Fast nonlinear deterministic forecasting of segmented stock indices using pattern matching and embedding techniques.

Georgios N. Banavas, Susan Denham, & Michael J. Denham
Centre for Neural & Adaptive Systems
School of Computing
University of Plymouth
Drake Circus, Plymouth PL4 8AA, Devon, UK
email: georgios.banavas@soc.plym.ac.uk

Abstract

We propose an automated system for out-of-sample predictions of a set of European stock indices. The system performs on piecewise linear representations of the time series. An automated segmentation algorithm converges to an optimum segmented time series representation, which achieves considerable data compression and allows variable sampling rate of the time series depending on different segments having different length. The minimum embedding segment dimension (MESD) algorithm, we propose, seeks for deterministic behavior of the processing data set. MESD returns the embedding dimension of the underlying dynamics of the series, measured in number of segments. Embedding dimension calculations have never been applied on segmented representations. We summarize the advantages of the method in the following: (i) it can detect high dimensional nonlinear deterministic behavior as being projected on a lower dimension segment space; (ii) it is computationally efficient; (iii) it converges to an optimum solution without a-priori parameterization. We use the minimum embedding segment dimension (MESD) as an indicator of the length of patterns that can be retrieved from the time series own past. Our pattern matching technique enables the matching of such historical patterns on others of similar shape which occur in different time scales. To define an appropriate similarity measure, we introduce the notation of Multiple Feature Sets (MFS) which employ Dynamic Time Warping (DTW), and first derivative and temporal features. An additional advantage of the system we propose is that the segmented representation scheme and the prediction model are both data driven and that the predictions are made using information only from the time-series own past without any a priori knowledge being injected into the model. We demonstrate that this approach may offer a useful decision support tool for stock market trading.

Keywords*: Time Series Segmentation, Embedding Segment Dimension, Dynamic Time Warping, Pattern Matching, Forecasting, Trading.*

1 Introduction

The aim of this paper is to perform out-of-sample trend prediction of financial stock indices. Forecasting indications about the standing and the momentum of the trend is attached to each prediction. Following the definition of time series segmentation as a process of dividing the data into distinct subsets which simplify the information processing, we address the problem of matching a query pattern, P^Q, extracted from the recent past of a stock index (SI), onto historical occurrences of similar shape but different duration and rescaled in a non-uniform manner. Past matches are used to predict future trend activity based on temporal ratios. The stock index from where the query pattern has been derived, is noted as the maternal time series M. Because we believe that the stock market generates its own patterns which do not always follow the shape of the known technical analysis chart patterns, we set the query pattern to illustrate the 'current' situation of the market [1]. The linguistic expression 'current' automatically inserts a parameter in our system that has to be fitted properly in order to perform successful predictions, [2]. To solve this problem, we introduce here the notation of Minimum Embedding Segment Dimension (MESD) and a method for that to be calculated. MESD indicates the number of successive segments which compose the 'current' pattern and also is used as evidence for deterministic segmented time series underlying dynamic behavior. The case that is examined here is the univariate case, where M is a univariate time series. The proposed system involves the following processing stages:

- Time series linear segmentation.
- Computation of the minimum embedding segment dimension for query pattern selection.
- Pattern matching on segmented representations by incorporating similarity measures based on Dynamic Time Warping (DTW) and Multiple Feature Sets (MFS).
- Trend Prediction based on temporal ratios.
- Stock index trading based on selective trading rules.

This paper is an extension of the work which has been undertaken by Banavas et al, (2000) [2]. Therefore, in order to present a complete system, some technical aspects similar to our previous work are quoted here too. All the processing stages summarized above are integrated parts of a time series analysis simulator, PROGNOSIS, which we have partly developed in the Centre for Neural & Adaptive Systems.

2 Segmenting Time Series

Time series segmentation can be defined as the process of dividing the data into distinct subsets which have common characteristics. Segmentation is applied in many domains such as image processing, speech recognition and scene analysis. The purpose of segmenting the data in most applications is to allow subsequent information processing using the data subsets. In financial time series, the aim of segmentation is to depict clearly local trends as determined from the data, by distinguishing them from noise. The global time series shape must, after a successful segmentation, be retained. Moreover, time series segmentation resembles the way that humans reproduce or draw series with high frequency fluctuations. In figure 1 we depict a part of a financial time series, a manual reproduction of it by the author and an automatic segmented representation as produced by PROGNOSIS.

