Parameter estimation and forecasting for multiplicative log-normal cascades

Andrés E. Leövey[1] and Thomas Lux[2,*]

[1]*Department of Economics, University of Kiel, Kiel 24118, Germany*
[2]*Department of Economics, University of Kiel, Kiel 24118, and Kiel Institute for the World Economy, Kiel 24105, Germany,
and Department of Economics, University Jaume I, Castellón 12071, Spain*

We study the well-known multiplicative log-normal cascade process in which the multiplication of Gaussian and log normally distributed random variables yields time series with intermittent bursts of activity. Due to the nonstationarity 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 probability density function to empirical data, cf. Castaing *et al.* [Physica D **46**, 177 (1990)]. More recently, alternative estimators based upon various moments have been proposed by Beck [Physica D **193**, 195 (2004)] and Kiyono *et al.* [Phys. Rev. E **76**, 041113 (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.

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:

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

(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,

$$x_i = \exp(\varepsilon_i)\xi_i,$$

(2)

where ξ_i and ε_i are mutually independent and follow normal distributions [denoted by $N(\cdot,\cdot)$ in the following]: $\xi_i \sim N(0,\bar{\sigma}^2)$ and $\varepsilon_i \sim N(\ln \sigma_0, \lambda^2)$.[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 nonlinear 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 $t + l$: $\delta_l X(t) = X(t + l) - X(t)$, self-similarity of the associated pdf amounts to

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

(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 multiscaling in a series, a unique scaling exponent H is not appropriate 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

$$\delta_{sl} X(st) =_{\text{Law}} M_s \delta_l X(t).$$

(4)

It can be shown that such a scale-dependent multiplicative random modulation of $P(\delta_l X)$ leads to a nonlinear scaling of absolute moments. The stochastic process of Eq. (2) is an example of a process characterized by such nonlinear 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 log normally distributed random variable $\exp(\varepsilon_s)$.

---

[*]lux@bwl.uni-kiel.de.
[1]Though not expressed explicitly, $\bar{\sigma}^2$ can clearly be accommodated in Eq. (1) via P_σ.

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. Another moment-based estimator has been proposed by Beck [13], who uses the standardized fourth moment (or "flatness" of the pdf). 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 [14] for the causal Markov-switching multifractal (MSM) model of Calvet and Fisher [15]. While our methodology is also based on moment matching, it differs from the approaches of Beck [13] and 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 moments proposed by Beck and Kiyono *et al.* are computed from the marginal distribution $P_{\lambda,\sigma_0}(x)$ of Eq. (1). These moments are exact for the multiplicative log-normal 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 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 previous approaches 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 Sec. VI presents empirical results for both parameter estimation and forecasting for a sample of financial data. Section VII concludes and the appendices collect explicit formulas for the particular moments used in our GMM and best linear forecasting approaches.

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 log-normal cascade as a compound process on the bounded interval $[1,2^n]$ by multiplying the sum of the log-normal weights with a normally distributed random variable ξ:

$$x_i \doteq \left[\prod_{j=1}^{n} M_j\left(\left\lfloor \frac{i-1}{2^{n-j}} \right\rfloor \right) \right] \xi_i$$

$$= \exp\left[\sum_{j=1}^{n} \omega_j\left(\left\lfloor \frac{i-1}{2^{n-j}} \right\rfloor \right) \right] \xi_i, \qquad (5)$$

where $\lfloor \cdot \rfloor$ represents the floor function and $\xi_i \sim N(0,\bar{\sigma}^2)$. It is common to select $\omega_j(\cdot) \sim N(\tilde{\mu},\tilde{\sigma}^2)$ so that the sum of $\omega_j(\cdot)$ is $N(n\,\tilde{\mu},n\,\tilde{\sigma}^2)$ distributed. Hence, $\sum_{j=1}^{n}\omega_j(\cdot)$ in Eq. (5) corresponds to ε_i in Eq. (2), and x_i 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 $\omega_j(\cdot) \sim N(-\lambda_0^2,\lambda_0^2)$ 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 $\omega_j(\cdot) \sim N(-\lambda_0^2,\lambda_0^2)$. 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.

To overcome the statistical difficulties that may arise from such a nonstationary 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 log-normal cascade process is consequently given by

$$x_t \doteq \exp\left[\sum_{j=1}^{n} \omega_j^{(m)}\left(\left\lfloor \frac{t - 2^n\,(m-1)-1}{2^{n-j}} \right\rfloor \right) \right] \xi_t, \quad (6)$$

where again $\xi_t \sim N(0,\bar{\sigma}^2)$.[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

