

Lux, Thomas

Working Paper

The Markov-Switching Multifractal Model of asset returns: GMM estimation and linear forecasting of volatility

Economics Working Paper, No. 2006-17

Provided in Cooperation with:

Christian-Albrechts-University of Kiel, Department of Economics

Suggested Citation: Lux, Thomas (2006) : The Markov-Switching Multifractal Model of asset returns: GMM estimation and linear forecasting of volatility, Economics Working Paper, No. 2006-17, Kiel University, Department of Economics, Kiel

This Version is available at:

<https://hdl.handle.net/10419/3927>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

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

by Thomas Lux



Christian-Albrechts-Universität Kiel

Department of Economics

Economics Working Paper

No 2006-17



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

April 5, 2006

Abstract

Multifractal 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 multifractal 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 incurred 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 multifractal models (Calvet et al., 1997; 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 rather than causal nature 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 multifractal model based on Poisson arrival 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 to this emergent literature in this paper is twofold: (i) we introduce an alternative GMM estimator of multifractal 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 or restrictions on the 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 multifractal 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 to various foreign exchange data shows that using a larger state space indeed provides further scope for improvement of long-run volatility predictions for some time series while the replacement of the Binomial model by the Lognormal specification seems to make almost no difference at all. The structure of the remainder is as follows: section 2 shortly reviews available literature on multifractal 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 The Markov-Switching Multifractal Model

Multifractal measures have a long history in physics dating back at least to the early seventies when Mandelbrot introduced them as models for the distribution of energy in turbulent dissipation (e.g., Mandelbrot, 1974). The main reason for considering multifractal processes 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. Multifractality, furthermore, implies that different powers of the measure have different decay rates of their autocovariances. While early empirical evidence of multifractality in financial data had already appeared in, e.g., Vassilicos et al. (1993), Calvet et al. (1997) were the first to develop a multifractal model for financial data. Their *Multifractal Model of Asset Returns* (MMAR) assumes that returns follow a compound process in which an incremental Brownian motion is subordinate to the cumulative distribution function of a multifractal measure. Calvet and Fisher (2002a) develop estimators and diagnostic tests for this model on the base of its scaling properties.

However, despite the attractiveness of its stochastic properties, practical applicability of the MMAR suffered from its combinatorial nature and its non-stationarity due to the restriction to a bounded interval. These limitations have been overcome by the iterative time series models introduced by Calvet and Fisher (2001, 2004). Calvet and Fisher (2001) define a continuous-time multifractal model with random times for the changes of multipliers (Poisson multifractal) 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 multifractal, 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-Switching Multifractal process* (MSM). Returns are modelled as:

$$x_t = \sigma_t \cdot u_t \quad (1)$$

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 $M_t^{(1)}, M_t^{(2)}, \dots, M_t^{(k)}$ and a constant scale factor σ :

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

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 $1 - \gamma_i$. The transition probabilities are specified by Calvet and Fisher (2001) as:

$$\gamma_i = 1 - (1 - \gamma_k)^{(b^i - k)}, \quad (3)$$

with parameters $\gamma_k \in [0, 1]$ and $b \in (1, \infty)$. Estimation of this model, then, involves the parameters γ_k and b as well as those characterizing the distribution of the components $M_t^{(i)}$. In the present paper, we explore two specifications for the distribution of multipliers: the Binomial and Lognormal MSM models. Following Calvet and Fisher (2004) the *Binomial MSM* is characterized by Binomial random draws taking the values m_0 and $2 - m_0$ ($1 \leq m_0 < 2$) with equal probability (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. Conditional state probabilities can be used to compute forecasts of future volatility and these are shown to outperform forecasts derived from GARCH, FIGARCH, and two-state Markov switching GARCH models. While this approach is optimal for the Binomial MSM as well as multinomial generalizations, its general applicability is limited in two respects: First, it is not applicable for models with an infinite state space, i.e. continuous distributions of the multipliers. Second, current computational limitations make choices of k beyond 10 unfeasible even for the Binomial case because of the implied evaluation of a $2^k \times 2^k$ transition matrix in each iteration.¹

In the following we consider both specifications with a continuous distribution of multipliers and large numbers of multipliers, $k > 10$. As an example of a multifractal model with a continuous state space we consider the *Lognormal MSM* model with volatility components drawn from a Lognormal distribution (cf. Calvet and Fisher (2002a) for the earlier non-stationary Lognormal MMAR). In this model, multipliers are determined by random draws from a Lognormal distribution with parameters λ and s , i.e.

$$M_t^{(i)} \sim LN(-\lambda, s^2). \quad (4)$$

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

$$\exp(-\lambda + 0.5s^2) = 1, \quad (5)$$

from which a restriction on the shape parameter can be inferred: $s = \sqrt{2\lambda}$. Hence, the distribution of volatility components is parameterized by a one-parameter family of Lognormals with the normalization restricting the choice of the shape parameter.

Given the successful performance of the Binomial MSM *with* $k \leq 10$ investigated in Calvet and Fisher (2004), it appears certainly worthwhile to explore whether one would get similar or better results by increasing the number of volatility components or adopting a continuous distribution of the multipliers. To this end, we introduce a GMM estimator as a flexible and versatile estimation method for multifractal 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 (1) and (2). In order to concentrate on the estimation of the parameters of the distribution of volatility components and to confine the overall number of parameters to be estimated via GMM, we restrict the specification of transition probabilities by fixing *ex ante* their parameters, b and γ_k , and focusing on estimation of the parameters of the distribution of volatility components. Following the recommendation of an anonymous referee, we imposed the restrictions $b = 2$ and $\gamma_k = 0.5$. Note that the later restriction implies that the volatility component at the highest frequency has a probability of one-half to be replaced

¹For multinomial specifications with more than two states, computational limitations would restrict the choice of k to even smaller numbers.

by a new draw. We have performed a limited sensitivity analysis by replacing this second parameter by $\gamma_k = 0.99$ which would amount to replacement of the component at the high-frequency end with a probability close to 1 (reminiscent of the earlier MMAR model in which the high-frequency component is replaced with certainty in every period).² We found a certain trade-off between the restriction on γ_k and results for estimated distributional parameters (reported below), but little overall difference in goodness-of-fit of the MSM model and its forecasting performance.³

3 Estimation of Markov-Switching Multifractal Models via GMM

Estimation of multifractal models has proceeded from adaption of the so-called “scaling estimator” in Calvet et al. (1997) for the combinatorial MMAR – which is still pervasively used in natural sciences – to the seminal proposal of maximum likelihood estimation for the discretized 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.

However, because of the restriction of MLE to discrete distributions of multipliers and computationally feasible values of k , it appears worthwhile to explore alternative ways of estimating MSM parameters. To this end, we follow some earlier attempts at investigating method of moments estimation. However, in contrast to the SMM (Simulated Method of Moments) approach in Calvet and Fisher (2002b), we use a GMM (Generalized Method of Moments, cf. Hansen, 1982) approach with analytically solvable moment conditions. We shortly review the basic GMM framework before turning to our particular set of moment conditions. In GMM, the vector of parameter estimates of a model, say φ , is obtained as:

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

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

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

with covariance matrix $\Xi = (\bar{F}_T' \bar{V}_T^{-1} \bar{F}_T)^{-1}$ in which φ_0 is the true parameter vector, $\hat{V}_T^{-1} = T \text{var} f_T(\hat{\varphi}_T)$ 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 Calvet et al. (1997) and Calvet and Fisher (2002a) because of its

²Note that one could also 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.

