Lux, Thomas

Working Paper

The Markov-switching multi-fractal model of asset returns: GMM estimation and linear forecasting of volatility

Economics Working Paper, No. 2004-11

Provided in Cooperation with:
Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Lux, Thomas (2004) : The Markov-switching multi-fractal model of asset returns: GMM estimation and linear forecasting of volatility, Economics Working Paper, No. 2004-11, Kiel University, Department of Economics, Kiel

This Version is available at:
https://hdl.handle.net/10419/3407

The Markov-Switching Multi-Fractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility

(A revised version of EWP 2003-13)

by Thomas Lux

Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2004-11

The Markov-Switching Multi-Fractal Model of Asset Returns: GMM Estimation and Linear Forecasting of Volatility

Thomas Lux
Department of Economics
University of Kiel

November 2004

Abstract

Multi-fractal processes have recently been proposed as a new formalism for modelling the time series of returns in finance. The major attraction of these processes is their ability to generate various degrees of long memory in different powers of returns - a feature that has been found in virtually all financial data. Initial difficulties stemming from non-stationarity and the combinatorial nature of the original model have been overcome by the introduction of an iterative Markov-switching multi-fractal model in Calvet and Fisher (2001) which allows for estimation of its parameters via maximum likelihood and Bayesian forecasting of volatility. However, applicability of MLE is restricted to cases with a discrete distribution of volatility components. From a practical point of view, ML also becomes computationally unfeasible for large numbers of components even if they are drawn from a discrete distribution. Here we propose an alternative GMM estimator together with linear forecasts which in principle is applicable for any continuous distribution with any number of volatility components. Monte Carlo studies show that GMM performs reasonably well for the popular Binomial and Lognormal models and that the loss incured with linear compared to optimal forecasts is small. Extending the number of volatility components beyond what is feasible with MLE leads to gains in forecasting accuracy for some time series.

Keyword: Markov-switching, multifractal, forecasting, volatility, GMM estimation

JEL Classification: C20, G12

Correspondence: Department of Economics, University of Kiel, Olshausenstr. 40; 24118 Kiel, Germany, Email: lux@bwl.uni-kiel.de

1 Introduction

The recent proposal of multi-fractal models (Mandelbrot, Fisher and Calvet, 1997; Mandelbrot, 1999; Calvet and Fisher, 2001, 2002a) has added an interesting new entry to the rich variety of volatility models available in financial econometrics (cf. Poon and Granger, 2003, for an up-to-date review). The essential new feature of this class of models is its multiplicative, hierarchical structure of volatility components with heterogeneous frequencies. Research on multifractal models originated from statistical physics where they had been proposed as models for turbulent flows (e.g. Mandelbrot, 1974). The main attraction in the financial sphere is the ability of these models to generate different degrees of long-term dependence in various powers of returns - a feature pervasively found in all financial data (cf. Ding, Engle and Granger, 1993; Andersen and Bollerslev, 1997; Lobato and Savin, 1999). Unfortunately, multifractal models used in physics were of a combinatorial nature rather than iterative processes and they suffered from non-stationarity due to the limitation to a bounded interval and the non-convergence of moments in the continuous-time limit. These major weaknesses were overcome by Calvet and Fisher (2001) who introduced a Markov-switching multi-fractal model based on Poisson arrivial times for which weak convergence to the continuous-time limit could be demonstrated. The interpretation as a Markov-switching process (albeit with a possibly huge number of states) also allowed maximum likelihood estimation of the parameters for cases with a discrete distribution of volatility components and forecasting based on the current conditional probabilities of volatility states. The implementation of this procedure in Calvet and Fisher (2004) showed that this new model provides gains in forecasting accuracy for medium and long horizons (up to 50 days) over forecasts from GARCH and FIGARCH models.

Our contribution in this paper is twofold: (i) we introduce an alternative GMM estimator of multi-fractal parameters which could be used in cases in which ML is not applicable or computationally infeasible, (ii) we propose linear forecasting which also is universally applicable and does not require particular specifications of the distribution and number of volatility components. We first explore the behavior of GMM plus linear forecasting against the benchmark of ML and optimal forecasts for the special case of the Binomial model used in Calvet and Fisher (2004). As it turns out, GMM is, of course, less efficient than ML but it is nicely behaved in that it has small biases and reasonable mean squared errors for the crucial multi-fractal parameters. Monte Carlo results are in good harmony with $T^{\frac{1}{2}}$ consistency and no problems of non-convergence or multiplicity of solutions are encountered in our simulations. Furthermore, despite the sometimes sizable difference in Monte Carlo MSEs between GMM and MLE, differences are less pronounced with respect to forecasting performance: GMM with linear forecasts typically only has a very slight disadvantage against ML-based optimal forecasts. Additional Monte Carlo runs also show that the performance of GMM and linear forecasts is not adversely affected by increasing the number of volatility components or by replacing the discrete Binomial distribution of multipliers by a continuous Lognormal distribution.

Our empirical application shows that using a larger state space indeed provides further scope for improvement of long-run volatility predictions while the replacement of the Binomial model by the Lognormal specification seems to make less of a difference. In the sample of assets studied here, we find the multi-fractal models to outperform GARCH and FIGARCH models for an exchange rate series (USD-DEM) and the price of gold, while for two applications to stock indices (the NYSE index and the German DAX) the performance is similar to that of these two benchmarks. The structure of the remainder is as follows: section 2 shortly reviews available literature on multi-fractal models, section 3 introduces the GMM estimator and compares its performance to MLE while section 4 deals with linear against optimal predictors. Section 5 presents our empirical findings before we provide concluding remarks in section 6. Two appendices provide detailed derivations of the moment conditions used for GMM estimation, and closed-form solutions of autocovariances needed to construct best linear forecasts.

2 Review of Multifractal Measures and Models

Multifractal measures have a long history in physics dating back at least to the early seventies when Mandelbrot proposed a probabilistic approach for the distribution of energy in turbulent dissipation (e.g., Mandelbrot, 1974). In this original setting, multifractals result from operations performed on probability measures.[1] The construction of a multifractal 'cascade' starts by assigning uniform probability to a bounded interval (e.g., the unit interval [0,1]). In a first step, this interval is split up into two subintervals receiving fractions m_0 and $1 - m_0$, respectively, of the total probability mass of unity of their mother interval. In the simplest case, both subintervals have the same length (i.e., 0.5), but other choices are possible as well. To continue, the two subintervals of the first stage of the cascade are split up again into similar subintervals (of length 0.25 each in the simplest case) receiving again fractions m_0 and $1 - m_0$ of the probability mass of their 'mother' intervals. In principle, this procedure is repeated *ad infinitum*. With this recipe, a heterogeneous, *fractal* distribution of the overall probability mass results which even for the most elementary cases has a perplexing visual resemblance to time series of volatility in financial markets.

Many variations of the above generating mechanism of a simple Binomial multi-fractal could be thought of: instead of always assigning probability m_0 to the left-hand decendent, this assignment could as well be randomized. Furthermore, one could think of more than two subintervals to be generated in each step (leading to multinomial cascades) or of using random numbers for m_0 instead of the same constant value. A popular example of the later generalisation is the Lognormal multifractal model which draws its probabilities for new branches of the cascade from a Lognormal distribution (cf. Mandelbrot, 1974; Mandelbrot et al., 1997; Calvet and Fisher, 2002a).

[1] Evertz and Mandelbrot (1992) and Harte (2001) are recommendable introductory sources to multifractal measures and their applications in the natural science.

An application of the above formalism to financial data has been first proposed by Mandelbrot et al. (1997). In the multi-fractal model of asset returns (MMAR) by Mandelbrot et al. (1997) and Calvet and Fisher (2002a), returns $x(t)$ are assumed to follow a compound process:

$$x(t) = B_H[\theta(t)] \tag{1}$$

in which an incremental fractional Brownian motion with index H, $B_H[\cdot]$, is subordinate to the cumulative distribution function $\theta(t)$ of a multi-fractal measure. Imposing the restriction $H = 0.5$, the MMAR would share the martingale property of most standard asset pricing models.

The main attraction of multifractals in the financial context is that they share certain properties which are known to be universal characteristics of asset returns as well: they have hyperbolically decaying autocovariances (long memory) and fat tails. Multifractility, furthermore, implies that different powers of the measure have different decay rates of their autocovariances. Calvet and Fisher (2002a) show that this feature carries over to absolute moments of returns in the MMAR (eq. 1). This behavior is in harmony with empirical findings of different degrees of long-term dependence in various powers of returns (Ding, Engle and Granger, 1993; Lobato and Savin, 1998) while alternatives like FIGARCH or long-memory stochastic volatility models are of a unifractal nature and are, therefore, unable to account for Ding et al.'s findings, i.e. they have the same decay rate for all moments.[2]

Despite the attractiveness of its stochastic properties, practical applicability of the MMAR suffers from its combinatorial nature and its non-stationarity due to the restriction to a bounded interval. These limitations have been overcome by the analogous iterative time series models introduced by Calvet and Fisher (1999, 2001, 2004). They consider a continuous-time multi-fractal model with random times for the changes of multipliers (Poisson multi-fractal) and demonstrate weak convergence of a discretized version of this process to its continuous-time limit.

This approach preserves the hierarchy of volatility components of MMAR but dispenses with its restriction to a bounded interval. In the discretized version of the Poisson multi-fractal, the volatility dynamics can be interpreted as a Markov-switching process with a large number of states. As long as the state space from which volatility components are drawn is finite maximum likelihood can be used for parameter estimation and Bayesian probability updating allows forecasting of future volatility. Forecasting algorithms developed for this model have been successfully applied for forecasting exchange rate volatility in Calvet and Fisher (2004).

In the following, we shortly review the building blocks of this Markov-

[2]The phenomenology of multifractality has attracted widespread attention in the recently emerging 'econophysics' literature (cf. Vassilicos et al., 1993; Vandewalle and Ausloos, 1998; Schmitt et al., 1999). While different statistical techniques are used, results are broadly equivalent to those first reported by Ding, Engle and Granger (1993) in the economics literature.

Switching Multi-Fractal process (MSM). Returns are modeled as:

$$x_t = \sigma_t \cdot u_t \tag{2}$$

with innovations u_t drawn from a standard Normal distribution N(0,1) and instantaneous volatility being determined by the product of k volatility components or multipliers $M_t^{(1)}$, $M_t^{(2)}$ ..., $M_t^{(k)}$ and a constant scale factor σ :

$$\sigma_t^2 = \sigma^2 \prod_{i=1}^{k} M_t^{(i)} \tag{3}$$

Each volatility component is renewed at time t with probability γ_i depending on its rank within the hierarchy of multipliers and remains unchanged with probability $1 - \gamma_k$. The transition probabilities are specified by Calvet and Fisher (2004) as:

$$\gamma_i = 1 - (1 - \gamma_1)^{(b^{k-1})} \tag{4}$$

with parameters $\gamma_1 \in [0, 1]$ and $b \in (1, \infty)$. Estimation of this model, then, involves the parameters γ_1 and b as well as those characterizing the distribution of the components $M_t^{(i)}$. Calvet and Fisher (2004) assume a Binominal distribution with parameters m_0 and $2 - m_0$ (thus, guaranteeing an expectation of unity for all $M_t^{(i)}$). The model, then, is a Markov switching process with 2^k states whose parameters can be estimated via maximum likelihood. Estimates can be used to construct forecasts of future volatility and these are shown to outperform forecasts derived from GARCH, FIGARCH, and two-state Markov switching GARCH models. Applicability of this approach is, however, limited in two respects: First, it is not applicable for models with an infinite state space, i.e. continuous distributions of the multipliers. For example, one could not use maximum likelihood for estimating an iterative analogue of the Lognormal multifractal model of Mandelbrot et al. (1997). Second, current computational limitations make choices of k beyond 10 unfeasible for the Binominal case because of the implied evaluation of a $2^k \times 2^k$ transition matrix in each iteration.

Given the successful performance of the particular version of MSM investigated in Calvet and Fisher (2004), it appears certainly worthwhile to explore whether one would get similar or better results with more flexible and general specifications. To this end, we introduce a GMM estimator as a flexible and versatile estimation method for multi-fractal parameters and compare its performance to that of ML (where applicable), and we complement GMM estimation by linear forecasting on the base of estimated parameter values which can be used as a substitute for Calvet and Fisher's Bayesian forecasts if conditional state probabilities are not available. Both approaches are applicable to a wide variety of specifications of (2) and (3). Here we will focus on increasing k beyond the computational limitations of ML estimation and on replacing the Binominal distribution of multipliers by a Lognormal one. In order to concentrate on these

particular extensions, we restrict the specification of transition probabilities by fixing them ex ante as:

$$\gamma_i = 2^{-(k-i)} \tag{5}$$

which leads to an even closer iterative analogue of the Binominal and Log-normal multi-fractal models introduced by Mandelbrot et. al. (1997) and Calvet and Fisher (2002a). Note that this specification remains very close to that of eq. (4) which, for small k, is approximately the same as: $\gamma_i = \gamma_1 b^{i-1}$. The same is obtained in our specification for the choice $b = 2$ together with $\gamma_1 = 2^{-(k-1)}$. This specification also implies that replacement happens with certainty for the component of highest frequency, $\gamma_k = 1$, and it imposes a structure of replacement probabilities in which the next lower component is renewed twice as often (on average) as its predecessor on the next higher level. Note that the high-frequency component at $i = k$ could also be interpreted as a contribution to the innovation in (2) rather than a part of the local volatility process and would, then, be equivalent to the choice of a fat-tailed conditional distribution.[3]

3 Estimation of Markov-Switching Multifractal Models via GMM

Estimation of multi-fractal models has proceeded from adaption of the so-called "scaling estimator" in Mandelbrot et al. (1997) – which is still pervasively used in natural science – to the seminal proposal of maximum likelihood estimation for the discretised MSM in Calvet and Fisher (2004). MLE is optimal in those cases in which it can be applied and has the added advantage of providing conditional probabilities of the unobserved volatility states which can be exploited for optimal forecasts on the base of the transition matrix of the model.