---

[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.
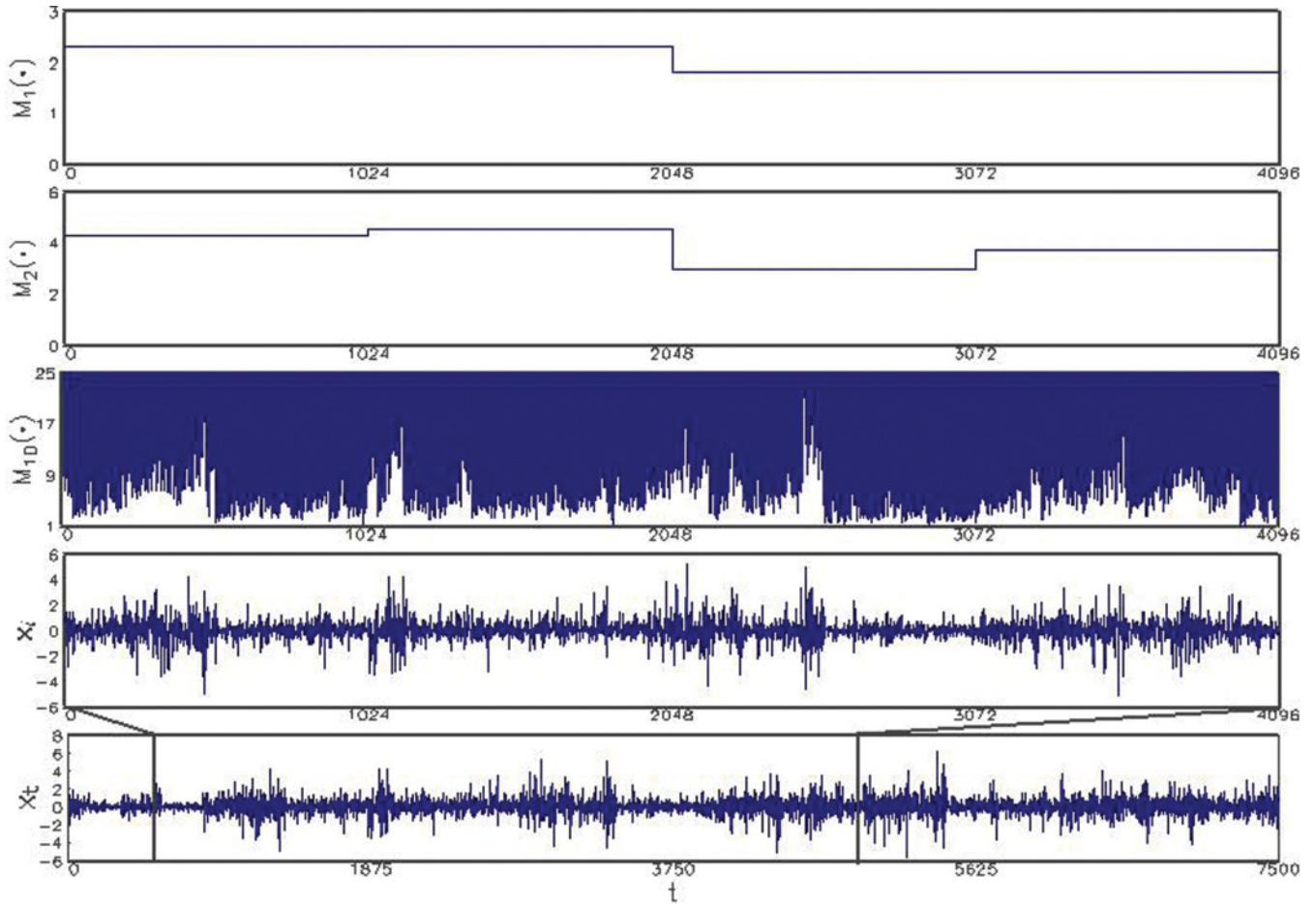
FIG. 1. (Color online) Illustration of one sample of the $\{x_t\}_{t=1}^{\infty}$ process of Eq. (6) with $n = 12$. (From top to bottom) The first level of draws of log-normal random variables, the second level, the 10th level, the corresponding bounded $\{x_i\}_{i=1}^{2^n}$ process according to Eq. (5), and a sample of 7500 points of the stationary $\{x_t\}_{t=1}^{\infty}$ process of Eq. (6). 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.

signals) that have clearly defined deterministic (e.g., sinusoidal) components (cf. Gardner *et al.* [16]).[3]

Note that the time series in Eq. (6) 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 $\{t_1 + z, \ldots, t_k + z\}$ 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 starting points.[4] The bottom panel of Fig. 1 shows a sample of 7500 observations of the $\{x_t\}_{t=1}^{\infty}$ process as a result of concatenating three $n = 12$-level bounded cascades.

Despite the nonstandard 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. (5) over a bounded interval only, the nonstationarity of the process would have followed trivially. As a consequence, standard "regularity conditions" (cf. Harris and Mátyás [17]) 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. (6). Due

---

[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.

[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 $\omega_j(t)$.

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:

$$\widehat{\varphi}_T = \arg\min_{\varphi \in \Phi} \; f_T(\varphi)' \, \Omega_T \, f_T(\varphi), \tag{7}$$

with Φ being the parameter space, $f_T(\varphi)$ the vector of differences between a set of sample and analytical moments, Ω_T a positive definite and possibly random weighting matrix, and $\widehat{\varphi}_T$ is the solution to this optimization problem [i.e., the argument (arg) that minimizes the objective function $f_T(\varphi)' \, \Omega_T \, f_T(\varphi)$, the weighted difference between sampled and analytical moments]; cf. Hansen [18]. Using log-absolute moments in the implementation of $f_T(\varphi)$, Lux [14] 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 log-normal 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 [14] autocovariances of the overlapping log differences $\zeta_{t,\ell}, \zeta_{t+1,\ell}, \ldots$. A closer look at these yields

$$\begin{aligned} \zeta_{t,\ell} &\doteq \ln|x_t| - \ln|x_{t-\ell}| \\ &= \sum_{j=1}^{n} [\omega_j(t) - \omega_j(t-\ell)] + \ln|\xi_t| - \ln|\xi_{t-\ell}|. \end{aligned} \tag{8}$$

As one can see, these log-absolute differences remain unaffected by $\bar{\sigma}$, the scale factor in Eq. (6) 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:

$$\mathfrak{M}(\ell, p) = \mathbb{E}\big[\zeta_{t+\ell,\ell}^{p} \, \zeta_{t,\ell}^{p}\big], \tag{9}$$

together with a raw moment like $\mathbb{E}[x_t^2] = \bar{\sigma}^2$ for the identification of $\bar{\sigma}$ [note that we have standardized the distribution of $\omega_j(\cdot)$ in a way to guarantee that the second moment of the first term on the right-hand side of Eq. (6) is equal to unity; cf. also Appendix B]. 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. (9).

The estimator $\widehat{\lambda}_q^2$ 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):

$$\widehat{\lambda}_q^2 = \frac{2}{q(q-2)}\left[\ln\left(\frac{\sqrt{\pi}\,\mathbb{E}[|x|^q]}{2^{q/2}} \right) - \ln\Gamma\left(\frac{q+1}{2} \right) \right], \tag{10}$$

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 *not* 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 $\widehat{\lambda}_q^2$ is not derived from the exact pdf of the cascade process, it will in all likelihood be an inconsistent estimator for such a model. The same applies to the traditional $\widehat{\chi}^2$ estimator of Castaing *et al.* [1] and the flatness estimator $\widehat{F}_2$ of Beck [13]. 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 $\omega_j(t) \sim N(-\lambda_0^2, \lambda_0^2)$ in Eq. (6), and so as stated before, $\widehat{\lambda}_q^2$ 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 $\widehat{\lambda}_q^2$ 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 the aforementioned estimators. To this end, we first apply Kiyono *et al.*'s [12] standardization for the generation of the data. In the following, we choose $q = 0.5$ for the $\widehat{\lambda}_q^2$ estimator and consider 20 bins with equal probability mass for the $\widehat{\chi}^2$ estimator; that is, the pdf of Eq. (1) is binned using a varying split of the support $[x_{j-1}, x_j]$, for $j = 1, \ldots, 20$, so that each bin contains a probability mass of $F_{\lambda,\sigma_0}(x_j) - F_{\lambda,\sigma_0}(x_{j-1}) \approx 0.05$, where F represents the cumulative distribution function associated to Eq. (1).[5]

We apply both estimators for various sample lengths T_i, namely $T_1 = 2500$, $T_2 = 5000$, and $T_3 = 10\,000$. The GMM procedure aims at exploiting the intermittency at different cascade levels, and therefore, the moments in Eq. (9) depend on the choice of the number and values of lags ℓ. After many trials, for which results are not presented here, we found that using three lags leads to a good compromise between computational speed and quality of the estimates. In 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.* [20]).

