Comment on "Tests of scaling and universality of the distributions of trade size and share volume: Evidence from three distinct markets"

Éva Rácz, Zoltán Eisler, and János Kertész
Institute of Physics, Budapest University of Technology and Economics, Budafoki út 8, Budapest H-1111, Hungary
(Received 26 March 2008; published 24 June 2009)

In a recent publication, Plerou and Stanley [Phys. Rev. E 76, 046109 (2007)] use the Meerschaert-Scheffler estimator to verify the "inverse half-cubic law" of trade size distributions. We show that this procedure systematically underestimates these tail exponents.

DOI: 10.1103/PhysRevE.79.068101 PACS number(s): 89.75.Da, 89.65.−s, 05.45.Tp

In a pioneering study Gopikrishnan *et al.* [1] extended the statistical analysis of financial data on the distribution of the volume q, the number of shares exchanged in a single transaction. They concluded that this decays asymptotically as a power law $P(q>x)\propto x^{-\zeta_q}$ with $\zeta_q\approx 1.5$, a value in the Levy stable regime. A later study [2], however, found significantly higher values for data aggregated over 15 min, typically $\zeta_Q\approx 2.3$, outside the Levy regime.

In a recent paper [3], Plerou and Stanley carried out their analysis and found $\zeta_q\approx 1.49-1.65$ for three different stock markets. Reference [3] argued that the discrepancy between the previous works [1,2] in determining the tail exponents could be traced back to the use of the Hill estimator (HE) [4], which contains a cutoff parameter that is usually determined in a qualitative fashion. In contrast, [3] uses the Meerschaert-Scheffler estimator (MSE) [5], which is cutoff independent and defined as

$$\zeta_q^{\mathrm{MS}} = \frac{2(\gamma+\ln N)}{\gamma+\ln_+\sum_{i=1}^{N}(q_i-\langle q\rangle)^2}. \qquad (1)$$

q_i are independent observations of q and $\langle q\rangle$ is its mean; γ=0.5772 is Euler's constant and $\ln_+(x)=\max(0,\ln x)$.

The use of the MSE raises several problems in this specific case. First of all, the MSE is an estimator of the tail exponent only if the exponent is less than 2; otherwise (i.e., if the variance is finite), it converges to 2.0. In addition, the rate of convergence is quite slow compared to HE. These two factors result in the MSE being a questionable tool to distinguish between the cases $\zeta>2$ and $\zeta<2$. Furthermore, this estimator is not invariant to changes of the scale: for any fixed $A>1$, $\zeta_{q/A}^{\mathrm{MS}}>\zeta_q^{\mathrm{MS}}$. In [3] the data are rescaled with the mean absolute deviation.

In order to demonstrate that the MSE-based procedure yields misleading results, we used it on computer-generated Pareto distributed data with different tail exponents. Sample sizes were chosen to be comparable to that of financial data concerning liquid stocks on the New York Stock Exchange in the period studied in [3] (1994–1995), where $\langle\zeta_q\rangle=1.65\pm 0.01$ was obtained for the top 116 stocks. Our results are shown in Fig. 1. The curves indicate an "average response curve," namely, $\langle\zeta^{\mathrm{MS}}\rangle$ over 100 Pareto distributed data samples with an exponent ζ^{true}. It is worth noting that the equation $\langle\zeta^{\mathrm{MS}}\rangle(\zeta^{\mathrm{true}})=\zeta^{\mathrm{true}}$ is solved by $\zeta^{\mathrm{true}}\approx 1.5$ and with $d/d\zeta^{\mathrm{true}}\langle\zeta^{\mathrm{MS}}\rangle(1.5)<1$. Consequently, the estimator rounds any ζ^{true} value toward 1.5.

Reference [3] argued that it is the threshold dependence of the Hill estimator which causes problems in the case of aggregated data, i.e., Q. Our findings suggest that this is an important issue even if dealing with unaggregated data. Using a cutoff optimized extension to the Hill estimator introduced in [6], we obtained $\langle\zeta_q\rangle=1.98$ with a standard deviation of 0.24 for the top 116 stocks; these values are indicated in Fig. 1.

Support by OTKA T049238 is acknowledged.

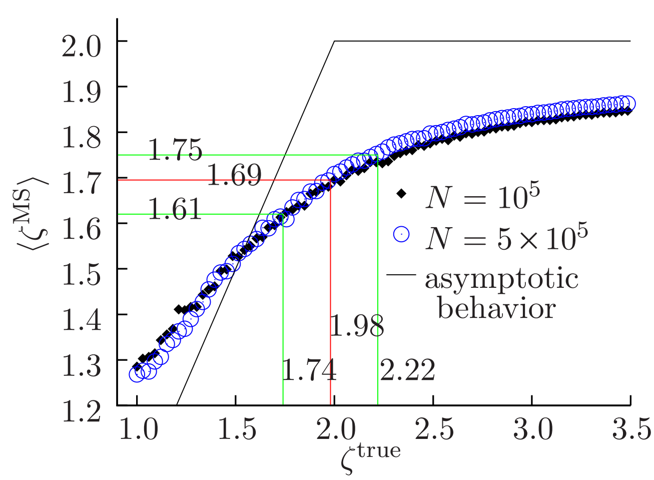

FIG. 1. (Color online) MSE estimates of the tail exponent of generated Pareto distributed data. Each point corresponds to the average of 100 runs. The sample sizes are indicated in the legend. The solid black line shows the asymptotic ($N\to\infty$) behavior of the MSE. The empirical value of $\langle\zeta_q\rangle=1.98$ with the errors as calculated from the self-consistent Hill estimator and the corresponding MSE values are indicated (average: gray lines, error bounds: light gray lines, online: red and green lines).

[1] P. Gopikrishnan, V. Plerou, X. Gabaix, and H. E. Stanley, Phys. Rev. E **62**, R4493 (2000).
[2] Z. Eisler and J. Kertész, Eur. Phys. J. B **51**, 145 (2006).
[3] V. Plerou and H. E. Stanley, Phys. Rev. E **76**, 046109 (2007).
[4] B. M. Hill, Ann. Stat. **3**, 1163 (1975).
[5] M. M. Meerschaert and H. Scheffler, J. Stat. Plan. Infer. **71**, 19 (1998).
[6] A. Clauset, C. R. Shalizi, and M. E. J. Newman, e-print arXiv:0706.1062, SIAM Review (to be published).