Sequential Estimation and Control of Time-Varying Unit Root Processes with an Application to S&P Stock Price

Carlo Grillenzoni

Faculty of Planning, University of Venice, Venice, Italy

Abstract: Stock price series typically behave like random walks, i.e. first-order auto-regressive models whose coefficients (roots) are on the unit circle. This paper investigates time-varying unit roots (TVUR, i.e. roots which wander about unity), and shows that their pattern is related to troughs and peaks of the observed series. Under the assumption of smooth evolution, exponentially weighted least-squares (EWLS) can track roots which wander on the unit circle and so can detect turning points sequentially. This allows to implement effective strategies of investment, which also provide optimization criteria for selecting the tuning coefficients. Extensive application to Standard & Poor index and comparison with other methods show the validity of the method.

Keywords: Control statistics; Financial series; Random walks; Recursive estimators; Sequential detection; Turning points.

Subject Classifications: 62L12; 60G40.

Address correspondence to Carlo Grillenzoni, Facoltá di Pianificazione, Universitá di Venezia, Santa Croce 1957, 30135 Venezia, Italy; E-mail: carlog@iuav.it

1. INTRODUCTION

Recent crisis in financial markets and the consequent instability of stock values, have produced huge capital losses at investors, involving, in certain cases, bankruptcy. There were also important effects on the real side of economy, with the beginning of a downward phase for the business cycle. This storm has increased the need for reliable forecasting models, as well as effective techniques of surveillance and decision. The implementation of timely alarm signals is a fundamental goal both for individual traders and managers of mutual funds, who pursue maximization of fund share values. In these cases, the golden rule for obtaining capital gains is *buy low* and *sell high*, as regard as the price level of equities. The optimal trading policy then coincides with the early detection of turning points in financial time series, i.e. the periods where the phases of expansion and recession begin.

The random-walk assumption (i.e. the autoregressive (AR) model with unit-root) is widely accepted and tested in the long run for many financial time series. It is unsuitable, however, for forecasting turning points because, in such processes, peaks/troughs occur at completely random periods. In the last decades, owing to the analysis of *non-linear* time series, the assumption of constant parameters has been relaxed. Indeed, many nonlinear AR processes (e.g. bilinear, threshold, exponential) can be viewed as linear AR models with time-varying parameters (TVP). There are also theoretical reasons, related to the representation of stochastic processes, which lead to TVP models (see White's theorem in Granger, 2008). Since exact non-linearity is difficult to identify, models with changing coefficients has been increasingly developed in econometrics and statistics. As a consequence, also the idea of time-varying unit roots (TVUR) has been accepted

Following the approach of random coefficients, Leybourne et al. (1996) and Granger and Swanson (1997) proposed the idea of stochastic unit roots (STUR), which can also be treated with Bayesian inference (see Jones and Marriott, 1999). Grillenzoni (1999) followed a semi-parametric approach, where the roots wander smoothly around the unit value. By using recursive least squares (RLS), he derived a sequential framework for testing for unit roots in real time. More recently, Basci and Caner (2005) considered threshold unit-root processes, where the variability of the root is driven by a threshold variable. Steland (2007) investigated a change-point testing problem where the root is unity only on a certain interval of time, and is stationary elsewhere. All of these approaches were only concerned with testing problems, rather than forecasting and control.

In Grillenzoni (1999) there was the insight that the level of nonstationary series is related to the path of unit roots. In particular, when a root is greater than 1 −and the starting value of the series is positive− it generates trends, whereas below 1 it generates stable behaviors. Therefore, points where time-varying roots cross the unit circle, usually correspond to turning points in the observed series. This remark is important for implementing sequential techniques to identify peaks and troughs in financial data. In this paper we exploit the idea of detecting turning points by monitoring recursive parameter estimates, and we use this information to build investment strategies for stock values.

In econometrics, the identification of turning points usually proceeds by smoothing series with nonparametric filters and then applying the trend-cycle decomposition (e.g., Canova 2007). Technical analysis in finance just compares one-sided moving averages having different sample size. Sequential analysis, and methods of change point detection, compute statistics on the prediction errors of dynamic models (e.g., Vander Wiel 1996). In this context, Frisén (2008) has developed a semiparametric likelihood ratio (LR) based on changes in monotonicity of the trend function. This method actually merges nonparametric smoothing and surveillance statistics, and has been applied to economic and epidemiological data.

The plan of the work is as follows: Section 2 presents models and estimators and explains the approach of turning point detection based on TVUR. Section 3 deals with the selection of smoothing coefficients and proposes an approach based on gain maximization. Section 4 compares the results with techniques based on surveillance statistics of prediction errors. Throughout, an application to the Standard & Poor's (S&P) index is carried out to illustrate and evaluate the various methods.

2. MODEL REPRESENTATION AND ESTIMATION

Analysis of non-stationary time series has a long history and has been developed at various levels. Asymptotically divergent processes, such as AR models with characteristic roots greater than one, were studied since 50s (e.g., White 1958). It was found that common estimators have faster convergence rates, and only the exact unit-root case involves nonstandard distributions (e.g., Fuller 1996). Non-stationarity related to time-varying parameters (TVP) has a more recent history. Many authors have preferred stochastic modelings, applying them also to unit-roots, see Granger and Swanson (1997). This approach has close connections with non-linear time series, and shares similar difficulties as concerned the analysis of model stability and distribution of estimates (e.g., Yoon 2006). Deterministic TVPs do not involve big analytical difficulties. For example, optimal test statistics based on piecewise linear and smoothly changing parameters have been developed in process surveillance (see Andersson et al., 2006).

2.1. Deterministic Parameters

In this paper we consider an AR(1) model whose parameter wanders about the unit value; it also has heteroskedastic innovations with unknown distribution:

$$\begin{aligned} Y_t &= \phi_t \, Y_{t-1} + e_t \,, \qquad e_t \sim \mathrm{ID}(0, \sigma_t^2) \,, \qquad Y_0 = C > 0 \qquad (2.1) \\ \bar{\phi} &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \phi_t = 1 \,, \qquad \bar{\sigma}^2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2 < \infty \end{aligned}$$

More precisely: *i*) the sequence $\{\phi_t\}$ is deterministic and has a time-average $\bar{\phi}$ on the unit circle; *ii*) the innovations $\{e_t\}$ are independently distributed (ID) with finite variances. The process (2.1) can be defined *doubly nonstationary*, in the sense that it has both time-varying coefficients and roots locally greater than one. On average, Y_t behaves like a random walk; however, fluctuations of $\{\phi_t\}$ around the unit value determine more complex patterns, such as local stability as well as local trends. These features strongly depend on the path of the root, on which we do not make specific assumptions; for example, ϕ_t can be either locally smooth or suddenly changing. It follows that model (2.1) is fundamentally semi-parametric, and thus may also represent nonlinear dynamics of Y_t (see Granger, 2008).

To show the effect of parameter changes on the trend of the series, we consider the following switching parameter model

$$Y_t = \begin{cases} \phi_1\, Y_{t-1} + e_t\,, & t \le \mathrm{t}_0\,; \quad \phi_1 > 1, \\ \phi_2\, Y_{t-1} + e_t\,, & t > \mathrm{t}_0\,; \quad \phi_2 < 1 \end{cases}$$

Solving for the difference equation, with the initial condition $Y_0 = C$, yields $Y_t = \phi_1^t\, C + \sum_{k=1}^t \phi_1^k\, e_{t-k}$ if $t \le \mathrm{t}_0$; and for $t > \mathrm{t}_0$ the expected value becomes

$$\mu_t = \mathrm{E}\Big(Y_t\Big) = \phi_2^{(t-\mathrm{t}_0)}\, \phi_1^{\mathrm{t}_0}\, C\,, \quad t > \mathrm{t}_0$$

At time t_0, the slope of the trend μ_t inverts and yields a *turning-point* in the realizations of Y_t. It should also be noted that sign of the slope (positive or negative) depends on the sign of the initial condition C. For real time series, Y_0 is selected according to the initial values of the series.

