Tsinaslanidis, P. (2018) Subsequence dynamic time warping for charting: bullish and bearish class predictions for NYSE stocks. Expert Systems with Applications, 94. pp. 193-204. ISSN 0957-4174.

https://doi.org/10.1016/j.eswa.2017.10.055

Subsequence dynamic time warping for charting: Bullish and bearish class predictions for NYSE stocks

Prodromos E. Tsinaslanidis

Business School, Canterbury Christ Church University, North Holmes Road, Canterbury, Kent, CT1 1QU, UK

Abstract

Advanced pattern recognition algorithms have been historically designed in order to mitigate the problem of subjectivity that characterises technical analysis (also known as 'charting'). However, although such methods allow to approach technical analysis scientifically, they mainly focus on automating the identification of specific technical patterns. In this paper, we approach the assessment of charting from a more generic point of view, by proposing an algorithmic approach using mainly the dynamic time warping (DTW) algorithm and two of its modifications; subsequence DTW and derivative DTW. Our method captures common characteristics of the entire family of technical patterns and is free of technical descriptions and/or guidelines for the identification of specific technical patterns. The algorithm assigns bullish and bearish classes to a set of query patterns by looking the price behaviour that follows the realisation of similar, in terms of price and volume, historical subsequences to these queries. A large number of stocks listed on NYSE from 2006 to 2015 is considered to statistically evaluate the ability of the algorithm to predict classes and resulting maximum potential profits within a test period that spans from 2010 to 2015. We find statistically significant bearish class predictions that generate on average significant maximum potential profits. However, bullish performance measures are not significant.

Email address: `prodromos.tsinaslanidis@canterbury.ac.uk` (Prodromos E. Tsinaslanidis)

Keywords: Technical analysis, Pattern recognition, Dynamic time warping

1. Introduction

Technical analysis (TA) or charting is a primarily graphical assessment of the historical evolution of trading-related price paths (Tsinaslanidis and Zapranis, 2016). It is based on the belief that history tends to be repeated, and thus these price paths exhibit regularities. Proponents of TA argue that these regularities can be profitably exploited to extrapolate future price movements (Campbell et al., 1993). Apparently, this conflicts with the weak-form efficient market hypothesis (Fama, 1970), in which current prices fully reflect all historical information and any trading system/strategy based solely on historical information (such as TA) can not systematically generate statistically significant excess returns. Technicians (or chartists) use a wide range of technical tools, such as technical indicators, candlesticks and technical patterns, to derive trading signals. The efficacy of these tools has been assessed in many empirical research works, whereby mixed results were reported.[1]

Yet, despite efforts undertaken, the task of assessing the overall efficacy of charting still remains a challenging area of research. First, subjectivity embedded mainly in the identification and interpretation of technical patterns, places significant barriers in assessing the performance of this trading practice. Many empirical works can be found in the bibliography that aimed to mitigate this problem by developing algorithmic pattern recognisers which allowed to extrapolate a set of technical patterns that met specific criteria and assess statistically their performance. For example, two of the first attempts to automate the recognition of the '*head-and-shoulders*' pattern can be found in Neftci (1991) and Osler and Chang (1995), where several precise criteria were set, whose fulfilment was a necessary condition for the pattern's confirmation.

[1] For a comprehensive literature review regarding the profitability of TA the reader may see Park and Irwin (2007) and Nazário et al. (2017), whilst for a focus on technical patterns see Tsinaslanidis and Zapranis (2016) and references therein.

For the identification of variations of '*bull flags*', template matching techniques were implemented in (Leigh et al., 2002a,b; Bo et al., 2005; Cervelló-Royo et al., 2015; Arévalo et al., 2017), whilst neural networks and genetic algorithm were used by Leigh et al. (2002c). A template matching technique was also adopted by Wang and Chan (2009) for identifying '*rounding bottoms*', whilst Zapranis and Tsinaslanidis (2012) proposed a rule-based algorithm for identifying this technical pattern.

Secondly, although there are many studies proposing algorithms for the identification of specific technical patterns (such as those referred above), or bundles of technical patterns (such as Lo et al., 2000; Dawson and Steeley, 2003), valid inferences for the predictability of TA as a trading method in general are difficult to be made. We could argue that technical patterns currently known and used by the industry and the academia are not exhaustive. Although there is a wide variety of technical patters, we have not found any evidence or statement supporting that technical patterns reported by this time in the bibliography are (or better, should be) the only ones. This implies that even proponents of TA would have to agree that existed technical patterns cannot be used to capture every possible regularity in the price paths of financial instruments. Thus, for exploring the overall predictability of TA (for a given dataset) it is required to evaluate either jointly the predictive performance of a relative large set of technical tools or assess the predictability of a prediction scheme that is designed to capture principles of TA rather than focusing on individual technical tools.

Motivated mainly by the latter point, this study aims to approach charting from a more generic viewpoint and advance the research on this field in several ways. First, we formally state a set of five charting characteristics. Subsequently, an algorithmic framework is proposed which is designed to capture these common characteristics of the entire family of technical patterns, instead of focusing on a specific, or a bundle of specific technical patterns.

The proposed algorithmic framework is mainly based on dynamic time warping (DTW) and two of its modifications; the derivative DTW and the subsequence DTW. DTW and its modifications are algorithmic techniques that initially became popular in the context of speech recognition (Sakoe and Chiba, 1978). However, they have been used in other scientific areas as well, including finance (Wang et al., 2012; Tsinaslanidis and Kugiumtzis, 2014). DTW is mainly used to find an optimal alignment between two given (time-dependent) sequences, derivative DTW (Keogh and Pazzani, 2001) is a modification of DTW which considers the estimated local derivatives of the data, allowing the implementation of DTW not only to sequences that differ in length (time) but also sequences that differ in price level, whilst subsequence DTW (Müller, 2007) allows the user to identify subsequences within a longer sequence, which are 'similar' to another shorter query sequence.

In this paper, price series of 640 NYSE stocks are considered for the period 2006-2015 (see section 3). In the proposed algorithm, starting and ending points of candidate query patterns are defined by the perceptually important points (PIPs) algorithm (Fu et al., 2008) and volume 'peaks' respectively. The search of the query patterns is confined to the test period that spans from 2010-2015. DTW and its two modifications are being used to find historical subsequences which are the most 'similar' to each query in terms of price and volume. The search of similar patterns is being carried in the whole dataset, rather than in the individual series that the query pattern belongs to. It is worth noting that these historical subsequences can occur at any time either during the period 2006-2009 or during the test period (after the end of 2009) as long as they occur before the realisation of the corresponding query. For each query pattern, class predictions are made for the future price evolution, by looking the price behaviour after the formation of the identified historical subsequences.

We initially assess the ability of the algorithm to predict bullish and bearish patterns and we find a superior performance in the latter case. The profitability of the algorithm is also assessed and we find that only bearish signals generate statistically significant returns. Our results indicate that TA may in principal add value in the trading decision processes.

The rest of this paper is organised as follows. In section 2, we present and justify a collection of five characteristics of charting that our proposed algorithm captures. Section 3 presents our dataset, whilst section 4 presents our methodology. Empirical results regarding the predictive performance of the algorithm are presented in section 5, whilst the algorithms profitability is statistically assessed in section 6. Finally, section 7 provides a discussion and concludes.

2. Inferences from charting

The aim of the algorithmic pattern recognition scheme, proposed in this paper, is to set the ground for a comprehensive assessment of the predictive performance of charting without the consideration of particular technical patterns. In order to achieve this, this scheme has to be free of specific technical descriptions and/or guidelines for the identification of particular technical patterns. Instead, it has to capture generic technical characteristics that can be inferred and justified from the relevant bibliography on TA or common practices that technicians follow. This section presents the justification of the five inferences from charting that we are considering in this study.