³We have also tried the simpler specification $\gamma_i = 2^{-(k-i)}$, also without much of an effect on the overall outcome of our Monte Carlo simulations and empirical application.

non-stationarity violating the required regularity conditions. This problem, fortunately, does not carry over to the MSM model which rather fits into the class of Markov-switching models with standard asymptotic behavior. As has been pointed out by Calvet and Fisher (2004), 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 $t \rightarrow \infty$ 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 multifractal processes, therefore,

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

holds only over an interval $1 \ll \tau \ll b^k$ with b the number of states (i.e. $b = 2$ in the binomial MSM) and k the number of multipliers.

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.

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:

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

As is shown in Appendix A, the transformed variable $\xi_{t,T}$, in fact, only has non-zero auto-covariances over a limited number of lags. With non-overlapping observations, we only observe non-zero auto-covariances at lag one, i.e. for $\xi_{t,T}$ and $\xi_{t+T,T}$, while for overlapping log differences $\xi_{t,T}, \xi_{t+1,T}, \dots$ auto-covariances would be non-zero over T lags. In our applications below, overlapping observations are used. One may note that moments of $\xi_{t,T}$ only depend on the parameters of the volatility process while the standard deviation of the Normally distributed increments, σ , drops out when computing log differences:

$$\begin{aligned} \xi_{t,T} &= \ln \left| \sigma u_t \left(\prod_{i=1}^k M_t^{(i)} \right)^{\frac{1}{2}} \right| - \ln \left| \sigma u_{t-T} \left(\prod_{i=1}^k M_{t-T}^{(i)} \right)^{\frac{1}{2}} \right| \\ &= 0.5 \sum_{i=1}^k (\varepsilon_t^{(i)} - \varepsilon_{t-T}^{(i)}) + \ln |u_t| - \ln |u_{t-T}| \end{aligned}$$

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

The lack of inclusion of the scale parameter σ in the above parameter set makes it necessary to add another moment condition such as the second moment of x_t that exactly identifies σ .

In order to exploit the temporal scaling properties of the multifractal 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:

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

for $q = 1, 2$ and $T = 1, 5, 10, 20$ together with $E[x_t^2] = \sigma^2$ for identification of σ . Since the eight moment conditions for m_0 or λ are not affected by σ , the covariance matrix of the parameters should be block-diagonal and estimated values of σ should be essentially identical to the sample standard deviation.

Note that our moment conditions differ from those used in previous literature. Calvet and Fisher (2002b) had used a variety of moment-scaling properties, parameters of log-log regressions of the sample auto-covariogram, slope parameters from log-periodogram regressions, high-frequency auto-covariances and tail index estimates. They report that the covariogram and tail index estimates showed the lowest mean squared errors. Earlier versions of this paper had also applied high-frequency auto-covariances following the example of the stochastic volatility literature (Andersen and Sørensen, 1996). However, experimentation indicated that the performance of these moment conditions was quite sensitive to the underlying parameter values of the multifractal model. In contrast, as will be shown in our Monte Carlo simulations below, using log-moments leads to relatively homogeneous behavior across the parameter values m_0 and λ and even more so across different choices of k . This nice behavior of log-moments might be attributed to the resulting linear connection of (log) volatility components instead of their multiplicative relationship in the raw data.

An alternative Simulated Method of Moments (SMM) approach has been pursued recently by Calvet, Fisher and Thompson (2006) for estimation of the parameters of a bi-variate MSM model. They use a particle filter to optimize the simulated likelihood of the bivariate Binomial model within a two-step estimation approach that can be interpreted as a SMM estimator. In comparison to this approach, our GMM estimator is based on analytical moments which considerably reduces computational demands.

It might be worthwhile to also try various combinations of our moment conditions for log increments and other moments used in previous research which might provide additional scope for increasing the efficiency of GMM. However, a comprehensive investigation into potential gains from various sets of alternative moment conditions is beyond the scope of the present paper and might be left for future research.

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). Variance-covariance matrices of moment conditions and, hence, the weighting matrices of the subsequent iteration are computed numerically. Inspection of the variance-covariance matrices shows that both in our Monte Carlo simulations and in the empirical application, the block-diagonal structure is apparent as the covariance between the estimators of m_0 or λ and σ is close to zero.

We start with the Binomial Model with a limited number of multipliers, $k = 8$, and subsequently increase the number of volatility components and also switch to the Lognormal model as an example of a specification with continuous state space.⁴ Our first experiments serve to establish the efficiency of the GMM

⁴Earlier 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 (2002a), in fact, devised 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

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 parameters of the transition probabilities (3) as $b = 2$ and $\gamma_k = 0.5$ so that we only have to estimate two parameters compared to four in Calvet and Fisher’s somewhat more general approach. 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 ML estimates of m_0 are hardly affected at all by the choice of parameters, while RMSEs appear to increase for σ for higher m_0 . This is plausible since with increasing m_0 one generates enhanced fluctuations of the product of volatilities which might interfere 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.⁵ Obviously, variability of estimates for m_0 with GMM is much higher, ranging from 4 to 10 times that of the ML estimates. The relative difference is the higher, the smaller the sample size 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 exhibit a decrease of both the bias and MSEs when proceeding from $m_0 = 1.3$ to $m_0 = 1.5$. A closer look reveals that the larger bias for smaller m_0 is caused by a certain number of cases in which the estimator gets locked-in at the lower boundary ($m_0 = 1$) of the admissible parameter space. The farther the true parameter is above the boundary the fewer are these cases so that for $m_0 = 1.4$ or 1.5 practically all estimates are located in the interior of the admissible parameter space and the GMM estimators become essentially as unbiased as ML. In contrast, the quality of the estimates of the scale parameter σ is almost identical under both methods with only a very slight advantage for ML against GMM. Interestingly, the average bias of the Monte Carlo estimates is moderate throughout and quickly approaches zero for the larger sample sizes.⁶ 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 an increasing number of cascade components,

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 nine 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 increased at reasonable computational costs.

⁵However, one might note that obtaining our GMM estimates with nine moment conditions requires only a small fraction of the time needed for ML estimation.

⁶In 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 (with a slightly slower convergence rate in the case $m_0 = 1.3$ due to the influence of boundary solutions). This underscores that the log transformation is useful in improving the quality of GMM estimates as compared to earlier experiments using moments of the raw data.

$k = 10, 15$ 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 efficiency of multifractal 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 .⁷

Table 2 about here

In the next step, we consider the Lognormal MSM, for which MLE is not applicable at all because of its continuous state space of volatility components. Moment conditions for this model are spelled out in Appendix B. Note that the admissible parameter space for the location parameter 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 heterogeneity 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 $10, 15$ and 20 . As can be seen, results are not too different from those obtained with the Binomial model. Biases are negligible for all parameter values of our Monte Carlo study. Interestingly, the danger of lock-in of estimators for λ at the lower boundary $\lambda = 0$ seems to be non-existent so that the performance of GMM is almost the same for the whole range of our λ s. Again results for λ are also almost insensitive with respect to k , but RMSEs for σ 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 used 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-nomial cases with a finite state space. Again, it would be useful to have methods at hand for specifications beyond the confines of multi-nomial 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 also be computed on the base of

⁷Note also that for $m_0 = 1.3$ biases are even smaller for $k \geq 10$ than for $k = 8$ due to a reduced number of estimates at the boundary $m_0 = 1$.