Previous example extends to continuously varying parameters by investigating the function $\mu_t = \prod_{i=1}^t \phi_i\, C$. Turning points of Y_t depend on those of μ_t, which in turn lie in correspondence of the values ϕ_t=1. As an example, we simulate the model (2.1) with sample size T=1000, Gaussian errors $e_t \sim \mathrm{IN}(0,3)$, initial value Y_0=30 and sinusoidal parameters $\phi_t = 1 + .01\,\sin(t/40)$, which yields $.99 < \phi_t < 1.01$. Figure 1 shows that fluctuations of the root around the unit value strongly influence the level of the series. In particular, expansions of Y_t are always anticipated by excursions $\phi_t > 1$, and periods t_i where $\phi_{t_i} = 1$ correspond to turning points of Y_t.

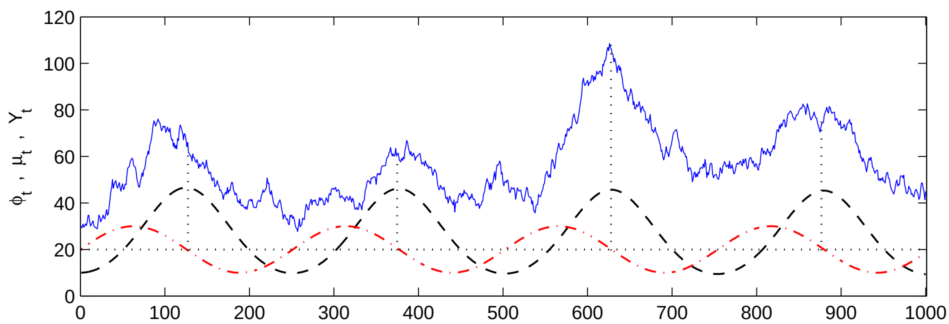

Figure 1. Simulation of the process (2.1) with sinusoidal parameters: series Y_t (solid), mean μ_t (dashed), parameter ϕ_t (dash-dot) (μ_t, ϕ_t are rescaled).

2.2. Stochastic Parameters

To show the analytical difficulties raised by stochastic coefficients, we consider the white noise model: $\phi_t = (1 + a_t)$, where $a_t \sim \mathrm{IID}(0, \sigma_a^2)$ is independent of e_t (e.g., Granger and Swanson, 1997). This model allows useful probabilistic properties to the process $\{Y_t\}$ (e.g., Yoon 2006); however, it is quite unrealistic for real phenomena. Indeed, economy usually changes slowly and random walk or AR models are more sensible parameterizations, namely

$$\begin{aligned} X_t &= \theta\, X_{t-1} + a_t\,, & a_t \sim \mathrm{IN}(0, \sigma_a^2)\,, \quad X_0 = 0 \\ Y_t &= (1 + X_t)\, Y_{t-1} + e_t\,, & e_t \sim \mathrm{IN}(0, \sigma_e^2)\,, \quad Y_0 = C \end{aligned} \tag{2.2}$$

For $\theta = 1$, the realizations of Y_t are asymptotically divergent; since real time series cannot grow indefinitely, suitable conditions of stability must be introduced. Intuitively, there must be an inverse relationship between the coefficients θ, σ_a. The condition $\theta < (1 - \sigma_a)^2$ allows non-divergence of Y_t (e.g., Grillenzoni 1993), and means that the process X_t in (2.2) must always be stationary.

The deterministic parameter case has a more simple treatment. For models as (2.1), the condition of stability is that the average value $|\,\bar{\phi}\,| < 1$. This implies that the series $|\,\phi_t\,|$ is uniformly bounded and is greater than 1 only for a finite number periods. In this case the mean value $\mu_t = \prod_{i=1}^{t} \phi_i\, C$ does not diverge.

Optimal estimation of the process $\phi_t = (1+X_t)$ of the model (2.2) is provided by the Kalman filter. Under the assumption of Gaussian a_t, e_t, the same filter enables to compute the likelihood function of the coefficients θ, σ_a^2, σ_e^2. Unfortunately, the maximum likelihood estimator may encounter problems of convergence due to the redundancy of the noise variances, and the non-smoothness of the likelihood function (e.g., Grillenzoni 1993). For these reasons, Bayesian inference is preferable, see Jones and Marriott (1999), and Kwiatkowski (2006).

2.3. Estimation Algorithms

Given the semi-parametric nature of the model (2.1), one can use the weighted least squares (WLS) estimator. Exponential weighting gives more weight to recent observations and is realized by a coefficient $0 < \lambda \leq 1$ which is inversely proportional to the rate of nonstationarity of the model. In sequential form the estimator is

$$\begin{aligned}
\hat{\phi}_t(\lambda) &= \sum_{i=2}^{t} \lambda^{t-i}\, Y_i\, Y_{i-1} \Big/ \sum_{i=2}^{t} \lambda^{t-i}\, Y_{i-1}^2 \\
\hat{\sigma}_t^2(\lambda) &= \sum_{i=2}^{t} \lambda^{t-i} \left(Y_i - \hat{\phi}_i\, Y_{i-1} \right)^2 \Big/ \sum_{i=2}^{t} \lambda^{t-i}
\end{aligned}$$

In recursive form it is widely used in system identification and has some connections with the Kalman filter (see Ljung, 1999). This is evident by noting $\sum_{i=2}^{t} \lambda^{t-i} \to_t 1/(1-\lambda)$ and rewriting the algorithm as follows

$$\begin{aligned}
\hat{e}_t &= Y_t - \hat{\phi}_{t-1}\, Y_{t-1} \\
R_t &= \lambda\, R_{t-1} + Y_{t-1}^2 \\
\hat{\phi}_t &= \hat{\phi}_{t-1} + R_t^{-1}\, Y_{t-1}\, \hat{e}_t \\
\hat{\sigma}_t^2 &= \lambda\, \hat{\sigma}_{t-1}^2 + (1-\lambda)\, \hat{e}_t^2
\end{aligned} \tag{2.3}$$

where $\hat{e}_t$ are prediction errors and R_t is the denominator of $\hat{\phi}_t(\lambda)$.

Algorithm (2.3) requires the initial values $\hat{\phi}_0, R_0, \hat{\sigma}_0$; these have relative importance when $\lambda \ll 1$ and can be obtained with OLS on a sub-set of data. Last equation of (2.3) is the adaptive variance of prediction errors and resembles the autoregressive conditional heteroskedastic (ARCH) modeling of σ_t^2. The resemblance is, however, superficial being ARCH a class of parametric models. Performance of the algorithm (2.3) in *tracking* the unknown sequence ϕ_t has been studied under several conditions of parameter variation (see Guo and Ljung, 1995). Assuming smooth evolution, the WLS estimator can be designed to reach the minimum MSE.