In their application to a Binomial Markov-Switching model, Calvet and Fisher also demonstrate that volatility forecasts over various time horizons derived from this model (from one day to fifty days ahead) improve upon forecasts from GARCH, two-state Markov Switching GARCH and FIGARCH models. While ML estimates and forecasts derived from identification of conditional state probabilities are optimal, their applicability is, however, restricted to MSM models with discrete state space. This excludes interesting classes like the Log-normal model with a continuous distribution of volatility components. From a practical point, the computational burden of a $b^k \times b^k$ tansition matrix (b: the number of states, k: the number of volatility components) would also make MLE unfeasible for multi-nomial models ($b > 2$) even with a relatively small number of components. Current technology would, therefore, restrict MLE to models with $k \leq 10$ in the case of a Binomial distribution ($b = 2$) and even smaller numbers for multinomial specification with a larger number of states.

[3]One could, in fact, go further and assume additional stages of the cascade at non-observable frequencies beyond the one at which data are available. Any assumption on the number of unobservable submerged components could be used to compute their expected contribution to the marginal distribution at the frequency of available data.

It, therefore, appears worthwhile to explore alternative ways of estimating MSM parameters and compare their efficiency (where possible) to that of the MLE benchmark.

To this end, we adopt a GMM (Generalized Method of Moments, cf. Hansen, 1982) approach using analytical solutions for a set of appropriate moment conditions. In GMM, the vector of parameter estimates of a model, say φ, is obtained as:

$$\widehat{\varphi}_T = \arg\min_{\varphi\in\Phi} f_T(\varphi)' A_T f_T(\varphi) \tag{6}$$

with Φ the parameter space, $f_T(\varphi)$ the vector of differences between sample moments and analytical moments, and A_T a positive definite and possibly random weighting matrix. As is well-known, $\hat{\varphi}_T$ is consistent and asymptotically Normal if suitable 'regularity conditions' are fulfilled (sets of which are detailed, for example, in Harris and Mátyás, 1999). $\hat{\varphi}_T$ then converges to

$$T^{1/2}(\hat{\varphi}_T - \varphi_0) \sim N(0, \Xi) \tag{7}$$

with covariance matrix $\Xi = (\bar{F}_T' \bar{V}_T^{-1} \bar{F}_T)^{-1}$ in which φ_0 is the true parameter vector, $\hat{V}_T^{-1} = T var f_T(\hat{\varphi}_T)$ is the covariance matrix of the moment conditions, $\hat{F}_T(\varphi) = \frac{\partial f_T(\varphi)}{\partial \varphi}$ 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.

Applicability of GMM would have been cumbersome for the MMAR approach of Mandelbrot et al. (1997) and Calvet and Fisher (2002a) because of its non-stationarity violating the required regularity conditions. This problem, fortunately, does not carry over to MSM models which rather fit into the class of Markov switching models with standard asymptotic behavior (cf. Calvet and Fisher, 2001). 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 (cf. Beran, 1994). MSMs 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. In the case of Markov-Switching multi-fractal processes, therefore,

$$Cov(|x_t^q|, |x_{t+\tau}^q|) \propto \tau^{2d(q)-1} \tag{8}$$

holds only over an interval $1 \ll \tau \ll b^k$ with b,k defined as above.

Although applicability of regularity conditions is not hampered by this type of "long memory on a bounded interval", the proximity to 'true' long memory might rise practical concerns. For example, note that with $b = 2$ and $k = 15$, the extent of the power law scaling might exceed the size of most available data for daily financial prices. In finite samples, application of GMM to Markov-

Switching multifractals could, then, yield poor results since usual estimates of the covariance matrix V_T might show large pre-asymptotic variation.[4]

Our practical solution to this potential problem is using log differences of absolute returns together with the pertinent analytical moment conditions, i.e. to transform the observed data x_t into:

$$\xi_{t,T} = \ln|x_t| - \ln|x_{t-T}|. \tag{9}$$

As is shown in Appendix A, the transformed variable $\xi_{t,T}$, in fact, only has non-zero auto-covariances at the first lag. One may note that by using moments of $\xi_{t,T}$, GMM only allows to estimate the parameters of the volatility process while the standard deviation of the Normally distributed increments, σ, drops out when computing log differences:

$$\xi_{t,T} = \ln\left|\sigma u_t(\prod_{i=1}^{k} M_t^{(i)})^{\frac{1}{2}}\right| - \ln\left|\sigma u_{t-T}(\prod_{i=1}^{k} M_{t-T}^{(i)})^{\frac{1}{2}}\right|$$

$$= 0.5\sum_{i=1}^{k}(\varepsilon_t^{(i)} - \varepsilon_{t-T}^{(i)}) + \ln|u_t| - \ln|u_{t-T}|$$

with $\varepsilon_t^{(i)} = \ln(M_t^{(i)})$.

The lack of inclusion of σ in the parameter set estimated by GMM does, however, not really pose a problem since as we show below, we can use the sample standard deviation as an estimation of σ which is only marginally less efficient than the pertinent ML estimates.

In order to exploit the temporal scaling properties of the multi-fractal model, our GMM estimator uses moment conditions providing information over various time horizons. In particular, we select covariances of the powers of $\xi_{t,T}$, i.e. moments of the following type:

$$Mom(T,q) = E\left[\xi_{t+T,T}^{q} \cdot \xi_{t,T}^{q}\right] \tag{10}$$

for $q = 1, 2$ and $T = 1, 5, 10, 20$.

We proceed by reporting results of several Monte Carlo studies designed to explore the performance of our new GMM estimator. Due to the moderate computational demands of GMM, we were able to use an iterative GMM scheme in which a new weighting matrix is computed and the whole estimation process is repeated until convergence of both the parameter estimates and the weighting matrix is obtained (cf. Hansen, Heaton and Yaron, 1996). We start with the Binomial Model with a limited number of multipliers, $k = 8$, and subsequently

[4]Poor results of GMM estimates in previous versions of this paper and related SMM (simulated model of moments) estimations in Calvet and Fisher (2002b) might have to be explained by this approximately hyperbolic behavior of autocovariances.

increase the number of volatility components and also switch to the Lognormal model as an example of a specification with continuous state space.[5]

Our first experiments serve to establish the efficiency of the GMM estimates vis-à-vis the ML approach of Calvet and Fisher (2004) for a setting that is close to the Monte Carlo experiments reported in their paper. To this end, we choose multipliers $m_0 = 1.3$, 1.4 and 1.5 and sample sizes $T_1 = 2500$, $T_2 = 5000$, and $T_3 = 10000$. The only difference of our simulation set-up is that we fixed the transition probabilities as $2^{-(k-i)}$ so that we only have to estimate two parameters compared to four in Calvet and Fisher's somewhat more general approach (cf. eq. 4). Results are displayed in Table 1. Comparison with the pertinent table in Calvet and Fisher (2004) shows that in the more parsimonious two-parameter model both m_0 and σ can be estimated somewhat more efficiently than with four parameters. Furthermore, Table 1 indicates that biases and root mean squared errors of the estimates of m_0 are hardly affected at all by the choice of parameters, while RMSEs appear to increase for σ for higher underlying parameter m_0. This is plausible since with increasing m_0 one generates enhanced fluctuations of the product of volatilities which interfers with the estimation of the constant scale factor.

Table 1 about here

Comparison of MLE and GMM estimates shows, of course, that the later are less efficient.[6] For reasons given above GMM provides estimates of the multifractal parameter m_0 only. Instead of a GMM estimate of σ we, therefore, report the statistics of elementary estimates of σ from the sample standard deviation in Table 1. This mainly serves to show that the efficiency of the σ estimates from MLE is pretty much identical to that of the sample standard deviation so that if needed we could use the later without much loss of efficieny.

More interesting is comparison of the estimated m_0's. Obviously, variability of estimates with GMM is much higher, ranging from 2.5 to 6 times that of the ML estimates. The relative difference is the higher, the smaller the sample size

[5]Earlier versions of this paper also included comparisons with the traditional "scaling estimator" as well as comparisons of GMM with varying numbers of moment conditions. These results have been skipped to preserve space. As concerns the scaling estimator, its performance typically is far worse than GMM, let alone MLE. Since the scaling estimator also exploits the structure of various moments of the data, its poor performance can be explained by inefficient use of information contained in moments compared to GMM. Calvet and Fisher (2002b), in fact, deviced a "simulated method of moment" adaption of the original scaling approach which should make more efficient use of the pertinent moments but was also discarded in their subsequent development of ML techniques. Experiments with varying numbers of moments in our GMM approach indicated monotonic behavior with a steady reduction of simulated mean squared errors when increasing the number of moments. Because of decreasing returns in the gains in precision, using eight moment conditions appeared to yield a satisfactory balance between computational speed and the quality of the estimates. One might, however, keep in mind that the efficiency of GMM could still be somewhat increased at resonable computational costs.

[6]However, one might note that obtaining our GMM estimates with eight moment conditions requires only a small fraction of the time needed for ML estimation.

and the lower the 'true' value m_0. While biases and MSEs of ML estimates of m_0 were essentially independent of the true parameter, GMM estimates show a reduction in MSEs by about fifty percent when proceeding from $m_0 = 1.3$ to $m_0 = 1.5$. Interestingly, the average bias of the Monte Carlo estimates is moderate throughout and practically zero for the larger sample sizes.[7] It is also worthwhile to point out that the GMM estimator is quite well-behaved in that we encounter no problems of non-convergence or breakdown of the estimation in all our Monte Carlo simulations. This is in contrast to GMM estimation of standard stochastic volatility models which are plagued by a non-negligible frequency of 'crashes', cf. Andersen and Sørensen (1996).

Now, we proceed into territory in which MLE becomes unpractical, at least for the purpose of simulation studies. Table 2 shows GMM results for the same set of parameters but with increasing number of cascade components, k=10, 15 and 20. Here we restrict ourselves to only one sample size, $T = 5000$, in order to conserve space. Comparison with the pertinent entries in Table 1 shows that the efficieny of multi-fractal parameter estimates is practically insensitive with respect to the number of components. Inspection of our moment conditions detailed in Appendix A reveals that this is probably so because high-level multipliers are expected to change only very infrequently and, therefore, would only contribute small increments to log differences. On the other hand, these nearly constant entries should make estimation of the scale factor σ more cumbersome as it would be hard to distinguish between long-lived high-level multipliers and an entirely constant factor. This ambiguity is clearly reflected in the blow-up of the FSSEs and RMSEs for σ with increasing k. However, the almost complete insensitivity of the estimates of m_0 with respect to the number of components might be viewed as a very welcome feature as it implies that estimation of m_0 is hardly affected by potential misspecification of k.

Table 2 about here

In the next step, we consider one particular variant of a multi-fractal process, for which MLE is not applicable at all because of its continuous state space of volatility components. In pertinent literature, the most popular continuous model is the Lognormal, cf. Mandelbrot et al. (1974) and Calvet and Fisher (2002a). In the spirit of these seminal contributions, we now specify the multipliers to be random draws from a Lognormal distribution with parameters λ and s, i.e.

$$M_t^{(i)} \sim LN(-\lambda, s^2) \tag{11}$$

Normalisation via $E[M_t^{(i)}] = 1$ leads to

$$exp(-\lambda + 0.5s^2) = 1, \tag{12}$$

[7] In view of our concerns about the proximity of the MSM with its 'long memory on a bounded interval' to processes with pure long-term dependence, it is also interesting to note that our estimates are in harmony with $T^{\frac{1}{2}}$ consistency. This undercores that the log transformation is useful in improving the quality of GMM estimates as compared to earlier experiments using moments of the raw data.

from which a restriction on the shape parameter can be inferred: $s = \sqrt{2\lambda}$. Hence, we end up with a one-parameter family of multi-fractal models as in the Binomial case. Moment conditions for this model are spelled out in Appendix B. Note that the admissible parameter space for the location papameter λ is $\lambda \in [0, \infty)$ where in the borderline case $\lambda = 0$ the volatility process collapses to a constant (the same if $m_0 = 1$ in the Binomial model). Simulations indicate that increasing λ leads to increasing heterogenity in volatility with $0 < \lambda < 0.2$ giving roughly realistic appearances of the resulting time series. In our Monte Carlo simulations reported in Table 3 we cover this interval by considering $\lambda = 0.05$, 0.10 and 0.15. Again, we only use one sample size, $T = 5,000$ and numbers of multipliers k equal to 8, 10, 15 and 20. As can be seen, results are not too different from those obtained with the Binomial model. Biases are moderate again, and results for λ are almost insensitive with respect to k, but RMSEs for σ (again obtained as the sample standard deviation) increase monotonically with k. All in all, the picture from both the Binomial and Lognormal Monte Carlo runs shows that GMM seems to work as well in the continuous case as with a discrete distribution of volatility components.

