ESWA 9297 ARTICLE IN PRESS No. of Pages 13, Model 5G
15 May 2014

Expert Systems with Applications xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A prediction scheme using perceptually important points and dynamic time warping

Q1 Prodromos E. Tsinaslanidis [a,*], Dimitris Kugiumtzis [b]

[a] Canterbury Christ Church University Business School, North Holmes Road, Canterbury, Kent CT1 1QU, UK
[b] Department of Electrical and Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

ARTICLE INFO

Keywords:
Perceptually important points
Dynamic time warping
Nonlinear prediction
Efficient market hypothesis

ABSTRACT

An algorithmic method for assessing statistically the efficient market hypothesis (EMH) is developed based on two data mining tools, perceptually important points (PIPs) used to dynamically segment price series into subsequences, and dynamic time warping (DTW) used to find similar historical subsequences. Then predictions are made from the mappings of the most similar subsequences, and the prediction error statistic is used for the EMH assessment. The predictions are assessed on simulated price paths composed of stochastic trend and chaotic deterministic time series, and real financial data of 18 world equity markets and the GBP/USD exchange rate. The main results establish that the proposed algorithm can capture the deterministic structure in simulated series, confirm the validity of EMH on the examined equity indices, and indicate that prediction of the exchange rates using PIPs and DTW could beat at cases the prediction of last available price.

1. Introduction

The efficient market hypothesis (EMH) gained much attention by the academia since its official introduction in the late 70s (Fama, 1970; Mandelbrot, 1966; Roberts, 1967). Generally, for a specific time period, a market is called efficient if prices fully reflect all available information. Defining historical prices, public available information and inside information as three subsets of the set of all available information, results in three forms of EMH, i.e. weak, semi-strong and strong form. Random walk hypothesis is aligned with the weak-form EMH. According with the above theories the best estimation we can make regarding the future price (return) is the current price (zero) conditioning the historical price path. Dealing here with scalar time series analysis we consider the weak-form EMH.

On the contrary, advocates of technical analysis (TA) assert that it is possible to forecast the future evolution of a financial price series and thus gain systematically abnormal returns by using historical price paths as available information. Thus, TA can be considered as an "economic test" (Campbell, Lo, & MacKinlay, 1997) of the random walk hypothesis and the weak form EMH. Tools of technical analysis can be mainly classified firstly into technical indicators, such as Relative Strength Index (RSI), Moving Averages (MA) and Moving Average Convergence Divergence (MACD), secondly into technical patterns such as "Head and Shoulders" (Osler, 1998; Savin, Weller, & Zvingelis, 2007; Zapranis & Tsinaslanidis, 2010), "Saucers" (Wang & Chan, 2009; Zapranis & Tsinaslanidis, 2012b) and thirdly into candlesticks (Caginalp & Laurent, 1998).[1] Trading strategies can be designed, by adopting the aforementioned tools, which return trading signals as well as support and resistance levels (Osler, 2000; Zapranis & Tsinaslanidis, 2012a). The majority of technical studies examine usually individual or small bundles of technical tools.

In this study, we implement an algorithmic approach in order to assess statistically the null hypothesis of weak-form EMH, by adopting perceptually important points (PIPs) and dynamic time warping (DTW). PIPs are used in order to identify significant points on a financial series. These points segment the series dynamically into subsequences of unequal length. Then our effort focuses in finding similar historical subsequences and then make predictions based on the manner that these best matches evolved in the past. To implement this we employ DTW, which can be used to measure the similarity between two time series of unequal length. By this method we intend to simulate the generalised manner a technician tries to make predictions by finding similar price paths evolutions

* Corresponding author. Tel.: +44 (0) 1227 767700.
E-mail addresses: prodromos.tsinaslanidis@canterbury.ac.uk (P.E. Tsinaslanidis), dkugiu@auth.gr (D. Kugiumtzis).

[1] For a comprehensive description of technical analysis and its tools we indicatively suggest (Achelis, 1995; Bulkowski, 2002; Edwards & Magee, 1997; Pring, 2002).

http://dx.doi.org/10.1016/j.eswa.2014.04.028
0957-4174/© 2014 Published by Elsevier Ltd.

occurred in the past. The technician identifies subjectively, based on own experience, the significant points to define the paths, while in the proposed approach PIPs are used for this segmentation and DTW for measuring the similarity between them.

PIPs were introduced by Chung, Fu, Luk, and Ng (2001) to exploit salient points from a price series and have also been used to identify specific technical patterns in (Fu, Chung, Luk, & Ng, 2007). In the context of data mining, PIPs have been used mainly for purposes of dimension reduction (time series representation), as a dynamic approach for time series segmentation (Fu, Chung, & Ng, 2006; Jiang, Zhang, & Wang, 2007) and for clustering reasons (Fu, Chung, Luk, & Ng, 2004) (for a comprehensive review see (Fu, 2011)).

Dynamic time warping (DTW) is an algorithmic technique mainly used to find an optimal alignment between two given (time-dependent) sequences under certain restrictions (Muller, 2007). First introduced in 1960s, DTW initially became popular in the context of speech recognition (Sakoe & Chiba, 1978), and then in time series data mining, in particular in pattern recognition and similarity measurement (Berndt & Clifford, 1994). We implement DTW for measuring similarities between the target subsequence and historical subsequences of the examined price series, as defined by PIPs. This is actually a subsequence matching problem. Finding salient points and then similar historical subsequences is aligned with the manner a technician tries to exploit information from the past and make forecasts.

The performance of the proposed approach is assessed on simulated time series generated by superimposing a chaotic deterministic time series on a stochastic trend. Subsequently we apply the same approach to real financial series composed of 18 major world equity indices and the GBP/USD currency pair.

The rest of the paper is organized as follows. In Section 2, the methodology is presented, including PIPs, DTW and the prediction scheme. In Section 3, the performance of this approach is assessed on simulated series, and in Section 4 it is applied to financial time series. Finally, discussion and conclusions are given in Section 5.

2. Methodology

2.1. Perceptually important points

First, we present the algorithm constructing PIPs to identify significant points. The algorithm starts by characterizing the first and the last observation as the first two PIPs. Subsequently, it calculates the distance between all remaining observations and the two initial PIPs, and signifies as the third PIP the one with the maximum distance. The fourth PIP is the point that maximizes its distance to its adjacent PIPs (which are either the first and the third, or the third and the second PIP). The algorithm stops when the required by the user number of PIPs is identified.

Three metrics are generally used for the distance in the PIPs algorithm, namely the Euclidean distance (ED) d_E, the perpendicular distance (PD) d_P and the vertical distance (VD) d_V. Let $\{p_1, p_2, \ldots, p_l\}$ be the price time series of length l, and two adjacent PIP $\boldsymbol{x}_t = (t, p_t)$ and $\boldsymbol{x}_{t+T} = (t+T, p_{t+T})$. The Euclidean distance d_E of each of the intermediate points $\boldsymbol{x}_i = (i, p_i)$, for $i \in \{t+1, \ldots, t+T-1\}$ from the two PIPs is defined as

$$d_E(\boldsymbol{x}_i, \boldsymbol{x}_t, \boldsymbol{x}_{t+T}) = \sqrt{(t-i)^2 + (p_t - p_i)^2} + \sqrt{(t+T-i)^2 + (p_{t+T} - p_i)^2}. \quad (1)$$

For the two other distances, we consider first the line connecting the two PIPs $\boldsymbol{x}_t = (t, p_t)$ and $\boldsymbol{x}_{t+T} = (t+T, p_{t+T})$, $z_i = si + c$, and (i, z_i) the points on the line, where the slope is $s = \frac{p_{t+T} - p_t}{T}$ and the constant term is $c = p_t - \frac{p_{t+T} - p_t}{T} t$. Then the perpendicular distance d_P of any intermediate point $x_i = (i, p_i)$, between the two PIPs from the line is

$$d_P(\boldsymbol{x}_i, \boldsymbol{x}_t, \boldsymbol{x}_{t+T}) = \frac{|si + c - p_i|}{\sqrt{s^2 + 1}} \quad (2)$$

and the vertical distance d_V of $\boldsymbol{x}_i$ to the line is

$$d_V(\boldsymbol{x}_i, \boldsymbol{x}_t, \boldsymbol{x}_{t+T}) = |si + c - p_i|. \quad (3)$$

For any of the three distances, denoted collectively d, the new PIP point, $x_i^* = (i^*, p_{i^*})$, is the one that maximizes the distance d at i^*

$$i^* = \text{argmax}_i(d(\boldsymbol{x}_i, \boldsymbol{x}_t, \boldsymbol{x}_{t+T})), \quad (4)$$

where "argmax" stands for the argument of maximum.

Fig. 1 presents five PIPs identified with each of the three distances on the S&P 500 index at two different time periods. Apparently, the distance metrics do not always give the same PIPs.

2.2. Dynamic time warping

Dynamic time warping (DTW) is an efficient scheme giving the distance (or similarity) of two sequences $Q \equiv \{q_1, q_2, \ldots, q_N\}$ and $Y \equiv \{y_1, y_2, \ldots, y_M\}$, where their lengths N and M may not be equal. An example of two sequences Q and Y is illustrated in Fig. 2.

First, a distance between any two components q_n and y_m of Q and Y is defined, e.g. the Euclidean distance $d(q_n, y_m) = (q_n - y_m)^2$, forming the distance (or cost) matrix $\boldsymbol{D} \in \mathbb{R}^{N \times M}$ (see Fig. 3).