Unlike $\hat{\phi}_0, \hat{\sigma}_0$, the design of the weighting factor λ is crucial for the performance of (2.3). As the bandwidth in kernel smoothers, it can be selected with a cross-validation approach by minimizing the sum of squared innovations:

$$\hat{\lambda}_T = \arg\min \sum_{t=1}^{T} \left(\hat{e}_t / \hat{\sigma}_{t-1} \right)^2 \tag{2.4}$$

Under Gaussianity, a simple method to test for the variability of unit-roots on the period $[1,\, T]$, consists of comparing the final value of the loss function in (2.4) with $\sum_{t=2}^{T} (y_t/\sigma_y)^2$, where $y_t = (Y_t - Y_{t-1})$, by means of F-statistic.

2.4. Testing and Detection

In the context of statistical surveillance, an important problem is testing for the unit-root hypothesis in a sequential way. Under the null $\phi_t = 1$, the model (2.1) becomes a pure random walk and classical methods could be used. However, test statistics must be adapted to the TVUR alternative, which requires adaptive estimators. As regards the Student statistic, Banerjee et al. (1992) referred to the *rolling* regression, i.e. sequential OLS with constant sample size. For the algorithm (2.3), Grillenzoni (1999) proposed an adaptive solution which converges weakly as $t \to \infty, \lambda \to 1$

$$S_t(\lambda) = \sqrt{R_t\, \hat{\sigma}_t^{-2}}\, \Big[\hat{\phi}_t(\lambda) - 1\Big] \quad \to^{\mathrm{L}} \quad \int_0^1 W(s)\, \mathrm{d}W(s) \left[\sqrt{2} \int_0^1 W^2(s)\, \mathrm{d}s\right]^{-1/2}$$

where $W(s)$ is a standard Brownian motion. Except for the term $1/\sqrt[4]{2}$, the above coincides with the Dickey-Fuller distribution which is used in classical tests for unit-root (see Fuller, 1996 p. 642). To directly use the tabulated critical values, one must modify the above statistic as $\hat{S}_t(\lambda) = S_t(\lambda)\, \sqrt[4]{1+\lambda}$, which takes into account the greater variability induced by the exponential weighting.

Now, the important fact for the paper is that the above framework can be used to detect turning points in the series Y_t, namely the periods where the slope of the series inverts and new phases of expansion or recession begin. As in the comments to Figure 1, the first approach consists of identifying the points where the estimates $\hat{\phi}_t(\lambda)$ cross the unit circle. Points t_k where $\Big(\hat{\phi}_{t_k} > 1 \,\Big|\, \hat{\phi}_{t_k - 1} \leq 1\Big)$ are points of expansion, whereas in the opposite case they are points of recession. The rationale of this solution is that the root ϕ_t determines the local slope of the trend μ_t and its threshold is automatically defined by the value 1.

The second approach mitigates the previous idea in that uses the statistic $\hat{S}_t(\lambda)$ and upper, lower (U, L) critical values of the Dickey-Fuller distribution as tolerance bands for the decision. Given the test size $.01 \leq \alpha \leq .10$, and the acceptance region of the unit-root hypothesis $\big[L_{\alpha/2}\,,\, U_{1-\alpha/2}\big]$, one can detect an expansion if $\hat{S}_t$ goes above U, and a recession if $\hat{S}_t$ goes below L, namely

$$\begin{array}{lcl} \text{Expansion} \quad t_k & : & \Big(\hat{S}_{t_k} > U_{1-\alpha/2} \,\Big|\, \hat{S}_{t_k - 1} \leq U_{1-\alpha/2}\Big) \\ \text{Recession} \quad t_h & : & \Big(\hat{S}_{t_h} < L_{\alpha/2} \,\Big|\, \hat{S}_{t_h - 1} \geq L_{\alpha/2}\Big) \end{array} \tag{2.5}$$

In practice, the interval $[L, U]$ is the critical region for the occurrence of turning points, whereas it is consistent with the stability of the trend slope.

2.5. S&P Case Study

We begin here an application to the Standard & Poor's (SP) index of the New York Stock Exchange (NYSE), which is the leading indicator of many stock prices. We consider the daily SP500 in the period January 4, 1999 - December 30, 2009, for a total of T=2767 observations (about 251 per year). Series Y_t is displayed in Figure 2a, where the consequences of the crises in 2002 and 2008 are apparent. One of the issues to check is whether the major peak of Sept. 2007 could be timely identified or, at least, the big fall of Sept. 2008 early detected.

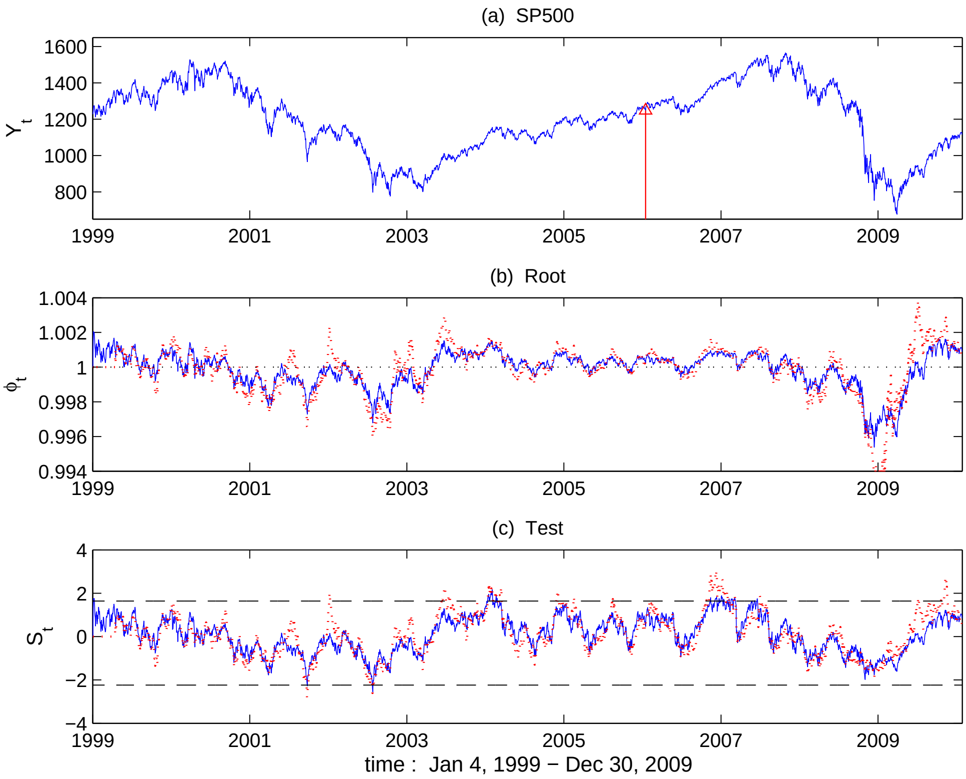

Figure 2. (a) Daily SP500 index in the period [1999, 2009], with the point Dec 30, 2005 (△); (b) Estimates of ϕ_t obtained with WLS, λ=.99 (solid), and OLS, N=100 (dotted); (c) Student statistics under the null ϕ_t=1, and Dickey-Fuller 95% critical values (dashed).

