Kiel

Working Papers

Kiel Institute
for the World Economy

Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades

by Andrés E. Leövey and Thomas Lux

No. 1746| December 2011

Web: www.ifw-kiel.de

Kiel Working Paper 1746 | December 2011

Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades

Andrés E. Leövey and Thomas Lux

Abstract:

We study the well-known multiplicative Lognormal cascade process in which the multiplication of Gaussian and Lognormally distributed random variables yields time series with intermittent bursts of activity. Due to the non-stationarity of this process and the combinatorial nature of such a formalism, its parameters have been estimated mostly by fitting the numerical approximation of the associated non-Gaussian pdf to empirical data, cf. Castaing et al. [Physica D, 46, 177 (1990)]. More recently, an alternative estimator based upon *q*th order absolute moments has been introduced by Kiyono et al. [Phys. Rev. E 76 41113 (2007)]. In this paper, we pursue this moment-based approach further and develop a more rigorous Generalized Method of Moments (GMM) estimation procedure to cope with the documented difficulties of previous methodologies. We show that even under uncertainty about the actual number of cascade steps, our methodology yields very reliable results for the estimated intermittency parameter. Employing the Levinson-Durbin algorithm for best linear forecasts, we also show that estimated parameters can be used for forecasting the evolution of the turbulent flow. We compare forecasting results from the GMM and Kiyono et al.'s procedure via Monte Carlo simulations. We finally test the applicability of our approach by estimating the intermittency parameter and forecasting of volatility for a sample of financial data from stock and foreign exchange markets.

Keywords: Random Lognormal cascades, GMM estimation, best linear forecasting, volatility of financial returns.

JEL classification: C20, G12

Thomas Lux
Kiel Institute for the World Economy
24100 Kiel, Germany
Phone: +49 431-8814 278
E-mail: thomas.lux@ifw-kiel.de
E-mail: lux@bwl.uni-kiel.de

Andrés E. Leövey
Chair of Monetary Economics and International Finance at the University of Kiel
Olshausenstraße 40, 24118 Kiel, Germany
E-mail: eleoevey@economics.uni-kiel.de

Parameter Estimation and Forecasting for Multiplicative Lognormal Cascades*

Andrés E. Leövey† Thomas Lux,‡

December 4, 2011

Abstract

We study the well known multiplicative Lognormal cascade process in which the multiplication of Gaussian and Lognormally distributed random variables yields time series with intermittent bursts of activity. Due to the non-stationarity of this process and the combinatorial nature of such a formalism, its parameters have been estimated mostly by fitting the numerical approximation of the associated non-Gaussian pdf to empirical data, cf. Castaing *et al.* [Physica D, 46, 177 (1990)]. More recently, an alternative estimator based upon qth order absolute moments has been introduced by Kiyono *et al.* [Phys. Rev. E 76 41113 (2007)]. In this paper, we pursue this moment-based approach further and develop a more rigorous Generalized Method of Moments (GMM) estimation procedure to cope with the documented difficulties of previous methodologies. We show that even under uncertainty about the actual number of cascade steps, our methodology yields very reliable results for the estimated intermittency parameter. Employing the Levinson-Durbin algorithm for best linear forecasts, we also show that estimated parameters can be used for forecasting the evolution of the turbulent flow. We compare forecasting results from the GMM and Kiyono *et al.*'s procedure via Monte Carlo simulations. We finally test the applicability of our approach by estimating the intermittency parameter and forecasting of volatility for a sample of financial data from stock and foreign exchange markets.

JEL Clasification: C20, G12

PACS Clasification: 05.40.-a, 02.50.-r, 87.15.Ya, 47.11.St

Keywords: Random Lognormal cascades, GMM estimation, best linear forecasting, volatility of financial returns.

*Financial support from the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

†Chair of Monetary Economics and International Finance at the University of Kiel, Germany. Email: eleoevey@economics.uni-kiel.de.

‡Corresponding author. Chair of Monetary Economics and International Finance at the University of Kiel and Kiel Institute for the World Economy, Germany & Banco de España Chair in Computational Economics, University Jaume I, Castellón, Spain, Email: lux@bwl.uni-kiel.de.

I Introduction

Castaing *et al.* [1] have introduced the following seminal approach for the characterization of the probability density function (pdf) of velocity differences in fully developed turbulent flows:

$$P_{\lambda,\sigma_0}(x) = \int_0^\infty \frac{1}{\sqrt{2\pi}\lambda} \exp\left(-\frac{\ln^2(\sigma/\sigma_0)}{2\lambda^2}\right) P_\sigma\left(\frac{x}{\sigma}\right) \frac{d\sigma}{\sigma^2}, \tag{1}$$

where λ and σ_0 are positive parameters characterizing the pdf of the variable σ and P_σ is the pdf of a stationary and zero mean random variable x. Both λ and σ_0 determine not only the second moment associated with $P_{\lambda,\sigma_0}(x)$ but also the kurtosis. When $\sigma_0 = 1$ and $\lambda > 0$, $P_{\lambda,\sigma_0}(x)$ represents a mixture of distributions with a variance greater than one and excess kurtosis. On the other hand, in the limit $\lambda \to 0$, we observe $\sigma \to \sigma_0$ and $P_{\lambda,\sigma_0}(x)$ becomes a standard, mesokurtic Gaussian distribution. Equation (1), therefore, covers a whole spectrum of processes that can be used to describe more complex fluctuations than those originating from a Gaussian source. Stochastic processes corresponding to the pdf in eq. (1) could be of the form:

$$x_i \;=\; \exp\left(\varepsilon_i\right) \cdot \xi_i, \tag{2}$$

where $\xi_i \sim \mathrm{N}(0, \bar{\sigma}^2)$ and $\varepsilon_i \sim \mathrm{N}(\ln\sigma_0, \lambda^2)$ is independent of ξ.[1] The resulting intermittency generated from processes of the type of eq. (2) has been found to approximate quite well the fluctuations observed in data from various fields, such as from hadron collision [2], solar wind [3] as well as human heartbeat [4, 5] fluctuations, high-resolution satellite images [6] and, finally, in data of stock index [7] and foreign exchange rate [8, 9] fluctuations.

It is also well-known that the phenomenological approach by Castaing *et al.* [1] allows for non-linear scaling of absolute moments or multifractality of the underlying data-generating process. Considering a continuous-time process $X(t)$ with increments between times t and $t+l$:

[1]Though not expressed explicitly, $\bar{\sigma}^2$ can clearly be accommodated in eq. (1) via P_σ.

$\delta_l X(t) = X(t+l) - X(t)$, self-similarity of the associated pdf amounts to:

$$P(\delta_l X) \;=\; s^H P(s^H \delta_{sl} X), \tag{3}$$

with H the pertinent (Hurst) exponent for the renormalization of the pdf under changes of the scale s ($s > 0$). In order to account for multi-scaling in a series, a unique scaling exponent H is not appropiate so that one has to extend the previous approach. As originally suggested by Mandelbrot [10, 11], by replacing the constant factor s^H in eq. (3) by a random factor M_s depending on the scale, we obtain:

$$\delta_{sl} X(st) \;=_{Law}\; M_s \, \delta_l X(t). \tag{4}$$

It can be shown that such a scale-dependent multiplicative random modulation of $P(\delta_l X)$ leads to a non-linear scaling of absolute moments. The stochastic process of eq. (2) is an example of a process characterized by such non-linear scaling and, consequently, the pdf of eq. (1) is a potential outcome of such a stochastic extension of the notion of a self-similar process. Considering a cascade scale l and a finer scale sl ($s < 1$) in eq. (4), the pdf of eq. (1) indeed characterizes their relationship, with the random factor M_s being represented by the Lognormally distributed random variable $\exp(\varepsilon_s)$.

In the tradition of Castaing *et al.* [1], practical implementations of eq. (1) have mostly resorted to numerical approximations of the shape of the pdf minimizing the χ^2 statistics with respect to the empirical pdf to obtain parameter estimates. To avoid certain problems related to this method, Kiyono *et al.* [12] suggest an alternative procedure based on $\mathbb{E}[\,|x|^q]$, the qth order absolute moments. In this paper, we introduce a new alternative estimation procedure based on a Generalized Method of Moments (GMM) framework and demonstrate its superior performance. Our approach is motivated by a similar estimator that has been proposed in Lux [13] for the causal Markov-Switching Multifractal (MSM) model of Calvet and Fisher [14]. While our methodology is also based on moment matching, it differs from the approach of Kiyono *et al.* [12] in two important aspects: First, our moments are computed with respect to the joint distribution of x_i at different points of the cascade and, as such, they are exact moments of the underlying process. In contrast, the qth absolute moment of Kiyono *et al.* is computed from the marginal distribution $P_{\lambda,\sigma_0}(x)$ of eq. (1). This moment is exact for the multiplicative Lognormal model of eq. (2) with independent draws ε_i but not for a model with added cascade-like structure. Secondly, by using a GMM approach, we use more than one moment condition and systematically exploit the degree of uncertainty in various moments.

For the practical use of parameter estimates, we develop a forecasting scheme based on the best linear forecast algorithm that dispenses with the necessity to work with an approximation to the pdf of the coarse scale l process. We finally test its out-of-sample accuracy via both Monte Carlo simulations and provide an empirical application. The remainder of this manuscript is structured as follows. Section II introduces a detailed description of the process. Section III details the estimation methodology and compares our GMM estimates with the simple moment approach of Kiyono *et al.* [12] via Monte Carlo simulations. Section IV shows how the estimator behaves under misspecification concerning the number of cascade levels. Section V introduces the best linear forecast algorithm, and section VI presents empirical results for both parameter estimation and forecasting for a sample of financial data. Section VII concludes and the Appendix collects explicit formulas for the particular moments used in our GMM approach.