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

$$\hat{X}_{n+h} = \sum_{i=1}^n \phi_{ni}^{(h)} X_{n+1-i} = \phi_n^{(h)} \mathbf{X}_n, \quad (11)$$

with the vectors of weights $\phi_n^{(h)} = (\phi_{n1}^{(h)}, \phi_{n2}^{(h)}, \dots, \phi_{nn}^{(h)})'$ being any solution of $\mathbf{\Gamma}_n \phi_n^{(h)} = \gamma_n^{(h)}$ and $\gamma_n^{(h)} = (\gamma(h), \gamma(h+1), \dots, \gamma(n+h-1))'$ denoting the auto-covariances for the data-generating process of X_t at lags h and beyond, and $\mathbf{\Gamma}_n = [\gamma(i-j)]_{i,j=1,\dots,n}$ 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 realizations 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 although 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 wishes 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 Binomial and Lognormal models in Appendices A and B, respectively. Implementing (11), we have to consider the zero-mean time series:

$$X_t = x_t^2 - E[x_t^2] = x_t^2 - \hat{\sigma}^2 \quad (12)$$

where $\hat{\sigma}$ is the estimate of the scale factor σ in eq. (2).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 Binomial 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 minimize 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. the sample standard deviation 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

⁸Note that $\hat{\sigma}$ only appears in the mean value of eq. (12), but it drops from the coefficients $\phi_{ni}^{(h)}$ where by construction it appears in both the denominator and the numerator.

as detailed in Calvet and Fisher (2004), “BL1” uses best linear forecasts on the base of GMM parameter estimates, and “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 a 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 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 only slight difference between the three methods is that BL1 produces a number of outliers with standardized MSE equal to one. These corresponds to those cases with parameters estimates $\widehat{m}_0 = 1$ for which multifractality vanishes and predicted volatility is identical to historical volatility. However, the overall performance (in terms of mean and median predictive performance) is hardly affected by these cases and the boundary estimates do hardly occur any more for higher values of m_0 . Although linear forecasts are geared towards minimization of MSE, the difference between ML based Bayesian forecasting and linear forecasting are practically the same with the MAE criterion. On the other hand, it also appears worth mentioning that ML seems to come along with a somewhat 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 4 also indicates that forecast quality depends on the multifractal 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 at long time horizons.

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 dealt with on the base of ML based methods. To this end, Table 5 shows results for the Binomial 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 6.5 percent at $k = 8$ to 15.1 percent with $k = 20$. For the twenty day horizon we observe an increase from 2.0 percent ($k = 8$) which probably is irrelevant in an applied context to 10.4 percent ($k = 20$) which might be more interesting. We have also included long time horizons of fifty and one hundred lags which still have improvements of about 8 to 9 percent compared to the MSE of the sample variance. Results are fairly homogeneous across multifractal parameters as well as for both the MSE and MAE criterion, but forecastability seems to be higher for higher values of m_0 . It seems particularly worthwhile 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 not appear to be a major obstacle to successful prediction of future volatility.

Pretty similar results are obtained with the Lognormal specification. Gains in predictability are about the same as with the Binomial model for increasing numbers of volatility components. Investigating a larger set of parameter values (available upon request), we found that 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 homogeneity 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 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 accessibility 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 five different foreign exchange markets: The Deutsche Mark exchange rate against the U.S.\$ extended by the EURO from 1999 (DEM/EUR) and the exchange rates of the British Pound (UKP), the Canadian Dollar (CND), Japanese Yen (YEN) and Swiss Franc (CHF), all against the U.S. Dollar. All time series start on 1 January 1979 and extend until 31 December 2004. All data were obtained from Datastream. Most series consist of buying rates computed at noon in New York (YEN, UKP, CND), while the DEM/EURO - U.S. Dollar exchange rates are based on snapshots of USD quotes from multi-contribution sources taken at 4 p.m. London time, and the Swiss Franc exchange rate is the one compiled by the Swiss National Bank at 11 a.m. GMT. Investigation of these foreign exchange rates is interesting as these time series (with the exception of CHF) have also been used in the recent study by Calvet and Fisher (2004) who found improvements of the MSM model over GARCH type models in out-of-sample forecasts of volatility. We have also studied the performance of our GMM approach for some other financial data (the stock market indices DAX of the German stock market, the New York Stock Exchange Composite Index, and the price of gold) and report on pertinent results below (details are available upon request).

Our samples cover twenty-six years starting from 1 January 1979 and ending on 31 December 2004. We use the data of the years 1979 to 1996 for in-sample estimation and leave the eight remaining years for out-of-sample evaluation of volatility forecasts. This gives about 4,600 in-sample observations and 2000 out-of sample entries (with slight variations of the numbers across markets due to differers in the number of active days). The empirical log-likelihood, AIC and BIC criterion values for ML estimation are shown in Appendix C alongside with

pertinent GARCH and FIGARCH results. Except for the Canadian Dollar, the MSM model is preferred over the alternative models by these statistics, which nicely confirms results reported in Calvet and Fisher (2004).

Table 6 reports in-sample parameter estimates for the multifractal 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. (1) 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.

Inspecting the results, a number of observations are remarkable⁹: in particular, for the GMM approach, parameter estimates for m_0 and λ are practically the same with different numbers of multipliers 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 at small numbers for k which becomes slower and slower until, eventually, a lock-in at a constant value can be found somewhere between $k = 10$ and $k = 15$. As can be seen from Table 6 estimates still undergo slight variations when proceeding from $k = 5$ to 10 but they remain practically 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 additional high-level 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 multifractal models with different numbers of components once k increases beyond a certain threshold.

While GMM parameter estimates are very homogeneous over different specifications of k , the ML and GMM parameter estimates show no particular coherence across time series. Table 6 also shows the probability of Hansen's test statistics $J_T = f_T(\hat{\varphi})' \hat{A}_T f_T(\hat{\varphi})$, which apparently does not change at all beyond $k = 10$. This observation also suggests that the similarity of models with a different number of volatility components 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 Calvet and 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 sensitivity 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 comparison of high k and k' even if the true k is the larger one if these tests were computationally feasible.¹⁰ Model selection among competing specifications of multifractal models with different numbers of multipliers, therefore, remains a challenging task. In view of the inconclusive results, we choose to take an agnostic approach and to investigate the performance of var-

⁹For the alternative setting with fixed parameters $b = 2$ and $\gamma_k = 0.99$ the multifractal parameters (m_0 and λ) were typically slightly below those reported here, while overall goodness-of-fit and forecasting capacity were almost the same. It appears plausible that a higher frequency of replacement at the high-frequency end leads to more volatility and, therefore, gives rise to a compensatory reduction of the heterogeneity of the multipliers.

¹⁰Calvet 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 find 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.

ious specifications in predicting future volatility.

Table 7 about here

As another remarkable finding in Table 6, 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, practically identical for both time series under all specifications of the number of volatility components. 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, for sufficiently high k , seem to provide enough flexibility for capturing the heterogenous volatility dynamics. Note also that estimated MSM models can fit our selection of moments quite well: except for the case of the Swiss Franc the J -test statistics would not allow to reject the MSM as the data-generating process on the base of our chosen moment function.

