

Sequential smoothing for turning point detection with application to financial decisions

Carlo Grillenzoni*

A fundamental problem in financial trading is the correct and timely identification of turning points in stock value series. This detection enables to perform profitable investment decisions, such as buying-at-low and selling-at-high. This paper evaluates the ability of sequential smoothing methods to detect turning points in financial time series. The novel idea is to select smoothing and alarm coefficients on the gain performance of the trading strategy. Application to real data shows that recursive smoothers outperform two-sided filters at the out-of-sample level.

Keywords: capital gain; double exponential; Hodrick-Prescott; Kalman filter; kernel smoothing, local regression, Standard & Poor index

**Correspondence to: C. Grillenzoni, Dept. of Planning, University IUAV, Venice, Italy. E-mail: carlog@iuav.it*

1. Introduction

A fundamental problem in the analysis of nonstationary time series is the detection of *turning points* (TP), i.e., identification of periods where the series changes the sign of its slope (e.g., Zellner *et al.* [24]). This topic is different from forecasting, which is concerned with pointwise prediction of future values, nevertheless it is crucial for planning control actions. For example, in macro-economics, knowing the beginning of a recession leads to increased government expenditures and money supply. In financial markets, it leads to selling equities, where correctness and timeliness of the detection are fundamental for maximizing gains.

There are many approaches to the analysis of TP in economic time series; generally speaking they can be classified in three groups. The first consists of *smoothing* series with adaptive filters; subsequently, first and second differences of the extracted signals are evaluated as the order conditions of deterministic functions (e.g., Canova [4, Ch. 3]). Another detection approach, mostly used in technical analysis of finance, compares the pattern of moving averages with different window sizes; if short-term averages cross the long-term ones, then a TP is detected. The second group deals with *local* estimates of regression parameters, and uses them as indicators of the local slope of the series. A change in the sign of slope coefficients indicates the occurrence of a turning point. Some methods refer to linear models and recursive least-squares (e.g., Ljung [17]); others are more sophisticated and consider Markov switching regression models (see Marsh [18]). The third group comes from sequential analysis of change-point problems, and uses hypothesis testing of mean shifts in probability distributions (e.g., Erghashev [8]). Typical test statistics are likelihood ratio (LR) and cumulative sums (CUSUM), which are widely employed in quality control of industrial manufacturing.

The aforementioned methods involve various technical problems. For example, smoothing methods identify turning points on the basis of trend-cycle components estimated on the entire data-set. This approach yields problems of accuracy of endpoint estimates and timeliness when on-line detection is used. Time-varying parameters require complex estimators, such as Kalman filter (KF) and expectation-

maximization (EM) algorithms, whose statistical properties are not well investigated in conditions of nonstationarity and nonlinearity (e.g., Teräsvirta *et al.* [20]). As regards testing methods, they treat turning points as if they were change points in independent or stationary sequences. However, the loss of optimality of test statistics in the presence of autocorrelation and nonstationarity may be considerable (see Vander Wiel [21]).

In this paper we only focus on smoothing methods, and we compare their detection performance. The methods are nonparametric (as the local polynomial regression) and parametric (as the latent component model); one-sided (as the double exponential) and two-sided (as Hodrick-Prescott). Statistical smoothers are natural tools for turning point identification because they filter out the random elements of the series and allow to monitor the trend component. However, they must be adapted to on-line data processing in order to achieve timeliness (e.g., Wildi and Elmer [23]). Some methods, as exponential smoothing (ES), can be easily cast in recursive form; instead, two-sided filters (kernel regression and splines), require *sequential* adaptation and suitable selection of the window size. In all cases, the smoothing coefficients (bandwidths) of the various filters must be selected to fulfill the final goals of financial activity.

Trading strategies usually pursue the rule of buying-at-low and selling-at-high price; hence, they require early identification of local troughs and peaks of stock values. In this paper we check the ability of various smoothers to detect turning points and we evaluate their performance on the consequent investment decisions. The fundamental idea is to select smoothing and alarm coefficients through the maximization of the capital gain computed on past data. Next, the out-of-sample profit can be evaluated on future observations with the same framework. Since maximum gain occurs in correspondence of actual turning points, it follows that the proposed solution also pursues unbiased detection.

The plan of the work is as follows: Section 2 presents smoothing methods and shows their connections. Section 3 explains turning point detection and the implied trading system. Section 4 applies the methodology to financial time series.

2. Representation and estimation

To introduce the problem in a general way, we assume that the series are generated by a non-linear and non-stationary process Y_t . The representation which is suitable in this case is the nonparametric one of Teräsvirta *et al.* [20]

$$Y_t = f(t, \mathbf{y}_t) + v(t, \mathbf{y}_t) a_t, \quad a_t \sim \text{IID}(0, 1)$$

where $f(\cdot)$ is the conditional mean, $v^2(\cdot)$ is the conditional variance, $\mathbf{y}'_t = [Y_{t-1} \dots Y_{t-r}]$, $r > 0$ is the "state" vector and $\{a_t\}$ is an independent and identically distributed (IID) sequence. Further, following the approach of *additive* models of Hastie and Tibshirani [15], we also assume that the regression function can be decomposed into the sum of two components

$$f(t, \mathbf{y}_t) = g(t) + m(\mathbf{y}_t)$$

where the first deals with the trend-cycle (low frequency) component, and the second is nearly stationary in covariance.

