Eur. Phys. J. B **17**, 333–339 (2000)

THE EUROPEAN PHYSICAL JOURNAL B

EDP Sciences

Gambling on the Budapest stock exchange

I.M. Jánosi[a]

Department of Physics of Complex Systems, Eötvös University, PO Box 32, 1518 Budapest, Hungary
and
John von Neumann Institut für Computing (NIC), Forschungszentrum Jülich, 52425 Jülich, Germany

Received 5 May 2000

Abstract. The statistical properties of the total yield are analyzed for an assembly of gamblers in an erratic period on the Budapest stock exchange. Random trading results in a log-normal limit distribution of a surprisingly large width, while the simplest profit realizing strategy narrows down the peak around a positive average value. The effect of transaction costs, the statistics of extremes, and patterns of successful trading are also investigated. In spite of the very simple approach, we present strong indications that large trading activity (*e.g.* day trading) is a rather risky way of capital investment. A comparison with the yield distribution of 32 public investment funds in the given period does not reflect the presence of a sophisticated investment strategy in the background.

PACS. 02.50.-r Probability theory, stochastic processes, and statistics – 89.90.+n Other topics of general interest to physicists

1 Introduction

Electronic trading on financial markets enormously increased the efficiency of "money industry", with huge sums being earned and lost on computer networks on all around the world. Prices and indices are recorded several times a minute in liquid financial markets. The large amount of available data and the complexity of market structures has attracted considerable interest of economists, mathematicians and also of physicists in recent years [1,2]. Research has focused on detailed statistical analysis of price fluctuations [3–8], modeling markets as complex interactive systems [9–14] and finding analogies between economic and other phenomena such as turbulence [15–17] or biological adaptation [18].

Considering the huge amount of data to be analyzed, it is a common dream to produce "automated strategies", *i.e.* intelligent computer programs which are able to evaluate quickly on-line information and to "propose" trading activity. The application of such softwares is desirable only if the risk is lower and, in optimal case, the profit is higher than for the "pure human" decision making process. While the literature on risk minimization is huge [19], contributions on possible automatized trading strategies are scarce. Recent counterexamples are the works of Andersen, Gluzman and Sornette [20], and Molgedey and Ebeling [21], where the identification of important patterns seems to be solved with significant predictive power. Nevertheless any attempt to find a trading algorithm needs a measure for evaluating the success rate. Here we propose such a measure by analyzing the yield for a large assembly of gamblers, *i.e.* "investors" without any or a simple profit realizing strategy.

The analysis is based on a real time series of the index of the Budapest Stock Exchange (BSE), a typical example for emerging markets, which was reestablished in June 1990. Securities trading is entirely electronic. Since 1 of April 1997 the index (BUX) has been calculated and recorded continuously with a time resolution of 5 s during the main trading hours. In Figure 1 we show the evolution of the BUX index as a function of trading time in the period from 21 April 1997 to 17 December 1998. (The trading time is calculated by eliminating the intervals when the market is closed.) This period is chosen for the analysis, because the first index value in the time series coincides with the last one, therefore it can serve as a simple reference for the following analysis. Detailed statistical characterization of the BUX data is given in references [22,23].

2 Random trading

Let us assume that an investor buys 100 BUX packages, *i.e.* a portfolio consisting the BUX stocks with proper weights, at the beginning of the period (21 April 1997), when the index was 5803.90. If no transaction is performed, the value of the packages is 100×5803.96 at the end of the period under consideration (17 December 1998). This means actually a net loss because of inflation, but for the sake of simplicity we consider the result of such an investor as zero net yield. A gambler, *i.e.* a random trader