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 the auto-correlations 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 auto-correlations at over 2700 lags, i.e. about 10 years. Since higher k implies a longer power-law range of the autocorrelations, the longer dependence in volatility might improve forecasting 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 multifractal 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 7 and 8 explore this issue for our selection of foreign exchange rates. For all five time series we compare forecasts over horizons varying from one day over 5, 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.¹¹ MSEs and MAEs are again reported relative to those of historical volatility.

Results are not entirely homogenous, but are quite encouraging. First, inspecting out-of-sample MSE (cf. Table 7), we see that in all cases but one some or all versions of MSM do outperform GARCH and FIGARCH. The one exception is the Canadian Dollar where FIGARCH has a slight advantage over MSM at most time horizons (which is in nice conformity with the in-sample dominance of the FIGARCH model for this series). Multistage forecasts typically exhibit an increasing advantage of MSM against GARCH/FIGARCH at longer

¹¹Estimated 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 multifractal model with a well-established benchmark. It is, however, worth pointing 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.

prediction horizons. The Diebold-Mariano test for differences in predictive accuracy allows to reject GARCH in favor of MSM in the Canadian Dollar and both, GARCH and FIGARCH in favor of MSM for the Deutsche Mark/EURO, British Pound and Swiss Franc at long horizons. For the Japanese Yen, the MSM's relative mean squared errors are the lowest over all time horizons, but differences are not significant at the 5 percent confidence level. For the MAE criterion (cf. Table 8) results are very similar. We find an overall dominance of MSM against (FI)GARCH in four time series: Deutsche Mark/EURO, Canadian Dollar, British Pound and Swiss Franc. The one outlier here is the Japanese Yen, for which the baseline GARCH model has the lowest MAE. Note, however, that the best predictor for both the Canadian Dollar and Japanese Yen exchange rate volatility under MAE is the historical average and all time series models lead to a deterioration against this naive prediction (i.e. MAEs are larger than one) in the two cases.

So far our results essentially confirm the findings reported in Calvet and Fisher (2004). Now turn to a detailed assessment of the new linear forecasts based on GMM parameter estimates for Binomial models with $k > 10$ and Lognormal models with their continuous distribution of multipliers. The first question on which we would let our data speak is whether a higher number of components improves forecasting performance. Considering the development of MSEs and MAEs with increasing numbers of multipliers in Table 7 and 8, we find that the quality of forecasts improves in the majority of cases. Significant differences between ML and GMM based forecasts are found at the 5 percent confidence level for the Canadian Dollar, British Pound and Swiss Franc under the MSE criterion and for all time series except the Canadian Dollar under the MAE criterion. The effect of the number of multipliers on forecast quality is underscored by Fig. 2 which depicts MSEs for the British Pound/U.S.Dollar exchange rate 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 other exchange rates. Note that we would normally not expect GMM10 to dominate the optimal forecasts from ML10 because of the higher sampling variability of GMM estimates and the suboptimality of linear forecasts. Under this perspective, a somewhat inferior behavior of GMM10 against ML10 would not be surprising. However, in some cases, even at $k = 10$ GMM estimates dominate those from maximum likelihood, while for other series increasing k appears to amount to an overcompensation of the 'natural' disadvantage of GMM versus ML.

Tables 7 and 8 about here

Turning to our second objective, it seems worthwhile to remark that the Lognormal model has practically the same performance as the Binomial for each choice of k at all time horizons so that the added flexibility of the continuous distribution does not seem to provide an advantage over the simpler discrete structure. This practically identical performance out-of-sample is in harmony with their identical goodness-of-fit in-sample under the J statistic.

We have also investigated other asset prices along the above lines. For two stock market indices (the German DAX and the New York Stock Exchange Composite Index) and the price of gold we also estimated multifractal parameters and investigated their forecasting performance for the same in-sample and out-sample horizons. The in-sample fit of the multifractal model with $k = 10$ estimated via ML again outperformed the GARCH and FIGARCH models. In contrast, however, the J test based on GMM parameters estimates rather suggested that the underlying MSM model could be rejected with the pertinent

moment conditions for all three series. Nevertheless, the comparison of forecasting performance turned again out rather favorable for MSM: Under the MSE criterion, we found the estimated MSM models to dominate over GARCH and FIGARCH albeit with fewer significant cases (at the 5 percent level) than with exchange rates. For the DAX and the price of gold, GMM with higher numbers of multipliers also improved upon ML10 forecasts. For the NYSE index, in contrast, Bayesian forecasts were slightly better than GMM based ones. They also dominated GARCH and FIGARCH forecasts but were not significant at the 5 percent level. The picture was less positive under the MAE criterion, however: for both stock markets, all time series models performed worse than historical volatilities, while in the case of gold, FIGARCH dominated over all other models under the MAE criterion. These somewhat more mixed results certainly call for a more comprehensive analysis of the performance of multifractal models in different types of financial markets.

Before concluding this section, it appears worthwhile to point out another remarkable feature in Table 7 and 8: In quite a number of cases, we see a U-shaped development of MSEs and MAEs at the lower end of our forecast horizons, i.e. the performance increases when moving from the one-period horizon to multi-step forecasts over five days and deteriorates thereafter. While one could argue that such behaviour could be due to the lack of parameters modulating short-run dependencies in the MSM framework, the observation of a similar behaviour with FIGARCH in some cases (e.g. Canad. Dollar and UK Pound) casts doubts on such an explanation. Theoretically, all of our candidate models would, of course, lead to a monotonic decline of the quality of forecasts with forecasting horizon as shown in the Monte Carlo simulations in sec. 5 for MSM. While we do not have a ready explanation for this phenomenon at hand, it seems worthwhile to explore in future research whether this is a pervasive feature of financial data or whether it is a particularity of our data sets and the choice of out-of-sample time horizon.

Fig. 2 about here

6 Conclusion

The recent proposal of multifractal 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 parsimony 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 multifractal 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 (2006), we propose a GMM approach for estimating multifractal parameters, which is less computation intensive and which is still applicable in cases where ML becomes unfeasible. Monte Carlo experiments show that this GMM estimator is nicely behaved for various settings with discrete and continuous state space. Since Bayesian 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 the majority of foreign exchange rates investigated, this promise indeed materializes itself. While better performance of MF vis-à-vis (FI)GARCH had already been demonstrated by Calvet and Fisher (2004), we show that specifications with a larger state space could provide further improvements in forecasting accuracy. In contrast, replacing the simple discrete Binomial model by the continuous Lognormal MSM had practically no effect at all: both goodness-of-fit as well as forecasting capability were almost exactly the same for both models. The confirmation of the previous positive results and the further improvements by some alternative specifications documented in this paper underline that multiplicative volatility models with a hierarchy of components should be a promising area of further empirical research.

7 Acknowledgement

I very much appreciate the helpful comments and suggestions by an anonymous referee. The able research assistance by Ruipeng Liu and Timur Yusupov is also gratefully appreciated. The major part of the final version of this manuscript was completed during a visit at the School of Finance and Economics at University of Technology Sydney whose generous hospitality is also gratefully acknowledged. I am furthermore appreciative for financial support by the European Commission under STREP Contract no. 516446.