Many authors define the turning points directly on Y_t , or its realizations (e.g., Zellener *et al.* [24]); however, this approach is problematic because Y_t is a stochastic process. Since $g(t)$ is deterministic, it enables a precise definition of turning points as local *troughs* t_i and local *peaks* s_i of the function itself, namely

$$\begin{aligned} t_i &: g(t_i - b_i) \geq \dots \geq g(t_i - 1) > g(t_i) < g(t_i + 1) < \dots < g(t_i + d_i) \\ s_i &: g(s_i - p_i) \leq \dots \leq g(s_i - 1) < g(s_i) > g(s_i + 1) > \dots > g(s_i + q_i) \end{aligned} \quad (1)$$

for some $(b_i, d_i; p_i, q_i) > 1$. Given the sampling interval $[1, T]$, we assume that the double sequence $\{t_i, s_i\}$ contains $n \ll T/2$ pairs, which can be ordered as

$$1 \leq t_1 < s_1 < t_2 < \dots < t_i < s_i < \dots < s_{n-1} < t_n < s_n \leq T$$

The aim of this paper is identification/detection of the periods $\{t_i, s_i\}$ by means of smoothing methods for time series. There are two main classes of smoothers, depending on the fact they have a one-sided or a two-sided structure.

A1. Kernel regression. The most general kernel approach is Local Polynomial Regression (LPR, see Fan and Gijbels [9]), which can also be applied to *irregularly* observed series. In a fixed design context this yields the model

$$Y_{t_k} = g(t_k) + y_{t_k}, \quad 1 \leq t_k \leq T, \quad k = 1, 2, \dots, N < T$$

where the "innovation" y_t includes the components $\{m_t, a_t\}$ and, therefore, it may not be IID. However, the autocorrelation of y_t does not change the bias of kernel smoothers, and only influence their variance (see Beran and Feng [3]).

The LPR smoother in a possible continuous domain $t \in [1, T]$ is defined as the weighted least squares estimator

$$\hat{g}_{\text{LPR}}(t) = \arg \min_g \left\{ \frac{1}{N\sigma} \sum_{k=1}^N K\left(\frac{t-t_k}{\alpha}\right) \left[(Y_{t_k} - g) - \sum_{h=1}^p g_h (t-t_k)^h \right]^2 \right\} \quad (2)$$

where $K(\cdot)$ is a symmetric density function (which provides local weighting), $\alpha > 0$ is the bandwidth and g stands for the constant term g_0 of the polynomial of order p . For $p=0$ one has the simple kernel smoother (Nadaraya-Watson type) and for $p=1$ one has the local linear regression (LLR). Major advantage of the latter is the *automatic* boundary correction property, which significantly reduces the bias of estimates $\hat{g}(\cdot)$ at the endpoints t_1, t_N . This feature enables to potentially use the method for early detection of the turning points defined in (1).

At the computational level, the PLR estimator can be explicitly defined as a weighted least squares (WLS) algorithm. Letting $\mathbf{g}' = [g, g_1 \dots g_p]$ the vector of parameters to be estimated in (2), and $\mathbf{t}'_k = [1, (t-t_k) \dots (t-t_k)^p]$ the vector of regressors, it can be seen that

$$\hat{\mathbf{g}}_N(t) = \left[\sum_{k=1}^N K_\alpha(t-t_k) \mathbf{t}_k \mathbf{t}'_k \right]^{-1} \sum_{k=1}^N K_\alpha(t-t_k) \mathbf{t}_k Y_{t_k}, \quad t \in [1, T]$$

A robust version of this algorithm has been developed by Grillenzoni [12]; it is both in-sensitive to anomalous observations (outliers) and sensitive to sudden changes in the level of the series. This feature is useful for detecting turning points in correspondence of structural breaks, which frequently occur in financial series.

A2. Spline smoothing. Spline is another class of non-parametric smoothers for irregularly observed data. The method looks for estimates which achieve a

compromise between data-fitting and function smoothness. Typically

$$\hat{g}_{\text{SS}}(t) = \arg \min_g \left\{ \sum_{k=1}^N [Y_{t_k} - g(t_k)]^2 + \gamma \int [g^{(h)}(t)]^2 dt \right\}$$

where $g^{(h)}$ is the h -th order derivative and $\gamma > 0$ is a tuning coefficient that control the trade-off between fitting and smoothness. The solution \hat{g}_{SS} is a piecewise polynomial of degree $(2h-1)$ between the points t_k . For $h=2$ one has the *cubic* spline, and for regularly observed time series the second component of the penalty function can be computed as second difference. This actually corresponds to the filter of Hodrick and Prescott [16]

$$\hat{g}_{\text{HP}}(t) = \arg \min_g \left\{ \sum_{t=1}^T (Y_t - g(t))^2 + \gamma \sum_{t=2}^{T-1} [g(t+1) - 2g(t) + g(t-1)]^2 \right\} \quad (3)$$

where for $\gamma = \sigma_{(y)}^2 / \sigma_{[d]}^2$ (the ratio of the variances of the two components of (3)), the corresponding smoothing spline is optimal in MSE sense.

At computational level, the estimator (3) can be rewritten in vector form as

$$\hat{\mathbf{g}}_{\text{HP}} = \mathbf{F}_T(\gamma) \mathbf{y}_T, \quad \mathbf{F}_T(\gamma) = \left(\mathbf{I}_T + \gamma \mathbf{H}'_T \mathbf{H}_T \right)^{-1}$$

where \mathbf{H}_T is a $(T-2) \times T$ block diagonal matrix, with the triple $(1, -2, 1)$ on the main diagonals, and $\mathbf{y}'_T = [Y_1, Y_2 \dots Y_T]$ (see Danthine and Girardin [7]). As a result, the weights of the matrix $\mathbf{F}_T(\gamma)$ depend on the dimension T , and its rows have a-symmetric path at the borders. This feature yields inaccuracy of endpoint estimates; to improve them, one can augment the vector \mathbf{y}_T with out-of-sample forecasts \hat{Y}_{T+k} (see Mise *et al.* [19]).