II The Process

To illustrate our procedure, we will first concretize the hypothesized data generating process. Although several ways to simulate intermittent fluctuations exist, we follow here the algorithm of Kiyono *et al.* [12]) for the generation of a cascade with n levels, and consider a fixed grid of 2^n points defining a sequence of uniform time intervals. In the first cascade step, we take the whole discrete set $[1, 2^n]$ and divide it into two sets of the same length. To each subset $[1, 2^{n-1}]$ and $[2^{n-1}+1, 2^n]$ we uniformly assign a random weight $M_1(k) = \exp[\omega_1(k)]$ $(k = 0, 1)$. In the next step, we further divide $[1, 2^{n-1}]$ and $[2^{n-1}+1, 2^n]$ into two new sets each, and assign in the same fashion the random weights $M_2(k) = \exp[\omega_2(k)]$ $(k = 0, 1, 2, 3)$. This procedure is repeated for $j = 1, \ldots, n$ leading to the final sequence of products of weights $\prod_{j=1}^{n} M_j(k)$ attached to the data points $\{1, \ldots, 2^n\}$. We obtain the Lognormal cascade as a compound process on the bounded interval $[1, 2^n]$ by multiplying the sum of the Lognormal weights with a Normally distributed random variable $\xi : x_i \doteq \left[\prod_{j=1}^{n} M_j\left(\left\lfloor \frac{i-1}{2^{n-j}} \right\rfloor\right)\right] \cdot \xi_i = \exp\left[\sum_{j=1}^{n} \omega_j\left(\left\lfloor \frac{i-1}{2^{n-j}} \right\rfloor\right)\right] \cdot \xi_i$, where $\lfloor\cdot\rfloor$ represents the floor function and $\xi_i \sim \mathrm{N}(0, \bar{\sigma}^2)$. It is common to select $\omega_j(\,\cdot\,) \sim \mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$ so that the sum of $\omega_j(\,\cdot\,)$ is $\mathrm{N}(n\,\tilde{\mu}, n\,\tilde{\sigma}^2)$ distributed. Eq. (II), therefore, fits into the framework of eq. (2) with $\sigma_0 = \exp(n\,\tilde{\mu})$ and $\lambda^2 = n\,\tilde{\sigma}^2$. Note, however, that here the ε_i are not independent draws but are correlated via the cascade structure. In the presentation of their estimator, Kiyono *et al.* assume that $\bar{\sigma} = 1$ and that $\sigma_0 = \exp(-\lambda^2)$, which in our context would be equivalent to require that $\omega_j(\,\cdot\,) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$ with $\lambda_0^2 = \frac{\lambda^2}{n} = -\tilde{\mu} = \tilde{\sigma}^2$. Figure 1 shows an illustration of a $n = 12$-level cascade with standardized factors $\omega_j(\,\cdot\,) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$. In the top three panels we exhibit draws at the first level $M_1(k) = \exp[\omega_1(k)]$ $(k = 0, 1)$, the second level $M_2(k) = \exp[\omega_2(k)]$ $(k = 0, 1, 2, 3)$ and the 10th level $M_{10}(k) = \exp[\omega_{10}(k)]$ $(k = 0, \ldots, 9)$, respectively, while in the fourth panel an outcome of the corresponding 'time series' $\{x_i\}_{i=1}^{2^n}$ is displayed.

(Please insert Figure 1 here)

To overcome the statistical difficulties that may arise from such a non-stationary construction, we go one step further and allow for an infinite sequence of independent cascades following the same generative principle, concatenating these series of sequences one after the other. This assumption leads to a sequence of data points $\{\ldots, m\,2^n + 1, m\,2^n + 2, \ldots, (m+1)\,2^n, (m+1)\,2^n + 1, (m+1)\,2^n + 2, \ldots\}$, with $m = 0, 1, \ldots$ an infinite sequence of repetitions of the same process of generation of a stochastic cascade of length 2^n. Our time series of measurements of the multiplicative Lognormal cascade process is consequently given by

$$x_t \doteq \exp\left[\sum_{j=1}^{n} \omega_j^{(m)}\left(\left\lfloor \tfrac{t-2^n\,(m-1)-1}{2^{n-j}} \right\rfloor\right)\right] \cdot \xi_t, \tag{5}$$

where again $\xi_t \sim \mathrm{N}(0, \bar{\sigma}^2)$.[2] The 'multipliers' $\omega_j^{(m)}(\,\cdot\,) = \ln M_j^{(m)}(\,\cdot\,)$ are assumed to be new draws for each newly started cascade, so that the process $\{x_t\}_{t=1}^{\infty}$ does not exhibit any obvious periodic structure, which distinguishes our algorithm from so-called cyclo-stationary processes (e.g. weather signals) that have clearly defined deterministic (e.g. sinusoidal) components (cf. Gardner *et al.* [15]).[3]

Note that, the time series in eq. (5) can also be described by eq. (2), taking into account the particular structure of the conditional distribution for the draws $\omega_j^{(m)}(\,\cdot\,)$ (or equivalently, ε_i) as imposed by the cascade structure. There are two ways to look at our infinite cascade process: First, under knowledge of the actual position, the joint distribution of observations at some time points $\{t_1, \ldots, t_k\}$ and $\{t_1+\mathrm{z}, \ldots, t_k+\mathrm{z}\}$ would clearly be different. This holds independently of whether any sequence would extend beyond the boundary of a single cascade or not. However, under ignorance of the current position, both sequences could be considered to be draws from a stationary process and would, thus, be characterized by the same joint distribution (and, of course, by the same moments). We adopt this second perspective and consider data samples being drawn from this infinite repetition of independent random cascade processes at arbitrary

[2] As the identification of the repetition number of the cascade is irrelevant for the variable ξ, this sequence can simply be indexed by time t.

[3] In our case the independent draws of the 'multipliers' have an effect that would be similar to reshuffling of the seasons in annual data.

starting points.[4] The bottom panel of Figure 1 shows a sample of 7,500 observations of the $\{x_t\}_{t=1}^{\infty}$ process as a result of concatenating three $n = 12$-level bounded cascades.

Despite the non-standard nature of the $\{x_t\}_{t=1}^{\infty}$ process (i.e. the application of a combinatorial construction in a time series context), our process is stationary under the second perspective (which corresponds to the limited information available to the empirical researcher), and many standard procedures for statistical inference become now available. On the contrary, when considering the original process from eq. (II) over a bounded interval only, the non-stationarity of the process would have followed trivially. As a consequence, standard 'regularity conditions' (cf. Harris and Mátyás [16]) for many standard methods of statistical inference would have been violated. As we will see in the following, our approach allows us to compute exact conditional and unconditional moments for our GMM estimation procedure that universally apply to any set of observations arising from the process $\{x_t\}_{t=1}^{\infty}$ of eq. (5). Due to the analytical structure of these moments (cf. Appendix A), standard regularity conditions such as differentiability and boundedness of the moments are now clearly satisfied.

III Estimation Methodology

GMM is a very general statistical approach for estimation of the parameters of a model. Given a set of analytical moments, the vector of parameter estimates, say φ, is obtained as the result of the minimization of an objective function of the following form:

$$\widehat{\varphi}_T = \arg\min_{\varphi\in\Phi} f_T(\varphi)' \, \Omega_T f_T(\varphi),$$

with Φ being the parameter space, $f_T(\varphi)$ the vector of differences between a set of sample and analytical moments, and Ω_T a positive definite and possibly random weighting matrix; cf. Hansen [17]. Using log-absolute moments in the implementation of $f_T(\varphi)$, Lux [13] has applied this estimation method to the iterative MSM model, demonstrating that it provides reliable parameter estimates even for small sample sizes. In the following, we will apply a similar approach in our analysis of multiplicative Lognormal cascades.

Let us consider the log-absolute difference $\zeta_{t,\ell} \doteq \ln|x_t| - \ln|x_{t-\ell}|$, with ℓ representing the lag at which the difference is taken. In order to exploit the scaling properties of the cascade process, we select as in Lux [13] autocovariances of the overlapping log differences $\zeta_{t,\ell}$, $\zeta_{t+1,\ell}, \ldots$. A closer look at these yields:

$$\begin{aligned} \zeta_{t,\ell} &\doteq \ln|x_t| - \ln|x_{t-\ell}| \\ &= \sum_{j=1}^{n} \left[\omega_j(t) - \omega_j(t-\ell)\right] + \ln|\xi_t| - \ln|\xi_{t-\ell}|. \end{aligned} \tag{6}$$

[4] For this reason, we will drop from this point on the notation of m in $\omega_j^{(m)}(\cdot)$ and we will identify the 'multipliers' simply by $\omega_j(t)$.

As one can see, these log-absolute differences remain unaffected by $\bar{\sigma}$, the scale factor in eq. (5) that is typically needed to match the order of magnitude of the data under scrutiny. Our moment conditions will consist, for $p = 1, 2$, of the following set of moments:

$$\mathfrak{Mom}(\ell, p) = \mathbb{E}[\zeta^p_{t+\ell,\ell} \cdot \zeta^p_{t,\ell}], \tag{7}$$

together with a raw moment like $\mathbb{E}[x_t^2] = \bar{\sigma}^2$ for the identification of $\bar{\sigma}$. With this device, the resulting estimates of $\bar{\sigma}$ from GMM are identical to the sample standard deviation of the $\{x_t\}_{t=1}^{\infty}$ process and the covariance matrix between both sets of parameters would be block diagonal. Appendix A contains the explicit derivations for the moments introduced in eq. (7).

The simpler estimator $\widehat{\lambda}^2_q$ of Kiyono *et al.* [12, eq.(5)] is derived from $\mathbb{E}[\,|x|^q]$, the absolute moment of power q for the marginal pdf of eq. (1):

$$\widehat{\lambda}^2_q = \frac{2}{q(q-2)}\left[\ln\left(\frac{\sqrt{\pi}\,\mathbb{E}[\,|x|^q]}{2^{q/2}}\right) - \ln\Gamma\left(\frac{q+1}{2}\right)\right] \tag{8}$$

(where $q \neq 0, 2$) after standardizing the mentioned pdf by setting $\sigma_0 = \exp(-\lambda^2)$ in eq. (2). Note that eq. (1) is <u>not</u> the pdf of the ensemble of observations from a cascade process as it applies strictly only for independent draws of ε_i in eq. (2). Given the stage of the cascade, it, however, characterizes the marginal pdf of the process at any position t. Since $\widehat{\lambda}^2_q$ is not derived from the exact pdf of the cascade process, it will in all likelihood be an inconsistent estimator for such a model. As we will see, this conjecture is confirmed by our Monte Carlo simulations below. In our cascade-setting, the mentioned standardization implies, on the other side, that $\omega_j(t) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$ in eq. (5), and so as stated before, $\widehat{\lambda}^2_q$ captures the overall intermittency $\lambda^2 = n\lambda_0^2$. In practice, $\mathbb{E}[\,|x|^q]$ is calculated from a zero-mean unit-variance series so that before being able to compute this moment, the series $\{x_t\}_{t=1}^{T}$ must be detrended and consequently standardized by the ad-hoc sample standard deviation estimator $\widehat{\sigma}$. The value of q is arbitrary *a priori*, but as the authors suggest, one can numerically compare the root mean squared errors (RMSE) of $\widehat{\lambda}^2_q$ under different q and select the optimal one.

We proceed by reporting results of several Monte Carlo studies designed to explore the applicability of our GMM estimator and its performance in comparison to $\widehat{\lambda}^2_q$ of eq. (8). To this end, we apply Kiyono *et al.*'s [12] standardization, choose $q = 0.5$, and leave the results for $\widehat{\lambda}^2_q$ under alternative choices of q for Appendix B.[5]

We apply both estimators for various sample lengths T_i, namely $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. The GMM procedure aims at exploiting the intermittency at different cascade levels, and therefore, the moments in eq. (7) depend on the choice of the number and values of lags ℓ. After many trials, for which results are not presented here, we find that using three lags leads to a good compromise between computational speed and quality of the estimates. In

[5] The table in Appendix B presents only the results for $\lambda_0^2 = 0.05$ under different q but is sufficient to show the relative increase in efficiency in terms of RMSE when $q = 0.5$ is used with a low number of cascade steps n. As n increases, the $\widehat{\lambda}^2_q$ estimator behaves very similar under any q due to the bias explained hereafter.

short, the values $\ell = 1, 14, 64$ are chosen to capture the intermittency generated by the last seven cascade levels.[6] We use the iterative GMM version instead of the simple two-step GMM scheme, where a new weighting matrix Ω_T is computed and the whole estimation process is repeated until convergence of both the parameter estimates and the weighting matrix is obtained (Hansen *et al.* [18]).

