *Table of Integrals, Series, and Products*. Academic Press, London. Sixth Edition.

Granger, C. W. J. and Joyeux, R. (1980). A introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1:15–29.

Greene, M. T. and Fielitz, B. D. (1977). Long-term dependence in common stock market. *Journal of Financial Economics*, 4:339–349.

Guillaume, D., Dacorogna, M., Dave, R., Muller, U., Olsen, R., and Pictet, O. (2000). From the birds eye view to the microscope: A survey of new stylized facts of the intraday foreign exchange markets. *Finance and Stochastics*, 1:95–131.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50:1029–1054.

Harris, D. and Mátyás, L. (1999). *Generalized Method of Moments Estimation*. Cambridge, University Press. Chap. 1.

Harte, D. (2001). *Multifractals: Theory and Applications*. Chapman and Hall, London.

Harvey, A. C. (1993). Long memory in stochastic volatility. Unpublished manuscript.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, 61:247–264.

Helms, B., Kaen, F., and Rosenman, R. (1995). Memory in commodity futures markets. *Journal of Futures Markets*, 4:559–567.

Holton, G. A. (2003). *Value-at-Risk: Theory and Practice*. Academic Press, California.

Hosking, J. R. M. (1981). Fractional differencing. *Biometrika*, 68:165–176.

Hurst, E. (1951). Long term storage capacity of reservoirs. *Transactions on American Society of Civil Engineering*, 116:770–808.

Hurvich, C. M., Deo, R. S., and Brodsky, J. (1998). The mean squared error of geweke and porter-hudaks estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19:19–46.

Iori, G. (2002). A micro-simulation of traders' activity in the stock market: the rule of heterogeneity , agents' interactions and trade friction. *Journal of Economic Behaviour and Organisation*, 49:269–285.

Jacquier, E., Polson, N., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–389.

Jensen, M. (1999). Using wavelets to obtain a consistent ordinary least squares estimator of the long memory parameter. *Journal of Forecasting*, 18:17–32.

Kim, S., Shephard, N., and Chilb, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH model. *Review of Economic Studies*, 65:361–393.

Kirman, A. (1993). Ants, rationality, and recruitment. *Quarterly Journal of Economics*, 108:137–156.

Kirman, A. and Teyssière, G. (2001). Microeconomic models for long memory in the volatility of financial time series. *Studies in Nonlinear Dynamics and Econometrics*, 5:281–302.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3:73–84.

Kwiatkowski, D., Phillips, P., Schmidl, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative era unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178.

Lee, D. and Schmidt, P. (1996). On the power of the kpss test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics*, 73:285–302.

Leland, W. E., Taqqu, M. S., Willinger, W., and Wilson, D. V. (1994). On the self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2:1–15.

Liesenfeld, R. (1998). Dynamic bivariate mixture models: modeling the behavior of prices and trading volume. *Journal of Business and Economic Statistics*, 16:101–109.

Liesenfeld, R. (2001). A generalized bivariate mixture model for stock price volatility and trading volume. *Journal of Econometrics*, 104:141–178.

Liesenfeld, R. and Jung, R. C. (2000). Stochastic volatility models: conditional normality versus heavy-tailed distributions. *Journal of Applied Econometrics*, 15:137–160.

Liesenfeld, R. and Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 207:1–27.

Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59:1279–1313.

Lo, A. W. and MacKinlay, A. C. (2001). *A Non-Random Walk Down Wall Street*. Princeton University Press.

Lobato, I. N. and Velasco, C. (2000). Long memory in stock-market trading volume. *Journal of Business and Economic Statistics*, 18:410–427.

Lux, T. (1995). Herd behaviour, bubbles and crashes. *Economic Journal*, 105:881–896.

Lux, T. (1996a). Long-term stochastic dependence in financial prices: Evidence from the German stock market. *Applied Economics Letters*, 3:701–706.

Lux, T. (1996b). The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. *Applied Financial Economics*, 6:463–475.

Lux, T. (1998). The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behavior and Organization*, 33:143–165.

Lux, T. (2001). Turbulence in financial markets: The surprising explanatory power of simple cascade models. *Quantitative Finance*, 1:632–640.

Lux, T. (2004). Detecting multi-fractal properties in asset returns: The failure of the "scaling estimator". *International Journal of Modern Physics*, 15:481–491.

Lux, T. (2008). The Markov-switching multi-fractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business and Economic Statistics*, 26:194–210.

Lux, T. and Kaizoji, T. (2007). Forecasting volatility and volume in the Tokyo stock market: Long-memory, fractality and regime switching. *Journal of Economic Dynamics and Control*, 31:1808 – 1843.

Lux, T. and Marchesi, M. (1999). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397:498–500.

Lux, T. and Marchesi, M. (2000). Volatility clustering in financial markets: A micro-simulation of interacting agents. *International Journal of Theoretical and Applied Finance*, 3:67–702.

Madhusoodanan, T. P. (1998). Long-term dependence in Indian stock market. *Journal of Financial Studies*, 5:33–53.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, 35:394–419.

Mandelbrot, B. (1967a). How long is the coast of Britain? statistical self-similarity and fractional dimension. *Science*, 156:936–938.

Mandelbrot, B. (1967b). The variation of the prices of cotton, wheat, and railroad stocks, and of some finacial rates. *Journal of Business*, 40:393–413.

Mandelbrot, B. (1971). When can price be arbitrage efficiency? a limit to the validity of the random walk and martingale models. *Review of Economics and Statistics*, 53:225–236.

Mandelbrot, B. (1974). Intermittent turbulence in self similar cascades: Divergence of high moments and dimension of carrier. *Journal of Fluid Mechanics*, 62:331–358.