(i) *Technical patterns are globally valid*. This means that technical patterns such as '*head-and-shoulders*' and '*flags*' are valid for all securities price series. In the context of TA, when a technical pattern is 'confirmed' technicians expect a particular future price behaviour regardless the securities series on which the pattern was identified. As an example, a '*head-and-shoulders*' pattern signals a downwards movement of the price series, whether this series corresponds to a stock (belonging to any sector), a currency rate or a market index. This 'global' validity of technical patterns, can also be inferred and further supported by the bibliography. For instance, Park and Irwin (2007), categorised 137 technical trading studies into 'stock markets', 'foreign exchange markets' and 'futures markets' studies, according to the markets they had concentrated on.

(ii) *Patterns are sequences of regional locals*. An explicit statement in support of this argument can be found in (Neftci, 1991, p. 550) where the author states: '*This article shows that most patterns used by technical analysts need* *to be characterized by appropriate sequences of local minima and/or maxima and will lead to nonlinear prediction problems*'. Various proposed algorithms for the identification of technical patterns involve the assignment of specific criteria to sequences of regional locals. Such locals can be identified with various methods. For example, Lo et al. (2000) use a kernel regression to smooth price series and then they consider the derivative of the smoothed series in respect to time. They compared the signs of neighbouring derivatives and a regional local was identified when those signs differed. Finally, Lucke (2003), identified local extrema by using a computer program which was originally designed to identify business cycle turning points, whilst Zapranis and Tsinaslanidis (2012), used a rolling window of fixed size to identify regional peaks.

(iii) *Increased volume confirms a pattern's formation.* There are many cases of technical patterns whose formation is confirmed by an increased level of volume. Pring (2002), when describing the trading signals that occur after price breaches a trend-line, highlights the importance of high volume at such breakouts with the following statement: '*It is this upward surge in the trading activity that confirms the validity of the breakout*'.The importance of high volume in confirming the formation of other technical patterns is also emphasised by the same author. In the context of TA, it can be argued that increased volume may indicate that most of the traders (or else most of the trades) are seeing the same thing in the market, and therefore the market confirms that a pattern has occurred.

(iv) *Patterns can occur at different sizes.* This statement implies that the width (time length) and the height of a particular technical pattern may vary. The variation that occurs in the time dimension is evident from the identification guidelines provided in the bibliography for various technical patterns. For example, concerning '*flag*' patterns, Bulkowski (2005) and Edwards et al. (2007) report a maximum width of 3 weeks, whereas Pring (2002) states that these patterns can be as long as 3 to 5 weeks. The variation in the possible height of a pattern can be easily inferred from the first argument presented above. Since a technical pattern may occur in any stock price series and given that each series fluctuates on different price levels, a rational conclusion to make is that a technical pattern may appear at different price levels, and thus realise different heights. Further support on this argument can be found in Cervelló-Royo et al. (2015) and Arévalo et al. (2017) by the manner the template grid for recognising 'flags' is fitted over price windows that correspond to different price levels.

(v) *Similar trading volume for each pattern.* This is another inference from the identification guidelines provided in the bibliography on TA, according to which the volume during the formation of a particular pattern should evolve in a particular manner. For example, in the case of the '*head-and-shoulders*' pattern, the volume trend gradually diminishes during the formation of the pattern and expands substantially on the breakout (Pring, 2002; Bulkowski, 2005). Similar volume patterns are also suggested for other technical patterns. Thus, it can be deduced that the volume evolution during the formation of each pattern has to be 'similar' too.

3. Data

The initial dataset was extracted from Bloomberg and consisted of adjusted daily closing prices for 1920 NYSE stocks for the requested period 3-Jan-2006 until 31-Dec-2015. We applied a filtering process that closely follows Marshall et al. (2009) and Sharma and Narayan (2014). More precisely, we have excluded stocks that had prices either less than \$5 or greater than \$500, stocks that had at least one year with a proportion of nontrades greater than 5% and stocks with at least four consecutive days of missing values. The aforementioned approaches ensure that results will not be influenced by stocks with unduly high or low prices and stocks with insufficient trading activity. After adopting the above filtering procedure, the finalised dataset consists of $\mathrm{T} = 2,517$ daily close prices of $\mathrm{I} = 640$ stocks, i.e. $\{C_{t,i}; t = 1, \ldots, \mathrm{T}; i = 1, \ldots, \mathrm{I}\}$. For these stocks we have also retrieved the following daily series: open prices $(O_{t,i})$, low prices $(L_{t,i})$ and high prices $(H_{t,i})$. This paper also considers the relative volume $(V_{t,i})$, which is the ratio of the number of shares traded to the number of shares outstanding (Campbell et al., 1993).[2] Remaining missing values were filled as follows. Remaining missing close values were filled with linear interpolation, open missing values were filled with the previous close price whilst high and low missing values, for a particular day t, were filled with $\max(O_t, C_t)$ and $\min(O_t, C_t)$ respectively.

4. Methods

Our methodology can be briefly described as follows. Let $\boldsymbol{Q}_{1\times N}$ be a price pattern with a length of N days, that has been identified during the period from time $t = t^* - N + 1$ until time $t = t^*$ on a particular stock, i. The expectation to be formed at time t^* on the further evolution of this price series will be based on other historical patterns, similar to $\boldsymbol{Q}$, that can appear at different series, on different price levels and with different lengths. In the rest of this paper we will interchangeably refer to these $\boldsymbol{Q}$ patterns as '*query patterns*' or simply $\boldsymbol{Q}$s, whilst the terms '*similar historical subsquences*' or simply '*target patterns*' will be used to refer to the historical patterns that are 'similar' to a query.

The derivative subsequence DTW algorithm is implemented to identify historical subsequences in the entire dataset, which are similar to a query pattern and occurred prior to time t^*. The subsequence version of DTW is designed to identify subsequences, similar to a query, that may differ in length. Furthermore, the derivative version of DTW identifies similar patterns that may

[2]The choice of using the relative volume instead of the number of shares traded can be justified for two reasons. First, in this paper volume 'peaks' will be used to determine the ending point of a candidate query pattern. The relative volume measures the trading volume relative to the capacity of the market to absorb volume. While, inferences for the liquidity of an asset can be made by examining the number of shares traded, a 'peak' on the relative volume series will indicate a relative low number of available shares and that a further, sudden increase in demand could affect substantially the price of the stock. Secondly, in this paper we will also consider the similarity between volume subsequences of different stocks. Thus, for comparison reasons, the relative volume has been chosen as a more appropriate volume measure than the total number of shares traded.

occur at different price levels. Apparently, combining the aforementioned DTW modifications deals with argument (iv), whilst searching for similar historical subsequences in the entire dataset deals with argument (i) (see section 2). Additionally, in order to find similar historical subsequencies, the similarity between the trading volume of $\boldsymbol{Q}$ and those of the target patterns is also considered, dealing with argument (v). Arguments (ii) and (iii) are embedded to our proposed methodology when identifying the starting and ending times of $\boldsymbol{Q}$s respectively.

A bullish or bearish class is assigned to each one of the target patterns after examining the price behaviour that follows their formation. The predicted class to be assigned to a specific query is determined by the type of target patterns that forms the majority. The proposed methodology described above is free of any specific technical pattern's characteristics and thus it deals with our argument that technical patterns which currently exist in the bibliography should not be expected to be exhaustive. If the main principle of TA is valid (that history tends to be repeated) then it should be expected that the proposed methodology will give accurate predictions most of the times.