The segmentation method employed here is the one proposed by Keogh, (1997) [8]. This approach is based on a local merging of adjacent segments algorithm, which starts from an initial, $\lfloor \frac{n}{2} \rfloor$, number of segments. All possible merges are tested and the one that gives the minimum residual error is the one selected for the next iteration step. The algorithm converges to an optimum number of segments, when the overall residual error exceeds some threshold. The process results in a compressed but reliable representation of the time series, as depicted in fig. 1.

3 The Minimum Embedding Segment Dimension

The question that many researchers ask themselves when observing financial time series, is whether the dynamics of the data generating process derive from deterministic chaos or from randomness. Suitable mathematical

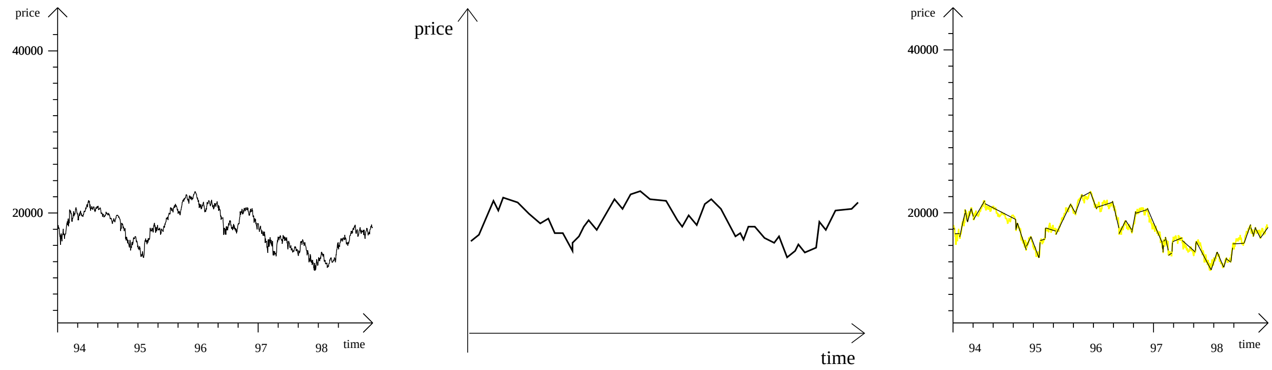


Figure 1. Left: The original JAPDOW stock index time-series. Center: A manual representation of the JAPDOW as obtained by the author. Right: A 50 linear segments JAPDOW stock index representation (93% data compression achieved).

theory development in distinguishing deterministic chaos from random processes has been introduced first by Takens, (1981), [15] and later by Sauer et al., (1991), [13]. Known as *time-delay embedding* this theory has been investigated by many researchers recently. Among others Cao et al., (1998), [4] and Soofi et al., (1999), [14] introduced different variants of calculating the optimal embedding dimension of financial time series. By studying their experiments on a set of exchange rate series, it can be revealed that the results on the embedding dimensions (ED), they claim, strongly depend on fixed parameters, like the constant sampling rate of the data. Cao et al., (1999), [5] for example, reported much higher EDs on daily exchange rates than Lisi et al., [10], who used monthly data. Here, we calculate the ED of segmented stock indices. We call it Minimum Embedding Segment Dimension (MESD) and it is measured in number of segments. This overcomes the problem of constant sampling rate and allows the search for attractors in the data, which derive from linear segments of different length. Finally, the MESD calculations are faster than its ancestors, which applied directly on raw data.

The Method Consider a time series $x_1, x_2, \cdots, x_N$ which has been divided into M segments ($M < N$), using the method given in 2. The segmented series is:

$$S = \{\vec{s_0}, \vec{s_1}, \vec{s_2}, \cdots, \vec{s_i}, \cdots, \vec{s_M}\} \quad , \tag{1}$$

where each segment is considered as a vector of successive time series values:

$$\vec{s_i} = \{x_j, \cdots, x_{j+k_i}\} \quad . \tag{2}$$

k_i is the time duration of each segment s_i. A time-delay vector z of 1 can be written as follows:

$$z_i(d) = \{s_i, s_{i+1}, \cdots, s_{i+(d-1)}\}, \quad i = 0, 1, 2, \cdots, M - (d-1) \quad , \tag{3}$$

where d is the systems embedding dimension. For simplicity, the time-delay parameter has be chosen to be 1. Following the notation in [6], β in eq. 4, is the ratio of the distances of the time-delay segment vectors z_i with

	UK	FRANCE	GERMANY	SPAIN	ITALY	GREECE
MESD	24	21	22	16	17	16

Table 1. The Minimum Embedding Segment Dimension (MESD) estimations of a set of European stock indices. The MESD is measured in number of segments.