(Please insert Table 1 and 2 here)

Table 1 and 2 present the outcomes from both methods, where we have normalized the results for $\widehat{\lambda}_q^2$ in Table 2 by the total number of cascade levels n for better comparability. As we can infer from both Tables, the $\widehat{\lambda}_q^2$ estimator starts out with a slight advantage over GMM in terms of RMSE at relatively small parameter values λ_0^2 or low cascade levels n. This appear plausible as these scenarios are closest to the case of independent ε_i for which $\widehat{\lambda}_q^2$ would be a consistent estimator. For a fixed number of cascade levels n, however, the bias of the simple moment estimator $\widehat{\lambda}_q^2$ increases considerably the higher λ_0^2 is. Equivalently, for a fixed λ_0^2, the bias of $\widehat{\lambda}_q^2$ increases with a higher number of cascade levels n, though at a slower pace. GMM, on the other hand, shows only very slight increases of RMSE when either λ_0^2 or n increases so that its advantage becomes more and more pronounced for high levels of intermittency and high number of cascade levels. The decrease of biases and sampling variability appears also much slower with increasing sample size for the $\widehat{\lambda}_q^2$ estimator than for the GMM estimator. While the latter seems to nicely satisfy squared-root consistency when doubling the sample size from T_1 to T_2 and from T_2 to T_3, the former almost never does so. For high λ_0^2 or n its RMSEs appear almost constant across sample sizes.

Table 1 and 2 also present the results for $\bar{\sigma}$. In Table 2 we show simply the sample standard deviation (as Kiyono *et al.*'s method only provides an estimate of $\widehat{\lambda}_q^2$). One observes that the GMM estimator in Table 1 agrees closely with the sample standard deviation in Table 2 which actually is to be expected under the block-diagonal structure of the covariance matrix of the moment conditions. In either case, the estimator seems to be more biased the higher both λ_0^2 and n, the number of cascade levels. As in Lux [13], this might be due to the fact that a higher λ_0^2 and n generate enhanced fluctuations of the product of volatilities, which might interfere with the estimation of the constant scale factor $\bar{\sigma}$. Somewhat surprisingly, Table 1 shows that in the case $\lambda_0^2 = 0.15$ and $n = 16$, an apparent violation of the squared-root consistency can be perceived for $\bar{\sigma}$ when doubling the sample size T. Additional analysis undertaken with larger sample sets, left aside in this paper for brevity, indicate that this behavior is only restricted to sample sizes T_1, T_2, and T_3, and we recover a 'nice' behavior for increasingly larger samples. The reason seems to be that: initially, for small sample sizes (T_1) in relation to the sample size of a bounded cascade with $n = 16$, the probability of encountering a major node at which many switches occur will be low so that the behavior of moments is quite regular. However, with medium sample sizes like T_3, the probability of meeting a node with many switches becomes much

[6] The value $\ell = 14$ is as good as any in $(8, 16)$.

larger and the remaining data points in that same sample may not be enough to compensate for this disruption. Thus, if n is very large, preasymptotic fluctuations of the quality of estimated parameters cannot be excluded even for data sets in the range of 10,000 observations.[7] We note, however, that this apparently only happens for both very large n together with a high intermittency parameter λ_0^2.

Nevertheless, with this particular caveat notwithstanding, the complete set of our simulations indicate that the GMM estimates are generally as well-behaved as they are expected to be.

IV Uncertainty of the number of cascade components

Our initial study (Tables 1 and 2) on GMM performance has been based on the assumption that we have exact knowledge about the relevant number of cascade steps. The lack of such knowledge introduces an additional source of uncertainty. To investigate the effect of such uncertainty we extend our previous analysis and generate samples of size $T = 10,000$ for a cascade of $n = 11$ levels with different values of λ_0^2. We apply then our GMM estimator for a range of hypothesized cascade levels from 8 to 14 and contemplate the change in the estimated value as well as in the objective function. Results of pertinent Monte Carlo simulations are presented in Table 3.

(Please insert Table 3 here)

As it turns out, the additional uncertainty does not impede the correct estimation of the intermittency parameter even if the cascade generating the data has a higher or lower number of components than the one used for estimation. As it can be seen from Table 3, the absolute percentage difference (APD) between the estimates is at most three percent, which occurs with low λ_0^2. In addition, the difference of the objective function compared to that of $n = 11$ increases with the difference of the assumed cascade steps from the true $n = 11$, with more change happening for lower than for larger values. The difference is more pronounced the higher λ_0^2 is. However, a large deviation between the minimized objective functions does not directly carry over to APDs, which appear to be smaller throughout the range of n considered. The reason for this is that our moment conditions focus on capturing the fluctuation generated at a cascade of size 2^ℓ so that for any higher cascade level, the number of anticipated switches decreases proportionally and, eventually, when the length of the cascade level is larger than the sample size T the number of added switches is at most one per level. As such, higher cascade levels add very little to the analytical moments, whereas the estimate of $\bar{\sigma}$ absorbs higher level cascade components to some extent and, therefore, shows a bias that increases with n (the same observation has been made for the MSM model in Lux [13]). In conclusion, our GMM

[7] In principle, it is quite plausible that the range of preasymptotic volatility of estimates scales with the cascade level n.

procedure seems to provide reliable estimates of the intermittency generating parameter λ_0^2 even with uncertainty regarding the number of cascade steps.

V Forecasting Methodology

Lux [13] has introduced best linear forecasts to predict out-of-sample fluctuations of realizations of the causal Markov-switching multifractal process of Calvet and Fisher [14, 19]. Given a zero-mean weakly stationary process $\{Y_t\}$, the standard approach for construction of best linear h-step forecasts amounts to predicting the realization of the process at time horizons h by

$$\widehat{Y}_{t+h} \,=\, \sum_{i=1}^{t} \phi_{t\,i}^{(h)}\, Y_{t+1-i} \,=\, \Phi_t^{(h)}\, \mathbf{Y}_t, \tag{9}$$

with the vector of weights $\Phi_t^{(h)} = (\phi_{t\,1}^{(h)}, \phi_{t\,2}^{(h)}, \ldots, \phi_{t\,t}^{(h)})'$ being any solution to the system $\Gamma_t\, \Phi_t^{(h)} = \gamma_t^{(h)}$, where $\Gamma_t = [\gamma(i-j)]_{i,j=1,\ldots,t}$ is the variance-covariance matrix, and $\gamma_t^{(h)} = (\gamma(h), \gamma(h+1), \ldots, \gamma(t+h-1))'$ denotes the vector of t elements of lag h auto-covariances and beyond ([20]).

One consequence of the periodicity of size 2^n introduced to the series by the concatenation of cascades is that the long memory of the process is bounded by the length of that period. As such, its autocovariances would rapidly drop to zero after lag 2^n so that the inclusion of all available data, as one might consider when dealing with long-memory processes, should have no practical influence on the resulting forecasts beyond the maximum lag.

In the implementation of the procedure involving eq. (9), we use the iterative algorithm developed by Brockwell and Dahlhaus [21, algorithm 5]. For the implementation, one needs the autocovariances 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 $\mathbb{E}[\, x_{t+\ell}^2\, x_t^2]$. With this in mind, we define a series of zero-mean squared fluctuations:

$$Y_t \,\doteq\, x_t^2 - \mathbb{E}[\, x_t^2\,] = x_t^2 - \widehat{\sigma}^2, \tag{10}$$

where $\widehat{\sigma}$ is the estimate of the scale factor $\bar{\sigma}$ in eq. (5). Also, $\widehat{\sigma}^2$ appears only in the mean value of eq. (10), but it drops from the coefficients $\phi_{t\,i}^{(h)}$. Appendix C presents the pertinent formulae for the variance and auto-covariances of the intermittency generating part of a series of length T.

We again explore the performance of our proposed methodology via Monte Carlo simulations, assuming that one knows the exact number of cascade levels in the data generating process. We restrict ourselves to one sample of size $T = 7500$, where we use the first 5,000 entries for in-sample parameter estimation and the remainder for an assessment of the out-of-sample forecasting performance in terms of mean squared error (MSE) and mean absolute error (MAE). Both MSE and MAE are standardized relative to the MSE and MAE of the most naive forecast, that is, the sample variance or squared "historical volatility" of a random-walk (RW) during the in-sample period, for the same sample, so that values below one indicate an improvement against the constant variance forecast based on a RW. A closer look at Table 4 shows that forecasts based on the $\widehat{\lambda}_q^2$ estimate plus the sample standard deviation are fairly similar to those based on GMM. Indeed, both forecasts outperform the naive forecast at similar rates. This advantage of the model-based forecasts over the naive predictor initially increases with the degree of intermittency of the time series, i.e. λ_0^2, but declines at the upper end of the spectrum of values used in our Monte Carlo study. It seems worthwhile emphasizing that we have kept the in-sample period constant at $T = 5,000$ at all times. This means that the information used to estimate the parameters has not been updated over the out-of-sample period. The increase in biases in the estimate of $\bar{\sigma}$ with increasing n does not appear to constitute a major obstacle for the prediction of future fluctuations.

(Please insert Table 4 here)

The U-shape of the prediction accuracy with varying λ_0^2 is reminiscent of similar observations in Lux [13]. Apparently, there are two opposite forces at work here: with small λ_0^2, an increase of this parameter leads to a better forecasting performance simply because, then, the fluctuations become more pronounced, and the series shows more of a deviation from a random walk, while at very high values ($\lambda_0^2 = 0.15$) these fluctuations become more intermittent and less predictable given the sample size T available for estimation. This may also explain why the inconsistent estimator of Kiyono *et al.* [12] is even marginally better than the GMM estimator at higher values of λ_0^2: its strong downward bias (cf. Table 2) leads to smoother forecasts which on average might lead to somewhat smaller errors than forecasts based on a more accurate estimate.[8]

(Please insert Table 5 here)

Next, we consider the case with added uncertainty on the number of cascade levels n of the series. Table 5 presents the forecasting results of a series generated by $n = 11$ and $T = 5,000$ splitted into two subsamples of 2,500 for estimation and forecasting. The process has been analyzed for a sequence of cascade levels ranging from 8 to 50, for which a GMM estimation and subsequent out-of-sample exercises have been conducted. As we can see the MSEs and MAEs stabilize at the 'true' n: while using too low a number of cascade steps leads to suboptimal performance in forecasting, using even unbounded high levels of n is almost completely harmless (except for a slightly higher variability of MSEs and MAEs around their means as indicated by their standard errors). Since, in practice, n will be typically unknown, these results speak in favor of using deliberately large hypothesized values of n in empirical research.

[8]While the lack of an advantage of the GMM estimates compared to the simple (inconsistent) moment estimator of Kiyono *et al.* might appear disappointing, we should note that the use of the later for forecasting already implies quite some effort in computing exact moments (to implement eq. (10)). Hence, at this point a reliance on the simple estimator would be moot anyway.

VI Empirical Evidence

Starting with Ghashghaie *et al.* [8], a fair amount of effort has been spent particularly by physicists [12, 7, 22, 23, 24, 25] on the analogy between turbulence and financial markets. In this line of thought, the main goal has been to retrieve the functional form of the relationship among pdfs of price changes at different scales. Instead, we focus in this paper on forecasting turbulence (volatility) on the base of a cascade model of intermittent fluctuations.