[a] e-mail: janosi@krumpli.elte.hu

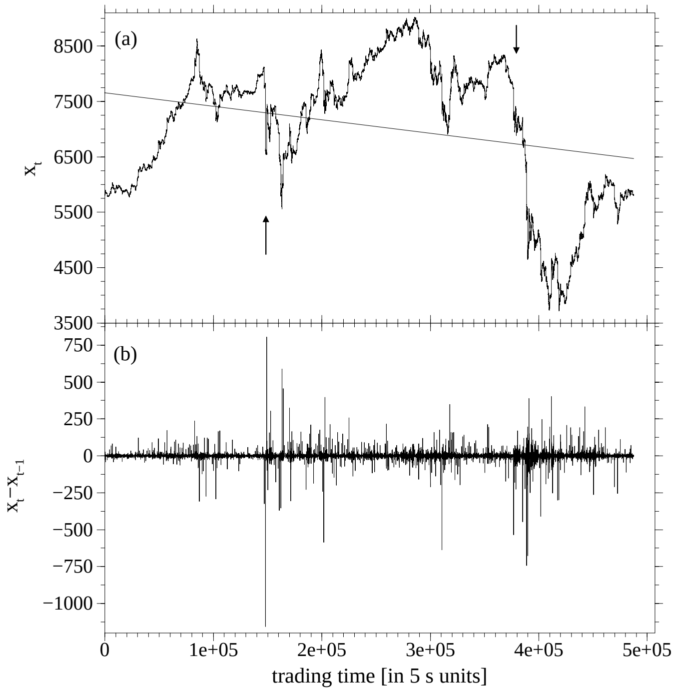

Fig. 1. BUX values (a) and changes of the index (b) as a function of trading time between 21.4.1997-17.12.1998 (487 643 data points). Arrows indicate two large financial crises: The first is the South-East Asian (November 1997) and the second is the Russian (August 1998) crisis. The solid line in (a) indicates an overall decreasing linear trend in the given period.

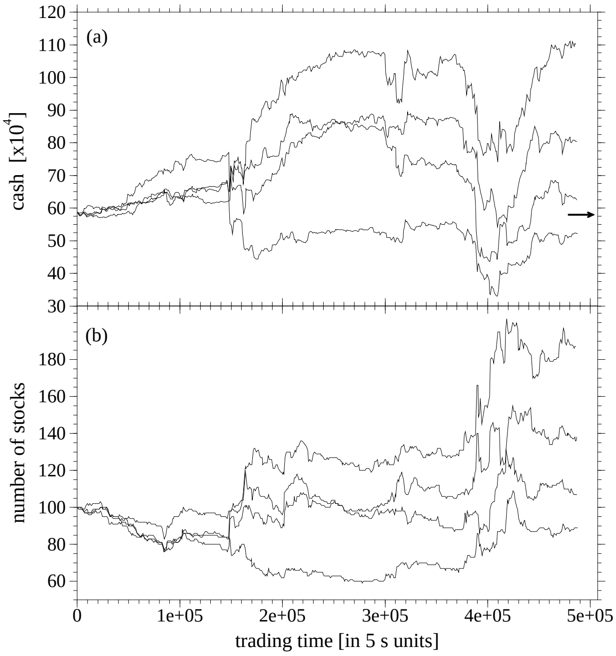

Fig. 2. (a) Instantaneous value of stocks and cash as a function of trading time for different random traders. At the beginning all had 100 BUX packages, the activity parameter $\Theta = 1\,000$ for each case. The arrow at the right side indicates the value of the package kept from the beginning without any trading. (b) Number of BUX packages sold as a function of trading time for the same random traders.

makes transactions in this period without any strategy and without any care of trends, other information, or even the instantaneous value of the index. The transactions are very simple:

1. The trader sells the packages at the instantaneous price after a random waiting interval of Δt_1, which has a uniform distribution between 0 and Θ measured in trading time (5 s steps).
2. After a second random interval Δt_2, during which nothing happens with the last amount of cash, the trader buys as many BUX packages as he can afford at the instantaneous price. The remainder is kept in cash.

Random "trajectories" can be generated by iterating the above steps until the end of the whole period, when the result can be evaluated by accounting the total amount of cash obtained by selling all the packages left. The only parameter of the model is Θ, which characterizes the trading activity, *i.e.* the average time between transactions. No transaction fees, interest rate or other ingredients are considered at the first step. We assume further that the individual (small volume) actions of such traders do not contribute to the evolution of the index itself, and the market is liquid enough to make possible the desired transactions at the instantaneous prices.

In Figure 2 we show a few examples of random trading "trajectories" as a function of trading time. Different random sequences can be realized simply by starting the runs from different random number seeds. Note that the final cash value at the end of the period may well exceed the instantaneous BUX value indicated by an arrow in Figure 2a. The activity parameter $\Theta = 1\,000$ means an average waiting time of 2 500 s, which is equivalent with 2-3 transactions/day (the opening of BSE was restricted to 100-110 minutes a day in the given period).

An interesting characteristics is the probability distribution $P(y)$ of the total yield y at the end of the gambling period, which is defined as the final value of cash and stocks divided by the value of the initial investment. We determined this function from 50 000 different realizations for several activity parameters Θ, see Figure 3. The results show that the smooth distribution for "active" traders of relatively low Θ values is strongly asymmetric, and can be well fitted by a log-normal distribution

$$P(y) = \frac{1}{\sqrt{2\pi} y \sigma_y} \exp\left[-\frac{1}{2}\left(\frac{\ln(y) - m}{\sigma_y}\right)^2\right], \qquad (1)$$