This section presents the main parts of our methodology and is organised as follows. Section 4.1 describes the procedure for identifying the $\boldsymbol{Q}$s. Section 4.2 presents the subsequence DTW algorithm. Finally, section 4.3 presents the labelling procedure that is adopted to assign bullish or bearish classes to the $\boldsymbol{Q}$s and their target patterns, and the manner predictions are made for $\boldsymbol{Q}$s.

4.1. Defining the query pattern

The first step of the proposed methodology involves the specification of the query pattern, $\boldsymbol{Q}$. More precisely, a procedure is required to identify its starting and ending points. Lets assume that the query pattern for a particular stock, spans over a time interval from $t = t^* - N + 1$ until time $t = t^*$. For defining the ending point, t^*, argument (iii) is considered, which suggests that increased volume is usually required to confirm a pattern's formation. Pring (2002) states as a major technical principle that volume is always measured relative to its recent past. Thus, we use a short-term rolling window of 30 days to identify times where the relative volume distances more than three standard deviations above its mean. The times that the above condition is satisfied will define all ending points of the $\boldsymbol{Q}$ patterns that will be considered later in the empirical part of our analysis.

Subsequently, the starting point of the query pattern has to be defined as well. As we have argued in section 2, patterns are sequences of regional locals. Thus, it can be inferred that the starting point of a pattern should be also a local extrema. Given an ending time, t^*, for a particular query, we implement the (PIPs) algorithm to identify the starting point of a pattern. PIPs is a promising method to identify salient points on a time series and they have been used in many applications on time series data mining, but also in finance applications to identify technical patterns (Fu et al., 2007; Chen et al., 2013).

The algorithm for identifying PIPs on a price series, starts by characterising the first and the last observation as PIPs. Afterwards, the Euclidean distances between the initial two PIPs and all the remaining observations is being calculated. The observation with the maximum distance is being characterised as the third PIP. The algorithm proceeds and characterises as the fourth PIP the observation that maximises its distance from its two adjacent PIPs. The algorithm usually terminates when the requested by the user number of PIPs are identified.

In this study, after identifying the ending time, t^*, for a particular query pattern of stock i, we search for PIPs on the closing prices of stock i, by using the observations from time $t = 1$ until time $t = t^*$. The algorithm stops when the first PIP is identified within a time frame of $[t^* - 4 \text{ months} : t^* - 2 \text{ weeks}]$, which preceded time t^*. By using this time frame, $\boldsymbol{Q}$ patterns are restricted to have a length between 20 and 80 trading days. These values were chosen based on the average lengths of celebrated technical patterns that were provided in the bibliography (Bulkowski, 2005; Tsinaslanidis and Zapranis, 2016).

4.2. Subsequence Dynamic Time Warping

DTW is an algorithmic technique which is mainly used to find the optimal alignment between two given (time-dependent) sequences, under certain restrictions (Müller, 2007). Having a query series, $\boldsymbol{Q}_{1\times N} = \{q_n\}_{n=1}^{N}$, and a set of K target series, $\boldsymbol{Y}_{1\times M_k} = \{y_m\}_{m=1}^{M_k}$, with various lengths M_k, for $k = 1, \ldots, K$, the DTW algorithm can be used to find which one of the target series is most similar (or better has the minimum alignment cost) to $\boldsymbol{Q}$. Furthermore, the optimal alignment of these two similar sequences can be achieved with the use of the optimal alignment (or warping) path, Z^*. More precisely, Z^* is a sequence of elements $\{z_1^*, z_2^*, \ldots, z_\lambda^*, \ldots, z_\Lambda^*\}$ where each z_λ^* indicates a pair of observations, one from each series $\boldsymbol{Q}$ and $\boldsymbol{Y}$, to be aligned, i.e. $z_\lambda^* = (n, m)$.

However, in the context of TA, when assessing the performance of a particular technical pattern, the task that someone needs to accomplish, is to search for a particular pattern in all stocks that compose the dataset under consideration. This is actually a subsequence matching problem where the user needs to identify subsequences within a longer sequence (or longer sequences), $\boldsymbol{Y}_{1\times M}$ ($\boldsymbol{Y}_{1\times M_k}$), which are 'similar' to another shorter sequence, $\boldsymbol{Q}_{1\times N}$, where $N \ll M$ ($N \ll M_k$). Apparently, $\boldsymbol{Q}$ represents the technical pattern under consideration and $\boldsymbol{Y}$ are the stock price series composing the dataset. In this paper we deal with this problem by implementing the derivative subsequence DTW algorithm. The steps of this algorithm are subsequently described.

Lets assume that our task is to identify subsequences of $\boldsymbol{Y}_{1\times M} = \{y_m\}_{m=1}^{M}$ which are similar to the query sequence $\boldsymbol{Q}_{1\times N}$, where $N \ll M$. Following Tsinaslanidis and Zapranis (2016), initial price series are smoothed by adopting the *Savitzky-Golay* filter with a rolling window of 21 trading days and a cubic polynomial model (Savitzky and Golay, 1964). Smoothing series in pattern recognition tasks is a usual practice. Lo et al. (2000), argue that by this process, nonlinear relations are extracted by 'averaging out' the noise. The same authors also argue that smoothing mimics the way human cognition extracts regularities from noisy data and thus the skills of technical analysts in identifying particular technical patterns via visual assessment. Series $\boldsymbol{Q}$ and $\boldsymbol{Y}$ which will be inputs to the derivative subsequence DTW, result by standardising the derivative, as defined in Keogh and Pazzani (2001), of the smoothed price series.

The first step involves the calculation of a distance (or cost) matrix, $\boldsymbol{D}_{N\times M}$, where each element, $d_{n,m}$, represents the distance between q_n and y_m, i.e. $d_{n,m} = |q_n - y_m|$. The second step involves the calculation of the accumulated distance (or cost) matrix, $\widetilde{\boldsymbol{D}}_{N\times M} = \left\{\tilde{d}_{n,m}\right\}$, as

$$\tilde{d}_{n,m} = \begin{cases} d_{n,m}, & \text{if } n = 1 \\ d_{n,m} + \tilde{d}_{n-1,m}, & \text{if } n \neq 1 \text{ and } m = 1 \\ d_{n,m} + \min\left\{\tilde{d}_{n-1,m}, \tilde{d}_{n,m-1}, \tilde{d}_{n-1,m-1}\right\}, & \text{if } n \neq 1 \text{ and } m \neq 1. \end{cases} \tag{1}$$

After adopting Eq. (1), the row vector $\tilde{\boldsymbol{d}}_{N,m}$ will include the total costs of all M optimal warping paths. Each one of these paths should satisfy the boundary condition $z_1 = (1, \alpha)$ and $z_\Lambda = (N, \omega)$, where $1 \leq \alpha \leq \omega \leq M$. In other words, α and ω indicate the start and end of a subsequence of $\boldsymbol{Y}$ which is candidate to be similar to $\boldsymbol{Q}$. Here, similarity is defined if the warping cost is less than a user-defined cost threshold, $\tau_{\mathrm{p}} \in \mathbb{R}$. The subscript in the threshold, τ_{p}, is used to distinguish it from two other thresholds that we introduce later in this paper; τ_{v} (section 4.3) and τ_{n} (section 5.3). Thus, the set $\Omega^* \equiv \{\omega_l^*\}_{l=1}^{L}$ with the L most similar subsequences' ends results from

$$\Omega^* = \arg_m \left(\tilde{\boldsymbol{d}}_{N,m} < \tau_{\mathrm{p}}\right). \tag{2}$$

The final step involves the identification of the set $A^* \equiv \{\alpha_l^*\}_{l=1}^{L}$ consisting of the starting points for each one of these L similar subsequences. Let Z^l be the optimal warping path ending at time ω_l^*, i.e. $z_\Lambda^l = (N, \omega_l^*)$. Given $z_\lambda^l = (n, m)$, $z_{\lambda-1}^l$ is defined as

$$z_{\lambda-1}^l = \begin{cases} (n-1, 1), & \text{if } m = 1 \\ \arg\min\left\{\tilde{d}_{n-1,m}, \tilde{d}_{n,m-1}, \tilde{d}_{n-1,m-1}\right\}, & \text{otherwise.} \end{cases} \tag{3}$$

The process described in Eq. (3) terminates when $n = 1$, whereby $z_\lambda^l = (1, \alpha_l^*)$. Thus, the two sets, A^* and Ω^*, contain the starting and ending points of the historical subsquences which are consider to be most similar to the query pattern. Appendix A provides an example of the aforementioned procedure.