B1. Kalman filtering. So far we have considered two-sided filters which estimate $\hat{g}(t)$ by smoothing past and future observations $Y_{t \pm k}$ with symmetric weights (with the exception of endpoints). We now consider one-sided filters which process data recursively and may involve parametric representations. The first one is the unobserved components (UC) model, which assume that the trend function has a *double* random walk structure (e.g., Harvey and Koopman [14])

$$Y_t = g_t + y_t, \quad y_t \sim N(0, \sigma_y^2)$$

$$\begin{aligned}
g_t &= g_{t-1} + h_{t-1} + e_{1t}, & e_{1t} &\sim \text{IN}(0, \sigma_1^2) \\
h_t &= h_{t-1} + e_{2t}, & e_{2t} &\sim \text{IN}(0, \sigma_2^2)
\end{aligned} \tag{4}$$

where y_t, e_{1t}, e_{2t} are mutually independent and normal (IN). In the system (4), g_t represents the level component of the series, while h_t stands for its slope. Both may be useful for identifying the turning points of Y_t .

By defining the arrays $\mathbf{A} = [1, 0; 1, 1]$, $\mathbf{b}' = [0, 1]$, $\mathbf{\Sigma} = \text{diag}([\sigma_1^2, \sigma_2^2])$ and $\mathbf{x}'_t = [h_t, g_t]$, $\mathbf{e}'_t = [e_{1t}, e_{2t}]$, the system (4) can be cast in state-space form as

$$\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{A} \mathbf{x}_t + \mathbf{e}_t, & \mathbf{e}_t &\sim \text{IN}(0, \mathbf{\Sigma}) \\
Y_t &= \mathbf{b}' \mathbf{x}_t + y_t, & y_t &\sim \text{N}(0, \sigma_y^2)
\end{aligned}$$

Now, by applying the forward Kalman filtering (see Ljung [17] or Grillenzoni [11]) one can obtain a recursive estimator which is simpler than that used by the authors who developed the model (4)

$$\begin{aligned}
\mathbf{k}_t &= \mathbf{A} \mathbf{P}_{t-1} \mathbf{b} (\mathbf{b}' \mathbf{A} \mathbf{b} + \sigma_y^2)^{-1} \\
\mathbf{P}_t &= \mathbf{A} \mathbf{P}_{t-1} \mathbf{A}' - \mathbf{k}_t (\mathbf{b}' \mathbf{A} \mathbf{b} + \sigma_y^2) \mathbf{k}'_t + \mathbf{\Sigma} \\
\hat{\mathbf{x}}_t &= \mathbf{A} \hat{\mathbf{x}}_{t-1} + \mathbf{k}_t (Y_t - \mathbf{b}' \hat{\mathbf{x}}_{t-1})
\end{aligned} \tag{5}$$

where \mathbf{k}_t is the filter gain and \mathbf{P}_t is the dispersion matrix. As regards the estimation of the trend component, it is provided by the second element of the state vector; namely, $\hat{g}_{\text{UC}}(t) = \hat{x}_{2t}$.

B2. Exponential smoothing. The most simple recursive smoother is the exponentially weighted moving average (EWMA). An extension, which is suitable for trend models, is the *double* exponential smoother (ES):

$$\begin{aligned}
\hat{s}_t &= \lambda \hat{s}_{t-1} + (1 - \lambda) Y_t \\
\hat{g}_t &= \lambda \hat{g}_{t-1} + (1 - \lambda) \hat{s}_t
\end{aligned} \tag{6}$$

where $\lambda \in (0, 1]$ is a weighting factor which gives more weight to recent observations. The first equation can be applied to forecasting as $\hat{Y}_{t+1} = \hat{s}_t$, and a multivariate robust version has been discussed in Croux *et al.* [6].

Also double exponential smoothing is concerned with forecasting. As in the linear trend model, the forecast function is sum of level and slope components. In the Brown's approach these can be estimated with the statistics of (6)

$$\begin{aligned} \text{level } \hat{a}_t &= (2\hat{s}_t - \hat{g}_t) \\ \text{slope } \hat{b}_t &= (\hat{s}_t - \hat{g}_t)(1 - \lambda)/\lambda \\ \hat{Y}_{t+k} &= \hat{a}_t + \hat{b}_t k \end{aligned} \tag{7}$$

Further, recomposing (6) and (7) leads to the Holt-Winters algorithm

$$\begin{aligned} \hat{a}_t &= \lambda(\hat{a}_{t-1} + \hat{b}_{t-1}) + (1 - \lambda)Y_t \\ \hat{b}_t &= \lambda\hat{b}_{t-1} + (1 - \lambda)(\hat{a}_t - \hat{a}_{t-1}) \end{aligned}$$

see Chatfield *et al.* [5]. For the trend estimator, one can use either \hat{g}_t in (6) or \hat{a}_t in (7), where the latter is less smooth than the first.

