Eur. Phys. J. B **50**, 147–151 (2006)
DOI: 10.1140/epjb/e2006-00120-3

THE EUROPEAN PHYSICAL JOURNAL B

Small scale behavior of financial data

A.P. Nawroth[a] and J. Peinke

Institut für Physik, Carl-von-Ossietzky Universität Oldenburg, 26111 Oldenburg, Germany

Received 13 October 2005 / Received in final form 30 November 2005
Published online 12 April 2006 –

Abstract. A new approach is presented to describe the change in the statistics of the log return distribution of financial data as a function of the timescale. To this purpose a measure is introduced, which quantifies the distance of a considered distribution to a reference distribution. The existence of a small timescale regime is demonstrated, which exhibits different properties compared to the normal timescale regime for timescales larger than one minute. This regime seems to be universal for individual stocks. It is shown that the existence of this small timescale regime is not dependent on the special choice of the distance measure or the reference distribution. These findings have important implications for risk analysis, in particular for the probability of extreme events.

PACS. 89.65.Gh Economics; econophysics, financial markets, business and management

1 Introduction

The origin of mathematical finance dates back to Bachelier's famous thesis *Théorie De La Spéculation* (see [1]). As a central point in this work a normal distribution was assumed for financial returns. This assumption was for several reasons later changed by other authors to a normal distribution for the log return r [2]. The log return r is defined in the following way:

$$r(\tau) := \log(P(t+\tau)) - \log(P(t)) \tag{1}$$

where $P(t)$ denotes the price of the investment at time t. For alternative distributions to the log return distribution we refer to [3–8]. Other attempts [9] focused on the mechanism that may produce such distributions. There remains a general problem to determine the correct family of distributions, based on an appropriate underlying stochastic process, incorporating the dependence of the shape of the distribution on the timescale.

In the following we focus on the distribution (or the so called probability density function — pdf), which is in general dependent on the value of the log return itself as well as on the considered timescale. The question of the dependence of the shape of the distribution on the time scale was already posed in [2]. Considering changes of the form of distributions requires to distinguish between changes due to the mean value, due to the standard deviation, see e.g. [10], and due to the shape, see e.g. [11,12]. A discussion of the importance of risk measures like VaR and their connection to the underlying distribution can be found in [13,14].

[a] e-mail: andreas.nawroth@uni-oldenburg.de

When considering individual stocks, for very large time scales the normalized distribution is quite similar to a Gaussian distribution. For small timescales a Non-Gaussian fat-tailed distribution is obtained. An interesting question now arises. Is this transition from a fat-tailed distribution towards a Gaussian a smooth and uniform process? A general non-parametric method, utilizing a Fokker-Planck equation in timescale, has been proposed, which provides a general description of how the shape of the distribution evolves with changing timescale [15]. Although this approach is very general, it is based on assumptions that are partially no longer fulfilled for very small time scales (typically smaller than several minutes). Therefore here a specific non-parametric approach is presented, which provides insight into timescales covering seconds and minutes.

2 Data

In this study tick-by-tick data sets are used, in order to cover timescales as small as possible. The financial data sets were provided by the Karlsruher Kapitalmarkt Datenbank (KKMDB) [16]. The data sets contain all transactions on IBIS and XETRA in the corresponding period. The data sets used in this study span from the beginning of 1993 till the end of 2003 and contain $3{-}4 \times 10^6$ data points. Only stocks with a continuous history of trading in this period are considered. Results are presented for the three stocks with the largest number of trades in this period. These three stocks are Bayer, Volkswagen (VW) and Allianz. In order to investigate changes of the shape of the distribution, we analyze in general normalized distributions and therefore look at

the normalized return variable R

$$R = \frac{r - \overline{r}}{\sqrt{\overline{r^2} - \overline{r}^2}}, \tag{2}$$

where the average is taken over the whole data set. In order to compare the findings for stocks to other systems, the same analysis is performed for a turbulence data set. The data set was obtained by measuring the local longitudinal and transversal velocity component of a fluid in the turbulent wake behind a cylinder with a Taylor-based Reynolds number of 180 and contains 31×10^6 data points. For more details see [17].