In our empirical analysis we will also consider the similarity between the volume evolution during the formation of a query pattern and the one realised during the formation of each one of the identified target patterns. Apparently, the corresponding volume series of the above patterns will differ in length. Thus, the DTW algorithm is adopted, and the cost of each time-warping alignment is calculated. This cost will be used as a similarity measure in the sense that the greater this cost is the more different the volume behaviours will be. Considering not only the price but also the volume similarity, is already justified in the fifth argument that has been presented in section 2.

4.3. Labelling and predictions

This section presents the labelling procedure that is followed in order to assign a class to each one of the identified query patterns and their historical target patterns. The class is assigned by considering the price evolution after the realisation of each pattern and each pattern is labelled as bullish, bearish or neutral. Classes assigned to target patterns will be used to make predictions for the queries, whilst classes assigned to queries patterns will be treated as actual classes and they will be used to assess the efficacy of these predictions.

The labelling procedure is designed as follows. Following Bulkowski (2005), ultimate high (low) is defined as the highest high (low) before prices decline (rise) by at least 20%. After the ending point, $t = \omega_l^*$, of a target pattern, l, we observe the following price behaviour and we identify the time that ultimate high and low occur (t^{uh} and t^{ul} respectively). Apparently, the horizon at which we examine this behaviour is bounded from time $t = \omega_l^* + 1$ until time $t = t^*$ which is the ending point of the query pattern. Examining the price behaviour after the realisation of a target pattern until the ending point of a query, ensures that predicting the class of the latter will not be influenced by future figures.

Three classes are considered in the labelling procedure; bullish, bearish and neutral. A bullish (bearish) pattern should signal a long (short) position, whilst a neutral pattern does not signal any position. Recall that the ending times of the patterns have been symbolised as t^* for a query and as ω_l^* for the target patterns. Regarding target patterns, the return generated from the open price at time $t = \omega_l^* + 1$ until the time t^{uh} (Eq. (4)) and t^{ul} (Eq. (5)), are calculated considering high and low daily prices accordingly. Regarding query patterns, Eqs. (4) and (5) are used after substituting ω_l^* with t^*, whilst t^{uh}, $t^{\mathrm{ul}} > t^*$.

$$r^{\mathrm{o}\to\mathrm{uh}} = \frac{H_{t^{\mathrm{uh}}} - O_{\omega_l^*+1}}{O_{\omega_l^*+1}} \tag{4}$$

$$r^{\mathrm{o}\to\mathrm{ul}} = \frac{L_{t^{\mathrm{ul}}} - O_{\omega_l^*+1}}{O_{\omega_l^*+1}} \tag{5}$$

Subsequently each pattern is labelled according to the following rule. Using the symbols $\vee$ and $\wedge$ for the logical 'Or' and 'And' respectively, a class is assigned to each pattern according to

$$\begin{cases} \text{Bullish}, & \text{if } \left(r^{\mathrm{o}\to\mathrm{uh}} \geq 5\%\right) \wedge \left\{\left(t^{\mathrm{uh}} < t^{\mathrm{ul}}\right) \vee \left[\left(t^{\mathrm{uh}} > t^{\mathrm{ul}}\right) \wedge \left(r^{\mathrm{o}\to\mathrm{ul}} > -5\%\right)\right]\right\} \\ \text{Bearish}, & \text{if } \left(r^{\mathrm{o}\to\mathrm{ul}} \leq -5\%\right) \wedge \left\{\left(t^{\mathrm{uh}} > t^{\mathrm{ul}}\right) \vee \left[\left(t^{\mathrm{uh}} < t^{\mathrm{ul}}\right) \wedge \left(r^{\mathrm{o}\to\mathrm{uh}} < 5\%\right)\right]\right\} \\ \text{Neutral}, & \text{otherwise.} \end{cases} \tag{6}$$

More precisely, for a pattern to be characterised as bullish it is required that it should be possible for an investor to generate a positive return at least 5% by taking a long position at the opening price immediately after the formation of the pattern and closing this position at the ultimate high whilst the latter should either occur before the time the ultimate low occurs or, if it does not, $r^{\mathrm{o}\to\mathrm{ul}}$ should fail to generate a return lower than -5%. The choice of 5% is based on the definition of '*5% failures*' in Bulkowski (2005). More precisely, Bulkowski described a failure when prices break out in the expected, according to the pattern, direction but reach the ultimate high or low less than 5% away from the breakout. He also assumes that the 5% move is sufficient to cover the cost of trading. The interpretation for the bearish class is analogous. Finally, patterns classified as neutral do not provide any significant profit since both $r^{\mathrm{o}\to\mathrm{uh}}$ and $r^{\mathrm{o}\to\mathrm{ul}}$ are not greater or equal to the threshold of 5%.

For a given value of τ_{p} the subsequence DTW will identify a set of historical subsequences which will be the most similar to a particular query pattern. This set will be composed by patterns that may have been realised in any of the price series of our dataset, at any time prior the formation of the query pattern. Adopting also the DTW algorithm on the volume series, allows the introduction of an additional threshold, dubbed τ_{v}, that will be used for defining the maximum acceptable volume similarity cost. Setting lower values for τ_{v} will filter further the initial set of target patterns, reducing their number for each query case. The predicted class to be assigned to a each $\boldsymbol{Q}$ pattern will be the class that the majority of the corresponding target patterns had been allocated to. However, the proportion of target patterns allocated to the neutral class is minor since it occurs only when the pattern is identified at the end of the training period and prices follow a 'parallel' trend. Thus, no query patterns are allocated to the neutral class and predictions are confined between the bullish and bearish classes. The labelling procedure is also adopted to query patterns in order to identify their actual class. Hence, for a given number of $\boldsymbol{Q}$s the overall accuracy of the algorithm can be computed by calculating the proportion of the correct class assignments (for both bullish and bearish classes) to the total number query patterns considered. Bullish and bearish precisions will also be estimated which are performance measures that focus on each class separately (see section 5).

5. Main results

The whole test period we consider in this study spans from 1-Jan-2010 until 31-Dec-2015 (2010-2015). This test period will be used to assess the performance of our proposed methodology. Recall that the preceding period, from 3-Jan-2006 until 31-Dec-2015, is also considered but only for the identification of target patterns. This is to ensure that there will be sufficient historical data to identify target patterns, especially for these queries that will appear at the beginning of the test period. Section 5.1 provides a first, overall picture of the proposed algorithm's ability to predict bullish and bearish classes, for various combinations of τ_{p} and τ_{v}. Subsequently, the whole test period is split into two subperiods of equal length; the 1^{st} subperiod (2010-2012) and the 2^{nd} subperiod (2013-2015). Section 5.2 repeats the procedure adopted in section 5.1 for the first subperiod. In addition, three characteristic combinations of τ_{p} and τ_{v} are identified which produce optimal results. Finally, Section 5.3 focuses on the second subperiod, and presents more detailed results for the aforementioned threshold combinations.