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. Sørensen (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) Multifractality 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 B. Mandelbrot (1997) Cowles Foundation Discussion Papers no. 1164-1166, Yale University, available at <http://www.ssrn.com>.
- Calvet, L., A. Fisher and S. Thompson (2006) Volatility Comovement: A Multi-Frequency Approach, *Journal of Econometrics*, **131**, 179-215.
- Chung, C.-F. (2002) *Estimating the Fractionally Integrated GARCH Model*. Mimeo: Academia Sinica.
- Diebold, F. and R. Mariano (1995) Comparing Predictive Accuracy, *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.

- 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.
- 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.
- Poon, S.-H. and C. Granger (2003) Forecasting Volatility in Financial Markets: A Review, *Journal of Economic Literature*, **41**, 468-539.
- 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 together with the second moment of the raw data. In order to apply GMM, we have to derive analytical solutions for this set of moment conditions. Let

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

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

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

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

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

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

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.

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

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

In general,

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

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 $(\frac{1}{2}\gamma_i)^2$, the above expression becomes:

$$\left(\frac{1}{2}\gamma_i\right)^2 \{2\ln(m_0)\ln(2-m_0) - \ln^2(m_0) - \ln^2(2-m_0)\}. \quad (\text{A5})$$

We, therefore, arrive at

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

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

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

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:

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

The calculations become slightly more involved when considering the auto-covariances of squared log increments:

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

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:

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

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

- (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

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

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

- (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 $(\frac{1}{2}\gamma_i)^2(\frac{1}{2}\gamma_j)^2$ but we also have to take into account that each of these terms appears two times in the expansion of $\eta_{t+1,1}^2\eta_{t,1}^2$ in (A8).

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

$$E[\eta_{t+1,1}^2\eta_{t,1}^2] = \left[\sum_{i=1}^k \left(\left(\frac{1}{2}\gamma_i \right) \sum_{j=1}^k \left(\frac{1}{2}\gamma_j \right) \right) + 2 \sum_{i=1}^k \left(\left(\frac{1}{2}\gamma_i \right)^2 \sum_{j=1, j \neq i}^k \left(\frac{1}{2}\gamma_j \right)^2 \right) \right] \chi. \quad (\text{A9})$$

For arbitrary T, we get:

$$E[\eta_{t+T,T}^2\eta_{t,T}^2] = \left\{ \sum_{i=1}^k \left(\frac{1}{2} \left(1 - (1 - \gamma_i)^T \right) \sum_{j=1}^k \frac{1}{2} \left(1 - (1 - \gamma_j)^T \right) \right) + 2 \sum_{i=1}^k \left(\frac{1}{4} \left(1 - (1 - \gamma_i)^T \right)^2 \sum_{j=1, j \neq i}^k \frac{1}{4} \left(1 - (1 - \gamma_j)^T \right)^2 \right) \right\} \chi. \quad (\text{A10})$$

Turning to the log innovations of the compound process,

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

we find:

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

and

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

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:

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

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

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

and

$$E [\mu_{t+T} \mu_t] = \prod_{i=1}^k \left\{ \left(1 - (1 - \gamma_i)^T \right) 0.5 m_0 (2 - m_0) + \left((1 - \gamma_i)^T + 0.5 \left(1 - (1 - \gamma_i)^T \right) \right) \left(0.5 m_0^2 + 0.5 (2 - m_0)^2 \right) \right\}. \quad (\text{A15})$$

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. $M_t^{(i)} \sim LN(-\lambda, s^2)$. In order to normalize the product of the k volatility components, we assume that $E[M_t^{(i)}] = 1$. This leads to:

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

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

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

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

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

For arbitrary lags T , one obtains accordingly:

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

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 $E[\eta_{t+1,1}^2\eta_{t,1}^2]$:

- (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 occurrence is $(\gamma_i)^2$. Solving the expectation in the non-zero case we obtain

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

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: $\kappa_1 = \left\{\sum_{i=1}^k (\gamma_i)^2\right\} \cdot 6s^4$

- (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

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

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

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

- (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 $\kappa_3 = \left\{ \sum_{i=1}^k \left((\gamma_i)^2 \sum_{j=1, j \neq i}^k (\gamma_j)^2 \right) \right\} 2 \cdot s^4$.

Putting all these components together, we end up with

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

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

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

In order to compute the auto-covariances 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 $E \left[(M_{i,t})^2 \right] = \exp(-2\lambda + 2s^2) = \exp(2\lambda)$

we arrive at

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

Furthermore,

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

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,

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

Appendix C: GARCH and FIGARCH Parameter Estimates

		μ	ρ	ω	β	α	φ	d	Logl	AIC	BIC
DEM/EUR	GARCH	0.001 (0.009)	0.025 (0.016)	0.010 (0.002)	0.904 (0.009)	0.076 (0.008)			-4480.00	8970.00	9002.21
	FIGARCH	0.001 (0.007)	0.028 (0.014)	0.011 (0.005)	0.685 (0.079)		0.218 (0.058)	0.515 (0.120)	-4476.18	8964.36	9003.01
	MF(10)								-4375.51	8755.02	8767.90
CND	GARCH	0.001 (0.003)	0.045 (0.016)	0.002 (0.000)	0.878 (0.013)	0.105 (0.011)			-56.93	123.86	155.96
	FIGARCH	0.001 (0.004)	0.043 (0.017)	0.002 (0.001)	0.601 (0.094)		0.250 (0.072)	0.488 (0.076)	-47.04	106.07	144.59
	MF(10)								-90.92	177.84	165.01
UKP	GARCH	0.001 (0.009)	0.054 (0.016)	0.004 (0.001)	0.939 (0.007)	0.053 (0.006)			-4320.50	8650.99	8683.09
	FIGARCH	0.001 (0.007)	0.055 (0.016)	0.012 (0.005)	0.652 (0.076)		0.274 (0.049)	0.433 (0.073)	-4324.65	8661.30	8699.82
	MF(10)								-4214.71	8433.43	8446.26
YEN	GARCH	-0.005 (0.009)	0.033 (0.016)	0.015 (0.003)	0.906 (0.014)	0.058 (0.008)			-4338.88	8687.77	8719.86
	FIGARCH	-0.002 (0.006)	0.031 (0.014)	0.038 (0.021)	0.540 (0.161)		0.398 (0.148)	0.253 (0.057)	-4336.64	8685.28	8723.79
	MF(10)								-4155.19	8314.39	8327.22
CHF	GARCH	0.006 (0.010)	-0.023 (0.015)	0.015 (0.003)	0.902 (0.009)	0.077 (0.008)			-5305.42	10620.84	10653.07
	FIGARCH	0.006 (0.010)	-0.020 (0.011)	0.015 (0.005)	0.701 (0.072)		0.204 (0.055)	0.546 (0.115)	-5301.96	10615.91	10654.59
	MF(10)								-5235.89	10475.79	10488.68

Note: GARCH and FIGARCH estimates are obtained on the base of quasi-maximum likelihood. In both models we included a constant and AR(1) component in the level of returns denoted by μ and ρ (we also removed linear dependence before computing multifractal 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): $h_t = \omega + \beta_1 h_{t-1} + (1 - \beta_1 L - (1 - \varphi_1 L)^d)x_t^2$. 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. For comparison, we also show the log likelihood, AIC and BIC values for the Binomial ($k = 10$) multifractal model (for the pertinent parameter estimates, cf. Tab. 6. Bold entries show the preferred model under the pertinent criterion.