Mandelbrot, B. (1997). Three fractal models in finance : discontinuity, concentration, risk. *Economic notes : economic review of Banca Monte dei Paschi di Siena. Oxford Blackwell*, 26:171–211.

Mandelbrot, B. (2001). Scaling in financial prices: Tails and dependence. *Quantitative Finance*, 1:113–123.

Mandelbrot, B., Fisher, A., and Calvet, L. (1997). A multifractal model of asset returns. Cowles Foundation for Research and Economics. Manuscript.

Mandelbrot, B. and Taqqu, M. S. (1979). Robust R/S analysis of long-run serial correlation. *Bulletin of the International Statistical Institute*, 48:69–104.

Mandelbrot, B. and Wallis, J. R. (1969a). Computer experiments with fractional gaussian noise. *Water Resources Research*, 5:228–267.

Mandelbrot, B. and Wallis, J. R. (1969b). Robustness of the rescaled range R/S in the measurement of non noncyclic long-run statistical dependence. *Water Resources Research*, 5:967–988.

Marmol, F. (1997). Searching for fractional evidence using unit root tests. Universidad Carlos III, Madrid. Mimeo.

Matsushita, R., Gleria, I., Figueiredo, A., and Da Silva, S. (2003). Fractal structure in the Chinese yuan/US dollar rate. *Economic Bulletin*, 7:1–13.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, 45:239–265.

Mills, T. (1993). Is there long-term memory in UK stock returns? *Appllied Financial Economics*, 3:303–306.

Moody, J. and Wu, L. (1996). Improved estimates for the rescaled range and hurst exponents. Appears in "Neural Networks in the Capital Markets", World Scientific, London.

Mulligan, R. F. (2000). A fractal analysis of foreign exchange markets. *International Advances in Economic Research*, 26:33–49.

Mulligan, R. F. (2004). Fractal analysis of highly volatile markets: an application to technology equities. *The Quarterly Review of Economics and Finance*, 44:155–179.

Näslund, B. (1990). A fractal analysis of capital structure. *Research paper: Ekonomiska Forskningsinstitutet, Handelshoegskolan in Stockholm No.6425.*

Nath, G. C. (2002). Long memory and indian stock market - an empirical evidence. WDM Dept, Mumbai. Manuscript.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance*, 3:15–102.

Parisi, G. and Frisch, U. (1985). Turbulence and predictability in geophysical fluid dynamics and climate dynamics. In Ghil, M., editor, *A multifractal model of intermittency*, pages 84–92. North-Holland.

Peng, C. K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotidies. *Physical Review E*, 49:1685–1689.

Peters, E. E. (1994). *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics.* John Wiley and Sons, US.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filter. *Journal of the American Statistical Association*, 94:590–599.

Rachev, S. T. and Mittnik, S. (2000). *Stable Paretian Models in Finance*. John Wiley & Sons, Chichester.

Ray, B. K. and Tsay, R. (2001). Long-range dependence in daily stock volatilities. *Journal of Business and Economic Statistics*, 18:254–262.

Rubin, D. B. (1987). Comment: A noniterative Sampling/Importance Resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association*, 82:543–546.

Schmitt, F., Schertzer, D., and Lovejoy, S. (2000). Multifractal fluctuations in finance. *International Journal of Theoretical and Applied Finance*, 3:361–364.

Shephard, N. G. (1993). Fitting non-linear time series models, with application to stochastic variance models. *Journal of Applied Econometrics*, 8:135–152.

Shiller, R. J. (1989). *Market volatility*. MIT Press, Cambridge.

Sibbertsen, P. and Kramer, W. (2006). The power of the kpss - test for cointegration when residuals are fractionally integrated. *Economics Letters*, 91:321–324.

So, M. P. (1999). Time series with additive noise. *Biometrika*, 86:474–482.

Sowell, F. (1992). Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, 53:165–188.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes - a study of daily sugar price 1961-79. In Anderson, O. D., editor, *Time series analysis: Theory and Practice*, pages 203–226. North-Holland, Amsterdam.

Taylor, S. J. (1986). *Modelling Financial Time Series*. John Wiley, UK.

Teverovsky, V., Taqqu, M. S., and Willinger, W. (1999). A critical look at Lo's modified R/S statistic. *Journal of Statistical Planning and Inference*, 80:211–227.

Tolvi, J. (2003). Long-term momory in a small stockt market. *Economics Bulletin*, 7:1–13.

Upton, D. and Shannon, D. (1979). The stable paretian distribution, subordinated stochastic processes, and asymptotic lognormality: an empirical investigation. *Journal of Finance*, 34(4):1031–1039.

Weron, R., Weron, K., and Weron, A. (1999). A conditionally exponential decay approach to scaling in finance. *Physica A*, 267:551–561.

Willinger, W., Taqqu, S., and Teverovsky, V. (1999). Stock market prices and long-range dependence. *Finance and Stochastics*, 3:1–13.

Wright, J. H. (1999). A new estimator of the fractionally integrated stochastic volatility model. *Economics Letter*, 63:295–303.

Yamai, Y. and Yoshiba, T. (2002). On the validity of Value-at-Risk: comparative analysis with expected shortfall. *Monetary and Economic Studies*, 20:57–86.

Zumbach, G. (2004). Volatility processes and volatility forecast with long memory. *Quantitative Finance*, 4:70–86.

Ich erkläre hiermit an Eides Statt, dass ich meine Doktorarbeit "Multivariate Multifractal Models: Estimation of Parameters and Applications to Risk Management" selbstständig 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 zitiert habe.

(Ruipeng Liu)

Melbourne, September 2008