where y denotes the yield (final wealth divided by the initial one), m and σ_y are the parameters of the distribution. Up to the value $\Theta \approx 25\,000$, there is a very weak dependence on Θ, the best fits give the values $m = 0.006 \pm 0.001$ and $\sigma_x = 0.29 \pm 0.01$. The most probable value y_p, *i.e.* the location of the peak is well below one: $y_\mathrm{p} \approx 0.93 \pm 0.01$. The median y_med defined by

$$P(y < y_\mathrm{med}) = P(y > y_\mathrm{med}) = \frac{1}{2} \qquad (2)$$

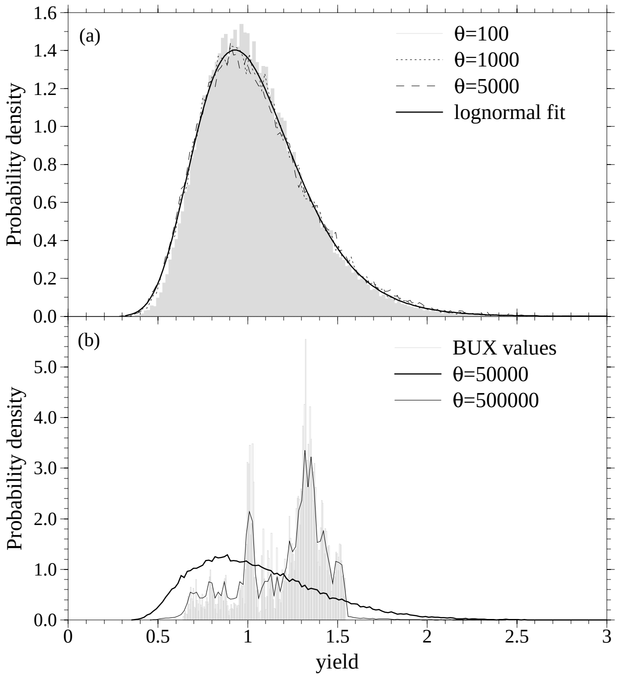

Fig. 3. (a) Normalized probability density distributions $P(y)$ of the total yield y for active traders of different Θ values. The histograms were obtained from 50 000 random trajectories, each. (b) The same as above, for less active traders of large Θ values. Gray histogram shows the BUX distribution normalized with the initial value.

is obtained by numerical integration: $y_{\rm med} \approx 1.00 \pm 0.008$. The average value for the yield y is

$$\langle y\rangle = \exp\left(m + \frac{\sigma_y^2}{2}\right) \approx 1.05 \pm 0.01. \qquad (3)$$

This average value completely agrees with the total wealth of all traders normalized by the total initial investment computed directly in the simulations.

These characteristics of the empirical distribution support the common opinion that gambling is a risky investment with 50% probability of win or loss, however in a lucky situation one can double or triple the initial capital. (Actually, in our samples 0.8-1.2% of the gamblers "earned" more than 100% of the initial investment.) According to the simple rules the total bankruptcy has a vanishing probability, however the introduction of transaction costs changes this situation drastically, see below.

If the activity parameter Θ increases, the smooth log-normal shape gradually changes to a strongly peaked, irregular distribution. If Θ is comparable to the length of the time series, only one or two transactions happen in the whole period, which corresponds to a point-sampling of the index curve. Thus the yield distribution for large Θ is closer and closer to the price distribution itself (expressed in units of the initial value), as it is apparent in Figure 3b.

The origin of the log-normal distribution is clear. The total yield after N transactions can be considered as a product of random variables

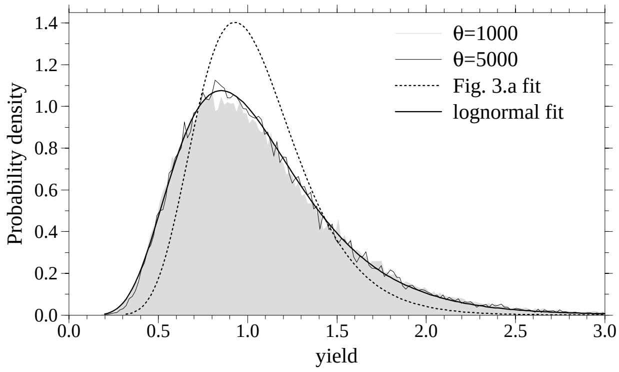

Fig. 4. The same as Figure 3a, but for the flipped time series.

$$y = \xi_1\xi_2 \ldots \xi_N, \qquad (4)$$

where $\xi_i = (x_i - x_{i-1})/x_i$, and the average time between the ith and the preceding $(i-1)$th transactions is $\Theta/2$. The logarithm of y is the finite sum

$$\ln(y) = \ln(\xi_1) + \ln(\xi_2) + \ldots + \ln(\xi_N). \qquad (5)$$

If N is large enough, ξ_i are independent, and each of them can be characterized by an identical probability distribution of finite first and second moments, the limit distribution for $\ln(y)$ converges to a Gaussian according to the central limit theorem. Several statistical analysis of stock prices concluded that price changes have very short time correlations [1,2,23]. Thus the total yield for a series of sufficiently separated transactions can be considered as a product of independent stochastic variables, therefore the probability distribution should converge to a log-normal distribution.