Table I shows the results from our GMM approach, while Table II presents the outcomes from the older methods, where we have normalized the results in Table II by the total number of cascade levels n for better comparability. As we can infer from both tables, all estimators start out very similarly with

---

[5]Results for alternative choices other than the ones presented in the tables are available upon request, cf. Ref. [19].
[6]The value $\ell = 14$ is as good as any in (8,16).

TABLE I. Monte Carlo results for GMM estimator.[a]

n		$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.15$		
		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	0.047	0.049	0.050	0.142	0.146	0.148
	RMSE	0.009	0.008	0.005	0.017	0.011	0.007	0.024	0.016	0.011
	$\bar{\sigma}$	1.000	1.002	1.000	1.000	0.995	0.998	0.969	0.967	0.993
	RMSE	0.039	0.027	0.019	0.094	0.066	0.048	0.216	0.161	0.139
16	$\bar{\lambda}_0^2$	0.010	0.011	0.011	0.047	0.049	0.050	0.142	0.146	0.148
	RMSE	0.010	0.007	0.005	0.016	0.011	0.007	0.024	0.017	0.012
	$\bar{\sigma}$	0.976	0.979	0.993	0.916	0.900	0.917	0.611	0.661	0.740
	RMSE	0.224	0.207	0.182	0.521	0.433	0.372	0.676	0.761	0.840

[a]All simulations are based on a process with $\xi \sim N(0,1)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. Sample lengths are as follows: $T_1 = 2500$, $T_2 = 5000$, and $T_3 = 10\,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}$ are the corresponding means of the estimated parameters. RMSE denotes the root mean squared error. GMM was executed using lags $\ell = 1, 14, 64$. For each case, 400 Monte Carlo runs have been carried out.

a slight advantage of the moment-based estimators of Kiyono *et al.* and Beck in terms of RMSE at relatively small parameter values λ_0^2 or low cascade levels n. This appears plausible as these scenarios are closest to the case of independent ε_i for which the latter would be a consistent estimator. For a fixed number of cascade levels n, however, the bias of the raw moment-based estimators increases considerably the higher λ_0^2 gets. Interestingly, the inconsistent $\hat{\chi}^2$ estimator still performs surprisingly well and even marginally better than our GMM estimator for $\lambda_0^2 = 0.15$ and $n = 8$. This indicates that with a

TABLE II. Monte Carlo results for $\hat{\chi}^2$, $\hat{\lambda}_q^2$, and $\hat{F}_2$ estimators.[a]

E_i	n		$\lambda_0^2 = 0.01$			$\lambda_0^2 = 0.05$			$\lambda_0^2 = 0.15$		
			T_1	T_2	T_3	T_1	T_2	T_3	T_1	T_2	T_3
$\hat{\chi}^2$	8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.050	0.050	0.050	0.149	0.151	0.152
		RMSE	0.004	0.003	0.002	0.008	0.006	0.004	0.019	0.013	0.009
		$\bar{\sigma}$	1.004	1.004	1.001	1.024	1.019	1.024	1.050	1.050	1.062
		RMSE	0.049	0.036	0.024	0.112	0.076	0.057	0.214	0.162	0.127
	16	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.034	0.036	0.040	0.119	0.125	0.129
		RMSE	0.005	0.004	0.003	0.020	0.018	0.015	0.057	0.048	0.040
		$\bar{\sigma}$	0.983	0.981	0.990	0.857	0.919	0.910	0.778	0.856	0.952
		RMSE	0.226	0.195	0.181	0.417	0.411	0.351	0.709	0.655	0.753
$\hat{\lambda}_q^2$	8	$\bar{\lambda}_0^2$	0.010	0.010	0.010	0.049	0.049	0.050	0.142	0.144	0.147
		RMSE	0.002	0.002	0.001	0.007	0.005	0.004	0.025	0.019	0.017
	16	$\bar{\lambda}_0^2$	0.007	0.007	0.008	0.033	0.036	0.040	0.090	0.100	0.111
		RMSE	0.004	0.003	0.003	0.019	0.016	0.013	0.064	0.055	0.044
$\hat{F}_2$	8	$\bar{\lambda}_0^2$	0.009	0.010	0.010	0.044	0.047	0.047	0.099	0.103	0.110
		RMSE	0.002	0.002	0.001	0.013	0.013	0.010	0.057	0.052	0.046
	16	$\bar{\lambda}_0^2$	0.006	0.007	0.008	0.026	0.030	0.033	0.052	0.059	0.066
		RMSE	0.004	0.004	0.003	0.025	0.022	0.019	0.099	0.092	0.085
$\hat{\sigma}$	8	$\bar{\sigma}$	1.000	1.001	1.000	0.997	0.992	0.996	0.973	0.970	0.993
		RMSE	0.037	0.026	0.019	0.084	0.063	0.046	0.223	0.156	0.141
	16	$\bar{\sigma}$	0.973	0.979	0.992	0.879	0.899	0.932	0.637	0.680	0.758
		RMSE	0.228	0.206	0.184	0.463	0.429	0.397	0.910	0.901	0.701

[a]All simulations are based on a process with $\xi \sim N(0,1)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. Sample lengths are as follows: $T_1 = 2500$, $T_2 = 5000$, and $T_3 = 10\,000$. $\bar{\lambda}_0^2$ and $\bar{\sigma}$ are the corresponding means of the estimated parameters, while RMSE denotes the root mean squared error. E_i denotes the estimation method: $\hat{\chi}^2$, Kiyono *et al.*'s estimator $\hat{\lambda}_q^2$, Beck's flatness or fourth moment estimator $\hat{F}_2$, and the simple sample standard deviation $\hat{\sigma}$. The $\hat{\chi}^2$ estimator was calculated using 20 bins with equal probability mass, that is, with a varying support size $[x_{j-1}, x_j]$, for $j = 1, \ldots, 20$, so that in each bin a probability mass of $F_{\lambda, \sigma_0}(x_j) - F_{\lambda, \sigma_0}(x_{j-1}) \approx 0.05$ is obtained. $\hat{\lambda}_q^2$ was calculated using $q = 0.5$ after the series was filtered by the sample estimate $\hat{\sigma}$. The same filtered series was employed for the $\hat{F}_2$ estimator, too. All entries referring to the parameter λ_0^2 were obtained by normalizing the resulting estimates by n. For each case, 400 Monte Carlo runs have been carried out.

limited number of cascade steps, the resulting pdf is still not too different from that of Eq. (1). However, the limitations of this approach become apparent when looking at the case $n = 16$, which shows sizable biases as compared to those from the GMM approach. Note that the relatively good performance of the $\widehat{\chi}^2$ estimator is in contrast to results displayed in Kiyono *et al.* [12]. Experiments with different settings indicate that the relatively favorable results of the $\widehat{\chi}^2$ estimator in Table II are due to our use of bins with equal probability mass, while other bin structures would typically give worse outcomes.