Our analysis is based on data from seven different foreign exchange markets: the Canadian Dollar (CND), the Japanese Yen (YEN), the Swedish Krona (SEK), and the Swiss Franc (CHF), all against the U.S. Dollar, and on the other hand, the Australian Dollar (AUD), the Deutsche Mark- extended by the EURO since 1999- (DEM/EUR), and the British Pound (UKP). Further, we have analyzed the price of gold in U.S. Dollars. All time series start on the 2nd of January of 1979 and extend until the 2nd of July of 2010.

Due to the slight variation in the number of active trading days among markets, we use the first $\approx$ 21.8 years of data for in-sample estimation and leave the remaining years for out-of-sample evaluation of volatility forecast. This gives exactly 5,500 in-sample observations for each asset and never less than 2000 observations for the out-of-sample analysis.

Though not compulsory for our GMM methodology, we employ the mentioned standardization for ω_j so that $\omega_j(t) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$.[9] Table 6 reports in-sample parameter estimates for the intermittency parameter λ_0^2 and for $\bar{\sigma}$, together with their standard errors and the corresponding probability of the Hansen's test statistics $J_T = f_T(\widehat{\varphi})' \, \Omega_T f_T(\widehat{\varphi})$, where the estimation procedure has been repeated for $n = 8, \ldots, 20$. At a significance level of 0.05, the J-test statistics would allow to reject the multifractal cascade as the data generating process only for the Swiss Francs (CHF), on the base of our chosen moment functions. Further, we can see that while the number of cascade levels at which the lowest objective function was obtained varies from asset to asset, the Maximum APD among different n remains always below 2.5 %, except for the case of the Swiss Franc (CHF).

(Please insert Table 6 here)

Finally, the forecasting procedure has also been applied for all model specifications $n = 8, \ldots, 20$. The forecasting results under the MSE and MAE criteria for the highest cascade level are presented in Table 7 and 8, respectively. Also presented are the forecasting results of a fitted GARCH(1,1) model for each asset.[10] Though we abstain from presenting all details here, we find that the differences in forecast ability for different n have been marginal, with the forecasts for $n = 20$ being very close to that of all other n. This agrees with results reported in Lux [13] and Calvet and Fisher [19] who arrive at a similar conclusion regarding the saturation

[9] The set of suggested moments in our GMM procedure allows the alternative specification $\bar{\sigma} = \exp(-n\,\tilde{\mu})$, together with an additional estimator for $\tilde{\sigma}^2$, the variance of each ω_j in eq. (5). In this case, however, the covariance matrix of the parameters would no longer be block diagonal.

[10] Estimation results for the GARCH(1,1) are not presented here but are available upon request.

of forecasting performance beyond a certain threshold.

(Please insert Tables 7 and 8 here)

As one can observe in Table 7 and 8, our procedure performs quite well for most of the series, particularly for the MSEs where results are almost always statistically significantly better than RW forecasts at the 99% level of the test statistic for nested models of Clark and West [26, 27].[11] In some cases, statistical significance was found even when the reported MSE is slightly higher than one. The reason is that if the two forecasts are highly correlated the series with a lower variance of squared-forecasting errors should be preferred even if its mean is slightly worse. As concerns comparison between the $n = 20$ Lognormal cascade and GARCH(1,1), most often GARCH(1,1) has a slight advantage for the smaller lags, while the cascade model provides better forecasts for larger horizons. Given that the GARCH(1,1) model has only short-term dependence (and two estimated parameters for the exact structure of this short-term dependence), while the cascade model is designated to capture dependence over larger horizons, these results very much coincide with our expectations. However, differences in both directions are mostly non-significant under the modified Diebold and Mariano [28] test statistic (Harvey *et al.* [29]) at the 95 % level. Nevertheless, it is noteworthy that the cascade forecasts typically dominates over the longer horizons which shows the added-value of the long-term dependence for the multiplicative structure of volatility.

The MAE results in Table 8 display more significant results, with the $n = 20$ Lognormal cascade forecasts performing almost always better than GARCH(1,1), with a statistical significance of 95%. Note that under the MAE criterion, the cascade models mostly also dominates over GARCH(1,1) at short horizons, and even significantly so. There is also a difference in the significance of results against the RW-based forecast in this table with a somewhat smaller number of improvement for both the cascade and the GARCH(1,1) models. This, however, may be based on the nature of the modified Diebold and Mariano test employed here, as the MAE logic of averaging L_1 distances precludes us from employing the Clark and West [26, 27] adjustments for nested models. Overall, while the results for MAEs are not entirely homogeneous they appear quite encouraging particularly as concerns potential improvements against the GARCH(1,1) benchmark.

VII Conclusion

We have proposed in this paper a GMM approach for estimation of Lognormal cascade processes, which compares favorably with the recent $\widehat{\lambda}_q^2$ moment estimator developed by Kiyono *et al.*. Our numerical analysis suggests that the GMM based estimator is indeed consistent and asymptotical Normally distributed. Further, our methodology allows us to retrieve the cascade

[11]Best-linear and GARCH forecasts can be considered to nest the naive forecasts from a random-walk (RW). Best-linear multifractal forecasts and GARCH forecasts are, however, not nested so that the non-adjusted version of the test applies.

parameter value even when the number of levels of the cascade is unknown.

To apply the estimates obtained for forecasting the future evolution of a cascade, we have developed a forecast methodology based on the Levinson-Durbin algorithm for best linear forecasts. Our methodology circumvents the statistical problems related to the definition of a cascade process on a bounded interval by allowing for a new initialization of the process each time the endpoint of the cascade is reached. We also show that the size of the interval, i.e. the number of cascade steps, has virtually no influence on the estimated intermittency parameter. The predictive power of forecasts based on past realizations is similarly relatively insensitive to the number of steps beyond some threshold. Somewhat surprisingly, using more precise GMM estimates yield virtually the same forecasting performance as combining the inconsistent Kiyono *et al.*'s estimator with best linear forecasts based on accurate moment conditions (a feature probably due to a lucky interplay between the bias of the estimator by Kiyono *et al.* and the volatility of forecasts for different parameters and cascade levels).

The applicability of our procedure is confirmed by an extensive simulation analysis. Our empirical application consists in the estimation of the intermittency parameter and the forecasting of volatility for various foreign exchange markets and the gold market. Our results suggest that cascade models, even with their grid-bound nature of volatility components, capture a non-trivial part of the variability of price fluctuations. This support previous findings for the causal Markov-Switching Multifractal model (Calvet and Fisher [14]). However, our use of the combinatorial structure of models of turbulence in physics demonstrates that similar results can be obtained without giving up the time-honored generating mechanisms for turbulent flows in statistical physics. Our approach might, therefore, be valuable for observations of turbulent processes in other areas that are not easily cast into a causal time series framework.

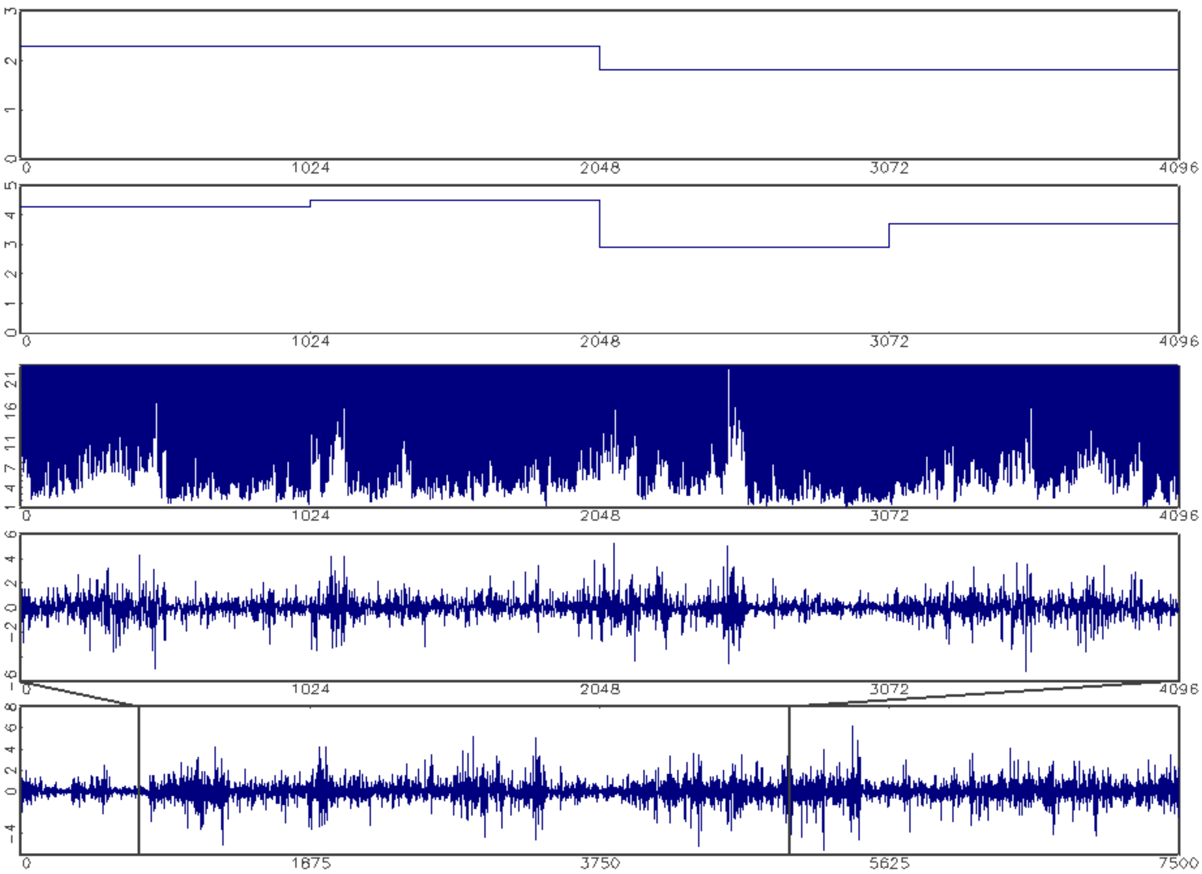

Figure 1: Illustration of one sample of the $\{x_t\}_{t=1}^{\infty}$ process of eq. (5) with $n = 12$. From top to bottom: The first level of draws of Lognormal random variables, the second level, the 10th level, the corresponding bounded $\{x_i\}_{i=1}^{2^n}$ process according to eq. (II), and a sample of 7500 points of the stationary $\{x_t\}_{t=1}^{\infty}$ process of eq. (5). The standardization of the pdf in eq. (1) suggested by Kiyono *et al.* for the construction of the cascades has been followed, with parameter value $\lambda_0^2 = 0.035$. For better visualization of the samples in the last two panels, $\bar{\sigma} = 2^{-n}$ was chosen to scale the overall magnitude of intermittency.

Table 1: Monte Carlo results for the GMM estimator.