It is obvious that the statistics of price changes determine primarily the total yield. In reference [23] the cumulative distribution functions for the return are evaluated for different time lags, the difference between negative and positive changes is apparent, but not very informative. The distribution for fixed step returns can not be directly related to our "gambling test", since here the interval between the consecutive transactions is a random variable itself. In order to check the effect of the asymmetry in the price-change distribution, one can change the role of negative and positive jumps by flipping the original time series around the horizontal axis determined by the first and last values. (Note that in this case the trend shown in Figure 1a is transformed to be positive, however the price change distribution is simply flipped around the vertical axis.) The result is shown in Figure 4. The limit distribution is wider, the fitted parameters are $m = -0.004 \pm 0.002$ and $\sigma_y = 0.42 \pm 0.02$. This means an average yield of $\langle y\rangle \approx 1.08 \pm 0.013$, which is almost the same as in the original case. We can conclude from this result that the asymmetry in the price change distribution affects primarily the width of the limit distribution, but the 50% chance for win or loss remains the same. Note that the larger width increases the proportion of lucky gamblers of more than 100% profit to be roughly 3.5%.

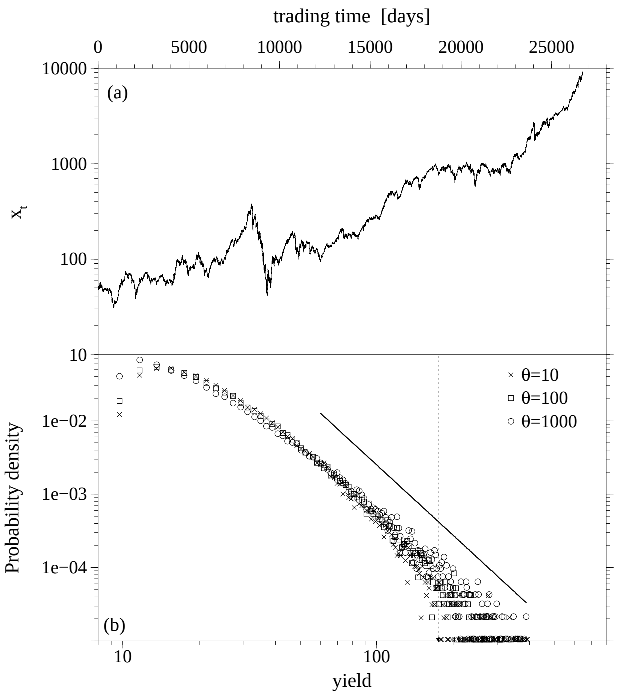

Fig. 5. (a) Daily values of the Dow Jones Industrial Average from 2 January 1901 to 24 April 1998 (the vertical scale is logarithmic). (b) Probability distribution of the total yield for random traders at three different activity parameters Θ (measured in units of days). The vertical dotted line corresponds to the yield of no trading at all, *i.e.* the final value of 100 packets. Thin solid line indicate a power-law tail with an exponent -3.2 ± 0.2.

The pure log-normal limit distribution is characteristic at weak overall trends only. As an example, we show the result for similar gambling tests for the famous Dow Jones daily series plotted in Figure 5a. The limit distribution does not depend on the activity parameter again, but the shape is not log-normal any more. It is remarkable that random trading is a very bad "strategy" at a strong increasing trend, a negligible minority earns more than the investors "sitting" on their stocks.

3 The effect of transaction costs

Up to this point we can conclude that active gambling might result in a positive yield. The limit distribution is insensitive to the trading activity parameter Θ in a wide range, which indicates the lack of correlations between price changes. It is obvious, however, that transaction costs affect strongly the total yield.