5.1. Results for the whole test period

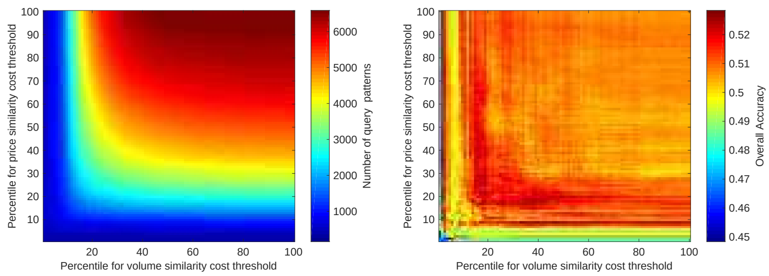
The total number of $\boldsymbol{Q}$s in the whole test period is 6,588. Having identified these queries, it is now possible to implement the subsequence DTW algorithm. However, since the price threshold τ_{p} (see Eq. (2)) has not been defined yet, we initially request from the algorithm to identify for each query pattern, the three most similar historical subsequences in each one of the 640 stock price series. Thus, for each $\boldsymbol{Q}$ pattern, $640 \times 3 = 1,920$ target patterns are identified, and the corresponding price and volume similarity costs to their $\boldsymbol{Q}$s are calculated. The term 'price' similarity cost for a target pattern, l, refers to the warping cost $\tilde{d}_{N,\omega_l^*}$ that results from the optimal alignment between this target pattern and its query, after adopting the subsequence DTW algorithm (see section 4.2), whilst the term 'volume' similarity cost refers to the alignment (warping) cost between the corresponding volume series of these patterns which results after implementing the DTW algorithm.

The performance of the proposed methodology is assessed for 10,000 combinations of τ_{p} and τ_{v} which is presented on an 100×100 grid in Fig. 1. The empirical distributions of the similarity costs for all target patterns were highly skewed to the right. Thus, for a better illustration, the grid is structured by using the percentiles of the cost's empirical distributions, rather than using equal distance values within the minimum and maximum values of these distributions.

Each percentile represents a similarity cost threshold that is used in order to consider the class assigned to a target pattern in predicting the class of its $\boldsymbol{Q}$ pattern. In other words, a target pattern is considered only if its price similarity and its volume similarity to the $\boldsymbol{Q}$ pattern are less than the values indicated by the corresponding percentiles. In general, when higher percentiles are used, more target patterns are considered for predicting each $\boldsymbol{Q}$ pattern's class, and thus the performance of more $\boldsymbol{Q}$ patterns is likely to be assessed.

The top-right corner of Fig. 1(a) corresponds to the extreme case where the 100^{th} percentiles are used and the number of $\boldsymbol{Q}$ patterns considered takes its maximum value of 6,588. Moving towards lower percentiles, the filters adopted become stricter and less target patterns are considered per case. Apparently, if there are no target patterns for a particular query, there cannot be a prediction for its class, and such cases are discarded. The overall effect to the number of $\boldsymbol{Q}$s considered after applying these filters is illustrated in Fig. 1(a).

In a similar manner, Fig. 1(b) shows the overall accuracy for each threshold combination. It is clear that when strict price similarity thresholds are considered (roughly less than the 10^{th} percentile) the algorithm performs poorly showing and overall accuracy of less than 50%. Relaxing this threshold by using values between the 10^{th} and 30^{th} percentiles allows the algorithm to consider more target patterns and its predictive performance is being improved. At this level of τ_{p}, it is also obvious that adopting a τ_{v} from the range between the 15^{th} and 50^{th} percentiles enhances even further the performance of the algorithm since it considers target patterns which are similar to the $\boldsymbol{Q}$s in terms of both the price and the volume. Relaxing further the threshold values and moving towards the upper right corner of the grid shows that the performance of the algorithm is deteriorated but still showing an overall accuracy that is marginally above the level of 50%. This performance behaviour could be mainly attributed to the fact that while increasing the number of target patterns for making predictions, at some point the algorithm starts to select target patterns that do not contribute positively to the predictive performance. These additional tar-



(a) Number of query patterns.

(b) Overall Accuracy.

Fig. 1. The different number of query patterns that are considered for various combinations of price and volume similarity cost thresholds, and the corresponding overall accuracies for the period 2010-2015.

get patterns are less similar to the queries compared to those that had been obtained by adopting stricter threshold values.

It is important to note that the overall accuracies presented in Fig. 1(b) are not comparable each other in terms of their significance since they result from different numbers of $\boldsymbol{Q}$ patterns. Thus, we proceed with the following hypothesis test where we test the null hypothesis that the true probability of the algorithm predicting bullish and bearish trends correctly is 50%, against the alternative hypothesis that this true probability is greater than 50%. The p-values from this test are illustrated in Fig. 2 and were calculated by

$$p\text{-value} = 1 - \mathrm{F}_{\mathrm{Bino}}(k; n, p), \tag{7}$$

where F_{Bino} is the binomial cumulative distribution function, k is the number of successful predictions out of the n number of $\boldsymbol{Q}$ patterns and $p = 0.5$. For better illustration, in Fig. 2, black dots signify the cases where p-value < 0.05 and the null hypothesis is rejected for a 95% confidence level. It is clear that significant cases appear when τ_{p} takes values roughly greater than the 20^{th} percentile. It is also worth noting that when volume similarities are not considered (100^{th} percentile for volume similarity cost), the performance of the algorithm is not significant. Clearly, these results highlight the importance of considering volume

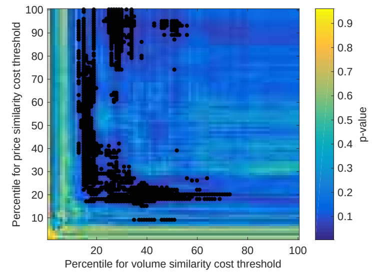


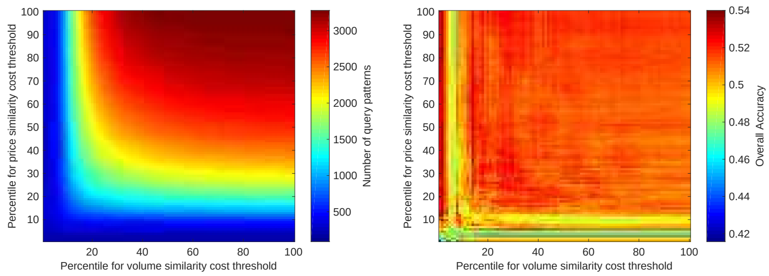
Fig. 2. The p-values resulting from testing the null hypothesis that the algorithm has an overall accuracy of 50% against the alternative that its overall accuracy is greater than 50%, for various combinations of τ_{p} and τ_{v}. Statistically significant cases for a 5% significance level are emphasised with black dots.

similarity in charting, and supports the fifth argument presented in section 2.

