Empirical Scaling Laws and the Aggregation of Non-stationary Data*

Lo-Bin Chang
Department of Applied Mathematics, National Chiao Tung University

Stuart Geman
Division of Applied Mathematics, Brown University

Abstract. Widely cited evidence for scaling (self-similarity) of the returns of stocks and other securities is inconsistent with virtually all currently-used models for price movements. In particular, state-of-the-art models provide for ubiquitous, irregular, and oftentimes high-frequency fluctuations in volatility ("stochastic volatility"), both intraday and across the days, weeks, and years over which data is aggregated in demonstrations of self-similarity of returns. Stochastic volatility renders these models, which are based on variants and generalizations of random walks, incompatible with self-similarity. We show here that *empirical* evidence for self-similarity does not actually contradict the analytic lack of self-similarity in these models. The resolution of the mismatch between models and data can be traced to a statistical consequence of aggregating large amounts of non-stationary data.

*The authors gratefully acknowledge the hospitality and support of the Academia Sinica Mathematics Department, Taipei, Taiwan, partial support of the Center of Mathematical Modeling & Scientific Computing and the National Center for Theoretical Science, Hsinchu, Taiwan, and financial support from the Office of Naval Research under contract N000141010933, the National Science Foundation under grant DMS-1007593, the Defense Advanced Research Projects Agency under contract FA8650-11-1-7151, and the National Science Council under grant 100-2115-M-009-007-MY2.

key words: random-walk models, self-similarity, stochastic volatility, market time

1 Background

The 1900 dissertation of Louis Bachelier, on The Theory of Speculation (Bachelier, 1900), proposed a random-walk model for security prices. The basic model, elaborated to accommodate heavy-tailed distributions and stochastic volatilities, still provides a compelling and nearly universally accepted foundation for a theory of price movements. At the same time, a salient and much-discussed feature of the data is the remarkably precise self-similarity of the returns, relative to the return interval, of many of these securities. This was first observed by Mandelbrot (1963) and has since been found in multiple data sets involving a range of securities and time periods (cf. Evertsz, 1995, Mantegna & Stanley, 1995, Guillaume et al., 1997, Gopikrishnan et al., 1999, Podobnik et al., 2000, Ivanov et al., 2001, Wang and Hui, 2001, Gencay et al., 2001, Xu & Gencay, 2003, Ivanov et al., 2004, Matteo et al., 2005, Matteo, 2007, Glattfelder et al., 2011, and Podobnik et al., 2011, to name a few). With a straightforward calculation we will conclude that state-of-the-art models of price movements do not generate self-similar processes (see §2), and are therefore at odds with the empirical scaling of returns. We will then show (§3) that scaling of *empirical* distributions is likely to be a statistical consequence of the aggregation of large amounts of *non-stationary* data.

Bachelier's remarkable thesis included a first construction of Brownian motion, and proposed a suitably scaled version as a model for the price dynamics of securities: $S(t) = S(0) + \sigma w(t)$, in which w is a "standard" Brownian motion and σ is the standard deviation of the change in price after one unit of time. The model has evolved, incrementally, to better accommodate theoretical and empirical constraints. For example, the realization that the scale of an ensuing price increment is typically and logically proportional to the current price, rather than independent of it, leads to the geometric (instead of linear) Brownian motion:

$$R(t) \doteq \ln S(t) - \ln S(0) = \sigma w(t) \tag{1.1}$$

after correcting for a possible drift associated with risk-free investment.

Additionally, the common observation that returns are too peaked and heavy-tailed to be consistent with the normal distribution led Mandelbrot (1963) to seek a replacement for the Brownian motion, while preserving the compelling argument that increments of prices arise from large numbers of small influences. As stable processes are the only possible limits of rescaled sums of independent random variables (the "generalized central limit theorem," Lévy, 1925), and as the resulting theoretical return distributions are a better, and often excellent, fit to empirical returns, Mandelbrot proposed models of the same form as (1.1) but with $w(t)$ interpreted more generally as an α-stable Lévy process, $\alpha \in (0, 2]$. The special case $\alpha = 2$ recovers ordinary Brownian motion.