In Figure 6 we show the limit distributions for different activity parameters by taking into account 1% transaction cost, which is an average for small investors at the Hungarian trading agencies. The width of the parabolas on the double logarithmic plot is almost the same for each case, however their center shifts drastically to smaller values at a high activity. In Figure 7 the average yields are plotted as a function of average number n of transactions

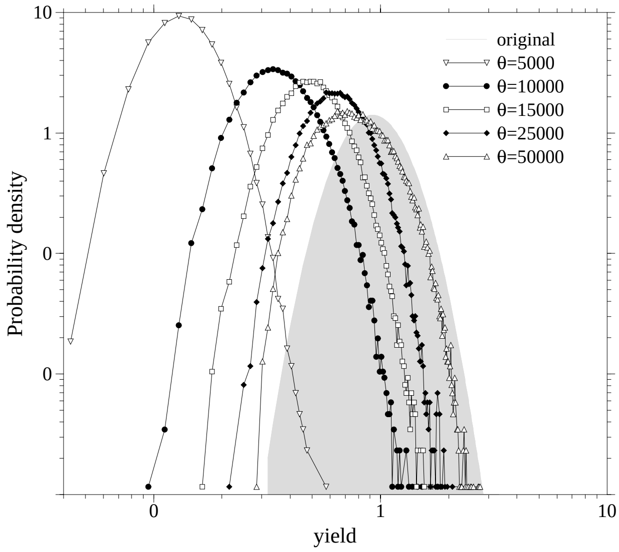

Fig. 6. The effect of 1% transaction cost on the limit distribution for different activity parameters Θ. (Note the double logarithmic scale.)

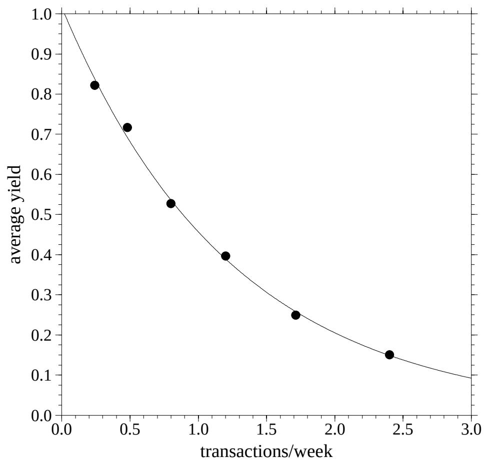

Fig. 7. The average yield as a function of average number of transactions per trading week. The solid line shows an exponential fit.

per trading week (the transformation from Θ is trivial). As expected, the average values decrease exponentially with the increasing trading activity, the empirical fit shown as solid line is $\langle y \rangle = 1.01 \exp(-0.8n)$.

It is clear from this curve that *e.g.* "day-trading" is a gamble for members of the stock market, who benefit from almost zero transaction costs. Nevertheless the limit distribution for $\Theta = 50\,000$, which represents roughly one transaction a month, is not very far from the cost-less limit distribution for random day-traders.

4 How to win

Let us show next the statistics of the luckiest gamblers. The question we pose: Is there any apparent pattern in the activity leading to large wins? As a representative example, we show the results for a mild activity $\Theta = 50\,000$.

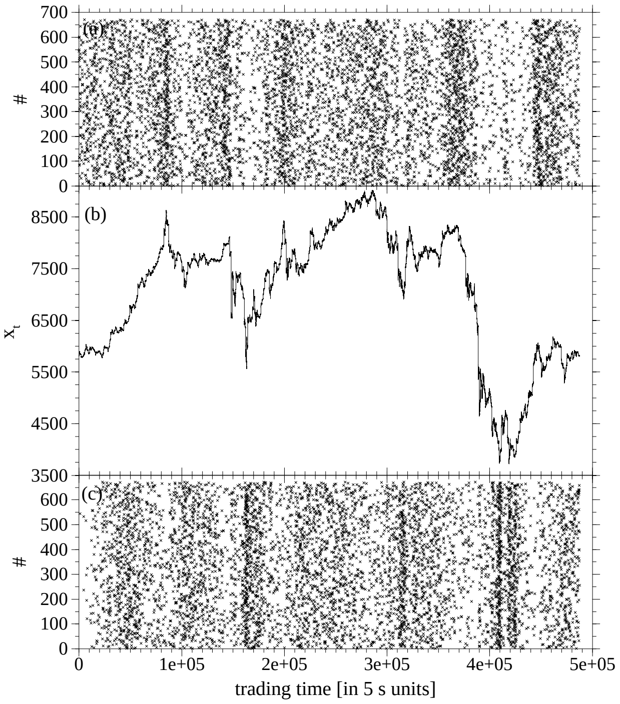

Fig. 8. The time of selling (a) and buying (c) for the 671 gamblers who won more than 100% of the initial capital ($\Theta = 50\,000$). The vertical axis show the serial number of the gamblers. (b) is the same as Figure 1a.