		$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.025$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.1$			$\lambda_0^2 = 0.15$		
n		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
8	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.024	0.025	0.026	0.047	0.049	0.050	0.094	0.098	0.099	0.142	0.146	0.148
	FSSE	.009	.008	.005	.012	.009	.005	.017	.010	.007	.019	.013	.009	.023	.016	.011
	RMSE	.009	.008	.005	.012	.009	.005	.017	.011	.007	.019	.013	.009	.024	.016	.011
	$\bar{\sigma}$	1.000	1.002	1.000	0.999	1.003	1.000	1.000	0.995	0.998	1.002	1.007	0.999	0.969	0.967	0.993
	FSSE	.039	.027	.019	.061	.044	.028	.094	.066	.048	.161	.133	.073	.214	.158	.139
	RMSE	.039	.027	.019	.061	.044	.028	.094	.066	.048	.161	.133	.073	.216	.161	.139
10	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.023	0.026	0.026	0.047	0.048	0.049	0.094	0.098	0.098	0.142	0.146	0.148
	FSSE	.009	.008	.005	.013	.009	.006	.016	.011	.007	.021	.014	.009	.024	.015	.011
	RMSE	.009	.008	.005	.013	.009	.006	.017	.011	.007	.021	.014	.009	.025	.016	.011
	$\bar{\sigma}$	0.993	1.001	1.001	0.999	1.000	0.997	0.995	0.987	0.987	0.986	0.981	0.991	0.950	0.934	0.976
	FSSE	.069	.047	.036	.105	.077	.056	.162	.122	.084	.268	.199	.150	.357	.292	.214
	RMSE	.069	.047	.036	.105	.077	.056	.162	.122	.085	.268	.200	.151	.360	.299	.215
13	$\bar{\lambda}_0^2$	0.011	0.011	0.011	0.024	0.025	0.025	0.044	0.048	0.049	0.094	0.097	0.099	0.142	0.146	0.148
	FSSE	.010	.007	.005	.013	.009	.006	.016	.010	.007	.020	.013	.010	.022	.016	.012
	RMSE	.010	.007	.005	.013	.009	.006	.017	.010	.007	.021	.014	.010	.024	.016	.012
	$\bar{\sigma}$	0.997	0.998	1.000	0.976	0.986	0.992	0.942	0.954	0.981	0.871	0.917	0.974	0.797	0.821	0.875
	FSSE	.164	.128	.090	.236	.193	.148	.386	.301	.204	.515	.423	.359	.599	.475	.468
	RMSE	.164	.128	.090	.237	.194	.148	.390	.304	.205	.531	.431	.360	.632	.508	.485
16	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.023	0.025	0.026	0.047	0.049	0.050	0.093	0.098	0.099	0.142	0.146	0.148
	FSSE	.010	.007	.005	.013	.009	.006	.016	.011	.007	.020	.014	.009	.023	.017	.012
	RMSE	.010	.007	.005	.013	.009	.006	.016	.011	.007	.021	.015	.009	.024	.017	.012
	$\bar{\sigma}$	0.976	0.979	0.993	0.960	0.944	0.937	0.916	0.900	0.917	0.814	0.808	0.862	0.611	0.661	0.740
	FSSE	.223	.206	.182	.351	.301	.279	.514	.422	.362	.803	.638	.570	.552	.681	.798
	RMSE	.224	.207	.182	.354	.306	.287	.521	.433	.372	.824	.666	.586	.676	.761	.840

NOTE: All simulations are based on a process with $\xi \sim \mathrm{N}(0,1)$, $\omega_j \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. GMM was executed using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out.

Table 2: Monte Carlo results for $\widehat{\lambda}_q^2$ estimator.

		$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.025$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.1$			$\lambda_0^2 = 0.15$		
n		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
8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.025	0.025	0.025	0.049	0.049	0.050	0.097	0.100	0.099	0.142	0.144	0.147
	FSSE	.002	.002	.001	.004	.003	.002	.007	.005	.004	.015	.014	.009	.024	.018	.017
	RMSE	.002	.002	.001	.004	.003	.002	.007	.005	.004	.015	.014	.009	.025	.019	.017
	$\bar{\sigma}$	1.000	1.001	1.000	0.999	1.002	1.000	0.997	0.992	0.996	1.002	1.007	0.997	0.973	0.970	0.993
	FSSE	.037	.026	.019	.058	.042	.027	.084	.063	.046	.154	.127	.071	.222	.153	.141
	RMSE	.037	.026	.019	.058	.043	.027	.084	.063	.046	.154	.127	.071	.223	.156	.141
10	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.024	0.025	0.025	0.048	0.048	0.049	0.093	0.095	0.097	0.136	0.138	0.144
	FSSE	.002	.002	.001	.005	.003	.002	.009	.007	.005	.017	.014	.011	.030	.023	.019
	RMSE	.002	.002	.001	.005	.003	.002	.009	.007	.005	.019	.015	.012	.033	.026	.019
	$\bar{\sigma}$	0.993	1.000	1.001	0.998	0.999	0.996	0.991	0.985	0.986	0.982	0.981	0.990	0.963	0.934	0.979
	FSSE	.067	.046	.035	.103	.076	.055	.155	.119	.083	.262	.194	.150	.389	.287	.210
	RMSE	.067	.046	.035	.103	.076	.055	.155	.119	.084	.263	.195	.151	.391	.294	.211
13	$\bar{\lambda}_0^2$	0.008	0.009	0.009	0.021	0.023	0.023	0.040	0.043	0.046	0.077	0.084	0.089	0.112	0.120	0.129
	FSSE	.002	.002	.002	.005	.005	.004	.010	.008	.008	.017	.017	.015	.024	.022	.024
	RMSE	.003	.002	.002	.007	.005	.004	.014	.011	.009	.029	.023	.019	.045	.037	.032
	$\bar{\sigma}$	0.997	0.997	1.000	0.975	0.985	0.991	0.942	0.953	0.980	0.869	0.914	0.974	0.801	0.823	0.875
	FSSE	.161	.126	.089	.229	.191	.146	.379	.297	.205	.496	.412	.359	.578	.468	.467
	FSSE	.161	.126	.089	.231	.191	.147	.383	.300	.206	.513	.421	.360	.611	.500	.484
16	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.017	0.019	0.020	0.033	0.036	0.040	0.064	0.071	0.075	0.090	0.100	0.111
	FSSE	.002	.002	.002	.005	.004	.004	.008	.007	.008	.016	.015	.016	.022	.022	.021
	RMSE	.004	.003	.003	.009	.008	.007	.019	.016	.013	.039	.033	.030	.064	.055	.044
	$\bar{\sigma}$	0.973	0.979	0.992	0.943	0.942	0.966	0.879	0.899	0.932	0.754	0.807	0.882	0.637	0.680	0.758
	FSSE	.226	.205	.183	.328	.298	.261	.447	.416	.391	.633	.633	.743	.834	.843	.658
	RMSE	.228	.206	.184	.333	.303	.263	.463	.429	.397	.679	.661	.752	.910	.901	.701

NOTE: All simulations are based on a process with $\xi \sim \mathrm{N}(0,1)$, $\omega_j \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}$ are the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. Kiyono *et al.*'s estimator $\widehat{\lambda}_q^2$ was calculated using $q = 0.5$ after the series was filtered by the sample estimate of $\bar{\sigma}$. The results for $\widehat{\lambda}_q^2$ have been normalized by n. For each case, 400 Monte Carlo runs have been carried out.

Table 3: Monte Carlo results for GMM with cascade-level uncertainty added

λ_0^2		n = 8	9	10	11	12	13	14
0.01	$\bar{\lambda}_0^2$	0.011	0.011	0.011	0.011	0.011	0.011	0.011
	FSSE	.005	.005	.005	.005	.005	.005	.005
	RMSE	.005	.005	.005	.005	.005	.005	.005
	APD	0.031	0.030	0.024	0.000	0.023	0.025	0.026
	$\bar{\sigma}$	1.001	1.001	1.002	1.002	1.002	1.002	1.002
	FSSE	.045	.045	.045	.045	.045	.045	.045
	RMSE	.045	.045	.045	.045	.045	.045	.045
	APD	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Qmin.	0.994	0.997	0.999	1.000	1.000	1.001	1.001
0.05	$\bar{\lambda}_0^2$	0.050	0.050	0.050	0.050	0.050	0.050	0.050
	FSSE	.007	.007	.007	.007	.007	.007	.007
	RMSE	.007	.007	.007	.007	.007	.007	.007
	APD	0.006	0.003	0.001	0.000	0.001	0.001	0.001
	$\bar{\sigma}$	0.998	0.999	0.999	0.999	0.999	0.999	0.999
	FSSE	.117	.117	.117	.117	.117	.117	.117
	RMSE	.117	.117	.117	.117	.117	.117	.117
	APD	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	Qmin.	0.991	0.990	0.996	1.000	1.003	1.004	1.005
0.15	$\bar{\lambda}_0^2$	0.151	0.149	0.149	0.149	0.148	0.148	0.148
	FSSE	.012	.012	.012	.012	.012	.012	.012
	RMSE	.012	.012	.012	.012	.012	.012	.012
	APD	0.014	0.006	0.002	0.000	0.001	0.002	0.002
	$\bar{\sigma}$	0.956	0.960	0.962	0.963	0.964	0.964	0.964
	FSSE	.282	.285	.287	.288	.288	.288	.288
	RMSE	.286	.288	.289	.290	.290	.290	.290
	APD	0.008	0.003	0.001	0.000	0.001	0.001	0.001
	Qmin.	1.156	1.025	1.002	1.000	1.002	1.003	1.004

NOTE: All simulations are based on a process with $n = 11$ cascade levels, $\xi \sim \mathrm{N}(0, 1)$, $\omega_j \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. The sample length is $T = 10,000$. GMM was executed for each number of cascade levels in the table using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out. $\bar{\lambda}_0$ is the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. For each iteration, the absolute difference between an estimate with cascade level n and the estimate with $n = 11$ as a proportion to the latter has been calculated. The Absolute Percentage Difference (APD) denotes the mean of such series. Qmin is the average of the objective value at the optimum for each n divided by the one with $n = 11$.