5.2. Results for the subperiod 2010-2012

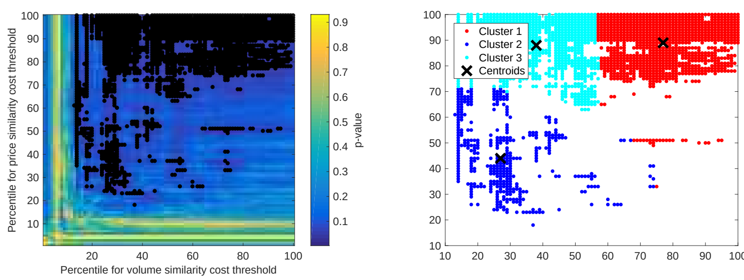
Regarding the subperiod 2010-2012, Fig. 3(a) is similar to Fig. 1(a) showing that when similarity thresholds take lower values, fewer $\boldsymbol{Q}$ patterns are considered, since those for which the algorithm does not find any target patterns are being discarded. Fig. 3(b) illustrates the algorithm's overall accuracy for various combinations of τ_{p} and τ_{v} which takes values above 50% in the vast majority of cases. The cases were the algorithm performs poorly (less than 50%) are those were very strict similarity thresholds (values less than those that correspond to the 10^{th} percentiles) are adopted.

Fig. 4(a) presents the corresponding p-values of the same one-side hypothesis test used in Fig. 2 and highlights the significant cases for a significance level $a = 5\%$. Applying a k-means clustering method on the significant cases of this grid allows us identify three clusters (Fig. 4(b)) with their centroids. These three cases represent combinations of high price and volume cost thresholds (cluster 1), low price and volume cost thresholds (cluster 2) and finally price cost threshold with high values but volume cost thresholds with low values (cluster 3). Table 1 presents the percentiles of the thresholds τ_{p} and τ_{v} of these three characteristic cases along with their actual values. In the following sections, the



(a) Number of query patterns. (b) Overall Accuracy.

Fig. 3. As for Fig. 1 but for the period 2010-2012.



(a) (b)

Fig. 4. p-values (as for Fig. 2 but for the period 2010-2012) are illustrated in Fig. 4(a). Fig. 4(b) illustrates three clusters of the significant cases and their centroids.

performance of the algorithm will be assessed in more detail, for these three threshold combinations, for the period 2013-2015.

Table 1. Percentiles of τ_p and τ_v for the three centroids presented in Fig. 4(b) along with their values.

Centroid of Cluster	Price Threshold		Volume Threshold	
	PCTL[a]	Value	PCTL	Value
1	89^{th}	22.5402	77^{th}	42.0205
2	44^{th}	11.0251	27^{th}	28.0557
3	88^{th}	21.8989	38^{th}	31.4274

[a] Percentile.

5.3. Results for the subperiod 2013-2015

In this section we scrutiny further the performance of the algorithm for the period 2013-2015 considering the three threshold combinations identified in the first subperiod (see Table 1). These combinations correspond to the three centroids of the three clusters of significant cases that were identified in section 5.2. Now, the algorithm is not restricted to search for the three most similar subsequences in each stock series for each $\boldsymbol{Q}$ pattern. Rather, it searches for all historical subsequences in all stock series for each $\boldsymbol{Q}$ pattern that satisfy a particular threshold combination. Apparently this is a more realistic approach allowing class predictions to be based on more 'relevant' information.

In order to assess the predictive performance per case, the overall accuracy is calculated as the proportion of the correct predictions to the total number query patterns considered. Furthermore, the overall accuracy is broken down by examining the bullish and bearish precisions. More precisely, the bullish (bearish) precision is the percentage of $\boldsymbol{Q}$ patterns classified as bullish (bearish) whose true class is bullish (bearish). In addition, we test the null hypothesis that these performance measures equal to 50% against the alternative that they are greater than 50% by using Eq. (7) to calculate the corresponding p-values. Results are presented in table 2.

During the subperiod 2010-2012, the overall accuracy of these three threshold combinations was statistically significant greater than 50%. However, Table 2

Table 2. Predictive performance for the three centroids of Fig. 4(b).

	Overall	Bullish	Bearish
Panel A: Centroid 1			
Correct Predictions	1,525	630	895
Total Predictions	3,080	1,347	1,733
Performance Measure (%)	49.51	46.77	51.64
p-value	0.6994	0.9905	0.0818
Panel B: Centroid 2			
Correct Predictions	910	397	513
Total Predictions	1,814	843	971
Performance Measure (%)	50.17	47.09	52.83
p-value	0.4533	0.9509	0.0361
Panel C: Centroid 3			
Correct Predictions	1,457	605	852
Total Predictions	2,942	1,289	1,653
Performance Measure (%)	49.52	46.94	51.54
p-value	0.6907	0.9851	0.1004

shows that the corresponding performance measures for the period 2013-2015 are not significant. Breaking down the overall accuracy into bullish and bearish precisions it is clear that this is mainly attributed to the poor bullish predictive performance, where all cases generate bullish precisions lower than 50%. On the contrary, bearish precisions are greater than 50% with relatively low p-values.

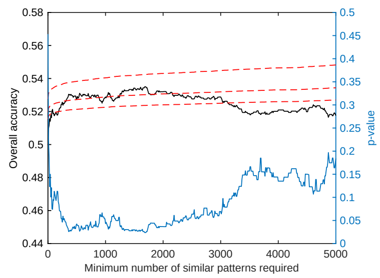
For brevity reasons, in the rest of this section we will focus on the combination of thresholds that corresponds to the centroid of the second cluster and explore further its performance.[3] Recall that the second centroid was the one

[3]The corresponding results for centroids 1 and 3 are available upon request.

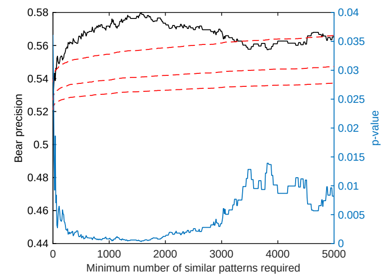
where lower values for τ_p and τ_v were used. This means that predictions for query patterns had to rely on historical subsequence that were more similar to the queries compared with the required similarity in centroids 1 and 3. In other words, the algorithm is more selective in which queries to make predictions and thus the total number of predictions is significantly lower than the corresponding total number of predictions made in the other two cases.

A further examination of the target patterns, based on which predictions were made, showed something interesting. Recall that the algorithm made predictions for 1,814 queries. However, for each prediction the number of target patterns considered, differed significantly. More precisely, there were many cases where the prediction of a particular $\boldsymbol{Q}$ pattern's class was based on just a few number of target patterns (1-10), whereas in other cases the number of target patterns was thousands. In practice, we could argue, that a technician would not consider the cases where there was not significant number of historical subsequences similar to the query pattern under consideration. In order to gain better insights in the effect that the number of target patterns has, we introduce a new threshold, τ_n, which represents the minimum required number of target patterns for making a prediction.

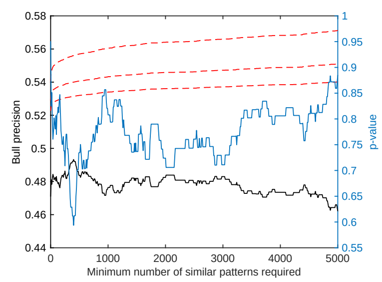
Fig. 5 presents the aforementioned performance measures for various values of τ_n (from 0 until 5,000 with a step of 10). More precisely, Figs. 5(a), 5(b) and 5(c) present the effect of this threshold to the overall accuracy, the bearish precision and the bullish precision respectively. One-tailed critical values that signify the rejection region for the null hypothesis are also provided and were estimated by, $1 - F_{\mathrm{Beta}}^{-1}(\alpha, n-k, k+1)$, where F_{Beta}^{-1} is the Beta inverse cumulative distribution function, α is the significance level, n is the number of trials and k is the number of successes. Setting $k = n/2$ and for significance levels 1%, 5% and 10% shows whether the null hypothesis can be rejected for confidence levels 99%, 95% and 90% respectively. It can also be seen that these critical values distance more from the level of 50% as higher values of τ_n are considered. The reason for this is that when using higher values for τ_n each performance measure is calculated on less predictions, and thus the performance measures



(a) Overall Accuracy.