Equivalently, for a fixed λ_0^2, the bias of the moment-based estimators increases with a higher number of cascade levels n. 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 inter-mittency and a high number of cascade levels. The decrease of biases and sampling variability for the moment-based and $\widehat{\chi}^2$ estimators appears also much slower with increasing sample sizes than for the GMM estimator at high levels n. 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 do so. For high λ_0^2 or n RMSEs of the traditional methods appear almost constant across sample sizes.

Tables I and II also present the results for $\bar{\sigma}$. In Table II, only one sample standard deviation for the moment-based estimators is presented given that these only provide an estimate of λ_0^2. One observes that the GMM estimator in Table I agrees closely with the sample standard deviation in Table II, which actually is to be expected under the block-diagonal structure of the covariance matrix of the moment conditions. In all cases, the estimator seems to be more biased the higher both λ_0^2 and n, the number of cascade levels. As in Lux [14], 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 I shows that in the case $\lambda_0^2 = 0.15$ and $n = 16$, an apparent violation of the square-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 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 I and II) 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 III.

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 III, the absolute percentage difference (APD) between the estimates is at most 3%, 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 [14]). In conclusion, our GMM 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 [14] has introduced best linear forecasts to predict out-of-sample fluctuations of realizations of the causal Markov-switching multifractal process of Calvet and Fisher [15,21]. 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

----

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

TABLE III. Monte Carlo results for GMM with cascade-level uncertainty added.[a]

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
	RMSE	0.005	0.005	0.005	0.005	0.005	0.005	0.005
	APD	0.031	0.030	0.024	0.000	0.023	0.025	0.026
	$Q_{\min}$	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
	RMSE	0.007	0.007	0.007	0.007	0.007	0.007	0.007
	APD	0.006	0.003	0.001	0.000	0.001	0.001	0.001
	$Q_{\min}$	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
	RMSE	0.012	0.012	0.012	0.012	0.012	0.012	0.012
	APD	0.014	0.006	0.002	0.000	0.001	0.002	0.002
	$Q_{\min}$	1.156	1.025	1.002	1.000	1.002	1.003	1.004

[a]All simulations are based on a process with $n = 11$ cascade levels, $\xi \sim N(0,1)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, 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, while RMSE denotes the root mean squared error. 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. $Q_{\min}$ is the average of the objective value at the optimum for each n divided by the one with $n = 11$.

horizons h by

$$\widehat{Y}_{t+h} = \sum_{i=1}^{t} \phi_{t\,i}^{(h)} Y_{t+1-i} = \Phi_t^{(h)} \mathbf{Y}_t, \tag{11}$$

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 autocovariances and beyond (Ref. [22]).

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. (11), we use the iterative algorithm developed by Brockwell and Dahlhaus [[23], 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:

$$Y_t \doteq x_t^2 - \mathbb{E}[x_t^2] = x_t^2 - \widehat{\sigma}^2, \tag{12}$$

where $\widehat{\sigma}$ is the estimate of the scale factor $\bar{\sigma}$ in Eq. (6). Also, $\widehat{\sigma}^2$ appears only in the mean value of Eq. (12), but it drops from the coefficients $\phi_{ti}^{(h)}$. Appendix B presents the pertinent formulas for the variance and autocovariances 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 5000 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 an RW. We confine ourselves to a comparison of the quality of forecasts based on the GMM estimator, on the one hand, and on the Kiyono *et al.*'s estimator, on the other. A closer look at Table IV 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 $T = 5000$ 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.

The U shape of the prediction accuracy with varying λ_0^2 is reminiscent of similar observations in Lux [14]. Apparently, there are two opposite forces at work here: With small λ_0^2, an increase of this parameter leads to a better forecasting

TABLE IV. Monte Carlo assessment of best linear forecasts based on GMM and $\widehat{\lambda}_q^2$ estimation.[a]

| | | | $n = 10$ | | | | | | $n = 16$ | | | | | |
| | | | $\widehat{\lambda}_q^2$ | | | GMM | | | $\widehat{\lambda}_q^2$ | | | GMM | | |
	T	$\|\lambda_0^2$	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15	0.01	0.05	0.15
MSE	1		0.954	0.911	0.936	0.959	0.911	0.938	0.912	0.835	0.828	0.919	0.835	0.850
			(0.023)	(0.039)	(0.056)	(0.025)	(0.039)	(0.054)	(0.075)	(0.166)	(0.228)	(0.073)	(0.165)	(0.195)
	10		0.973	0.964	0.982	0.975	0.964	0.983	0.931	0.883	0.880	0.935	0.883	0.908
			(0.020)	(0.030)	(0.030)	(0.020)	(0.031)	(0.028)	(0.075)	(0.170)	(0.211)	(0.072)	(0.169)	(0.161)
	20		0.980	0.975	0.987	0.981	0.975	0.989	0.938	0.895	0.890	0.942	0.895	0.920
			(0.018)	(0.026)	(0.024)	(0.018)	(0.026)	(0.022)	(0.074)	(0.171)	(0.201)	(0.071)	(0.169)	(0.149)
	100		0.994	0.993	0.996	0.994	0.993	0.998	0.956	0.916	0.918	0.958	0.917	0.946
			(0.010)	(0.014)	(0.010)	(0.009)	(0.015)	(0.011)	(0.074)	(0.165)	(0.164)	(0.070)	(0.161)	(0.114)
MAE	1		0.965	0.897	0.901	0.968	0.897	0.897	0.936	0.806	0.833	0.941	0.807	0.855
			(0.052)	(0.112)	(0.127)	(0.049)	(0.113)	(0.126)	(0.140)	(0.278)	(0.325)	(0.134)	(0.273)	(0.232)
	10		0.977	0.952	0.970	0.979	0.952	0.964	0.948	0.857	0.908	0.952	0.859	0.924
			(0.045)	(0.096)	(0.088)	(0.042)	(0.097)	(0.091)	(0.137)	(0.278)	(0.297)	(0.131)	(0.268)	(0.189)
	20		0.983	0.967	0.981	0.984	0.967	0.975	0.954	0.875	0.925	0.957	0.877	0.938
			(0.040)	(0.084)	(0.072)	(0.038)	(0.085)	(0.080)	(0.135)	(0.272)	(0.280)	(0.128)	(0.262)	(0.173)
	100		0.994	0.992	0.997	0.995	0.993	0.992	0.968	0.913	0.960	0.970	0.915	0.963
			(0.021)	(0.042)	(0.030)	(0.020)	(0.046)	(0.060)	(0.122)	(0.241)	(0.221)	(0.115)	(0.227)	(0.133)