Table 3 about here

4 Best Linear vs. Optimal Forecasts

ML estimation comes along with identification of conditional probabilities of the current states of the volatility components. Together with the transition matrix of the model, these conditional probabilities can be uses to compute one-step and multi-step forecasts according to Bayes' rule. Since ML is restricted to discrete distributions, this elegant and optimal way of generating forecasts for multifractal volatility is also restricted to the multi-nominal cases with a finite state space. Again, it would be useful to have methods at hand for specifications beyond the confines of multi-nominal models. One alternative for cases in which MLE is not applicable is to use best linear forecasts. Since these do not require state probabilities as an input, they could be computed also on the base of GMM parameter estimates. The standard approach for construction of linear forecasts is outlined, for example, in Brockwell and Davis (1991, c.5). Assuming that the data of interest follow a stationary process $\{X_t\}$ with mean zero, the best linear h-step forecasts are obtained as

$$\widehat{X}_{n+h} = \sum_{i=1}^{n} \phi_{ni}^{(h)} X_{n+1-i} = \boldsymbol{\phi}_n^{(h)} \mathbf{X}_n \tag{13}$$

with the vectors of weights $\boldsymbol{\phi}_{\ n}^{(h)} = (\phi_{n1}^{(h)}, \phi_{n2}^{(h)}, ..., \phi_{nn}^{(h)})'$ being any solution of $\boldsymbol{\Gamma}_n \boldsymbol{\phi}_{\ n}^{(h)} = \boldsymbol{\gamma}_{\ n}^{(h)}$ and $\boldsymbol{\gamma}_{\ n}^{(h)} = (\gamma(h), \gamma(h+1), ..., \gamma(n+h-1))'$ denoting the autocovariances for the data-generating process of X_t at lags h and beyond,

and $\mathbf{\Gamma}_n = [\gamma(i-j)]_{i,j=1,\ldots,n}$ the pertinent variance-covariance matrix. It is well-known, that this approach produces the best linear estimators under the criterion of minimization of mean squared errors. It is also well known, that for long-memory processes, one should use as much information as available, i.e. the vector $\mathbf{X}_n$ should contain all past realisations of the process. One might argue that for the MSM, its 'long memory on a bounded interval' would lead to an optimal choice of a number of 2^k past observations (as auto-covariances would rapidly drop to zero thereafter). However, in our application we simply use all available data as with a 'true' long-memory process altough very long lags might have no practical influence on the resulting forecasts. The computational demands of these predictors is immensely reduced by using the generalized Levinson-Durbin algorithm developed recently by Brockwell and Dahlhaus (2004, in particular their algorithm 6).

What is needed to implement linear forecasts is analytical solutions for the auto-covariances of the quantity one whishes to predict. In our case, our aim is to predict squared returns, x_t^2, as a proxy of volatility which requires analytical solutions for $E[x_{t+n}^2 x_t^2]$. These are also given for the Binominal and Lognormal models in Appendices A and B, respectively. Implementing (13), we have to consider the zero-mean time series:

$$X_t = x_t^2 - E[x_t^2] = x_t^2 - \widehat{\sigma}^2 \tag{14}$$

where $\widehat{\sigma}$ is the sample standard deviation of the time series under consideration (which as we have shown above, seems to be almost as efficient an estimator as the ML estimate of σ).[8]

Again, our aim is to first explore how much is lost by using linear instead of optimal forecasts and, then, to investigate the performance of linear forecasts in cases in which optimal forecasting is infeasible or unpractical.

Our first example parallels the comparison of MLE and GMM for the Binominal model with $k = 8$ in Table 1. To conserve space, we restrict ourselves to one sample size $T = 10,000$ using half of the data for in-sample parameter estimation and the remainder for assessment of the out-of-sample forecasting performance in terms of mean-squared error (MSE) and mean absolute error (MAE). Including MAEs seems interesting since the best linear estimators are those among all linear forecasts which minimise mean squared errors. It, therefore, appears worthwhile to also explore their performance with respect to a different criterion. Both MSE and MAE are standardized relative to the MSE and MAE of the most naive forecast, i.e. $\widehat{\sigma}^2$ or 'historical volatility' during the in-sample period, for the same sample, so that values below 1 indicate an improvement against the constant variance forecast by the pertinent model. Three different forecasting procedures are shown in Table 4: "ML" uses Bayesian updating together with ML parameter estimates as detailed in Calvet and Fisher (2004), "BL1" uses best linear forecasts on the base of GMM parameter estimates, and

[8] Note that $\widehat{\sigma}$ only appears in the mean value of eq. (14), but it drops from the coefficients $\phi_{ni}^{(h)}$ where by construction it enters in both the denominator and the numerator so that it cancels out.

"BL2" uses best linear forecasts together with ML parameter estimates. The later variant has been added to see how much of a potential loss of efficiency might be due to the use of linear forecasts vs. Bayesian updating and how much to GMM vs. ML. However, as it turned out, we never observe much of loss of efficiency anyway from using either BL1 or BL2.

As can be seen from the average MSEs and MAEs in Table 4, as expected, ML mostly comes out as the most efficient method but its advantage against BL1 and BL2 is tiny, with the later on average reaching about 99 percent of the efficiency of ML for most parameters and forecasting horizons. We also illustrate the full distribution of Monte Carlo MSEs for one representative case ($m_0 = 1.3$) in Fig. 1. As can be seen from the box-plots, the entire distribution of MSEs seems to be quite similar for our three different approaches.

The difference between ML based Bayesian forecasting and linear forecasting is sligthly higher with the MAE criterion at short lags. On the other hand, it also appears worth mentioning that ML seems to come along with higher variability than BL1 and BL2 under MAE comparison so that one would have to compare higher average gain with higher dispersion of forecasting quality when choosing between these three methods. Table 1 also indicates that forecast quality depends on the multi-fractal parameter m_0: increasing m_0 from 1.3 to 1.4, we observe decreasing MSEs and particularly MAEs. This might be a consequence of the more pronounced volatility clustering with higher m_0. Note, however, that this trend continues for MAEs when proceeding to $m_0 = 1.5$, while MSEs rather deteriorate slightly.

Table 4 and Figure 1 about here

In parallel with our investigation of GMM estimation in sec. 3, we now extend the scope of our Monte Carlo analysis to cases that can not (or only at prohibitive computational cost) be dealed with on the base of ML based methods. To this end, tables 5 and 6 show results for the Binominal model with numbers of volatility components $k = 10$, 15 and 20 and similar experiments for the continuous parameter case of the Lognormal MSM. In both cases, we only have one method available, GMM estimation together with linear forecasts (the former BL1).

The basic message of Table 5 is that with a higher number of volatility components, squared returns have a larger predictable component. For example, for one-day horizons and $m_0 = 1.3$, the predictable component (the decrease of our relative MSEs) increases from 5.3 percent at $k = 8$ to 15.6 percent with $k = 20$. For the twenty day horizon we observe an increase from 1.5 percent ($k = 8$) which probably is irrelevant in an applied context to 11 percent ($k = 20$) which might be more interesting. We have also included long time horizons of fifty and one hundred lags which still have predictable components of about 8 to 9 percent of MSE. Results are fairly homogeneous across multi-fractal parameters as well as for both the MSE and MAE criterion, but forecastability is higher throughout with higher values of m_0. It is particularly worth emphasizing that we have kept the in-sample period constant at $T = 5000$ in all these

experiments. This means that the information used to estimate the parameters has *not* been increased with k. One might also note that the increase in biases and estimation variability of the constant component σ with increasing k (cf. Tables 2 and 3) does apparently not restrict predictability in any way.

Pretty similar results are obtained with the Lognormal specification (cf. Table 6). Gains in predictability are about the same as with the Binominal model for increasing numbers of volatility components. Interestingly, the variation of MSE and MAE for the Lognormal model is often non-monotonic under variation of λ. In particular, we observe a rather pronounced U shape in MSEs with forecastability improving for $\lambda = 0.10$ against $\lambda = 0.05$ but decreasing again for $\lambda = 0.15$. We conjecture that this variability is due to the lower degree of homogenity in the volatility clustering of the Lognormal process in which multipliers are drawn randomly from an infinite support. Higher λ also leads to a higher dispersion of the Lognormal variates via $s = \sqrt{2\lambda}$ so that the volatility of volatility increases, which might explain the deterioration of forecastability for higher λ.

Table 5 and 6 about here

5 Empirical Evidence

We now turn to empirical data in order to see whether using GMM plus linear forecasting helps improve on previous ML estimates together with optimal forecasts. Potential gains could result from the accessability of richer specifications, i.e. allowing for additional multipliers beyond the constraints of about k $\leq$ 10 and from using distributions with continuous rather than discrete state space for the multipliers.

Our empirical analysis uses data from four different financial markets: the New York Stock Exchange composite index, the German Stock Price Index DAX, the US$ - Deutsche Mark Exchange rate and the price of gold. Stock market series were obtained from the New York and Frankfurt Stock Exchanges, while the exchange rate and precious metal series were provided by the financial database at the University of Bonn. Our samples cover twenty years starting from 1 January 1979 and ending on 31 December 1998. We use the data of the years 1979 to 1996 for in-sample estimation and leave the two remaining years for out-of-sample evaluation of volatility forecasts. This gives about 4,400 in-sample observations and 500 out-of sample entries (with slight variations of the numbers across markets due to diffences in the number of active days).

Table 6 reports in-sample parameter estimates for the multi-fractal parameters m_0 or λ from both the Binomial and Lognormal specification. For the sake of brevity we skip the second parameter σ in eq. (3) as there is less genuine interest in this scale factor. For the Binomial model, results based on ML estimation with $k = 10$ as well as GMM with $k = 5, 10, 15$ and 20 are shown, whereas the lack of applicability of ML estimation only leaves us with the later

choice in the case of the Lognormal model. However, for the German DAX, the estimation of multi-fractal parameters appeared to be somewhat cumbersome anyway as both for the Binomial and Lognormal model, estimates from our standard setting with eight moments got stuck at the lower bounds of the parameter space ($m_0 = 1$ and $\lambda = 0$, respectively). Since this would have amounted to Normally distributed returns without temporal dependence, we re-estimated the multi-fractal parameters of the DAX with moments defined according to (10), but using $T = 1, 5, 10, 20, 50, 100$ and 200, i.e. six more moment conditions with auto-covariances extending over longer time lags which produced non-trivial estimates.

Inspecting the results, a number of observations are remarkable: (i) in the case of the Binomial Model, ML and GMM estimates are mostly very similar, the only exeption being the results of the DAX. (ii) for the GMM approach, results obtained for different numbers of multipliers are virtually identical for both the Binomial and Lognormal model. In fact, if one monitors the development of estimated parameters m_0 and λ with increasing k, one finds strong variation initially with a pronounced decrease of the estimates which becomes slower and slower until, eventually, a lock-in at a constant value can be found somewhere between $k = 5$ and $k = 10$. As can be seen from Table 7 estimates still undergo slight variations when proceeding from $k = 5$ to 10 but they remain entirely unchanged thereafter. It is not too difficult to see why this happens: inspection of moment conditions in the Appendix (e.g. eq. (A7) and (A8) for the Binomial model, and (B3) and (B4) for the Lognormal model) shows that higher multipliers make a smaller and smaller contribution to the moments so that their numerical values would stay almost constant. This small influence of additional elements in the hierarchy of multipliers also suggests that it should be hard or impossible to distinguish between multi-fractal models with different numbers of components once k increases beyond a certain threshold.

Table 7 also shows the probability of Hansen's test statistics $J_T = f_T(\hat{\varphi})' \hat{A}_T f_T(\hat{\varphi})$, which also does not change at all beyond $k = 10$. This observation also suggests that the similarity of different models with respect to their moments hampers model selection on the base of the objective function. We tried model comparisons along the line of the SMM approach in a previous version of Calvet al Fisher (2002b) combining the minimized moment functions with simulated weighting matrices for varying numbers k', but results were practically constant across all choices of k and k' beyond some threshold. It is plausible that the lack of sensibility of moments on k beyond a certain value would carry over to the weighting matrices as well so that the discriminatory power of such comparisons should be extremely limited. We would expect a similar pattern to apply to likelihood ratio tests for comparision of high k and k' even if the true k is the larger one if these tests were computationally feasible.[9] Model selection among competing specifications of multi-fractal model

[9] Calvet and Fisher (2004) report results of a likelihood ratio test of models with 10 components against models with $k = 1$ to 9 multipliers. Though they report monotonically increasing p - values, at the 5 percent level only models with up to 4 or 5 components can be rejected for various daily exchange rate series.

with different numbers of multipliers, therefore, remains a challenging task. In view of the undecisive results, we choose to take an agnostic approach and to investigate the performance of various specifications in predicting future volatility.