The goal is to find the optimal alignment path between Q and Y of minimum overall cost (cumulative distance). A valid path is a sequence of elements $Z \equiv \{z_1, z_2, \ldots, z_K\}$ with $z_k = (n_k, m_k)$, $k = 1, \ldots, K$, denoting the positions in the distance matrix D that satisfy the boundary, monotonicity and step size conditions. The boundary condition ensures that the first and the last element of Z are $z_1 = (1,1)$ and $z_K = (N, M)$, respectively (i.e. the bottom left and the top right corner of D, see Fig. 3). The other two conditions ensure that the path always moves up, right or up and right of the current position in D, i.e. $z_{k+1} - z_k \in \{(1,0), (0,1), (1,1)\}$.

To compute the total distance of each valid path, first the cost matrix of accumulated distances $\tilde{\boldsymbol{D}} \in \mathbb{R}^{N \times M}$ is constructed with initial condition $\tilde{d}(1,1) = d(1,1)$, and accumulated distance for every other element of $\tilde{\boldsymbol{D}}$ defined as

$$\tilde{d}(n,m) = d(n,m) + \min\left\{\tilde{d}(n-1,m), \tilde{d}(n,m-1), \tilde{d}(n-1,m-1)\right\}, \quad (5)$$

where $\tilde{d}(0,m) = \tilde{d}(n,0) = +\infty$ in order to define the accumulated distances for all elements of $\tilde{\boldsymbol{D}}$ (see Fig. 4). At this stage we keep the indexation regarding the adjacent cell with the minimum distance, and then starting from $\tilde{d}(N,M)$ we identify backwards the optimal path. In particular, if the optimal warping path is a sequence of elements $Z^* \equiv \{z_1^*, z_2^*, \ldots, z_K^*\}$ with $z_K^* = (N, M)$, then conditioning on $z_k^* = (n, m)$, we choose z_{k-1}^* as

$$z_{k-1}^* = \begin{cases} (1, m-1), & \text{if } \ n = 1 \\ (n-1, 1), & \text{if } \ m = 1 \\ \text{argmin}\{\tilde{d}(n-1,m-1), \tilde{d}(n-1,m), \tilde{d}(n,m-1)\}, & \text{otherwise.} \end{cases} \quad (6)$$

The process terminates when $n = m = 1$ and $z_k^* = (1,1)$ (Muller, 2007). The optimal path for our example is illustrated in Figs. 3–5 with the white solid line. Having identified the optimal path we can align the initial sequences Q and Y by warping their time axis (Fig. 6).

2.3. The prediction scheme using PIPs and DTW

The prediction scheme combines the use of PIPs and DTW in order to make predictions regarding the future evolution of the

ESWA 9297
15 May 2014
ARTICLE IN PRESS
No. of Pages 13, Model 5G

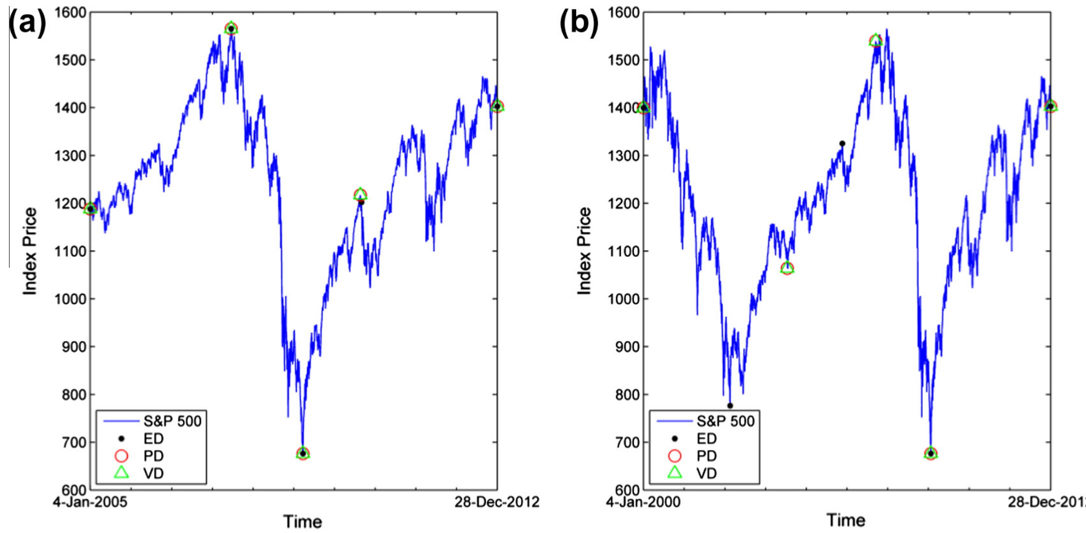

Fig. 1. Five PIPs identified with three different distance metrics (ED, PD and VD) on the S&P 500 index for the time periods from 4-Jan-2005 to 28-Dec-2012 (a), and from 4-Jan-2000 to 28-Dec-2012 (b).

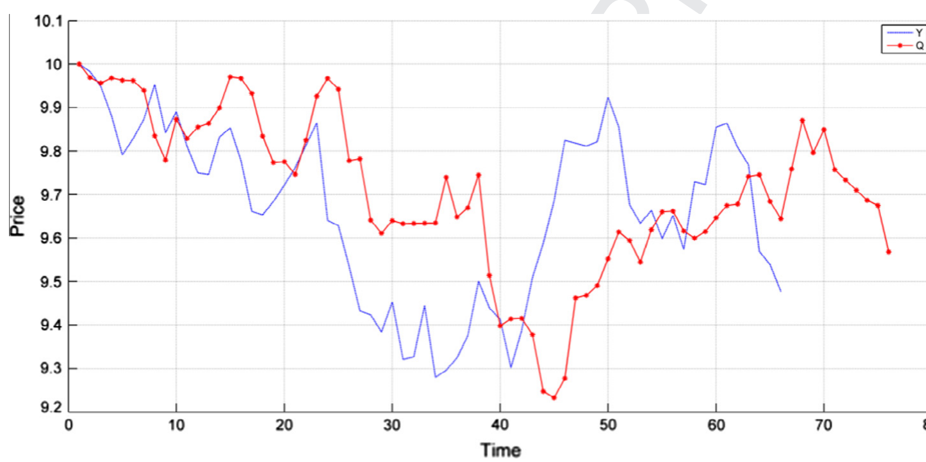

Fig. 2. Two sequences Q and Y of different lengths.

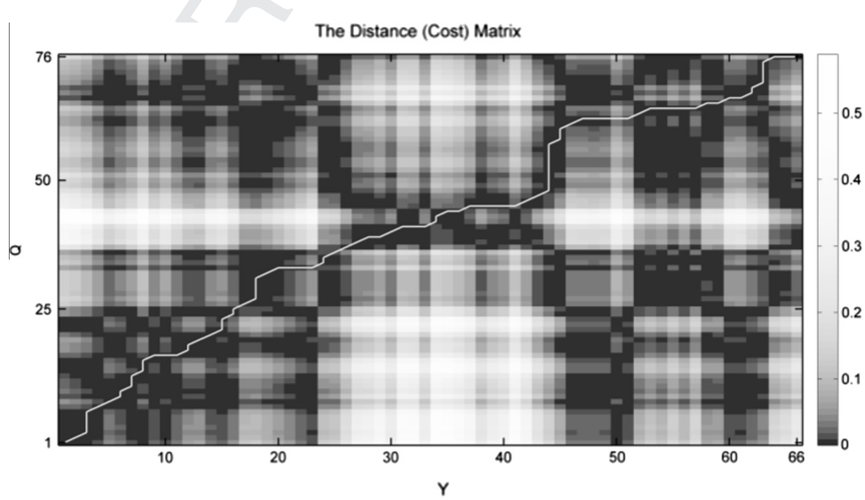

Fig. 3. Colormap of the distance (cost) matrix of sequences Q and Y. The white solid line is the optimal warping path (discussed later in this section).

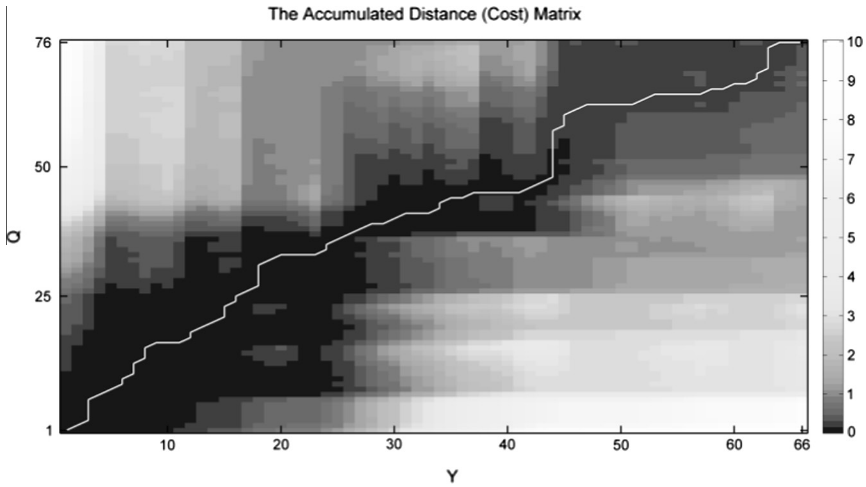

Fig. 4. Colormap of the accumulated distance (cost) matrix of sequences Q and Y, and the optimal warping path (white line).