[a]The table shows mean squared errors (MSE) and mean absolute errors (MAE) for two different cascade sizes n. MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from an RW using the in-sample variance. All entries are averaged over 400 Monte Carlo runs (with standard errors given in parentheses). 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 5000 entries for parameter estimation and an adjacent out-of-sample period of 2500 entries for evaluation of forecasting performance were selected. Parameter values have been estimated with Kiyono's $\widehat{\lambda}_q^2$ and the GMM 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.

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 II) leads to smoother forecasts which on average might lead to somewhat smaller errors than forecasts based on a more accurate estimate.[8]

Next, we consider the case with added uncertainty on the number of cascade levels n of the series. Table V presents the forecasting results of a series generated by $n = 11$ and $T = 5000$ splitted into two subsamples of 2500 for estimation and forecasting. The process has been analyzed for a sequence of cascade levels ranging from 8 to 50, for which GMM estimation and subsequent out-of-sample forecasting 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 unboundedly 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.

VI. EMPIRICAL EVIDENCE

Starting with Ghashghaie *et al.* [8], a fair amount of effort has been spent particularly by physicists [7,12,24–27] on the analogy between turbulence and financial markets. In this strand of research, 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. Recently, multiscale descriptions have also been used for the prediction of financial time series (cf. Nawroth *et al.* [28]).

Our analysis is based on data from seven different foreign exchange markets: the Canadian dollar (CND), the Japanese yen (YEN), the Swedish krona (SEK), the Swiss franc (CHF), the Australian dollar (AUD), the Deutsche mark—extended by the Euro since 1999—(DEM-EUR), and the British pound (UKP), all against the US dollar. Furthermore, we have analyzed the price of gold in US dollars. All time series start on January, 2, 1979 and extend until July 2, 2010.

---

[8]While the lack of an advantage of the GMM estimates compared to the inconsistent moment estimator of Kiyono *et al.* might appear disappointing, we should note that the use of the latter for forecasting already implies quite some effort in computing exact moments [to implement Eq. (12)]. Hence, at this point a reliance on the moment estimator would be moot anyway.

TABLE V. Monte Carlo assessment of best linear forecasts with cascade-level uncertainty.[a]

| | | | | | | | n | | | | |
	T	8	9	10	11	12	13	14	$\cdots$	20	$\cdots$	50
MSE	1	0.896	0.893	0.891	0.890	0.890	0.890	0.890		0.890		0.890
		(0.073)	(0.078)	(0.081)	(0.083)	(0.084)	(0.085)	(0.085)		(0.086)		(0.086)
	10	0.954	0.947	0.843	0.941	0.940	0.940	0.939		0.939		0.940
		(0.051)	(0.064)	(0.072)	(0.078)	(0.081)	(0.083)	(0.084)		(0.085)		(0.085)
	20	0.970	0.961	0.956	0.953	0.951	0.951	0.951		0.951		0.951
		(0.038)	(0.054)	(0.066)	(0.074)	(0.078)	(0.081)	(0.082)		(0.084)		(0.084)
	100	1.000	0.992	0.983	0.977	0.974	0.972	0.972		0.971		0.971
		(0.013)	(0.018)	(0.036)	(0.052)	(0.063)	(0.070)	(0.073)		(0.077)		(0.077)
MAE	1	0.890	0.884	0.881	0.880	0.878	0.878	0.878		0.878		0.878
		(0.126)	(0.146)	(0.163)	(0.175)	(0.185)	(0.190)	(0.193)		(0.198)		(0.199)
	10	0.948	0.940	0.934	0.931	0.930	0.929	0.929		0.928		0.928
		(0.082)	(0.111)	(0.136)	(0.157)	(0.172)	(0.182)	(0.187)		(0.195)		(0.197)
	20	0.967	0.957	0.950	0.947	0.945	0.944	0.944		0.943		0.943
		(0.061)	(0.090)	(0.119)	(0.143)	(0.162)	(0.174)	(0.180)		(0.190)		(0.192)
	100	0.999	0.991	0.984	0.979	0.976	0.975	0.975		0.975		0.975
		(0.038)	(0.040)	(0.063)	(0.094)	(0.122)	(0.141)	(0.152)		(0.168)		(0.172)

[a]All simulations are based on a process with $n = 11$ cascade levels, $\xi \sim N(0,1)$, $\omega_j \sim N(-\lambda_0^2, \lambda_0^2)$, and $\sigma = 1$. MSE and MAE are given in % of the pertinent MSEs and MAEs of a naive forecast from an RW 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 5000 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 2500 entries for parameter estimation and an adjacent out-of-sample period of 2500 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.

Due to the slight variation in the number of active trading days among markets, we use the first ≈ 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 5500 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 $\omega_j(t) \sim N(-\lambda_0^2, \lambda_0^2)$.[9] Table VI 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 Hansen's test statistics $J_T = T f_T(\widehat{\varphi})' \Omega_T f_T(\widehat{\varphi})$, 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 one to reject the multifractal cascade as the data-generating process only for the Swiss francs (CHF), on the basis of our chosen moment functions. Furthermore, 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).

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 Tables VII and VIII, 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 [14] and Calvet and Fisher [21] who arrive at a similar conclusion regarding the saturation of forecasting performance beyond a certain threshold.

As one can observe in Tables VII and VIII, 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 [29,30].[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 $n = 20$ log-normal 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

---

[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. (6). 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.

[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 nonadjusted version of the test applies.

