Contents

1. Introduction

The notion of market efficiency, broadly understood as impossibility of obtaining abnormal returns, has been a subject of much discussion starting from the seminal paper by Fama (1970). There are substantial difficulties in testing the hypothesis of market efficiency and despite the vast amount of theoretical and empirical studies, decisive answer is yet to be found. The first immediate difficulty is to define what can be considered as "normal" return, i.e. to specify the correct market model. But even after neglecting the possibility of time-varying risk premia and ignoring all sorts of operational market inefficiencies, like transactions costs and infrequent trading, when we are reduced to testing the constant expected returns property, it is still not clear, how significant coefficients in regression-based tests can be transformed into profit. There are a number of other caveats in the market efficiency testing, see Granger and Timmermann (2002) for an excellent survey.

In the recent years there has been an increasing attention to the issue of the profitability as opposed to a simple coefficient significance testing in the context of market efficiency, see Pesaran and Timmermann (1995) for an analysis of the predictability of stock returns in the context of its economic significance. The drawback of this approach is that the very fact that some strategy was able to obtain an abnormal profit during a certain period of time does not imply inefficiency, firstly because of potential data mining biases and secondly, due to difficulties in testing the significance of these type of results.

The primary purpose of this thesis is to construct a formal forecast quality measurement tool which in case of financial time series has a clear economic interpretation. This forecast quality measure also allows us to construct a new test for the constant expected returns property.

In Section 2 we argue that the conventional measures, like MSPE are probably not the most relevant in the context of financial time series. We introduce a simple trading strategy and construct a new Hausman-type test in the spirit of Pesaran and Timmermann (1992) for the constant expected returns property, which is based on the profitability of this trading strategy over some "flipping coin" portfolio.

In view of testing the constant expected returns property, we are interested in securing the best performing model available, since it increases the power of our test. It is generally accepted that various financial time series exhibit nonlinear patterns in their behavior. Abhyankar, Copeland and Wong (1997) document numerous published results on nonlinearity testing in financial data and it is noted that Brock-Dechert-Scheinkman(BDS) test almost invariably rejects the null of i.i.d for the returns process. However, the question of whether it is possible to build an econometric model which is able to catch this nonlinearity to the extent of obtaining good out of sample forecasts is yet unresolved. The bulk of the modelling attempts can be roughly divided into two parts: parametric and non-parametric ones, each of which has its own benefits and drawbacks. Parametric models usually enjoy a $\sqrt{n}$ or even better convergence rate but suffer from the possible misspecification. Non-parametric ones are free from the latter disadvantage, but the rate of convergence is usually worse. In both cases most of the researchers find that these models are unable to improve upon the simple out of sample forecast produced from the usual assumption that financial data follows constant expected returns pattern.

The variety of the parametric methods employed in the literature is remarkable, but we mention only some recent development: Aslanidis(2002) investigates the out-of sample forecasting performance of various threshold models for the UK stock market indexes.

Among the nonparametric models of particular interest are the local linear regression technics: Numerous studies on stock market indexes [Hsieh(1991), LeBaron(1998)], exchange rates [Meese and Rose (1990, 1991), Mizrach (1992), Diebold and Nason(1990), Satchell and Timmermann(1995)] suggest that taking the mean squared prediction error (MSPE) as a measure of forecast quality, it is impossible to improve significantly upon the simple "no change" forecast. On the positive side we mention Ramaswamy(1998) who argues, that MSPE is probably not the right measure of the economic forecast quality and is able to improve upon the "no change" forecast in terms of profitability using a relatively simple nearest neighbor Nadaraya-Watson estimator and Barinov et al(1999) who, among other estimators, investigate Nadaraya-Watson estimator in view of speculative decision-making.

In Section 3 we outline the general local polynomial regression framework and suggest a new empirical procedure for selecting bandwidth in the spirit of "nearest neighbor" technics. The procedure is rather specific for financial time series and is not supposed to have any outstanding theoretical properties in general.

In Section 4 some Monte Carlo simulation is performed to assess this procedure. We find that its performance in terms of classical measures approximately matches that of conventional nearest neighbor approach.

In Section 5 we describe daily RTS and weekly SP500 data which were used in the empirical part of our work and report some preliminary results. In particular, the results of the BDS test applied to the residuals suggest a strong nonlinearity and provide an additional motivation for the study of nonlinear models in our work.

In Section 6 the estimation results are reported and discussed. For both RTS daily and SP500 weekly data we are able to reject the null of constant expected return. We find that the power of our test is significantly greater (in case of RTS data) and approximately equal to (in case of SP500 data) to the power of the Pesaran and Timmermann test.

In Section 7 the results of the previous section are used to track the evolution of market efficiency through time. For RTS data our results mostly agree with those encountered in the literature and give evidence on the improving efficiency of Russian financial markets. For SP500 data we observe a non-trivial dynamics of the level of market efficiency throughout the last 50 years, in particular, we provide some evidence for the fact that it decreased substantially during the last decade.

The author would like to thank his thesis advisor Stanislav Anatoliev for valuable discussions concerning the material presented here.

2. Forecast evaluation and trading strategies

This section is the main theoretical contribution of this thesis. Based on the properly normalized profitability of the simple trading strategy, which takes some model's forecast as input, we construct a new measure of the out-of-sample forecast quality and the test for the constant expected returns property.

For motivating our model's forecast quality measure, it is necessary to discuss possible goals which the construction of such a model should pursue. Throughout the financial literature, the most frequently used (and questioned) model of financial time series is that of constant expected returns. The usual objective is, therefore, to build a model which will improve the quality of one-step forecast over the 'naive' forecast obtained via the above-mentioned models in terms of a certain loss functional. The choice of this loss functional turns out to be a very interesting subject. While we actually need certain structural market models to claim the impossibility of improving MSPE or mean absolute prediction error (MAPE), the most reasonable assumption is that of impossibility of obtaining a statistically significant profit in excess of the available riskless rate. Hellström (1998) argues that "trading" and "statistical" forecast problem should be separated and different quality measures should be used depending on what is the objective for constructing a model. A more formal approach is found in Leitch, Tanner (1991) who investigate the following question: "why profit maximizing firms buy professional forecasts when statistics such as MSPE or MAPE often indicate that a naive model will forecast about as well?". First they analyze the performance of several forecasting models(linear only) and services for the three-month Treasury bill rate and it appears that although the "naive" forecast is the best in terms of conventional forecast quality measures, it is inferior to the professional service forecast in terms of the profit obtained via various trading strategies based on the corresponding forecast. Splitting the data set into several periods they also find that the correlation between the MSPE and profits is positive in most of the cases. Ramaswamy(1998) also offers examples where the model with smaller MSPE is less profitable, and uses total return over the investment period to formulate an optimization problem and to compute optimal weights for the Nadaraya-Watson estimator.

