Nonparametric Retrospection and Monitoring of Predictability of Financial Returns

Stanislav Anatolyev

Nonparametric retrospection and monitoring of predictability of financial returns

by

Stanislav Anatolyev[*]

New Economic School, Moscow

Abstract

We develop and evaluate sequential testing tools for a class of nonparametric tests for predictability of financial returns that includes, in particular, the directional accuracy and excess profitability tests. We consider both the retrospective context where a researcher wants to track predictability over time in a historical sample, and the monitoring context where a researcher conducts testing as new observations arrive. Throughout, we elaborate on both two-sided and one-sided testing, focusing on linear monitoring boundaries that are continuations of horizontal lines corresponding to retrospective critical values. We illustrate our methodology by testing for directional and mean predictability of returns in a dozen of young stock markets in Eastern Europe.

1   Introduction

Economists have been getting more and more concerned with possible structural instabilities in economic relationships which may invalidate conclusions obtained using conventional econometric tools. More than a decade ago, econometricians revived old CUSUM-type fluctuation tests that allowed one to track structural shifts in model parameters in time in order to detect deviation from constancy (e.g., Ploberger, Krämer and Kontrus, 1989). More recently, there has been a new burst of interest to developing tools of sequential testing for practitioners who make decisions in real time. This was started by Chu, Stinchcombe, and White (1996) and Chu, Hornik, and Kuan (1995), and continued in Leisch, Hornik, and Kuan (2000), Altissimo and Corradi (2003), Zeileis, Leisch, Kleiber, and Hornik (2005), Inoue and Rossi (2005), and Andreou and Ghysels (2006), among others. This resulted in a number of sequential tests designed both for static and for dynamic models, both for conditional means and for conditional variances. Most of this work is targeted towards parametric models.

In this paper, we develop and evaluate sequential testing tools for a certain class of nonparametric tests for predictability of financial returns. This class is quite large and allows testing for hypotheses of non-predictability of various features of a series of interest. Two representatives of this class are the directional accuracy test of Pesaran and Timmermann (1992) and the excess profitability test of Anatolyev and Gerko (2005). Testing for stability of predictability is important; see the discussion in Pesaran and Timmermann (2004) where it is shown that ignoring structural instability may have serious consequences for the quality of directional forecasting.

We consider both retrospective tests where a researcher wants to track predictability over time in a historical sample, and monitoring tests where a researcher conducts testing as new observations arrive. It is worth noting that the literature does not usually consider these tasks together; considering both is the first novelty introduced in this paper. Underlying it is a scenario that a researcher after having carried out a retrospective test goes on to the monitoring stage. Moreover, the retrospective boundaries (horizontal lines corresponding to retrospective critical values) continuously translate into the monotonically growing monitoring boundaries. The continuity of the boundaries is an appealing property as the first observation of the monitoring period should not affect dramatically the inference about the null. Our second novelty is that we develop both two-sided and one-sided testing, with the emphasis put on the latter as more appropriate in the context of testing for predictability (cf. Inoue and Rossi, 2005). We focus on the use of the supremum functional over empirical

processes to construct test statistics which is most widely used functional in the rest of the literature.

In the monitoring context, a widely discussed issue is the shape of monitoring boundaries. Some authors follow Chu, Stinchcombe, and White (1996) who suggest complicated boundaries which however lead to an analytic form of critical values. Zeileis, Leisch, Kleiber, and Hornik (2005) proposed more intuitively appealing linear boundaries which tend to distribute the size throughout the monitoring period more evenly. In Monte–Carlo exercises reported in Zeileis, Leisch, Kleiber, and Hornik (2005) and Andreou and Ghysels (2006) the linear boundaries performed well. We concentrate on such linear boundaries for two reasons, one being the mentioned intuitive appeal, and the other being that they also lead to analytic critical values that can be obtained even more easily.

Note that in most of the work on sequential stability testing, the emphasis is usually put on testing for stability rather than testing whether a particular hypothesis holds throughout the sample. In this sense, the closest to the present work is the paper by Inoue and Rossi (2005) who also sequentially track deviations of some parameter combinations from a hypothesized value. The main consequence is that the asymptotic analog of emerging empirical processes is the Wiener process (and functions thereof) rather than the Brownian Bridge. Inoue and Rossi (2005), however, do not consider one-sided testing, and their framework is, as mentioned above, parametric, albeit nonlinear. Finally, their monitoring boundaries are inherited from Chu, Stinchcombe, and White (1996) which means that the ability to detect structural changes is skewed from late changes to early changes.

We illustrate our methodology by testing for directional and mean predictability of returns in a dozen of young stock markets in Eastern Europe. Such markets are an ideal polygon for applying predictability tests as it is documented using other econometric tools that the pattern of predictability there is changing (e.g., Rockinger and Urga, 2000).

The paper is organized as follows. In Section 2 we review the class of one-shot tests for predictability and its special cases. Sequential tests are developed in Section 3. In Section 4, simulation evidence is discussed, while the empirical application is presented in Section 5. All proofs are collected in the Appendix. Throughout, $\lfloor a \rfloor$ denotes taking an integer part of a, and $\Rightarrow$ denotes weak uniform convergence in the space of cadlag functions.

3

2 One-shot predictability tests

Let y_t represent some economic variable, and x_t be a continuously distributed forecast of y_t that depends only on the data from $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$, or, more generally, from the extended information set $\mathcal{I}_{t-1} \supset \{y_{t-1}, y_{t-2}, \cdots\}$ which may include other historical variables. We are interested in testing the null hypothesis

$$H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const},$$

where $g(u)$ is a given stationary function. The predictability test is based on the contrast

$$A^{g,h} - B^{g,h} \equiv \frac{1}{T}\sum_t h(x_t)g(y_t) - \left(\frac{1}{T}\sum_t h(x_t)\right)\left(\frac{1}{T}\sum_t g(y_t)\right), \qquad (2.1)$$

where the function $h(u)$ is chosen by the researcher. A popular choice is $h(u) = \text{sign}(u)$, in which case setting $g(u) = \text{sign}(u)$ leads to the directional accuracy (DA) test for conditional sign independence of Pesaran and Timmermann (1992), while setting $g(u) = u$ leads to the excess profitability (EP) test for conditional mean independence of Anatolyev and Gerko (2005). The DA test is routinely used as a predictive-failure test in constructing forecasting models, or for evaluating the quality of predictors; see, for example, Pesaran and Timmermann (1995), Franses and van Dijk (2000), and Qi and Wu (2003). When y_t is a logarithmic return on some financial asset or index, the EP statistic can be interpreted as a normalized return of the position implied by a simple trading strategy that issues a buy signal if a forecast of next period return is positive and a sell signal otherwise, over a certain benchmark (see Anatolyev and Gerko, 2005 for details). These two example of special interest will be intensively tackled throughout, although we develop testing algorithms for the general framework.

Let us impose

Assumption 1
(i) The series y_t and its forecast x_t are continuously distributed, strictly stationary, and strongly mixing with mixing coefficients $\alpha(j)$ satisfying $\sum_{j=1}^{\infty} \alpha(j)^{1-1/\nu} < \infty$ for some $\nu > 1$.
(ii) The forecast x_t is $\mathcal{I}_{t-1}$-measurable.
(iii) The functions $g(u)$ and $h(u)$ are measurable, and $E\left[|g(y_t)|^{2\nu q}\right]$ and $E\left[|h(x_t)|^{2\nu p}\right]$ exist and are finite for ν from (i), and for some q and p such that $q^{-1} + p^{-1} = 1$.

The moment condition in assumption 1(iii) is sufficient, but not necessary. With a choice of bounded $h(u)$, as is the case for the DA and EP tests, it is possible to set $p = \infty$ and $q = 1$, so that the moment condition on $g(y_t)$ is quite mild.

Let us introduce the following notation for future use:

$$
\begin{aligned}
M_g &= E\left[g(y_t)\right], & V_g &= \mathrm{var}\left[g(y_t)\right], \\
M_h &= E\left[h(x_t)\right], & V_h &= \mathrm{var}\left[h(x_t)\right],
\end{aligned}
$$

and

$$
m_y = E\left[\mathrm{sign}(y_t)\right], \quad m_x = E\left[\mathrm{sign}(x_t)\right], \quad V_y = \mathrm{var}\left[y_t\right].
$$

We will base our tests on the following result which we will generalize to the context of sequential testing in the next Section.

Lemma 1 *Suppose $h(u)$ and $g(u)$ satisfy the regularity conditions specified in Assumption 1. Consider the contrast (2.1). Under $H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \mathrm{const}$,*

$$
\sqrt{T}\left(A^{g,h} - B^{g,h}\right) \xrightarrow{d} \mathrm{N}(0, V^{g,h})
$$

as $T \to \infty$, where

$$
V^{g,h} = V_h V_g + C_1 - 2M_h C_2,
$$

where $C_1 = \mathrm{cov}\left[h(x_t)^2, g(y_t)^2\right]$ and $C_2 = \mathrm{cov}\left[h(x_t), g(y_t)^2\right]$.