TABLE VI. Empirical parameter estimates.[a]

	Asset							
	CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
$\widehat{\lambda}_0^2$	0.0154	0.0357	0.0167	0.0119	0.0357	0.0238	0.0191	0.0339
(SE)	(0.006)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.007)
M-APD	0.021	0.006	0.003	0.165	0.017	0.011	0.005	0.003
$\widehat{\sigma}$	0.2781	0.6945	0.6556	0.7533	0.6208	0.7144	0.6546	1.3499
(SE)	(0.009)	(0.021)	(0.038)	(0.017)	(0.032)	(0.018)	(0.020)	(0.107)
J_{prob}	0.243	0.843	0.718	0.005	0.500	0.695	0.550	0.260
n_{min}	8	8	19	8	8	20	20	12

[a]Empirical estimates for standardized cascade shocks $\omega_j(t) \sim N(-\lambda_0^2, \lambda_0^2)$ and overall order of fluctuation magnitude $\bar{\sigma}$, obtained via GMM from a sample of 5500 entries for each asset. Each column shows the estimate corresponding to the lowest objective function obtained for the range of cascade sizes $n = 8, \ldots, 20$ 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.

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 nonsignificant under the modified Diebold and Mariano [31] test statistic (Harvey *et al.* [32]) at the 95 % level. Nevertheless, it is noteworthy that the cascade forecasts typically dominate 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 VIII display more significant results, with the $n = 20$ log-normal cascade forecasts performing almost always better than GARCH(1,1), with a statistical significance of 95%. Note that under the MAE criterion, the cascade model 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 [29,30] 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 log-normal cascade processes, which compares

TABLE VII. Empirical forecast: MSE.[a]

	T	Asset							
		CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
GMM20	1	0.801[c]	0.956[c]	0.852[c]	0.943[c]	0.786	0.925[c]	0.890[c]	0.911[c]
	5	0.808[c]	0.970[c]	0.840[c]	0.951[c]	0.851[c]	0.900[c]	0.877[c]	0.894[c]
	20	0.855[c]	0.989[c,d]	0.867[c]	0.980[c]	0.927[c]	0.914[c]	0.917[c]	0.917[c]
	50	0.913[c,d]	0.996[c]	0.919[c]	0.986[c]	0.968[c]	0.952[c]	0.962[c]	0.964[c]
	100	0.923[c]	0.999[c]	0.970[c]	1.015[c]	0.973[c]	0.988[c]	1.008[c]	0.984[c]
GARCH	1	0.792[c]	0.958[c]	0.836[c,d]	0.942[c]	0.788	0.917[c]	0.874[c,d]	0.903[c]
	5	0.794[c]	0.971[c]	0.827[c,d]	0.953[c]	0.902[b]	0.900[c]	0.877[c]	0.899[c]
	20	0.847[c]	0.996[c]	0.851[c]	0.985[c]	1.056[c]	0.913[c]	0.908[c]	0.917[c]
	50	0.933[c]	1.006[c]	0.899[c]	0.994[c]	1.203[c]	0.961[c]	0.948[c]	1.009[c]
	100	0.952[c]	1.011[c]	0.948[c]	1.013[c]	1.252[c]	1.000	0.997[c]	1.050[c]

[a]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 $\ell = 1, 14, 64$.
[b]Denotes an improvement against RW model which is significant at the 95% level.
[c]Denotes an improvement against RW model which is significant at the 99% level.
[d]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 [29,30]. Comparisons against GARCH(1,1) are based on the modified Diebold and Mariano [31] test statistic by Harvey *et al.* [32].

TABLE VIII. Empirical forecast: MAE.[a]

	T	Asset							
		CND	YEN	SEK	CHF	AUD	DM	UKP	Gold
GMM20	1	1.094	0.949[c]	1.051[d]	0.923[c,d]	1.090[d]	0.885[c,d]	0.911[c,d]	0.838[c,d]
	5	1.085	0.949[c,d]	1.045[d]	0.916[c,d]	1.093[d]	0.872[c,d]	0.905[c,d]	0.838[c,d]
	20	1.080	0.958[b,d]	1.035[d]	0.925[c,d]	1.107[d]	0.882[c,d]	0.910[c,d]	0.860[c,d]
	50	1.070	0.964	1.044[d]	0.932[b,d]	1.120[d]	0.900[c,d]	0.918[b,d]	0.887[c,d]
	100	1.061	0.969	1.060[d]	0.942	1.119[d]	0.915[b,d]	0.942[d]	0.904[d]
GARCH	1	1.103	0.950[c]	1.060	0.939[c]	1.172	0.908[c]	0.920[c]	0.849[c]
	5	1.088	0.957[c]	1.061	0.939[c]	1.219	0.907[c]	0.921[c]	0.857[c]
	20	1.073	0.974	1.068	0.961[c]	1.352	0.950[c]	0.937[b]	0.900[b]
	50	1.037[d]	0.988[b]	1.088	0.983[b]	1.641	1.016	0.955	0.991
	100	1.013[d]	0.998	1.119	0.998	2.082	1.064	0.994	1.091

[a]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 $\ell = 1, 14, 64$.

[b]Denotes an improvement against RW model which is significant at the 95% level.

[c]Denotes an improvement against RW model which is significant at the 99% level.

[d]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 [31] test statistic by Harvey *et al.* [32].

favorably with previously proposed $\widehat{\chi}^2$ and raw moment-based estimators. Our numerical analysis suggests that the GMM estimator is indeed consistent and asymptotical normally distributed. Furthermore, our methodology allows us to retrieve the cascade parameter value with high accuracy 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 end point 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 yields 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 nontrivial part of the variability of price fluctuations. This supports previous findings for the causal Markov-switching multifractal model (Calvet and Fisher [15]). 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.

APPENDIX A: DESCRIPTION OF THE ANALYTICAL MOMENTS

We first consider the following definition:

$$\eta_{t,\ell} \doteq \sum_{j=1}^{n} [\omega_j(t) - \omega_j(t - \ell)], \tag{A1}$$

where obviously it holds that $\mathbb{E}[\eta_{t,\ell}] = 0$. Lux [14] shows that the moments of $\zeta_{t,\ell}$, defined as in Eq. (8), can be written as

$$\mathbb{E}[\zeta_{t+\ell,\ell} \, \zeta_{t,\ell}] = \mathbb{E}[\eta_{t+\ell,\ell} \, \eta_{t,\ell}] + \mathbb{E}^2[\ln |\xi_i|] - \mathbb{E}[(\ln |\xi_t|)^2], \tag{A2}$$

and

$$\mathbb{E}\big[\varsigma_{t+\ell,\ell}^2 \, \varsigma_{t,\ell}^2\big] = \mathbb{E}\big[\eta_{t+\ell,\ell}^2 \, \eta_{t,\ell}^2\big] - 4\{\mathbb{E}\big[\eta_{t,\ell}^2\big] - \mathbb{E}[\eta_{t+\ell,\ell} \, \eta_{t,\ell}]\}\,\{\mathbb{E}^2[\ln |\xi_i|] - \mathbb{E}[(\ln |\xi_t|)^2]\} + 3\,\mathbb{E}^2[(\ln |\xi_t|)^2]$$
$$- 4\mathbb{E}[\ln |\xi_t|]\,\mathbb{E}[(\ln |\xi_t|)^3] + \mathbb{E}[(\ln |\xi_t|)^4]. \tag{A3}$$