Table 7 about here

As another remarkable finding in Table 7, the similarity of the probability of the J statistic (and therefore, also the optimized value of this statistic) for the Binomial and Lognormal model stands out. The numbers are, in fact, perfectly identical for the NYCI, the DAX and the USD-DEM for $k \geq 15$, and are also very close for the price of Gold. This shows that both the Binomial and Lognormal can fit our selection of moments equally well. Therefore, allowing for a continuous distribution of multipliers seems not to improve upon the performance of the discrete case with two states only, where 2^k possible combinations, then, seem to provide enough flexibility for capturing the heterogenous volatility dynamics.

Parsimony of model design might suggest to restrict oneself to a relatively small number of multipliers, given the inconclusive results of specification tests for high values k. Note, however, that increasing k does not come along with additional parameters. On the other hand, higher k implies a larger region of apparent long memory. In empirical data, hyperbolic decline of auto-correlation of absolute and squared returns is observed over many orders of magnitude without any apparent cut-off and significantly positive auto-correlations have been found at extremely long lags. For the daily S&P 500 returns, Ding, Granger and Engle (1993) found significant autocorrelations at over 2700 lags, i.e. something like 10 years. Since higher k implies a longer power-law range of the autocorrelations, the longer dependence in volatility might improve forcasting performance. As has been shown in section 3, with the same number of in-sample observations, mean squared errors and absolute errors decline at all forecasting horizons for increasing k while the quality of the estimates of the multi-fractal parameters m_0 and λ remains essentially unchanged. These findings provide the perspective that higher choices of k could also improve volatility forecasts for empirical data. Tables 8 and 9 explore this issue for our selection of financial market data. For all four time series we compare forecasts over horizons varying from one day over 5, 10, 20, 50 to 100 days. The models we use are: GARCH and FIGARCH as benchmarks from the traditional time series literature, the Binomial MSM with $k = 10$ estimated by ML as well as the Binomial and Lognormal MSM with $k = 10$, 15 and 20 estimated by GMM.[10]MSEs and MAEs are again reported relative to those of historical volatility.

[10]Estimated GARCH and FIGARCH models can be found in Appendix C. Despite certain recently emphasized ambiguities in the parametrization of FIGARCH models (cf. Chang, 2002; Zumbach, 2004), we stick to the original framework by Baillie et al. (1996) in order to compare the multi-fractal model with a well-established benchmark. It is, however, worth printing out that evidence on the forecasting improvement of FIGARCH vis-à-vis simple GARCH is surprisingly sparse. Vilasuso (2002) and Zumbach (2004) seem to be the only available references on the subject and have somewhat divergent findings on this point.

As it turns out, results are somewhat different for the two stock indices, on the one hand, and the exchange rate and precious metal price, on the other hand. Starting with the bad news: for both stock indices, results are not too exciting. For the DAX and NYCI, little variation of MSEs and MAEs is actually obtained across all models. While the benchmark multi-fractal estimates obtained via maximum likelihood are better than GARCH or FIGARCH at some horizons, no additional gains are obtained from increasing k. Furthermore, all time series models are worse than the sample variance under the MAE criterion. In retrospect, this similar and relatively poor performance of all competing models might have to be explained by the huge increase of volatility in the out-of-sample period 1997/98 which is probably hard to crasp for any time-series approach.[11]

Tables 8 and 9 about here

The picture is totally different for the two remaining series. Here the multifractal models typically beat GARCH and FIGARCH with an increasing advantage for longer horizons (confirming similar results by Calvet and Fisher, 2004). Furthermore, increasing k beyond 10 levels yields further improvements of forecast quality with GMM estimates at $k = 15$ and 20 providing the lowest relative MSEs and MAEs over all time horizons. Interestingly, comparison of the results obtained with ML and optimal forecasts versus GMM and linear forecasts at $k = 10$ reveals that the latter are mostly slightly worse. While this is to be expected from the higher sampling variability of GMM estimates and the suboptimality of the linear forecasts, increasing k appears to amount to an overcompensation of this drawback. As can be seen from Tables 8 and 9, many of the linear forecasts at $k = 15$ or 20 are even significantly better at the 99 percent level than their counterparts at $k = 10$ (and almost all are significant at least at the 95 percent level). The effect of the number of multipliers on forecast quality is underscored by Fig. 2 which depicts MSEs for the price of Gold obtained with linear forecasts on the base of ten multifractal models with k ranging from 2 to 20. As can be seen, increasing k yields a monotonic improvement which saturates at about $k \approx 15$. The same behavior can be found for the exchange rate USD-DEM. Lastly, it seems worthwile to remark that the Lognormal model has almost the same performance as the Binomial for each choice of k so that the added flexibility of the continuous distribution does not seem to provide an advantage over the simpler discrete structure which is in harmony with their identical goodness-of-fit under the J statistic.

Fig. 2 about here

[11] In the framework of the multi-fractal model, it might amount to the change of a high-level multiplier which might have been constant over much of the in-sample period.

6 Conclusion

The recent proposal of multi-fractal models has added a new family of stochastic models to the already abundant variety of candidate processes for financial returns. Its attractiveness stems from its pasimony which with a very limited number of parameters allows to capture the phenomenology of returns and their fluctuations (i.e. volatility persistence, hyperbolic decay of autocorrelations and heavy tails of returns). Some cumbersome properties of the multi-fractal approach inherited from the physics literature have meanwhile been overcome by the introduction of Markov-switching multifractals which have nice asymptotic properties and whose parameters can be estimated by standard econometric methods. While maximum likelihood estimation has been explored in Calvet and Fisher (2004) - and simulated ML via a particle filter algorithm in Calvet, Fisher, and Thompson (2004)-, I propose a GMM approach for estimating multi-fractal parameters, which is less computation intensive and which is still applicable in cases where ML becomes infeasible. Monte Carlo experiments show that our GMM estimator is nicely behaved for various settings with discrete and continous state space. Since Baysian forecasting becomes unfeasible with ML we combine our GMM approach with best linear forecasts instead. Monte Carlo comparisons show that the loss in forecasting accuracy of GMM plus linear forecasts compared to ML plus optimal forecasts is, in fact, quite small (the percentage deviation being much smaller than that of mean squared errors of estimated parameters). Extending the state space beyond what is practical with ML (and probably simulated ML as well) could, in principle, lead to higher forecasting accuracy through the increase of the predictable component of volatility with a higher number of multipliers.

Our empirical application shows, that at least for some time series, this promise indeed materializes itself. While we found a similar performance of GARCH, FIGARCH and various multi-fractal specifications for two stock indices, in the case of the USD-DEM exchange rate and the precious metal price (Gold), the multi-fractal model dominates both the GARCH model and its fractionally integrated version. While better performance of MF vis-*à*-vis (FI)GARCH had already been demonstrated for USD-DEM and other exchange rates by Calvet and Fisher (2004), we show that increasing the state space yields further improvments in forecasting accuracy. Our results for the Gold price also show that these are not restricted to foreign exchange markets alone. The somewhat less supportive evidence for stock indices might either point to differences in the stochastic nature of domestic asset markets or might be caused by the particularities of our out-of-sample period which contained an episode of sharply increasing volatility in the pertinent markets. In any case, the confirmation of the previous positive results and the further improvements by alternative specifications documented in this paper underline that multiplicative models with a hierarchy of volatility components are a promising area of empirical research.

References

Andersen, T.G. and T. Bollerslev (1997) Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long Run in High Frequency Data,*Journal of Finance*, **52**, 975-1005.

Andersen, T.G. and B.E. Sϕrensen (1996) GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study, *Journal of Business and Economics Statistics*, **14**, 328-352.

Baillie, R.T., T. Bollerslev and H.O. Mikkelsen (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **74**, 3-30.

Beran, J. (1994) *Statistics for Long-Memory Processes.* New York: Chapman & Hall.

Brockwell, P.J. and R. Dahlhaus (2004) Generalized Levinson-Durbin and Burg Algorithms, *Journal of Econometrics*, **118**, 129-149.

Brockwell, P.J. and R.A. Davis (1991) *Time Series: Theory and Methods.* New York, Springer.

Calvet, L. and A. Fisher (2001) Forecasting Multifractal Volatility, *Journal of Econometrics*, **105**, 27-58, working paper version: NYU Stern Working Paper FIN-99-017 (1999).

Calvet, L. and A. Fisher (2002a) Multi-Fractality in Asset Returns: Theory and Evidence, *Review of Economics & Statistics*, **84**, 381-406.

Calvet, L. and A. Fisher (2002b) *Regime Switching and the Estimation of Multifractal Processes*, Mimeo: Harvard University.

Calvet, L. and A. Fisher (2004) Regime Switching and the Estimation of Multifractal Processes,*Journal of Financial Econometrics*, **2**,49-83.

Calvet, L., A. Fisher and S. Thompson (2004) *Volatility Comovement: A Multi-Frequency Approach*, Mimeo: Harvard University.

Chung, C.-F. (2002) *Estimating the Fractionally Integrated GARCH Model.* Mimeo: Academia Sinica.

Diebold, F. and R. Mariano (1995) Comparing Predictive Accurancy, *Journal*

of Business and Economics Statistics, **13**, 253-263.

Ding, Z., C.W.J. Granger and R.F.Engle (1993) A Long Memory Property of Stock Market Returns and a New Model, *Journal of Empirical Finance*, **1**, 83-106.

Evertz, C.J.G. and B. Mandelbrot (1992) Multifractal Measures, in: Peitgen, H.-O., H. Jürgens and D. Saupe, eds., *Chaos and Fractals: New Frontiers of Science*, Berlin: Springer.

Hansen, L.P (1982) Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, **50**, 1029-1054.

Hansen, L.P., J. Heaton and A. Yaron (1996) Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business & Economics Statistics*, **14**, 262-280.

Harris, D. and L. Mátyás (1999) Introduction to the Generalized Method of Moments Estimation, Chap. 1 in: L. Mátyás, ed., *Generalized Method of Moments Estimation*, Cambridge: University Press.

Harte, D. (2001) Multifractals: Theory and Applications. London: Chapman and Hall.

Lobato, I.N. and N.E. Savin (1998) Real and Spurious Long-Memory Properties of Stock Market Data, *Journal of Business & Economics Statistics*, **16**, 261-283.

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., A. Fisher and L. Calvet (1997) *A Multifractal Model of Asset Returns*. Mimeo: Cowles Foundation for Research in Economics.

Poon, S.-H. and C. Granger (2003) Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature*, **41**, 468-539.

Schmitt, F., D. Schertzer and S. Lovejoy (1999), Multifractal Analysis of Foreign Exchange Data, *Applied Stochastic Models and Data Analysis*,**15**, 29-53.

Vandewalle, N. and M. Ausloos (1998) Multi-affine Analysis of Typical Currency Exchange Rates, *European Physical Journal B*, **4**, 257-261.

Vassilicos, J.C., A. Demos and F. Tata (1993) No Evidence of Chaos But Some Evidence of Multifractals in the Foreign Exchange and the Stock Market, in: Crilly, A.J., R.A Earnshaw and H. Jones, eds., *Applications of Fractals and Chaos*. Berlin: Springer.

Vilasuso, J. (2002) Forecasting Exchange Rate Volatility, *Economics Letters*, **76**, 59-64.

Zumbach, G. (2004) Volatility Process and Volatility Forecast with Long Memory, *Quantitative Finance*, **4**, 70-86.

Appendix

A Moments of Binomial Model

In order to apply GMM and to compare its performance with that of the ML approach, we have to compute closed-form solutions for selected moments of the Binomial model. As pointed out in the main text, we use a selection of moments of log increments of the raw data. Let

$$\mu_t = \prod_{i=1}^{k} M_t^{(i)} \tag{A1}$$

denote the volatility process and $\eta_{t,T}$ its log increments:

$$\eta_{t,T} = \ln(\mu_t) - \ln(\mu_{t-T}) = \sum_{i=1}^{k} \ln(M_t^{(i)}) - \sum_{i=1}^{k} \ln(M_{t-T}^{(i)}). \tag{A2}$$

It is readily apparent that $E[\eta_{t,T}] = 0$ for all T. Let us now consider the first auto-covariance of $\eta_{t,1}$:

$$E[\eta_{t+1,1}\eta_{t,1}] = E\left[\left(\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\right)\left(\sum_{j=1}^{k}\left(\varepsilon_t^{(j)} - \varepsilon_{t-1}^{(j)}\right)\right)\right] \tag{A3}$$

with $\varepsilon_t^{(i)} = \ln\left(M_t^{(i)}\right)$.