their corresponding nearest neighbors, when moving from embedding dimension d to $d+1$. That mathematically is illustrated as follows:

$$\beta(i,d) = \frac{\|z_i(d+1) - z_{NN_i}(d+1)\|}{\|z_i(d) - z_{NN_i}(d)\|}, \quad i = 0, 1, 2, \cdots, M-d \quad . \tag{4}$$

$\| \ \|$ is a distance norm defined as the standard deviation (std) of the distances of the corresponding time series parts indicated by individual segments. This arises from the fact that similar vectors have point distances characterized by low standard deviation, [7]. Time series parts are compared using the DTW distance metric (see [3, 2]). So:

$$\|z_x(d) - z_y(d)\| = std(D_{DTW}(\vec{s_1^x}, \vec{s_1^y}), D_{DTW}(\vec{s_2^x}, \vec{s_2^y}), \cdots, D_{DTW}(\vec{s_d^x}, \vec{s_d^y})) \quad . \tag{5}$$

D_{DTW} is the minimum distance metric defined in [3, 2]. NN is an index which indicates the position of nearest neighbor of the time-delay vector z, on the segmented time series. NN is the same in both the numerator and denominator of eq.3.

According to the embedding theorems of [15] and [13], d is chosen to be the system's ED when two time-delay vectors mapped in the d-dimensional reconstructed space, will remain mapped in the $d+1$-dimensional space. In other words the β-ratio defined in eq.4 will not be close to one for an non-suitable minimum embedding segment dimension d. However, because β might be different for every segments time-delay vector z_i, we adopt the quantity defined by Cao, (1998), [4] to overcome this problem.

$$E(d) = \frac{1}{N-d} \sum_{i=1}^{N-d} \beta(i,d) \tag{6}$$

$E(d)$ in eq.6 depends solely on the value of dimension d (time delay has been set to one). To investigate the variations of $E(d)$, Cao introduce the fraction $W(d) = E(d+1)/E(d)$. The value of d for which the fraction $W(d)$ stops changing substantially, indicates that the MESD of the underlying system is $d+1$. By applying the above described algorithm on segmented representations of a set of stock indices, we have found deterministic behavior for the European stock indices test set we use. The MESDs on the whole data set is given in table 1.

4 Pattern Matching

In this section, we present the way that pattern matching is performed on segmented time series. The key idea is to define a similarity measure based on compound feature information as derived from each individual segment of the query pattern P^Q. As said before, the number of segments for the pattern P^Q is assessed by the identification of the MESD of the underlying time series. All features are attached to each segment as depicted in fig. 2. More specifically, the first feature is the distance between corresponding time series subsets[1] calculated by using the Dynamic Time Warping distance norm (see [2]). Second, the slope of each segment is compared against the one

[1] The start and the end point of each time series subset, $\vec{s_i}$ is identified by the corresponding segment, s_i. A way of applying DTW directly on segmented time series has been proposed in [9].

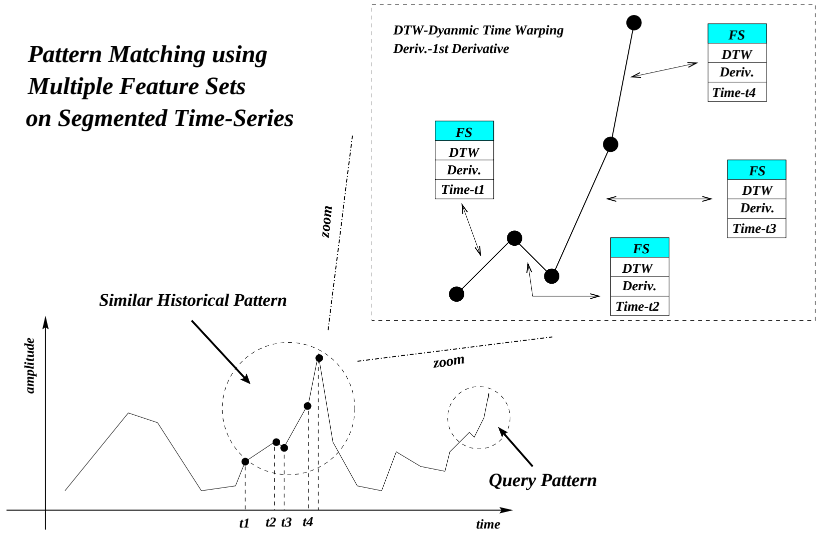


Figure 2. Pattern Matching using Multiple Feature Sets on Segmented Time-Series.