Here we aim to unite the "trading" and "statistical" approaches and introduce a test for the null hypothesis of the return series having a constant expectation with respect to a certain information set $\mathcal{I}_{t-1}$. The test is designed in the spirit of the directional accuracy test of Pesaran and Timmermann (1992) and we outline its construction for the sake of completeness. The null hypothesis of this test is the independence of y_t and $\mathrm{E}[y_t \mid \mathcal{I}_{t-1}]$,where $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}...\}$.Clearly, if this does hold, the success ratio, defined as

$$SR = \frac{1}{T}\sum_{t=1}^{T} \mathbf{I}_{[0,+\infty)}(y_t \, \mathrm{E}[y_t \mid \mathcal{I}_{t-1}]) \tag{2.1}$$

cannot differ much from the expected success ratio that would be obtained in case y_t and $\mathrm{E}[y_t \mid \mathcal{I}_{t-1}]$ are independent. The estimate of the latter, denoted SRI, can be expressed as $SRI = P\hat{P} + (1-P)(1-\hat{P})$, where P and $\hat{P}$ are the shares of positive values among y_t and $\mathrm{E}[y_t \mid \mathcal{I}_{t-1}]$, correspondingly. The variances of SR and SRI are easily estimated and the test statistic of the Hausman type is

$$DA = \frac{SR - SRI}{\sqrt{\mathrm{var}(SRI) - \mathrm{var}(SR)}} \tag{2.2}$$

is distributed as $\mathrm{N}(0,1)$ under the null.

It should be noted, that the results of this test frequently get misinterpreted in the literature. The actual H_0 is independence of y_t and $\hat{y}_t|t-1$ at all lags and leads. Under the null of independence of of y_t and $\mathrm{E}[y_t \mid \mathcal{I}_{t-1}]$, the latter may by true only approximately and some correction of the size of the test is needed, which is the subject of further research.

Our test is an extension of the test of Pesaran and Timmermann and is based on the out of sample profitability of the trading strategy corresponding to an arbitrary forecasting model. By trading strategy here we mean a rule which issues a buy signal if the next day return forecast is greater than 0 and sell otherwise.

We start with a stationary time series y_t ("returns") and we are interested in testing the H_0 : $\mathrm{E}[y_t \mid \mathcal{I}_{t-1}] = \text{const}$, where $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}...\}$ is a certain information set. Suppose $s_t \subseteq \mathcal{I}_{t-1}$

is a stationary time series, which is used for issuing the trading signal of the form described above. Strictly speaking, under the null we ask that sign s_t be independent from y_t for all lags and leads, however, just like in the case of the Pesaran-Timmermann test, the results appear to be close to exact in case sign s_t is meaningful estimator of sign y_t. Note that the signals of the form "buy if previous day return is positive and sell otherwise", which are mimicked by setting $s_t = y_{t-1}$ are ruled out.

The expected daily return of the trading strategy based on this signal is $\mathrm{E}[\mathrm{sign}(s_t)y_t]$,and our test is based on the following two estimators for this return, which are both consistent under the null:

$$A_T = \frac{1}{T}\sum_t \mathrm{sign}(s_t)y_t \tag{2.3}$$

and

$$B_T = (\frac{1}{T}\sum_t \mathrm{sign}(s_t))(\frac{1}{T}\sum_t y_t) \tag{2.4}$$

Note, that A_T represents an average daily return of this trading strategy, and B_T can be interpreted as an expected return of a "flipping coin" strategy with the same proportion of "buy" and "hold" signals, as the strategy based on s_t. Under the null $\mathrm{plim}(A_T - B_T) = 0$ and it remains to compute the variance of $(A_T - B_T)$.

Our approach is based on the idea of the Pesaran-Timmermann test: the estimator B_T is efficient under the null and inconsistent under the alternative, thus, Hausman-type considerations are available. Unfortunately we were unable to prove this statement *per se*, so we resort to the direct computation of the joint distribution of A_T and B_T.

$$\mathrm{var}(A_T) = \frac{1}{T}(\mathrm{E}[\mathrm{sign}(s_t)^2 y_t^2] - \mathrm{E}[\mathrm{sign}(s_t)y_t]^2) = \frac{1}{T}(\mathrm{E}[y_t^2] - (2p_s - 1)^2\,\mathrm{E}[y_t]^2) \tag{2.5}$$

$$\text{(2.6)} \qquad \begin{gathered} \operatorname{var}(B_T) = \mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t))^2(\frac{1}{T}\sum_t y_t)^2] - \mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t))(\frac{1}{T}\sum_t y_t)]^2 = \\ \mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t))^2]\,\mathrm{E}[(\frac{1}{T}\sum_t y_t)^2] - \mathrm{E}[\frac{1}{T}\sum_t \operatorname{sign}(s_t)]^2\,\mathrm{E}[\frac{1}{T}\sum_t y_t]^2 = \\ (\operatorname{var}[\frac{1}{T}\sum_t \operatorname{sign}(s_t)] + \mathrm{E}[\frac{1}{T}\sum_t \operatorname{sign}(s_t)]^2)(\operatorname{var}[\frac{1}{T}\sum_t y_t] + \mathrm{E}[\frac{1}{T}\sum_t y_t]^2) - (2p_s-1)^2\,\mathrm{E}[y_t]^2 = \\ (\frac{4}{T}p_s(1-p_s) + (2p_s-1)^2)(\frac{1}{T}\,\mathrm{E}[y_t^2] + \frac{T-1}{T}\,\mathrm{E}[y_t]^2) - (2p_s-1)^2\,\mathrm{E}[y_t]^2 = \\ \mathrm{E}[y_t^2](\frac{1}{T}(2p_s-1)^2 + \frac{4}{T^2}p_s(1-p_s)) + \mathrm{E}[y_t]^2(-\frac{1}{T}(8p_s^2 - 8p_s + 1) - \frac{4}{T^2}p_s(1-p_s)) \end{gathered}$$

$$\text{(2.7)} \qquad \begin{gathered} \operatorname{cov}(A_T, B_T) = \mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t)y_t)(\frac{1}{T}\sum_t \operatorname{sign}(s_t))(\frac{1}{T}\sum_t y_t)] - \\ \mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t)y_t)]\,\mathrm{E}[(\frac{1}{T}\sum_t \operatorname{sign}(s_t))(\frac{1}{T}\sum_t y_t)] = \\ \frac{1}{T^2}\,\mathrm{E}[y_t^2] + \frac{T-1}{T^2}(2p_s-1)^2\,\mathrm{E}[y_t^2] + \frac{T-1}{T^2}\,\mathrm{E}[y_t]^2 + \frac{(T-1)^2}{T^2}(2p_s-1)^2\,\mathrm{E}[y_t]^2 - (2p_s-1)^2\,\mathrm{E}[y_t]^2 = \\ \mathrm{E}[y_t^2](\frac{1}{T}(2p_s-1)^2 + \frac{4}{T^2}p_s(1-p_s))) + \mathrm{E}[y_t]^2(-\frac{1}{T}(8p_s^2 - 8p_s + 1) - \frac{4}{T^2}p_s(1-p_s)) = \operatorname{var}(B_T) \end{gathered}$$

Thus, indeed, $\operatorname{var}(A_T - B_T) = \operatorname{var}(A_T) - \operatorname{var}(B_T) = 4p_s(1-p_s)\frac{T-1}{T^2}\operatorname{var}(y_t)$ and it follows, that under the null the following estimator of $\operatorname{var}(A_T - B_T)$ is valid:

$$\text{(2.8)} \qquad \hat{V}(A_T - B_T) = 4\frac{1}{T^2}\hat{p}_s(1-\hat{p}_s)(\sum_t (y_t - \bar{y})^2).$$

The test statistic is constructed in a standard way and under the null we have

$$\text{(2.9)} \qquad \frac{A_T - B_T}{\hat{V}^{1/2}(A_T - B_T)} \xrightarrow{d} \mathrm{N}(0,1)$$

This statistic will be referred to as "profitability statistic".