Specialization of Theorem 1 to the two special cases of DA and EP tests yields

Corollary 1

(i) Under the null of conditional sign independence, i.e. $H_0^{DA} : E\left[\mathrm{sign}(y_t)|\mathcal{I}_{t-1}\right] = \mathrm{const}$,

$$
\sqrt{T}\left(A^{DA} - B^{DA}\right) \xrightarrow{d} \mathrm{N}(0, V^{DA})
$$

as $T \to \infty$, where

$$
V^{DA} = \left(1 - m_x^2\right)\left(1 - m_y^2\right).
$$

(ii) Under the null of conditional mean independence, i.e. $H_0^{EP} : E\left[y_t|\mathcal{I}_{t-1}\right] = \mathrm{const}$,

$$
\sqrt{T}\left(A^{EP} - B^{EP}\right) \xrightarrow{d} \mathrm{N}(0, V^{EP})
$$

as $T \to \infty$, where

$$
V^{EP} = \left(1 - m_x^2\right) V_y - 2m_x \,\mathrm{cov}\left[\mathrm{sign}(x_t), y_t^2\right].
$$

To construct the test statistic, the contrast (2.1) may be pivotized using

$$\hat{V}^{g,h} = \hat{V}_h \hat{V}_g + \hat{C}_1 - 2\hat{M}_h \hat{C}_2,$$

where $\hat{M}_h$, $\hat{V}_h$, $\hat{V}_g$, $\hat{C}_1$ and $\hat{C}_2$ are empirical analogs of corresponding population quantities. For example, for the DA and EP tests,

$$\hat{V}^{DA} = \left(1 - \hat{m}_x^2\right)\left(1 - \hat{m}_y^2\right),$$

$$\hat{V}^{EP} = \left(1 - \hat{m}_x^2\right)\hat{V}_y - 2\hat{m}_x\hat{C},$$

where

$$m_y = \frac{1}{T}\sum_t \text{sign}(y_t), \quad m_x = \frac{1}{T}\sum_t \text{sign}(x_t),$$

$$\hat{V}_y = \frac{1}{T}\sum_t y_t^2 - \left(\frac{1}{T}\sum_t y_t\right)^2,$$

$$\hat{C} = \frac{1}{T}\sum_t \left(\text{sign}(x_t) - \hat{m}_x\right)y_t^2.$$

3 Sequential tests

3.1 Sequential testing and boundaries

In the sequential context, the null hypothesis of interest is the conditional independence of $g(y_t)$ throughout the entire period, i.e. that

$$H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const for all } t. \tag{3.2}$$

Note that we do not require that the const in (3.2) be the same across time; all we want to test is that $g(y_t)$ cannot be predicted by information at $t-1$. Thus, the emerging tests may not be able to detect deviations of the risk premium from a constant value.

Let us continue denoting the size of the historical sample by T. Then, if we do retrospective testing of H_0^g on the historical sample, t in (3.2) runs from 1 to T. If we monitor H_0^g further, t in (3.2) runs from $T+1$ to infinity. We choose the boundaries to be linear for both retrospection and monitoring periods: horizontal lines corresponding to retrospective critical values which continuously translate into linear monitoring boundaries going upward (see Fig.1).

The underlying scenario is the following: a researcher has a historical sample in hands and carries out a retrospective test; then he/she goes on to the monitoring stage as new
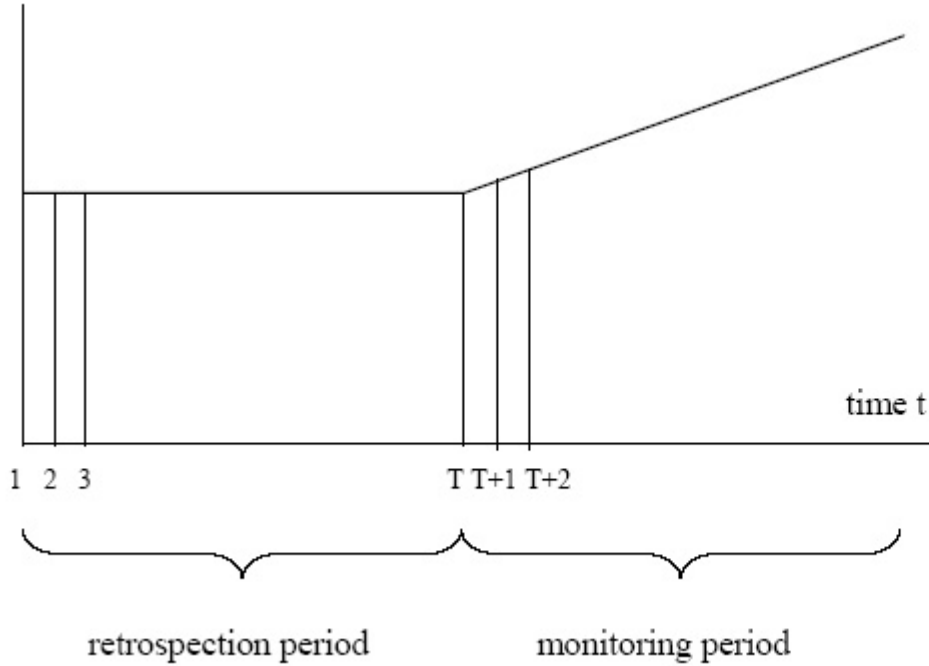
Figure 1: Sequential testing: periods and boundaries

observations begin to arrive. The continuity of the boundaries makes sense as first several observations in the monitoring period should not affect dramatically the inference about the null. With linear boundaries, this continuity is possible to impose provided that the test sizes are equal in the retrospective and monitoring stages. Technically, this happens due to the property

$$\Pr\left\{\sup_{r\geq 1}\left(w\left(r\right)-\lambda r\right)\geq 0\right\}=\Pr\left\{\sup_{r\geq 1}\frac{w\left(r\right)}{r}\geq\lambda\right\}=\Pr\left\{\sup_{0<r\leq 1}w\left(r\right)\geq\lambda\right\},$$

and to a similar property for $\left|w\left(r\right)\right|,$ where $\lambda>0$ is a constant, and $w\left(r\right)$ is a univariate standard Wiener process on $[0,+\infty)$, a limiting process for the sequential test statistic to be developed below.

3.2   Asymptotics for partial contrasts

For a generic series $a_t,$ $t=1,2,\cdots,T,T+1,\cdots$, let us introduce the notation

$$\bar{a}_\tau=\frac{1}{\lfloor T\tau\rfloor}\sum_{t=1}^{\lfloor T\tau\rfloor}a_t,$$

where $\tau\geq 0$. When a_t is a product of several series, $a_t=b_tc_t$, say, then we write $\bar{a}_\tau$ also as $\overline{bc}_\tau$.

7

Kuan and Chen (1994) discovered that fluctuation tests are better sized in finite samples when test statistics are pivotized using data in the expanding window, thus we use such estimates throughout. To this end, let us denote by $\hat{V}^{g,h}_\tau$ the value of $\hat{V}^{g,h}$ computed using the data from 1 to $\lfloor T\tau \rfloor$:

$$\hat{V}^{g,h}_\tau = \left(\overline{h^2}_\tau - \bar{h}^2_\tau \right) \left(\overline{g^2}_\tau - \bar{g}^2_\tau \right) + \overline{h^2 g^2}_\tau - \overline{h^2}_\tau \overline{g^2}_\tau - 2\bar{h}_\tau \left(\overline{hg^2}_\tau - \bar{h}_\tau \overline{g^2}_\tau \right).$$

In particular,

$$\hat{V}^{DA}_\tau = \left(1 - \overline{\mathrm{sign}(x)}^2_\tau \right) \left(1 - \overline{\mathrm{sign}(y)}^2_\tau \right)$$

and

$$\hat{V}^{EP}_\tau = \left(1 - \overline{\mathrm{sign}(x)}^2_\tau \right) \left(\overline{y^2}_\tau - \bar{y}^2_\tau \right) - 2\overline{\mathrm{sign}(x)}_\tau \left(\overline{\mathrm{sign}(x)y^2}_\tau - \overline{\mathrm{sign}(x)}_\tau \overline{y^2}_\tau \right).$$

The empirical process for sequential tests corresponding to the one-shot test based on $g(u)$ and $h(u)$, is

$$P_\tau = \tau \sqrt{\frac{T}{\hat{V}^{g,h}_\tau}} \left(\overline{gh}_\tau - \bar{g}_\tau \bar{h}_\tau \right).$$

Because the usual time t is related to τ by $t = \lceil \tau T \rceil$, we have

$$P_\tau = \frac{t}{\sqrt{T \hat{V}^{g,h}_\tau}} \left(\overline{gh}_\tau - \bar{g}_\tau \bar{h}_\tau \right). \tag{3.3}$$

Theorem 1 *Suppose the null hypothesis*

$$H^g_0 : E\left[g(y_t) | \mathcal{I}_{t-1} \right] = \mathrm{const} \text{ for all } t$$

holds, and $h(u)$ and $g(u)$ satisfy the regularity conditions specified in Assumption 1. Then we have that as $T \to \infty$,

$$P_\tau \Rightarrow w(\tau),$$

where $w(r)$ is a univariate standard Wiener process on $[0, +\infty)$.