Several papers have shown the connections between HP, UC, ES methods, and their relationships with other filters, see Harvey and Jäger [13] and Gómez [10]. The general conclusion is that they are asymptotically equivalent when the innovation process y_t is white noise and the tuning coefficients are properly selected; for example, HP and UC converge to the same estimates when $\sigma_1=0$ and $\gamma=(\sigma_y/\sigma_2)^2$. The connection between HP and ES follows by noting that exponential weights arise from the optimization problem (3) by using first difference in the second term, i.e. a matrix \mathbf{H} with diagonal elements $(1, -1)$. Further, the level component of the UC model, corresponds to the simple ES filter s_t when $\sigma_2 = 0$ and $(\sigma_y/\sigma_1)^2 = \lambda^2/(1 - \lambda)$; whereas the Kalman filter solution converges to the Holt-Winters algorithm when $(\sigma_y/\sigma_2)^2 = \lambda^4/(1 - \lambda)^2$ (e.g., Harvey and Koopman [14]). Apart from formal relationships, empirical applications have shown the similarity of trend estimates provided by the various methods (e.g., Alexandrov *et al.* [1]). On the contrary, significant differences have been obtained for the cycle component, which sometimes lead to prefer HP and UC methods (see Van Ruth *et al.* [22]).

In our presentation, we just noted that the non-parametric smoothers can work in conditions of missing data, but are inaccurate at endpoints; instead, recursive

methods need regularly observed series and suitable initial conditions for $\hat{\mathbf{x}}_0$, s_0 , etc.. In real-life applications, turning points must be identified *sequentially*, as new data Y_t become available. This implies that also two-sided methods must be managed sequentially, starting from the first observation. This management rises the problem of initial values, as in the case of recursive methods. For *all* filters we solve the initialization problem by adding to the data-set an initial sub-sample of size N , just rescaled to the level of the first observation, namely

$$Y_{-j}^* = Y_{N-j} - (Y_N - Y_1), \quad j = 1, 2 \dots N \ll T \quad (8)$$

The size of N is important for the quality of estimates on the left border and, in the KF (5), it is associated to the period required to reach the *steady-state*.

Owing to the structure of their weights, the derivation of the recursive version of two-sided smoothers is not possible. To speed up calculation of LPR and HP, we use sequential versions based on a moving subsample of size $N < T$. For regularly observed series, they are given by the first entry of the vectors

$$\begin{aligned} \hat{\mathbf{g}}_N(t) &= \left[\sum_{j=t-N+1}^t K_\alpha(t-j) \mathbf{t}_j \mathbf{t}'_j \right]^{-1} \sum_{j=t-N+1}^t K_\alpha(t-j) \mathbf{t}_j Y_j \\ \hat{\mathbf{g}}_{t,N} &= \left(\mathbf{I}_N + \gamma \mathbf{H}'_N \mathbf{H}_N \right)^{-1} \mathbf{y}_{t,N}, \quad t = 1, 2 \dots T \end{aligned} \quad (9)$$

where $\mathbf{t}'_j = [1, (t-j) \dots (t-j)^p]$ and $\mathbf{y}_{t,N} = [Y_t \dots Y_{t-N+1}]'$. In Hodrick-Prescott filter, the weights of the matrix \mathbf{F}_N also depend on the series length, therefore the window size N should be designed together with γ .

The signal extraction capability of the filters (2)-(6) crucially depends on their smoothing coefficients. A general method for selecting $\alpha, \gamma, \sigma, \lambda$ is the cross-validation (CV), or its robust version (see Grillenzoni [12]). In sequential form, the CV approach corresponds to the prediction error (PE) criterion

$$(\hat{N}, \hat{\theta})_{\text{PE}} = \arg \min_{\theta} \sum_{t=2}^T [Y_t - \hat{g}_\theta(t-1)]^2, \quad \theta = \alpha, \gamma, \sigma, \lambda$$

where the estimates $\hat{g}(\cdot)$ are obtained with the data up to time $(t-1)$, in recursive or sequential form. For the parametric model (3), the maximum likelihood method can be applied to $\boldsymbol{\sigma}' = [\sigma_y, \sigma_1, \sigma_2]$. However, some constraints on the coefficients, such

as $\sigma_1 = \sigma_2$, may be necessary to prevent identification and convergence problems (see Grillenzoni [11]). In the next section we will define a selection strategy which stems from the operational use of the models.

3. Turning point detection and financial decisions

Time series of stock prices and exchange rates are strongly nonstationary and non-linear, but the methods discussed in Section 2 are suitable for trend estimation. Financial investors are usually interested in timely identification of the turning points of Y_t , so as to take profitable decisions, such as *buying low* and *selling high*. We assume that they follow investment strategies of *long* type, i.e. where the returns are obtained from the increments of stock values. A complete trading-cycle is defined as a buy action which is coupled with the subsequent sale; the corresponding capital *gain* is the price difference in the two periods.

As in the definition (1), we assume that the trend function $g(t)$ has n pairs of turning points $\{t_i, s_i\}$ in the sampling interval $[1, T]$, and realizations of the process Y_t have turning points which are close to $\{t_i, s_i\}$. This means that the variance of the innovation series $y_t = Y_t - g(t)$ is small compared to that of Y_t . The total gain in the period $[1, T]$ is provided by

$$G_T(t_i, s_i) = \sum_{i=1}^n (Y_{s_i} - Y_{t_i})$$

and its expected value is $\sum_i [g(s_i) - g(t_i)]$. In the inter-cycle periods $s_i < t < t_{i+1}$, we assume that money is invested in low-risk activities whose returns cover the transaction costs on equities. For the sake of simplicity, we do not consider *short* strategies because they are risky and not very popular.