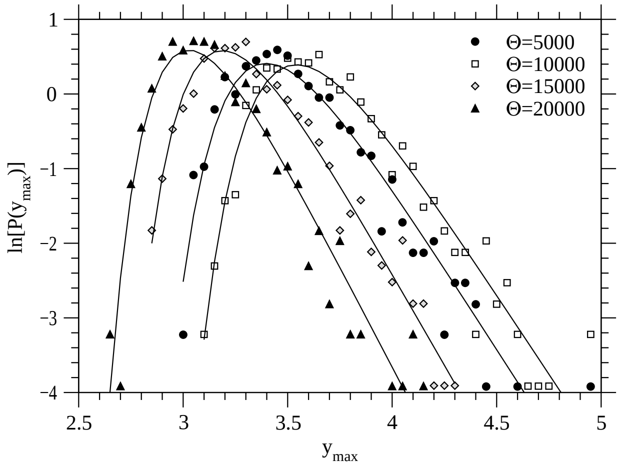

Fig. 9. Logarithm of normalized probability density distributions for the maximal yield $y_{\rm max}$ for 1 000 different samples of 50 000 random traders. Thin solid lines show the fits by equation (6) for different Θ values.

(No transaction costs considered.) The best 671 of 50 000 gamblers doubled the initial capital, they are identified simply by a serial number from 1 to 671. We plotted in Figure 8a the time instants when they sold the packages, in Figure 8c when they bought stocks. (Fig. 8b shows the time series of the BUX.) The first observation is that some stripes are apparent close to local maxima (Fig. 8a) and local minima (Fig. 8c) of the time series. This is in agreement with the golden rule: "Sell when it is expensive and buy when it is cheap." Nevertheless the patterns are not strongly located, there is a pronounced trading activity all along the time axis. One interpretation for the success of distributed trading activities might be that there are a few crucial moves not to miss, and the other transactions do not play an important role. In order to locate the "golden" moves, we generated a much larger sample, namely 5×10^7 trajectories for $\Theta = 50\,000$ (which means roughly 8-12 transactions in the whole period). The criterium of superiority was sharpened too, activities of the 3 gamblers getting 400% of the initial investment were evaluated only. No unique pattern was found.

As for the distribution of the extreme values, we evaluated the maximal and minimal total yields for 10^3 different realizations of finite assemblies of 5×10^4 gamblers at different activity parameters Θ. The result for maximal values is shown in Figure 9. All the curves can be satisfactorily fitted by the so called Gumbel-I distribution originally identified by Fisher and Tippett [24]:

$$P(y_{\rm max}) = K\left({\rm e}^{z-{\rm e}^z}\right), \qquad (6)$$

where $z = -b(y_{\rm max} - c)$. The most probable value c for the given sample size apparently depends on the activity parameter Θ, actually the change $3.0 < c(\Theta) < 3.7$ (not shown here) is nonmonotous with a local maximum at around $\Theta = 10^4$ and a local minimum at $\Theta \approx 3 \times 10^4$. No similar dependence for the minimal values $y_{\rm min}$ was found.

5 Minimal strategy

It is obvious that moves of investors are not independent from the time-evolution of price history and many other factors contributing to the expectations. The simplest possible strategy is probably a finite threshold profit-realization with the following rules:

1. The investor follows the time evolution of the index, and sells the stocks when the profit exceeds a given threshold k%. (*E.g.* $k = 5$% means a total yield of $y = 1.05$ after the first transaction.)
2. Having a cash-position, the investor buys stocks whenever their possible number exceeds the last value in the previous stock-position.

These rules are fully deterministic, thus a statistical evaluation of such a strategy is not feasible. Furthermore a direct application of them is not "realistic" at all, because nobody can follow the changes and perform transactions with the given time-resolution (5 seconds!). Therefore we keep randomizing with the activity parameter Θ for both steps above, in order to simulate limited information access, random decision making, and similar probabilistic factors. Note that the number of ways for defining strategies is unlimited, but the introduction of further thresholds, algorithmic treatment of technical patterns, etc., raises the set of model parameters.