of its potential neighbor and finally the time information labels are attached to each pattern in a way to overcome the problem of finding more than one pattern having the same similarity. In such a case, we make use of the time labels to select the most recent match. The feature set, described above, is integrated in the following similarity function D_G:

$$D_G(P^Q, P) = \frac{1}{d} \sum_{i=1}^{d} \frac{w_{DTW} D_{DTW}(\vec{s_i^Q}, \vec{s_i}) + w_{SLOPE} D_{SLOPE}(s_i^Q, s_i)}{d} \quad , \tag{7}$$

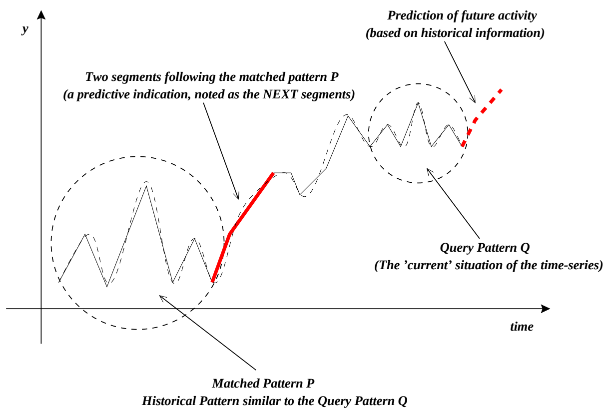
where $D_{DTW}(\vec{s_i^Q}, \vec{s_i})$ is the minimum distance norm calculation, according to the optimum path found via dynamic programming using the Dynamic Time Warping algorithm (see [2]), $D_{SLOPE}(s_i^Q, s_i)$ is squared difference of the slopes of the segments. s_i and $\vec{s_i}$ refer to the segment and its corresponding time series subset at position i respectively.

In D_G, eq. 7, distance measures for each feature F are computed separately. Each constituent feature similarity is then added with certain normalized weighting factors. Satisfactory weighting can be found experimentally. Here, we have selected $w_{DTW} = 0.6$ and $w_{SLOPE} = 0.4$ ($w_{DTW} + w_{SLOPE} = 1$). For this weighting set, desired matches are achieved when similarity is greater than 60%

5 Trend Forecasting

The idea of using pattern matching for predicting financial time series, has been systematically introduced in [1, 2]. Once the *current* situation of the market (query pattern) is geometrically similar to what happened sometime ago, we hypothesize that the local reaction of the market after the historical matched occurrence, will indicate some shape and magnitude information about the market's future activity. Economic theory or technical analysis evaluation of this information might also be possible. Fig. 3 depicts a graphical explanation of our prediction hypothesis.

This kind of prediction system thinking becomes more valuable, because of affinities in the market's behavior, derived solely from well established economic market interpretation theory, from broadly used technical analysis tools or even from the basic rule of *demand and supply*. Imponderable factors, such as political guidance or intervention and insider trading, are not considered here.



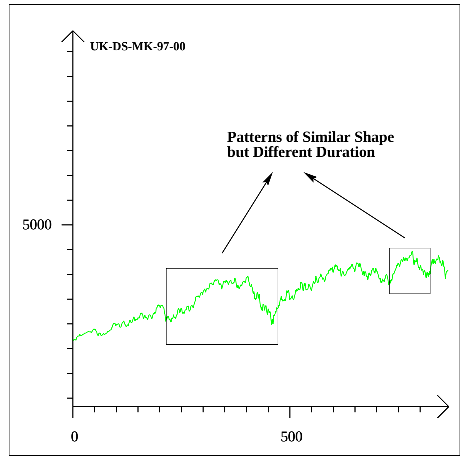


Figure 3. How trend prediction can be made using multiple feature set pattern matching. Left: An artificial example of *Head & Shoulders* **patterns occurring in different resolutions. Right: Similar pattern identified on the UK datastream market index.**

Technically speaking, each trend prediction is normalized in duration, according to the relative time ratios of patterns P and Q (see fig. 3). Another technical issue of the prediction approach is the verification of whether the last point of the univariate time series is the end or part of the current trend. Our analysis on that issue, based again on temporal ratios, verifies whether the last segment has finished its activity or not (see [2]). Finally, predictions are accompanied by two confidence measures:

- **Matching Confidence -** The similarity value (%) measured between the query pattern, Q, and its best match P.
- **Prediction Confidence -** A measure (%) of how well the k segments (lines) that follow the best match, P, fit the actual time series.