Further refinements are dictated by the fact that volatilities, modeled by the scaling factor σ (which is a standard deviation only in the case $\alpha = 2$), are almost never constant ($\sigma = \sigma(t)$, "stochastic volatility" cf. Shephard, 2005 and Shapira, 2011). And in fact the evidence is for very rapid fluctuations in $\sigma(t)$ (e.g. the left-hand panel in Figure 1 is typical). A parsimonious extension of (1.1), whether or not $\alpha = 2$, is through the stochastic integral

$$R(t) = \ln S(t) - \ln S(0) = \int_0^t \sigma(s) dw(s) \tag{1.2}$$

which falls out of the same thought experiment that took us from discrete and small price movements to the stable process $w(t)$, except that a step at time t has scale proportional to $\sigma(t)$ rather than σ.

Many lines of thought lead to more-or-less the same thing. For example, the function $\sigma(t)$ can be thought of as itself a stochastic process, dependent or independent of w, or as a given deterministic (perhaps historical) volatility trajectory. Many authors prefer to think of $\sigma(t)$ as a proxy for, or measure of, market activity or "market time," and in fact under very general conditions the result of a random time change can also be expressed by (1.2), cf. Clark (1973), Geman et al. (2001) and Veraart & Winkel (2010). We will assume that either $\sigma(t)$ is deterministic or, if stochastic, it is independent of w, in which case we will condition on a sample path of $\sigma(t)$ so that the two situations amount to the same thing. In any case, it would be a mistake to think of $\sigma(t)$ as statistically *stationary*, given the prototypical intraday volatility profile (including high values in the opening and closing thirty minutes) and the overall rise in volatility with rise in volume through the years and decades over which return profiles are studied.

The question we wish to examine is an apparent incompatibility between the class of models embodied by equation (1.2) and the widely cited evidence for scaling of returns on stock prices and other financial processes.

2 Are Random-Walk Models Consistent with Empirical Scaling?

Assume (1.2) for $t \in [0, T]$ and consider the sequence of returns, $R_1^{(h)}, R_2^{(h)}, \ldots, R_N^{(h)}$, over intervals of length h, where

$$R_k^{(h)} = \ln \frac{S(kh)}{S((k-1)h)} = \int_{(k-1)h}^{kh} \sigma(s)dw(s), \quad k = 1, 2, \ldots, N$$

and $N = N^{(h)} = \lfloor T/h \rfloor$ (i.e. T over h rounded down to the nearest integer). The observations of Mandelbrot and others suggest that the distribution of $R_k^{(h)}$ is nothing more than a scaled version of the distribution of $R_k^{(1)}$. In other words, for some scaling parameter α the probability mass function of $R_k^{(h)}$, say $f_k^{(h)}(r)$, is the same as the probability mass function of $h^{1/\alpha}R_k^{(1)}$, in which case $f_k^{(h)}(r) = h^{-1/\alpha} f_k^{(1)}(h^{-1/\alpha} r)$, by change of variables. Indeed, in a particularly innovative and convincing approach, Mantegna & Stanley (1995) compare $f_k^{(h)}(0)$ (the probability of zero return over h minutes) to $f_k^{(1)}(0)$ (the corresponding probability over 1 minute) and find that by setting $\alpha = 1.40$ *the entire distribution of returns,* $f_k^{(h)}(r)$*, on the S&P 500 from the five-year period 1984–1989, is virtually indistinguishable from the scaled distribution of one-minute returns,* $h^{-1/\alpha} f_k^{(1)}(h^{-1/\alpha} r)$, over the three orders of magnitude $h = 1$ minute to $h = 1{,}000$ minutes.

If, in 1.2, $w(t)$ is an α-stable Lévy process, as would be assumed of a random-walk model, then $\mathcal{L}\{w(ht)\} = \mathcal{L}\{h^{1/\alpha} w(t)\}$ (henceforth $w(ht) \sim h^{1/\alpha} w(t)$, for short), meaning that the two stochastic processes, $w(ht)$, $t \geq 0$, and $h^{1/\alpha} w(t)$, $t \geq 0$, are statistically indistinguishable. We would appear to have an explanation for the empirical scaling of returns: the returns simply inherit the scaling characteristics of the α-stable process w, with, for example, something approximating $\alpha = 1.4$ in the model generating the data examined by Mantegna & Stanley. But, concerning returns,

$$R_k^{(h)} = \int_{(k-1)h}^{kh} \sigma(s)dw(s) = \int_{(k-1)}^{k} \sigma(hs)dw(hs) \sim h^{1/\alpha} \int_{(k-1)}^{k} \sigma(hs)dw(s) \neq h^{1/\alpha} R_k^{(1)} \quad (2.1)$$

unless $\sigma(t)$ is constant.