Since the points $\{t_i, s_i\}_1^n$ are unknown, sequential methods must be used for their identification; the natural approach is to estimate $g(t)$ with smoothing methods and then apply the definition (1). However, timeliness of the detection requires one-step signaling, and this increases the probability of wrong decisions. To reduce the number of false alarms, a tolerance value $\kappa > 0$ can be introduced; the detection

rules then become

$$\begin{aligned} \text{trough } t_i &: \left[\hat{g}_\theta(t_i) > \hat{g}_\theta(t_i - 1) + \kappa \right] \cap \left[\hat{g}_\theta(t_i - 1) < \hat{g}_\theta(t_i - 2) - \kappa \right] \\ \text{peak } s_i &: \left[\hat{g}_\theta(s_i) < \hat{g}_\theta(s_i - 1) - \kappa \right] \cap \left[\hat{g}_\theta(s_i - 1) > \hat{g}_\theta(s_i - 2) + \kappa \right] \end{aligned} \quad (10)$$

where θ denotes the type of smoother, and $\hat{g}_\theta(t_i)$ is estimated with data up to time t_i . To better understand the meaning of (10), one may rewrite the inequalities in terms of $[\hat{g}(t) - \hat{g}(t - 1)]$. Unlike the detection rules based on first and second differences (e.g., Canova [4, Chap. 3]), the advantage of (10) is that it directly identifies the kind (trough or peak) of a turning point. Further, it stresses the role of the tolerance coefficient κ , which reduces the number of false alarms.

Rather than fitting and forecasting, in this paper we are interested to check the *control* capability of smoothing methods. This means evaluating the effectiveness of investment decisions which stem from the scheme (10). The selection of design coefficients θ, κ follows the same principle, and can be based on the function G_T . To be specific, we first define the maximum gain (MG) detected points as

$$(\hat{t}_i, \hat{s}_i)_{\text{MG}} = \arg \max_{t,s} G_T(t_i, s_i)$$

and since, from (10), they depend on the coefficients θ, κ we have

$$(\hat{\theta}, \hat{\kappa})_{\text{MG}} = \arg \max_{\theta, \kappa} G_T[t_i(\theta, \kappa), s_i(\theta, \kappa)] \quad (11)$$

In practice, we design the smoothers on the basis of their maximum profitability, which implicitly means timely identification of the turning points. Obviously, also the window size N can be inserted into the program (11).

The approach followed so far has implicitly assumed that the quantity of money invested at each period t is constant. However, if only the initial capital is fixed, and it is entirely invested and disinvested at each trading, then the implied gain function becomes

$$R_T(t_i, s_i) = \prod_{i=1}^n Y_{s_i} / Y_{t_i} \quad (12)$$

The reference value of R_T is 1, but since $\log(R_T) = G_T[\log(Y_t)]$, it follows that the maximization of (12) leads to the same solution as (11).

S&P application. We apply the methodology to the Standard and Poor's (S&P) index of the New York Stock Exchange, which is the leading indicator of many world stock price indexes. We consider the daily SP500 series in the period [1999, 2009], which consists of $T=2767$ observations (about 251 data per year). The series is displayed in Figure 1(a), where the dramatic consequences of the crises of 2002 and 2008 are evident. Figure 1(b) reports the smoothed series generated by the filters of Section 2, with heuristically selected coefficients; namely: LLR($\alpha=30$), HP($\gamma=10^5$), KF($\sigma_y=33$, $\sigma_{1,2}=0.005$), ES($\lambda=0.96$). The one-sided filters KF, ES were initialized as in (8), with a sub-sample of size $N=250$. As one can see, the methods LPR, HP are nearly equivalent, whereas the others have greater variance and bias, in terms of delay with respect to the two-sided estimates.

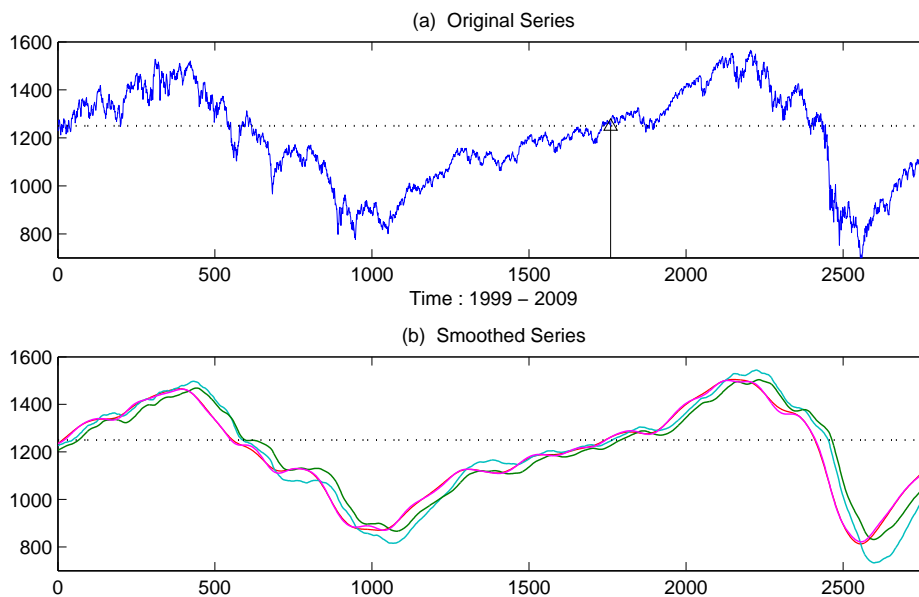


Figure 1. (a) S&P500 series in the period [1999, 2009] and $t_1=30$ Dec. 2005; (b) Smoothed series: two-sided filters (red), one-sided recursive filters (green).