Finally, note that the better our forecasting model is, the greater the expected value of this statistic under alternative hypothesis will be, thus, in view of increasing the power of our test we are interested in securing the best model available.

Under the assumption of a constant risk premia and no operational market imperfections, these kind of tests are usually referred to as tests of "informational market efficiency". In particular, if we take $\mathcal{I}_t = \{y_t, y_{t-1}...\}$, we obtain a test of the "weak form market efficiency", and if we add to $\mathcal{I}_t$ some other variables publicly known as of the moment $t-1$, a test of the semi-strong efficiency of the market will be obtained. The notion of informational market efficiency can be interpreted as a statistically testable proxy for the market efficiency *per se,* that is the impossibility of obtaining abnormal profits on the basis of some information. The main benefit of our approach to testing informational market efficiency is that unlike regression-based or volatility tests it is based on the out-of sample forecasting profitability thus not only is free of "spurious regression"-type biases but also shows direct connections to the underlying notion of market efficiency.

3. Local polynomial regression

In this section we outline the general local polynomial regression framework and describe our bandwidth selection procedure.

First we fix some notations. For the model

$$y_t = m(y_{t-1}, y_{t-2}...y_{t-k}) + \varepsilon_t \tag{3.1}$$

we denote $z_t = (y_{t-1}, y_{t-2}...y_{t-k})'$, $z_{t,i} = y_{t-i}$. The forecast at a point z produced by local polynomial regression can, in the most general form, be written out using the solution to the following minimization problem

$$\sum_{s=k+1}^{t-1} (y_s - P_l(z - z_s))^2 \mathbf{K}(z - z_s, h(z)) \to \min_{P_l}, \tag{3.2}$$

where P_l is a polynomial of degree l, K is a kernel function and $h(z)$ is a (possibly non-scalar) bandwidth. The coefficients of P_l are estimated straightforwardly by GLS and the estimated intercept of this polynomial is the point forecast $\hat{y}(z)$. In this paper we are mostly interested in the case when the number of lags k can be greater than 1, thus fitting polynomial of degree higher than 1 becomes cumbersome. Moreover, in view of the point forecasting problem fitting polynomials of higher order is also not crucial. Thus, the minimization problem becomes

$$\sum_{s=k+1}^{t-1} (y_s - \alpha - (z - z_s)'\beta)^2 \mathbf{K}(z - z_s, h(z)) \to \min_{\alpha,\beta} \tag{3.3}$$

There is a substantial amount of subjectivity in the treatment of the $\mathbf{K}(z - z_s, h(z))$ term above.

The selection of the underlying one-dimensional kernel function is certainly not the most important choice to be made, since the loss of effectivity due to the non-optimal kernel function selection is negligible. In the empirical literature the most commonly used kernels are uniform, triangle, tricube and standard normal.

In case $k > 1$ we have to choose the way of constructing multidimensional kernel functions out of one-dimensional ones. There two main ways to do that.

First, product kernel functions:

$$\mathbf{K}(z - z_s, h(z)) = \prod_{i=1}^{k} K(\frac{z_i - z_{s,i}}{h_i(z)}) \tag{3.4}$$

Second, norm kernel functions:

$$\mathbf{K}(z - z_s, h(z)) = K(\left\| (\frac{z_1 - z_{s,1}}{h_1(z)}, \frac{z_2 - z_{s,2}}{h_2(z)}, ... \frac{z_k - z_{s,k}}{h_k(z)}) \right\|), \tag{3.5}$$

where by$\|\bullet\|$we denote any norm on R^k and K is a one-dimensional kernel function.

It appears that the most important choice concerns the way the function $h(z)$ is constructed. In case of financial data we have excessively non-uniform design of the regressor space, which means that adjusting bandwidth to the boundary of the regressor space becomes crucial. Indeed, it is clear that in this case the variance term plays the main role near the boundary. This is especially true in case of local linear regression, where we have a better control over the boundary bias, thus increasing bandwidth near the boundary in view of decreasing the variance appears to be reasonable. Thus it can be suggested that various global bandwidth selection technics, including crossvalidation, are not likely to be the best choice in this case. Other problems arise when we switch to the multidimensional locally linear regression, since we have to increase bandwidth at the boundary of the regressor space for each of the regressors, the latter complication usually being ignored.

A family of the "nearest neighbors" algorithms offers a possible solution to this problem. The idea is that the optimal bandwidth function $h(z)$ which is supposed to catch the behavior of the regressors in the neighborhood of z can be estimated using the set of k nearest neighbors of z among the available sample. More precisely, first, the set of neighbors of the point z is selected according to a certain metric, and then every component of the function $h(z)$ is set to equal the distance from the point z to its farthest neighbor or some other measure of this set of neighbors.

Nearest neighbor methods appear to be the most popular in the empirical literature, since unlike various global bandwidth selection methods they at least allow for the automatic bandwidth adjustment at the boundary of the regressor space. Another reason is offered by Cleveland

and Loader(1996b), who note that these methods offer control over the number of points being smoothed, which is convenient from the practical point of view. In Cleveland and Loader(1996a) it is also argued that there seems to be a substantial gap between the recommendations of the asymptotic theory and the problems encountered in finite samples. In particular, the global bandwidth selection turn out to perform worse than the nearest neighbors methods in practice, which contradicts the asymptotic theory.

Before describing the contribution of the current paper we record the usual settings of this algorithm encountered in the empirical work. Diebold and Nason(1990), Meese and Rose(1990,1991) use tricube kernel, Euclidean norm kernel function and the distance to the farthest neighbor as the bandwidth, while the number of neighbors is selected manually. Mizrach (1992) and Barkoulas, Baum(1996) employ triangle kernel, Euclidean distance and the sum of distances to all neighbors as the bandwidth, while the number of neighbors is also selected manually. It should be noted that in the latter case the actual weight system is not a continuous function of the distance from z, since observations not belonging to the set of neighbors are assigned a zero weight, while the weight for the farthest of neighbors is non-zero.

Here we propose an algorithm which is based on the following observation: Suppose that one of the components of z, say, z_i, is relatively close to the boundary of the corresponding one-dimensional regressor space. It is natural to assume, that in this case the corresponding component of the optimal bandwidth, $h_i(z)$, should be greater than in the case when z_i is in the interior of the regressor space. Although this observation appears to be widely discussed in the literature on kernel density estimation since mid-sixties, there are few results in the case of local linear regression. We mention Yang and Tschering(1999) who use elaborate plug-in technics to estimate the optimal bandwidth vector. The nearest neighbors procedure also allows one to take this into account and we propose two variants of the algorithm (which may well be known in case of density estimation).

1. Set $X_k(z)$ to be the set of k nearest neighbors of the point z in the L^∞ metric. Set

$$h_i(z) = \sup_{h \in X_k(z)} (|z_i - h_i|) \tag{3.6}$$

2. Let $X_{k,i}(z)$ be the set of k nearest neighbors for the point z_i among $z_{s,i}$. Set

$$h_i(z) = \sup_{h \in X_{k,i}(z)} (|z_i - h|) \tag{3.7}$$

A thorough evaluation of this proposals in the general prediction context falls out of the scope of this paper, and it should be noted that here we have no hope for the asymptotic optimality of this selection in the general case, since the information about the dependent variable is not used in the selection of h. Some simulation study results are provided in the next section.