These confidence measures are used to verify improvement on the matching and the prediction accuracy and to drive an implicit trading strategy.

6 Trading

Basically, a trading system is a hybrid of two main modules. A prediction module, which optimizes its performance to output accurate predictions and a trading module, which according to a set of trading rules and to the predictions originated from the prediction module, takes trading positions (buy-sell-hold). Other more sophisticated types of trading systems have been proposed [11, 17, 12, 16]. Here, we base trading on the segment predictions using simple trend following trading rules. If no change in the trend from the previous segment is predicted, the current trading position is maintained and unnecessary brokerage costs are avoided. The trading rule, we employ, is summarized as follows:

$$if \quad SL(s_{t-1}) < SL(s_t) > SL(s_{t+1}) \quad sell - go\ short \quad , \tag{8}$$

$$if \quad SL(s_{t-1}) > SL(s_t) < SL(s_{t+1}) \quad buy - go\ long \quad . \tag{9}$$

$SL(s_{t-1}), SL(s_t), SL(s_{t+1})$ are the slopes of the previous, current and predicted segments respectively[2]. $SL(s_i)$ is one if the segment s_i has an upwards trend and zero if it moves downwards. Each transaction takes place at time

[2] $SL(s_{t+1})$ is replaced with $SL(s_{t+a})$ for a-segs ahead predictions.

t, guided by the duration of the predicted segment. The trading rules of eq. 8,9 are further enriched by making use of the slope indication for each predicted segment. We force the trading system to buy(sell) more(less) stocks, depending on how aggressive the segment has been forecasted to be. Assuming that α is the slope of the predicted segment and α_G is the slope of the regression line fitted on the whole time series, we set the following trading constraints:

$$if \quad \frac{|\alpha|}{|\alpha_G|} \leq 1, \quad sell/buy\ one\ stock \quad , \tag{10}$$

$$if \quad \frac{|\alpha|}{|\alpha_G|} > 1, \quad sell/buy\ \lfloor \frac{\alpha}{\alpha_G} \rfloor\ stocks \quad . \tag{11}$$

Brokerage costs are considered to be 1% of the transaction price. Finally, a stop loss criterion is considered for the system. Everytime that losses exceed a 15% threshold, open trading positions close in order to prevent extensive losses. The automatic trader then waits for a new trading signal in order to take further action. The trading rules system is activated when the prediction confidence is more than 70%. The profit achieved is accumulated all over the testing interval.

7 Results

The evaluation set used to test the system previously described, is composed of six European stock indices (SIs). To increase objectivity about the generalization properties of the system, we chose to work with three SIs from north European countries and another three from countries washed from the Mediterranean. These are the UK, France, Germany, Spain, Italy and Greece. Data have been downloaded from the DATASTREAM[3] database, as it is processed by the DATASTREAM team. The time series range from 01.01.1990 to 26.04.2000, in daily representations (2693 working days). The compression achieved with the segmentation algorithm is on average 84%. The results, we present here have been obtained over the half of the time series points.

7.1 Deterministic Behavior

All the European indices employed in this study, reveal deterministic behavior, according to our embedding segment dimension (ESD) measures. In fig. 4, we depict the variation of $E(d)$ (eq. 6) while d varies between 2 and 75. It can be seen there that $E(d)$ stops changing substantially after a value D_{min}. D_{min} corresponds to the selected ESD. That behavior may be some evidence that the data in process are not generated by purely random processes. Purely random generated series do not follow patterns such as those depicted in fig. 4. However, additional statistical evaluation is needed to strongly support this conclusion. As shown in table 1, the ESD for the European indices oscillates between 25 and 35 segments. An attempt to transform the ESD in days will establish the claim made by [5], that the embedding dimension (ED) of financial data is high. According to our measures the ED of the European data set is greater than 50-70 days. The fact the the ESD calculations are reduced by at least a factor of two, makes our system faster than its predecessors.

7.2 Prediction Results

Table 2 shows the accuracy of the system in predicting one to three segments ahead. The query pattern length is fixed using the ESD indications for each of the time series. Each prediction is accompanied with matching confidence and average prediction confidence measures.

Predictions with accuracy more than 55% have been achieved. Specifically, the system performs generally better when applied on mature indices of north European stock markets. Finally, the matching confidence is constantly over sixty percent. That indicates the realiability of the matching algorithm.