As can be seen in Eqs. (A2) and (A3), the moments in Eq. (9) require one to compute $\mathbb{E}[\eta_{t,\ell}^2]$, $\mathbb{E}[\eta_{t+\ell,\ell} \, \eta_{t,\ell}]$, and $\mathbb{E}[\eta_{t+\ell,\ell}^2 \, \eta_{t,\ell}^2]$. 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, $\eta_{t,\ell}^2$ elements would be different from zero if the pertinent ω_j has different realizations in $t + \ell$ and t. Therefore, the expected value of $\eta_{t,\ell}^2$ is computed as

$$\mathbb{E}\big[\eta_{t+\ell,\ell}^2\big] = \left(\frac{2}{2^n}\right) \sum_{j=1}^n \{\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]\}\,\{\mathbb{I}(2^{n-j} > \ell)(2^j\,\ell) + \mathbb{I}(2^{n-j} \leqslant \ell)\,2^n\}$$
$$= \left(\frac{2\,\tilde{\sigma}^2}{2^n}\right) \sum_{j=1}^n \{\mathbb{I}(2^{n-j} > \ell)(2^j\,\ell) + \mathbb{I}(2^{n-j} \leqslant \ell)\,2^n\}, \tag{A4}$$

where $\mathbb{I}(\cdot)$ represents the indicator functions, ω_j is distributed $N(\tilde{\mu}, \tilde{\sigma}^2)$, and 2^{n-j} accounts for the length of each subinterval at cascade step j. To understand the computation of Eq. (A4) note that 2^j is the number of different log-normal 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. (A4), we distinguished between the cases $2^{n-j} > \ell$ and $2^{n-j} \leqslant \ell$. 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 $2^{n-j} \leqslant \ell$ 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

$$\eta_{t+\ell,\ell} \, \eta_{t,\ell} = \left(\sum_{j=1}^n [\omega_j(t + \ell) - \omega_j(t)]\right)$$
$$\times \left(\sum_{s=1}^n [\omega_s(t) - \omega_s(t - \ell)]\right). \tag{A5}$$

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. (A5) becomes

$$\eta_{t+\ell,\ell} \, \eta_{t,\ell} = \sum_{j=1}^n \big[\omega_j(t + \ell)\,\omega_j(t) - \omega_j^2(t) - \omega_j(t + \ell)$$
$$\times \omega_j(t - \ell) + \omega_j(t)\,\omega_j(t - \ell)\big].$$

Furthermore, 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

$\mathbb{E}[(\omega_j)^2] - \mathbb{E}^2[(\omega_j)]$:

$$\mathbb{E}[\eta_{t+\ell,\ell} \, \eta_{t,\ell}]$$
$$= -\left(\frac{\tilde{\sigma}^2}{2^n}\right) \sum_{j=1}^n \{\mathbb{I}(2^{n-j} \leqslant \ell)\,2^n + \mathbb{I}(2^{n-j} > \ell)$$
$$\times [\mathbb{I}(2^{n-j} < 2\ell)\,[\mathbb{I}(j > 1)\,2^j\,(2\ell - 2^{n-j})]]\}. \tag{A6}$$

Equation (A6) can be understood following the sequence of different cases we distinguish: First, if $2^{n-j} \leqslant \ell$, 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 $\eta_{t,\ell}^2$ 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

$$\eta_{t+\ell,\ell}^2 \, \eta_{t,\ell}^2 = \left(\sum_{j=1}^n [\omega_j(t + \ell) - \omega_j(t)]\right)^2$$
$$\times \left(\sum_{s=1}^n [\omega_s(t) - \omega_s(t - \ell)]\right)^2, \tag{A7}$$

which requires one 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,

$$(\omega_j(t + \ell) - \omega_j(t))^2 (\omega_j(t) - \omega_j(t - \ell))^2. \tag{A8}$$

We count here the same number of entries as in the case of $\eta_{t+\ell,\ell} \, \eta_{t,\ell}$. 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. (A8), once the number of cases is counted, is

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

Together, we obtain

$$\kappa_1 = 6\,\tilde{\sigma}^4 \sum_{j=1}^{n} \{\mathbb{I}(2^{n-j} \leqslant \ell)\,2^n + \mathbb{I}(2^{n-j} > \ell)$$
$$\times [\mathbb{I}(2^{n-j} < 2\ell)\,[\mathbb{I}(j > 1)\,2^j(2\ell - 2^{n-j})]]\}. \quad \text{(A9)}$$

(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

$$(\omega_j(t+\ell) - \omega_j(t))^2(\omega_s(t) - \omega_s(t-\ell))^2. \quad \text{(A10)}$$

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 nonzero for the higher index, it is the lower index that determines whether the complete expression [Eq. (A10)] is vanishing or not.[12] Therefore, we define for $h > l$,

$$\psi(h) \doteq \mathbb{I}(2^{n-h} \leqslant \ell)\,[\mathbb{I}(2^{n-l} > \ell)\,2^l\,\ell + \mathbb{I}(2^{n-l} \leqslant \ell)\,2^n]$$
$$+ \mathbb{I}(2^{n-h} > \ell)\,\mathbb{I}(2^{n-h} < 2\ell)\,2^l\,(2\ell - 2^{n-h}). \quad \text{(A11)}$$

In Eq. (A11) we encounter the following cases: First, if $2^{n-h} \leqslant \ell$ and $2^{n-l} > \ell$, a number of 2^l boxes has each ℓ nonzero contributions for the double differences. Furthermore, if both $2^{n-h} \leqslant \ell$ and $2^{n-l} \leqslant \ell$, 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 nonzero entries, the expectation of Eq. (A10) is defined as

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

Together with the number of cases, we obtain

$$\kappa_2 = 4\,\tilde{\sigma}^4 \sum_{j=1}^{n} \sum_{s=1,\,s \neq j}^{n} [\mathbb{I}(j > s)\,\psi(j) + \mathbb{I}(j < s)\,\psi(s)]. \quad \text{(A12)}$$

(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,

$$(\omega_j(t+\ell) - \omega_j(t))(\omega_j(t) - \omega_j(t-\ell))$$
$$\times (\omega_s(t+\ell) - \omega_s(t))(\omega_s(t) - \omega_s(t-\ell)). \quad \text{(A13)}$$

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.