(b) Bearish Precision.



(c) Bullish Precision.

Fig. 5. Performance measures for $\tau_{\mathrm{p}} = 11.0251$, $\tau_{\mathrm{v}} = 28.0557$ and for various values of τ_{n}. Dashed lines are the one-tailed critical values that signify the rejection area for significance levels of 1%, 5% and 10%. The right vertical axis shows the p-values.

should be improved further to reject the null for a given confidence level. Finally, p-values are also illustrated for a better understanding of the performance of the algorithm since they provide the marginal significance levels where we would be indifferent between rejecting and not rejecting the null hypothesis.

Fig. 5(a), shows that the overall accuracy of the algorithm becomes significantly greater than 50% (for a confidence level of 90%) if the user sets τ_{n} to take values between roughly 250 and 3,000. When τ_{n} takes even higher values the performance of the algorithm deteriorates as indicated by the corresponding high p-values. The effect that this threshold has at the performance of the algorithm can be interpreted as follows. Initially when τ_{n} takes very small values, the overall accuracy is calculated by considering also cases where predictions are made based on just a few target patterns. This contradicts with the main assumption of TA, that history tends to be repeated. In other words, for a query subsequence to be considered as a chart pattern there must be evidence from the historical price series that 'enough' similar subsequences occurred. Setting a higher value for τ_{n} discards these problematic cases enhancing at the same time the performance of the algorithm. However, setting even higher values for τ_{n} excludes $\boldsymbol{Q}$ patterns where the number of target patterns was enough to make accurate prediction, and thus the performance starts to exacerbate. Another interpretation that could be given to the latter point is based on the '*self destructive*' nature of a technical trading rule (Timmermann and Granger, 2004). According to this, a forecasting pattern, will self destruct after the time that is broadly used. When τ_{n} takes higher values (greater than roughly 3,000 in our results), $\boldsymbol{Q}$ patterns that remained for assessment are those for which the algorithm found a great number of similar reference patterns. It can be argued that these cases may represent patterns that have been occurred so many times in the past that the market started to recognise them and thus their predictive power eroded or even vanished.

Regarding the bearish precision, the algorithm predicts significantly better than 50% even if no τ_{n} is used (see Table 2). However, the use of this threshold enhances even further the bearish precision moving the p-values below 1% when the τ_{n} takes values again between roughly 250 and 3,000 (Fig. 5(b)). It is also worth to note the bearish precision of 58.08% that is produced when $\tau_{\mathrm{n}} = 1,570$ compared to the precision of 52.83% when no τ_{n} is used. Bearish precision deteriorates for even greater values of τ_{n}, similarly to the effect that larger values of τ_{n} have on the overall accuracy. However, p-values still remain lower than 1.5%. Finally, Fig. 5(c) illustrates the performance of bullish predictions which is not significantly greater than 50% even though it does improve slightly when τ_{n} is set to take values around 400.

The effect that τ_{n} has on the other two cases (centroids 1 and 3) is similar, in the sense that as it takes greater values the performance measures are initially improved whilst adopting even greater values affects these measures negatively. However, in the case of centroid 2, the performance of the algorithm is superior.

6. Additional results

This section presents a profitability analysis in order to see how the above performance measures are interpreted in terms of potential profits. For this assessment we adopt the following trading rule. When the algorithm makes a bullish (bearish) prediction, a long (short) trading position is taken at the open price of the following day. This position is closed on the earlier time between the time a stop loss condition is triggered and the time the ultimate high (low) is realised. In our analysis we have considered various values for stop loss orders plus a case without a stop loss condition. Apparently, it is practically impossible for someone to systematically close the initial position at the ultimate locals. However this approach will illustrate the average maximum potential profit that can be exploited through the proposed algorithm although the probability of realising this in practice is zero. Fig. 6 illustrates the average maximum trading profit following bearish signals, when τ_{p} and τ_{v} take values of the second centroid, for various values of τ_{n} and for stop loss orders of 1%, 3%, 5% and 7%. Increasing the value of the stop loss seems to initially decrease the average maximum profitability.[4] However, the case where no stop loss condition is used can be treated as a case where the stop loss order is set at a very high level that is never reached. Hence, using greater than 7% stop loss orders increases the maximum potential profitability up to the extreme case where no stop loss is considered. Overall, Fig. 6 shows that this average maximum profitability fluctuates mainly between 12.5% and 15%.

[4]This is aligned with other empirical results such as those reported by Arévalo et al. (2017) and Wu et al. (2017). Although different trading strategies had been used in these studies, the aforementioned authors found superior profitability for smaller stop loss levels.

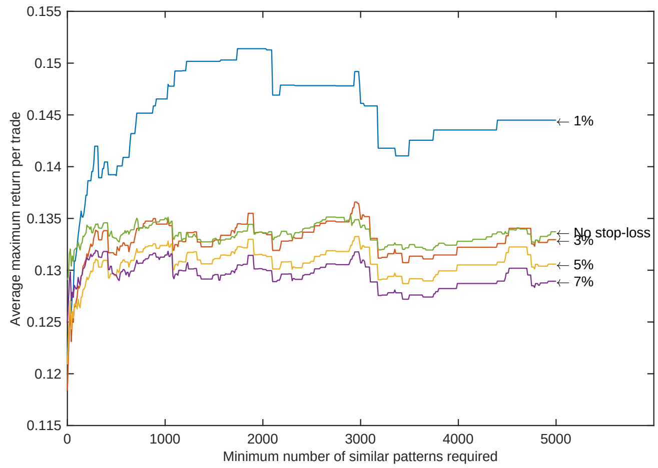


Fig. 6. Profitability of bearish signals when τ_{p} and τ_{v} take values of the 2^{nd} centroid.

In order to assess statistically these returns we adopt the following procedure. Each return reported in Fig. 6 resulted by adopting three different thresholds. The first two, τ_{p} and τ_{v}, are constant whilst the third one, τ_{n}, takes the values presented in section 5.3. Recall that, increasing the value of τ_{n}, the number of predictions is decreased. Thus, for every case we randomly select with replacement a number of days which equals to the number of the corresponding predictions made. The ultimate lows after these days are identified by adopting the procedure described in section 4.3. The aforementioned trading rule is adopted and the average maximum profit per trade is calculated. This procedure is repeated 1,000 times and the profitability resulting from our methodology is compared with the distribution of the average maximum returns that was generated by taking randomly short positions. With this comparison, simulated p-values are estimated. These p-values are fractions indicating the proportion of average maximum profits which resulted from random short positions that are greater than the mean maximum return realised from the initial methodology. The proposed algorithm would be considered profitable at a significance level of $\alpha\%$ if these simulated p-values are less than $\alpha\%$. The p-values when stop loss orders of 1%, 3%, 5% and 7% are all lower than 1% indicating that our results for bearish signals are statistically significant. The case of no stop-loss is also statistically significant although the p-values fluctuate below the 5% level. However, the profitability of bullish signals is not statistical significant where p-values take high values. The same analysis has been carried for the other two centroids and results were similar. More precisely, maximum potential profits from bearish signals were statistically significant, whilst those from bullish signals were not.