Figure 1(a) also displays the point $t_1=30$ Dec 2005, which separates the in-sample period [1999, 2005] (with $T_1=1760$ data), from the out-of-sample period [2006, 2009] (with $T_2=1007$ data). Selection of the coefficients θ, κ with the approach (11) is based on in-sample data, whereas the remaining data serve for evaluation. In this context, the two-sided filters were implemented sequentially as in (8)-(9), using

the window size N of the initiation sub-sample. Owing to the non-smooth pattern of the objective function G_T , the choice of initial values of N, θ, κ is very important. These are selected by evaluating the function G_{T_1} on a grid of values as in Figure 2, which is conditioned on the choice $\kappa=0.05$.

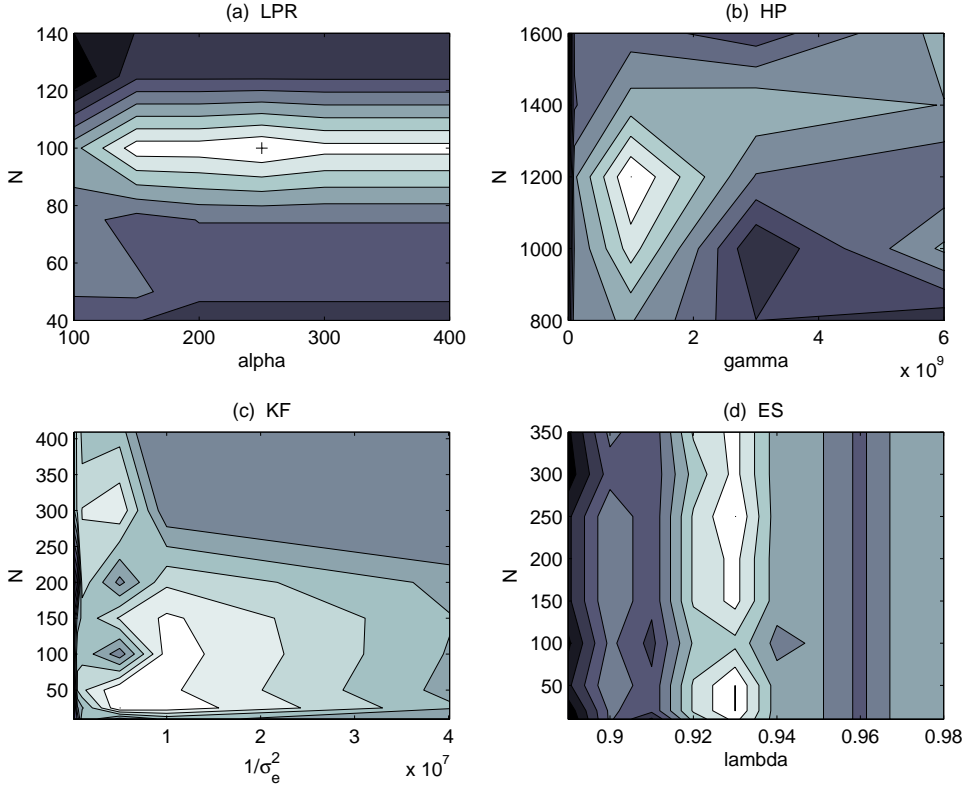


Figure 2. Contour of the gain function $G_{T_1}(N, \theta | \kappa = 0.05)$, where $\theta = \alpha, \gamma, \sigma_e^2, \lambda$, on $T_1=1760$ observations of the period [1999, 2005], of four smoothing methods.

Figure 2(a) shows that the gain performance of LLR is maximized by the value $N=100$, whereas the ES method is relatively independent of N . These facts can be explained with the structure of the smoothers (9) and (6). On the basis the initial values identified in Figure 2, the optimization (11) was carried out with numerical methods. The results are reported in Table I; it can be seen that one-sided smoothers outperform the others by 2-3 times at out-of-sample level. The KF was implemented with the constraints $\sigma_{e_1}=\sigma_{e_2}$, and $\hat{\sigma}_y$ is estimated on the initiation sub-sample Y_t^* . Unlike ES, the performance of KF is hindered by its parametric complexity, which also involves the highest number of trading cycles n . The HP filter needs the largest

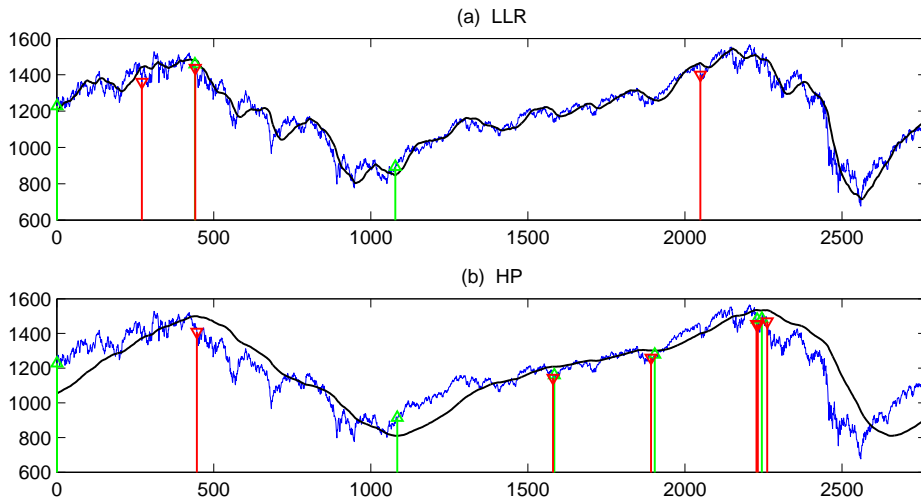
value of N and, therefore, it is extremely slow. As is natural, the value of n is inversely proportional to the size of the tolerance coefficient κ .