---

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

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 $\eta_{t+\ell,\ell}\,\eta_{t,\ell}$, so for $h > l$ we define the counting formula as

$$\varphi(l) = \sum_{l=1}^{n} \{\mathbb{I}(2^{n-l} \leqslant \ell)\,2^n + \mathbb{I}(2^{n-l} > \ell)$$
$$\times [\mathbb{I}(2^{n-l} < 2\ell)\,[\mathbb{I}(l \geqslant 1)\,2^l\,(2\ell - 2^{n-l})]]\}. \quad \text{(A14)}$$

Given the previous explanations, the components of Eq. (A14) are easily explained.

Using these results and the fact that the expectation of Eq. (A13) is $2\,\tilde{\sigma}^4$ we find

$$\kappa_3 = 2\,\tilde{\sigma}^4 \sum_{j=1}^{n} \sum_{s=1,\,s \neq j}^{n} [\mathbb{I}(j > s)\,\varphi(s) + \mathbb{I}(j < s)\,\varphi(j)]. \quad \text{(A15)}$$

Putting (A9), (A12), and (A15) together we finally obtain

$$\mathbb{E}\big[\eta_{t+\ell,\ell}^2\,\eta_{t,\ell}^2\big] = \left(\frac{1}{2^n}\right)(\kappa_1 + \kappa_2 + \kappa_3). \quad \text{(A16)}$$

As the last element we need, the log-absolute moments of the standard normal variates ξ_i in Eq. (A3) can be easily obtained by using the Gamma function and its derivatives.

APPENDIX B: FORECASTING FORMULAS

For computing linear forecasts of $x_t^2 - \widehat{\sigma}^2$ in Eq. (12), we need the second moment and the autocovariances of the x_t. Let us define

$$\vartheta_t \doteq \exp\left[2 \sum_{j=1}^{n} \omega_j^{(m)}\left(\left\lfloor \frac{t - 2^n(m-1) - 1}{2^{n-j}} \right\rfloor\right)\right]$$
$$= \exp\left[2 \sum_{j=1}^{n} \omega_j(t)\right] = \left[\prod_{j=1}^{n} M_j(t)\right]^2, \quad \text{(B1)}$$

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

$$\mathbb{E}[\vartheta_t^2] = \mathbb{E}\left[\exp\left[4 \sum_{j=1}^{n} \omega_j(t)\right]\right] = \mathbb{E}\left[\left(\prod_{j=1}^{n} M_j(t)\right)^4\right]$$
$$= \exp\left(4\,n\,\lambda_0^2\right), \quad \text{(B2)}$$

while the autocovariance of the volatility process is given by

$$
\mathbb{E}[\vartheta_{t+\ell}\,\vartheta_t] = \mathbb{E}\left[\prod_{j=1}^{n}\left[M_j^2(t)M_j^2(t+\ell)\,\mathbb{I}(M_j(t)\neq M_j(t+\ell)) + M_j^4(t)\,\mathbb{I}(M_j(t)=M_j(t+\ell))\right]\right]
$$

$$
= \prod_{j=1}^{n}\left[\mathbb{E}\left[M_j^2(t)M_j^2(t+\ell)\,\mathbb{I}(M_j(t)\neq M_j(t+\ell))\right] + \mathbb{E}\left[M_j^4(t)\,\mathbb{I}(M_j(t)=M_j(t+\ell))\right]\right]
$$

$$
= \prod_{j=1}^{n}\left[\mathbf{1}\,\mathbb{I}(2^{n-j}\leqslant\ell) + \left[\exp\left(4\,\lambda_0^2\right)\frac{2^n-2^j\,\ell}{2^n} + \mathbf{1}\,\frac{2^j\,\ell}{2^n}\right]\mathbb{I}(2^{n-j}>\ell)\right].
$$

[1] B. Castaing, Y. Gagne, and E. J. Hopfinger, Physica D **46**, 177 (1990).
[2] S. Carius and G. Ingelman, Phys. Lett. B **252**, 647 (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**, 058101 (2005).
[6] A. Arneodo, N. Decoster, and S. G. Roux, Phys. Rev. Lett. **83**, 1255 (1999).
[7] K. Kiyono, Z. R. Struzik, and Y. Yamamoto, Phys. Rev. Lett. **96**, 068701 (2006).
[8] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature (London) **381**, 767 (1996).
[9] A. Fisher, L. E. Calvet, and B. Mandelbrot, Multifractality of Deutschemark/US Dollar Exchange Rates, Discussion Paper 1166, Cowles Foundation, New Haven, 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**, 041113 (2007).
[13] C. Beck, Physica D **193**, 195 (2004).
[14] T. Lux, J. Business and Economic Statistics **26**, 194 (2008).
[15] L. E. Calvet and A. Fisher, J. Econometrics **105**, 27 (2001).
[16] W. A. Gardner, A. Napolitano, and L. Paura, Signal Process. **86**, 639 (2006).
[17] D. Harris and L. Mátyás, (Cambridge University Press, Cambridge, 1999).
[18] L. P. Hansen, Econometrica **50**, 1029 (1982).
[19] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.85.046114 for additional tables.
[20] L. P. Hansen, J. Heaton, and A. Yaron, J. Business and Economic Statistics **14**, 262 (1996).
[21] L. E. Calvet and A. Fisher, J. Financial Econometrics **2**, 49 (2004).
[22] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods* (Springer, New York, 1991).
[23] P. J. Brockwell and R. Dahlhaus, J. Econometrics **118**, 129 (2004).
[24] J. F. Muzy, J. Delour, and E. Bacry, The European Physical Journal B - Condensed Matter and Complex Systems **17**, 537 (2000).
[25] A. Arneodo, J. F. Muzy, and D. Sornette, The European Physical Journal B - Condensed Matter and Complex Systems **2**, 277 (1998).
[26] M. Ausloos and K. Ivanova, Phys. Rev. E **68**, 046122 (2003).
[27] N. Kozuki and N. Fuchikami, Physica A **329**, (2003).
[28] A. P. Nawroth, R. Friedrich, and J. Peinke, New J. Phys. **12**, 083021 (2010).
[29] T. E. Clark and K. D. West, J. Econometrics **135**, 155 (2006).
[30] T. E. Clark and K. D. West, J. Econometrics **138**, 291 (2007).
[31] F. X. Diebold and R. S. Mariano, J. Business and Economic Statistics **13**, 253 (1995).
[32] D. Harvey, S. Leybourne, and P. Newbold, Int. J. Forecasting **13**, 281 (1997).