3 Method

A non-parametric approach to the detection of a change in shape of a distribution is a direct measurement of the distance between two distributions. $p_{ref}(R)$ denotes the distribution for a reference timescale and $p(\tau, R)$ the distribution for another timescale. Firstly this allows verification of the frequently proposed assumption of a constant shape with respect to the timescale. Secondly if the shape is not constant this provides a quantitative measure of the size of the change in the shape of the distribution. Therefore a measure is needed to quantify the distance between two distributions. Here, the Kullback-Leiber-Entropy is used, which is defined as [18]

$$d_K(p(\tau,R), p_{ref}(R)) := \int_{-\infty}^{+\infty} dR\, p(\tau,R) \cdot \ln\left(\frac{p(\tau,R)}{p_{ref}(R)}\right). \tag{3}$$

In order to demonstrate the independence of our results on the particular choice of the measure we also use the weighted mean square error in logarithmic space

$$d_M(p(\tau,R), p_{ref}(R)) := \tag{4}$$

$$\frac{\int_{-\infty}^{+\infty} dR\, (p(\tau,R) + p_{ref}(R))(\ln p(\tau,R) - \ln p_{ref}(R))^2}{\int_{-\infty}^{+\infty} dR\, (p(\tau,R) + p_{ref}(R))(\ln^2 p(\tau,R) + \ln^2 p_{ref}(R))}.$$

Furthermore the chi-square distance is used as a third measure

$$d_C(p(\tau,R), p_{ref}(R)) := \frac{\int_{-\infty}^{+\infty} dR\, (p(\tau,R) - p_{ref}(R))^2}{\int_{-\infty}^{+\infty} dR\, p_{ref}(R)}. \tag{5}$$

Using these distance measures it is possible to determine the distance of a log return distribution calculated for a certain timescale from a reference distribution.

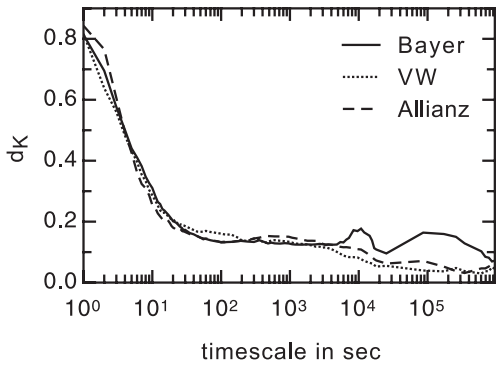

Fig. 1. Kullback-Leiber distance to the Gaussian distribution for three stocks.

4 Evidence of a new universal small timescale regime

For very large timescales the distribution is quite close to a Gaussian, therefore the Gaussian distribution is taken as reference distribution. In Figure 1 the Kullback-Leiber distance to the Gaussian distribution for three individual stocks is shown. It is evident, that the behavior changes considerably for timescales smaller than 100 s. For such small scales the pdfs of financial data are considerably different from the Gaussian distribution.

In a second step the distribution of the smallest scale of the considered asset is chosen as a reference distribution. In Figure 2 the distance d_K to the smallest timescale for the three stocks is shown, together with the one sigma error (dotted lines). The error estimate was calculated by means of sub-samples of the data set to estimate the distribution of the distance measure. Again a transition behavior is seen, indicating a change in the stochastic behavior in the region 10–100 s. In all three cases the first region may be characterized by a linear increase of the distance measure d_K. The linear fit for this first region is drawn as a solid line in Figure 2 (Note the use of semilog plots).

In order to verify if the region displaying linear behavior is dependent on the chosen reference timescale, the analysis has been redone for different reference timescales. As an illustration, the results for Volkswagen are shown in Figure 3a. For all these reference distributions the extent of the linear region (more precisely the upper bound) does not change. This and similar results for the other assets indicate, that the linear region is independent of the timescale that was chosen for the reference distribution. Next we discuss the influence of different distance measures (Eqs. (3)–(5)). As an example the distance to the smallest timescale for VW is shown in Figure 3b. Similar results were obtained for other stocks. For comparison all distance measures were rescaled to the interval $[0, 1]$ in Figure 3b. For all three distance measures a division of the timescale in two parts characterized by the different functional behavior in these parts is evident.