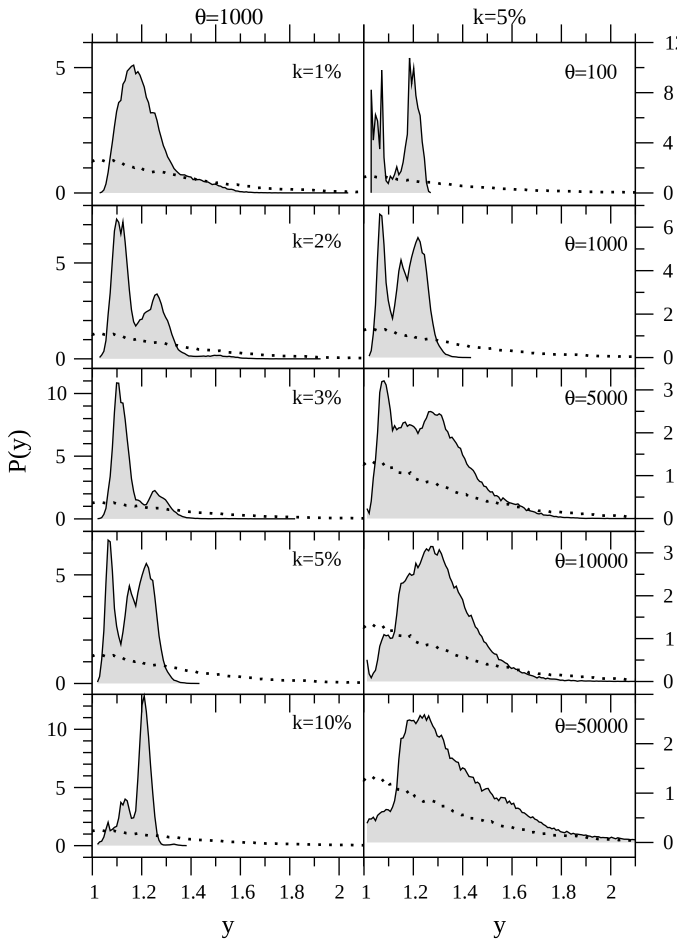

Fig. 10. Normalized limit distribution of the total yield y for the minimal strategy gamble described in the text for 1.5×10^5 samples each. Left column: The activity parameter is fixed ($\Theta = 1\,000$), five different selling thresholds are shown. Right column: The threshold is fixed ($k = 5\%$), Θ is changing. Note the different vertical scales. Dotted line indicates the log-normal distribution for fully random trading.

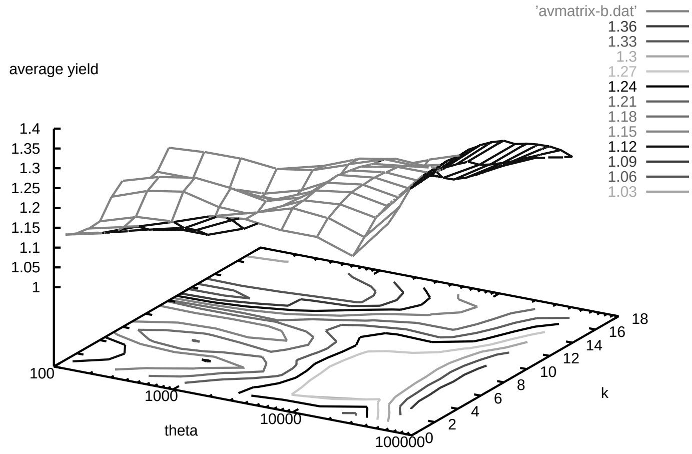

Fig. 11. Average yield as a function of model parameters Θ and k in samples of 1.5×10^5 gamblers. Note the logarithmic Θ-axis.

This simple two-parameter gamble increases the complexity of the limit distributions for total yield y, as it is illustrated in Figure 10. Some trends are clearly visible: The higher the threshold k at a fixed activity parameter Θ the narrower the distribution (Fig. 10 left column), while a decreasing trading activity (larger Θ) at a fixed threshold k increases the width of the peaks (Fig. 10 right column). It is also clear that such a profit realization strategy can not lead to a total yield smaller than 1 at the particular choice of the trading interval (the final index value is identical with the initial one).

The shape of the probability density distributions changes also strongly at different parameter pairs, therefore its characterization with a couple of numbers is difficult. Nevertheless we attempted to evaluate the model-parameter space by calculating the average value, width, and extremes as a function of Θ and k. A representative result for the average total yield $\langle y \rangle$ is shown in Figure 11. The strongly irregular surface partly reflects the change of shape of the underlying distribution (see Fig. 10). This surface is far from being a plane, nevertheless there is a global increasing (logarithmic) trend at decreasing activities for a rather wide threshold range $1\% < k < 10\%$. Similarly to the fully random case, very high trading activity results in definitely smaller average yields than a moderate behavior.

As for the transaction cost, the effect is similar to the fully random case discussed above with only one exception. Namely, Rule 2 means in this case that an agent can buy stocks if the amount of cash covers also the transaction cost. Thus the lower limit of total yield at the end is 1. At small Θ values (high activity) the distributions shown in Figure 10 sharpen further around average values 1.07-1.15, low activity ($\Theta > 20\,000$) has a small effect on them.

6 Discussion

Unfortunately we do not have access to real data on the yields of individual investors, thus a direct comparison with our simulations is not possible. However, we can show the data for 32 public investment funds being active in the same period (Fig. 12), since their yield is published regularly. (Note that these data are not corrected with inflation, similarly to our treatment.) Although the statistics is very poor, we found that the histogram is most similar to the distribution of index values (expressed in the initial price) itself. As Figure 3b shows, this yield distribution is characteristic for fully random trading at quite low activities (large Θ values).