To the contrary, the weight of evidence is that $\sigma(t)$ is far from constant, at least over the return intervals at which scaling behavior is often demonstrated. Consider, for example, the 101-minute sample trajectory displayed in the left-hand panel of Figure 1, which was drawn from a GARCH model with $p = q = 10$ (e.g. Shephard, 2005), estimated from the entire year of one-minute returns of IBM in 2005. (Lower-order fits, with smaller p and or q, estimated over different securities in different eras, all look and behave similarly.) Fluctuations are extreme, even over very short intervals. The GARCH model assumes Gaussian noise ($\alpha = 2$), but there is plenty of statistical evidence for equally rapid changes in volatility whether or not $\alpha = 2$.

In light of equation (2.1), how can models that are consistent with (1.2) be reconciled with evidence that returns are stable (i.e. scaling) random variables: $R^{(h)} \sim h^{1/\alpha} R^{(1)}$, for suitable α?

In fact, even when returns are generated artificially from (1.2) they appear to scale, as in the experimental results shown in the right-hand panel of the figure, where the GARCH data was used in (1.2) to produce prices and their returns at multiple intervals.[1] We offer the following theorem, together with the ensuing discussion about rate of convergence, as a plausible explanation.

3 Analytic Results and Discussion

We consider the class of models defined by (1.2), where $w(t)$ is a generalized random walk ("α-stable Lévy processes"), and where $\sigma(t)$ is assumed to be either deterministic or independent of the random walk w. In either case we assume that $\sigma(t)$ is continuous.

We are interested in studying the empirical return distribution, which is an aggregation of the observed returns over the interval $t \in [0, T]$. We anticipate showing that the collection of returns $R_k^{(h)}$, $k = 1, 2, \ldots, N^{(h)}$, will generate a histogram (empirical distribution) indistinguishable from that of $h^{1/\alpha)} R_k^{(1)}$, $k = 1, 2, \ldots, N^{(1)}$, regardless of the return interval h. Equivalently, our goal is to show that the collections $h^{-1/\alpha} R_k^{(h)}$, $k = 1, 2, \ldots, N^{(h)}$, produce the same empirical distribution. With this in mind, we define the cumulative distribution

[1] A similar experiment can be performed with $\alpha \neq 2$ by re-interpreting the GARCH-generated volatility trajectory, $\sigma(t)$, $t \geq 0$, as the scaling function of an α-stable process, as in equation (1.2). This involves generating α-stable variables, which can be done with modern software easily and efficiently. We performed this experiment systematically for α values between 0.5 and 2.0; in each case the results were virtually identical to those shown in the Figure 1.

function of the scaled returns $h^{-1/\alpha} R_k^{(h)}$ by

$$\hat{F}^{(h)}(r) = \frac{1}{N^{(h)}} \sum_{k=1}^{N^{(h)}} \mathbb{1}_{(-\infty,r)}(h^{-1/\alpha} R_k^{(h)})$$

where $\mathbb{1}_A(x)$ is the "indicator function," indicating $x \in A$ with 1 and $x \notin A$ with 0. For each r and h, $\hat{F}^{(h)}(r)$ is the fraction of the scaled returns that are less than r.

The theorem below identifies a single cumulative distribution function, $F(r)$, as the limit of each of the empirical distributions, $\hat{F}^{(h)}(r)$, thereby offering a resolution of the mismatch between observations of self-similarity and models of the form (1.2).

Using $\sigma(t)$, define a random volatility V by $V = \sigma(U)$, where U has the uniform distribution on $[0, T]$, and let W be independent of V with $W \sim w(1)$, where $w(t)$ is the α-stable process in (1.2). Let $F(r) = Prob\{VW < r\}$. Then

Theorem.

$$\sup_r |\hat{F}^{(h)}(r) - F(r)| \to 0$$

as $h \downarrow 0$.

Various forms of convergence of the empirical distributions $\hat{F}^{(h)}(r)$ to $F(r)$ are equivalent. Please see the appendix for precise statements about the convergence in the theorem, some of the equivalent formulations, and for the proof itself along with a detailed discussion and analysis of the proof.