3.3 Retrospective tests

Anatolyev and Gerko (2005) observe the evolution of the level of mean predictability at the American stock market throughout the last half of the 20th century by computing the EP statistic from data in a moving window. Although the results obtained are remarkable, the comparison of the maximum level with the conventional critical values for one-shot tests constitutes, strictly speaking, an invalid testing procedure. In this subsection we derive a formal recursive estimates procedure so that the overall size of this test is controlled.

Using the supremum functional, we obtain the asymptotic size α one-sided sup-RE test

$$\text{Reject if } \max_{t=2,\ldots,T} P_\tau \geq q_\alpha^{(1)},$$

and the asymptotic size α two-sided sup-RE test

$$\text{Reject if } \max_{t=2,\ldots,T} |P_\tau| \geq q_\alpha^{(2)},$$

where $q_\alpha^{(j)}$ is a critical value for the j-sided sup-RE test with significance level α.

It is widely known (e.g., Karatzas and Shreve (1988, problem 8.2)) that for $\lambda > 0$,

$$\Pr\left\{ \sup_{0\leq r\leq 1} w\left(r\right) \geq \lambda \right\} = 2\left(1 - \Phi\left(\lambda\right)\right),$$

where $\Phi\left(\circ\right)$ is the CDF of the standard normal distribution. Hence, $q_\alpha^{(1)}$ can be easily found as a solution to the equation

$$\Phi\left(q_\alpha^{(1)}\right) = 1 - \frac{\alpha}{2},$$

so the α-quantiles for the one-sided sup-RE test are equal to conventionally used α-quantiles for two-sided one-shot tests. Next, from Erdös and Kac (1946),

$$\Psi(\lambda) \equiv \Pr\left\{ \sup_{0\leq r\leq 1} |w\left(r\right)| \leq \lambda \right\} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{\pi^2 \left(2k+1\right)^2}{8\lambda^2} \right).$$

Hence, $q_\alpha^{(2)}$ can be easily found as a solution to the equation

$$\Psi\left(q_\alpha^{(2)}\right) = 1 - \alpha.$$

In the following table we document the critical values for popular levels of significance. Note that for small α, as typically is the case, $q_\alpha^{(1)} \approx q_{2\alpha}^{(2)}$. This reflects a low probability of the standard Wiener process' hitting both $-\lambda$ and $+\lambda$ when λ is large enough (i.e. when α is small enough).

One-sided			Two-sided		
10%	5%	1%	10%	5%	1%
1.645	1.960	2.576	1.960	2.241	2.807

3.4 Monitoring tests

Using the supremum functional, we obtain the asymptotic size α one-sided sup-RE test

$$\text{Reject if } \max_{t=T+1,T+2,\ldots} \left(P_\tau - b_\alpha^{(1)}\left(t\right)\right) \geq 0,$$

9

and the asymptotic size α two-sided sup-RE test

$$\text{Reject if } \max_{t=T+1,T+2,\dots} \left(|P_\tau| - b_\alpha^{(2)}(t) \right) \geq 0,$$

where $b_\alpha^{(1)}(t)$ and $b_\alpha^{(2)}(t)$ are upper boundaries.

We base our recursive monitoring one-sided tests on the boundaries of the type

$$b_\alpha^{(j)}(t) = \lambda_\alpha^{(j)} \frac{t}{T},$$

which tend to distribute the size throughout the monitoring period evenly when the underlying process has growing variance (Zeileis, Leisch, Kleiber, and Hornik, 2005). From the results of Robbins and Siegmund (1970, example 1) we obtain that

$$\lim_{T \to \infty} \Pr \left\{ \max_{t=T+1,T+2,\dots} \left(P_\tau - \lambda \frac{t}{T} \right) \geq 0 \right\} = 2\left(1 - \Phi(\lambda)\right).$$

Hence, $\lambda_\alpha^{(1)}$ can be found as a solution of the equation

$$\Phi\left(\lambda_\alpha^{(1)}\right) = 1 - \frac{\alpha}{2}.$$

For the two-sided test,

$$b_\alpha^{(2)}(t) = \lambda_\alpha^{(2)} \frac{t}{T},$$

and $\lambda_\alpha^{(2)}$ solves

$$\Psi\left(\lambda_\alpha^{(2)}\right) = 1 - \alpha.$$

Note that the same equations are used by the retrospective RE tests in the previous subsection. Hence, we can consult the same tables to get values of $\lambda_\alpha^{(j)}$. Note also that $b_\alpha^{(j)}(T) = q_\alpha^{(j)}$, which provides continuity of the boundaries.

4 Simulation evidence

In this Section, we use Monte–Carlo simulations to check on actual sizes of the developed tests in finite samples, and to study their power properties. Throughout, for the sake of simplicity, we set the predictor x_t to be the total return from two previous periods, i.e. $x_t = y_{t-2} + y_{t-1}$. This is an easy way to construct a predictor, and it is always available. The forecasting power of this predictor, however, may not be large, so that in practical situations the power properties of the tests reported below may be even higher when other predictors are used such as coming from estimation of parametric or nonparametric autoregressions. In

what follows, we report actual rejection frequencies of one-sided sequential analogs of the DA and EP tests corresponding to the nominal size of 5%. The simulation results are collected in Table 1 for retrospective tests and in Table 2 for monitoring tests.

For retrospective tests, we generate $R = 10,000$ times the series of y_t's of length T, where $T = 50, 100, 200$, or 500 when the size is checked, and $T = 200$ or 500 when the power is investigated, according the data generating processes (DGPs) described below. The parameters in DGPs A, B and C are calibrated using the weekly S&P500 index during the period from 1954 to 1973 as in Anatolyev and Gerko (2005); other DGPs are created artificially using these parameters as benchmarks.

The first two DGPs, A and B, are used to check on size (except for testing for directional accuracy in DGP B: Christoffersen and Diebold (2006) have recently shown that conditional heteroskedasticity induces sign predictability even when there is no mean predictability).

$$\text{DGP A} \quad y_t \;=\; 0.001526 + \varepsilon_t, \; \varepsilon_t \sim iid \; \text{N}(0, 0.000025), \tag{4.4}$$

$$\text{DGP B} \quad y_t \;=\; 0.002483 + \varepsilon_t, \; \varepsilon_t = \sigma_t \eta_t, \; \eta_t \sim iid \; \text{N}(0, \sigma_t^2), \tag{4.5}$$

$$\sigma_t^2 \;=\; 0.0000223 + 0.1773 \cdot \varepsilon_{t-1}^2 + 0.7397 \cdot \sigma_{t-1}^2$$

The following DGPs C, D and E in several variations are used to investigate power. While in DGP C there is the same non-zero amount of predictability throughout the sample period, in DGPs D the predictability is observed only during subperiods in the middle or towards the beginning or the end of the sample. Finally, in DGPs E there is a continuous transition from no predictability to higher and even higher predictability, or vice versa. Extra factors 3 and 2 attached to the autoregressive parameter serve to equalize the "amount" of predictability across the DGPs.

$$\text{DGP C} \quad y_t \;=\; 0.001526 + 0.1256 \cdot y_{t-1} + \varepsilon_t, \; \varepsilon_t \sim iid \; \text{N}(0, 0.00025), \tag{4.6}$$

$$\text{DGP D}_k \quad y_t \;=\; 0.001526 + 3 \cdot 0.1256 \cdot \mathbb{I}_{\{t \in T_k\}} \cdot y_{t-1} + \varepsilon_t, \tag{4.7}$$

$$\varepsilon_t \;\sim\; iid \; \text{N}(0, 0.00025), \; k = 1, 2, 3,$$

$$\text{DGP E}_k \quad y_t \;=\; 0.001526 + 2 \cdot 0.1256 \cdot \frac{t \mathbb{I}_{\{k=1\}} + (T-t)\,\mathbb{I}_{\{k=2\}}}{T} \cdot y_{t-1} + \varepsilon_t, \tag{4.8}$$

$$\varepsilon_t \;\sim\; iid \; \text{N}(0, 0.00025), \; k = 1, 2,$$

where $\mathbb{I}_{\{\circ\}}$ is an indicator function, and T_k contains time periods from the k's third of the sample. That is, T_1 contains observations from the first third of the sample, T_2 – those from the second third of the sample, and T_3 – those from the last third of the sample. Hence, in DGPs D_1 through D_3 the predictability is observed during one of the three periods, and is

not observed during the other two. In contrast, the predictability is continuously escalating as time passes in DGP E_1, but is vanishing as time passes in DGP E_2.

First let us look at panel A of Table 1 which contains actual sizes when data are serially independent, without any predictability. One can immediately see that the sequential tests are very well-sized, especially for larger samples. The DA tests tend to be a little undersized, while the EP tests tend to be instead a little oversized. Next consider panel B corresponding to the GARCH process which is mean unpredictable. The actual sizes for EP tests are a little smaller compensating for some oversizedness but sometimes overshooting the target. More importantly, the DA tests display practically the same sizes as for DGP A, even though under DGP B the series is, in contrast to DGP A, sign predictable. Next, if we compare size distortions across alternatives, those for two-sided tests seem to exceed those for one-sided tests in the case of EP, but the opposite is true in the case of DA.