[3] Datastream International Limited, Monmouth House, 58-64 City Road, London EC1Y 2AL

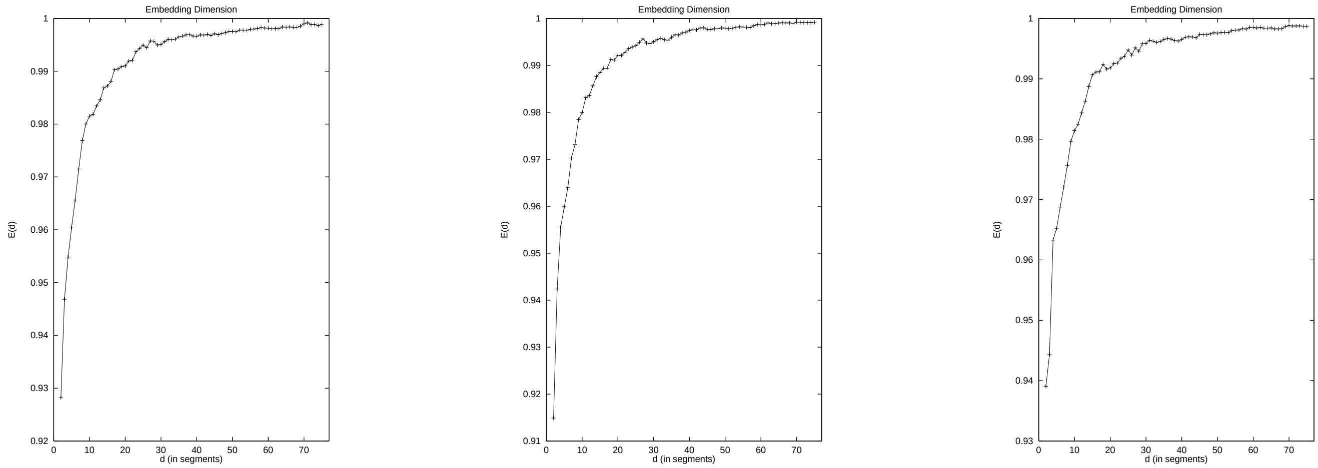


Figure 4. Three examples of the ESD curves on a set of DATASTREAM stock indices. **Left:** FRANCE. **Center:** GERMANY. **Right:** UK.

	UK	FRANCE	GERMANY	SPAIN	ITALY	GREECE
1 s-a	59.22%	60.23%	58.57%	56.48%	57.63%	56.16%
2 s-a	58.31%	57.33%	56.19%	57.07%	55.88%	59.07%
3 s-a	59.42%	57.52%	55.00%	57.27%	57.85%	58.68%
MC	64.32%	63.68%	64.49%	63.51%	66.83%	63.27%
APC	62.02%	61.95%	64.36%	61.58%	67.00%	64.21%

Table 2. Trend prediction accuracy results. Tests have been made for 1, 2, 3 segments ahead (s-a) predictions. The matching confidence (MC) and the average prediction confidence (APC) are displayed in the fourth and fifth table rows respectively.

7.3 Trading

Following the strategy described in section 6, we trade on each of the stock indices independently, starting from the 01.01.1996 till the 26.04.2000. That corresponds to 1128 working days. The strategy we follow, extracts on average 120 trading actions per index. This number has been achieved after applying the *stop loss* criterion described above. The average (%) profits gained for each index are represented in table 3. Prediction confidence greater than 60% has been taken as the constraint for the automatic trader to enter the market. In table 4, we clearly show the improvement gained in terms of profit, when enhancing the trading strategy with the slope rules given by the equations 10 and 11. Representative profit curves on the whole test set are depicted in fig. 5.

The profit results show that both the total and the average profit gained, constantly has a positive sign. The number of transactions indicate that the *automatic trader* acts only for almost 10% of the time applied. That saves the trading strategy from additional brokerage costs and avoids the risk of everyday trading. Finally, the profit curves of fig. 5 are characterized by a clear upwards trend.

	UK	FRANCE	GERMANY	SPAIN	ITALY	GREECE
Profit (Total)	45.88%	16.42%	28.25%	22.58%	31.21%	59.81%
Transactions	166	172	168	169	171	172
Profit (Aver.)	25.18%	12.18%	8.82%	7.45%	9.77%	23.78%

Table 3. Profit gained while trading six European stock indices from 01.01.1996 to the 26.04.2000. No trend slope predictive information has been incorporated in the trading strategy (see section 6).

	UK	FRANCE	GERMANY	SPAIN	ITALY	GREECE
Profit (Total)	58.86%	22.53%	35.65%	29.66%	39.98%	63.72%
Profit (Aver.)	31.30%	16.22%	10.78%	9.50%	12.12%	27.87%

Table 4. Profit gained while trading six European stock indices from 01.01.1996 to the 26.04.2000. The trading strategy takes advantage of trend slope predictive information which indicates the scale of the aggression of the forecasted trends (see section 6).