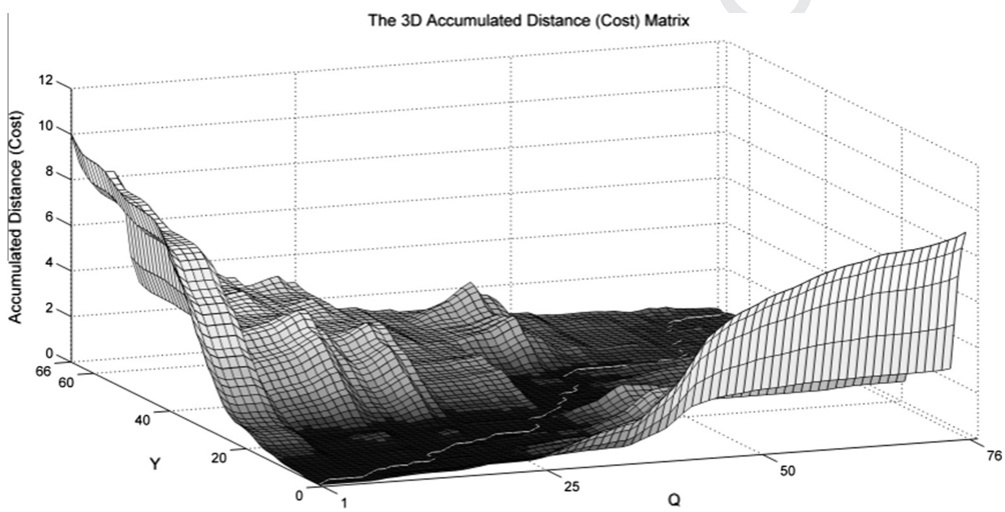

Fig. 5. 3D illustration of Fig. 4.

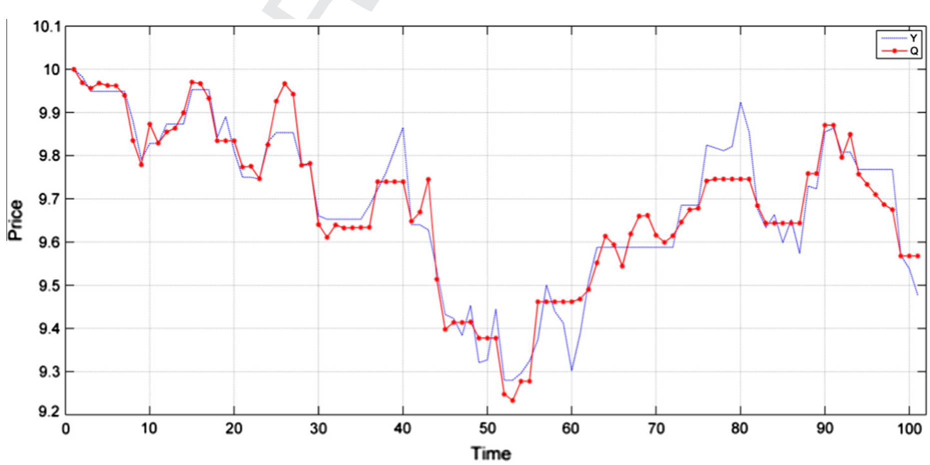

Fig. 6. Sequences Q and Y aligned with DTW.

series. First, PIPs are constructed to dynamically segment the examined time series. Then for each target time, the DTW algorithm is called to measure the similarity between the target (query) subsequence and each of the past subsequences. The observations ahead of the most similar subsequences are used to make predictions ahead the target time. The prediction scheme is explained in detail below.

For a given target time u, $u_0 \leqslant u < l$, where l is the time series length and u_0 is the starting time of the test set, the first u observations form the training set on which the PIPs are computed with

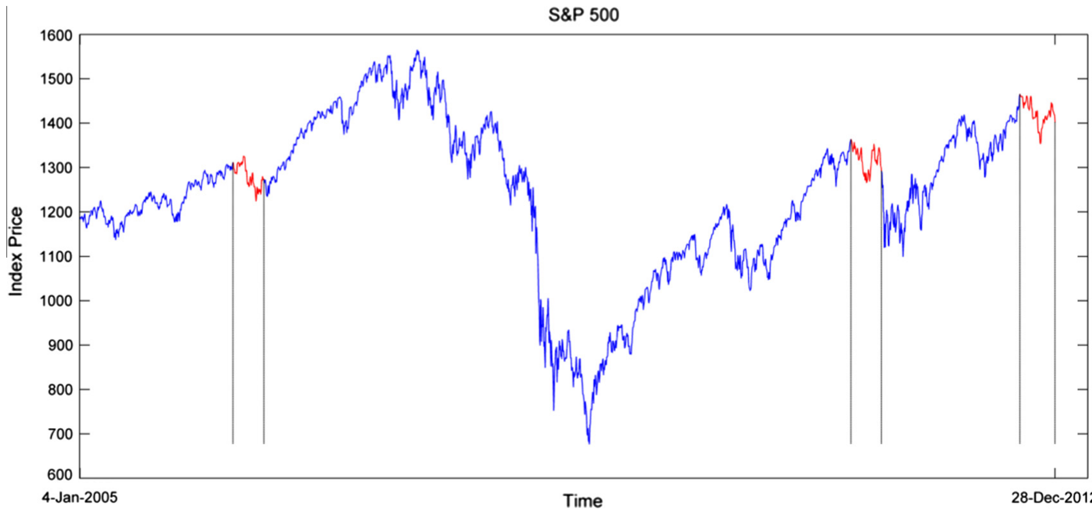

Fig. 7. Two most similar subsequences identified in S&P 500 index (from 4-Jan-2005 to 28-Dec-2012) with DTW. The target sequence (Q) is from 14-Sep to 28-Dec 2012, and the two most similar subsequences identified Y_1 and Y_2 are from 5-Apr to 6-Jul 2006 and from 29-Apr to 29-Jul 2011 respectively. Y_2 is the best match and is the one presented restructured with Q at Fig. 2. Six ED-PIPs were used to define the subsequences in this graph.

one of the three different distance measures presented in Section 2.1. The objective is to make predictions at T times ahead the target time u, i.e. $u+1,\ldots,u+T$. The number of estimated PIPs is set so as to achieve an average time interval between successive PIPs, which here is set to 10 days. The target subsequence is defined by the last two PIPs, and the past subsequences are defined by all other pairs of successive PIPs. The subsequences may be at different magnitude levels, and therefore they are translated so that their starting value matches a given value. In view of the financial applications, this value is set to 10. All subsequences to be compared are first converted to logarithmic returns and then they are restructured to a non-stationary series with an initial price[2] of 10. The expressions for the logarithmic return and the restructured price $\bar{p}_{t+T}$ at time $t+T$ for a predefined by the user initial price p_t^* (in our case $p_t^*=10$) are given below.

$$r_{t+1} = \ln(p_{t+1}/p_t), \tag{7}$$

$$\bar{p}_{t+T} = \exp\left\{\sum_{i=1}^{T} r_{t+i} + \ln(p_t^*)\right\} \tag{8}$$

The total average cost (TAC) of the optimal path derived by DTW for each comparison is

$$\mathrm{TAC} = d_{Z^*}/K, \tag{9}$$

where $d_{z^*} = \tilde{d}(N,M)$ is the total cost of the optimal path of length K. After the comparison between the target sequence and all its candidates has been made TACs are scaled between 0 and 1. Denoting $\mathbf{TAC} = [\mathrm{TAC}_1, \mathrm{TAC}_2, \ldots, \mathrm{TAC}_J]$ the vector containing TACs for each of the J comparisons, **STAC** contains the corresponding scaled total average costs, each defined as

$$\mathrm{STAC}_j = \frac{\mathrm{TAC}_j - \min(\mathbf{TAC})}{\max(\mathbf{TAC}) - \min(\mathbf{TAC})}, \tag{10}$$

where $j = 1,2,\ldots,J$. Subsequently, we define as the most similar subsequences those that have STAC below a given threshold (say 0.01). For example, Fig. 7 illustrates the two best matches for a specific target subsequence, when six ED-PIPs were used.

The next stage is to make predictions, say for the next 10 days. To do so, we restructure the N best matches plus the prices observed in the following 10 trading days by using (7) and (8) and by setting the initial price the last price of the target sequence. The prediction is the weighted average of the two, 10-days restructured price paths, where the weight is the similarity of each subsequence to the target sequence. The greater the similarity the lower the STAC, so we have

$$\hat{p}_{t(T)} = \sum_{i=1}^{N}\left[\frac{1-\mathrm{STAC}_i}{\sum_{i=1}^{N}(1-\mathrm{STAC}_i)}\right]\bar{p}_{i,t+T}. \tag{11}$$

In (11), $\hat{p}_{t(T)}$ is the price prediction we make for T steps ahead conditioning the last available price at time t, STAC_i is the scaled total average cost of the ith similar subsequence out of the N most similar subsequences occurred by the aforesaid process, and $\bar{p}_{i,t+T}$ is the restructured price T steps after the last observation of the ith similar subsequence. We chose to base the prediction on a few best matches rather than one best match because the prediction tends to be more stable and not affected by noise.

Subsequently, the target time is increased by one and the procedure presented in this Section is repeated until $u = l - 10$. We leave out the last 10 observations in order to assess the predictions made for these days. Finally, the predictive performance is assessed by two measures: the normalized (by) persistence root mean square error (NPRMSE) and the independent prediction of change in direction (IPOCID) (Zapranis & Refenes, 1999):

$$\mathrm{NPRMSE}(T) = \sqrt{\frac{\sum\left(\hat{p}_{t(T)} - p_{t+T}\right)^2}{\sum\left(p_t - p_{t+T}\right)^2}}, \tag{12}$$

$$\mathrm{IPOCID}(T) = \frac{100}{n-1}\sum_{t=1}^{n} d_{t(T)}, \tag{13}$$

$$d_{t(T)} = \begin{cases} 1, & \text{if} \quad (p_{t+T} - p_{t+T-1})(\hat{p}_{t(T)} - \hat{p}_{t(T-1)}) > 0 \\ 0, & \text{if} \quad (p_{t+T} - p_{t+T-1})(\hat{p}_{t(T)} - \hat{p}_{t(T-1)}) \leqslant 0 \end{cases}, \tag{14}$$

where the sum runs over the times in the test set, $n = l - 10 - u_0$.