Average analysis on the whole sample confirms the random-walk hypothesis, because OLS estimation provides $\hat{\phi}_T$=.999, $\hat{\sigma}_e$=15.23 and S_T=-.51 (the Student statistic under the null ϕ_t=1). However, the cross-validation (2.4) yields the value $\hat{\lambda}_T$=.37, which indicates parameter variability. Figure 2b confirms that $\hat{\phi}_t$ (2.3) wanders significantly even with mild weighting, as λ=.99. For the sake of comparison, it also exhibits the estimates $\bar{\phi}_t$ obtained with a rolling regression of sample-size N=100 (see Banerjee et al., 1992). Figure 2c provides the corresponding Student statistics $\hat{S}_t$, $\bar{S}_t$, together with Dickey-Fuller 95% critical values for N=100.

Owing to the sample-size relationship N=1/(1-λ), the rectangular window of size N=100 should be *equivalent* to the exponential window with λ=.99. However, one can see that OLS estimates in Figure 2b,c are less smooth than those of WLS. This is due to the fact that WLS actually involves a greater amount of past data, because $\lambda^N \neq 0$. The pattern of the root confirms that values $\hat{\phi}_t > 1$ roughly correspond to periods of expansion (buy) of Y_t, whereas the others correspond to phases of recession (sell). By following the indications of Figure 2b, one could detect the fall risk for stock prices just at the beginning of 2008. Statistics in Figure 2c confirm these insights and lead to *local* rejections of the unit-root hypothesis. This test legitimates the use of the TVUR model (2.1) for the SP500 series.

3. DESIGN AND SELECTION OF COEFFICIENTS

Previous application has shown that *heuristic* selection of the coefficient λ and of the limits U, L may not provide a close correspondence between the level of estimates $\hat{\phi}_t$, $\hat{S}_t$ and the location of turning points of Y_t. In particular, Figure 2 shows that $\hat{\phi}_t$ crosses the unit value too frequently and $\hat{S}_t$ is not a timely indicator by following the rule (2.5). Investment strategies require suitable design of algorithms and decision rules in order to achieve profitability (e.g., Lam and Yam, 1997; and Bock et al., 2008). In particular, the optimal design of a trading system should be based on the economic results of the system itself. In this section we follow this principle to optimize the method of buying when $\phi_t > 1$ and selling when $\phi_t < 1$.

3.1. Maximum Gain

Assuming only a viewpoint of *long* market, i.e. where the gain is obtained by increments of stock values (*short* rule can be conceived as well), the optimal trading problem is to identify the buy-sell sequence

$$\{t_i, s_i\}_1^n \quad : \quad 1 \le t_1 < s_1 < t_2 \ldots < t_n < s_n \le T$$

which maximizes the *total* gain G_T

$$\begin{aligned} G_T(t_i, s_i) &= \sum_{i=1}^{n} \Big(Y_{s_i} - Y_{t_i} \Big) \\ \text{i.e.} \quad \{\hat{t}_i, \hat{s}_i\} &= \text{arg max } G_T(t_i, s_i) \end{aligned} \tag{3.1}$$

or its mean value G_T/n, where n is the number of trading cycles (buy & sell) in the period $[1, T]$. Clearly, the sought sequence coincides with that of turning points of expansion-recession of Y_t. As in Section 2, $\{t_i, s_i\}$ can tentatively be identified with the periods where $\hat{\phi}_t(\lambda)$ crosses the unit circle.

Random fluctuations of estimates may cause, however, false alarms or weak signals, i.e. when $\hat{\phi}_t$ crosses the threshold 1 just for few periods. To reduce this risk, one can smooth $\hat{\phi}_t$ with a one-sided moving average of few terms, or one may delay the decision to the second consecutive signal, etc. As in the equation (2.5), the definitive solution is provided by a tolerance band $(1 \pm \kappa)$ for values of $\hat{\phi}_t$ close to 1. In this case, the trading sequence is identified as follows: t_i the first time $\hat{\phi}_t$ goes above $1{+}\kappa$, and s_i the first time $\hat{\phi}_s$ goes below $1{-}\kappa$, namely

$$\begin{aligned} \text{Buy} \quad t_i \quad &: \quad \Big[\hat{\phi}_{t_i} > (1+\kappa) \Big| \hat{\phi}_{t_i-1} \le (1+\kappa) \Big] \\ \text{Sell} \quad s_i \quad &: \quad \Big[\hat{\phi}_{s_i} < (1-\kappa) \Big| \hat{\phi}_{s_i-1} \ge (1-\kappa) \Big] \end{aligned} \tag{3.2}$$

It follows that $\{t_i, s_i\}$ depend on (λ, κ) and, instead of the criterion (2.4), one can select the coefficients by maximizing the total gain, or its mean value

$$(\hat{\lambda}, \hat{\kappa}) = \text{arg max } \; G_T\Big[t_i(\lambda, \kappa), s_i(\lambda, \kappa) \Big] \tag{3.3}$$

Similarly, if the initial capital is continuously reinvested, one could also estimate the coefficients by maximizing the final *relative* gain, defined as

$$g_T(\kappa, \lambda) = \prod_{i=1}^{n} \Big(Y_{s_i} \Big/ Y_{t_i} \Big) \tag{3.4}$$

Taking logarithm it is clear that maximization of (3.4) is equivalent to (3.3). Given the inverse relationship between alarm delay and expected utility (e.g., Frisén 2008), it is clear that the rule (3.3) allows unbiased detection of turning points.

Objective function G_T is usually non-smooth and may have several local maxima. Since it has only two entries, it is possible to identify the position of the global optimum by exploring its surface on a grid of values of $\lambda\,,\kappa$. Subsequently, numerical optimization methods can refine the search. To test the out-of-sample performance, we carry out the optimization (3.3) only on the first $T_1 \ll T$ observations; next we compute the (out-of-sample) gain on the remaining $T_2 = T - T_1$ data. With this approach, we are interested to test the reliability of the method and to check the stability over time of the estimates $\hat{\lambda}, \hat{\kappa}$.

3.2. S&P Case Study

In the SP500 case study, we selected T_1=1760, which corresponds to Dec 30, 2005. Figure 2a shows that the training period [1, T_1] is sufficiently complete, in terms of expansion and recession phases, and the index change is negligible, i.e. $Y_{T_1} \approx Y_1$. As in other studies (e.g., Bock et al. 2008), we have placed the first buying signal at t_1=1 (Jan 4, 1999). Figure 3 plots the contour of the gain functions (3.1) and (3.4), evaluated on [1, T_1] for a grid of values of λ, κ. Their pattern is similar, relatively smooth and convex, and show an inverse relationship between the coefficients. This is natural because as λ decreases, the variability of estimates $\hat{\phi}_t$ increases; therefore, larger tolerance limits $\pm\kappa$ are needed.

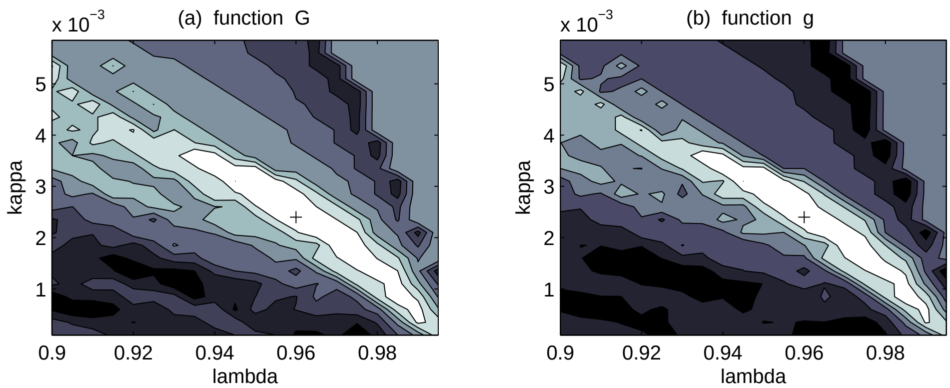

Figure 3. Contour of functions G_T (3.1) and g_T (3.4) for T=1760 data of SP500.