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	1.298	1.299	1.299	1.210	1.230	1.259	1.398	1.398	1.399	1.324	1.377	1.393	1.500	1.501	1.500	1.475	1.494	1.500
FSSE	0.018	0.013	0.009	0.162	0.143	0.118	0.018	0.012	0.009	0.162	0.093	0.065	0.017	0.012	0.009	0.090	0.050	0.024
RMSE	0.018	0.013	0.009	0.185	0.159	0.125	0.018	0.013	0.009	0.179	0.096	0.065	0.017	0.012	0.009	0.093	0.050	0.024
$\bar{\sigma}$	1.000	0.998	0.998	0.995	0.996	0.997	0.993	1.001	0.999	0.984	0.997	0.998	1.002	1.006	1.000	0.981	0.999	0.995
FSSE	0.082	0.057	0.039	0.087	0.060	0.043	0.115	0.078	0.057	0.123	0.085	0.061	0.159	0.097	0.064	0.163	0.105	0.076
RMSE	0.082	0.057	0.039	0.088	0.060	0.043	0.115	0.078	0.057	0.124	0.085	0.061	0.158	0.097	0.064	0.164	0.105	0.076

Note: All simulations are based on a multifractal 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.

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	\tilde{m}_0	1.275	1.395	1.497	$\bar{\sigma}$	0.990	0.987	0.971
	FSSE	0.107	0.049	0.035	FSSE	0.117	0.159	0.192
	RMSE	0.110	0.049	0.035	RMSE	0.118	0.159	0.193
k=15	\tilde{m}_0	1.277	1.391	1.492	$\bar{\sigma}$	0.995	0.922	0.843
	FSSE	0.105	0.058	0.039	FSSE	0.335	0.399	0.505
	RMSE	0.107	0.059	0.040	RMSE	0.334	0.406	0.528
k=20	\tilde{m}_0	1.273	1.392	1.492	$\bar{\sigma}$	0.866	0.796	0.694
	FSSE	0.109	0.059	0.036	FSSE	0.431	0.515	0.602
	RMSE	0.112	0.060	0.036	RMSE	0.451	0.553	0.675

Note: The table exhibits GMM parameter estimates for a Binomial multifractal 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.

Table 3: GMM Estimation of Lognormal Model

	$\lambda=0.05$			0.10			0.150			0.05			0.10			0.15		
k=10	$\bar{\lambda}$		0.049	0.099	0.148				$\bar{\sigma}$	0.997	0.990	0.981						
	FSSE		0.022	0.025	0.028				FSSE	0.118	0.180	0.230						
	RMSE		0.022	0.025	0.028				RMSE	0.118	0.180	0.231						
k=15	$\bar{\lambda}$		0.051	0.099	0.149				$\bar{\sigma}$	0.928	0.898	0.808						
	FSSE		0.023	0.024	0.027				FSSE	0.315	0.432	0.477						
	RMSE		0.023	0.024	0.027				RMSE	0.323	0.443	0.514						
k=20	$\bar{\lambda}$		0.052	0.099	0.147				$\bar{\sigma}$	0.906	0.782	0.671						
	FSSE		0.023	0.025	0.026				FSSE	0.462	0.602	0.649						
	RMSE		0.023	0.025	0.023				RMSE	0.433	0.663	0.698						

Note: GMM parameter estimates for Lognormal multifractal processes 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.

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

T	$m_0 = 1.3$			$m_0 = 1.4$			$m_0 = 1.5$		
	ML	BL1	BL2	ML	BL1	BL2	ML	BL1	BL2
1	0.924 (0.017)	0.935 (0.026)	0.928 (0.018)	0.900 (0.023)	0.908 (0.025)	0.908 (0.024)	0.890 (0.029)	0.902 (0.031)	0.902 (0.032)
5	0.952 (0.014)	0.959 (0.019)	0.955 (0.015)	0.941 (0.017)	0.948 (0.018)	0.948 (0.018)	0.942 (0.019)	0.951 (0.019)	0.951 (0.020)
10	0.964 (0.013)	0.970 (0.015)	0.967 (0.013)	0.958 (0.014)	0.964 (0.015)	0.963 (0.015)	0.961 (0.016)	0.968 (0.015)	0.968 (0.016)
20	0.976 (0.011)	0.980 (0.012)	0.978 (0.011)	0.973 (0.013)	0.977 (0.012)	0.977 (0.012)	0.975 (0.013)	0.980 (0.012)	0.980 (0.013)
1	0.941 (0.054)	0.951 (0.050)	0.945 (0.051)	0.894 (0.080)	0.906 (0.072)	0.906 (0.074)	0.856 (0.105)	0.878 (0.090)	0.880 (0.096)
5	0.961 (0.050)	0.969 (0.045)	0.965 (0.048)	0.933 (0.075)	0.942 (0.066)	0.942 (0.069)	0.915 (0.101)	0.931 (0.081)	0.934 (0.091)
10	0.971 (0.047)	0.977 (0.041)	0.974 (0.044)	0.950 (0.070)	0.958 (0.059)	0.959 (0.064)	0.939 (0.096)	0.951 (0.073)	0.955 (0.086)
20	0.981 (0.042)	0.985 (0.035)	0.984 (0.039)	0.967 (0.063)	0.972 (0.051)	0.974 (0.057)	0.961 (0.088)	0.969 (0.062)	0.973 (0.079)

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 and Lognormal Models

	$k = 10$		$k = 15$		$k = 20$		$k = 10$		$k = 15$		$k = 20$	
	$m_0 = 1.3$	$m_0 = 1.5$	$m_0 = 1.3$	$m_0 = 1.5$	$m_0 = 1.3$	$m_0 = 1.5$	$\lambda = 0.05$	$\lambda = 0.15$	$\lambda = 0.05$	$\lambda = 0.15$	$\lambda = 0.05$	$\lambda = 0.15$
MSE	1	0.906 (0.043)	0.875 (0.051)	0.856 (0.124)	0.810 (0.173)	0.849 (0.131)	0.895 (0.046)	0.876 (0.080)	0.852 (0.106)	0.833 (0.156)	0.838 (0.109)	0.806 (0.166)
	10	0.943 (0.035)	0.943 (0.045)	0.892 (0.125)	0.873 (0.181)	0.885 (0.132)	0.938 (0.043)	0.944 (0.071)	0.892 (0.108)	0.898 (0.159)	0.882 (0.113)	0.876 (0.173)
	20	0.955 (0.032)	0.957 (0.042)	0.905 (0.125)	0.886 (0.181)	0.896 (0.132)	0.951 (0.041)	0.957 (0.066)	0.906 (0.108)	0.911 (0.157)	0.896 (0.114)	0.890 (0.171)
	100	0.979 (0.025)	0.980 (0.034)	0.931 (0.124)	0.912 (0.177)	0.924 (0.130)	0.976 (0.034)	0.978 (0.053)	0.933 (0.108)	0.933 (0.148)	0.925 (0.114)	0.915 (0.161)
MAE	1	0.919 (0.106)	0.844 (0.173)	0.891 (0.200)	0.796 (0.288)	0.865 (0.208)	0.907 (0.115)	0.836 (0.181)	0.888 (0.203)	0.830 (0.286)	0.854 (0.213)	0.775 (0.298)
	10	0.948 (0.098)	0.919 (0.162)	0.919 (0.198)	0.867 (0.293)	0.893 (0.209)	0.941 (0.107)	0.915 (0.165)	0.920 (0.202)	0.906 (0.283)	0.887 (0.215)	0.853 (0.298)
	20	0.958 (0.092)	0.938 (0.150)	0.930 (0.195)	0.887 (0.287)	0.904 (0.207)	0.953 (0.101)	0.936 (0.152)	0.932 (0.199)	0.926 (0.274)	0.899 (0.213)	0.875 (0.291)
	100	0.981 (0.067)	0.974 (0.107)	0.955 (0.179)	0.929 (0.256)	0.930 (0.194)	0.977 (0.075)	0.972 (0.107)	0.958 (0.182)	0.962 (0.236)	0.929 (0.198)	0.919 (0.257)