[2] The initial price of ten is set arbitrarily, but it does not affect the results of this experiment. Ordinarily, two sequences are being normalized before they are compared with DTW. However, in this paper we are about to use financial data. Normalizing or displacing the initial series would change the daily returns of the price series. By following the proposed restructuring process we implicitly state that two identical (or similar) sequences of returns should represent the same pattern regardless of the price level observed. It is also noteworthy that setting the initial value of the restructured series at a different level would affect the total cost calculated by the DTW, but not the optimal path. Thus the scheme finding the most similar historical subsequences to the target sequence is not affected by the initial value.

Please cite this article in press as: Tsinaslanidis, P. E., & Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.04.028

By construction, NPRMSE $\geqslant 0$ and compares the predictive performance of the proposed algorithm with that of a naïve, benchmark model (12). NPRMSE can be used to assess the weak-form EMH since the benchmark model used for comparison makes predictions with the last observed price. For a given test sample we can identify three different scenarios:

- NPRMSE is significantly lower than 1 which indicates that the proposed algorithm outperforms the benchmark predictive model. Under this scenario, the lower its value the better the performance of the proposed prediction scheme is.
- NPRMSE is significantly greater than 1 which denotes that the benchmark model outperforms the examined predictive model.
- Finally, if NPRMSE is not significantly different than 1, we can infer that the proposed algorithm performs similarly with predictions made with the last available price.

For instance, when estimating NPRMSE(1), under the second and third scenario, we cannot reject the weak form EMH, since conditioning the historical price path, the best expectation of tomorrow's price is today's price.

IPOCID (13–14) is generally used in applications where predictions are made on a price path, and measures the ability of the examined model to predict changes regardless their size. It is expressed as a percentage and values towards 100% imply good predictive performance. In particular, it can be argued that values statistically significant greater than 50% imply that the proposed algorithm predicts changes in directions better than predicting with a "fair" level of 50%.

3. A simulation experiment

The purpose of this simulation experiment is to assess the predictive performance of the presented methodology on simulated price series. Each simulated time series is composed as a weighted sum of two time series, a stochastic trend with a weight of a% and deterministic time series with weight of $(1 - a)$%. The stochastic trend is simply a random walk and the input white noise has a standard deviation that is one tenth of the standard deviation of the deterministic time series. The deterministic time series is generated by the delay differential equation of Mackey–Glass with delay $\Delta = 30$. The Mackey–Glass delay differential equation defines a deterministic system that can have chaotic behavior of a complexity determined by the parameter of delay Δ (Mackey & Glass, 1977). For $\Delta = 30$, the fractal dimension of the chaotic attractor of the system is about 3 (Grassberger & Procaccia, 1983). The time series is obtained at a sampling time of 20 time points that produces discrete-like data, and it is used here to regard an hypothesis of an underlying chaotic deterministic mechanism mixed with stochastic trend at a rate determined by the parameter a.

Fig. 8 presents the distributions of the NPRMSE and the IPOCID of the 10 time step ahead predictions for three different values of the parameter a (0.5, 0.75 and 1), respectively. The sample distributions are shown as boxplots (box edges are the 25th and 75th percentiles, the horizontal line in the box denotes the median, the whiskers extend to the minimum and maximum of the sample if no outliers are detected, otherwise the outliers are singled out and denoted by crosses). Subsequences are determined by 2 and 3 subsequent PIPs and their similarity is quantified with the ED distance. For comparison, in Fig. 9 the same statistical results are shown for the prediction scheme that uses the same number of breakpoints as the number of PIPs but segmenting the time series at constant intervals, and this scheme is denoted CI. This prediction is equivalent to the nearest neighbor prediction after state space reconstruction using the constant interval as the embedding window (Kugiumtzis, 2002).

As shown in Fig. 8 when there is 50% contribution of the deterministic process, the proposed method can capture this information and exhibits some predictability at the first four time steps ahead as NPRMSE is lower than one. As the proportion of the stochastic trend increases, predictive performance worsens. At the extreme case of $a = 1$, the time series is actually a random walk and predicting with the current price outperforms our methodology. In addition, IPOCID is significantly greater than the fair level of 50%, and it reasonably decrease as we try to predict more days ahead. Again when $a = 1$, IPOCID fluctuates around the 50% level indicating the complete stochastic behavior of the simulated series. It is worth to mention that PIPs do not add value in the predictive performance as compared to the scheme CI of constant time intervals. On the contrary, for CI the NPRMSE is marginally lower than this obtained when PIPs are used. A possible explanation for this is that the chaotic time series does not have the signature assumed in the approach of PIPs, i.e. time varying patterns characterized by important breakpoints, but rather varying patterns at a time window that regards the time of orbits in the state space (Kugiumtzis, 1996). This system type is to be contrasted to the real financial series, where the performance of PIPs and CI differ, as will be shown in the next Section. Thus our limited simulation study indicates that real price series, affected by exogenous parameters and exhibiting salient points, cannot be explained as chaotic time series with stochastic trend.

4. Empirical results

The presented methodology was applied to two datasets of real financial price series. The first one consists of 18 major world indices (Table 1). Adjusted daily closing prices for the period 4-Jan-2000 until 28-Dec-2012 were downloaded from Bloomberg database. We applied a filter similar with the one adopted in (Lo, Mamaysky, & Wang, 2000; Marshall, Qian, & Young, 2009; Zapranis & Tsinaslanidis, 2012a) and remaining missing values were filled with linear interpolation. The second dataset consists of daily prices of the British pound to US dollar (GBP/USD) exchange rate for the period 1971–2012.

We considered the prediction schemes of PIPs with all three distance measures, ED, PD and VD, as well as the prediction scheme with constant time intervals (CI). All four prediction schemes were applied to the first data set using 2 and 3 breakpoints (PIPs for ED, PD and VD and points at constant intervals for CI). The prediction summary statistics NPRMSE and IPOCID were calculated at windows of 50 days for one step ahead prediction. Specifically, the first 200 observations are used to find similar subsequences and make prediction one day ahead. Then the training window in increased by one day, new PIPs are found and prediction is made for the time point 202. This procedure is repeated until the prediction at time point 250, and then NPRMSE and IPOCID are calculated on the basis of these 50 predictions. Subsequently, the training window slides by 50 days and the aforementioned procedure is repeated. Thus for a price series of length l, $(l - 200)/50$ NPRMSE and IPOCID are calculated, each characterizing the prediction at a period of 50 days given the past 200 days.

The results on NPRMSE and IPOCID suggest that for the first dataset weak-form EMH holds. The vast majority of NPRMSE values are above unity, which indicates that predicting with the last available price outperforms predictions made with any of the prediction schemes. In addition, IPOCID fluctuates around the 50% fair level of prediction. For example, the profiles of NPRMSE and IPOCID for the index S&P500 in Figs. 10 and 11 are at the levels above one and 0.5, respectively, bearing strong similarity to the results on the simulated pure stochastic paths. However, the NPRMSE from PIPs and any of the distance measures tend to be lower than the

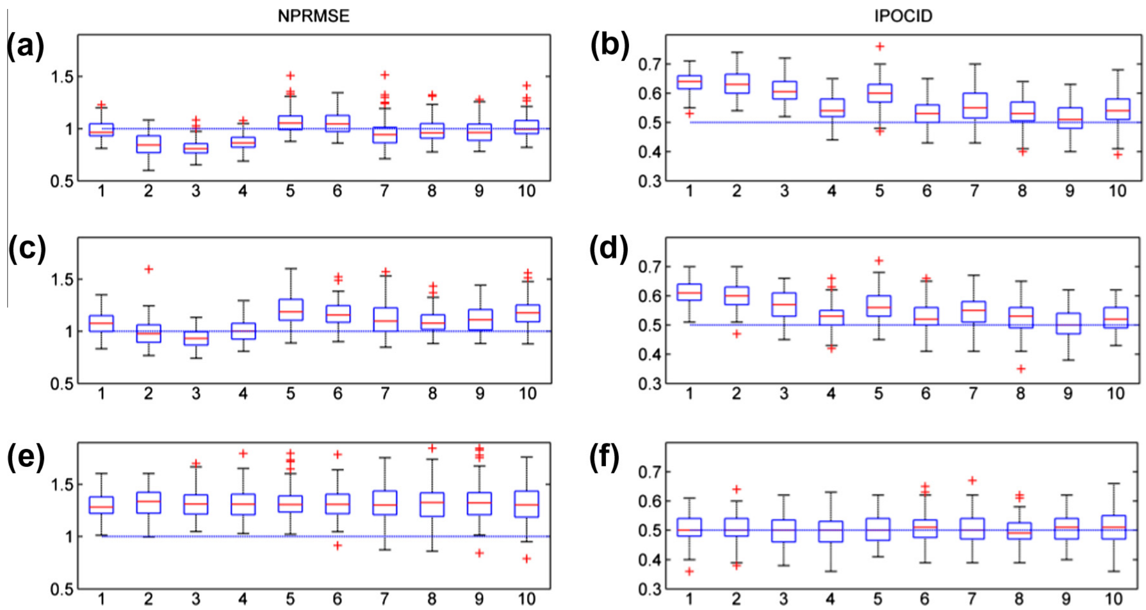

Fig. 8. NPRMSE and IPOCID calculated for 100 simulated price paths for predictions of 10 time steps ahead. Parameters: two PIPs with distance type ED. a and b $a = 0.5$, c and d $a = 0.75$ and e and f $a = 1$.