Let us now turn to power figures in panels C, D and E corresponding to DGPs with predictability of the autoregressive type. Overall, the EP tests are more powerful than the DA tests in detecting such predictability, which is in line with the analysis in Anatolyev and Gerko (2005, section 3). Naturally, the power increases quickly with the sample size. Also, power figures are significantly higher for one-sided tests than for two-sided ones.

For monitoring tests, we verify the test sizes and how these sizes are distributed along the monitoring period. For that purpose, we generate $R = 10,000$ times the series of y_t's of length $4T$, where $T = 50, 100, 200$, or 500, according to the DGPs A or B. Then we read off actual rejection frequencies using boundaries corresponding to the true size of 5%, happened within the periods $[T+1, \tau T]$, where τ equals $\frac{3}{2}$, 2, and 4.

One can see from Table 2 that the monitoring RE tests are very well sized, strictly smaller than 5% (except for DA testing in the case of DGP B where there is sign predictability), with the size exhausting pretty rapidly. For smaller sample sizes underrejection is slightly larger than for larger samples.

5 Application to returns from Eastern European stock markets

In this Section we apply the developed methodology to the analysis of predictability of stock market indexes in a dozen of former communist countries in Eastern Europe. The indexes together with some of their characteristics are listed in Table 3. All indexes are weekly, start on January of the year 1997 (7 series), 1998 (2 series), 1999 (1 series) or 2000 (2 series), and end on January 2005. The data are taken from Bloomberg.

The literature has documented a significant amount of predictability in such markets at the end of 20th century when these markets were very young; see Zalewska-Mitura and Hall (1999) and Rockinger and Urga (2000, 2001)[1]. At the same time, one could observe a movement towards non-predictability in most of them (Rockinger and Urga, 2000). By now several more years have passed, much more data have arrived, and it is interesting to test if that movement indeed has been taking place further.

Our empirical strategy is the following. To illustrate retrospective testing, we run retrospective RE tests over entire samples. To illustrate monitoring testing, we partition the sample into two parts, the earlier period representing a historical subsample, and then run the monitoring test acting in the position of a real time observer during the second period.

Figure 1 presents graphs of evolution of values of recursive estimates (3.3), together with horizontal lines corresponding to critical values for one-sided testing; left panels representing the DA test, the right panels representing the EP test. As can be easily seen, only Ukrainian, Lithuanian and Estonian stock indices exhibit a clear pattern of strong predictability of both types. For the Polish stock index, mean predictability is strongly rejected, but directional predictability is not. The Slovak, Slovenian and Romanian stock indices display marginal rejection at the 5% significance level for at least one of predictability criteria. Finally, the Russian, Czech, Hungarian, Croatian, and Latvian stock indices do not exhibit predictability of either type even at the 10% significance level.

Now we turn to illustrations of the monitoring tests, restricting ourselves only to the excess profitability statistic. Let us put ourselves into a position of a researcher who monitors in real time the stock market starting from two years after the beginning of the sample. Before monitoring, let us suppose that the researcher conducts retrospective tests on the historical subsample of two years. The results are presented in Figure 2.

The documented patterns are quite different across markets, and range from the situation where neither retrospective nor monitoring tests detected mean predictability (Slovenian, Romanian and Latvian stock markets) to the situation where retrospective tests on the historical subsample strongly reject conditional mean independence (Hungarian, Croatian, and Estonian stock markets). Most of series, however, exhibit intermediate patterns. For the Russian, Czech, and Slovak stock markets the retrospective tests either reject marginally

---

[1]In the literature, this predictability is often referred to as (weak form) "efficiency". However, in such markets there are a few market limitations (like transactions costs and short selling constraints) that do not support such interpretation. For discussions, see Pesaran and Timmermann (1995) and Timmermann and Granger (2004).

or does not reject at all, the same tendency continuing for the monitoring tests at the very beginning of the monitoring period. For the Ukrainian and Polish markets, the retrospective tests reject conditional mean independence at the end of the historical period, and so do the monitoring test at the beginning of the monitoring period. In contrast, in the Lithuanian stock market the retrospective tests do not detect predictability in historical subsamples, but the monitoring test indicates that strong mean predictability appeared in the market in 2003.

Overall, the patterns of predictability vary significantly across countries, in agreement with findings in Rockinger and Urga (2001) and Mateus (2004). Even markets in closely connected countries may display completely different patterns of predictability; one example is Russia and Ukraine; another example is the three Baltic states. Few stock markets do not tend to exhibit predictability at all, while for most it proves possible to detect predictability during some periods.

Acknowledgments

References

Altissimo, F. and V. Corradi (2003) Strong rules for detecting the number of breaks in a time series. *Journal of Econometrics* 117, 207–244.

Anatolyev, S. and A. Gerko (2005) A trading approach to testing for predictability. *Journal of Business and Economic Statistics* 23, 455–461.

Andreou, E. and E. Ghysels (2006) Monitoring for Disruptions in Financial Markets. *Journal of Econometrics* forthcoming.

Christoffersen, P.F. and F.X. Diebold (2006) Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* forthcoming.

Chu, C.S.J., Hornik, K., Kuan C.M. (1995) The moving-estimates test for parameter stability. *Econometric Theory* 11, 669–720.

Chu, C.S.J., Stinchcombe, M. and H. White (1996) Monitoring structural change. *Econometrica* 64, 1045–1065.

Erdös, P. and M. Kac (1946) On certain limit theorems of the theory of probability. *Bulletin of American Mathematical Society* 52, 292–302.

Franses, P. and D. van Dijk (2000) *Nonlinear Time Series Models in Empirical Finance.* Cambridge University Press.

Inoue, A. and B. Rossi (2005) Recursive Predictability Tests for Real Time Data. *Journal of Business and Economic Statistics* 23, 336–345.

Karatzas, I. and S.E. Shreve (1988) *Brownian motion and stochastic calculus.* Springer-Verlag: New York.

Kuan, C.M., Chen, M.Y. (1994) Implementing the fluctuation and moving-estimates tests in dynamic econometric models. *Economics Letters* 44, 235–239.

Leisch, F., Hornik, K., Kuan, C.M. (2000) Monitoring structural changes with the generalized fluctuation test. *Econometric Theory* 16, 835–854.

Mateus, T. (2004) The risk and predictability of equity returns of the EU accession countries. *Emerging Markets Review* 5, 241–266.

Pesaran, M.H. and A. Timmermann (1992) A simple nonparametric test of predictive performance. *Journal of Business and Economic Statistics* 10, 561–565.

Pesaran, M.H. and A. Timmermann (1995) Predictability of stock returns: robustness and economic significance. *Journal of Finance* 50, 1201–1228.

Pesaran, M.H. and A. Timmermann (2002) Market timing and return prediction under model instability. *Journal of Empirical Finance* 9, 495–510.

Pesaran, M.H. and A. Timmermann (2004) How costly is it to ignore breaks when forecasting the direction of a time series? *International Journal of Forecasting* 20, 411–425.

Phillips, P.C.B. and S.N. Durlauf (1986) Multiple time series regression with integrated processes. *Review of Economic Studies* 53, 473–495.

Ploberger, W., Krämer, W., Kontrus K. (1989) A new test for structural stability in the linear regression model. *Journal of Econometrics* 40, 307–318.

Qi, M. and Y. Wu (2003) Nonlinear prediction of exchange rates with monetary fundamentals. *Journal of Empirical Finance* 10, 623–640.

Robbins, H. and D. Siegmund (1970) Boundary Crossing Probabilities for the Wiener Process and Sample Sums. *Annals of Mathematical Statistics* 41, 1410–1429.

Rockinger, M. and G. Urga (2000) The Evolution of Stock Markets in Transition Economies. *Journal of Comparative Economics* 28, 456–472.

Rockinger, M. and G. Urga (2001) A Time Varying Parameter Model to Test for Predictability and Integration in the Stock Markets of Transition Economies. *Journal of Business and Economic Statistics* 19, 73–84.

Timmermann, A. and C. Granger (2004) Efficient market hypothesis and forecasting. *International Journal of Forecasting* 20, 15–27.

Zalewska-Mitura, A. and S.G. Hall (1999) Examining the first stages of market performance: a test for evolving market efficiency. *Economics Letters* 64, 1–12.

Zeileis A., F. Leisch, C. Kleiber and K. Hornik (2005) Monitoring structural change in dynamic econometric models, *Journal of Applied Econometrics* 20, 99–121.