Note: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for both the Binomial (left-hand side with parameter m_0) and Lognormal (right-hand side with parameter λ) models. 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: Empirical Parameter Estimates

		Binomial					Lognormal				
		ML10	GMM5	GMM10	GMM15	GMM20	$\hat{\lambda}$ (SE) (J_{prob})	GMM5	GMM10	GMM15	GMM20
DEM/EUR	$\hat{\eta}_0$ (SE) (J_{prob})	1.248 (0.013)	1.345 (0.062) (0.236)	1.342 (0.062) (0.223)	1.1342 (0.062) (0.223)	1.342 (0.062) (0.223)		0.065 (0.025) (0.235)	0.063 (0.025) (0.222)	0.063 (0.025) (0.221)	0.063 (0.025) (0.221)
CND	$\hat{\eta}_0$ (SE) (J_{prob})	1.289 (0.013)	1.301 (0.074) (0.535)	1.303 (0.073) (0.555)	1.303 (0.073) (0.555)	1.303 (0.073) (0.555)	$\hat{\lambda}$ (SE) (J_{prob})	0.048 (0.025) (0.536)	0.049 (0.025) (0.555)	0.049 (0.025) (0.556)	0.049 (0.025) (0.556)
UKP	$\hat{\eta}_0$ (SE) (J_{prob})	1.261 (0.014)	1.353 (0.052) (0.366)	1.352 (0.052) (0.377)	1.352 (0.052) (0.378)	1.352 (0.052) (0.378)	$\hat{\lambda}$ (SE) (J_{prob})	0.068 (0.022) (0.366)	0.068 (0.022) (0.378)	0.068 (0.022) (0.379)	0.068 (0.022) (0.379)
YEN	$\hat{\eta}_0$ (SE) (J_{prob})	1.315 (0.016)	1.442 (0.044) (0.496)	1.438 (0.044) (0.468)	1.438 (0.044) (0.467)	1.438 (0.044) (0.467)	$\hat{\lambda}$ (SE) (J_{prob})	0.113 (0.026) (0.494)	0.111 (0.026) (0.466)	0.111 (0.026) (0.465)	0.111 (0.026) (0.465)
CHF	$\hat{\eta}_0$ (SE) (J_{prob})	1.232 (0.012)	1.455 (0.046) (0.002)	1.470 (0.045) (0.003)	1.471 (0.044) (0.003)	1.471 (0.044) (0.003)	$\hat{\lambda}$ (SE) (J_{prob})	0.122 (0.029) (0.002)	0.132 (0.029) (0.003)	0.132 (0.029) (0.003)	0.132 (0.029) (0.003)

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.). The nine moment conditions detailed in eq. 10 have been used, i.e. $\text{Mom}(T, q)$ in eq. 10 with $q = 1, 2$ and $T = 1, 5, 10, 20$ together with $E[x_t^2]$. 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 7: Multifractal Forecasts: MSE

horizon	GARCH	FIGARCH	Binomial				Lognormal		
			ML10	GMM10	GMM15	GMM20	GMM10	GMM15	GMM20
DEM/EUR	1	0.977	0.976	0.982	0.981	0.981	0.983	0.983	0.983
	5	0.975	0.978	0.970	0.967*†	0.969*†	0.967*†	0.969†	0.970†
	20	0.978	0.978	0.973	0.972*†	0.973	0.972*	0.973	0.974
	50	0.994	0.995	0.984*†	0.983*†	0.985*†	0.983*†	0.984*†	0.984*†
	100	1.007	1.014	0.993*†	0.993*†	0.995*†	0.992*†	0.994*†	0.995*†
CND	1	0.821	0.822	0.823	0.822	0.820	0.823	0.821	0.821
	5	0.812	0.809	0.822	0.813‡	0.809†	0.813‡	0.810‡	0.810‡
	20	0.837	0.826*	0.861	0.844‡	0.834‡	0.844‡	0.834‡	0.833‡
	50	0.901	0.864*	0.900	0.884*†	0.871*†	0.885*†	0.871*†	0.871*†
	100	0.921	0.860*	0.923	0.901*†	0.878*†	0.902*†	0.879*†	0.876*†
UKP	1	0.818	0.813*	0.830	0.829	0.827	0.831	0.829	0.829
	5	0.813	0.804*	0.812	0.809	0.805*†	0.810	0.805*†	0.806*†
	20	0.850	0.828*	0.838*	0.836*	0.827*†	0.837*	0.827*†	0.827*†
	50	0.893	0.841*	0.848*	0.855*	0.838*†	0.857*	0.838*†	0.837*†
	100	0.961	0.856*	0.862*	0.877*	0.849*†	0.878*	0.850*†	0.849*†
YEN	1	0.929	0.916	0.910	0.903	0.903	0.902	0.902	0.902
	5	0.956	0.953	0.946	0.949	0.950	0.948	0.949	0.949
	20	0.968	0.965	0.958	0.955	0.954	0.956	0.955	0.954
	50	0.992	0.981	0.974	0.974*†	0.975	0.975*†	0.975	0.975
	100	1.002	0.997	0.987	0.992*†	0.995	0.992*†	0.994	0.995
CHF	1	0.983	0.981	0.981	0.995	0.995	0.998	0.998	0.998
	5	0.979	0.984	0.967*†	0.965*†	0.965*†	0.964*†	0.964*†	0.964*†
	20	0.981	0.980	0.965*†	0.962*†	0.960*†	0.962*†	0.960*†	0.960*†
	50	1.007	1.005	0.977*†	0.971*†	0.968*†	0.971*†	0.968*†	0.968*†
	100	1.027	1.033	0.980*†	0.974*†	0.969*†	0.974*†	0.969*†	0.969*†

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 95% level, † an improvement against the FIGARCH model which is significant at the 95% level (applied only to multifractal models), and ‡ an improvement against the ML10 Binomial multifractal model which is significant at the 95% level (only applied to other multifractal models). All comparisons are based on the test statistics of Diebold and Mariano (1995). Note that the Diebold-Mariano test statistic is not necessarily monotonic in the relative MSEs shown in the table.