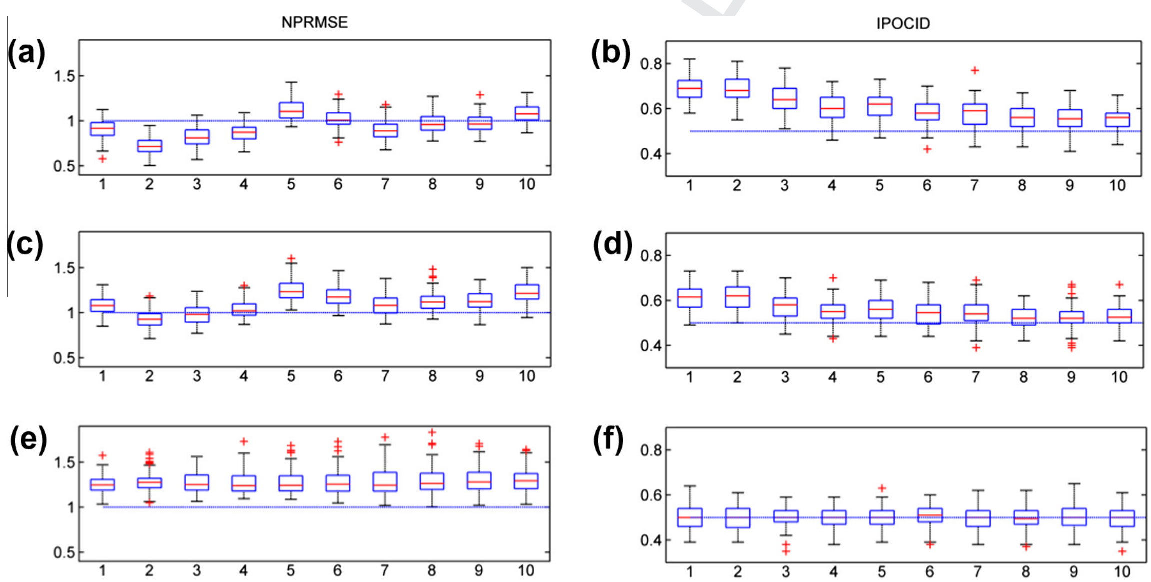

Fig. 9. NPRMSE and IPOCID calculated for 100 simulated price paths for predictions of 10 time steps ahead. Parameters: CI with two breakpoints. a and b $a = 0.5$, c and d $a = 0.75$ and e and f $a = 1$.

Table 1
Major world indices.

idx_i	Index name	idx_i	Index name
Panel A: Americas		*Panel B: EMEA*	
idx_1	DOW JONES (INDU)	idx_7	EURO Stoxx (SX5E)
idx_2	S&P 500 (SPX)	idx_8	FTSE 100 (UKX)
idx_3	NASDAQ (CCMP)	idx_9	CAC 40 (CAC)
idx_4	TSX (SPTSX)	idx_{10}	DAX (DAX)
idx_5	MEX IPC (MEXBOL)	idx_{11}	IBEX 35 (IBEX)
idx_6	IBOVESPA (IBOV)	idx_{12}	FTSE MIB (FTSEMIB)
Panel C: Asia/Pacific		idx_{13}	AEX (AEX)
idx_{16}	NIKKEI (NKY)	idx_{14}	OMX STKH30 (OMX)
idx_{17}	HANG SENG (HSI)	idx_{15}	SWISS MKT (SMI)
idx_{18}	ASX 200 (AS51)		

Note: In parenthesis the Bloomberg ticker is presented for every index.

Please cite this article in press as: Tsinaslanidis, P. E., & Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.04.028

NPRMSE from CI and this is observed both when 2 or 3 breakpoints are used (see Fig. 10). At some few time periods the NPRSE from PIPs is even smaller than one, whereas CI gives NPRMSE always well above one. Further we compare the distributions of NPRMSE and IPOCID for S&P500 over the whole time record in the four prediction schemes. Since these distributions are not always normal we apply two-sample, one tailed, Kolmogorov–Smirnov tests (larger and smaller tail for the NPRMSE and IPOCID ratios respectively). The p-values of the K–S tests are tabulated in Table 2. A small p-value for a prediction scheme i at the row and a prediction scheme j at the column of the table denotes that the cumulative density function (cdf) of the NPRMSE (IPOCID) of the i scheme is significantly at a larger (smaller) level than the respective cdf of the j scheme. The p-values smaller than 0.1 are highlighted in Table 2, indicating that the prediction scheme of PIPs with ED or PD distance measure provide better predictions than the predictions scheme CI when 2 PIPs are used, and CI and PIPs with VD when 3 PIPs are used. For IPOCID, these differences are less significant but still PIPs perform better that CI.

Similar results were obtained with many other of the 18 indices. To provide summary results, we count the financial indices for which the difference for a pair of prediction schemes is found statistically significant (p-value < 0.1). The scores for all pairs of prediction schemes and for NPRMSE and IPOCID, as well as 2 and 3 breakpoints, are given in Table 3. In particular, NPRMSE indicates that the prediction schemes with PIPs outperform the CI scheme,

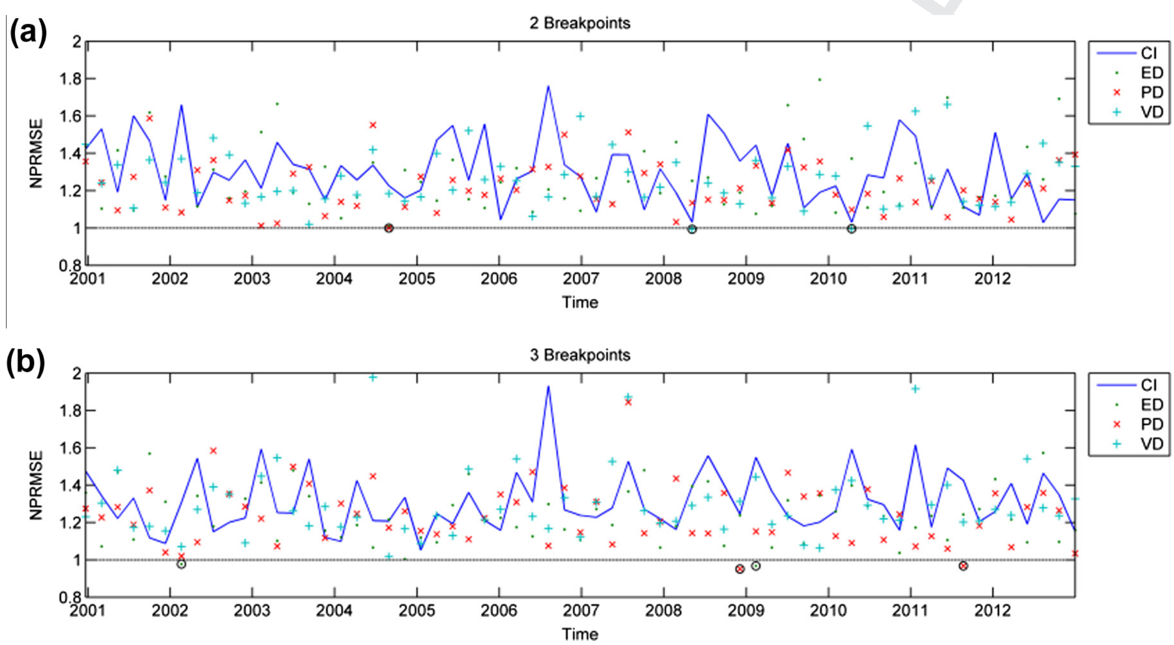

Fig. 10. The profile of NPRMSE over the whole time record of the S&P500 index with the four prediction schemes as shown in the legend and for 2 breakpoints (a) and 3 breakpoints (b).

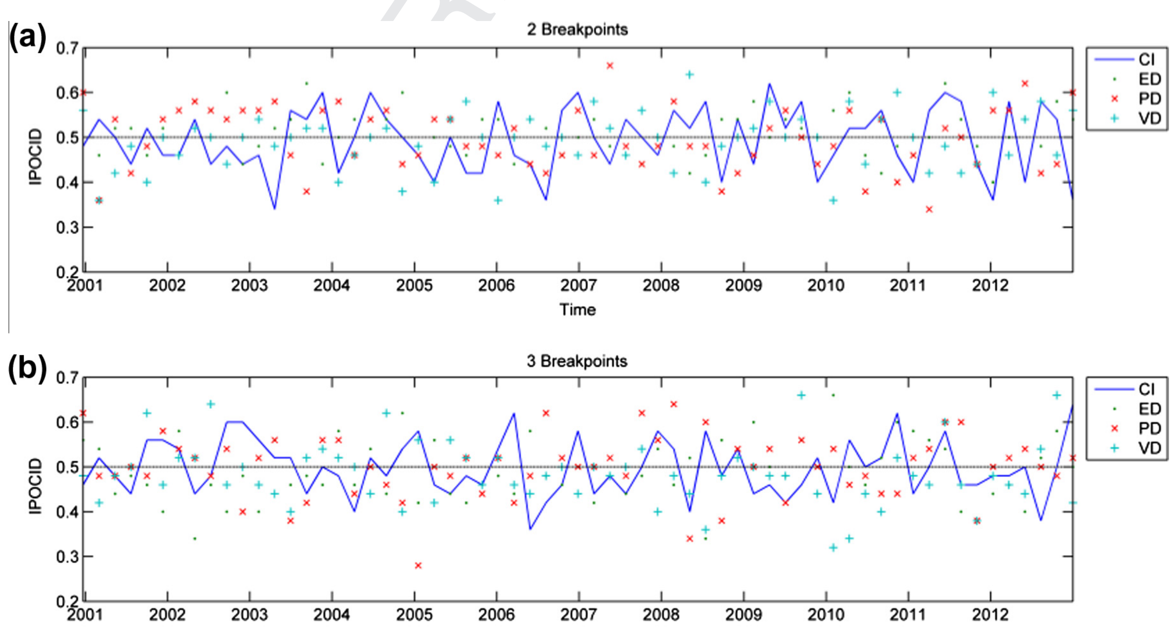

Fig. 11. As for Fig. 10 but for IPOCID.