A   Appendix

Proof. [of Lemma 1] Follows as a special case of Theorem 1 by setting $\tau_1 = 0$ and $\tau_2 = 1$.
∎

Lemma 2 *Suppose $h(u)$ and $g(u)$ satisfy the regularity conditions specified in Assumption 1. Then under*

$$H_0^g : E\left[g(y_t)|\mathcal{I}_{t-1}\right] = \text{const},$$

as $T \to \infty$, we have

$$\frac{1}{\sqrt{T}} V^{-1/2} \sum_{t=1}^{[\tau T]} \begin{pmatrix} h_t g_t - M_h M_g \\ g_t - M_g \\ h_t - M_h \end{pmatrix} \Rightarrow W(\tau)$$

where $W(\tau)$ is a trivariate standard Brownian motion, and the elements of V are given by

$$V_{11} = \text{var}\left[h(x_t)g(y_t)\right] + 2M_g \sum_{j=1}^{+\infty} \text{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right],$$

$$V_{22} = V_g,$$

$$V_{33} = V_h + 2\sum_{j=1}^{+\infty} \text{cov}\left[h(x_t), h(x_{t+j})\right],$$

$$V_{12} = \text{cov}\left[h(x_t), g(y_t)^2\right] + M_h V_g + M_g \sum_{j=1}^{+\infty} \text{cov}\left[g(y_t), h(x_{t+j})\right],$$

$$V_{13} = M_g V_h + M_g \sum_{j=1}^{+\infty} \text{cov}\left[h(x_t), h(x_{t+j})\right] + \sum_{j=1}^{+\infty} \text{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right],$$

$$V_{23} = \sum_{j=1}^{+\infty} \text{cov}\left[g(y_t), h(x_{t+j})\right].$$

Proof.   The conclusion follows directly from Phillips and Durlauf (1986, corollary 2.2),

with the elements of the long-run covariance V given by

$$
\begin{aligned}
V_{11} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})g(y_{t+j})\right] \\
&= \operatorname{var}\left[h(x_t)g(y_t)\right] + 2M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right], \\
V_{22} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[g(y_t), g(y_{t+j})\right] = V_g, \\
V_{33} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right] = V_h + 2\sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right], \\
V_{12} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), g(y_{t+j})\right] \\
&= \operatorname{cov}\left[h(x_t), g(y_t)^2\right] + M_h V_g + M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right], \\
V_{13} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right] \\
&= M_g V_h + M_g \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t), h(x_{t+j})\right] + \sum_{j=1}^{+\infty} \operatorname{cov}\left[h(x_t)g(y_t), h(x_{t+j})\right], \\
V_{23} &= \sum_{j=-\infty}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right] = \sum_{j=1}^{+\infty} \operatorname{cov}\left[g(y_t), h(x_{t+j})\right],
\end{aligned}
$$

where the law of iterated expectations and the statement of the null hypothesis are intensively used. $\blacksquare$

Proof. [of Theorem 1] Let us denote

$$
\mu = \begin{pmatrix} 1 & -M_h & -M_g \end{pmatrix}'.
$$

From Lemma 2, it follows that

$$
\sqrt{T}\left(\overline{gh}_\tau - \bar{g}_\tau \bar{h}_\tau\right) \Rightarrow \frac{\mu' V^{1/2} W(\tau)}{\tau}.
$$

When pivotized,

$$
P_\tau \Rightarrow \frac{\mu' V^{1/2} W(\tau)}{\sqrt{\mu' V \mu}} \stackrel{d}{=} w(\tau),
$$

because $\hat{V}_\tau^{g,h} \stackrel{p}{\to} V^{g,h} = \mu' V \mu$. $\blacksquare$

18

		Directional accuracy		Excess profitability	
DGP	T	RE_1	RE_2	RE_1	RE_2
A	50	3.2	4.1	6.7	7.7
	100	4.1	4.6	6.1	6.5
	200	4.0	4.9	5.4	6.4
	500	4.4	4.7	5.0	5.5
B	50	3.4	4.5	4.7	6.0
	100	4.0	4.4	4.9	5.9
	200	4.2	4.6	4.7	5.4
	500	5.5	4.6	4.5	4.8
C	200	17.2	11.4	22.6	15.4
	500	34.2	24.3	46.4	35.0
D_1	200	25.5	15.9	43.2	36.3
	500	51.7	37.6	74.4	66.1
D_2	200	23.4	15.2	36.9	27.7
	500	44.5	32.8	66.1	55.0
D_3	200	15.2	10.3	25.7	19.0
	500	30.0	22.1	49.4	39.6
E_1	200	15.2	10.1	22.9	16.6
	500	29.0	20.2	42.7	33.2
E_2	200	19.4	12.6	30.2	23.3
	500	37.9	26.7	55.2	44.7

Table 1. Actual size and power for retrospective tests, from Monte–Carlo simulations.

DGP	T	τ	Directional accuracy		Excess profitability	
			RE_1	RE_2	RE_1	RE_2
A	50	$\frac{3}{2}$	3.0	3.9	3.2	3.7
		2	3.2	4.0	3.4	3.8
		4	3.3	4.1	3.5	3.8
	100	$\frac{3}{2}$	3.5	4.0	3.5	3.8
		2	3.7	4.1	3.6	3.9
		4	3.8	4.2	3.7	3.9
	200	$\frac{3}{2}$	3.8	4.4	3.9	4.3
		2	4.1	4.5	4.1	4.4
		4	4.1	4.5	4.1	4.4
	500	$\frac{3}{2}$	4.2	4.3	4.3	4.6
		2	4.4	4.5	4.4	4.8
		4	4.4	4.5	4.5	4.8
B	50	$\frac{3}{2}$	3.0	4.1	2.6	3.7
		2	3.2	4.3	2.8	3.7
		4	3.2	4.3	2.8	3.8
	100	$\frac{3}{2}$	3.5	4.1	3.1	3.9
		2	3.7	4.2	3.3	4.1
		4	3.8	4.3	3.3	4.2
	200	$\frac{3}{2}$	4.1	4.4	3.6	4.0
		2	4.3	4.5	3.7	4.0
		4	4.4	4.6	3.8	4.1
	500	$\frac{3}{2}$	5.1	4.5	4.1	4.5
		2	5.3	4.7	4.2	4.6
		4	5.4	4.7	4.2	4.6

Table 2. Actual size for monitoring tests, from Monte–Carlo simulations.

Country	Series	Beginning	End	Sample size
Russia	RUX	Jan 1998	Jan 2005	369
Ukraine	PFTS	Jan 1998	Jan 2005	367
Poland	WIG	Jan 1997	Jan 2005	423
Czech Republic	PX50	Jan 1997	Jan 2005	422
Slovakia	SKSM	Jan 1997	Jan 2005	404
Hungary	BUX	Jan 1997	Jan 2005	422
Croatia	CROBEX	Jan 1997	Jan 2005	418
Slovenia	SBI	Jan 1997	Jan 2005	422
Romania	ROL	Jan 1999	Jan 2005	309
Lithuania	VILSE	Jan 2000	Jan 2005	266
Latvia	RIGSE	Jan 2000	Jan 2005	266
Estonia	TALSE	Jan 1997	Jan 2005	423