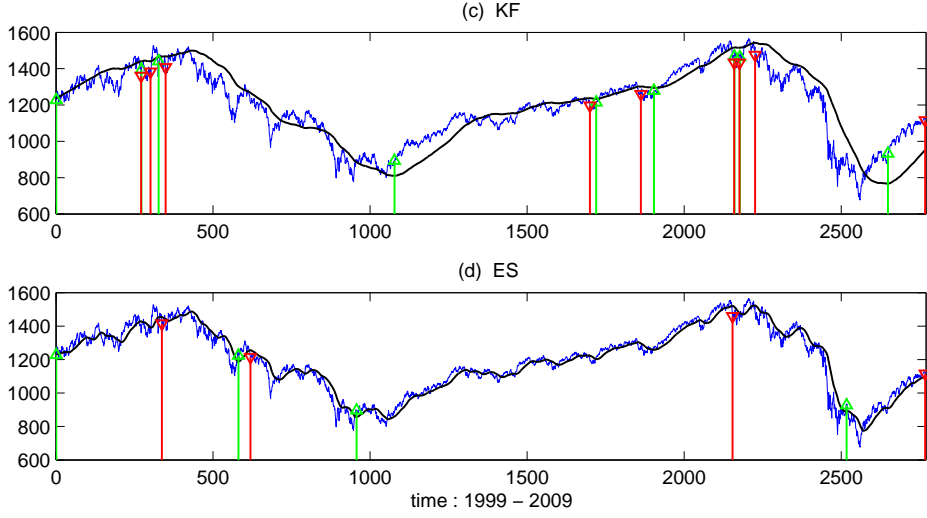
Table I. Results obtained by applying (11) to S&P series: N is defined in (8)-(9); $\theta = \alpha, \gamma, \sigma_e^2, \lambda$ are smoothing coefficients; κ is the threshold in (10); G_{T_1}, G_{T_2} are in-sample and out-of-sample gains, $T_1=1760$; n is the number of trading cycles.

| Method | Eq. | \hat{N} | $\hat{\theta}$ | $\hat{\kappa}$ | G_{T_1} | n | G_{T_2} |
|--------|------|-----------|-----------------|----------------|-----------|-----|-----------|
| LLR | (2) | 126 | 245 | .364 | 485 | 3 | 151 |
| HP | (3) | 1115 | $58 \cdot 10^7$ | .033 | 497 | 6 | 136 |
| KF | (5) | 15 | $2/10^6$ | .063 | 423 | 10 | 312 |
| ES | (6) | 32 | .913 | .121 | 529 | 4 | 419 |
| ESo | (13) | 10 | .981 | .110 | 477 | 4 | 301 |

Trend estimates generated with the coefficients of Table I are displayed in Figure 3, together with the corresponding buy and sell signals. Unlike the two-sided estimates of Figure 1(b), the results in Figures 3(a),3(b) are obtained with the sequence of right-endpoint values. In the out-of-sample period, all methods avoid the crash of 2008, but only the recursive filters capture the buy signal in 2009.

Figure 3. Trend estimates (black), and buy (green), sell (red) signals generated with the coefficients of Table I, of four sequential smoothers.





The good performance of ES suggests its use in the context of the so-called *oscillation* techniques (say ESo). Many traders try to identify the turning points of stock series, by comparing one-sided moving averages of different size. Typically, if the 50-days average crosses the 200-days average, then a turning point is identified. Applying this principle to the smoothers (6), one has the detection rule

$$\begin{aligned}
 \text{trough } t_i &: \hat{s}_\lambda(t_i) > \hat{g}_\lambda(t_i) + \kappa \\
 \text{peak } s_i &: \hat{s}_\lambda(s_i) < \hat{g}_\lambda(s_i) - \kappa
 \end{aligned} \tag{13}$$

where the coefficients λ, κ can be selected with the method (11). The results for S&P data are given in the last row of Table I. The gain performance of ESo is similar to that of KF, and its components are displayed in Figure 4.

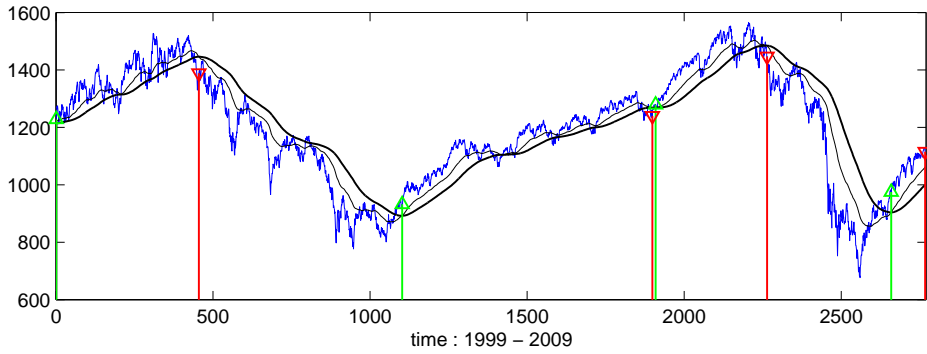


Figure 4. Trend estimates and buy-sell signals generated with the coefficients in the last row of Table I for the method (6)-(13): $\hat{s}(t)$ (thin), $\hat{g}(t)$ (solid).