Hence the properly scaled empirical distribution on returns is independent of the size of the return interval, for all intervals sufficiently small, i.e. the empirical distributions scale. Specifically, the empirical distribution on returns, standardized by $h^{-1/\alpha}$, approaches a mixture of distributions of α-stable variables. The mixture is multiplicative, and defined by the distribution of volatilities $\sigma(t)$, $t \in [0, T]$.

The assumed continuity of $\sigma(t)$ is a key part of the proof. Yet the extreme and rapid fluctuations in $\sigma(t)$, as seen for example through the 101-minute window in the left-hand panel of the figure, suggest that continuity *per se* is insufficient to explain the scaling properties of all but the smallest intervals. A close look at the proof of the theorem in fact highlights the *contributions* of these fluctuations to the near convergence of $\hat{F}^{(h)}$, already at large values of h. (See the appendix for a detailed discussion and analysis.) In other words, within the framework of pricing models of the type depicted in (1.2), non-constant volatility, and especially high-frequency fluctuations, *actually contribute to the empirical scaling of returns*, while at the same time ensuring that the returns themselves are not self similar. Empirical scaling is consistent with (1.2), and in fact can be viewed as support for rather than evidence against this class of models.

Finally, we remark on some additional practical and theoretical implications of non-stationary aggregation. Consider a trader who builds a strategy based, in part, on the idea that stock returns behave like a self-similar process. As indicated by the theorem, the trader can indeed estimate the scaling parameter, under a random-walk model, from aggregated data of stock returns. As for the practical implications, let $\mathcal{S} = \mathcal{S}(R^{(h)})$ be a function of the return sequence $R^{(h)}$ $(= (R_1^{(h)}, R_2^{(h)}, \cdots, R_N^{(h)}))$ representing a trading strategy, such as, for example, an event-triggered buy or sell order. If stock returns really were self-similar, then $\mathcal{S}(R^{(h)}) \sim \mathcal{S}(r^{\frac{1}{\alpha}} R^{(rh)})$, meaning that the single strategy would be equivalent to a multitude of strategies, each executed at its own time scale. But we have argued that *empirical* self-similarity is not the same thing as actual self-similarity, and that the latter is in fact inconsistent with stochastic volatility, at least in the random-walk model. In a finite-time experiment, the actual relationships among scales are subordinate to the actual trajectory of volatility, which is rarely well described as stationary, much less constant.

More generally, the practical challenge is to find invariant features in a dynamical system that is demonstrably non-stationary. It can not be a surprise that ever-more frequent introductions of new securities and derivatives, a relentless trend to higher volume, and the increasing pace of innovation in market making and trading strategies, render an assumption of temporal homogeneity less and less tenable. How are we to interpret, or even detect, empirical invariants? The challenge is to tease apart the non-stationary and oftentimes highly circumstantial trends (e.g. in volatility) from any preserved dimensions of the dynamics (e.g. a scaling exponent in a generalized random-walk model).

One approach is to make a weaker assumption: identify pairs of (generally multivariate) statistics under which the *conditional* distribution on one given the other can be reasonably assumed, or demonstrably approximated, to be homogeneous, e.g. independent of the particular day, week, month, or year of interest. This works well in neuroscience applications, where micro-electrode recordings from so-called repeated trials are never actual repeats in any strict statistical sense; the activities of billions of unrecorded neurons, reflecting the animal's fluctuating attention, mood, and general state of mind, are in no way stationary. But parameters can still be estimated and models can still be tested, by conditioning on variables that are observable but out of control of the experimenter (see Amarasingham et al. for a review).

In a study of various financial markets, Preis et al. (2011) demonstrate a remarkable invariant of the trajectory of volume activity in and around local price extrema, *when conditioning on* (and hence scaling by) the interval length to the previous extremum. Similarly, conditioning on a set of returns and then defining large returns ("excursions") *relative to the conditioned set*, can be shown to produce reliably invariant waiting times between excursions (see Hsieh et al., 2012, Chang et al., 2013). These and other methods of *conditional inference* can sidestep some of the barriers to coherent statistical analysis in complex non-stationary systems.

References

[1] A. Amarasingham, M. Harrison, N. Hatsopoulos, and S. Geman. Conditional modeling and the jitter method of spike resampling. *Journal of Neurophysiology*, 107:517–531, 2012.