Table 3. Characteristics of series of stock indexes.
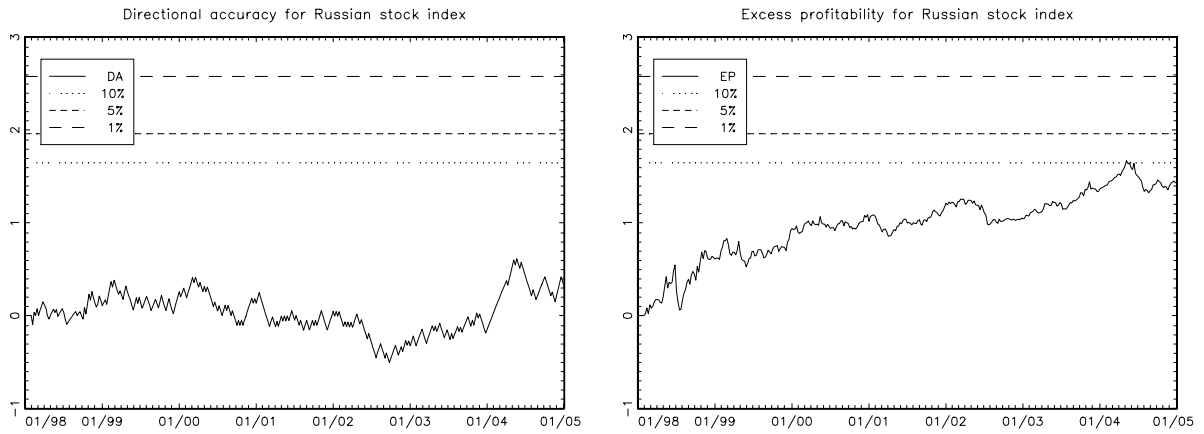
Figure 1(a). Retrospective test for the Russian stock index.
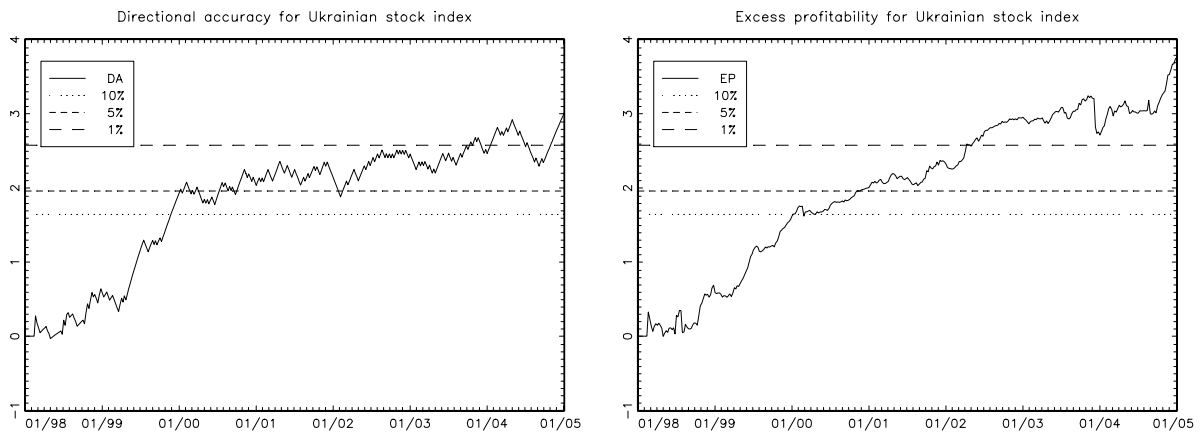
Figure 1(b). Retrospective test for the Ukrainian stock index.
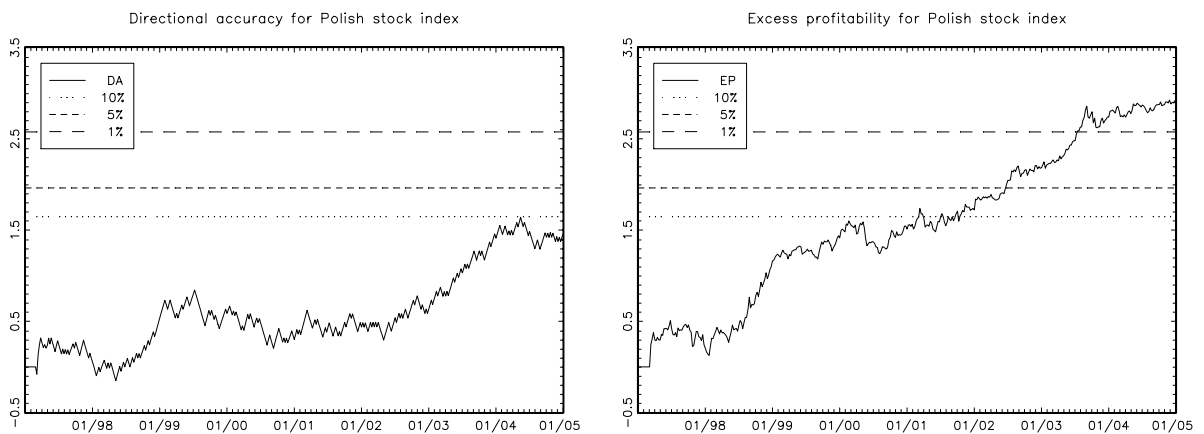
Figure 1(c). Retrospective test for the Polish stock index.
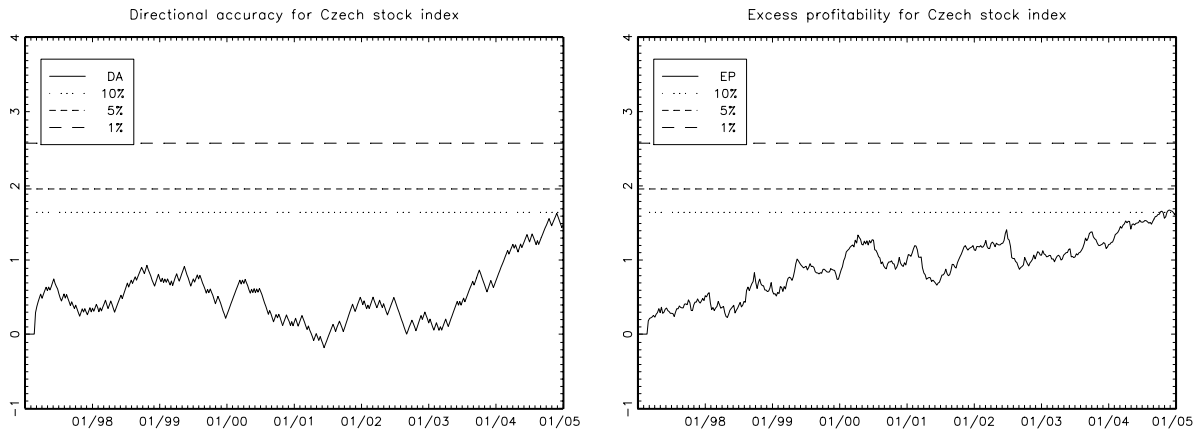
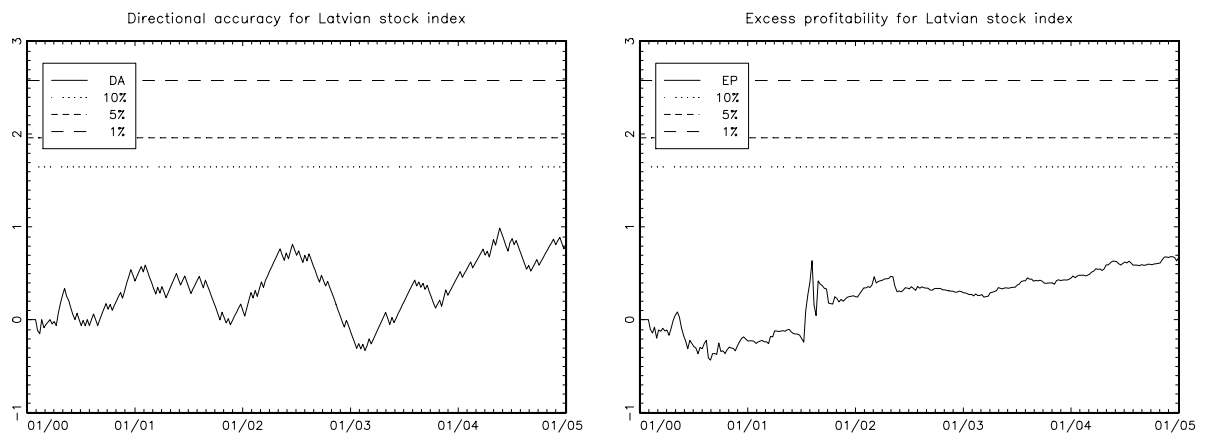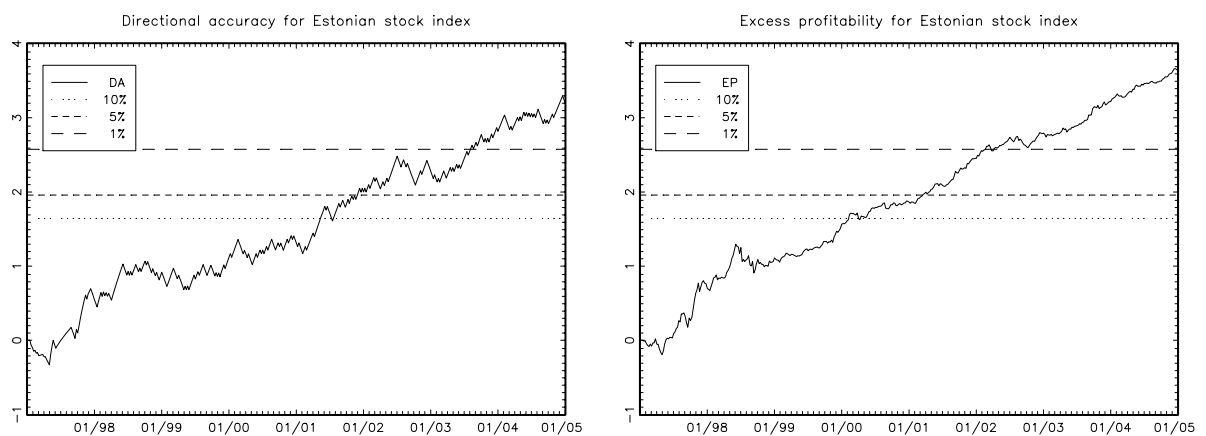Figure 1(d). Retrospective test for the Czech stock index.
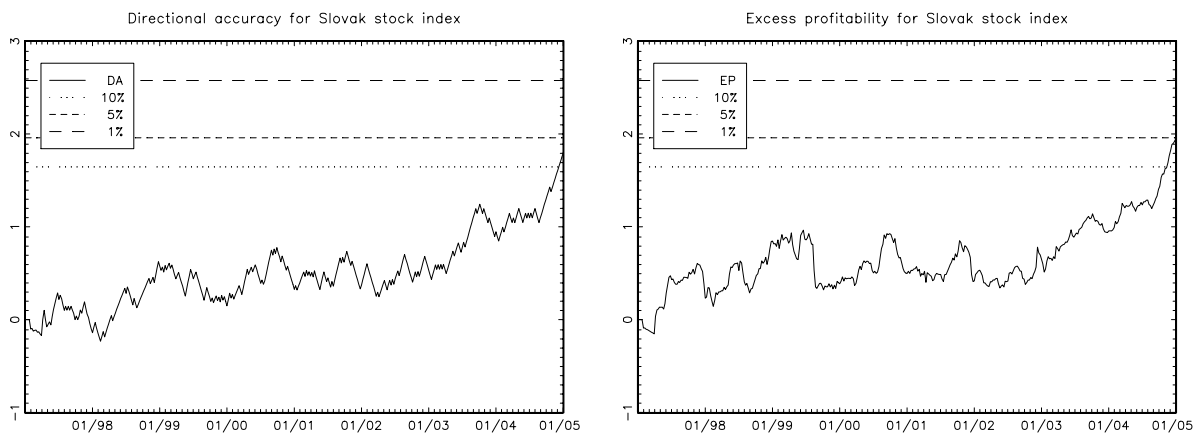
Figure 1(e). Retrospective test for the Slovak stock index.
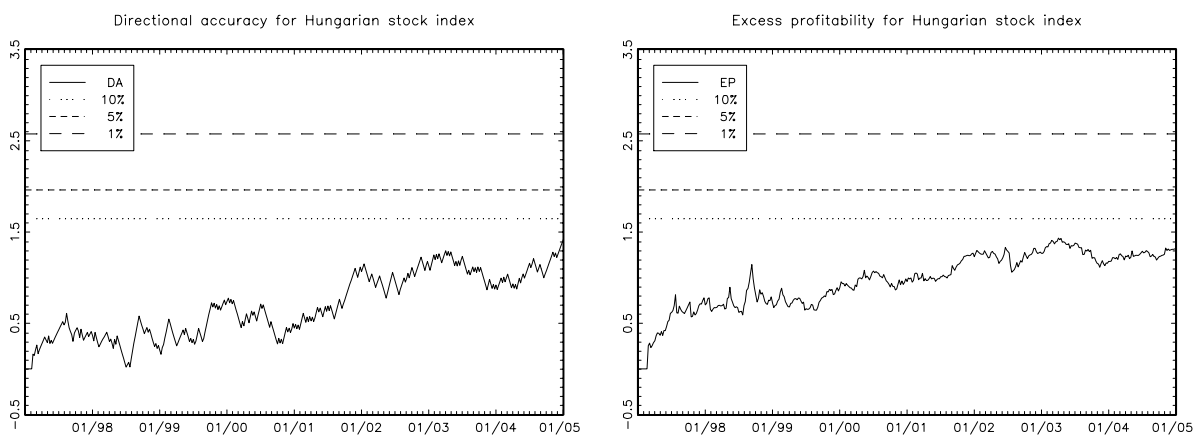
Figure 1(f). Retrospective test for the Hungarian stock index.
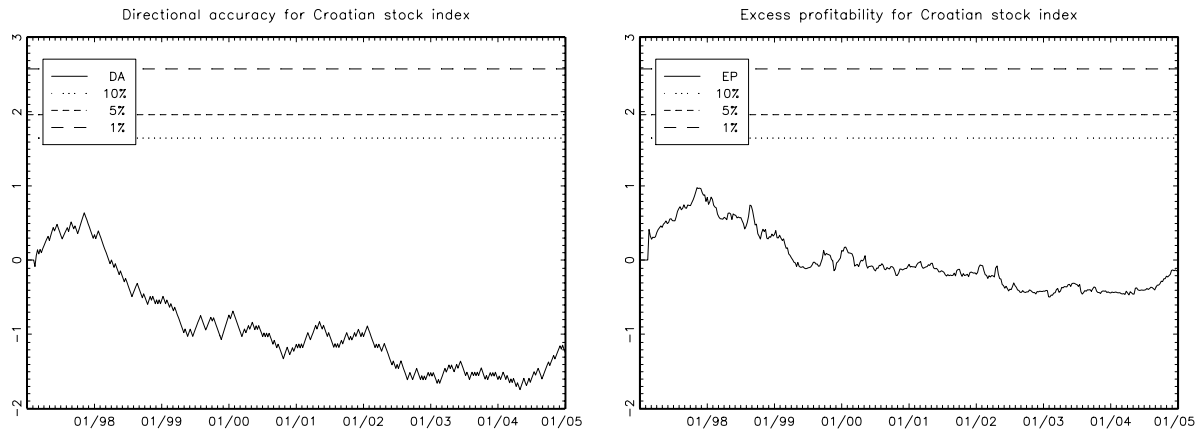
Figure 1(g). Retrospective test for the Croatian stock index.

