Physica A 392 (2013) 5046–5052

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

# Empirical scaling laws and the aggregation of non-stationary data[☆,☆☆]

Lo-Bin Chang[a,*], Stuart Geman[b]

[a] Department of Applied Mathematics, National Chiao Tung University, Taiwan

[b] Division of Applied Mathematics, Brown University, United States

## HIGHLIGHTS

- Stable process is described to accommodate the observations of heavy-tailed returns.
- Non-stationarity is formulated into the return process through stochastic volatility models.
- Random-walk models for the return sequence are derived from the return process.
- The incompatibility between the random-walk model and scaling laws is pointed out.
- An aggregation theorem is proposed to offer a resolution of the mismatch problem.

ARTICLE INFO

*Article history:*
Received 23 January 2013
Received in revised form 1 May 2013
Available online 2 July 2013

*Keywords:*
Random-walk models
Self-similarity
Stochastic volatility
Market time

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.



## 1. Background

The 1900 dissertation of Louis Bachelier, on The Theory of Speculation [1], 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 [2] and has since been found in multiple data sets involving



☆☆ 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.

* Corresponding author. Tel.: +886 35722088.
*E-mail address:* lobin_chang@math.nctu.edu.tw (L.-B. Chang).

0378-4371/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.06.049

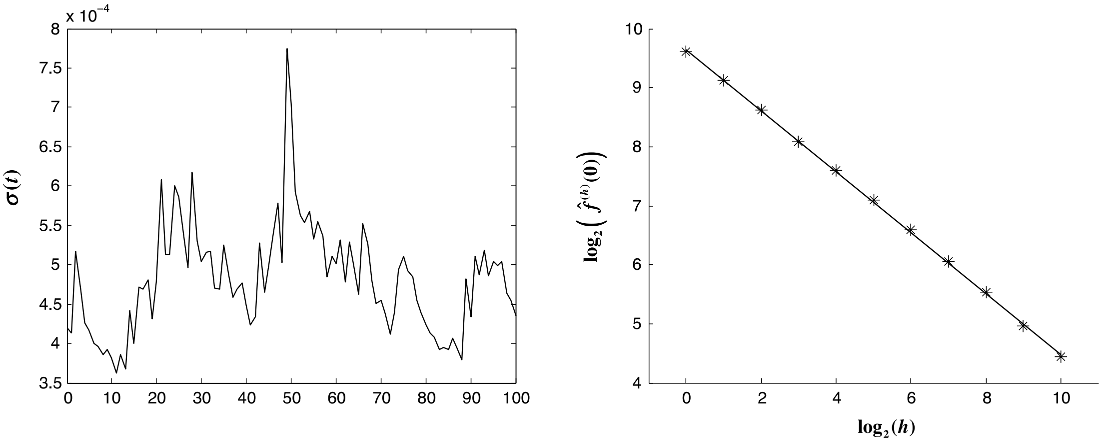


**Fig. 1.** Scaling of GARCH-simulated returns. One-minute returns on IBM for all of 2005 were used to fit a GARCH model, with autoregressive and moving average terms each of order 10 ($p = q = 10$). The model was used to generate $\sigma(t)$, for ten years of one-minute volatilities ($t = 1, 2, \ldots, 932, 400$). Left-hand panel shows a typical window with 101 consecutive values of $\sigma$. Volatilities were used to generate ten years of simulated one-minute returns, using (1.2) and a standard Brownian motion for $\omega$. 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 Ref. [4], 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.

a range of securities and time periods (cf. Refs. [3–16], 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 Section 2), and are therefore at odds with the empirical scaling of returns. We will then show (Section 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 $\sigma$ 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 [2] 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" [17]), 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 $\alpha$-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 $\sigma$ (which is a standard deviation only in the case $\alpha = 2$), are almost never constant ($\sigma = \sigma(t)$, "stochastic volatility" cf. Refs. [18,19]). And in fact the evidence is for very rapid fluctuations in $\sigma(t)$ (e.g. the left-hand panel in Fig. 1 is typical). A parsimonious extension of (1.1), whether or not $\alpha = 2$, is through the stochastic integral

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

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 $\sigma$.

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. Refs. [20–22]. 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