Please cite this article in press as: Tsinaslanidis, P. E., & Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.04.028

Table 2
Q2 Two sample, one tailed, Kolmogorov–Smirnov tests for S&P500 index.

	CI	ED	PD	VD		CI	ED	PD	VD
Panel A: 2 breakpoints, NPRMSE p-value					*Panel B: 3 breakpoints, NPRMSE p-value matrix*				
CI	1.0000	0.7659	1.0000	0.7659	CI	1.0000	1.0000	1.0000	0.9355
ED	**0.0597**	1.0000	0.7659	0.2591	ED	**0.0081**	1.0000	0.7659	**0.0381**
PD	**0.0381**	0.2591	1.0000	0.1888	PD	**0.0024**	0.5487	1.0000	**0.0081**
VD	0.2591	0.7659	0.9355	1.0000	VD	0.5487	0.9835	0.9355	1.0000
Panel C: 2 breakpoints, IPOCID p-value					*Panel D: 3 breakpoints, IPOCID p-value*				
CI	1.0000	0.6591	0.8607	0.5487	CI	1.0000	0.6591	0.8607	0.3440
ED	**0.0906**	1.0000	0.2591	0.2591	ED	0.9835	1.0000	0.9355	0.2591
PD	0.8607	0.3440	1.0000	0.2591	PD	0.4418	**0.0906**	1.0000	**0.0906**
VD	0.5487	0.9835	0.6591	1.0000	VD	0.9355	0.8607	0.9355	1.0000

as they decrease significantly the NPRMSE for most of the financial indices, with a maximum score of 15 out of 18 financial indices for the PD distance measure and 2 PIPs. This superiority exists but is less apparent regarding the IPOCID.

Another finding extracted from these results is that if the PIPs identification procedure simulates the manner technicians look for important points on a price series, this procedure adds value as a preliminary step in pattern recognition procedure.

We apply the same analysis on a second dataset consisting of daily prices of the GBP/USD exchange rates for the period 1972–2012. The algorithm's predictive behavior is superior on foreign exchange markets compared to that on stock markets. This superiority is also reported by Park and Irwin (2007) who mention that while technical trading strategies failed to yield economic profits in US stock markets after the 1980s, they generated economic profits in foreign exchange markets over the last few decades. However, this predictability seems to decline or vanish since the early 1990s. Here we focus on one particular price series and we scrutiny further our analysis by adopting a number of different parameters' combinations. In particular, we are using 4 different sizes of test samples $w_{train} = \{200, 400, 600, 800\}$, 5 different number of breakpoints $\{2, 3, 4, 5, 6\}$ and we introduce 4 different similarity thresholds $t_{sim} = \{1, 2, 3, 4\}$. By the introduction of this new parameter, we allow the algorithm to make predictions on an iteration only if the number of most similar historical subsequences (J) equals to or is greater than t_{sim} (i.e. $J \geqslant t_{sim}$). The last parameter used is the method under which the examined price series is being segmented, and it takes three variables $\{CI, ED, PD, VD\}$.

In addition we implement two statistical assessments to examine whether the values of NPRMSE (IPOCID) generated on each 50-days window are significantly lower (greater) than 1 (0.5). As we already described in Section 2.3 NPRMSE compares the predictive performance of the proposed algorithm (say method a) with that of a benchmark model where predictions are made with the last available price (method b). The NPRMSE measure is actually the ratio of the two Root Mean Squared Errors (RMSE), produced under the two prediction schemes:

$$\mathrm{NPRMSE}(T) = \sqrt{\frac{\sum(\hat{p}_{t(T)} - p_{t+T})^2}{\sum(p_t - p_{t+T})^2}} = \frac{\mathrm{RMSE}(T)_a}{\mathrm{RMSE}(T)_b}. \quad (15)$$

Testing whether a particular NPRMSE is significantly lower than 1 is equivalent to testing whether $\mathrm{RMSE}(T)_a < \mathrm{RMSE}(T)_b \iff \mathrm{RMSE}(T)_a - \mathrm{RMSE}(T)_b < 0$. But RMSE is a monotonic transformation of Mean Squared Error (MSE) so alternatively we can assess whether $\mathrm{MSE}(T)_a - \mathrm{MSE}(T)_b < 0$. But the difference between two MSEs is the mean of the differenced squared errors (16).

$$\mathrm{MSE}(T)_a - \mathrm{MSE}(T)_b = \frac{1}{n}\sum_{i=1}^{n} e_{i,a}^2 - \frac{1}{n}\sum_{i=1}^{n} e_{i,b}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(e_{i,a}^2 - e_{i,b}^2\right), \quad (16)$$

where $e_{i,a}^2$ and $e_{i,b}^2$ are the squared errors produced by the two methods for the ith prediction made out of n total predictions. However, instead of using an one tailed, paired t-test to assess whether $\mathrm{MSE}(T)_a - \mathrm{MSE}(T)_b < 0$, we implement the Diebold–Mariano test (DM-test hereafter) which allows forecast errors to be non-Gaussian, nonzero mean, serially correlated and contemporaneously correlated (Diebold & Mariano, 1995).[3] Under the null of equal predictive accuracy between the two methods the DM statistic follows a standard normal distribution.

We also implement Bernoulli trials in order to assess whether IPOCID values are statistically, significant greater than 0.5. In particular the term $\sum_{t=1}^{n} d_{t(T)}$ returns the number of successful cases out of n trials which follow a Binomial distribution.

Table 4 presents the breakdown of significant cases generated by using different parameters for a given size of training sample. Significant cases tabulated are those signified by adopting the DM-test for a significance level of 5%. For instance, when $w_{train} = 200$ the examined currency pair is split to 215 nonoverlapping 50-days windows where the predictive performance of the algorithm is being assessed on every possible parameter combination. Four different methods for the series segmentation, five different number of breakpoints and four different similarity thresholds result in 80 different parameter combinations. Totally, the number of 50-days windows, where the algorithm is called to make predictions, is $n_w = 17{,}200$. However the algorithm makes predictions for $n_p = 11{,}504$ cases due to the adoption of the t_{sim} parameter, and outperforms significantly the benchmark model at $n_s = 198$ cases giving a "success" ratio of 1.72%. The breakdown of n_s and n_p according to each parameter's value is being presented in Table 4.For example, out of 198 significant cases, 47, 58, 47 and 46 occurred by using CI, ED, PD and VD respectively. However, changing the value of a parameter affects also the number of cases where predictions were made. For instance, using ED instead of CI, increased the number of significant cases from 47 to 58 but also increased the cases where predictions were made from 2879 to 2979. For this reason, Table 4 also provides the equivalent percentage which shows the number of significant cases out of 100 cases where predictions were made. By using the other two types of PIPs (PD, VD) the number of significant cases was relatively the same, whilst the number of cases where predictions were made was decreased.

By increasing the number of breakpoints, the algorithm makes fewer predictions. This can be attributed to the fact that candidates have larger length and thus the chance of finding similar subsequences reduces. However using 4 and 5 breakpoints increased the number of significant cases resulting in the enhancement of the corresponding ratios. When we increased the size of the training sample, using more breakpoints affect negatively the generated

[3] One of the drawbacks of the proposed method that we subsequently acknowledge is its computational expensiveness. Dealing with this problem will allow the user to assess the predictive performance with bootstrap-based evaluative procedures (Efron, 1979; Efron, 1982) or surrogate data analysis (Theiler, Eubank, Longtin, Galdrikian, & Farmer, 1992) that we consider more appropriate evaluation techniques for the performance of the proposed predictive scheme.

Table 3
Significant differences of prediction schemes for 18 major world indices.

	CI	ED	PD	VD		CI	ED	PD	VD
Panel A: 2 breakpoints, NPRMSE, counts					*Panel B: 3 breakpoints, NPRMSE, counts*				
CI	0	2	1	4	CI	0	1	0	3
ED	11	0	0	4	ED	9	0	1	2
PD	15	9	0	9	PD	9	3	0	2
VD	10	3	0	0	VD	10	2	3	0
Panel C: 2 breakpoints, IPOCID, counts					*Panel D: 3 breakpoints, IPOCID, counts*				
CI	0	1	2	3	CI	0	2	1	3
ED	2	0	3	2	ED	0	0	0	0
PD	3	2	0	1	PD	3	3	0	3
VD	0	1	1	0	VD	5	2	1	0

Table 4
The breakdown of identified significant cases according to DM-test for a significance level of 5%.

Panel A: $w_{train} = 200$, $n_s = 198$, $n_p = 11{,}504$, $n_s/n_p = 1.72\%$, $n_w = 215 \times 80 = 17{,}200$					
PIPs type	47/2879	58/2979	47/2822	46/2824	
{CI, ED, PD, VD}	1.63%	1.95%	1.67%	1.63%	
Breakpoints	38/2399	34/2356	42/2315	53/2236	31/2198
{2, 3, 4, 5, 6}	1.58%	1.44%	1.81%	2.37%	1.41%
Threshold	0/4300	41/4266	128/2462	29/476	
{1, 2, 3, 4}	0.00%	0.96%	5.20%	6.09%	
Panel B: $w_{train} = 400$, $n_s = 206$, $n_p = 14{,}904$, $n_s/n_p = 1.38\%$, $n_w = 211 \times 80 = 16{,}880$					
PIPs type	55/3725	58/3812	46/3684	47/3683	
{CI, ED, PD, VD}	1.48%	1.52%	1.25%	1.28%	
Breakpoints	46/3018	41/3034	44/2995	43/2967	32/2890
{2, 3, 4, 5, 6}	1.52%	1.35%	1.47%	1.45%	1.11%
Threshold	0/4220	10/4219	70/3983	126/2482	
{1, 2, 3, 4}	0.00%	0.24%	1.76%	5.08%	
Panel C: $w_{train} = 600$, $n_s = 168$, $n_p = 16{,}003$, $n_s/n_p = 1.05\%$, $n_w = 207 \times 80 = 16{,}560$					
PIPs type	39/3978	38/4040	47/3992	44/3993	
{CI, ED, PD, VD}	0.98%	0.94%	1.18%	1.10%	
Breakpoints	41/3210	40/3233	35/3200	24/3184	28/3176
{2, 3, 4, 5, 6}	1.28%	1.24%	1.09%	0.75%	0.88%
Threshold	0/4140	8/4140	44/4110	116/3613	
{1, 2, 3, 4}	0.00%	0.19%	1.07%	3.21%	
Panel D: $w_{train} = 800$, $n_s = 130$, $n_p = 16{,}009$, $n_s/n_p = 0.81\%$, $n_w = 203 \times 80 = 16{,}240$					
PIPs type	39/3997	23/4023	34/3994	34/3995	
{CI, ED, PD, VD}	0.98%	0.57%	0.85%	0.85%	
Breakpoints	31/3206	26/3211	31/3205	22/3197	20/3190
{2, 3, 4, 5, 6}	0.97%	0.81%	0.97%	0.69%	0.63%
Threshold	1/4060	3/4060	29/4043	97/3846	
{1, 2, 3, 4}	0.02%	0.07%	0.72%	2.52%	

Table 5
Further breakdown of identified significant cases according to DM-test for a significance level of 5%, $w_{train} = 200$ and $t_{sim} = \{3, 4\}$.