8 Discussion & Outlook

In this study, we presented a complete automated system for univariate financial time series analysis, prediction and trading. We introduced the concept of ESD, as the ED of piecewise linear segment represented time series. ESD is measured in number of segments. The calculation of ESD on a set of European stock indices proved deterministic behavior. The forecasting component of our system was mainly derived from a pattern matching algorithm we presented in [2]. We performed afterwards, trend predictions accompanied by time indications about the duration of the upcoming trend and slope predictions which indicate the aggression of upwards or downwards movements. We summarized these forecasting information within an explicit trading strategy to achieve consistent profits. Brokerage costs have also been taken into account. The extension of this approach to multivariate financial data sets will be presented elsewhere.

References

[1] Banavas, G., (1999), 'Robust Direction Movement Forecasting using a Pattern Matching Method applied on Univariate Financial Time Series.', *International Symposium of Forecasting, ISF'99, Washington DC*.

[2] Banavas, G. N., Denham, S. & Denham, M. J., (2000), 'A Financial Stock Index Trend Prediction Approach with Temporal Considerations', *Computational Finance 2000, LBS, London*.

[3] Berndt, D. & Clifford, J., (1994), ' and J. Clifford Using Dynamic Time Warping to find patterns in time-series', In *AAAI94 workshop of knowledge discovery in databases*.

[4] Cao, L., (1998), 'Determing the Minimum Embedding Dimension of Input-Output Time Series Data', In *International Journal of Bifurcation and Chaos, Vol. 8, No. 7, pp. 1491-1504*.

[5] Cao, L. & Soofi, A. S., (1999), 'Nonlinear deterministic forecasting of daily dollar exchange rates', In *International Journal of Forecasting, Vol. 15, pp. 421-30*.

[6] Kennel, M., Brown, R. & Abarbanel, H., (1992), 'Determing Embedding Dimension for Phase-Space Reconstruction', In *Physical Review, A 45, pp. 3403-3411*.

[7] Keogh, E., (1997), 'A fast and robust method for pattern matching in time series database', In *Proceedings of WUSS 97, Universal City*.

[8] Keogh, E., (1997), 'Fast Similarity Search in the Presence of Longitudinal Scaling in Time Series Databases', In *Proceedings of the 9th International Conference on Tool with Artificial Intelligence, pp. 578-584, IEEE Press*.

Profit curves

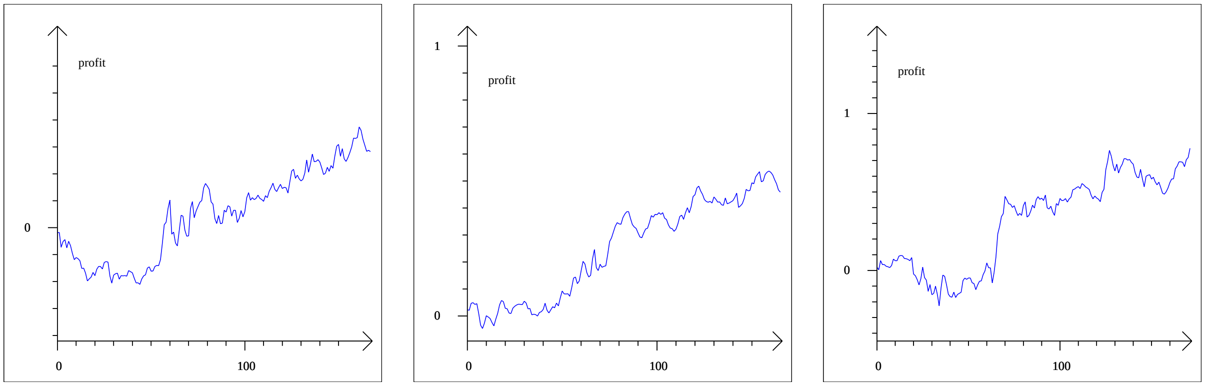


Figure 5. The evolution of profit (%) on the GERMANY(left), UK(center) and GREECE(right) datastream stock indices starting from 01.01.1996 - 26.04,2000.

[9] Keogh, E. & Pazzani, M., (1999), 'Scaling up dynamic time warping to massive datasets', In *3rd European Conference on Principles and Practice of Knowledge Discovery in Databases.*.