Here we just make some comments concerning their possible relevancy to the financial time series prediction task. It is well known that a substantial and easy detectable nonlinearity is present in these series. The fact that this easily detected nonlinearity so far haven't resulted in a creation of good prediction model can be explained in different ways. One of the explanations provided by Diebold and Nason (1990) is that the outliers may cause linearity test to reject the null, while being useless for the out of sample prediction. Here we propose a different view on the same point. It is natural to assume that the degree of market efficiency is not constant in time as is our ability to predict its behavior. Moreover, outliers in returns series correspond to the moments in time, when the market is likely to be less efficient. An easy example is a news under/overreaction, when the big market moves happen not only directly after the news announcement, but also for some time after, which clearly contradicts the market efficiency hypothesis. Thus, in view of the profitability of the corresponding trading strategy, improved prediction quality in these cases seems to be more important, thus providing the rationale for the proposed algorithms.

4. Simulation study

Our goal is mainly to compare the performance of the proposed bandwidth vector selection procedure and the classical nearest neighbors approach, but we also report the results of other kernel estimators.

We consider the following model with a simple nonlinearity in one of the variables:

$$y_n = x_{1,n}\mathbf{I}_{[0,+\infty)}(x_{1,n}) + x_{2,n} + x_{3,n} + \varepsilon_n \tag{4.1}$$

where $(x_{1,n}, x_{1,n}, x_{3,n})$ are i.i.d. and distributed as $\mathrm{N}(0, 100)$ and ε_n is independent of x's and is distributed as $\mathrm{N}(0, 100)$. The sample size is 100 and the number of Monte-Carlo repetitions is 100. The quality of forecast is estimated as MSPE computed over the points z_m, where the first coordinate ranges from -20 to 20 with the step equal to 0.4 (101 points in total) and the other coordinates are zero. The parameters of the algorithms were chosen *ex post* optimal, to ensure that all of the estimation methods are in the same conditions.

The results are reported in Table 1 below. The following should be noted: since the model itself is locally linear, all of the local linear regression estimators perform much better than the Nadaraya-Watson estimator due to the zero bias. The performance of the procedure suggested in Alg.2 is unsatisfactory, probably due to the fact, that it goes too far in dealing separately with each of the regressors, thus the control over the actual number of points being smoothed is weak. The best performing model in terms of MSPE is the classical nearest neighbors approach, but the difference among the remaining three models is not very significant.

5. The data and estimation settings

We analyze the RTS stock market index for the period 01.09.1995-18.04.2003 (1909 observations in total). This index is a basic and most frequently used indicator of the Russian stock market. As of 01.01.2003 it is comprised of the dollar prices of 59 shares of Russia's 35 leading companies, weighted according to their capitalization and expressed in terms of the initial period value. The detailed information concerning the way it is computed and the data itself are available via www.rts.ru.

The alternative index, MICEX, accounts for only 18 most liquid stocks and is denominated in roubles, thus it is less suitable for our needs.

We also use weekly data on SP500 stock market index for the period 03.01.1950-05.05.2003 to investigate, how sensitive our results are to the choice of dataset.

To obtain stationarity we take first differences of the logarithms of the index.

As was noted in the construction of the market efficiency test in Section 2, we are interested in securing the best performing model, since it increases the power of our test. Thus, it seems reasonable to report the results of the BDS test for the raw returns and the residuals of the simple linear model and see whether they call for the nonlinear models to be considered. The results for the RTS data are presented in Table 2. Not surprisingly, it can be seen that AR model fails to detect the dependence in the data caught by the BDS test, i.e. building a nonlinear model appears to be reasonable.

We employ various nonparametric kernel estimators and use linear model predictions as benchmarks. Parameters selection, in particular, bandwidth choice is known to be very important step in choosing the appropriate nonparametric model.

The kernel function was chosen to be usual standard normal, since the theoretical efficiency loss due to this selection (comparing to Epanechnikov kernel) is negligible (cf., for example, Härdle(1990), section 4.5) and, moreover, it allows for the smaller global bandwidth to be chosen due to the unbounded support.

In the multivariate case the product kernel was used.

The global bandwidth in the case of Nadaraya-Watson estimator and local linear regression was chosen quite arbitrarily. Crossvalidation produces rather unstable results which, when finite, didn't differ much from the finally selected bandwidth. On the other hand, excess manual parameter tweaking may result in data-snooping biases (cf., for example, Lo and MacKinlay (1990)).

In case of the nearest neighbors local regression we employed the first of the algorithms of bandwidth selection described in the previous section. The number of neighbors selected by the crossvalidation again tends to be greater than optimal, thus the number of neighbors was manually selected to equal 10 or 100.

The number of lags was chosen to be equal to 5 in case of daily RTS data and 1 in case of weekly SP500 data, since in the latter case it is hard to expect any dependence for the lags beyond the first one. We assess the quality of the one step forecast using rolling regression with block of constant length equal to 300 (1909-1-5-300=1603 predictions in total) in case of RTS data and 104, i.e. two years (2783-1-1-104=2677 predictions in total) in case of SP500 data.

Finally, in the construction of the 'profitability statistic' defined in section 2, we are setting $s_t = \hat{y}_t$, where $\hat{y}_t$ is the forecast produced by the corresponding model.

6. The results

The estimation results are reported in the Table 3. Linear model and 'naive' forecast are used as benchmarks to estimate the accuracy of forecasts produced by the Nadaraya-Watson estimator and various types of local linear regressions. We use either a global scalar bandwidth or Alg. 1 for the bandwidth vector selection described in Section 3. The empirical results of Alg. 2 are somewhat worse, probably due to the low control over the actual number of point weighted and are not reported (but are available upon request).

The first thing to be noted is that no model succeeded in improving over the forecast from the RW hypothesis in terms of MSPE. This is in line with the results encountered in the literature. However, the results of the directional accuracy test of Pesaran and Timmermann (1992) suggest a strong predictability even in the case of the usual linear model. The results become even more impressive if we take into account the profitability of the trading strategy based on the corresponding models. The last column in Table 3 reports the values of the profitability statistic which are highly significant in most cases (note, that raw cumulative returns, while being impressive, should be taken with a grain of salt, especially in case of RTS data, since transaction costs and market restrictions are taken into account.). Just as our simulation result suggest, the usual nearest neighbor algorithm tends to perform better in terms of MSPE than Alg.1. However, other forecast quality measures, like directional accuracy and profitability statistic speak in favor of our method. (Note, that in case of SP500, where we took only 1 lag of y, there is no difference between our bandwidth selection algorithm and the classical procedure, hence, the results are the same).

Fig. 1 and Fig. 2 show the graphs of the log cumulative returns (in case of h=0.03 and n=10) of the portfolios based on various forecasting models and our simple trading strategy. For the case of RTS index the trading strategies based on the different forecasting models turn out to be ordered according to our perception of their relative power. The best performing one is the local linear regression with the bandwidth chosen according to the Alg.1, then comes local linear regression with global bandwidth, Nadaraya-Watson with global bandwidth and a linear model. The only difference in case of SP500 is that Nadaraya-Watson estimator underperformed all linear and local linear estimators, which could be explained by the fact that it belongs to the class of locally constant estimators, thus, in case linear models are a better proxy for the true market model, this estimator has a substantial bias at the boundary of the regression space, which is the most crucial region for the profitability.