The optimization (3.3) provides $\hat{\lambda}$=.963, $\hat{\kappa}$=.0021 and G_{T_1}=440, with just two trading cycles: (t_1, s_1)=(1,449) and (t_2, s_2)= (1094,T_1). By extending the analysis to the whole period, one obtains (t_2, s_2)=(1094,2168) and (t_3, s_3)=(2588,T), with the out-of-sample gain G_{T_2}(.963,.0021)=+388. Further, the values of $\hat{\lambda}, \hat{\kappa}$ on the whole sample [1, T] do not change significantly. Figure 4 displays the estimates $\hat{\phi}_t(.963)$ and the alarm signals (t_i, s_i).

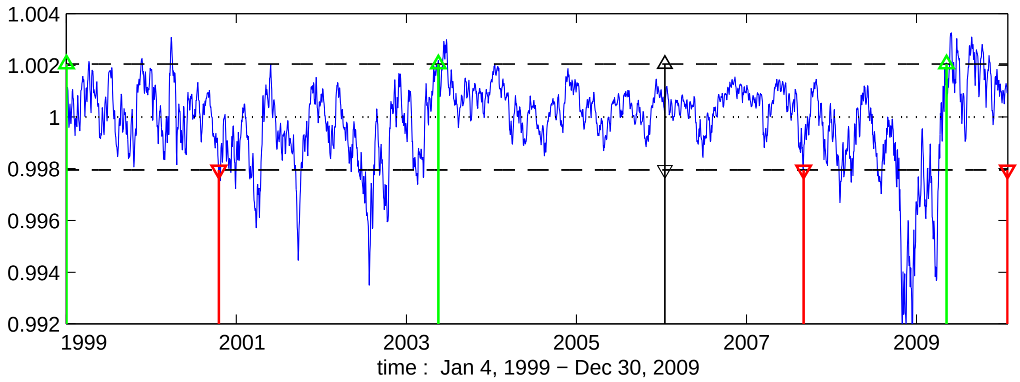

Figure 4. Graph of $\hat{\phi}_t(.963)$, bands $(1 \pm .0021)$ and trading points $t_i(\triangle)$, $s_i(\bigtriangledown)$.

With respect to the heuristic estimate $\hat{\phi}_t(.99)$ in Figure 2b, one can see that $\hat{\phi}_t$ in Figure 4 is more reactive and enables to track the positive rally of SP500 in April 2009. Curiously, inverting the decision rule (3.2) (i.e. buying when $\hat{\phi}_t > 1 - \kappa$ and selling when $\hat{\phi}_t < 1 + \kappa$) increases the in-sample gain as G_{T_1}(.83,.001)=857; however, this happens at the cost of inflating the number of trading signals as n=103, and making the out-of-sample gain very negative: G_{T_2}(.83,.001)= -356. These side effects show, once more, the importance of tolerance bands $1 \pm \kappa$ to reduce false alarms and to avoid wrong decisions.

As regards the evaluation of the relative gain function (3.4), out-of-sample it provides g_{T_2}(.963,.0021)=1.43, over 4 years. In terms of gross annualized return $1.43^{1/4} - 1$, it corresponds to more than +9%, which is very positive, because on the period Jan 2006 - Dec 2009, the index performance was negative: Y_T / Y_{T_1+1}=0.89. These conclusions do not change by adding the annual rate of dividends which, on SP500, is about +2% (see Standard & Poor's, 2008).

3.3. HSI Case Study

As a further application, which introduces other methods, we consider a case study discussed in Lam and Yan (1997) and Bock et al. (2008). It deals with the Hang Seng Index (HSI), of Hong Kong stock exchange, in the period Feb. 10, 1999 - June 26, 2002, for a total of T=829 observations. The authors consider the series in logarithm and select the model coefficients on the first T_1=71 observations. By comparing CUSUM statistics, adaptive filters, hidden Markov models and LRs, they find that best method is a semiparametric LR which tests the change in monotonicity of the trend function μ_t (see Andersson et al., 2006). In particular, by selecting the alarm limit on the T_2 segment, they obtain G_{T_2}= 0.32 with n=49 sell signals.

In our application, we trained the method (3.2) on the data of 1998, obtaining $\hat{\lambda}$=.625, $\hat{\kappa}$=.00006; on T_2 these yield G_{T_2}=0.51 with n=56 cycles. The performance remains good also by subtracting a transaction cost k=0.1% (0.2% per cycle), or by dividing the total return by n. Figure 5 displays graphical results; one can see that TVUR method can work in situations where frequent tradings are required.

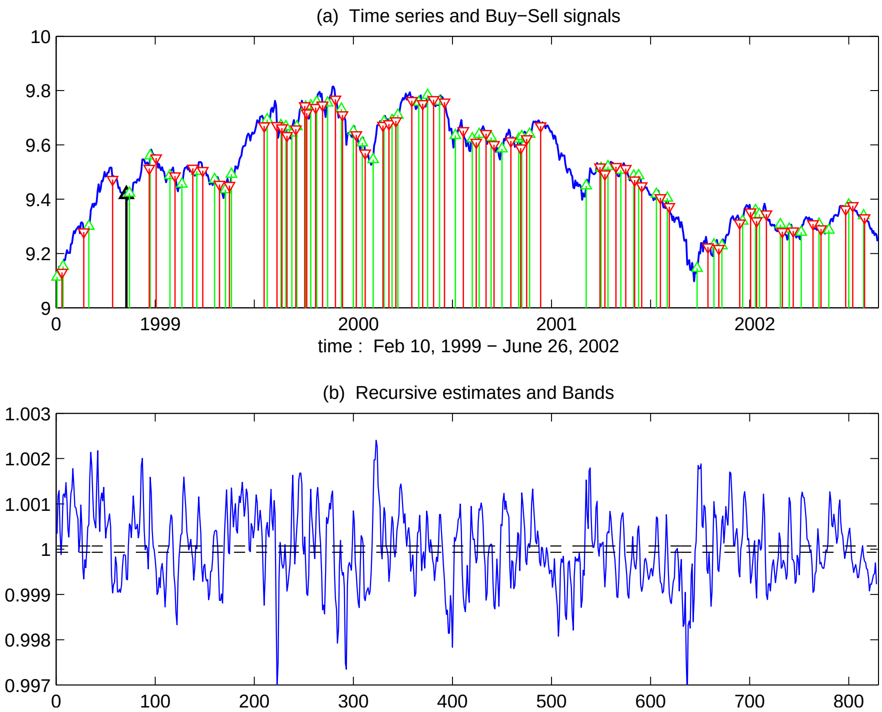

Figure 5. (a) Log of HSI from Feb. 10, 1999 - June 26, 2002, with trading points $t_i(\triangle)$, $s_i(\bigtriangledown)$ signaled by method (3.2); (b) Estimates $\hat{\phi}_t(.625)$ and bands $(1\pm.00006)$.