Table 4: Monte Carlo assessment of best linear forecasts based on GMM and $\widehat{\lambda}_q^2$ estimation

		$n = 10$						$n = 13$						$n = 16$					
		$\widehat{\lambda}_q^2$			GMM			$\widehat{\lambda}_q^2$			GMM			$\widehat{\lambda}_q^2$			GMM		
	T | λ_0^2	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15
MSE	1	.954	.911	.936	.959	.911	.938	.922	.857	.878	.928	.857	.887	.912	.835	.828	.919	.835	.850
		(.023)	(.039)	(.056)	(.025)	(.039)	(.054)	(.060)	(.134)	(.160)	(.060)	(.134)	(.153)	(.075)	(.166)	(.228)	(.073)	(.165)	(.195)
	10	.973	.964	.982	.975	.964	.983	.941	.906	.930	.944	.906	.939	.931	.883	.880	.935	.883	.908
		(.020)	(.030)	(.030)	(.020)	(.031)	(.028)	(.060)	(.135)	(.136)	(.059)	(.134)	(.126)	(.075)	(.170)	(.211)	(.072)	(.169)	(.161)
	20	.980	.975	.987	.981	.975	.989	.948	.919	.940	.951	.919	.949	.938	.895	.890	.942	.895	.920
		(.018)	(.026)	(.024)	(.018)	(.026)	(.022)	(.059)	(.133)	(.125)	(.058)	(.132)	(.116)	(.074)	(.171)	(.201)	(.071)	(.169)	(.149)
	100	.994	.993	.996	.994	.993	.998	.966	.942	.961	.968	.942	.968	.956	.916	.918	.958	.917	.946
		(.010)	(.014)	(.010)	(.009)	(.015)	(.011)	(.055)	(.121)	(.091)	(.054)	(.120)	(.091)	(.074)	(.165)	(.164)	(.070)	(.161)	(.114)
MAE	1	.965	.897	.901	.968	.897	.897	.938	.839	.870	.942	.839	.876	.936	.806	.833	.941	.807	.855
		(.052)	(.112)	(.127)	(.049)	(.113)	(.126)	(.118)	(.240)	(.246)	(.113)	(.237)	(.201)	(.140)	(.278)	(.325)	(.134)	(.273)	(.232)
	10	.977	.952	.970	.979	.952	.964	.951	.893	.946	.954	.893	.946	.948	.857	.908	.952	.859	.924
		(.045)	(.096)	(.088)	(.042)	(.097)	(.091)	(.114)	(.234)	(.210)	(.110)	(.230)	(.161)	(.137)	(.278)	(.297)	(.131)	(.268)	(.189)
	20	.983	.967	.981	.984	.967	.975	.956	.910	.961	.958	.910	.959	.954	.875	.925	.957	.877	.938
		(.040)	(.084)	(.072)	(.038)	(.085)	(.080)	(.111)	(.225)	(.192)	(.106)	(.221)	(.146)	(.135)	(.272)	(.280)	(.128)	(.262)	(.173)
	100	.994	.992	.997	.995	.993	.992	.971	.948	.986	.973	.948	.981	.968	.913	.960	.970	.915	.963
		(.021)	(.042)	(.030)	(.020)	(.046)	(.060)	(.095)	(.186)	(.137)	(.090)	(.182)	(.111)	(.122)	(.241)	(.221)	(.115)	(.227)	(.133)

NOTE: The table shows mean squared errors (MSE) and mean absolute errors (MAE) for three different cascade sizes n. MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from a RW using the in-sample variance. All entries are averaged over 400 Monte Carlo runs (with standard errors given in parenthesis). In each run, an overall sample of 7500 entries from a random starting point has been drawn. From that starting point on, an in-sample period of 5,000 entries for parameter estimation and an adjacent out-of-sample period of 2,500 entries for evaluation of forecasting performance were selected. Parameter values have been estimated with Kiyono's $\widehat{\lambda}_q^2$ and GMMs estimator respectively. GMM was executed using lags $\ell = 1, 14, 64$. The column T represents the forecast horizons, whereas the row λ_0^2 describes the selected intermittency values.

Table 5: Monte Carlo assessment of best linear forecasts with cascade-level uncertainty.

		n										
	T	8	9	10	11	12	13	14	···	20	···	50
MSE	1	.896	.893	.891	.890	.890	.890	.890		.890		.890
		(.073)	(.078)	(.081)	(.083)	(.084)	(.085)	(.085)		(.086)		(.086)
	10	.954	.947	.843	.941	.940	.940	.939		.939		.940
		(.051)	(.064)	(.072)	(.078)	(.081)	(.083)	(.084)		(.085)		(.085)
	20	.970	.961	.956	.953	.951	.951	.951		.951		.951
		(.038)	(.054)	(.066)	(.074)	(.078)	(.081)	(.082)		(.084)		(.084)
	100	1.000	.992	.983	.977	.974	.972	.972		.971		.971
		(.013)	(.018)	(.036)	(.052)	(.063)	(.070)	(.073)		(.077)		(.077)
MAE	1	.890	.884	.881	.880	.878	.878	.878		.878		.878
		(.126)	(.146)	(.163)	(.175)	(.185)	(.190)	(.193)		(.198)		(.199)
	10	.948	.940	.934	.931	.930	.929	.929		.928		.928
		(.082)	(.111)	(.136)	(.157)	(.172)	(.182)	(.187)		(.195)		(.197)
	20	.967	.957	.950	.947	.945	.944	.944		.943		.943
		(.061)	(.090)	(.119)	(.143)	(.162)	(.174)	(.180)		(.190)		(.192)
	100	.999	.991	.984	.979	.976	.975	.975		.975		.975
		(.038)	(.040)	(.063)	(.094)	(.122)	(.141)	(.152)		(.168)		(.172)

NOTE: All simulations are based on a process with $n = 11$ cascade levels, $\xi \sim \mathrm{N}(0,1)$, $\omega_j \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from a RW using the in-sample variance. All entries are averages over 400 Monte Carlo runs (with standard errors given in parenthesis). In each run, an overall sample of 5,000 entries from a random starting point with $\lambda_0^2 = 0.05$ have been drawn. From that starting point on, an in-sample period of 2,500 entries for parameter estimation and an adjacent out-of-sample period of 2,500 entries for evaluation of forecasting performance were selected. Employed parameter values were estimated with GMM, using lags $\ell = 1, 14, 64$. The column T represents the forecast horizons.

Table 6: Empirical parameter estimates

	Asset							
	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
$\widehat{\lambda}_0^2$	.0154	.0357	.0167	.0119	.0357	.0238	.0191	.0339
(SE)	(.006)	(.007)	(.006)	(.005)	(.007)	(.006)	(.006)	(.007)
Max APD	.021	.006	.003	.165	.017	.011	.005	.003
$\widehat{\sigma}$	.2781	.6945	.6556	.7533	.6208	.7144	.6546	1.3499
(SE)	(.009)	(.021)	(.038)	(.017)	(.032)	(.018)	(.020)	(.107)
J_{prob}	.243	.843	.718	.005	.500	.695	.550	.260
n_{min}	8	8	19	8	8	20	20	12

NOTE: Empirical estimates for standardized cascade shocks $\omega_j(t) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$ and overall order of fluctuation magnitude $\bar{\sigma}$, obtained via GMM from a sample of 5500 entries for each asset. Each column show the estimate corresponding to the lowest objective function obtained for the range of cascade sizes $n = 8, \dots, 20$ for each asset. GMM was executed using lags $\ell = 1, 14, 64$. SE are the standard errors of the pertinent estimates and the entry J_{prob} gives the probability of the corresponding J statistics. n_{min} is the number of cascade levels at which the lowest objective function was obtained. The Maximum Absolute Percentage Difference (M-APD) is taken between the $\widehat{\lambda}_0^2$ with the lowest and the highest objective function from the employed range.

Table 7: Empirical forecast: MSE

		Asset							
	T	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
GMM20	1	$.801^{**}$	$.956^{**}$	$.852^{**}$	$.943^{**}$	.786	$.925^{**}$	$.890^{**}$	$.911^{**}$
	5	$.808^{**}$	$.970^{**}$	$.840^{**}$	$.951^{**}$	$.851^{**}$	$.900^{**}$	$.877^{**}$	$.894^{**}$
	20	$.855^{**}$	$.989^{**\dagger}$	$.867^{**}$	$.980^{**}$	$.927^{**}$	$.914^{**}$	$.917^{**}$	$.917^{**}$
	50	$.913^{**\dagger}$	$.996^{**}$	$.919^{**}$	$.986^{**}$	$.968^{**}$	$.952^{**}$	$.962^{**}$	$.964^{**}$
	100	$.923^{**}$	$.999^{**}$	$.970^{**}$	1.015^{**}	$.973^{**}$	$.988^{**}$	1.008^{**}	$.984^{**}$
GARCH	1	$.792^{**}$	$.958^{**}$	$.836^{**\dagger}$	$.942^{**}$	.788	$.917^{**}$	$.874^{**\dagger}$	$.903^{**}$
	5	$.794^{**}$	$.971^{**}$	$.827^{**\dagger}$	$.953^{**}$	$.902^{*}$	$.900^{**}$	$.877^{**}$	$.899^{**}$
	20	$.847^{**}$	$.996^{**}$	$.851^{**}$	$.985^{**}$	1.056^{**}	$.913^{**}$	$.908^{**}$	$.917^{**}$
	50	$.933^{**}$	1.006^{**}	$.899^{**}$	$.994^{**}$	1.203^{**}	$.961^{**}$	$.948^{**}$	1.009^{**}
	100	$.952^{**}$	1.011^{**}	$.948^{**}$	1.013^{**}	1.252^{**}	1.000	$.997^{**}$	1.050^{**}

NOTE: Multifractal and GARCH(1,1) mean squared errors (MSE) in % of the pertinent MSEs of a 'naive' forecast from a RW using the historical variance. GMM20 denotes multifractal forecasts that were employed using a cascade level $n = 20$. Multifractal parameter values were estimated with the GMM estimator executed using lags $\ell = 1, 14, 64$.
* denotes an improvement against RW model which is significant at the 95% level.
** denotes an improvement against RW model which is significant at the 99% level.
† denotes an improvement significant at the 95% level (GMM20 against GARCH(1,1) and *vice versa*).
Comparisons against RW are based on the test statistic for nested models of Clark and West [26, 27]. Comparisons against GARCH(1,1) are based on the modified Diebold and Mariano [28] test statistic by Harvey *et al.* [29].

Table 8: Empirical forecast: MAE

		Asset							
	T	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
GMM20	1	1.094	.949**	1.051†	.923**†	1.090†	.885**†	.911**†	.838**†
	5	1.085	.949**†	1.045†	.916**†	1.093†	.872**†	.905**†	.838**†
	20	1.080	.958*†	1.035†	.925**†	1.107†	.882**†	.910**†	.860**†
	50	1.070	.964	1.044†	.932*†	1.120†	.900**†	.918*†	.887**†
	100	1.061	.969	1.060†	.942	1.119†	.915*†	.942†	.904†
GARCH	1	1.103	.950**	1.060	.939**	1.172	.908**	.920**	.849**
	5	1.088	.957**	1.061	.939**	1.219	.907**	.921**	.857**
	20	1.073	.974	1.068	.961**	1.352	.950**	.937*	.900*
	50	1.037 †	.988*	1.088	.983*	1.641	1.016	.955	.991
	100	1.013 †	.998	1.119	.998	2.082	1.064	.994	1.091

NOTE: Multifractal and GARCH(1,1) mean squared errors (MSE) in % of the pertinent MSEs of a 'naive' forecast from a RW using the historical variance. GMM20 denotes multifractal forecasts that were employed using a cascade level $n = 20$. Multifractal parameter values were estimated with the GMM estimator executed using lags $\ell = 1, 14, 64$.

* denotes an improvement against RW model which is significant at the 95% level.

** denotes an improvement against RW model which is significant at the 99% level.

† denotes an improvement significant at the 95% level (GMM20 against GARCH(1,1) and *vice versa*).

Comparisons against RW and GARCH(1,1) are based on the modified Diebold and Mariano [28] test statistic by Harvey *et al.* [29].