[10] Lisi, F. & Medio, A., (1997), 'Is a Random Walk the best Exchange-Rate Predictor?', In *International Journal of Forecasting, 13, 255-267*.

[11] Moody, J. & Wu, L. Z., (1996), 'Optimization of Trading Systems and Portfolios', In *Proceedings of the Neural Networks in the Capital Markets (NNCM'96)*.

[12] Moody, J., Wu, L., Liao, Y., & Saffell, M., (1998), 'Performance Functions and Reinforcement Learning for Trading Systems and Portfolios', In *Journal of Forecasting, vol.17, pp. 441-470*.

[13] Sauer, T, Yorke, J. A. & Casdagli, M., (1991), 'Embedology', In *Journal of Statistical Physics, 65, pp. 579-616*.

[14] Soofi, A. S. & Cao, L., (1999), 'Nonlinear Deterministic Forecasting of Daily Peseta-Dollar Exchange Rates', In *Economics Letters, 62, 175-180*.

[15] Takens F., (1981), 'Detecting Strange Attractors in Fluid Turbulence', In *Rand, D. A. & Young, L. S. (Eds.) Dynamical Systems and Turbulence, Lecture Notes in Mathematics, Vol. 898, Springer Verlag, Berlin*.

[16] Towers, N. & Burgess, A. N., (1998), 'Optimisation of Trading Strategies using Parameterised Decision Rules', In *Perspectives on Financial Engineering and Data Mining, PRoc. IDEA'98, edited by L. Xu et al., Springer-Verlag*.

[17] Xu, L., & Cheung, Y., (1997), 'Adaptive Supervised Learning Decision Networks for Traders and Portfolios', In *Journal of Computational Intelligence in Finance, (November-December): 11-16*.

A ProGNOSIS

Fig. 6, on the next page, is a snapshot of the *ProGNOSIS* user interface.

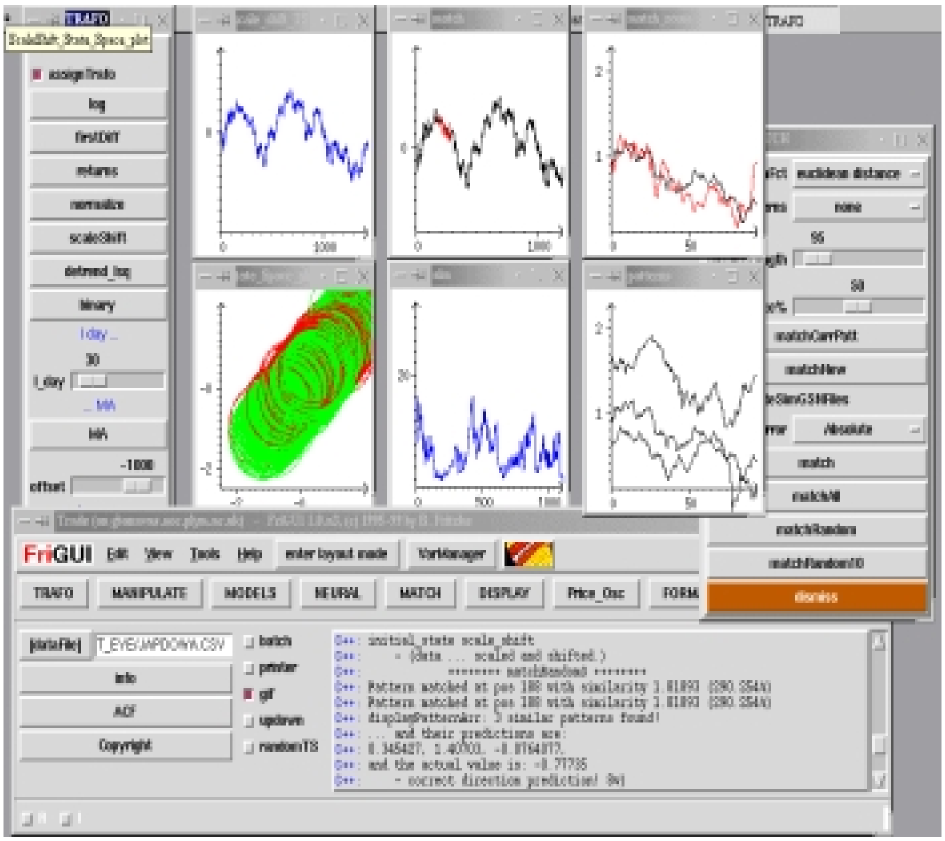


Figure 6. Snapshot of the ProGNOSIS simulator (ProGNOSIS.a9.9).