[2] L. Bachelier. Théory de la spéculation. *Annales Scientifiques de l'Ecole Normale Supérieure*, 3(17):21–86, 1900.

[3] C. Chang, S. Geman, F. Hsieh, and C.-R. Hwang. Invariance in the recurrence of large returns and the validation of models of price dynamics. submitted for publication, 2013.

[4] P.K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973.

[5] C. Evertsz. Fractal geometry of financial time series. *Fractals*, 3(3):609–616, 1995.

[6] H. Geman, D.B. Madan, and M. Yor. Time changes for Lévy processes. *Mathematical Finance*, 11(1):79–96, 2001.

[7] R. Gencay, F. Selcuk, and B. Whitcher. Scaling properties of foreign exchange volatility. *Physica A*, 289:249–266, 2001.

[8] J.B. Glattfelder, A. Dupuis, and R.B. Olsen. Patterns in high-frequency fx data: discovery of 12 empirical scaling laws. *Quantitative Finance*, 11:599–614, 2011.

[9] P. Gopikrishnan, V. Plerou, L. Amaral, M. Meyer, and H. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60:5305–5316, 1999.

[10] D. Guillaume, M. Dacorogna, R. Davé, Ulrich A. Müller, R. Olsen, and O. Pictet. From the birds eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics*, 1:95–129, 1997.

[11] F. Hsieh, S.-C. Chen, and C.-R. Hwang. Discovering stock dynamics through multidimensional volatility phases. *Quantitative Finance*, 12:213–230, 2012.

[12] P.C. Ivanov, B. Podobnik, Y. Lee, and H.E. Stanley. Truncated levy process with scale-invariant behavior. *Physica A*, 299:154–160, 2001.

[13] P.C. Ivanov, A. Yuen, B. Podobnik, and Y. Lee. Common scaling patterns in intertrade times of u. s. stocks. *Physical Review E*, 69:056107(7 pages), 2004.

[14] P. Lévy. *Calcul des probabilités*. Gauthier-Villars, Paris, 1925.

[15] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[16] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1995.

[17] T.D. Matteo. Multi-scaling in finance. *Quantitative Finance*, 7:21–36, 2007.

[18] T.D. Matteo, T. Aste, and M.M. Dacorogna. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking and Finance*, 29:827–851, 2005.

[19] B. Podobnik, P.C. Ivanov, and H.E. Stanley. Scale-invariant truncated levy process. *Europhysics Letters*, 52(5):491–497, 2000.

[20] B. Podobnik, A. Valentincic, D. Hovatic, and H.E. Stanley. Asymmetric levy flight in financial ratios. *Proceedings of the National Academy of Sciences*, 108:17883, 2011.

[21] T. Preis, J.J. Schneider, and H.E. Stanley. Switching processes in financial markets. *Proceedings of the National Academy of Sciences*, 108(19):7674, 2011.

[22] Y. Shapira, D. Y. Kenett, Ohad Raviv, and E. Ben-Jacob. Hidden temporal order unveiled in stock market volatility variance. *AIP Advances*, 1(2):022127, 2011.

[23] N. Shephard. *Stochastic Volatility*. Oxford University Press, Oxford, 2005.

[24] A.E. Veraart and M. Winkel. Time change. In Rama Cont, editor, *Encyclopedia of Quantitative Finance*. Wiley Online Library, 2010.

[25] B.H. Wang and P.M. Hui. The distribution and scaling of fluctuations for hang seng index in hong kong stock market. *The European Physical Journal B*, 20:573–579, 2001.

[26] Z. Xu and R. Gencay. Scaling, self-similarity and multifractality in fx markets. *Physica A*, 323:578–590, 2003.