A possible reason for the existence of different domains may be based on a specific relationship between consecutive increments on different timescales. One way to analyze this is to destroy all possible causal relationships of

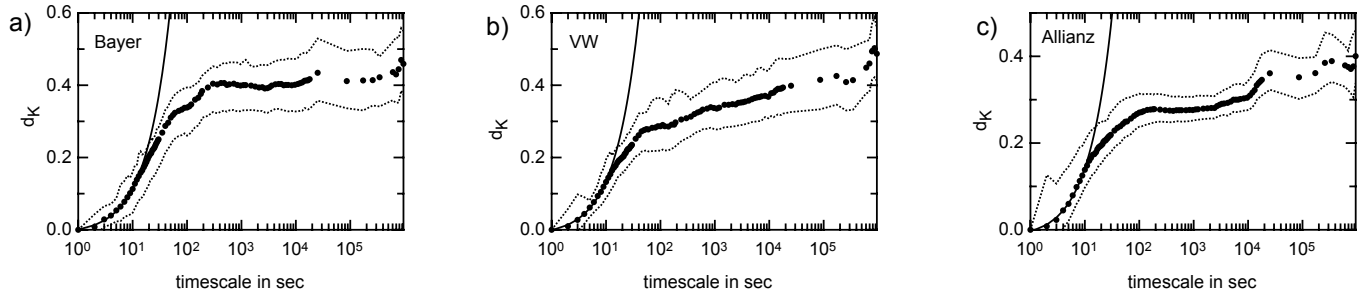

Fig. 2. The distance measure d_K for a reference distribution $p_{ref}(R) := p(\tau = 1\text{ s}, R)$ for the individual stocks. The dots represent the estimated value, the dotted lines the one sigma error bound and the solid line the linear fit for the first region.

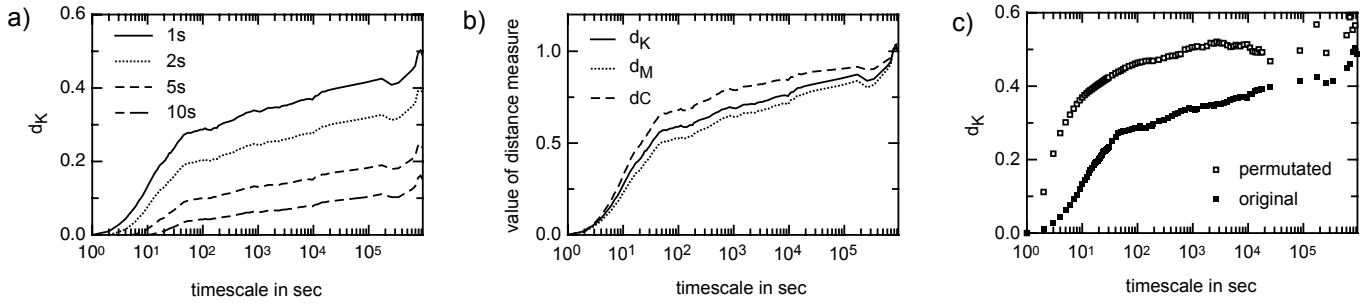

Fig. 3. (a) The distance measure d_K for Volkswagen. (b) Three different distance measures with $p_{ref}(R) = p(\tau = 1\text{ s}, R)$ for Volkswagen. (c) Comparison of the distance measure d_K with $p_{ref}(R) = p(\tau = 1\text{ s}, R)$ for the original and the permuted Volkswagen data set.