Apparent similarity between the graphs of log cumulative returns should be noted. The periods where the gains from the trading strategies are the most evident are the same for all strategies, which suggests that the results of our test are relatively robust to the choice of the forecast model.

It can clearly be seen, that the constant expected returns hypothesis for the daily returns of the RTS index as well as weekly returns of SP500 index is strongly rejected. It is interesting to observe, that while MSPE tends to decrease when h or the number of neighbors increases, other forecast quality measures behave differently. This can be explained by the following: taking the smaller number of neighbors (smaller h) is actually much more in the spirit of "finding similar prehistories" idea, i.e. is more likely to allow us to catch certain subtle dependencies in the data. In this case outliers can strongly influence the quality of the forecasts, as measured by MSPE. However, DA and "profitability" statistics are influenced by the prediction errors only to the extent that they may change the sign of the prediction, i.e. they are more robust. A possible explanation for the apparent impossibility of a significant improvement of MSPE out of sample can be suggested: Even if the data does contain predictable nonlinearity, it is clear that the stochastic component, which can not be predicted, dominates the data (cf.

Abhyankar et al(1997)), hence, its variance dominates the MSPE. Properly normalized out-of-sample profitability is free from this drawback and may be suggested as a right measure for the forecast quality in case of financial data.

7. The evolution of market efficiency

The raw results of the previous section, i.e. the daily return series of portfolios based on various trading strategies can be used not only for testing the hypothesis of market efficiency over the whole period, but also for tracking the evolution of market efficiency through time. This is obtained by computing the profitability statistic over all subperiods of fixed length of our sample. The length of this subperiod should be chosen to satisfy the following two conditions: it should not be too small, to include too much noise, and it should not be too large, to smooth out some interesting details. For the case of RTS daily data the compromise was obtained at the block length equal to 250, which corresponds to 1 year, and for weekly SP500 data, 520, which corresponds to 10 years. It is natural to choose the best performing model in terms of the profitability statistic, thus only the results for the case of the local linear regression with the bandwidth selected according to the Alg.l are reported.

The results for RTS and SP500, along with upper 5% confidence bands are reported in Fig 3 and 4, correspondingly. Note, that market inefficiency at a point in time is unobservable, so the values of the functions at each point, i.e. the values of the profitability statistic over some period of 1 year (10 years), should be interpreted as the measure of market inefficiency over the corresponding period. We are mostly interested in global minima and maxima, which are likely to account for structural breaks in the market efficiency.

For the case of RTS data our results are not completely in line with those of Hall and Urga(2000), who claim a substantial increase in the market efficiency starting from the first quarter of the year 1999. The first period for which constant expected returns are not rejected on 5% level by our test is 08.1999 - 07.2000 which is more in line with the results of Rockinger, Urga(2000) who notice that the predictability of the Russian market remains high up until the end of the year 1999. From the beginning of the year 2001 there seem to be no evident trend in the data and constant expected returns are not rejected.

For the case of SP500 data we can clearly observe 2 points, where the structural breaks in the market efficiency might have occurred. The first one corresponds to the global maximum of the profitability statistic over our sample, which occurs for the period from 1963 to 1973. Starting from the end of this period and up to the year 1992 an apparent trend towards improving market efficiency can easily be seen. In fact, for the period 1982-1992 the value of the profitability statistic is close to zero. However, after the year 1992 a reversal of this trend is observed and for the period 1993-2003 the hypothesis of constant expected returns is accepted only marginally. If this 11-year long trend continues, we are likely to see the hypothesis of weak informational market efficiency rejected for the SP500 data in the nearest future.

8. Conclusions and directions for further research

We find a corroborating evidence for the fact that the positive relation between the MSPE of a model and the profitability of the corresponding trading strategy is ambiguous at best and propose a new formal forecast quality measure which, in case of return series, has a clear economic interpretation as a normalized return of a certain trading strategy. Based on this measure, we constructed a test for the constant expected return property.

We investigated the performance of various non-linear nonparametric models in view of their forecasting power and suggested a new bandwidth selection procedure for the case of local linear regression.

In our empirical studies we find that although for both RTS daily and SP500 weekly data the hypothesis of constant expected returns is rejected, there is remarkable dynamics in the level of market efficiency throughout the observation period.

Among the most interesting directions of further research we note the following: It is interesting to note that obtaining profit via out of sample forecasting appears to be easier than improving MSPE of the 'naive' forecast. This is true despite the fact that the underlying minimized functional is actually the in-sample squared prediction error (with certain weights in case of local linear regression). Substituting this functional for the in-sample profitability might lead to the improvement of the profitability of the corresponding out of sample forecasts. The theoretical framework for these kind of estimators is developed in the Appendix A and will be a subject of further empirical studies.

Appendix A: Profitability-based estimators

Traditionally, the first thing to be considered while choosing the estimator is its asymptotic properties. Under the assumption of normality, this usually implies that the "in-sample loss function" is quadratic in the prediction error. However, depending on the expected application of the forecasts produced by the model, the actual "out of sample loss" may be quite arbitrary.

Here we consider the case which is likely to be quite common in case of financial data, namely, when we can describe the expected loss resulting from the forecast' error. Indeed, suppose that the forecast $\hat{y}_t$ is used for trading via the strategy "buy if $\hat{y}_t > 0$,sell otherwise". Then expected loss from this strategy is $\mathrm{E}[-\operatorname{sign}(\hat{y}_t)y_t]$ and it seems natural that the estimator $\hat{y}_t$ should minimize this loss instead of usually considered $\mathrm{E}[(\hat{y}_t - y_t)^2]$.Formally, consider the model:

$$y_t = x_t'\beta + \varepsilon_t \tag{8.1}$$

and the estimator

$$\hat{\beta}_p = \underset{b}{\operatorname{argmin}}\left(\frac{1}{N}\sum -\operatorname{sign}(x_t'b)y_t\right) \tag{8.2}$$

Note that the minimized function is in fact the incurred loss from using the simple trading strategy with the forecasts from the model. It is also easy to introduce transaction costs as well as more elaborate trading strategies into this picture.

Since $\hat{\beta}_p$ is defined only up to a multiplication by a positive constant, it is certainly not consistent, however, there are some reason to believe that it allows one to consistently estimate the quotients $\frac{\beta_i}{\beta_j}$. There are obvious theoretical as well as practical problems associated with this estimator, since the minimized function is not differentiable and we are planning to deal with them in a sequel.