	2	3	4	5	6	Total
Panel A: $t_{sim} = 3$						
CI	9(7.03%)	8(6.25%)	2(1.56%)	6(4.69%)	6(4.69%)	31(24.22%)
ED	7(5.47%)	4(3.13%)	5(3.91%)	13(10.16%)	6(4.69%)	35(27.34%)
PD	3(2.34%)	8(6.25%)	9(7.03%)	9(7.03%)	2(1.56%)	31(24.22%)
VD	3(2.34%)	8(6.25%)	9(7.03%)	9(7.03%)	2(1.56%)	31(24.22%)
Total	22(17.19%)	28(21.88%)	25(19.53%)	37(28.91%)	16(12.5%)	128(100%)
Panel B: $t_{sim} = 4$						
CI	2(6.9%)	2(6.9%)	2(6.9%)	2(6.9%)	0(0%)	8(27.59%)
ED	3(10.34%)	2(6.9%)	1(3.45%)	1(3.45%)	5(17.24%)	12(41.38%)
PD	0(0%)	0(0%)	2(6.9%)	2(6.9%)	1(3.45%)	5(17.24%)
VD	0(0%)	0(0%)	2(6.9%)	2(6.9%)	0(0%)	4(13.79%)
Total	5(17.24%)	4(13.79%)	7(24.14%)	7(24.14%)	6(20.69%)	29(100%)

Note: In parenthesis the corresponding percentages are illustrated.

success ratios. Intuitively, we can argue that short-term historical patterns are more likely to have short-term influence in the price evolution.

Finally, the last parameter used (t_{sim}) seems to be the most consistent of all. Increasing the similarity threshold increases the number of significant cases whilst decreasing the cases where predictions are made. This is consistent to all different sizes of training sample used with an exception of $w_{train} = 200$, $t_{sim} = \{3, 4\}$ where the number of significant cases dropped from 128 to 29. However, even in this exception due to the simultaneous significant decrease of the number of prediction the performance of the algorithm increased from 5.2% to 6.09%. It is also worth to mention that this enhancement is more apparent when the training sample is smaller.

Increasing the size of the training set exacerbates the performance of the proposed algorithm. This can be attributed to the fact that price evolution, after the completion of similar historical patterns, behaves less similarly as the time interval between these patterns increases. For example, adopting a training sample of 800 days means that the algorithm makes predictions with similar historical patterns that can be spotted three years before. Thus, we can infer that if there is any repetitive behavior in the price path evolution, there are more chances to identify it to the near past. Several surveys similarly report that trading profits from earlier profitable technical trading rules seem to vanish in more recent years (Olson, 2004; Sullivan, Timmermann, & White, 1999; Sullivan, Timmermann, & White, 2003). Perhaps, when using larger training samples which exacerbates the algorithms performance, a greater value for the t_{sim} should be adopted to compensate for this deterioration.

Subsequently, we focus on the best performing case where a train sample of 200 days and a similarity threshold of 3 and 4 are being used. Table 5 presents the allocation of the 128 and 29 successful cases to the segmentation method and the number of breakpoints used. Most significant cases are generated when the ED distance measure is used.

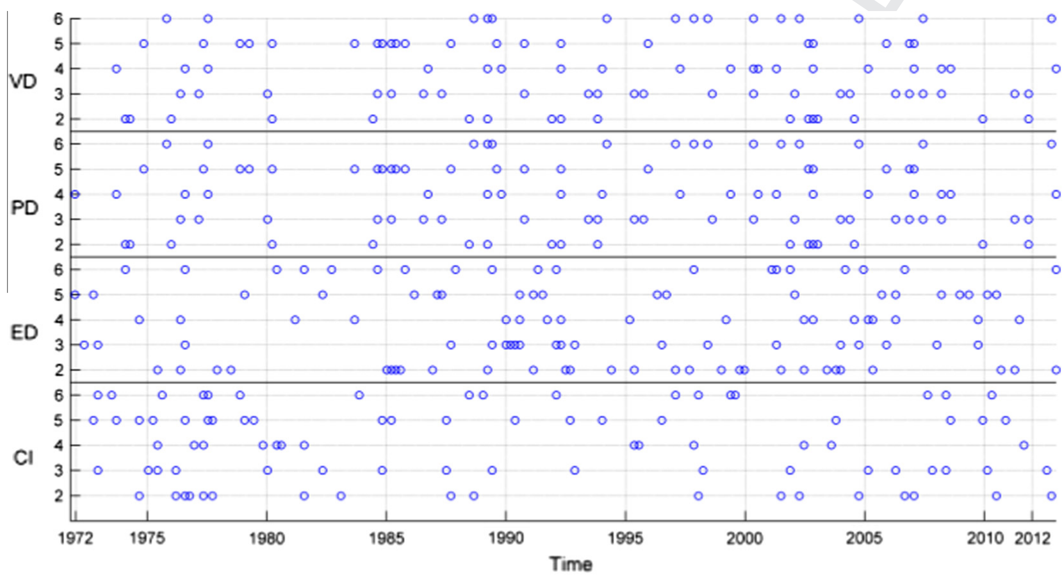

Fig. 12. Diachronically allocation of significant cases presented in Table 5 (panel A).

Table 6
The breakdown of identified significant cases according to Bernoulli trials for a significance level of 5%.

Panel A: $w_{train} = 200$, $n_s = 1176$, $n_p = 11{,}504$, $n_s/n_p = 10.22\%$, $n_w = 215 \times 80 = 17{,}200$					
PIPs type	291/2879	331/2979	279/2822	275/2824	
{CI, ED, PD, VD}	10.11%	11.11%	9.89%	9.74%	
Breakpoints	239/2399	236/2356	223/2315	256/2236	222/2198
{2, 3, 4, 5, 6}	9.96%	10.02%	9.63%	11.45%	10.10%
Threshold	276/4300	438/4266	382/2462	80/476	
{1, 2, 3, 4}	6.42%	10.27%	15.52%	16.81%	
Panel B: $w_{train} = 400$, $n_s = 1478$, $n_p = 14{,}904$, $n_s/n_p = 9.92\%$, $n_w = 211 \times 80 = 16{,}880$					
PIPs type	403/3725	381/3812	346/3684	348/3683	
{CI, ED, PD, VD}	10.82%	9.99%	9.39%	9.45%	
Breakpoints	310/3018	305/3034	302/2995	291/2967	270/2890
{2, 3, 4, 5, 6}	10.27%	10.05%	10.08%	9.81%	9.34%
Threshold	305/4220	388/4219	432/3983	353/2482	
{1, 2, 3, 4}	7.23%	9.20%	10.85%	14.22%	
Panel C: $w_{train} = 600$, $n_s = 1551$, $n_p = 16{,}003$, $n_s/n_p = 9.69\%$, $n_w = 207 \times 80 = 16{,}560$					
PIPs type	428/3978	369/4040	377/3992	377/3993	
{CI, ED, PD, VD}	10.76%	9.13%	9.44%	9.44%	
Breakpoints	320/3210	343/3233	290/3200	293/3184	305/3176
{2, 3, 4, 5, 6}	9.97%	10.61%	9.06%	9.20%	9.60%
Threshold	299/4140	379/4140	422/4110	451/3613	
{1, 2, 3, 4}	7.22%	9.15%	10.27%	12.48%	
Panel D: $w_{train} = 800$, $n_s = 1352$, $n_p = 16{,}009$, $n_s/n_p = 8.45\%$, $n_w = 203 \times 80 = 16{,}240$					
PIPs type	380/3997	335/4023	321/3994	316/3995	
{CI, ED, PD, VD}	9.51%	8.33%	8.04%	7.91%	
Breakpoints	325/3206	274/3211	245/3205	260/3197	248/3190
{2, 3, 4, 5, 6}	10.14%	8.53%	7.64%	8.13%	7.77%
Threshold	270/4060	312/4060	342/4043	428/3846	
{1, 2, 3, 4}	6.65%	7.68%	8.46%	11.13%	

Please cite this article in press as: Tsinaslanidis, P. E., & Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.04.028

Significant cases presented in Table 5 (panel A) are being presented diachronically in Fig. 12. One first obvious finding is that PD and VD measures generate almost the same significant cases. This is also consistent when other parameter combinations are used. In addition we can observe that CI outperforms PIPs until 1980, whilst PIPs identify more significant cases afterwards. In addition when using CI, the algorithm's performance decreases during periods of financial crisis that affected the examined currency pair (1987 Black Monday, 1992–1993 Black Wednesday). Many issues of the Wall Street journal allocate the start of the US sub-prime crisis in June 2007 (Wang, Xie, Han, & Sun, 2012), after which the predictive performance of the proposed algorithm also decreases. Overall, the ED measure seems to add value in the predictive performance of the algorithm as we move closer to present.