consecutive increments. This can be done by permuting all increments on a certain timescale (here the timescale of the reference distribution) and thereby creating a new time series with the same $p_{ref}(R)$. This new time series exhibits for small timescales a logarithmic increase in the distance measure d_K, see Figure 3c. Further there is no longer a division into two distinct timescale intervals with different functional behavior of the distance measure d_K. It is therefore evident, that the small timescale regime is due to functional relationships between consecutive increments. In order to investigate if the dependencies are linear, the autocorrelation function (ACF) of the non-uniformly sampled time series is calculated, cf. [19]. The estimator for the autocorrelation is defined in the following way

$$\hat{\rho}(\tau, \Delta\tau) := \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \left[r(\tau, t_i) - \bar{r}(\tau) \right] \right. \tag{6}$$

$$\times \left. \left[r(\tau, t_j) - \bar{r}(\tau) \right] b(t_j - t_i) \right]$$

$$\times \left[\left[\sum_{i=1}^{N} \sum_{j=1}^{N} \left[r^2(\tau, t_i) - \bar{r}^2(\tau) \right] b(t_j - t_i) \right] \right.$$

$$\left. \times \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \left[r^2(\tau, t_j) - \bar{r}^2(\tau) \right] b(t_j - t_i) \right] \right]^{-\frac{1}{2}}$$

$$b(t_j - t_i) = \begin{cases} 1 & \text{for} \quad |(t_j - t_i) - \Delta\tau| < \delta\Delta\tau \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $r(\tau, t_i)$ is the log return on the timescale τ at the time t_i and δ a small number. The results for the ACF, computed on a timescale of four seconds, are shown in Figures 4a, 4b. The computation of the ACF for smaller timescales becomes increasingly difficult due to the very small number of available log returns. In agreement with the literature [7,20], there is a negative autocorrelation for the smallest lag, while for larger lags the ACF yields values very close to zero. The ACF of the magnitude of the log returns is considered in Figure 4b. Here there is a strong positive autocorrelation for the smallest lag, which slowly decays for larger lags. However, for both ACFs and all the considered stocks there is no indication of a small timescale regime in the ACF. It therefore appears that the functional relationship between consecutive increments, which causes the small timescale regime, is of nonlinear nature.

5 Comparison with turbulence data

In [12,21] it has been shown, that finance and turbulence data display common properties. The analysis described above is therefore also performed with turbulence data in order to see if a small timescale regime is present in that case as well. In Figure 5 the distances d_K of the distribution of the velocity increments with respect to the Gaussian distribution (a) and with respect to a small scale reference distribution (b and c) for the turbulence data are shown. The qualitative behavior for larger timescales is similar to that observed for individual stocks, while for smaller timescales the behavior differs. It is important to note the difference in scale of the distance measure in Figures 1 and 5a.

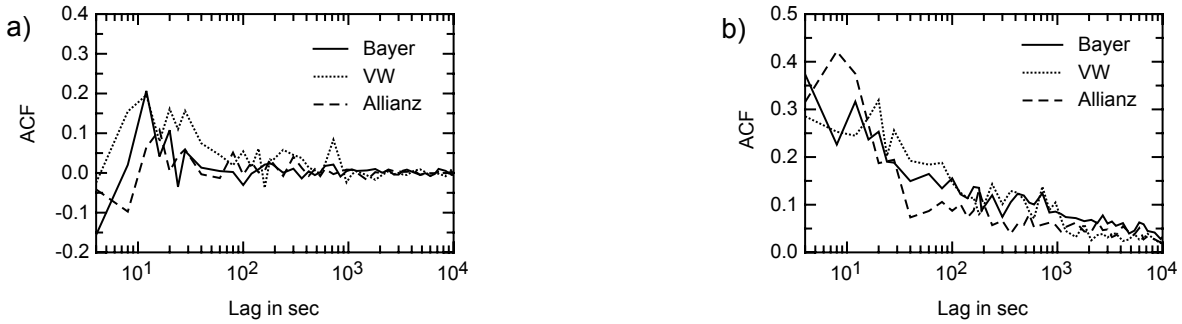

Fig. 4. (a) Autocorrelation function of the log returns for three individual stocks. (b) Autocorrelation function of the magnitude of the log returns for three individual stocks.