Appendix B: Tables and Figures

Table 1. The results of Monte-Carlo simulations

NW, $n=4$	LWR, $h=6$	NN, $n=6$	NNA1, $n=6$	NNA2, $n=30$
6,05	0,73	0,66	0,68	2,05

NW- Nadaraya-Watson estimator;

LWR -locally linear regression with global bandwidth;

NN-locally linear regression with the scalar bandwidth

NNA1-locally linear regression with the bandwidth vector selected according to Alg.1

NNA2-locally linear regression with the bandwidth vector selected according to Alg.2

Table 2. BDS test statistic for the daily returns of RTS index

BDS test dimension	Raw returns	AR residuals
2	14,82	12,82
3	17,11	15,61
4	18,48	17,15
5	19,92	18,79
6	21,45	20,46

The AR model was selected according to BIC (lags 1, 10, 12 are included)

The reported numbers are the values of z-statistic, which is distributed as N(0,1) under the null

Table 3a. Forecasting performance- RTS daily

	Sum squared prediction errors	DA-statistic	Cumulative return	Profitability statistic
RW	1,62		2,10	
OLS	1,65	3,36**	345,89	4,56**
NW, h=0.03	1,69	3,05**	362,22	4,66**
NW, h=0.1	1,61	1,71*	6,35	1,33*
LWR, h=0.03	2,43	3,67**	1680,21	5,81**
LWR, h=0.1	1,67	4,03**	811,8	5,24**
NNA1, n=10	1,70	4,15**	2090,73	5,98**
NNA1, n=100	1,63	4,31**	1462,77	5,69**
NN, n=10	1,68	3,80**	1426,72	5,68**

Table 3b. Forecasting performance - SP500 weekly

RW	1,07		38,97	
OLS	1,1	2,52**	50,94	2,48**
NW, h=0.03	1,09	0,18	9,82	0,13
NW, h=0.1	1,08	-0,05	10,80	0,10
LWR, h=0.03	1,16	2,31*	77,63	2,97**
LWR, h=0.1	1,10	2,92**	64,56	2,73**
NNA1, n=10	1,14	4,24**	138,50	4,03**
NNA1, n=100	1,10	2,63**	60,40	2,66**
NN, n=10	1,14	4,24**	138,50	4,03**

OLS-Linear model; NW- Nadaraya-Watson estimator;

LWR -locally linear regression with global bandwidth;

NNA1-locally linear regression with the bandwidth vector selected according to Alg.1

NN-locally linear regression with scalar bandwidth

By cumulative return in case of RW prediction we mean that of buy-and-hold strategy

DA-statistic is a directional accuracy statistic of Pesaran and Timmermann

Profitability statistic is defined in 2.9

All reported statistics are distributed as N(0,1) under the null.