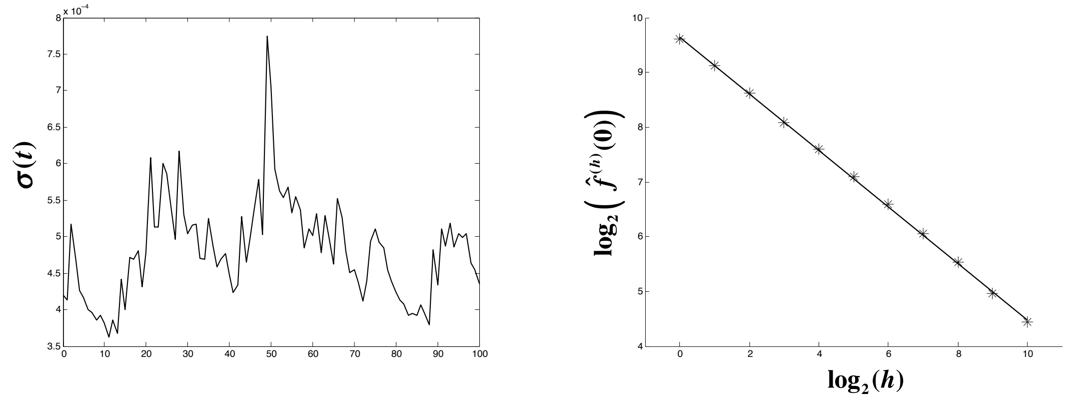

Figure 1: **Scaling of GARCH-simulated returns** One-minute returns on IBM for all of 2005 were used to fit a GARCH model, with autoregressive and moving average terms each of order 10 ($p = q = 10$). The model was used to generate $\sigma(t)$, for ten years of one-minute volatilities ($t = 1, 2, \ldots, 932{,}400$). Left-hand panel shows a typical window with 101 consecutive values of σ. Volatilities were used to generate ten-years of simulated one-minute returns, using (1.2) and a standard Brownian motion for ω. These were summed over disjoint intervals to produce the corresponding (simulated) returns $R_k^{(h)}$, $k = 1, 2, \ldots, N^{(h)}$, $h = 2^i$ for $i = 0, 1, \ldots, 10$. For each of the 11 return intervals, the ensemble of returns was used to estimate $f^{(h)}(0)$, the magnitude of the mass function at zero. Following the approach of Mantegna & Stanley (1995), if $R^{(h)} \sim h^{1/\alpha} R^{(1)}$ then $f^{(h)}(r) = h^{-1/\alpha} f^{(1)}(rh^{-1/\alpha})$, and hence $f^{(h)}(0) = h^{-1/\alpha} f^{(1)}(0)$ and $\log f^{(h)}(0) = -\frac{1}{\alpha} \log h + \log f^{(1)}(0)$. The fit is excellent, as seen in the right-hand panel where the least-squares regression line is superimposed on the pairs $(\log h, \log \hat{f}^{(h)}(0))$, $\hat{f}^{(h)}$ denoting a nonparametric, kernel, estimator of $f^{(h)}$. (We used the Matlab library function *ksdensity*.) The slope of the regression is about -0.52, which together with the good fit could be mistaken as evidence for $R^{(h)} \sim h^{1/\alpha} R^{(1)}$ with $\alpha = 2$. We did not include the superposition of the eleven histograms of scaled returns, $h^{-1/\alpha} R_k^{(h)}$, $k = 1, 2, \ldots, N^{(h)}$, where $h = 2^i$, $i = 0, 1, \ldots, 10$, since they are indistinguishable.

4 Appendix: Proof and Discussion of the Empirical Scaling Theorem

We consider the (drift-corrected) model for price movement given by

$$\ln S(t) - \ln S(0) = \int_0^t \sigma(s) dw(s) \qquad t \in [0, T] \tag{4.1}$$

where $S(t)$ is the price of a security at time t, $w(t)$ is an α-stable Lévy process ("generalized random walk"), and $\sigma(t)$ is either a deterministic function or, if stochastic, then it is independent of w.

The α-stable Lévy process, w, is characterized by a "location parameter" (playing the role of a mean or mode) which we will take to be zero, a "scale parameter" (which determines units, and can be absorbed into $\sigma(t)$) which we will take to be one, a "scaling parameter" α (sometimes called the "stability parameter," under which $w(ht) \sim h^{1/\alpha} w(t)$) which can be anything in the interval $\alpha \in (0, 2]$, and finally a skewness parameter β (which determines the range of possible steps as well as their symmetry, or lack of symmetry) which we we will take to be in the interval $\beta \in (-1, 1)$, except when $\alpha = 1$ or $\alpha = 2$ (corresponding to $w(1)$ being the Cauchy or Gaussian distribution, respectively), for which β is always zero.[2] As for $\sigma(t)$, we assume that it is either deterministic or independent of the random walk w. In either case we assume that $\sigma(t)$ is continuous.