4. COMPARISON WITH OTHER METHODS

In this section we compare previous results with those of surveillance and control methods. These originate from quality control charts of industrial manufacturing, and Lam and Yam (1997) and Blondell et al. (2002) have applied them to economics. Recently, Bock et al. (2008) have provided a comprehensive comparison of surveillance methods for finance, focusing on the trend model

$$Y_t = \mu_t + \varepsilon_t \ , \qquad \varepsilon_t \sim \mathrm{IN}(0, \sigma_\varepsilon^2) \tag{4.1}$$

where $\mu_t = \mathrm{E}(Y_t)$ is a smooth function with peaks and troughs at unknown periods, and $\{\varepsilon_t\}$ is independent normal. This differs from change-point models, where μ_t is locally constant with jumps and discontinuities which must be identified with sequential tests. For the model (4.1), Frisén (2008, Ch.3) has proposed a semi-parametric LR method which signals changes in monotonicity of μ_t. This function is estimated with restrictions of monotonicity under the null, and restrictions of unimodality (U-shaped) under the alternative. The method provides sequential detection of turning points, and represents a powerful alternative to nonparametric smoothing of Y_t and trend-cycle decomposition of μ_t (see Canova, 2007 Ch.3).

4.1. Control Statistics

As in adaptive control literature (see Vander Wiel, 1996; and Ljung, 1999), in this section we refer to the dynamic model (2.1), and we compute monitoring statistics on the prediction errors of the algorithm (2.3). This has the advantage of working with a series $\{\hat{e}_t\}$ which is nearly stationarity and independent; hence, it fulfills the conditions for the optimality of test statistics. The basic intuitive idea behind adaptive detection is that large prediction errors $\hat{e}_t^*$ tend to occur in correspondence of turning and change points. It then follows that a simple signaling rule consists of comparing the size of the errors with their standard deviation $\hat{\sigma}_e$.

Since structural changes generate patches of significant errors, cumulative sums (CUSUMs) of $\hat{e}_t$ are more robust indicators. Control charts usually adopt the two-sided form of such statistics (see Vander Wiel, 1996)

$$
\begin{aligned}
C_t^+ &= \max\Big[\,0\,,(\,C_{t-1}^+ + \hat{e}_t - \eta\,)\Big]\,, \quad C_0^+ = 0 \\
C_t^- &= \min\Big[\,0\,,(\,C_{t-1}^- + \hat{e}_t + \eta\,)\Big]\,, \quad C_0^- = 0
\end{aligned} \tag{4.2}
$$

where C_t^+ is sensitive positive shifts in the mean of e_t, and C_t^- is sensitive to negative changes; $\eta > 0$ is the tolerance value (typically $\eta = \sigma_e$). This scheme gives an alarm when $\max(C_t^+, -C_t^-)$ exceeds a threshold $\kappa > 0$; this value is selected so as to minimize the number of false alarms, or the delay in detecting a real change.

Application of the above framework to financial trading make necessary changes, because stock price series are clearly nonstationary.

1. The reference model should be (2.1), and the algorithm (2.3) provides the required prediction errors. Given heteroskedasticity, CUSUM statistics $C_t^\pm$ must be computed on the standardized errors $\hat{u}_t = \hat{e}_t/\hat{\sigma}_{t-1}$.

2. Because prediction errors tend to be negative when a recession starts, whereas they are positive when an expansion begins, the decision rules become:

$$
\begin{aligned}
\text{Buy} \quad t_i \ &: \ \Big(C_{t_i}^+ > \kappa \,\Big|\, C_{t_i-1}^+ \le \kappa\Big) \\
\text{Sell} \quad s_i \ &: \ \Big(C_{s_i}^- < -\kappa \,\Big|\, C_{s_i-1}^- \ge -\kappa\Big)
\end{aligned} \tag{4.3}
$$

3. The control coefficients (η, κ) can be selected automatically as in (3.3), by maximizing the in-sample gain on actual data. This avoids the design based on average run lengths (ARL), which involves complex computations and difficult interpretation. Further, Vander Wiel (1996) showed that when the process Y_t is random walk, then the ARL-performance of main control statistics (including the LR) worsens significantly.

Notice that the CUSUM statistic used in econometrics $C_t^0 = \sum_{i=1}^t \hat{e}_i$ tends to reproduce the random walk series Y_t. Therefore, in order to use C_t^0 in place of (4.2), its value must be reset to 0 whenever it passes its alarm limit κ

$$
C_t^0 = \begin{cases} C_{t-1}^0 + \hat{e}_t/\hat{\sigma}_{t-1}\,, & \text{if } |\,C_{t-1}^0| \le \kappa, \\ \hat{e}_t/\hat{\sigma}_{t-1}\,, & \text{if } |\,C_{t-1}^0| > \kappa \end{cases}
$$

The exponentially weighted moving average (EWMA) of $\hat{e}_t$ follows the same principle as CUSUMs, but gives more weight to recent errors. With respect to the two-sided indicator (4.2), the major problem of EWMA is its inertia in staying in the state (positive or negative) of a signaled alarm. This may cause loss of timeliness in detecting turning points which are close to each others. As a solution, one may adopt the *resetting* adjustment used for C_t^0. Thus, using the indicator function $I(\cdot)$, the modified EWMA statistic becomes

$$M_t = (1-\lambda)\, M_{t-1} \cdot I\Big(|M_{t-1}| \le \kappa\Big) + \lambda\; \hat{e}_t/\hat{\sigma}_{t-1}\ , \tag{4.4}$$

where $\kappa > 0$ is the alarm limit. The decision rule for EWMA is similar to (4.3), that is: Buy when $M_t > \kappa$ and Sell when $M_t < -\kappa$. The coefficients λ, κ can be selected with the gain criterion (3.3); for reasons of parsimony, it is sensible to use in (4.4) the same λ as (2.3). This could reduce the alarm capability of M_t, in which case it may be useful to exchange λ and $(1-\lambda)$ in the equation (4.4).

4.2. S&P Case Study

Before applying these schemes to the SP500 series, we test the performance of the so-called *Shewhart* method, which consists of monitoring individual prediction errors. As stated before, negative errors indicate beginning of a recession, therefore the decision rule is: Buy when $\hat{e}_t/\hat{\sigma}_{t-1} > \kappa$, etc.. Selection of the coefficients, by maximizing the in-sample gain, gives $\hat{\lambda}$=.95, $\hat{\kappa}$=2.9, with G_{T_1}=331 and n=2 cycles.

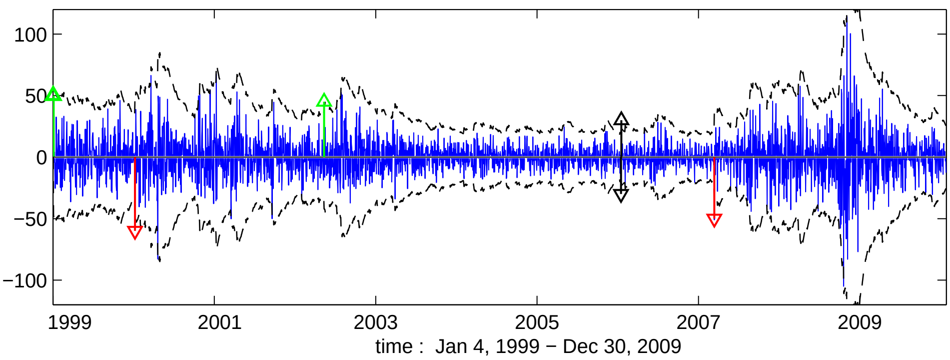

Figure 6. Prediction errors $\hat{e}_t(.95)$ and tolerance bands $\pm 2.9\hat{\sigma}_{t-1}$, computed with the algorithm (2.3) on the SP500 series.