Appendix A

We first consider the following definition:

$$\eta_{t,\ell} \doteq \sum_{j=1}^{n} \left[\omega_j(t) \ - \omega_j(t-\ell)\right],$$

where obviously it holds that $\mathbb{E}[\eta_{t,\ell}] = 0$. Lux [13] shows that the moments of $\zeta_{t,\ell}$, defined as in eq. (6), can be written as:

$$\mathbb{E}[\zeta_{t+\ell,\ell} \cdot \zeta_{t,\ell}] = \mathbb{E}[\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}] + \mathbb{E}^2[\ln|\xi_i|] - \mathbb{E}[(\ln|\xi_t|)^2] \tag{A.1}$$

and

$$\begin{aligned}\mathbb{E}[\zeta^2_{t+\ell,\ell} \cdot \zeta^2_{t,\ell}] = \mathbb{E}[\eta^2_{t+\ell,\ell} \cdot \eta^2_{t,\ell}] - 4\left\{\mathbb{E}[\eta^2_{t,\ell}] - \mathbb{E}[\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}]\right\} \cdot \left\{\mathbb{E}^2[\ln|\xi_i|] - \mathbb{E}[(\ln|\xi_t|)^2]\right\} \\ + 3\,\mathbb{E}^2[(\ln|\xi_t|)^2] - 4\mathbb{E}[\ln|\xi_t|] \cdot \mathbb{E}[(\ln|\xi_t|)^3] + \mathbb{E}[(\ln|\xi_t|)^4].\end{aligned} \tag{A.2}$$

As can be seen in eq. (A.1) and (A.2), the moments in eq. (7) require to compute $\mathbb{E}[\eta^2_{t,\ell}]$, $\mathbb{E}[\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}]$, and $\mathbb{E}[\eta^2_{t+\ell,\ell} \cdot \eta^2_{t,\ell}]$. These moments require two inputs, the probabilities for the renewal of the 'multipliers' ω_j, and the distribution of ω_j itself.

In the first case, $\eta^2_{t,\ell}$ elements would be different from zero if the pertinent ω_j has different realizations in $t+\ell$ and t. Therefore, the expected value of $\eta^2_{t,\ell}$ is computed as

$$\begin{aligned}\mathbb{E}[\eta^2_{t+\ell,\ell}] &= \left(\tfrac{2}{2^n}\right) \cdot \sum_{j=1}^{n} \left\{\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]\right\} \cdot \left\{\mathbb{I}(2^{n-j} > \ell) \cdot (2^j\,\ell) + \mathbb{I}(2^{n-j} \leq \ell) \cdot 2^n\right\} \\ &= \left(\tfrac{2\ \tilde{\sigma}^2}{2^n}\right) \cdot \sum_{j=1}^{n} \left\{\mathbb{I}(2^{n-j} > \ell) \cdot (2^j\,\ell) + \mathbb{I}(2^{n-j} \leq \ell) \cdot 2^n\right\},\end{aligned} \tag{A.3}$$

where $\mathbb{I}(\cdot)$ represents the indicator functions, ω_j is distributed $\mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$, and 2^{n-j} accounts for the length of each subinterval at cascade step j. To understand the computation of eq. (A.3) note that 2^j is the number of different Lognormal draws at cascade level j while 2^{n-j} is the number of successive elements with equal contributions at level j within a single sequence of a cascade with 2^n time-ordered observations. In eq. (A.3), we distinguished between the cases $2^{n-j} > \ell$ and $2^{n-j} \leq \ell$. In the first case, we have to account for the sequence of the ℓ-first consecutive equal contributions at level j. From these ℓ of the 2^{n-j} numbers, a difference of ℓ will reach into the next box and, hence, have a nonzero value. If $2^{n-j} \leq \ell$ all ℓ-differences at level j will lead out of the individual box so that nonzero values will be estimated for all 2^n admissible starting points.

Calculations become slightly more involved for the autocovariances of $\eta_{t,\ell}$. First of all, we know that

$$\eta_{t+\ell,\ell} \cdot \eta_{t,\ell} = \left(\sum_{j=1}^{n} \left[\omega_j(t+\ell) \; - \omega_j(t) \right] \right) \times \left(\sum_{s=1}^{n} \left[\omega_s(t) \; - \omega_s(t-\ell) \right] \right). \tag{A.4}$$

Because of independence of realizations of any pair of volatility components j and s, only summands with $j = s$ give nonzero contributions. As such, eq. (A.4) becomes

$$\eta_{t+\ell,\ell} \cdot \eta_{t,\ell} = \sum_{j=1}^{n} \left[\omega_j(t+\ell) \cdot \omega_j(t) \; - \omega_j^2(t) \; - \omega_j(t+\ell) \cdot \omega_j(t-\ell) \; + \omega_j(t) \cdot \omega_j(t-\ell) \right].$$

Further, different realizations in $t+\ell$ and t, and in t and $t-\ell$, that is, $\omega_j(t+\ell) \neq \omega_j(t) \neq \omega_j(t-\ell)$ at each cascade step j may now exist depending on the relationship between 2^{n-j} and 2ℓ, and 2^{n-j} and ℓ. We find, then, the autocovariances of $\eta_{t,\ell}$ using the first and second moments $\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]$:

$$\begin{aligned}\mathbb{E}[\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}] = -\left(\tfrac{\tilde{\sigma}^2}{2^n}\right) \cdot \sum_{j=1}^{n} \Big\{ & \mathbb{I}(2^{n-j} \leq \ell) \cdot 2^n + \mathbb{I}(2^{n-j} > \ell) \cdot \\ & \cdot \Big[\mathbb{I}(2^{n-j} < 2\ell) \cdot \Big[\mathbb{I}(j > 1) \cdot 2^j \cdot \left(2\ell - 2^{n-j}\right) \Big] \Big] \Big\}.\end{aligned} \tag{A.5}$$

Eq. (A.5) can be understood following the sequence of different cases we distinguish: First, if $2^{n-j} \leq \ell$, any ℓ-difference at level j involves two different random numbers, and so, all admissible values, 2^n, make a nonzero contribution. If $2^{n-j} > \ell$, at least $2^{n-j} < 2\ell$ and $j > 1$ must hold to have any nonzero entries. Their number can then be determined by the following considerations: The term 2^j is the number of boxes of size 2^{n-j} on the bounded interval while $2\ell - 2^{n-j}$ determines the number of possible starting points of nonzero double differences of two times the size ℓ. Calculations for the autocovariances of $\eta^2_{t,\ell}$ are more complex. We can arrive at the closed-form solutions for the total sample counterpart by identifying the nonzero entries in the cascade with respect to

$$\eta^2_{t+\ell,\ell} \cdot \eta^2_{t,\ell} = \left(\sum_{j=1}^{n} [\omega_j(t+\ell) \; - \omega_j(t)] \right)^2 \times \left(\sum_{s=1}^{n} [\omega_s(t) \; - \omega_s(t-\ell)] \right)^2 , \tag{A.7}$$

which requires to identify three different cases:

(1) $j = s$ and $\omega_j(t+\ell) \neq \omega_j(t) \neq \omega_j(t-\ell)$ leading to entries of the form

$$(\omega_j(t+\ell) \; - \omega_j(t))^2 \times (\omega_j(t) \; - \omega_j(t-\ell))^2 , \tag{A.8}$$

We count here the same number of entries as in the case of $\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}$. Using the identities $\mathbb{E}[(\omega_j)^3] = 3\,\tilde{\mu}\,\tilde{\sigma}^2 + \tilde{\mu}^3$ and $\mathbb{E}[(\omega_j)^4] = 3\,\tilde{\sigma}^4 + 6\,\tilde{\mu}^2\,\tilde{\sigma}^2 + \tilde{\mu}^4$, the expectation of eq. (A.8), once the number of cases is counted, is

$$\mathbb{E}[(\omega_j)^4] + 3\,\mathbb{E}^2[(\omega_j)^2] - 4\,\mathbb{E}[(\omega_j)^3] \cdot \mathbb{E}[(\omega_j)] \; = \; 6\,\tilde{\sigma}^4.$$

Together, we obtain

$$\begin{aligned} \kappa_1 = 6\,\tilde{\sigma}^4 \cdot \sum_{j=1}^{n} \; \Big\{ & \mathbb{I}(2^{n-j} \leq \ell) \cdot 2^n + \mathbb{I}(2^{n-j} > \ell) \cdot \\ & \cdot \Big[\mathbb{I}(2^{n-j} < 2\ell) \cdot \Big[\mathbb{I}(j > 1) \cdot 2^j \left(2\ell - 2^{n-j} \right) \Big] \Big] \Big\} \end{aligned} \tag{A.9}$$

(2) $j \neq s$ and $\omega_j(t+\ell) \neq \omega_j(t)$ and $\omega_s(t) \neq \omega_s(t-\ell)$ for the case of entries like

$$(\omega_j(t+\ell) \; - \omega_j(t))^2 \times (\omega_s(t) \; - \omega_s(t-\ell))^2 , \tag{A.10}$$

In this case, we can simplify our computations by considering only contributions for which the second index is lower than the first. The reason is that if a term is non-zero for the higher index, it is the lower index that determines whether the complete expression

((A.10)) is vanishing or not.[12] Therefore, we define for $h > l$

$$\psi(h) \doteq \mathbb{I}(2^{n-h} \leq \ell) \cdot \Big[\mathbb{I}(2^{n-l} > \ell) \cdot 2^l \cdot \ell + \mathbb{I}(2^{n-l} \leq \ell) \cdot 2^n\Big] + \\ + \mathbb{I}(2^{n-h} > \ell) \cdot \mathbb{I}(2^{n-h} < 2\ell) \cdot 2^l \cdot \Big(2\ell - 2^{n-h}\Big) . \tag{A.11}$$

In eq. (A.11) we encounter the following cases: First, if $2^{n-h} \leq \ell$ and $2^{n-l} > \ell$, a number of 2^l boxes has each ℓ nonzero contributions for the double differences. Further, if both $2^{n-h} \leq \ell$ and $2^{n-l} \leq \ell$, all countable elements, i.e. 2^n cases, make a nonzero contribution. If, finally, $\ell < 2^{n-h} < 2\ell$ we have 2^l boxes with $2\ell - 2^{n-h}$ nonzero elements each.