Because of independence of any pair of volatility components i and j, only summands with $i = j$ give non-zero contributions to the term on the right-hand side. Furthermore, it can easily be seen that these components, i.e.

$$E\left[\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right]$$

are themselves different from zero only if the relevant multiplier changes two times between $t-1$ and $t+1$.

In general,

$$\begin{aligned} &E\left[\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right] = \\ &E\left[\varepsilon_{t+1}^{(i)}\varepsilon_t^{(i)}\right] - E\left[\left(\varepsilon_t^{(i)}\right)^2\right] - E\left[\varepsilon_{t+1}^{(i)}\varepsilon_{t-1}^{(i)}\right] + E\left[\varepsilon_t^{(i)}\varepsilon_{t-1}^{(i)}\right]. \end{aligned} \tag{A4}$$

Since in order to get non-zero entries, we have to have: $\varepsilon_{t+1}^{(i)} = \varepsilon_{t-1}^{(i)} \neq \varepsilon_t^{(i)}$ and this happens with probability $(\frac{1}{2}\frac{1}{2^{k-i}})^2$, the above expression becomes:

$$\left(\frac{1}{2}\frac{1}{2^{k-i}}\right)^2 \left\{2\ln(m_0)\ln(2-m_0) - \ln^2(m_0) - \ln^2(2-m_0)\right\}. \tag{A5}$$

We, therefore, arrive at

$$E\left[\eta_{t+1,1}\eta_{t+1}\right] = \left\{2\ln\left(m_0\right)\ln\left(2-m_0\right) - \ln^2\left(m_0\right) - \ln^2\left(2-m_0\right)\right\} \cdot \sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\frac{1}{2}\right)^2. \quad \text{(A6)}$$

In passing, we note that autocovariance at higher lags $\tau > 1$, i.e.

$$E\left[\eta_{t+\tau,1}\eta_{t,1}\right] = E\left[\left(\sum_{i=1}^{k}\left(\varepsilon_{t+\tau}^{(i)} - \varepsilon_{t+\tau-1}^{(i)}\right)\right)\left(\sum_{i=1}^{k}\left(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right)\right]$$

are all equal to zero because of the independence of changes between $t-1$ and t and between $t+\tau-1$ and $t+\tau$ ($\tau > 1$), respectively.

Considering the autocovariances of the log changes over time intervals $T > 1$, we have to replace the probabilities for renewal after one time step by the pertinent probabilities for T steps which leads to:

$$E\left[\eta_{t+T,T}\eta_{t,T}\right] = \left\{2\ln\left(m_0\right)\ln\left(2-m_0\right) - \ln^2\left(m_0\right) - \ln^2\left(2-m_0\right)\right\} \cdot \sum_{i=1}^{k}\frac{1}{4}\left(1-\left(1-\frac{1}{2^{k-i}}\right)^T\right)^2. \quad \text{(A7)}$$

The calculations become slightly more involved when considering the autocovariances of squared log increments:

$$E\left[\eta_{t+T,T}^2\eta_{t,T}^2\right] = E\left[\left(\sum_{i=1}^{k}\left(\varepsilon_{t+T}^{(i)} - \varepsilon_{t}^{(i)}\right)\right)^2\left(\sum_{j=1}^{k}\left(\varepsilon_{t}^{(j)} - \varepsilon_{t-T}^{(j)}\right)\right)^2\right]. \quad \text{(A8)}$$

Again, one can arrive at a relatively simple formula by identifying the non-zero entries and their probabilities of occurrence. Let us start with T=1. Three cases are relevant here:

(1) $i = j$ and $\varepsilon_{t+1}^{(i)} \neq \varepsilon_{t}^{(i)} \neq \varepsilon_{t-1}^{(i)}$:

This leads to entries of the form:

$$\begin{aligned}(\varepsilon_{t+1}^{(i)} - \varepsilon_{t}^{(i)})^2(\varepsilon_{t}^{(i)} - \varepsilon_{t-1}^{(i)})^2 \\ &= \left(\ln\left(m_0\right) - \ln\left(2-m_0\right)\right)^2\left(\ln\left(2-m_0\right) - \ln\left(m_0\right)\right)^2 \\ &= \ln^4\left(m_0\right) + \ln^4\left(2-m_0\right) + 6\ln^2\left(m_0\right)\ln^2\left(2-m_0\right) \\ &\quad -4\ln^3\left(m_0\right)\ln\left(2-m_0\right) - 4\ln\left(m_0\right)\ln^3\left(2-m_0\right) \equiv \chi.\end{aligned}$$

Note that the relevant sequences $m_0 \to m_1 \to m_0$ or vice versa again happen with probabilities $(\frac{1}{2}\frac{1}{2^{k-i}})^2$.

(2) $i \neq j$ and $\varepsilon_{t+1}^{(j)} \neq \varepsilon_t^{(j)}$ and $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$.

It can easily be seen that for non-zero entries, this leads again to

$$\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2 = \left(\ln\left(m_0\right) - \ln\left(2 - m_0\right)\right)^4 = \chi$$

which is the same as in (1) and occurs for each pair $i \neq j$ with probabilities $(\frac{1}{2}\frac{1}{2^{k-i}})(\frac{1}{2}\frac{1}{2^{k-j}})$.

(3) $i \neq j$ and $\varepsilon_{t+1}^{(n)} \neq \varepsilon_t^{(n)} \neq \varepsilon_{t-1}^{(n)}$, $n = i, j$ in entries of the form

$(\varepsilon_{t+1}^j - \varepsilon_t^j)(\varepsilon_{t+1}^i - \varepsilon_t^i)(\varepsilon_t^j - \varepsilon_{t-1}^j)(\varepsilon_t^i - \varepsilon_{t-1}^i)$

which again is identical to $(ln(m_0) - ln(2 - m_0))^4 = \chi$. The probability of this to happen is $(\frac{1}{2}\frac{1}{2^{k-i}})^2(\frac{1}{2}\frac{1}{2^{k-j}})^2$ but we also have to take into account that each of these terms appears two times in the expansion of $\eta_{t+1,1}^2\eta_{t,1}^2$(A8).

Putting cases (i) to (iii) together we arrive at:

$$E\left[\eta_{t+1,1}^2\eta_{t,1}^2\right] =$$

$$\left[\sum_{i=1}^{k}\left(\left(\frac{1}{2}\frac{1}{2^{k-i}}\right)\sum_{j=1}^{k}\left(\frac{1}{2}\frac{1}{2^{k-j}}\right)\right) + 2\sum_{i=1}^{k}\left(\left(\frac{1}{2}\frac{1}{2^{k-i}}\right)^2\sum_{j=1,j\neq i}^{k}\left(\frac{1}{2}\frac{1}{2^{k-j}}\right)^2\right)\right]\chi. \tag{A9}$$

For arbitrary T, we get:

$$E\left[\eta_{t+T,T}^2\eta_{t,T}^2\right] = \left\{\sum_{i=1}^{k}\left(\tfrac{1}{2}\left(1 - \left(1 - \tfrac{1}{2^{k-i}}\right)^T\right)\sum_{j=1}^{k}\tfrac{1}{2}\left(1 - \left(1 - \tfrac{1}{2^{k-j}}\right)^T\right)\right)\right.$$

$$\left. + 2\sum_{i=1}^{k}\left(\tfrac{1}{4}\left(1 - \left(1 - \tfrac{1}{2^{k-i}}\right)^T\right)^2\sum_{j=1,j\neq i}^{k}\tfrac{1}{4}\left(1 - \left(1 - \tfrac{1}{2^{k-j}}\right)^T\right)^2\right)\right\}\chi. \tag{A10}$$

Turning to the log innovations of the compound process,

$$\xi_{t,T} = \ln|x_t| - \ln|x_{t-T}|,$$

we find:

$$E\left[\xi_{t+T,T}\xi_{t,T}\right] = E\left\{\left(0.5\sum_{i=1}^{k}\left(\varepsilon_{t+T}^{(i)} - \varepsilon_{t}^{(i)}\right) + ln|u_{t+T}| - ln|u_t|\right)\right.$$
$$\left.\left(0.5\sum_{i=1}^{k}\left(\varepsilon_{t}^{(i)} - \varepsilon_{t-T}^{(i)}\right) + ln|u_t| - ln|u_{t-T}|\right)\right\} \tag{A11}$$
$$= 0.25 \cdot E\left[\eta_{t+T,T}\eta_{t,T}\right] + \left(E\left[\ln|u_t|\right]\right)^2 - E\left[\left(\ln|u_t|\right)^2\right]$$

and

$$\begin{aligned} E\left[\xi_{t+T,T}^2\xi_{t,T}^2\right] &= E\left\{\left(0.5\sum_{i=1}^{k}\left(\varepsilon_{t+T}^{(i)} - \varepsilon_{t}^{(i)}\right) + ln|u_{t+T}| - ln|u_t|\right)^2\right. \\ &\quad \left.\left(0.5\sum_{i=1}^{k}\left(\varepsilon_{t}^{(i)} - \varepsilon_{t-T}^{(i)}\right) + ln|u_t| - ln|u_{t-T}|\right)^2\right\} \\ &= 0.25^2 \cdot E\left[\eta_{t+T,T}^2\eta_{t,T}^2\right] - \left\{E\left[\eta_{t,T}^2\right] - E\left[\eta_{t+T,T}\eta_{t,T}\right]\right\} \\ &\quad \cdot\left\{\left(E\left[\ln|u_t|\right]\right)^2 - E\left[\left(\ln|u_t|\right)^2\right]\right\} + 3 \cdot E\left[\left(\ln|u_t|\right)^2\right]^2 \\ &\quad -4 \cdot E\left[\ln|u_t|\right]E\left[\left(\ln|u_t|\right)^3\right] + E\left[\left(\ln|u_t|\right)^4\right]. \end{aligned} \tag{A12}$$

The log moments of the standard Normal variates u_t in (A11) and (A12) can be easily computed using the Gamma function and its derivatives.

We also note that the expectation of the squared log increment of the volatility process is:

$$E\left[\eta_{t+T,T}^2\right] = \sum_{i=1}^{k}\frac{1}{2}\left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right) \cdot \left(\ln\left(m_0\right) - \ln\left(2 - m_0\right)\right)^2 \tag{A13}$$

Furthermore, for computing linear forecasts we also need the second moment and the autocovariances of the volatility process itself which are easily derived as follows:

$$E\left[\mu_t^2\right] = \left(0.5(m_0^2 + (2 - m_0)^2)\right)^k \tag{A14}$$

and

$$E\left[\mu_{t+T}\mu_t\right]$$

$$= \prod_{i=1}^{k} \left\{ \left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right) 0.5 m_0 \left(2 - m_0\right) + \right. \tag{A15}$$

$$\left. \left(\left(1 - \frac{1}{2^{k-i}}\right)^T + 0.5\left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right)\right) \left(0.5 m_0^2 + 0.5\left(2 - m_0\right)^2\right) \right\}.$$

In (A15), the first term on the right-hand side gives the probability of observing different multipliers times the two different values m_0 and $2 - m_0$, whereas the second term gives the probability for observing the same multipliers at some level i at t and $t + T$ times the expectation of this common component.

B Moments of Lognormal Model

We now consider the case of a cascade with multipliers drawn from a Lognormal distribution with parameters λ and s, i.e. $M_t^{(i)} \sim LN\left(-\lambda, s^2\right)$. In order to normalize the product of the k volatility components, we assume that $E\left[M_t^{(i)}\right] = 1$. This leads to:

$$\exp\left(-\lambda + 0.5s^2\right) = 1 \Leftrightarrow s = \sqrt{2\lambda}.$$

Denoting the logs of the volatility components again by $\varepsilon_t^{(i)}$, we move on to derive the moment conditions fo this specification along the lines of Appendix A:

$$E\left[\eta_{t+1,1}\eta_{t,1}\right] = E\left\{\left(\sum_{i=1}^{k}\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\right)\left(\sum_{i=1}^{k}\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)\right)\right\}. \tag{B1}$$

Again we encounter non-zero contributions only if $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$ which leads to:

$$\begin{aligned} E\left[\eta_{t+1,1}\eta_{t,1}\right] &= \sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\right)^2 \cdot \left\{E[\varepsilon_t^{(i)}]^2 - E[(\varepsilon_t^{(i)})^2]\right\} \\ &= \left\{\sum_{i=1}^{k}\left(\frac{1}{2^{k-i}}\right)^2\right\}\left(\lambda^2 - \lambda^2 - s^2\right) \\ &= -\left\{\sum_{i=1}^{k}\left(\frac{1}{2^{k-1}}5\right)^2\right\}s^2. \end{aligned} \tag{B2}$$

For arbitrary lags T, one obtains accordingly:

$$E\left[\eta_{t+T,T}\eta_{t,T}\right] = -\left\{\sum_{i=1}^{k}\left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right)^2\right\}s^2. \tag{B3}$$

Turning to the auto-covariances of squared log increments, we can also take stock of our previous derivations. We can again distinguish between three different types of entries in $E\left[\eta_{t+1,1}^2\eta_{t,1}^2\right]$:

(1) entries of the form $\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)^2\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2$ which have non-zero value only if $\varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$. Their probability of occurance is $(\frac{1}{2^{k-i}})^2$ Solving the expectation in the non-zero case we obtain

$$\begin{aligned} &E\left[\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)^2\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2 \middle| \varepsilon_{t+1}^{(i)} \neq \varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}\right] \\ &= E[(\varepsilon_t^{(i)})^4] + 3E[(\varepsilon_t^{(i)})^2]^2 - 4E[(\varepsilon_t^{(i)})^3]E[\varepsilon_t^{(i)}] = 6s^4. \end{aligned}$$

The later result follows from the identities $E[(\varepsilon_t^{(i)})^3] = 3\lambda s^2 + \lambda^3$ and $E[(\varepsilon_t^{(i)})^4] = 3s^4 + 6\lambda^2 s^2 + \lambda^4$. Overall, putting together the probabilities of their occurrence and the non-zero value they give rise to yields a contribution: $\kappa_1 = \left\{ \sum_{i=1}^{k} \left(\frac{1}{2^{k-i}}\right)^2 \right\} \cdot 6s^4$

(2) entries of the form $\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2$ which are non-zero for $i \neq j$, $\varepsilon_{t+1}^{(j)} \neq \varepsilon_t^{(j)}$ and $\varepsilon_t^{(i)} \neq \varepsilon_{t-1}^{(i)}$. Since

$$E\left[\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)^2 \left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right)^2 | \varepsilon_{t+1}^{(l)} \neq \varepsilon_t^{(l)} \neq \varepsilon_{t-1}^{(l)}; l = i, j; i \neq j\right]$$

$$= 4E[(\varepsilon_t^{(i)})^2]^2 - 8E[(\varepsilon_t^{(i)})^2]E[\varepsilon_t^{(i)}]^2 + 4E[\varepsilon_t^{(i)}]^4 = 4s^4,$$

their overall contribution is: $\kappa_2 = \left\{ \sum_{i=1}^{k} \left(\frac{1}{2^{k-i}} \sum_{j=1, j\neq i}^{k} \frac{1}{2^{k-j}} \right) \right\} \cdot 4s^4$

(3) finally, entries of the 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.

Since $\left(\varepsilon_{t+1}^{(j)} - \varepsilon_t^{(j)}\right)\left(\varepsilon_{t+1}^{(i)} - \varepsilon_t^{(i)}\right)\left(\varepsilon_t^{(j)} - \varepsilon_{t-1}^{(j)}\right)\left(\varepsilon_t^{(i)} - \varepsilon_{t-1}^{(i)}\right) = s^4$ in this case, we obtain a contribution $\kappa_3 = \left\{ \sum_{i=1}^{k} \left((\frac{1}{2^{k-i}})^2 \sum_{j=1, j\neq i}^{k} (\frac{1}{2^{k-j}})^2 \right) \right\} 2 \cdot s^4$

Putting all these components together, we end up with

$$E\left[\eta_{t+1,1}^2 \eta_{t,1}^2\right] \quad = \kappa_1 + \kappa_2 + \kappa_3 \tag{B4}$$

and an appropriately modified analogous formula for $E\left[\eta_{t+T,T}^2 \eta_{t,T}^2\right]$

We also note that the second moment of $\eta_{t+T,T}$ is

$$E\left[\eta_{t+T,T}^2\right] = 2E\left[\left(\varepsilon_t^{(i)}\right)^2\right] - 2\left(E\left[\varepsilon_t^{(i)}\right]\right)^2 = \sum_{i=1}^{k} \left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right) \cdot 2s^2. \tag{B5}$$

In order to compute the autocovariances of the compound process, we only have to insert (B3), (B4) and (B5) into (A11) and (A12) of Appendix A.

Now turn to the moments of the volatility process itself. For the second moment of the product of volatility components we find:

$$E\left[\mu_t^2\right] = E\left[\left(\prod_{i=1}^{k} M_t^{(i)}\right)^2\right] = E\left[\left(M_t^{(i)}\right)^2\right]^k$$

since $E\left[\left(M_{i,t}\right)^2\right] = \exp\left(-2\lambda + 2s^2\right) = \exp\left(2\lambda\right)$

we have

$$E\left[\mu_t^2\right] = \exp\left(2\lambda \cdot k\right). \tag{B6}$$

Furthermore,

$$\begin{aligned} E\left[\mu_{t+1}\mu_t\right] &= E\left[\prod_{i=1}^{k} M_{t+1}^{(i)} \prod_{i=1}^{k} M_t^{(i)}\right] \\ &= \prod_{i=1}^{k}\left\{\frac{1}{2^{k-i}} + \left(1 - \frac{1}{2^{k-i}}\right) E\left[\left(M_t^{(i)}\right)^2\right]\right\} \\ &= \prod_{i=1}^{k}\left\{\frac{1}{2^{k-i}} + \left(1 + \frac{1}{2^{k-i}}\right) \exp\left(2\lambda\right)\right\}. \end{aligned} \tag{B7}$$

Note that the first term in the product stands for the probability of a change of the pertinent multiplier times the expectation of the product $M_{t+1}^{(i)} M_t^{(i)}$ which is unity.

Analogously,

$$E\left[\mu_{t+T}\mu_t\right] = \prod_{i=1}^{k}\left\{\left(1 - \left(1 - \frac{1}{2^{k-i}}\right)^T\right) + \left(1 - \frac{1}{2^{k-i}}\right)^T \exp\left(2 \cdot \lambda\right)\right\}. \tag{B8}$$

Appendix C: GARCH and FIGARCH Parameter Estimates

		μ	ρ	ω	β	α	φ	d	Logl	AIC	BIC
NYCI	GARCH	0.050	0.114	0.013	0.913	0.072			−5225.21	10460.43	10492.54
		(0.010)	(0.016)	(0.003)	(0.010)	(0.007)					
	FIGARCH	0.051	0.112	0.024	0.664		0.442	0.350	**-5199.82**	**10411.65**	**10450.19**
		(0.010)	(0.016)	(0.012)	(0.0971)		(0.100)	(0.106)			
DAX	GARCH	0.042	0.077	0.044	0.843	0.125			−6232.18	12474.36	12506.41
		(0.013)	(0.017)	(0.007)	(0.015)	(0.013)					
	FIGARCH	0.045	0.074	0.075	0.344		0.052	0.378	**-6200.93**	**12413.86**	**12452.32**
		(0.014)	(0.019)	(0.030)	(0.0801)		(0.044)	(0.119)			
USD-DEM	GARCH	−0.002	−0.042	0.018	0.876	0.095			**-4673.87**	**-9357.74**	**-9389.68**
		(0.010)	(0.016)	(0.003)	(0.011)	(0.009)					
	FIGARCH	−0.003	−0.040	0.022	0.601		0.215	0.467	−4676.51	9365.02	9403.35
		(0.011)	(0.024)	(0.011)	(0.127)		(0.063)	(0.161)			
GOLD	GARCH	−0.006	0.077	0.024	0.899	0.092			−6648.87	13307.74	13339.71
		(0.014)	(0.017)	(0.003)	(0.009)	(0.009)					
	FIGARCH	0.004	−0.095	0.067	0.600		0.382	0.408	**-6639.15**	**13290.30**	**13328.67**
		(0.023)	(0.022)	(0.075)	(0.397)		(0.260)	(0.241)			

Note: GARCH and FIGARCH estimates are obtained on the base of quasi-maximum likelihood. In both models we included a constand and AR(1) component in the level of returns denoted by μ and ρ (we also removed linear dependence before conputing multi-fractal estimates and historical volatility). The remaining parameters follow from the standard specification of GARCH(1,1): $h_t = \omega + \alpha_1 x_{t-1}^2 + \beta_1 h_{t-1}$ and FIGARCH(1,d,1): $h_t = \omega + \beta_1 h_{t-1} + (1 - \beta_1 L - (1 - \varphi_1 L)(1 - L)^d) x_t^2$. FIGARCH estimates are based on a truncation lag T = 1000 together with 1000 presample values set equal to the sample variance of the time series. Logl is the maximized log-likelihood and AIC and BIC are Akaike and Bayesian information criteria. In the case of the Gold price, reported FIGARCH parameters are those of a local maximum of the likelihood function which is dominated by a borderline solution with $d = 1$ which is prefered by both the AIC and BIC criteria, but whose forecasting performance is disastrous.

Table 1: MLE and GMM Estimation of Binomial Model Parameters ($k = 8$)

	$m_0 = 1.3$						$m_0 = 1.4$						$m_0 = 1.5$					
	MLE			GMM			MLE			GMM			MLE			GMM		
	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\bar{m}_0$	1.299	1.300	1.300	1.281	1.298	1.305	1.399	1.400	1.399	1.387	1.396	1.404	1.499	1.501	1.501	1.486	1.498	1.501
FSSE	0.017	0.012	0.009	0.095	0.060	0.040	0.017	0.011	0.008	0.070	0.043	0.027	0.016	0.012	0.008	0.049	0.030	0.021
RMSE	0.017	0.012	0.009	0.097	0.060	0.041	0.017	0.011	0.008	0.071	0.043	0.027	0.016	0.012	0.008	0.050	0.030	0.021
$\bar{\sigma}$	0.999	0.996	1.000	0.995	0.995	0.999	1.001	1.002	1.001	0.994	0.999	1.000	1.007	0.996	1.000	0.990	0.990	0.998
FSSE	0.072	0.047	0.035	0.076	0.049	0.037	0.090	0.064	0.043	0.094	0.069	0.046	0.111	0.084	0.059	0.122	0.088	0.065
RMSE	0.072	0.048	0.035	0.076	0.050	0.037	0.090	0.064	0.043	0.094	0.069	0.046	0.111	0.084	0.059	0.123	0.088	0.065

Note: All simulations are based on a multi-fractal process with $k = 8$ and $\sigma = 1$. Sample lengths are $T_1 = 2,500$, $T_2 = 5,000$ and $T_3 = 10,000$. $\bar{m}_0$ and $\bar{\sigma}$ denote the means of estimated parameters, FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. For each case, 400 Monte Carlo runs have been carried out. In the case of GMM, estimates of σ have been obtained from the elementary estimator of the sample standard deviation as the log increments used in GMM are scale free and, therefore, lack information on σ.

Table 2: GMM Estimation of Binomial Model ($k = 10, 15, 20$)

		$m_0 = 1.3$	$m_0 = 1.4$	$m_0 = 1.5$		$m_0 = 1.3$	$m_0 = 1.4$	$m_0 = 1.5$
k=10	$\bar{m}_0$	1.298	1.398	1.497	$\bar{\sigma}$	0.998	0.993	0.986
	FSSE	0.064	0.042	0.031	FSSE	0.096	0.132	0.171
	RMSE	0.064	0.042	0.031	RMSE	0.096	0.132	0.172
k=15	$\bar{m}_0$	1.297	1.402	1.498	$\bar{\sigma}$	0.964	0.924	0.891
	FSSE	0.061	0.041	0.030	FSSE	0.276	0.378	0.490
	RMSE	0.061	0.041	0.030	RMSE	0.278	0.385	0.502
k=20	$\bar{m}_0$	1.297	1.400	1.499	$\bar{\sigma}$	0.880	0.867	0.781
	FSSE	0.064	0.042	0.032	FSSE	0.423	0.676	0.716
	RMSE	0.064	0.042	0.032	RMSE	0.439	0.688	0.748

Note: The table exhibits GMM parameter estimates for a Binomial multi-fractal process based on samples of length 5,000. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. 400 Monte Carlo simulations have been run for each scenario. The underlying value of σ is 1 in all simulations. Estimates of σ have been obtained from the elementary sample standard deviation as the log increments used in GMM are scale free and, therefore, lack information on σ.

Table 3: GMM Estimation of Lognormal Model

		λ=0.05	0.10	0.150		0.05	0.10	0.15
	$\bar{\lambda}$	0.051	0.100	0.150	$\bar{\sigma}$	1.001	0.998	0.995
k=8	FSSE	0.020	0.021	0.024	FSSE	0.051	0.076	0.110
	RMSE	0.020	0.021	0.024	RMSE	0.051	0.076	0.110
	$\bar{\lambda}$	0.053	0.100	0.147	$\bar{\sigma}$	0.994	0.982	0.987
k=10	FSSE	0.020	0.023	0.024	FSSE	0.099	0.152	0.214
	RMSE	0.020	0.023	0.024	RMSE	0.099	0.153	0.214
	$\bar{\lambda}$	0.051	0.099	0.149	$\bar{\sigma}$	0.948	0.926	0.833
k=15	FSSE	0.021	0.022	0.023	FSSE	0.279	0.436	0.475
	RMSE	0.021	0.022	0.023	RMSE	0.283	0.442	0.503
	$\bar{\lambda}$	0.051	0.098	0.148	$\bar{\sigma}$	0.898	0.851	0.676
k=20	FSSE	0.020	0.023	0.023	FSSE	0.421	0.647	0.620
	RMSE	0.020	0.023	0.023	RMSE	0.433	0.663	0.698

Note: GMM parameter estimates for a Lognormal multi-fractal process based on samples of length 5,000. $\bar{\lambda}$ and $\bar{\sigma}$ denote the mean value of estimated parameters,FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. 400 Monte Carlo simulations have been run for each scenario. The underlying value of σ is 1 in all simulation. Estimates of σ have been obtained from the elementary sample standard deviation as the log increments used in GMM are scale free and, therefore, lack information about σ.