Define the return process $R_1^{(h)}, R_2^{(h)}, \ldots, R_N^{(h)}$, over intervals of length h, by

$$R_k^{(h)} = \ln \frac{S(kh)}{S((k-1)h)} = \int_{(k-1)h}^{kh} \sigma(s) dw(s), \quad k = 1, 2, \ldots, N$$

where $N = N^{(h)} = \lfloor T/h \rfloor$. And define the empirical cumulative distribution function of the the scaled returns $h^{-1/\alpha} R_k^{(h)}$ by

$$\hat{F}^{(h)}(r) = \frac{1}{N^{(h)}} \sum_{k=1}^{N^{(h)}} \mathbb{1}_{(-\infty, r)}(h^{-1/\alpha} R_k^{(h)})$$

where $\mathbb{1}_A(x)$ is the indicator function of the event $x \in A$. For each r and h, $\hat{F}^{(h)}(r)$ is the fraction of the scaled returns that are less than r.

Define the distribution of volatilities by the random variable $V = \sigma(U)$, where U has the uniform distribution on $[0, T]$, and let W be independent of V with $W \sim w(1)$, where $w(t)$ is the α-stable process in (1.2). Let $F(r) = Prob\{VW < r\}$. Then

Theorem.

$$\sup_r |\hat{F}^{(h)}(r) - F(r)| \to 0$$

in probability, as $h \downarrow 0$.

[2] For completeness, we could also allow $\beta = \pm 1$, when $\alpha \in (0, 1)$, though this would truncate the possible values of $w(1)$.)

Remarks:

1. In other words, for every $\epsilon > 0$:

$$\lim_{h \downarrow 0} Prob\{\sup_r |\hat{F}^{(h)}(r) - F(r)| > \epsilon\} = 0$$

2. Since $\sup_r(\cdot)$ is bounded 2, convergence in probability implies convergence of moments: $E[\sup_r |\hat{F}^{(h)}(r) - F(r)|^m] \to 0$ for all $m > 0$.

Proof. For any $\alpha \in (0, 2]$, $h > 0$, and $k = 1, 2, \ldots, N^{(h)}$, define

$$\bar{\sigma}_{k,\alpha}^{(h)} = \left\{ \frac{1}{h} \int_{(k-1)h}^{kh} \sigma(s)^{\alpha} ds \right\}^{1/\alpha}$$

Then

$$\{h^{-1/\alpha} R_k^{(h)}\}_{k=1}^{N^{(h)}} \sim \{\bar{\sigma}_{k,\alpha}^{(h)} W_k\}_{k=1}^{N^{(h)}}$$

where $W_1, W_2, \ldots, W_{N^{(h)}}$ are iid copies of $w(1)$.

Let $G(r) = Prob\{W < r\}$ be the distribution function of $w(1)$. (Then G is infinitely differentiable, and each of its derivatives is bounded.) In terms of G,

$$F(r) = \frac{1}{T} \int_0^T G(\frac{r}{\sigma(s)}) ds$$

By a well-known argument based on the continuity of F, it is enough to show that $\hat{F}^{(h)}(r) \to F(r)$ in probability, for each (fixed) r:

$$\begin{aligned} \hat{F}^{(h)}(r) - F(r) &= \hat{F}^{(h)}(r) - \frac{1}{T} \int_0^T G(\frac{r}{\sigma(s)}) ds \\ &\sim \frac{1}{N^{(h)}} \sum_{k=1}^{N^{(h)}} \left\{ 1_{\bar{\sigma}_{k,\alpha}^{(h)} W_k < r} - G(\frac{r}{\bar{\sigma}_{k,\alpha}^{(h)}}) \right\} &(4.2) \\ &+ \frac{1}{N^{(h)}} \sum_{k=1}^{N^{(h)}} \left\{ G(\frac{r}{\bar{\sigma}_{k,\alpha}^{(h)}}) - G(\frac{r}{\bar{\sigma}_{k,1}^{(h)}}) \right\} &(4.3) \\ &+ \frac{1}{N^{(h)}} \sum_{k=1}^{N^{(h)}} \left\{ G(\frac{r}{\frac{1}{h}\int_{(k-1)h}^{kh} \sigma(s) ds}) - \frac{1}{h} \int_{(k-1)h}^{kh} G(\frac{r}{\sigma(s)}) ds \right\} &(4.4) \\ &+ \left\{ \frac{1}{N^{(h)} h} \int_0^{N^{(h)} h} G(\frac{r}{\sigma(s)}) ds - \frac{1}{T} \int_0^T G(\frac{r}{\sigma(s)}) ds \right\} &(4.5) \end{aligned}$$