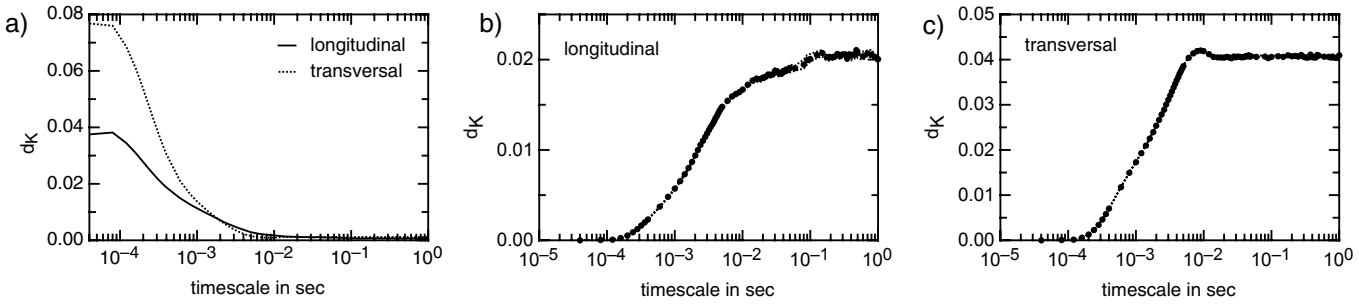

Fig. 5. Turbulence data: (a) the dependence of d_K to the Gaussian distribution on the timescale. (b), (c) The distance measure d_K for a reference distribution $p_{ref}(R) := p(\tau = 4 \times 10^{-5}\ \mathrm{s}, R)$. The dots represent the estimated value, the dotted lines the one sigma error bound.

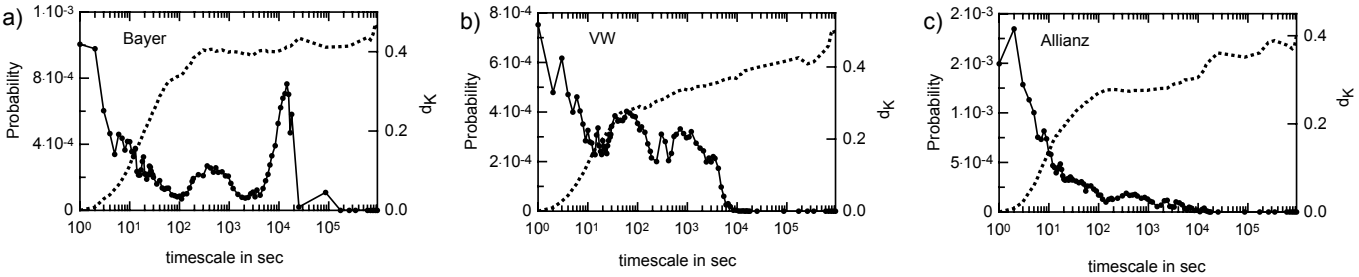

Fig. 6. Comparison between the probability mass beyond the 10th standard deviation (solid line) and d_K (dotted line).

6 Applications

How does the specific behavior of the small timescale regime translate into practical applications? The deviation from the Gaussian distribution is increasing much faster in the small timescale regime than in the normal timescale regime. A visual inspection shows, that the considered log return distributions deviate into the direction of fat-tailed distributions. Therefore the probability mass in the tails of the distribution should increase faster by entering the small timescale regime. In order to analyze this, the probability mass in the tails of the distribution, i.e. the probability mass beyond the 10th standard deviation, where left and right tail are considered together, is calculated and the results are compared to the distance measure d_K. The reference timescale is one second. The results are shown in Figure 6. In all three cases it is evident, that the change of the distance measure corresponds to a change of probability mass in the tails of the distribution. In the small timescale regime the increase in the probability mass in the tails of the distribution is very pronounced. The estimates of probability mass, for timescales larger than 10^3 s, are rather noisy, due to the effect that events are quite rare in this region.