*(**) indicates that the corresponding statistic is significant at 5%(1%) level

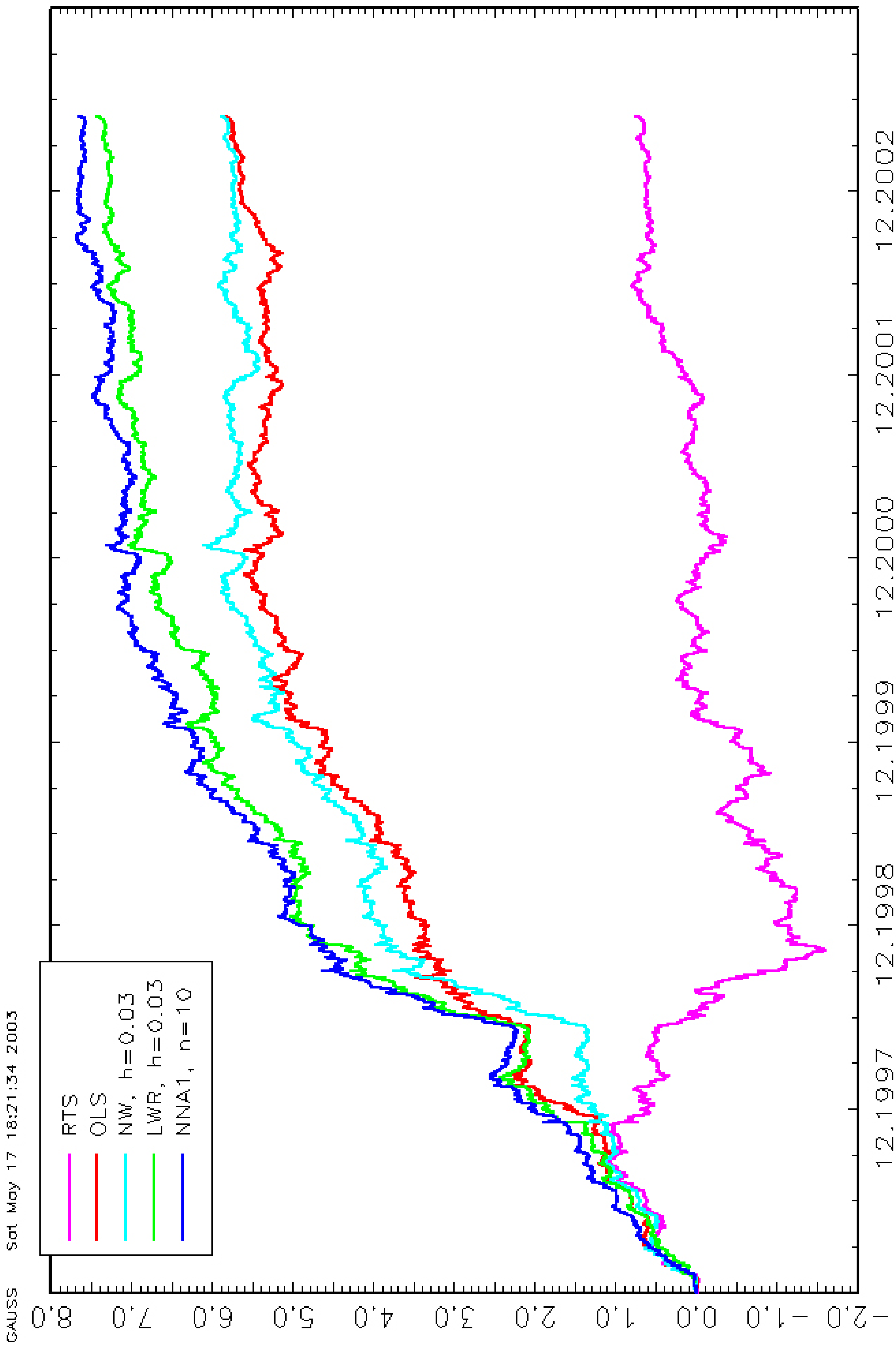

FIGURE 1. Log cumulative returns of the trading strategies - RTS daily

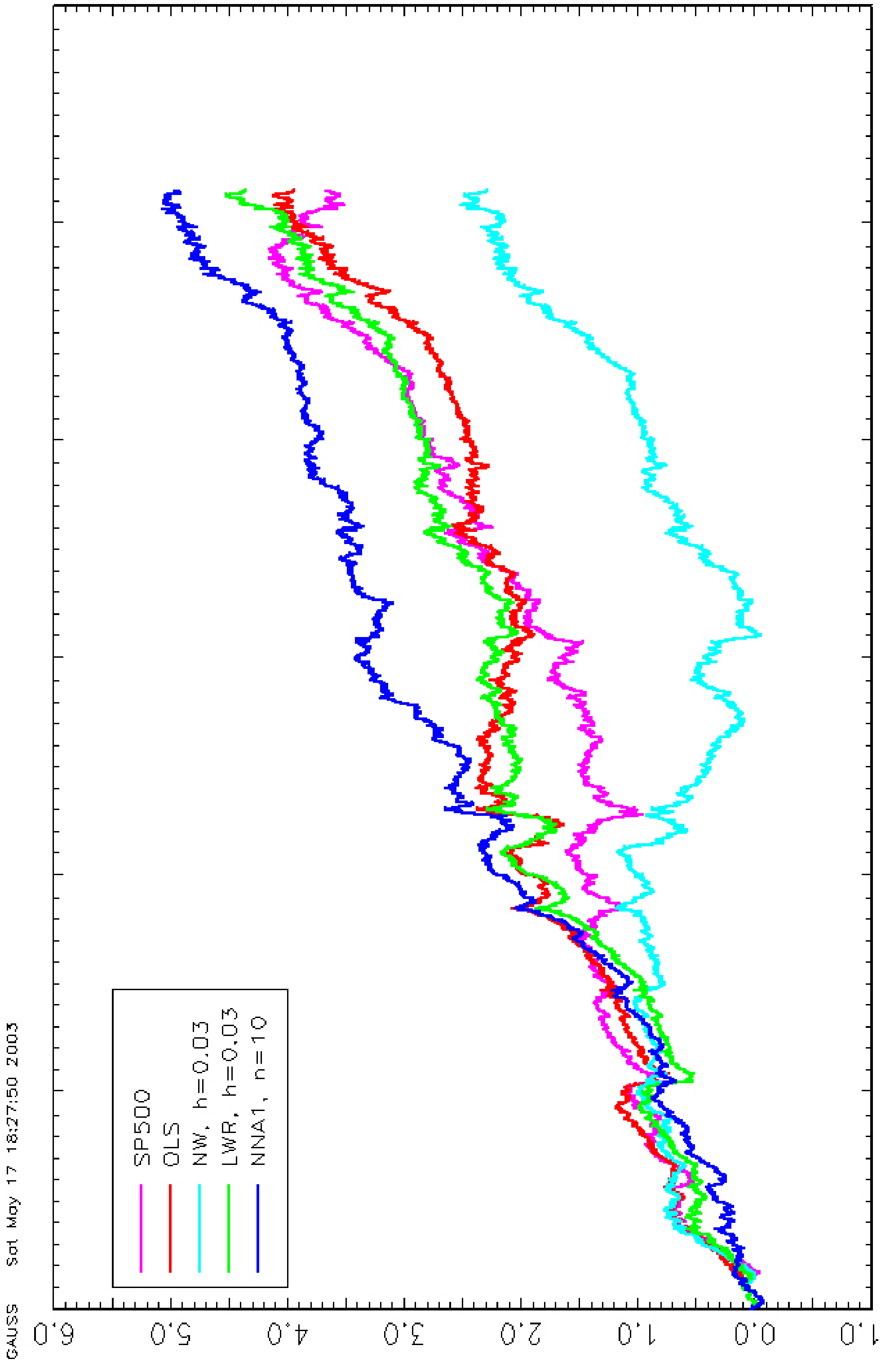

FIGURE 2. Log cumulative returns of the trading strategies - SP500 weekly

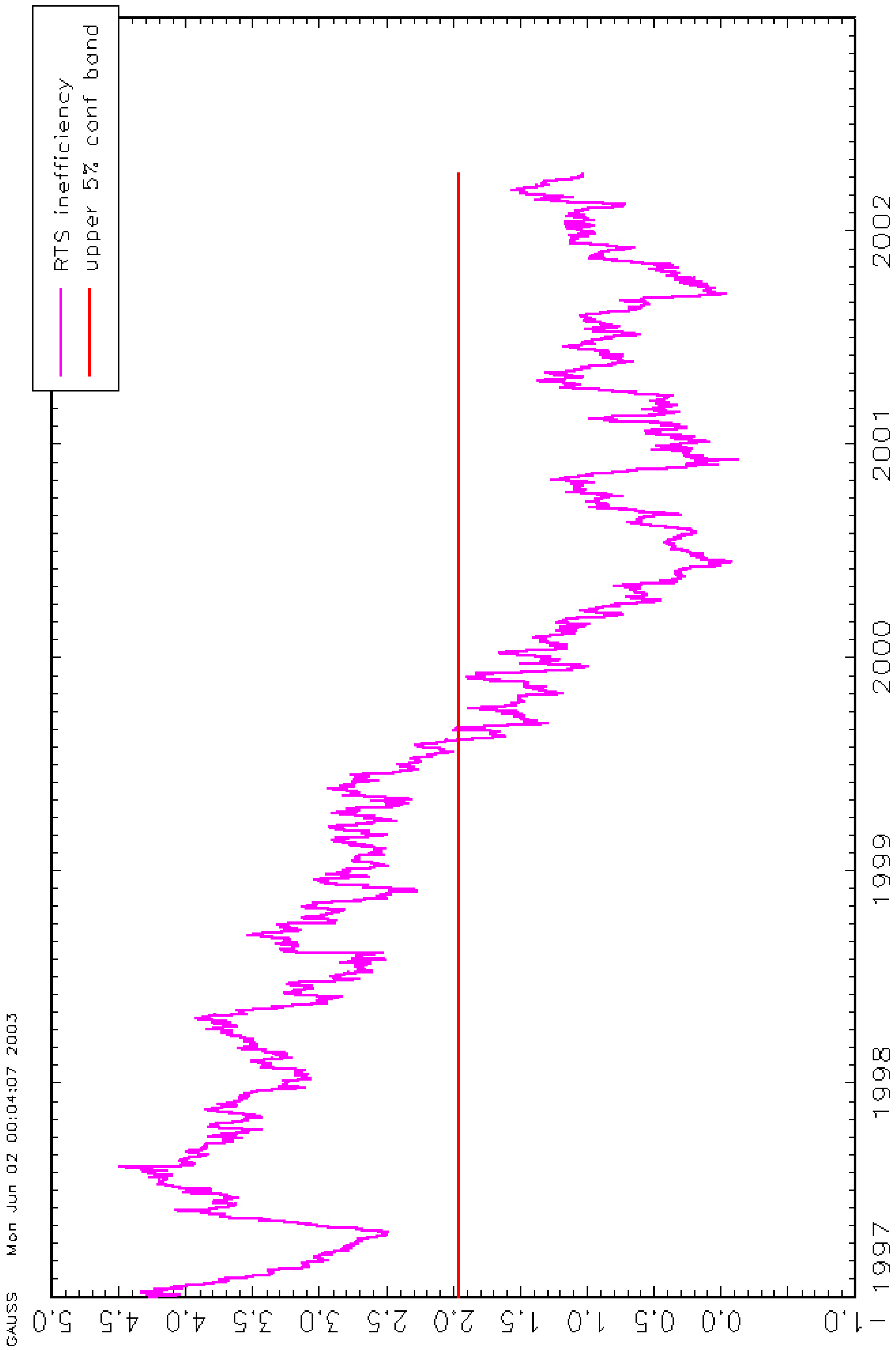

FIGURE 3. Market efficiency evolution - RTS daily

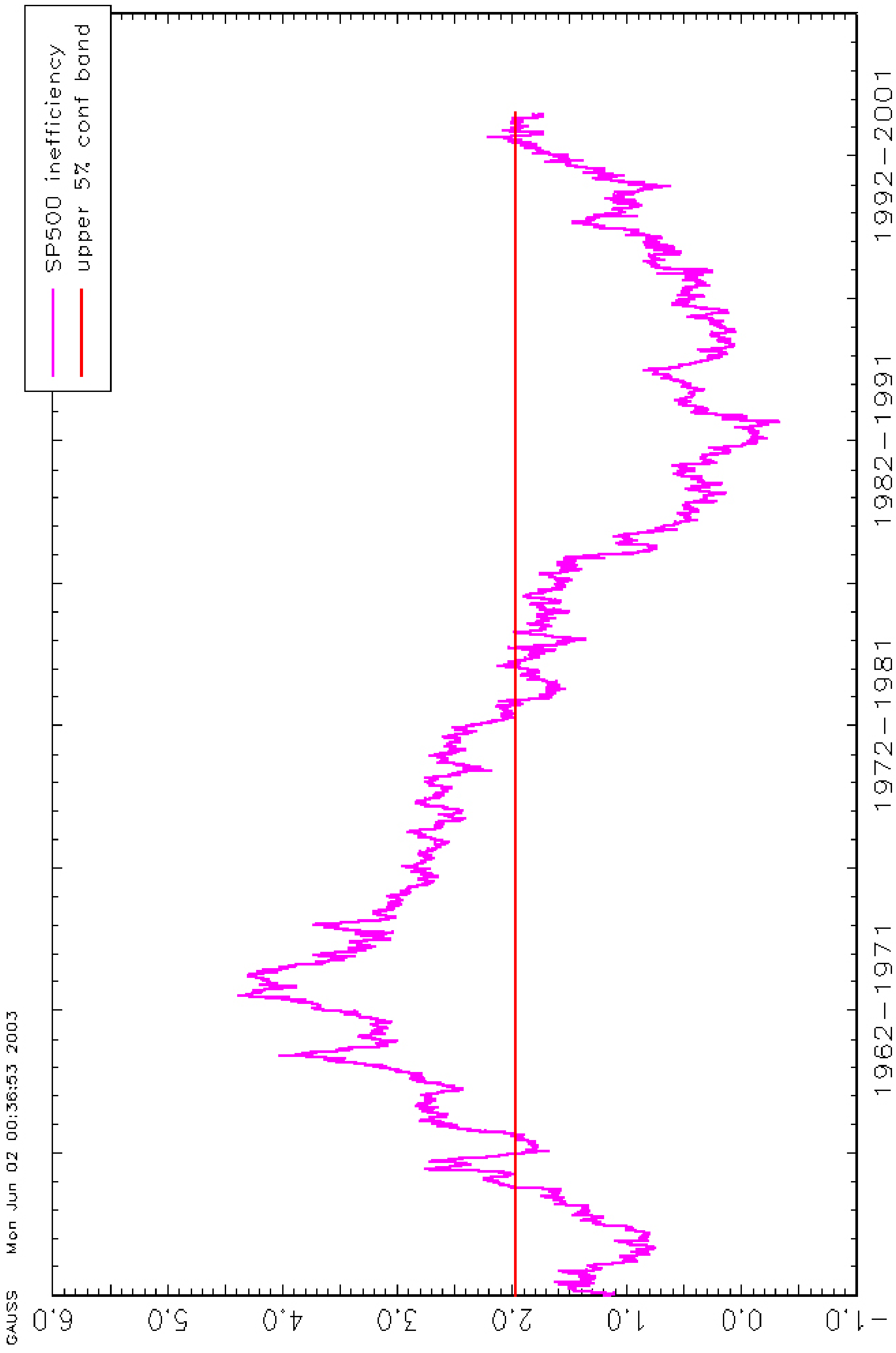

FIGURE 4. Market efficiency evolution - SP500 weekly

References

[1] A. Abhyankar, L. S. Copeland and W. Wong (1997): 'Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100', *Journal of Business & Economic Statistics*, **15**, 1-14.

[2] Aslanidis N., D.R. Osborn and M. Sensier (2002): 'Smooth Transition Regression Models in UK Stock Returns', SES/CGBCR, University of Manchester.

[3] V. Barinov, A. Pervozvansky and T. Pervozvanskaya (1999): 'State Debt Policy and Bonds Market Behaviour', EERC working paper, 99-05e.

[4] J. Barkoulas and C. F. Baum (1996), 'Essential Nonparametric Prediction of U.S. Interest Rates', *Boston College Working Papers in Economics*, 313.

[5] W. S. Cleveland and C. L. Loader (1996a): 'Smoothing by Local Regression: Principles and Methods', In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, 10-49. Springer, New York.

[6] W. S. Cleveland and C. L. Loader (1996b): 'Rejoinder to Discussion of "Smoothing by Local Regression: Principles and Methods"'. In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, 113-120. Springer, New York.

[7] F. X. Diebold and J.A. Nason (1990): 'Nonparametric exchange rate prediction', **28**,315-332.

[8] E. F. Fama (1970): 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance*, **25**, 383-423.

[9] S. Hall and G. Urga (2000): 'Testing for time-varying stock market efficiency using Russian stock prices', IRMI, CUBS.

[10] W. Härdle (1990): 'Applied Nonparametric Regression', Cambridge University Press, New York.

[11] T. Hellström (1998): 'A random walk through the stock market', Licentiate thesis, Ůmea university.