The out-of-sample performance is positive G_{T_2}=130, but is less than half of that of TVUR. This is due to the fact that Shewhart method does not pick up the expansion phase started on April 2009. Figure 6 plots the series $\hat{e}_t(\hat{\lambda})$ and the confidence bands $\pm\hat{\kappa}\,\hat{\sigma}_{t-1}$, together with the alarm signals.

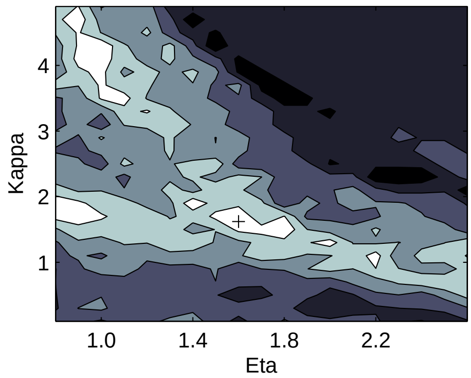

Figure 7. Contour of the function $G_T(\eta,\kappa|\lambda = .97)$ of the CUSUM method (4.2) for T_1=1760 data of SP500.

In applying the CUSUM scheme (4.2), Figure 7 shows the contour of the gain function $G_{T_1}(\eta,\kappa)$, conditioned on the a-priori value λ=.97. The surface has two local maxima at (1.6,1.8) and (1,4); the second one has a larger value of κ and, therefore, involves a smaller number of trading cycles. On the basis of these initial values, numerical minimization of (3.3) provides $\hat{\lambda}$=.965, $\hat{\kappa}$=1.34, $\hat{\eta}$=1.66 with G_{T_1}=330 and n=2. The out-of-sample performance is positive G_{T_2}=+130 and is identical to that of the Shewhart scheme. Figure 8 shows the CUSUM statistics C_t^+, C_t^-, together with the trading signals on the whole period [1,T].

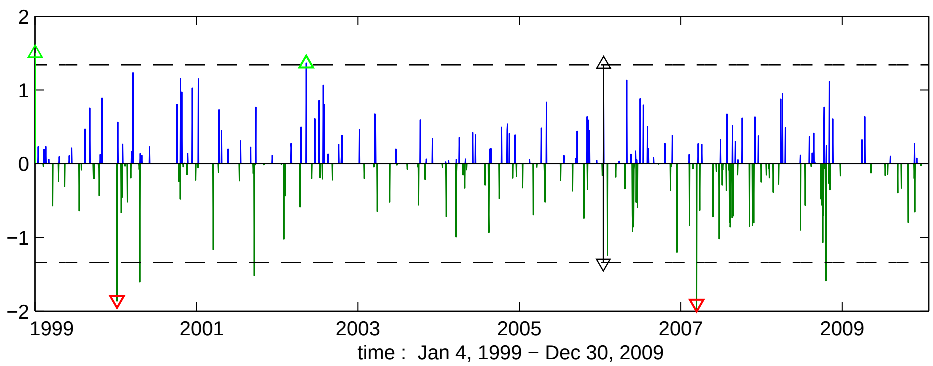

Figure 8. CUSUM statistics C_t^+, C_t^- in (4.2) and alarm bands $\pm\hat{\kappa}$ for SP500.

Last analysis concerns the EWMA statistic (4.4). Selection of the coefficients with the criterion (3.3) on the first T_1=1760 observations provides $\hat{\lambda}$=.950, $\hat{\kappa}$=2.73, with G_{T_1}=331 and $n = 2$ tradings. The out-of-sample performance is slightly inferior to the previous ones, since G_{T_2}=+113. Figure 9 provides the path of the statistic (4.4), with the trading signals on the whole sample.

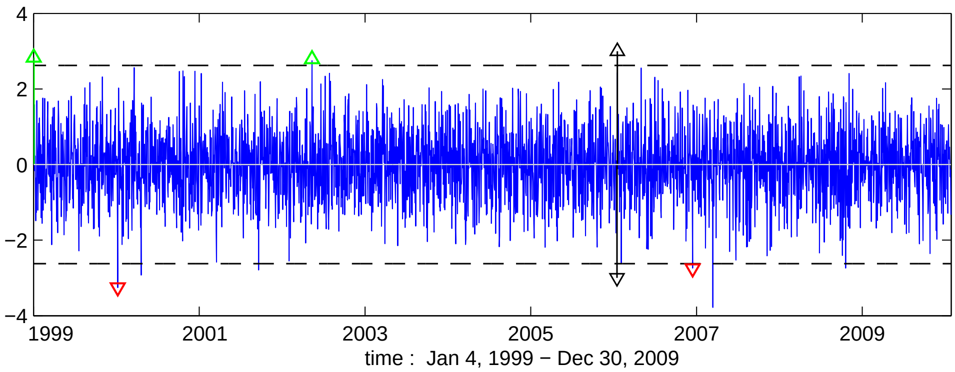

Figure 9. EWMA statistic M_t in (4.4) and alarm bands $\pm\hat{\kappa}$ for SP500.

4.3. Some Comparisons

Table 1 summaries the main numerical results of Sections 3 and 4 as concerned the Standard and Poor's case study. There are four main remarks:

1. The time-varying unit-root (TVUR) method outperforms the others, especially at out-of-sample level. This is a consequence of the stability of its coefficients, which enables to pick up the expansion phase started on April 2009.

2. Methods based on control statistics provide similar results. This is due to the equivalence of their alarm capability in the case of random walk processes (see Vander Wiel, 1996). Indeed, SP500 and the model (2.1) belong to the class of random walks. The relative difference of EWMA may be due to the fact that its coefficient λ is constrained to that of the algorithm (2.3).

3. The annualized gross return rates r_{out}, show that all methods are preferable to the naive investment strategy of buying and holding, even considering the return coming from dividends.

4. Inversion of the decision rule can provide a significant improvement of the in-sample gain; however, this happens at the cost of increasing the number of trading actions and deteriorating the out-of-sample performance.

Table 1. Summary of numerical results of the SP500 case study: $\hat{\lambda}, \hat{\kappa}, \hat{\eta}$ are coefficients which maximize $G_{\rm in}$; this is the in-sample gain (3.1) computed on the period [1999, 2005]; $G_{\rm out}$ is the out-of-sample gain on [2006, 2009]; n is the number of trading cycles; $g_{\rm in}, g_{\rm out}$ are the relative gains (3.4) and $r_{\rm out} = g_{\rm out}^{1/T_2} - 1$ is the annualized rate. S-inverse is the Shewhart method based on the inversion of the decision rule.

Method	$\hat{\lambda}$	$\hat{\kappa}$	$\hat{\eta}$	$G_{\rm in}$	n	$G_{\rm out}$	$g_{\rm in}$	$g_{\rm out}$	$r_{\rm out}$ %
TVUR (3.2)	.963	.0021	.	440	3	388	1.47	1.43	9.4
Shewhart	.95	2.90	.	331	2	130	1.31	1.10	2.4
CUSUM (4.3)	.965	1.34	1.66	331	2	130	1.31	1.10	2.4
EWMA (4.4)	.95	2.73	.	331	2	113	1.31	1.09	2.2
S-inverse	.965	1.91	.	693	20	−158	1.76	0.91	−2.3

As regards the last point, the rationale of inverting the decision rules (e.g. buying when prediction errors are negative) corresponds to the so-called "contrarian investing", which tries to anticipate the market tendencies. In the normal operating mode, the tolerance bands $\pm\kappa$ serve to reduce the number of false alarms and wrong detections; in the inverted mode, they increase the number of possible tradings. The risk of this strategy is well described by what happens in the out-of-sample context, where the gain performance becomes negative. However, this may be a consequence of the excessive length of the forecasting horizon and/or of the time-variability of the coefficients λ, κ. The strategy of updating the coefficient estimates as new data become available may reduce the negative side effects.