Table 4: Bayesian vs. Best Linear Forecasts for Binomial Model ($k = 8$)

		$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
	T	ML	BL1	BL2	ML	BL1	BL2	ML	BL1	BL2
MSE	1	0.942	0.947	0.946	0.923	0.932	0.932	0.919	0.932	0.932
		(0.03)	(0.014)	(0.013)	(0.016)	(0.017)	(0.017)	(0.022)	(0.023)	(0.023)
	5	0.963	0.967	0.967	0.955	0.961	0.961	0.957	0.965	0.965
		(0.012)	(0.012)	(0.012)	(0.014)	(0.014)	(0.014)	(0.016)	(0.015)	(0.015)
	10	0.974	0.977	0.976	0.969	0.974	0.974	0.972	0.978	0.978
		(0.011)	(0.010)	(0.010)	(0.012)	(0.011)	(0.011)	(0.013)	(0.012)	(0.013)
	20	0.984	0.985	0.985	0.981	0.984	0.984	0.983	0.987	0.987
		(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.010)	(0.009)	(0.010)
MAE	1	0.949	0.953	0.952	0.923	0.934	0.935	0.892	0.914	0.916
		(0.044)	(0.041)	(0.042)	(0.064)	(0.058)	(0.059)	(0.085)	(0.073)	(0.075)
	5	0.966	0.969	0.969	0.953	0.962	0.963	0.936	0.952	0.955
		(0.041)	(0.037)	(0.039)	(0.060)	(0.052)	(0.054)	(0.080)	(0.064)	(0.068)
	10	0.975	0.978	0.978	0.968	0.974	0.976	0.956	0.969	0.971
		(0.038)	(0.033)	(0.035)	(0.055)	(0.046)	(0.049)	(0.074)	(0.057)	(0.062)
	20	0.984	0.985	0.986	0.981	0.985	0.987	0.974	0.982	0.985
		(0.033)	(0.028)	(0.031)	(0.047)	(0.038)	(0.042)	(0.064)	(0.047)	(0.054)

Note: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for different forecast algorithms for the Binomial model. MSEs and MAEs are given in percentage of the pertinent MSEs and MAEs of a naive forecast using the in-sample variance. All entries are averages over 400 Monte Carlo runs (with standard errors given in parentheses). In each run, an overall sample of 10,000 entries has been split into an in-sample period of 5,000 entries for parameter estimation and an out-of-sample period of 5,000 entries for evaluation of forecasting performance. ML stands for parameter estimation based on the maximum likelihood procedure and pertinent inference on the probability of the 2^k states of the model. BL1 uses the best linear forecasts together with parameters estimated by GMM, while BL2 implements the best linear forecasting algorithm together with parameters estimated via maximum likelihood (using the same parameters as the ML algorithm).

Table 5: Linear Forecasts for Binomial Model, $k > 8$

		$k = 10$			$k = 15$			$k = 20$		
		$m_0 = 1.3$	$m_0 = 1.4$	$m_0 = 1.5$	$m_0 = 1.3$	$m_0 = 1.4$	$m_0 = 1.5$	$m_0 = 1.3$	$m_0 = 1.4$	$m_0 = 1.5$
MSE	1	0.918	0.904	0.906	0.852	0.848	0.827	0.844	0.836	0.829
		(0.028)	(0.037)	(0.047)	(0.114)	(0.105)	(0.176)	(0.121)	(0.140)	(0.172)
	5	0.940	0.936	0.943	0.874	0.879	0.863	0.866	0.865	0.863
		(0.026)	(0.034)	(0.043)	(0.116)	(0.107)	(0.181)	(0.124)	(0.144)	(0.177)
	10	0.952	0.950	0.957	0.886	0.893	0.877	0.878	0.880	0.877
		(0.025)	(0.033)	(0.040)	(0.118)	(0.107)	(0.182)	(0.125)	(0.145)	(0.178)
	20	0.963	0.962	0.968	0.898	0.906	0.890	0.890	0.893	0.889
		(0.024)	(0.031)	(0.038)	(0.118)	(0.108)	(0.181)	(0.127)	(0.147)	(0.177)
	50	0.977	0.977	0.980	0.914	0.924	0.904	0.906	0.909	0.903
		(0.021)	(0.027)	(0.033)	(0.119)	(0.108)	(0.180)	(0.128)	(0.147)	(0.176)
	100	0.986	0.986	0.987	0.925	0.934	0.913	0.917	0.920	0.911
		(0.018)	(0.023)	(0.028)	(0.118)	(0.108)	(0.175)	(0.129)	(0.147)	(0.173)
MAE	1	0.927	0.894	0.874	0.864	0.846	0.806	0.873	0.830	0.796
		(0.088)	(0.119)	(0.148)	(0.205)	(0.244)	(0.303)	(0.221)	(0.258)	(0.306)
	5	0.945	0.924	0.915	0.881	0.874	0.845	0.889	0.857	0.834
		(0.085)	(0.115)	(0.142)	(0.206)	(0.246)	(0.306)	(0.223)	(0.263)	(0.310)
	10	0.955	0.939	0.934	0.892	0.889	0.864	0.899	0.873	0.854
		(0.081)	(0.110)	(0.135)	(0.205)	(0.244)	(0.303)	(0.222)	(0.262)	(0.309)
	20	0.965	0.954	0.952	0.903	0.905	0.884	0.910	0.889	0.873
		(0.075)	(0.102)	(0.124)	(0.202)	(0.240)	(0.297)	(0.220)	(0.259)	(0.304)
	50	0.978	0.973	0.972	0.919	0.926	0.907	0.926	0.911	0.897
		(0.064)	(0.086)	(0.103)	(0.195)	(0.230)	(0.281)	(0.215)	(0.251)	(0.291)
	100	0.987	0.985	0.985	0.931	0.941	0.924	0.938	0.927	0.914
		(0.052)	(0.069)	(0.083)	(0.185)	(0.217)	(0.262)	(0.207)	(0.240)	(0.275)

Note: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for the Binomial model. MSEs and MAEs are given in percentage of the pertinent MSEs and MAEs of a naive forecast using the in-sample variance. All entries are averages over 400 Monte Carlo runs (with standard errors given in parentheses). In each run, an overall sample of 10,000 entries has been split into an in-sample period of 5,000 entries for parameter estimation and an out-of-sample period of 5,000 entries for evaluation of forecasting performance. All runs are based on the best linear forecasts together with parameters estimated by GMM. As can be seen, a higher number of cascade steps leads to higher degree of forecastability.

Table 6: Linear Forecasts for Lognormal Model

		$k=8$			$k=10$			$k=15$			$k=20$		
		$\lambda=0.05$	$\lambda=0.10$	$\lambda=0.15$	$\lambda=0.05$	$\lambda=0.10$	$\lambda=0.15$	$\lambda=0.05$	$\lambda=0.10$	$\lambda=0.15$	$\lambda=0.05$	$\lambda=0.10$	$\lambda=0.15$
MSE	1	0.943	0.931	0.935	0.915	0.904	0.913	0.851	0.846	0.852	0.865	0.835	0.842
		(0.018)	(0.026)	(0.029)	(0.030)	(0.053)	(0.059)	(0.110)	(0.143)	(0.156)	(0.098)	(0.156)	(0.171)
	5	0.965	0.963	0.968	0.939	0.939	0.947	0.873	0.878	0.885	0.888	0.866	0.876
		(0.015)	(0.019)	(0.018)	(0.028)	(0.050)	(0.053)	(0.112)	(0.146)	(0.158)	(0.102)	(0.159)	(0.173)
	10	0.975	0.975	0.980	0.951	0.952	0.961	0.886	0.893	0.898	0.901	0.880	0.889
		(0.013)	(0.015)	(0.013)	(0.026)	(0.047)	(0.050)	(0.114)	(0.147)	(0.157)	(0.101)	(0.160)	(0.172)
	20	0.985	0.985	0.988	0.962	0.965	0.972	0.898	0.904	0.910	0.914	0.892	0.900
		(0.010)	(0.012)	(0.009)	(0.024)	(0.044)	(0.046)	(0.114)	(0.147)	(0.155)	(0.102)	(0.161)	(0.170)
	50	0.994	0.994	0.996	0.977	0.978	0.982	0.914	0.919	0.922	0.930	0.908	0.912
		(0.006)	(0.007)	(0.005)	(0.020)	(0.038)	(0.040)	(0.114)	(0.145)	(0.151)	(0.102)	(0.160)	(0.166)
	100	0.998	0.998	0.998	0.986	0.986	0.998	0.926	0.928	0.931	0.942	0.917	0.920
		(0.004)	(0.004)	(0.003)	(0.016)	(0.031)	(0.033)	(0.113)	(0.141)	(0.144)	(0.102)	(0.157)	(0.162)
MAE	1	0.953	0.923	0.911	0.922	0.893	0.875	0.866	0.831	0.809	0.890	0.810	0.812
		(0.043)	(0.063)	(0.081)	(0.092)	(0.128)	(0.151)	(0.211)	(0.263)	(0.286)	(0.191)	(0.271)	(0.305)
	5	0.970	0.955	0.951	0.941	0.927	0.917	0.883	0.864	0.850	0.908	0.841	0.852
		(0.038)	(0.056)	(0.070)	(0.088)	(0.122)	(0.142)	(0.212)	(0.264)	(0.285)	(0.193)	(0.273)	(0.305)
	10	0.978	0.969	0.968	0.952	0.943	0.936	0.894	0.881	0.869	0.918	0.858	0.871
		(0.034)	(0.050)	(0.062)	(0.084)	(0.117)	(0.134)	(0.210)	(0.262)	(0.279)	(0.192)	(0.272)	(0.300)
	20	0.986	0.981	0.981	0.962	0.959	0.954	0.906	0.897	0.888	0.930	0.875	0.889
		(0.028)	(0.041)	(0.051)	(0.078)	(0.108)	(0.122)	(0.207)	(0.257)	(0.270)	(0.190)	(0.268)	(0.291)
	50	0.995	0.993	0.993	0.976	0.977	0.973	0.922	0.919	0.911	0.946	0.898	0.911
		(0.019)	(0.027)	(0.033)	(0.066)	(0.090)	(0.100)	(0.199)	(0.245)	(0.252)	(0.184)	(0.258)	(0.272)
	100	0.998	0.998	0.998	0.985	0.987	0.985	0.934	0.934	0.927	0.957	0.915	0.927
		(0.011)	(0.016)	(0.019)	(0.053)	(0.073)	(0.088)	(0.190)	(0.230)	(0.231)	(0.176)	(0.245)	(0.253)

Note: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for the Lognormal model. MSEs and MAEs are given in percentage of the pertinent MSEs and MAEs of a naive forecast using the in-sample variance. All entries are averages over 400 Monte Carlo runs (with standard errors given in parentheses). In each run, an overall sample of 10,000 entries has been split into an in-sample period of 5,000 entries for parameter estimation and an out-of-sample period of 5,000 entries for evaluation of forecasting performance. All runs are based on the best linear forecasts together with parameters estimated by GMM. In line with previous findings for the Binominal model, a higher number of cascade steps leads to higher degree of forecastability.

Table 7: Empirical Parameter Estimates

		Binomial					Lognormal				
		ML10	GMM5	GMM10	GMM15	GMM20		GMM5	GMM10	GMM15	GMM20
NYCI	$\hat{m}_0$	1.286	1.266	1.258	1.258	1.258	$\widehat{\lambda}$	0.037	0.035	0.035	0.035
	(SE)	(0.011)	(0.076)	(0.075)	(0.075)	(0.075)	(SE)	(0.022)	(0.021)	(0.021)	(0.021)
	(J_{prob})		(0.611)	(0.600)	(0.600)	(0.600)	(J_{prob})	(0.611)	(0.600)	(0.600)	(0.600)
DAX	$\hat{m}_0$	1.250	1.181	1.167	1.167	1.167	$\widehat{\lambda}$	0.017	0.014	0.014	0.014
	(SE)	(0.010)	(0.049)	(0.041)	(0.041)	(0.041)	(SE)	(0.009)	(0.007)	(0.007)	(0.007)
	(J_{prob})		(0.969)	(0.984)	(0.984)	(0.984)	(J_{prob})	(0.969)	(0.984)	(0.984)	(0.984)
USD-DEM	$\hat{m}_0$	1.267	1.283	1.276	1.276	1.276	$\widehat{\lambda}$	0.042	0.040	0.040	0.040
	(SE)	(0.012)	(0.066)	(0.066)	(0.066)	(0.066)	(SE)	(0.021)	(0.020)	(0.020)	(0.020)
	(J_{prob})		(0.410)	(0.388)	(0.388)	(0.388)	(J_{prob})	(0.410)	(0.388)	(0.388)	(0.388)
GOLD	$\hat{m}_0$	1.394	1.430	1.427	1.428	1.428	$\widehat{\lambda}$	0.106	0.105	0.105	0.105
	(SE)	(0.012)	(0.040)	(0.040)	(0.040)	(0.040)	(SE)	(0.023)	(0.022)	(0.022)	(0.022)
	(J_{prob})		(0.266)	(0.271)	(0.272)	(0.272)	(J_{prob})	(0.267)	(0.273)	(0.273)	(0.273)

Note: Empirical estimates of Binomial and Lognormal models are obtained via maximum likelihood with k=10 multipliers (ML10) and from GMM with k=5, 10, 15 and 20 (GMM5 etc.). Except for the case of the DAX, the eight moment conditions detailed in eq. 10 have been used. Since boundary solutions were obtained for the DAX with this setting, results shown in the table are those for a larger set of moments using Mom(T,q) in eq. 10 with $q = 1, 2$ and $T = 1, 5, 10, 20, 50, 100$ and 200. SE is the standard error of the pertinent estimates and the entry J_{prob} gives the probability of the pertinent J statistic in the case of GMM.