We do not want to overemphasize the observed similarity, especially because these funds invest not only in stocks, in fact some of them trade only bonds. However, the evolution of Hungarian bond prices undoubtedly interacts with the BUX index. Further, the total capital treated by the Hungarian public funds is very low

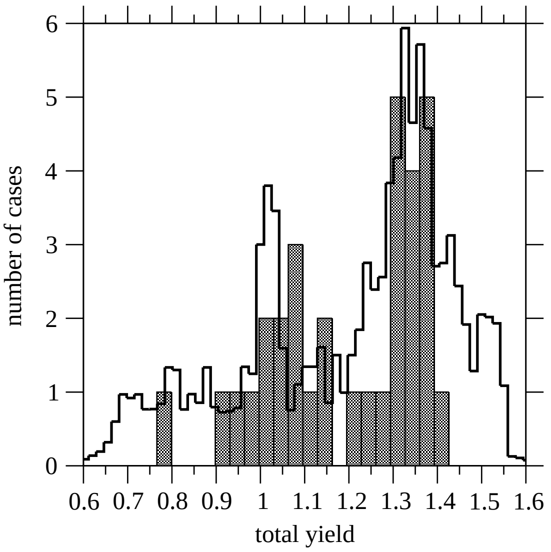

Fig. 12. Empirical histogram of the total yield for 32 public investment funds in the given period (gray bars). The solid line is the (unnormalized) index distribution for the same period shown also in Figure 3b.

compared to the large institutional investors and multinational companies, thus the assumption that their small volume transactions can not change drastically the prices, should be close to the reality. The indication of low trading activity coincides with the expectation, since these funds obey a careful investment strategy (according to their claims) without risky, *e.g.* day-trading transactions. Nevertheless the similarity of the histograms in Figure 12 does not indicate the presence of any sophisticated investment strategy for the public funds.

The author thanks Imre Kondor, Dietrich Stauffer, and Gábor Vattay for stimulating discussions and constructive remarks. This work has been partially supported by the Hungarian National Science Foundation (OTKA) under Grant No. T032437.

References

1. J.P. Bouchaud, M. Potters, *Théorie des Risques Financiers* (Alea-Saclay, Eyrolles, 1997) (in French).
2. R.N. Mantegna, H.E. Stanley, *Introduction to Econophysics: Correlation and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).
3. W. Li, Int. J. Bifurcation and Chaos **1**, 583 (1991).
4. R.N. Mantegna, Physica (Amsterdam) A **179**, 232 (1991).
5. R.N. Mantegna, H.E. Stanley, Nature **376**, 46 (1995).
6. N. Vandewalle, M. Ausloos, Physica (Amsterdam) A **246**, 454 (1997).
7. L. Laloux *et al.*, Phys. Rev. Lett. **83**, 1467 (1999).
8. V. Plerou *et al.*, Phys. Rev. Lett. **83**, 1471 (1999).
9. G.J. Stigler, J. Business **37**, 117 (1964).
10. H. Takayasu, H. Miura, T. Hirabayashi, K. Hamada, Physica (Amsterdam) A **184**, 127 (1992).
11. M. Levy, H. Levy, S. Solomon, J. Phys. I France **5**, 1087 (1995).
12. G. Caldarelli, M. Marsili, Y.C. Zhang, Europhys. Lett. **40**, 479 (1997).
13. T. Lux, M. Marchesi, Nature **397**, 493 (1999).
14. D. Stauffer, N. Jan, Physica (Amsterdam) A **277**, 215 (2000).
15. S. Ghasghaie, W. Breymann, J. Peinke, P. Talkner, Y. Dodge, Nature **381**, 767 (1996).
16. R.N. Mantegna, H.E. Stanley, Nature **383,** 587 (1996).
17. R.N. Mantegna, H.E. Stanley, Physica (Amsterdam) A **239**, 255 (1997).
18. M. Potters, R. Cont, J.P. Bouchaud, Europhys. Lett. **41**, 239 (1998).
19. See `www.amazon.com`, where dozens of textbooks are listed on portfolio theory and investment strategies.
20. J.V. Andersen, S. Gluzman, D. Sornette, Eur. Phys. J. B **14**, 579 (2000).
21. L. Molgedey, W. Ebeling, Eur. Phys. J. B **15**, 733 (2000).
22. J. Rotyis, G. Vattay, in *Econophysics: An Emerging Science*, edited by J. Kertész, I. Kondor (Kluwer, Amsterdam, 1999).
23. I.M. Jánosi, B. Janecskó, I. Kondor, Physica (Amsterdam) A **269**, 111 (1999).
24. R.A. Fisher, L.H.C. Tippett, Proc. Cambridge Phil. Soc. **24**, 180 (1928).