4. Conclusions

Investment decisions in financial markets require timely identification of turning points of stock values. Smoothing methods for nonstationary time series have a fundamental role in this analysis. In this paper we have considered four types of smoothers and we have discussed their relationships in a sequential implementation. Classical two-sided filters require a suitable moving window and present accuracy problems at endpoints. On the contrary, recursive smoothers have faster adaptability but their performance may depend on the parametric complexity. Exponential filtering is recursive in nature and involves just one coefficient; hence, it is suitable in situations which involve complex computations.

In the trading activity, smoothing coefficients must not necessarily optimize prediction error criteria (namely pointwise forecasting), but should be able to timely detect sparse turning points. The gain function G_T , based on the sum of the differences between significant peaks and troughs of past data, can be constructed for all filters. These functions are usually non-smooth and need efficient search algorithms to find their global optimum. Obviously, the simpler the filter structure, the more effective and stable the solution turns out.

The empirical application has shown that various smoothers are nearly equivalent on the in-sample interval (as measured by G_{T_1}), but differ significantly in the out-of-sample period [2006-2009]. In particular, recursive filters outperform the others by 2-3 times (in terms of G_{T_2}). In general, all filters are acceptable because provide a sell-signal during the year 2007, which avoid the crash of Sept 2008; however, only KF and ES detect the buy-signal in 2009.

Directions for further research deal with stability and robustness of the methodology in the presence of structural breaks (as that occurred in the fall 2008). In particular, one should analyze the sensitivity of coefficients to in-sample data and their efficacy at out-of-sample level. Since parameters should be continuously updated as new data become available, an important issue is the design of T_1 to be used in the optimization of G_T . If time-variability of the coefficients emerges, then adaptive modeling of θ_t, κ_t should be developed.

References

1. Alexandrov T, Bianconcini S, Bee Dagum E, Maass P, McElroy T. A review of some modern approaches to the problem of trend extraction. Research Report n.2008-3, Statistical Division, US Census Bureau, 2008.
2. Andersson E, Bock D, Frisén M. Some statistical aspects for detection of turning points in business cycles. *Journal of Applied Statistics* 2006; **33**: 257-278.
3. Beran J, Feng Y. Local polynomial fitting with long-memory, short memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics* 2002; **54**: 291-311.
4. Canova F. *Methods for Applied Macroeconomic Research*. Princeton University Press: Princeton, NJ, 2007.
5. Chatfield C, Koehler B, Ord K, Snyder D. A new look at models for exponential smoothing. *The Statistician* 2001; **50**: 147-159.
6. Croux C, Gelper S, Mahieu K. Robust exponential smoothing of multivariate time series. *Computational Statistics and Data Analysis* 2009; **86**: 1175-1188.
7. Danthine JP, Girardin M. Business cycles in Switzerland. A comparative study. *European Economic Review* 1989; **33**: 31-50.
8. Ergashev BA. Sequential detection of US business cycle turning points: Performances of Shirayayev-Roberts, CUSUM and EWMA procedures. EconWPA n.0402001, 2004.
9. Fan J, Gijbels I. *Local Polynomial Modelling and its Applications*. Chapman & Hall: London, 1996.
10. Gómez V. Three equivalent methods for filtering finite nonstationary time series. *Journal of Business and Economic Statistics* 1999; **17**: 109-116.
11. Grillenzoni C. ARIMA processes with ARIMA parameters. *Journal of Business and Economic Statistics* 1993; **11**: 235-254.

12. Grillenzoni C. Robust non-parametric smoothing of non-stationary time series. *Journal of Statistical Computation and Simulation* 2009; **79**: 379-393.
13. Harvey AC, Jäger A. Detrending, stylised facts and the business cycle. *Journal of Applied Econometrics* 1993; **8**: 231-247.
14. Harvey AC, Koopman SJ. Unobserved components models in economics and finance. *IEEE Control Systems Magazine* 2009; **8**: 71-81.
15. Hastie T, Tibshirani R. *Generalized Additive Models*. Chapman & Hall: London, 1990.
16. Hodrick RJ, Prescott EC. Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit, and Banking* 1997; **24**: 116.
17. Ljung L. *System Identification: Theory for the User*. Prentice Hall: New Jersey, 1999.
18. Marsh IW. High-frequency Markov switching models in the foreign exchange market. *Journal of Forecasting* 2000; **19**: 123-134.
19. Mise M, Kim TH, Newbold P. On suboptimality of the Hodrick-Prescott filter at time series endpoints. *Journal of Macroeconomics* 2005; **27**: 53-67.
20. Teräsvirta T, Tjøstheim D, Granger CWJ. *Modelling Nonlinear Economic Time Series*. Oxford University Press: New York, 2011.
21. Vander Wiel SA. Monitoring processes that wander using integrated moving average models. *Technometrics* 1996; **38**: 139-151.
22. Van Ruth F, Schouten B, Wekker R. The Statistics Netherlands' business cycle tracer. Report n. 2005-MIC-44, Statistics Netherlands, 2005.
23. Wildi M, Elmer S. Real-time filtering and turning-point detection. Available at <http://www.idp.zhaw.ch/>, 2008.
24. Zellner A, Hong C, Min CK. Forecasting turning points in international output growth rates. *Journal of Econometrics* 1991; **49**: 275-304