Table 8: Multi-fractal Forecasts: MSE

				Binomial				Lognormal		
	horizon	GARCH	FIGARCH	ML10	GMM10	GMM15	GMM20	GMM10	GMM15	GMM20
	1	0.943	0.921	0.934	0.929	0.929	0.929	0.928	0.928	0.928
	5	0.983	0.975	0.964	0.969	0.970	0.970	0.969	0.970	0.970
NYCI:	20	1.011	0.996	0.989	0.998	1.002	1.003	0.998	1.002	1.003
	50	1.011	1.018	0.995^{*}	1.007	1.012	1.013	1.007	1.012	1.013
	100	1.007	1.015	0.995	1.006	1.012	1.014	1.006	1.012	1.013
	1	0.779	0.788	0.840	0.853	0.850	0.849	0.852	0.848	0.849
	5	0.899	0.891	0.886	0.897	0.892	0.891	0.896	0.892	0.890
DAX:	20	0.937	0.935	0.934	0.940	0.936	0.935	0.940	0.936	0.935
	50	0.976	0.973	0.965	0.975	0.973	0.973	0.975	0.973	0.973
	100	0.983	0.976	0.969^{+}	0.983	0.978	0.979	0.983	0.978	0.979
	1	0.899	0.891	0.893	0.889	0.888	0.888	0.889	0.888	0.888
	5	0.899	0.888	0.885	0.884	0.879	0.878	0.883	0.879	0.877
USD-DEM:	20	0.957	0.932^{*}	0.920	0.916	$0.910^{\ddagger}$	$0.909^{*\ddagger}$	0.916	$0.910^{\ddagger}$	$0.909^{*\ddagger}$
	50	1.031	0.980^{*}	0.942^{*+}	0.939^{*+}	0.934^{*+}	0.936^{*+}	0.939^{*+}	0.934^{*+}	0.936^{*+}
	100	1.062	1.013^{*}	0.948^{*+}	0.948^{*+}	$0.939^{*+\ddagger}$	0.940^{*+}	0.949^{*+}	$0.939^{*+\ddagger}$	0.940^{*+}
	1	0.370	0.362	0.377	0.371	$0.356^{*\ddagger}$	$0.356^{*\ddagger}$	0.375	$0.358^{\ddagger}$	$0.358^{\ddagger}$
	5	0.386	0.375^{*}	0.385	0.395	0.367^{*+}	0.366^{*+}	0.402	0.371^{*}	0.371^{*+}
Gold:	20	0.484	0.429^{*}	0.429^{*}	0.463	$0.404^{*+\ddagger}$	$0.403^{*+\ddagger}$	0.474	0.412^{*+}	0.410^{*+}
	50	0.640	0.465^{*}	0.430^{*+}	0.517^{*}	0.416^{*+}	0.414^{*+}	0.533^{*}	0.427^{*+}	0.425^{*+}
	100	0.941	0.549^{*}	0.475^{*+}	0.608^{*}	0.465^{*+}	0.461^{*+}	0.625^{*}	0.480^{*+}	0.477^{*+}

Note: The table shows empirical mean squared errors of various models in relation to the MSEs of the benchmark of historical volatility. * denotes an improvement against the GARCH model which is significant at the 99% level, $^{+}$ an improvement against the FIGARCH model which is significant at the 99% level (applied only to multifractal models), and ‡ an improvement against the ML10 Binomial multifractal model which is significant at the 99% level (only applied to other multifractal models). All comparisons are based on the test statistics of Diebold and Mariano (1995).

Table 9: Multi-fractal Forecasts: MAE

				Binomial				Lognormal		
		GARCH	FIGARCH	ML10	GMM10	GMM15	GMM20	GMM10	GMM15	GMM20
NYCI:	1	1.084	1.046*	1.074	1.055*‡	1.058*	1.058*‡	1.056*	1.058*	1.058*
	5	1.079	1.042*	1.070	1.050‡	1.054	1.055*	1.050‡	1.054	1.055*
	20	1.119	1.052*	1.094	1.065*‡	1.074	1.079	1.065‡	1.074	1.079
	50	1.082	1.029*	1.083	1.046*‡	1.059*‡	1.060*	1.046*‡	1.055*‡	1.060*
	100	1.032	0.991*	1.061	1.012*‡	1.011*‡	1.014*‡	1.014*‡	1.012*‡	1.014*‡
DAX:	1	1.051	1.047	1.025	1.018	1.029	1.029	1.019	1.029	1.030
	5	1.071	1.061	1.024	1.022^{+}	1.038	1.041	1.023^{+}	1.038	1.041
	20	1.036	1.046	1.025	1.020^{+}	1.038	1.043	1.020^{+}	1.039	1.043
	50	1.015	1.040	1.026	1.023^{+}	1.037	1.042	1.023^{+}	1.038	1.042
	100	1.002	1.014	1.003	1.007	1.014	1.017	1.007	1.014	1.017
USD-DEM:	1	0.822	0.797*	0.792*	0.793*	0.779$^{*+\ddagger}$	0.777$^{*+\ddagger}$	0.792*	0.778$^{*+\ddagger}$	0.777$^{*+\ddagger}$
	5	0.850	0.815*	0.802^{*+}	0.806*	0.785$^{*+\ddagger}$	0.782$^{*+\ddagger}$	0.805*	0.785$^{*+\ddagger}$	0.782$^{*+\ddagger}$
	20	0.946	0.886*	0.841^{*+}	0.847^{*+}	0.820$^{*+\ddagger}$	0.816$^{*+\ddagger}$	0.847^{*+}	0.820$^{*+\ddagger}$	0.817$^{*+\ddagger}$
	50	1.031	0.937*	0.853^{*+}	0.867^{*+}	0.830$^{*+\ddagger}$	0.826$^{*+\ddagger}$	0.868^{*+}	0.831$^{*+\ddagger}$	0.827$^{*+\ddagger}$
	100	1.075	0.999*	0.873^{*+}	0.895^{*+}	0.849$^{*+\ddagger}$	0.842$^{*+\ddagger}$	0.895^{*+}	0.849$^{*+\ddagger}$	0.843$^{*+\ddagger}$
GOLD:	1	0.456	0.443*	0.446	0.480	0.440*	0.439*	0.488	0.447	0.446
	5	0.483	0.473	0.465*	0.516	0.459^{*+}	0.457^{*+}	0.528	0.470	0.468
	20	0.599	0.548*	0.525*	0.597	0.513^{*+}	0.510^{*+}	0.611	0.527^{*+}	0.524^{*+}
	50	0.759	0.607*	0.554^{*+}	0.602*	0.545^{*+}	0.541^{*+}	0.678*	0.562^{*+}	0.558^{*+}
	100	0.967	0.686*	0.606^{*+}	0.741*	0.595^{*+}	0.591^{*+}	0.755*	0.614^{*+}	0.610^{*+}

Note: The table shows empirical mean absolute errors of various models in relation to MAEs of the benchmark of historical volatility. * denotes an improvement against the GARCH model which is significant at the 99% level, $^{+}$ an improvement against the FIGARCH model which is significant at the 99% level (applied only to multifractal models), and ‡ an improvement against the ML10 Binomial multifractal model which is significant at the 99% level (only applied to other multifractal models). All comparisons are based on the test statistics of Diebold and Mariano (1995).

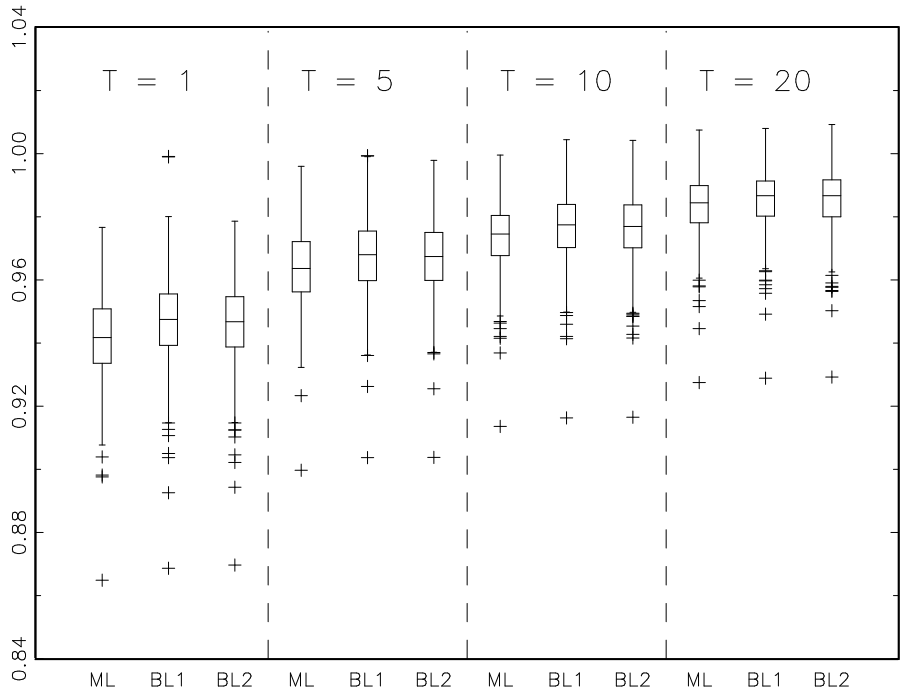

Figure 1: Box plot of the distributions of MSEs for different forecasting horizons over 400 Monte Carlo runs for the Binomial model with parameter $m_0 = 1.3$ (corresponding to the results given in the upper left corner of Table 4). The boxes show the median of the distribution surrounded by a box that spans the center half of the data set (the inter-quartile range). The whiskers extend 1.5 times the inter-quartile range with the values outside this range identified as outliers. The appearence of the box plot is virtually identical for other parameter values.

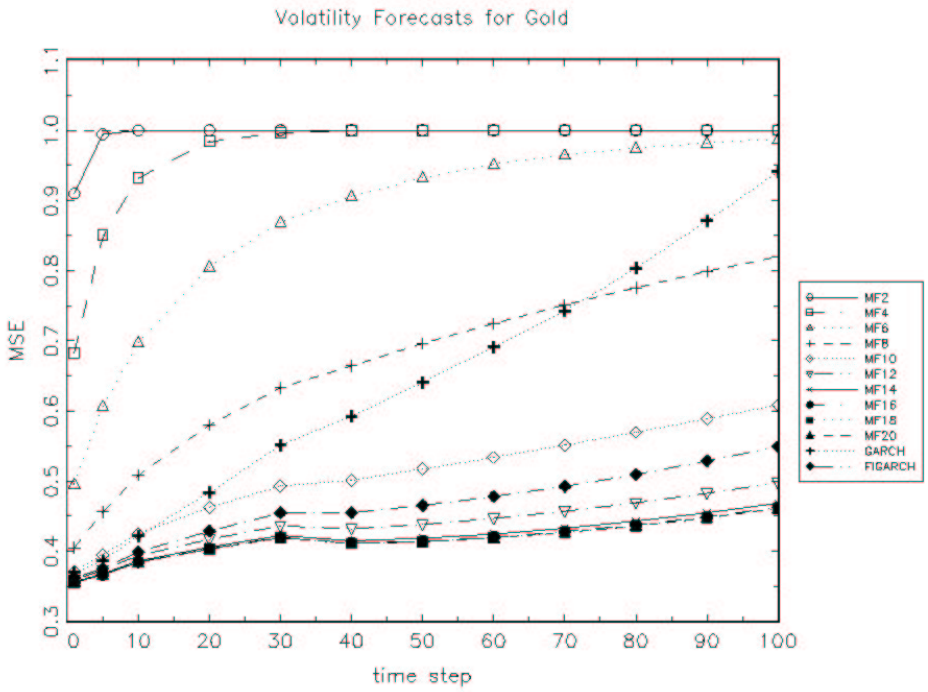

Figure 2: Mean squared errors of volatility forecasts from GARCH, FIGARCH and Binomial MF for the Gold price over various time horizons.