7 Conclusions

Summarizing, it has been demonstrated that the properties of the log return distribution of stocks do not change uniformly if one goes to smaller timescales. Instead, for small timescales a distinct regime is entered with different properties. In this small timescale regime, the shape of the distribution changes much faster than one would expect by extrapolating the behavior of the normal timescale

regime. This small timescale regime extends for individual stocks to our knowledge from timescales of around 1s to timescales of around 15 s. The small timescale regime can be characterized by a linear dependence of the Kullback-Leiber distance d_K on the timescale, if as a reference distribution a log return distribution on a very small timescale is chosen. In the normal timescale regime the dependence is much slower and can be assumed to be logarithmic or for very large timescales independent of the timescale. This result seems to be independent of the chosen reference distribution as long as it is a log return distribution on a sufficiently small timescale. If the Gaussian distribution is taken as a reference distribution, d_K is rising very fast with decreasing timescale in the small timescale regime, while it stays nearly constant in the normal timescale regime in accordance with [11]. This indicates a very fast deviation from a Gaussian-like shape in the small timescale regime. These results could be confirmed with different distance measures. Further it has been shown that this small timescale regime is a specific feature of the financial data investigated here. For turbulence data no such small timescale regime is observed, although financial and turbulence data sets exhibit similarities in the normal timescale regime. For very small timescales the shape of the distribution, in contrast to the findings for the financial data sets, changes slower than one would expect by extrapolating the behavior of the normal timescale regime. Two prominent candidates for this effect are the dissipation on small scales for turbulence [17] and the noise added by the measurement system. Furthermore the particular small timescale regime for individual stocks cannot be reproduced by trivial randomized data. As an application of this new approach it has been demonstrated that on entering the small timescale regime, a large increase in the probability mass in the tails of the distribution occurs, which could lead to very different risk characteristics in comparison to that of larger timescales.

References

1. J. Courtault, Y. Kabanov, B. Bru, P. Crepél, I. Lebon, A.L. Marchand, Mathematical Finance **10**, 341 (2000)
2. M. Osborne, Operations Research **7**, 145 (1959)
3. E. Fama, J. Business **38**, 34 (1965)
4. B. Mandelbrot, J. Business **36**, 394 (1963)
5. P.K. Clark, Econometrica **41**, 135 (1973)
6. R.N. Mantegna, H.E. Stanley, Nature **376**, 46 (1995)
7. J.P. Bouchaud, M. Potters, *Theory of Financial Risks* (Cambridge University Press, 2001)
8. B. Castaing, Y. Gagne, E.J. Hopfinger, Physica D **46**, 177 (1990)
9. T. Lux, M. Marchesi, Nature **397**, 498 (1999)
10. R. Cont, M. Potters, J.P. Bouchaud, Proc. CNRS Workshop on Scale Invariance (1997), e-print `arXiv:cond-mat/9607120`
11. V. Plerou, P. Gopikrishnan, L.A.N. Amaral, M. Meyer, H.E. Stanley, Phys. Rev. E **60**, 6519 (1999)
12. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Nature **381**, 767 (1996)
13. J. Hull, A. White, J. Derivatives **5**, 9 (1998)
14. A. Farias, J. Ornelas, J. Barbachan, Stochastic Finance 2004 (2004)
15. C. Renner, J. Peinke, R. Friedrich, Physica A **298**, 499 (2001)
16. T. Lüdecke, Discussion Paper No. 190 University of Karlsruhe (1998)
17. M. Siefert, J. Peinke, Phys. Rev. E **70**, 015302 (2004)
18. S. Kullback, *Information Theory And Statistics* (Dover Publications, 1968)
19. L.H. Benedict, H. Nobach, C. Tropea, Meas. Sci. Technol. **11**, 1089 (2000)
20. M.M. Dacorogna, R. Gençay, U. Müller, R.B. Olsen, O.V. Pictet, *An Introduction to High-Frequency Finance* (Academic Press, 2001)
21. R. Mantegna, H. Stanley, Physica A **239**, 255 (1997)