4.4. Simulation Study

In order to evaluate in depth the methods described so far, it is necessary to perform simulations experiments. We consider an AR(1) model with parameter ϕ_t=1+.01

$\sin(t/40)$, which wanders in the interval [.99, 1.01], with Y_0=10 and $e_t \sim \mathrm{IN}(0,1)$ Gaussian. A typical realization of Y_t is provided in Figure 1. We perform m=1000 replications of length T=1000, and we apply TVUR, Student, Shewhart and EWMA methods to sub-samples of size $T_1 = T_2 = 500$. The Student method is based on the statistic $\hat{S}_t$ of Section 2.4, and replaces U, L limits in (2.5) with $\pm\kappa$, to be selected with (3.3). Its advantage, with respect to the estimator $\hat{\phi}_t$, is the greater smoothness of $\hat{S}_t$, which could reduce false alarms. Mean values of estimates are reported in Table 2; they show the superiority of the first two methods, which have greater out-of-sample gains and a smaller number of trading cycles n. We have also applied inverted decision rules to TVUR and Shewhart, but their in-sample gains are inferior to those in Table 2, and out-of-sample gains are negative.

Table 2. Results of a simulation experiment with an AR(1) model with sinusoidal unit-root (e.g., Figure 1). The entries are mean values over m=1000 replications; their legend is given in Table 1, but $G_{\mathrm{in,out}}$ are computed on $T_{1,2}$=500 observations.

Method	$\bar{\lambda}$	$\bar{\kappa}$	$\bar{G}_{\mathrm{in}}$	$\bar{n}$	$\bar{G}_{\mathrm{out}}$
Student (2.5)	.741	.571	29.2	30.0	13.2
TVUR (3.2)	.778	.003	28.6	35.6	12.9
Shewhart	.812	1.46	27.7	59.2	3.6
EWMA (4.4)	.795	1.26	29.0	46.0	2.9

5. CONCLUSIONS

In this paper we have developed methods of turning point detection and trading strategies for financial time series. The original proposal is to monitor the path of recursively estimated AR roots on the unit circle. The second proposal is to select the smoothing coefficients λ, κ with a data driven approach based on gain maximization. Numerical results, on real and simulated data, have shown that TVP approach outperforms methods based on prediction errors. The reason is that the relationship between prediction errors and turning points is not causal, because significant errors may also be generated by outliers or heteroskedasticity.

The proposed methodology is flexible and is open to all changes which concern signal statistics and decision rules. For example, one may use any unit-root test statistic (as S_t) in place of the root estimate $\hat{\phi}_t$; or one may invert the decision rules in various way, etc.. However, open issues are still present in the selection and updating of the design coefficients λ, κ, η. In on-line trading, where the operating horizon is one-step-ahead, those coefficients should be re-estimated for each new observation. Unfortunately, there are not recursive algorithms for this updating, and numerical optimization (3.3) must be carried out for each period t.

Topics for further research also deal with the stability over time of the selected coefficients, and, therefore, with the optimal design of the in-sample size T_1. Despite the fact that our methodology is completely adaptive, in situations of high volatility it is sensible to use only the most recent data. This seems, indeed, the main lesson from the recent crisis in 2008.

ACKNOWLEDGMENTS

The author is very grateful to the editors and referees for a concrete aid in improving the quality of the article.

REFERENCES

Andersson, E., Bock, D., and Frisén, M. (2006). Some Statistical Aspects of Methods for Detection of Turning Points in Business Cycles, *Journal of Applied Statistics* 33: 257-278.

Banerjee, A., Lumsdaine R. L., and Stock, J. H. (1992). Recursive and Sequential Tests of the Unit Root and Trend Break Hypothesis: Theory and International Evidence, *Journal of Business & Economic Statistics* 10: 271-289.

Basci, E. and Caner, M. (2005). Are Real Exchange Rates Nonlinear or Nonstationary? Evidence from a New Threshold Unit Root Test, *Studies in Nonlinear Dynamics & Econometrics* 9: 1-15.

Blondell, D., Hoang, P., Powell, J., and Shi, J. (2002). Detection of Financial Time Series Turning Points: A New CUSUM Approach Applied to IPO Cycles, *Review of Quantitative Finance and Accounting* 18: 293-315.

Bock, D., Andersson, E., and Frisén, M. (2008). The Relation Between Statistical Surveillance and Technical Analysis in Finance, in *Financial Surveillance*, M. Frisén, ed., pp. 69-92, London: Wiley.

Canova, F. (2007). *Methods for Applied Macroeconomic Research*, Princeton: Princeton University Press.

Frisén, M. (2008). *Financial Surveillance*, London: Wiley.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*, New York: Wiley.

Granger, C. W. J. (2008). Non-linear Models: Where Do We Go Next - Time Varying Parameter Models?, *Studies in Nonlinear Dynamics & Econometrics* 12: 1-9.

Granger, C. W. J. and Swanson, N. R. (1997). An Introduction to Stochastic Unit Root Processes, *Journal of Econometrics* 80: 3562.

Grillenzoni, C. (1993). ARIMA Processes with ARIMA Parameters, *Journal of Business & Economic Statistic* 11: 235-250.

Grillenzoni, C. (1999). Adaptive Tests for Changing Unit Roots in Nonstationary Time Series, *Journal of Computational and Graphical Statistics* 8: 763-778.

Guo, L. and Ljung, L. (1995). Performance Analysis of General Tracking Algorithms, *IEEE Transactions on Automatic Control* 40: 1388-1402.

Jones, C. R. and Marriott, J. M. (1999). A Bayesian Analysis of Stochastic Unit Root Models, *Bayesian Statistics* 6: 785-794.

Kwiatkowski, J. (2006). A Bayesian Estimation and Testing of STUR Models with Application to Polish Financial Time-Series, *Dynamic Econometric Models* 7: 151-160.

Lam, K. and Yam, H. C. (1997). CUSUM Techniques for Technical Trading in Financial Markets, *Financial Engineering and Japanese Market* 4: 257-274.

Leybourne, S. J., McCabe, B. P. M., and Mills, T. C. (1996). Randomized Unit Root Processes for Modelling and Forecasting Financial Time Series: Theory and Applications, *Journal of Forecasting* 15: 253-270.

Ljung, L. (1999). *System Identification: Theory for the User*, 2nd Edition, New York: Prentice Hall.

Standard & Poor's (2008). S&P 500 Dividend Aristocrats, McGraw-Hill, available at www.standardandpoors.com/indices/sp-500-dividend-aristocrats

Steland, A. (2007). Weighted Dickey-Fuller Processes for Detecting Stationarity, *Journal of Statistical Planning and Inference* 137: 4011-4030.

Vander Wiel, S. A. (1996). Monitoring Processes that Wander Using Integrated Moving Average Models, *Technometrics* 38: 139-151.

White, J. (1958). The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case, *Annals of Mathematical Statistics* 29: 1188-1197.

Yoon, G. (2006). A Note on Some Properties of STUR Processes, *Oxford Bulletin of Economic and Statistics* 68: 253-260.