Figure 1(h). Retrospective test for the Slovenian stock index.
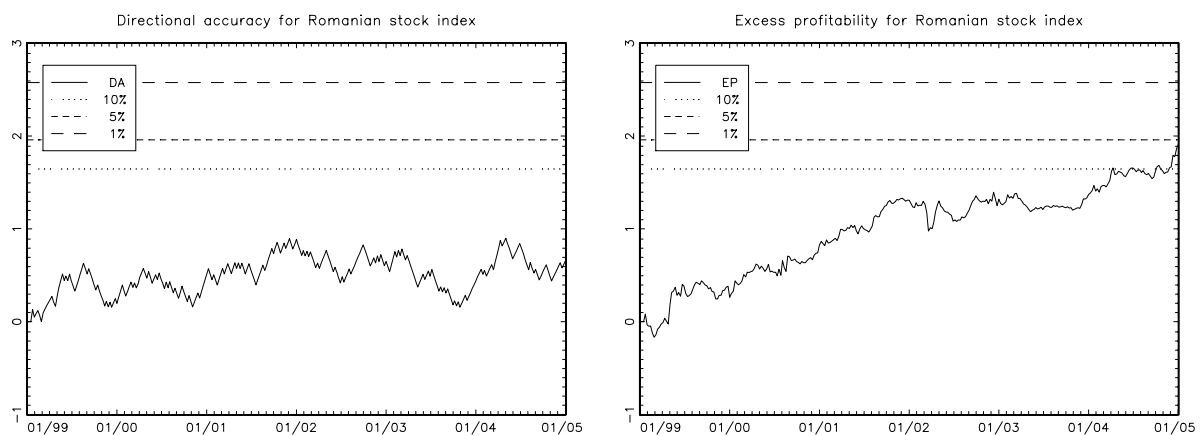
Figure 1(i). Retrospective test for the Romanian stock index.
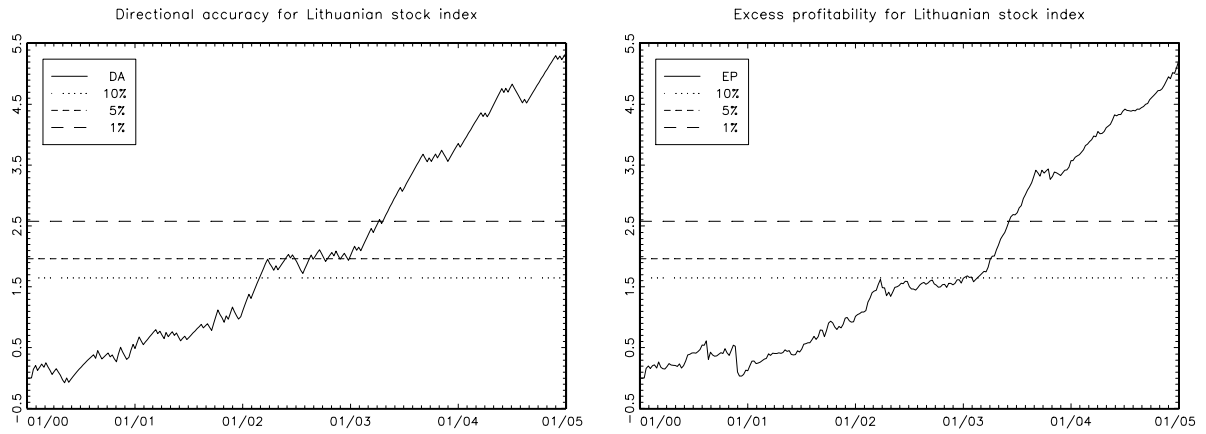
Figure 1(j). Retrospective test for the Lithuanian stock index.

Figure 1(k). Retrospective test for the Latvian stock index.