Table 8: Multifractal Forecasts: MAE

		GARCH	FIGARCH	Binomial				Lognormal		
				ML10	GMM10	GMM15	GMM20	GMM10	GMM15	GMM20
DEM/EUR	1	0.936	0.931*	0.945	0.926* [†]	0.921* [†]	0.920* [†]	0.928* [†]	0.922* [†]	0.921* [†]
	5	0.938	0.933*	0.936	0.916* [†]	0.908* [†]	0.907* [†]	0.916* [†]	0.908* [†]	0.908* [†]
	20	0.969	0.961*	0.950* [†]	0.934* [†]	0.922* [†]	0.921* [†]	0.935* [†]	0.923* [†]	0.922* [†]
	50	1.007	0.997	0.957* [†]	0.948* [†]	0.931* [†]	0.930* [†]	0.948* [†]	0.932* [†]	0.930* [†]
	100	1.035	1.037	0.961* [†]	0.960* [†]	0.940* [†]	0.938* [†]	0.960* [†]	0.941* [†]	0.939* [†]
CND	1	1.099	1.113	1.093 [†]	1.079* [†]	1.089* [†]	1.090* [†]	1.080* [†]	1.090* [†]	1.091* [†]
	5	1.081	1.098	1.068* [†]	1.057* [†]	1.072* [†]	1.075 [†]	1.057* [†]	1.072* [†]	1.075 [†]
	20	1.062	1.091	1.053 [†]	1.043* [†]	1.061 [†]	1.064 [†]	1.042* [†]	1.060* [†]	1.063 [†]
	50	1.054	1.097	1.040* [†]	1.035* [†]	1.058 [†]	1.062 [†]	1.035* [†]	1.057 [†]	1.061 [†]
	100	1.026	1.077	1.023 [†]	1.014* [†]	1.033 [†]	1.036 [†]	1.013* [†]	1.032 [†]	1.035 [†]
UKP	1	0.739	0.716*	0.732*	0.733*	0.716* [†]	0.714* [†]	0.734*	0.717* [†]	0.715* [†]
	5	0.747	0.719*	0.738*	0.742*	0.717* [†]	0.714* [†]	0.743*	0.718* [†]	0.716* [†]
	20	0.791	0.744*	0.763*	0.776*	0.737* [†]	0.733* [†]	0.779*	0.739* [†]	0.735* [†]
	50	0.858	0.766*	0.780*	0.807*	0.752* [†]	0.746* [†]	0.810*	0.755* [†]	0.749* [†]
	100	0.947	0.793*	0.803*	0.842*	0.772* [†]	0.764* [†]	0.845*	0.775* [†]	0.767* [†]
YEN	1	1.024	1.046	1.091	1.048 [†]	1.056 [†]	1.057 [†]	1.047 [†]	1.055 [†]	1.056 [†]
	5	1.029	1.052	1.102	1.053 [†]	1.066 [†]	1.068 [†]	1.050 [†]	1.062 [†]	1.064 [†]
	20	1.014	1.053	1.114	1.045* [†]	1.064 [†]	1.067 [†]	1.042 [†]	1.061 [†]	1.063 [†]
	50	1.007	1.057	1.128	1.046* [†]	1.071 [†]	1.074 [†]	1.043 [†]	1.067 [†]	1.070 [†]
	100	1.005	1.061	1.129	1.045* [†]	1.077 [†]	1.081 [†]	1.042 [†]	1.072 [†]	1.076 [†]
CHF	1	0.897	0.886*	0.894	0.888* [†]	0.882* [†]	0.881* [†]	0.892	0.886* [†]	0.886* [†]
	5	0.905	0.893*	0.893*	0.881* [†]	0.871* [†]	0.870* [†]	0.885* [†]	0.876* [†]	0.875* [†]
	20	0.947	0.927*	0.910* [†]	0.898* [†]	0.883* [†]	0.882* [†]	0.903* [†]	0.889* [†]	0.888* [†]
	50	1.000	0.977*	0.927* [†]	0.916* [†]	0.897* [†]	0.895* [†]	0.921* [†]	0.904* [†]	0.902* [†]
	100	1.038	1.032*	0.937* [†]	0.929* [†]	0.905* [†]	0.902* [†]	0.934* [†]	0.912* [†]	0.910* [†]

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 95% level, [†] an improvement against the FIGARCH model which is significant at the 95% level (applied only to multifractal models), and [‡] an improvement against the ML10 Binomial multifractal model which is significant at the 95% level (only applied to other multifractal models). All comparisons are based on the test statistics of Diebold and Mariano (1995).

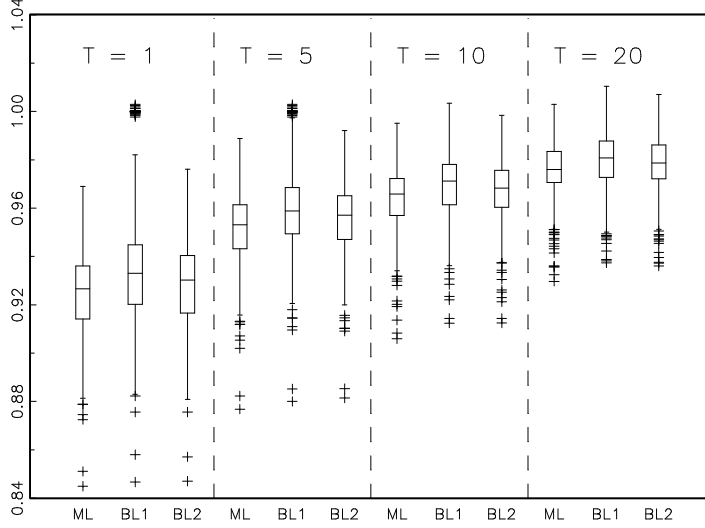


Figure 1: Box plot of the distributions of MSEs for different forecasting horizons over 400 Monte Carlo runs for the Binomial model with parameter $m_0 = 1.3$ (corresponding to the results given in the upper left corner of Table 4). The boxes show the median of the distribution surrounded by a box that spans the center half of the data set (the inter-quartile range). The whiskers extend to 1.5 times the inter-quartile range with the values outside this range identified as outliers. The appearance of the box plot is virtually identical for other parameter values, except for the accumulation of a few cases of relative MSEs equal to one for BL1 at small horizons (corresponding to an estimate $\hat{m}_0 = 1$ for which the multifractal model collapses with historical volatility).

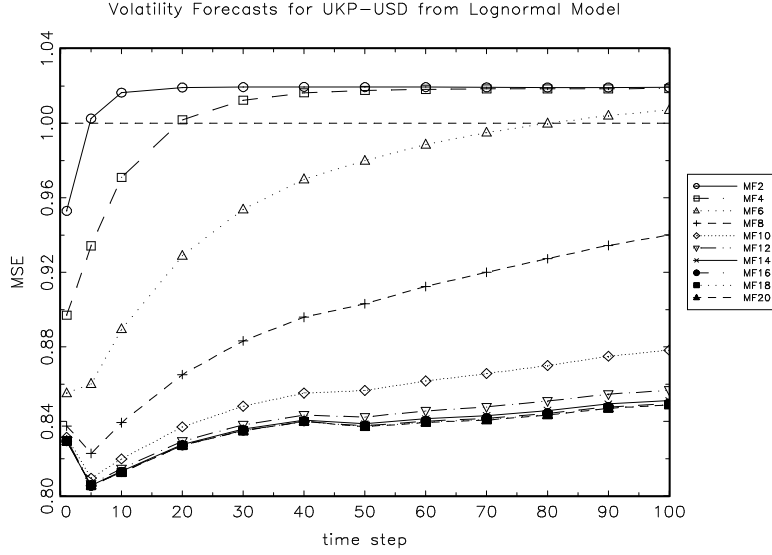


Figure 2: Mean squared errors over various time horizons of multifractal volatility forecasts of the British Pound / U.S. Dollar exchange rate. The underlying model is the Lognormal MSM with k ranging from 2 to 20 (a plot of the Binomial MSM results with the same numbers of multipliers would be undistinguishable from the one displayed here).