The expression in (4.2) has mean zero and variance at most $1/N^{(h)}$, and (4.5) is no bigger than $2/N^{(h)}$. As for (4.3) and (4.4), we note that the boundedness of $G(r)$ and of $G'(r)$ (i.e. the α-stable density function) imply the existence of a function $\gamma_{r,\alpha}(x) \geq 0$ such that $\lim_{x \to 0} \gamma_{r,\alpha}(x) = 0$ and

$$|G(\frac{r}{a}) - G(\frac{r}{b})| \leq \gamma_{r,\alpha}(|a - b|)$$

for any pair of non-negative numbers a and b. Set $a = \bar{\sigma}_{k,\alpha}^{(h)}$ and $b = \bar{\sigma}_{k,1}^{(h)}$ and note that, by the continuity of σ, $|a-b| \to 0$ uniformly in k, which takes care of (4.3). Finally, for each k replace

$$\frac{1}{h}\int_{(k-1)h}^{kh} G(\frac{r}{\sigma(s)})ds$$

by $G(\frac{r}{\sigma(s_k)})$ for some $s_k \in [(k-1)h, kh]$ (mean value theorem), and again use the continuity of σ to conclude that $|a-b| \to 0$, again uniformly in k, but this time with $a = \frac{1}{h}\int_{(k-1)h}^{kh}\sigma(s)ds$ and $b = \sigma(s_k)$. Hence (4.4) vanishes as $h \to 0$, and this completes the proof. □

As remarked in the main text, the experiments indicate that convergence is rapid, and in fact essentially complete even when $h = 1{,}024$. This observation, in the face of the rapid fluctuations of $\sigma(t)$ on a far shorter time scale, suggest that the continuity of $\sigma(t)$, in and of itself, is insufficient to completely explain the empirical results. Here we will take a closer look at the proof of the theorem, in order to highlight the contributions of the additional path properties of $\sigma(t)$, beyond just continuity, that explain the near convergence of $\hat{F}^{(h)}$ already at large values of h.

$|\hat{F}^{(h)}(r) - F(r)|$ is bounded by bounding the four expression (4.2), (4.3), (4.4), and (4.5). There is no mystery about the first and the fourth: the expressions in (4.2) and (4.5) are small simply because $N^{(h)}$ is large, the first by virtue of the law of large numbers, and the second because of a deterministic bound. Rapid convergence in (4.3) and (4.4) is less obvious. We conclude with a brief analysis of these terms, based on the presumed properties of typical volatility trajectories.

Beginning with (4.3), we note that the contribution is small provided that $\bar{\sigma}_{k,\alpha}^{(h)}$ is close to $\bar{\sigma}_{k,1}^{(h)}$. Depending on whether $\alpha > 1$ or $\alpha < 1$, one or the other term is larger by virtue of Jensen's inequality. The difference between the two depends on the distribution of $\sigma(t)$, $t \in [(k-1)h, kh]$, and not directly on the continuity or smoothness of σ. For most intervals this distribution is unimodal, in which case the uniform distribution on $\sigma([(k-1)h, kh])$ gives an upper bound on the percentage error in replacing $\bar{\sigma}_{k,\alpha}^{(h)}$ by $\bar{\sigma}_{k,1}^{(h)}$; this turns out to be quite small for all $\alpha \in (0, 2]$. A similar conclusion holds in (4.4)—the magnitude of the expression depends on the distribution of σ and not on the smoothness, *per se*. In particular, (4.4) can be re-written as the difference between $E[G(r/X)]$ and $E[G(r/Y)]$, where

$$X = \frac{1}{h}\int_{(K-1)h}^{Kh}\sigma(s)ds$$

with K uniformly distributed on $\{1, 2, \ldots, N^{(h)}\}$, and where $Y = \sigma(S)$ with S uniformly distributed on $[0, hN^{(h)}]$. Notice that non-stationarity or at least fluctuations in σ on time scales larger than h (e.g. a ten-year trend in increasing volatility) actually works to make the distributions of X and Y similar, and therefore contributes to making the contribution from (4.4) small.