[12] D. Hsieh (1991): 'Chaos and Nonlinear Dynamics: Application to Financial Markets', *Journal of Finance*, **46**, 5, 1839-1877.

[13] C. Granger, A.G. Timmermann (2002): 'Efficient Market Hypothesis and Forecasting', C.E.P.R. Discussion Papers, 3593.

[14] B. LeBaron (1988): The Changing Structure of Stock Returns, Working Paper, University of Wisconsin.

[15] G. Leitch and J. E. Tanner (1991): 'Economic forecast evaluation: profits versus the conventional error measures', *American Economic Review*, **81**, 580-590.

[16] A. W. Lo and A. C. MacKinlay (1990): 'Data-snooping biases in tests of financial assets pricing models' , *Review of Financial Studies*, **3**, 431-468.

[17] R. A. Meese and A. K. Rose (1990): 'Nonlinear, Nonparametric, Nonessential Exchange Rate Estimation', *American Economic Review*, **80**, 2, 192-196.

[18] R. A. Meese and A. K. Rose (1991): 'An Empirical Assessment of Non-linearity in Models of Exchange Rate Determination', *Review of Economic Studies*, **58** , 603-619.

[19] B. Mizrach (1992): 'Multivariate nearest-neighbor forecasts of EMS exchange rates', *Journal Of Applied Econometrics*, **7**, 151-163.

[20] M. H. Pesaran and A. Timmermann (1992): 'A Simple Nonparametric Test of Predictive Performance', *Journal of Business and Economic Statistics,* **10**, 561-565.

[21] M. H. Pesaran and A. Timmermann (1995): 'Predictability of Stock Returns: Robustness and Economic Significance', *Journal of Finance,* **50**, 1201-1228.

[22] S. Ramaswamy (1998):'One-step prediction of financial time series models', Bank for international settlements, Monetary and Economic Department, Basle, Switzerland, working paper 57.

[23] M. Rockinger and G. Urga (2000): 'Evolution of Stock Markets in Central and Eastern Europe', *Journal of Comparative Economics,* **28**, 456-472.

[24] S. Satchell and A. Timmermann (1995): 'An assessment of the economic value of non-linear foreign exchange rate forecasts', *Journal of Forecasting*, **14**, 477-497.