Similarly to Tables 4 and 6 presents the breakdown of significant cases where the algorithm predicts changes regardless their size in a frequency greater than 50. Significant cases tabulated are those signified by adopting Bernoulli trials for a significance level of 5%. As expected, parameters affect similarly the predictive performance of the proposed predictive scheme with the difference that significant cases are apparently more than those identified when the DM-test is used for the evaluation. This is reasonable, since predicting changes in directions regardless their size is a relaxed version of predictions made on price changes considering the size.

5. Discussion and conclusions

In this paper, we proposed an algorithmic, nonlinear prediction scheme, implemented for assessing statistically the efficient market hypothesis (EMH) on simulated and real financial price series. PIPs and DTW were combined for this purpose. Initially, we applied the proposed scheme to simulated series composed by weighted sum of a random walk and a deterministic time series generated by the chaotic Mackey–Glass (MG) system in order to verify the ability of the proposed approach to model successfully the deterministic part. Subsequently, we used two datasets, the set of 18 major world indices and the GBP/USD exchange rates. For the first dataset, the prediction scheme did not provide evidence for rejecting EMH as the statistic for the prediction error was at the level (or worse!) than the persistent prediction, i.e. predicting the future value with the current value. However, the predictive performance of the proposed algorithm is better when adopted on the examined currency pair. We scrutiny further our analysis by introducing more values for the examined parameters, and using larger sizes of training samples. In addition, we apply an additional parameter dubbed similarity threshold (t_{sim}) whereby the algorithm makes predictions only when the number of similar historical subsequences is greater than or equal to t_{sim}. Without the adoption of this parameter, the algorithm is forced to find nearest neighbors and make predictions on a daily basis which reduces its overall predictive performance. An interesting finding is that the performance is improved by the introduction of this last parameter. This implies that future research should direct attention to examining the criteria that deal with the selection of the historical similar subsequences.

The construction of the proposed prediction scheme is to the best of our knowledge novel and designed in the spirit of TA. The PIPs implemented in the prediction scheme seem to add some value, as at cases they could provide better prediction than the current value prediction. However, our results should be cautiously interpreted. In order to reject the EMH we should construct trading strategies that generate systematically abnormal returns, considering also other aspects like the number of transactions, transaction costs and embedded risk. In addition, the proposed methodology is computationally expensive, and efforts could be spent to optimize the search of PIPs and computation of DTW, but certainly this is not a computational issue for daily predictions. Further, the proposed scheme relies on two important free parameters that are not investigated or optimized in this study, namely the number of PIPs to be extracted from the time series or alternatively the average number of samples between PIPs, and the number of best matches to be used for prediction (alternatively this number can be derived by the threshold on the standardized distances). Reducing the computational expensiveness of the proposed algorithm, will also allow the user to assess its predictive performance with bootstrap-based evaluative procedures techniques and/or surrogate data analysis and also apply parameter optimization techniques for selecting the important parameters and assess their consistency over time. The aforementioned issues are out of the scope of this paper, and are left for future investigation. However the proposed methodology combines well known tools of data mining and introduces them in the scientific field of finance. We believe that the proposed prediction scheme, either in its current status or enhanced can have practical and valuable implications in the academia and financial industry in general. In particular, academics can use the proposed algorithm as a tool for testing the weak-form EMH whereas practitioners may use it as a basis on which they can design trading strategies in the future, after assessing its performance, enhancing it (if necessary), or modifying it according to their idiosyncratic needs.

Acknowledgement

We would like to thank but not implicate two anonymous referees for their constructive comments that substantially improved the final version of this article. All errors remain, of course our own.

References

Achelis, S. B. (1995). *Technical analysis from A to Z*. Chicago: Probus Publishing.
Berndt, D. J., & Clifford, J. (1994). Using dynamic time warping to find patterns in time series. In *Association for the advancement of artificial intelligence, workshop on knowledge discovery in databases (AAAI)* (pp. 229–248).
Bulkowski, T. N. (2002). *Trading classic chart patterns*. New York: John Wiley & Sons Inc.
Caginalp, G., & Laurent, H. (1998). The predictive power of price patterns. *Applied Mathematical Finance, 5*, 181–205.
Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. New Jersey: Princeton University Press.
Chung, F. L., Fu, T. C., Luk, R., & Ng, V. (2001). Flexible time series pattern matching based on perceptually important points. In *International joint conference on artificial intelligence workshop on learning from temporal and spatial data* (pp. 1–7).
Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253–263.
Edwards, R. D., & Magee, J. (1997). *Technical analysis of stock trends* (7th ed.). Boston: John Magee inc.
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*, 1–26.
Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance, 25*, 383–417.
Fu, T. C. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence, 24*, 164–181.
Fu, T. C., Chung, F. L., Luk, R., & Ng, C. M. (2007). Stock time series pattern matching: Template-based vs. rule-based approaches. *Engineering Applications of Artificial Intelligence, 20*, 347–364.
Fu, T. C., Chung, F. L., Luk, R., & Ng, C. M. (2004). Financial Time Series Indexing Based on Low Resolution Clustering. In *Workshop at the 4th international conference on data mining* (pp. 5–14).
Fu, T. C., Chung, F. L., & Ng, C. M. (2006). Financial time series segmentation based on specialized binary tree representation. In *International conference on data mining* (pp. 3–9).
Grassberger, P., & Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica D: Nonlinear Phenomena, 9*, 189–208.
Jiang, J., Zhang, Z., & Wang, H. (2007). A New segmentation algorithm to stock time series based on PIP approach. In *International conference on wireless communications, networking and mobile computing* (pp. 5609–5612).
Kugiumtzis, D. (1996). State space reconstruction parameters in the analysis of chaotic time series – the role of the time window length. *Physica D, 95*, 13–28.

Kugiumtzis, D. (2002). State space local linear prediction. In A. Soofi & L. Cao (Eds.), *Modelling and forecasting financial data, techniques of nonlinear dynamics* (pp. 95–113). Kluwer Academic Publishers.

Lo, A. W., Mamaysky, H., & Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance, 55*, 1705–1765.

Mackey, M., & Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science, 197*, 287–289.

Mandelbrot, B. (1966). Forecasts of future prices, unbiased markets, and martingale models. *Journal of Business, 39*.

Marshall, B. R., Qian, S., & Young, M. (2009). Is technical analysis profitable on US stocks with certain size, liquidity or industry characteristics? *Applied Financial Economics, 19*, 1213–1221.

Muller, M. (2007). *Information retrieval for music and motion*. Springer-Verlag.

Olson, D. (2004). Have trading rule profits in the currency markets declined over time? *Journal of Banking and Finance, 28*, 85–105.

Osler, C. L. (1998). Identifying Noise Traders: The head-and-shoulders pattern in US equities. In *Federal reserve bank of New York staff reports* (Vol. 42).

Osler, C. L. (2000). Support for resistance: Technical analysis and intraday exchange rates. *FRBNY Economic Policy Review*.

Park, C.-H., & Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys, 21*, 786–826.

Pring, M. J. (2002). *Technical analysis explained: The successful investor's guide to spotting investment trends and turning points* (4th ed.). New York: McGraw-Hill Book Company.

Roberts, H. (1967). Statistical versus clinical prediction of the stock market. Unpublished manuscript, Center for Research in Security Prices, University of Chicago.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 43–49.

Savin, G., Weller, P., & Zvingelis, J. (2007). The predictive power of "head-and-shoulders" price patterns in the US stock market. *Journal of Financial Econometrics, 5*, 243–265.

Sullivan, R., Timmermann, A., & White, H. (1999). Data-snooping, technical trading rule performance, and the bootstrap. *Journal of Finance, LIV*, 1647–1691.

Sullivan, R., Timmermann, A., & White, H. (2003). Forecast evaluation with shared data sets. *International Journal of Forecasting, 19*, 217–227.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., & Farmer, J. D. (1992). Testing for nonlinearity in time series: The method of surrogate data. *Physica D: Nonlinear Phenomena, 58*, 77–94.

Wang, J., & Chan, S. (2009). Trading rule discovery in the US stock market: An empirical study. *Expert Systems with Applications, 36*, 5450–5455.

Wang, G. J., Xie, C., Han, F., & Sun, B. (2012). Similarity measure and topology evolution of foreign exchange markets using dynamic time warping method: Evidence from minimal spanning tree. *Physica A, 391*, 4136–4146.

Zapranis, A., & Tsinaslanidis, P. (2010). Identification of the head-and-shoulders technical analysis pattern with neural networks. In *20th international conference on artificial neural networks – ICANN 2010* (Vol. 6354, pp. 130–136). Thessaloniki, Greece: Springer.

Zapranis, A., & Refenes, A.-P. (1999). Principles of neural model identification. In *Selection and adequacy: With applications to financial econometrics*. Springer-Verlag London Limited.

Zapranis, A., & Tsinaslanidis, P. E. (2012a). Identifying and evaluating horizontal support and resistance levels: An empirical study on US stock markets. *Applied Financial Economics, 22*, 1571–1585.

Zapranis, A., & Tsinaslanidis, P. E. (2012b). A novel, rule-based technical pattern identification mechanism: Identifying and evaluating saucers and resistant levels in the US stock market. *Expert Systems with Applications, 39*, 6301–6308.

Please cite this article in press as: Tsinaslanidis, P. E., & Kugiumtzis, D. A prediction scheme using perceptually important points and dynamic time warping. *Expert Systems with Applications* (2014), http://dx.doi.org/10.1016/j.eswa.2014.04.028