Figure 1(l). Retrospective test for the Estonian stock index.
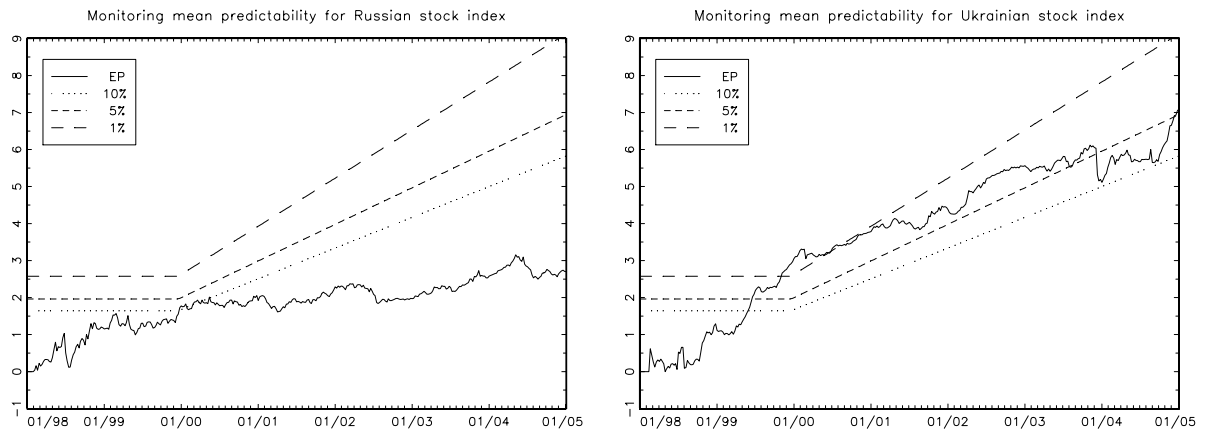
Figure 2(a). Sequential tests for the Russian and Ukrainian stock indexes.
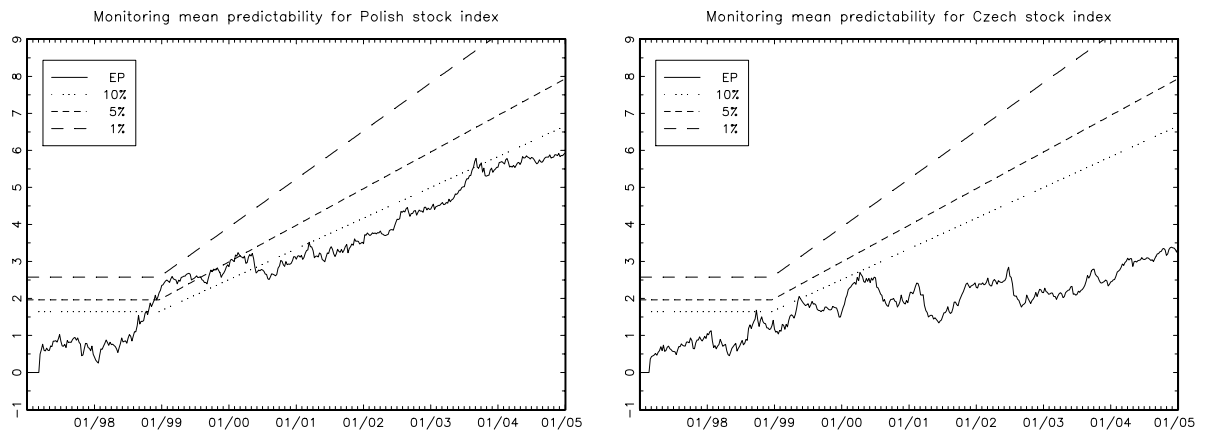
Figure 2(b). Sequential tests for the Polish and Czech stock indexes.
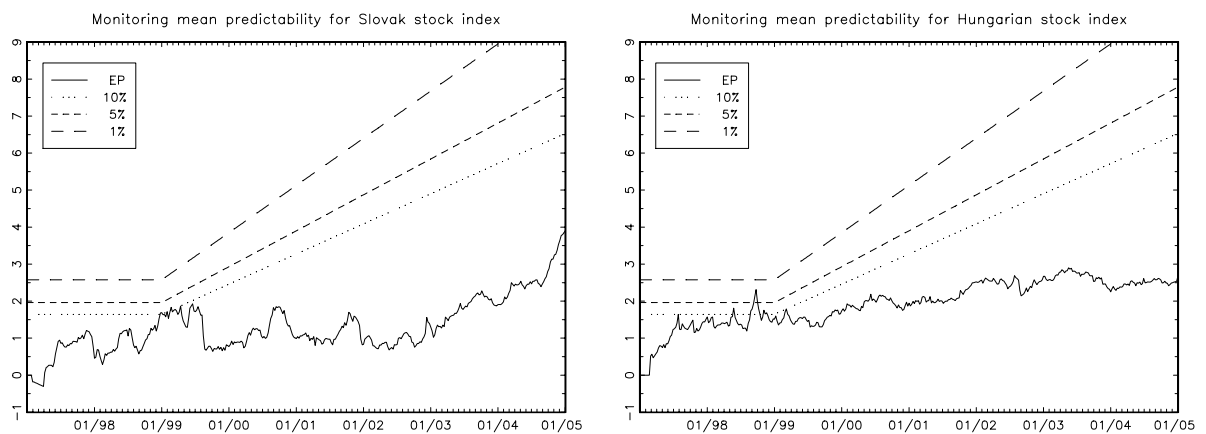
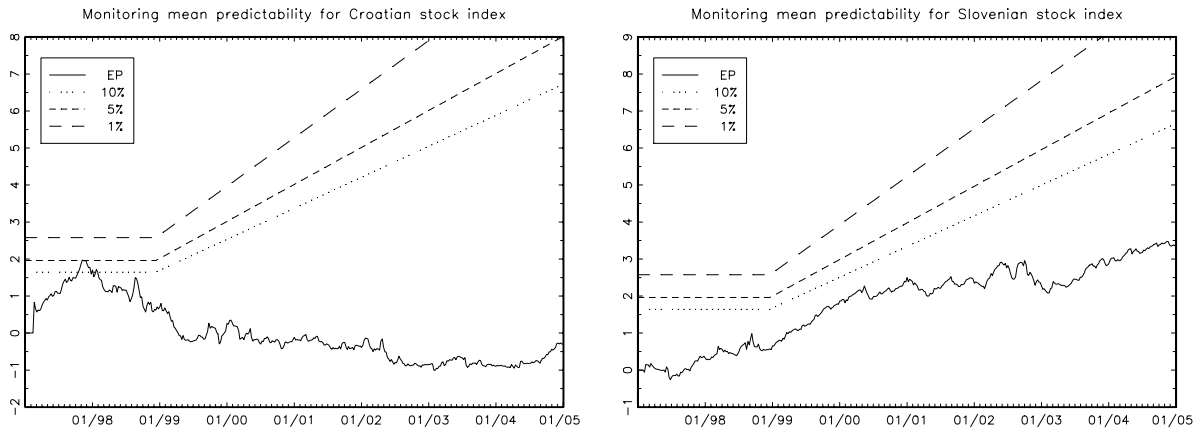Figure 2(c). Sequential tests for the Slovak and Hungarian stock indexes.

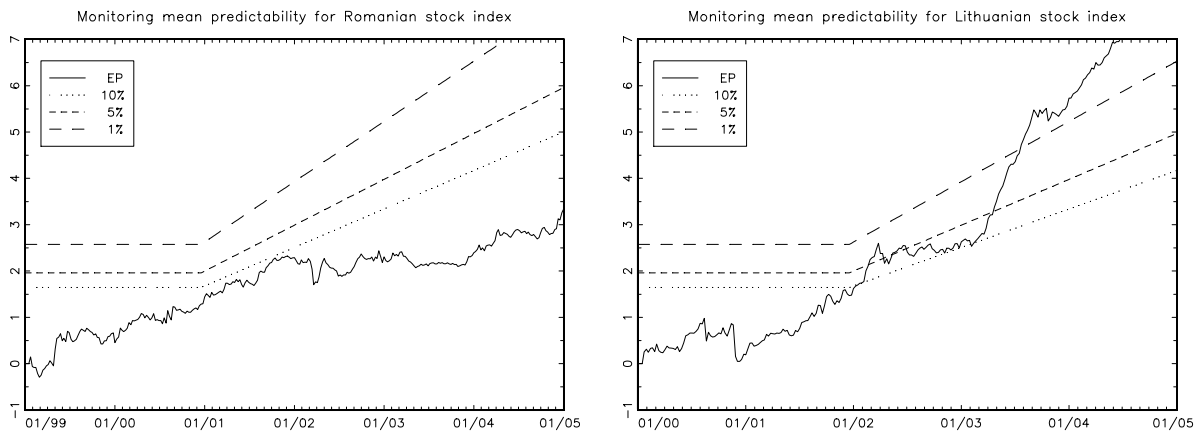Figure 2(d). Sequential tests for the Croatian and Slovenian stock indexes.

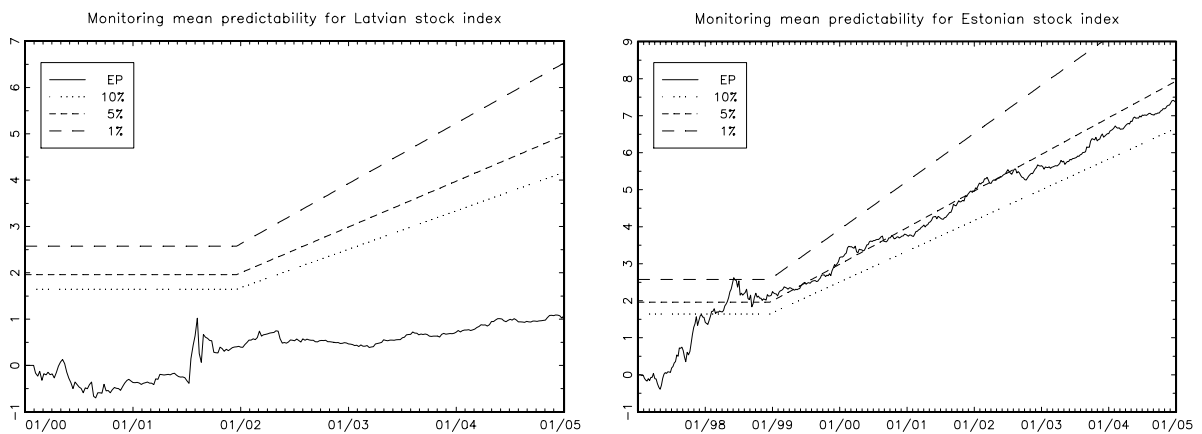Figure 2(e). Sequential tests for the Romanian and Lithuanian stock indexes.

Figure 2(f). Sequential tests for the Latvian and Estonian stock indexes.