Furthermore, for non-zero entries, the expectation of eq. (A.10) is defined as

$$4\,\mathbb{E}^2[(\omega_j)^2] - 8\,\mathbb{E}[(\omega_j)^2] \cdot \mathbb{E}^2[(\omega_j)] + 4\,\mathbb{E}^4[(\omega_j)] \;=\; 4\,\tilde{\sigma}^4$$

Together with the number of cases, we obtain

$$\kappa_2 = \; 4\,\tilde{\sigma}^4 \cdot \sum_{j=1}^{n} \sum_{\substack{s=1 \\ s \neq j}}^{n} \; [\mathbb{I}(j > s) \cdot \psi(j) + \mathbb{I}(j < s) \cdot \psi(s)] \, , \tag{A.12}$$

(3) $j \neq s$ and $\omega_m(t+\ell) \neq \omega_m(t) \neq \omega_m(t-\ell)$ for $m = j, s$ leads to entries of the format

$$(\omega_j(t+\ell) \; - \omega_j(t)) \times (\omega_j(t) \; - \omega_j(t-\ell)) \times \\ \times (\omega_s(t+\ell) \; - \omega_s(t)) \times (\omega_s(t) \; - \omega_s(t-\ell)) \, . \tag{A.13}$$

As we can see, we have here a double term at cascade step j and another double term at a higher or lower cascade step s. Since both pairs of terms must not disappear simultaneously, once we are in, let's say, $\mathbb{I}(j < s)$, the total number of double terms is fully determined by the number of double terms with the index j. In summary, we have the same number of double terms for index j as in $\eta_{t+\ell,\ell} \cdot \eta_{t,\ell}$, so for $h > l$ we define the

[12] Please remember that a lower index means a position in a longer subinterval, so if the term is non-zero there, it must also be non-zero at a higher index level or shorter subinterval respectively.

counting formula as

$$\varphi(l) = \sum_{l=1}^{n} \Big\{ \mathbb{I}(2^{n-l} \leq \ell) \cdot 2^n + \mathbb{I}(2^{n-l} > \ell) \cdot \\ \cdot \Big[\mathbb{I}(2^{n-l} < 2\ell) \cdot \Big[\mathbb{I}(l \geq 1) \cdot 2^l \cdot \Big(2\ell - 2^{n-l} \Big) \Big] \Big] \Big\} . \tag{A.14}$$

Given the previous explanations, the components of eq. (A.14) are easily explained. Using these results and the fact that the expectation of eq. (A.13) is $2\tilde{\sigma}^4$ we find

$$\kappa_3 = 2\,\tilde{\sigma}^4 \cdot \sum_{j=1}^{n} \sum_{\substack{s=1 \\ s \neq j}}^{n} \left[\mathbb{I}(j > s) \cdot \varphi(s) + \mathbb{I}(j < s) \cdot \varphi(j) \right] . \tag{A.15}$$

Putting (A.9), (A.12), and (A.15) together we finally obtain

$$\mathbb{E}[\eta^2_{t+\ell,\ell} \cdot \eta^2_{t,\ell}] \;=\; \left(\tfrac{1}{2^n} \right) \cdot (\kappa_1 + \kappa_2 + \kappa_3). \tag{A.14}$$

As the last element we need, the log-absolute moments of the standard Normal variates ξ_i in eq. (A.2) can be easily obtained by using the Gamma function and its derivatives.

Appendix B

We display here results for the $\widehat{\lambda}^2_q$ estimator of Kiyono *et al.* under different q values. Results have been normalized by n to ease comparison to the value from Table 2.

Table B.1: Monte Carlo results for $\widehat{\lambda}_q^2$ estimator with different q values and $\lambda_0^2 = 0.05$.

		$q=-0.5$			$q=0.01$			$q=0.5$			$q=1$			$q=1.5$		
k		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
8	$\bar{\lambda}_0^2$	0.049	0.050	0.050	0.049	0.049	0.050	0.049	0.049	0.050	0.049	0.049	0.050	0.049	0.049	0.050
	FSSE	.010	.008	.005	.007	.005	.004	.007	.005	.004	.008	.006	.005	.009	.006	.005
	RMSE	.010	.008	.005	.007	.005	.004	.007	.005	.004	.008	.006	.005	.009	.006	.005
10	$\bar{\lambda}_0^2$	0.048	0.049	0.049	0.048	0.049	0.049	0.048	0.048	0.049	0.047	0.048	0.049	0.047	0.048	0.048
	FSSE	.009	.008	.005	.008	.007	.004	.009	.007	.005	.009	.008	.005	.010	.008	.006
	RMSE	.009	.008	.005	.008	.007	.005	.009	.007	.005	.010	.008	.005	.011	.009	.006
13	$\bar{\lambda}_0^2$	0.040	0.044	0.047	0.041	0.043	0.046	0.040	0.043	0.046	0.040	0.043	0.046	0.040	0.042	0.045
	FSSE	.010	.009	.008	.010	.008	.007	.010	.008	.008	.010	.009	.008	.011	.009	.008
	RMSE	.014	.011	.008	.013	.010	.008	.014	.011	.009	.014	.011	.009	.015	.012	.010
16	$\bar{\lambda}_0^2$	0.034	0.037	0.040	0.034	0.037	0.040	0.033	0.036	0.040	0.033	0.036	0.039	0.033	0.036	0.039
	FSSE	.008	.008	.008	.008	.008	.008	.008	.007	.008	.008	.008	.008	.008	.008	.008
	RMSE	.018	.015	.013	.018	.015	.013	.019	.016	.013	.019	.016	.014	.019	.017	.014

NOTE: All simulations are based on a process with $\xi \sim \mathrm{N}(0,1)$, $\omega_j \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, $\lambda_0^2 = 0.05$ and $\sigma = 1$. Sample lengths are: $T_1 = 2,500$, $T_2 = 5,000$, and $T_3 = 10,000$. $\bar{\lambda}_0^2$ is the corresponding mean of the estimated parameters. FSSE and RMSE denote the finite sample standard error and root mean squared error, respectively. Kiyono *et al.*'s estimator $\widehat{\lambda}_q^2$ was calculated using different values of q after the series was filtered by the estimate of $\bar{\sigma}$ (not shown here). The results for $\widehat{\lambda}_q^2$ have been normalized by n. For each case, 400 Monte Carlo runs have been carried out.

Appendix C

For computing linear forecasts of $x_t^2 - \widehat{\sigma}^2$ in eq. (10), we need the second moment and the autocovariances of the x_t. Let us define

$$\vartheta_t \doteq \exp\left[2 \cdot \sum_{j=1}^{n} \omega_j^{(m)}\left(\left\lfloor \tfrac{t-2^n(m\text{-}1)-1}{2^{n-j}} \right\rfloor\right)\right] = \exp\left[2 \cdot \sum_{j=1}^{n} \omega_j\,(t)\right] = \left[\prod_{j=1}^{n} M_j\,(t)\right]^2, \qquad \text{(B.1)}$$

for $m = 0, 1, \ldots$, and $\exp[\omega_j\,(t)] = M_j\,(t)$. Due to the standardization of the pdf in eq. (1) by $\sigma_0 = \exp(-\lambda^2)$ in eq. (2), it follows that $\omega_j\,(t) \sim \mathrm{N}(-\lambda_0^2, \lambda_0^2)$, which conveys that $\mathbb{E}[\,\vartheta_i\,] = 1$ and the cascade level conserves mass on average. The second moment of the volatility process is

$$\mathbb{E}[\,\vartheta_t^2\,] = \mathbb{E}\left[\exp\left[4 \cdot \sum_{j=1}^{n} \omega_j(t)\right]\right] = \mathbb{E}\left[\left(\prod_{j=1}^{n} M_j(t)\right)^4\right] = \exp(4\,n\,\lambda_0^2),$$

while the autocovariance of the volatility process is equivalent to

$$\begin{aligned}
\mathbb{E}[\vartheta_{t+\ell}\,\vartheta_t] &= \mathbb{E}\left[\prod_{j=1}^{n}\left[M_j^2(t) M_j^2(t+\ell)\;\mathbb{I}_{(M_j(t) \neq M_j(t+\ell))} + M_j^4(t)\;\mathbb{I}_{(M_j(t) = M_j(t+\ell))}\right]\right] \\
&= \prod_{j=1}^{n}\left[\mathbb{E}\left[M_j^2(t) M_j^2(t+\ell)\;\mathbb{I}_{(M_j(t) \neq M_j(t+\ell))}\right] + \mathbb{E}\left[M_j^4(t)\;\mathbb{I}_{(M_j(t) = M_j(t+\ell))}\right]\right] \\
&= \prod_{j=1}^{n}\left[\mathbf{1} \cdot \mathbb{I}(2^{n-j} \leq \ell) + \left[\exp(4\,\lambda_0^2) \cdot \tfrac{2^n - 2^j\,\ell}{2^n} + \mathbf{1} \cdot \tfrac{2^j\,\ell}{2^n}\right] \cdot \mathbb{I}(2^{n-j} > \ell)\right].
\end{aligned}$$

References

[1] B. Castaing, Y. Gagne, and E. J. Hopfinger, Physica D **46** (1990).

[2] S. Carius and G. Ingelman, Phys Lett B **252** (1990).

[3] L. Sorriso-Valvo, V. Carbone, P. Veltri, G. Consolini, and R. Bruno, Geophys. Res. Lett **26**, 1801 (1999).

[4] K. Kiyono, Z. R. Struzik, N. Aoyagi, S. Sakata, J. Hayano, and Y. Yamamoto, Phys Rev. Lett **93**, 178103 (2004).

[5] K. Kiyono, Z. R. Struzik, N. Aoyagi, F. Togo, and Y. Yamamoto, Phys Rev. Lett **95**, 58101 (2005).

[6] A. Arneodo, N. Decoster, and S. G. Roux, Phys Rev. Lett **83** (1999).

[7] K. Kiyono, Z. R. Struzik, and Y. Yamamoto, Phys Rev. Lett **96**, 68701 (2006).

[8] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature **381**, 767 (1996).

[9] A. Fisher, L. Calvet, and B. Mandelbrot, Multifractality of Deutschemark / US Dollar Exchange Rates, Discussion Paper 1166, Cowles Foundation, 1997.

[10] B. B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer, New York, 1997.

[11] B. B. Mandelbrot, *Multifractals and* 1/f *Noise: Wild Self-Affinity in Physics*, Springer, New York, 1999.

[12] K. Kiyono, Z. R. Struzik, and Y. Yamamoto, Phys Rev. E **76**, 41113 (2007).

[13] T. Lux, J. of Business & Economic Statistics **26**, 194 (2008).

[14] L. Calvet and A. Fisher, J. of Econometrics **105**, 27 (2001).

[15] W. A. Gardner, A. Napolitano, and L. Paura, Signal Processing **86**, 639 (2006).

[16] D. Harris and L. Mátyás, University Press, Cambridge, 1999.

[17] L. P. Hansen, Econometrica **50**, 1029 (1982).

[18] L. P. Hansen, J. Heaton, and A. Yaron, J. of Business & Economic Statistics **14**, 262 (1996).

[19] L. Calvet and A. Fisher, J. of Financial Econometrics **2**, 49 (2004).

[20] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, New York, 1991.

[21] P. J. Brockwell and R. Dahlhaus, J. of Econometrics **118**, 129 (2004).

[22] J. F. Muzy, J. Delour, and E. Bacry, The European Physical Journal B - Condensed Matter and Complex Systems **17** (2000).

[23] A. Arnodo, J. F. Muzy, and D. Sornette, The European Physical Journal B - Condensed Matter and Complex Systems **2** (1998).

[24] M. Ausloos and K. Ivanova, Phys Rev. E **68**, 046122 (2003).

[25] N. Kozuki and N. Fuchikami, Physica A **329** (2003).

[26] T. E. Clark and K. D. West, J. of Econometrics **135**, 155 (2006).

[27] T. E. Clark and K. D. West, J. of Econometrics **138**, 291 (2007).

[28] F. X. Diebold and R. S. Mariano, J. of Business and Economic Statistics **13**, 134 (1995).
[29] D. Harvey, S. Leybourne, and P. Newbold, International Journal of Forecasting **13**, 281 (1997).