7. Conclusions and discussion

Heretofore, charting has been a common tool for making trading decisions whilst its efficacy has been under academic scrutiny. However, the subjective nature of TA, which mainly resides in the identification and interpretation of specific technical patterns, places significant barriers in assessing the predictability of TA. In order to mitigate this problem, various pattern recognition techniques, that identify specific technical patterns, have been historically developed, which remove a part of this subjectivity and allow for more objective assessments of the predictability of such technical patterns.

The contribution of this paper is two-fold. First, to the best of our knowledge, we formally state and justify for the first time a set of five common characteristics that technicians consider in practice. Secondly, we propose an algorithmic pattern recognition scheme which captures these characteristics. Our methodology is mainly based on the DTW algorithm and two of its modifications; the subsequence DTW and the derivative DTW. The approach followed in this study differs from that used in other studies. More precisely, our proposed methodology is not designed to identify specific technical patterns. Rather, it is designed to capture these common characteristics of the entire family of technical patterns and assess the performance of charting from a more general perspective.

While assessing the performance of the proposed algorithm, three different thresholds had been used; the maximum accepted price and volume similarity cost thresholds (τ_{p} and τ_{v} respectively), and the minimum number of target patterns required (τ_{n}) for predicting the class of a query pattern. We have explored the overall accuracy of the algorithm for various combinations of the first two thresholds and identified three clusters where the algorithm performed statistically significant: (a) when τ_{p} and τ_{v} take relative high values, (b) when they both take relative low values and (c) when τ_{p} takes relative high values but τ_{v} takes low values. For the three centroids representing these three clusters the performance of the algorithm was assessed in more detail. Our results indicate that the class predictability of the proposed algorithm is statistically significant only in the case of bearish classes. We found that maximum potential profits generated after the bearish predictions were also statistically significant. Finally, we showed that the algorithm's performance measures (especially those for bearish classes) can be further improved with the introduction of the third threshold,τ_{n}. We believe that this study can have valuable and practical implications in the academia and the financial industry in general. Our results suggests that TA in principal may add value in the trading decisions. We hope that this study will motivate future research to move towards more comprehensive assessments of the predictive performance of charting. We also believe, that our proposed algorithmic scheme may be used to design new pattern recognition trading rules with the view to supporting trading decision systems in the future.

Acknowledgements

We would like to thank two anonymous referees for their constructive comments and suggestions that substantially improved the final version of this article.

References

Arévalo, R., García, J., Guijarro, F., Peris, A., 2017. A dynamic trading rule based on filtered flag pattern recognition for stock market price forecasting. Expert Systems with Applications 81, 177–192.

Bo, L., Linyan, S., Mweene, R., 2005. Empirical study of trading rule discovery in china stock market. Expert Systems with Applications 28 (3), 531–535.

Bulkowski, T. N., 2005. Encyclopedia of chart patterns, 2nd Edition. John Wiley & Sons.

Campbell, J. Y., Grossman, S. J., Wang, J., 1993. Trading volume and serial correlation in stock returns. The Quarterly Journal of Economics 108 (4), 905–939.

Cervelló-Royo, R., Guijarro, F., Michniuk, K., 2015. Stock market trading rule based on pattern recognition and technical analysis: Forecasting the djia index with intraday data. Expert Systems with Applications 42 (14), 5963–5975.

Chen, C.-H., Tseng, V. S., Yu, H.-H., Hong, T.-P., 2013. Time series pattern discovery by a pip-based evolutionary approach. Soft Computing 17 (9), 1699–1710.

Dawson, E. R., Steeley, J. M., 2003. On the existence of visual technical patterns in the uk stock market. Journal of Business Finance & Accounting 30 (1-2), 263–293.

Edwards, R. D., Magee, J., Bassetti, W. C., 2007. Technical analysis of stock trends. CRC Press.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. The Journal of Finance 25 (2), 383–417.

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 (3), 347–364.

Fu, T.-c., Chung, F.-l., Luk, R., Ng, C.-m., 2008. Representing financial time series based on data point importance. Engineering Applications of Artificial Intelligence 21 (2), 277–300.

Keogh, E. J., Pazzani, M. J., 2001. Derivative dynamic time warping. In: Proceedings of the 2001 SIAM International Conference on Data Mining. SIAM, pp. 1–11.

Leigh, W., Modani, N., Purvis, R., Roberts, T., 2002a. Stock market trading rule discovery using technical charting heuristics. Expert Systems with Applications 23 (2), 155–159.

Leigh, W., Paz, N., Purvis, R., 2002b. Market timing: a test of a charting heuristic. Economics Letters 77 (1), 55–63.

Leigh, W., Purvis, R., Ragusa, J. M., 2002c. Forecasting the nyse composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. Decision support systems 32 (4), 361–377.

Lo, A. W., Mamaysky, H., Wang, J., 2000. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. The Journal of Finance 55 (4), 1705–1765.

Lucke, B., 2003. Are technical trading rules profitable? evidence for head-and-shoulder rules. Applied Economics 35 (1), 33–40.

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 (15), 1213–1221.

Müller, M., 2007. Information retrieval for music and motion. Vol. 2. Springer.

Nazário, R. T. F., e Silva, J. L., Sobreiro, V. A., Kimura, H., 2017. A literature review of technical analysis on stock markets. The Quarterly Review of Economics and Finance.

Neftci, S. N., 1991. Naive trading rules in financial markets and wiener-kolmogorov prediction theory: a study of" technical analysis". Journal of Business, 549–571.

Osler, C., Chang, P. K., 1995. Head and shoulders: not just a flaky pattern. Federal Reserve Bank of New York Staff Papers (4).

Park, C.-H., Irwin, S. H., 2007. What do we know about the profitability of technical analysis? Journal of Economic Surveys 21 (4), 786–826.

Pring, M. J., 2002. Technical analysis explained: The successful investor's guide to spotting investment trends and turning points, 4th Edition. McGraw-Hill Professional.

Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing 26 (1), 43–49.

Savitzky, A., Golay, M. J., 1964. Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry 36 (8), 1627–1639.

Sharma, S. S., Narayan, P. K., 2014. New evidence on turn-of-the-month effects. Journal of International Financial Markets, Institutions and Money 29, 92–108.

Timmermann, A., Granger, C. W., 2004. Efficient market hypothesis and forecasting. International Journal of forecasting 20 (1), 15–27.

Tsinaslanidis, P. E., Kugiumtzis, D., 2014. A prediction scheme using perceptually important points and dynamic time warping. Expert Systems with Applications 41 (15), 6848–6860.

Tsinaslanidis, P. E., Zapranis, A. D., 2016. Technical Analysis for Algorithmic Pattern Recognition. Springer.

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: Statistical Mechanics and its Applications 391 (16), 4136–4146.

Wang, J.-L., Chan, S.-H., 2009. Trading rule discovery in the us stock market: An empirical study. Expert Systems with Applications 36 (3), 5450–5455.

Wu, M.-E., Wang, C.-H., Chung, W.-H., 2017. Using trading mechanisms to investigate large futures data and their implications to market trends. Soft Computing 21 (11), 2821–2834.

Zapranis, A., Tsinaslanidis, P. E., 2012. 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 (7), 6301–6308.

Appendix A Subsequence DTW

This section presents an illustrative example for the methodology described in section 4.2. Fig. A.1 presents the daily, close price series for the Air Products & Chemicals Inc[5] traded in NYSE for the period from 3-Jan-2006 until 17-Jan-2014 (2,025 observations). It also presents the query pattern that will be used in the subsequence DTW algorithm and the three most similar historical subsequences that will be identified. The query pattern has been identified following the procedure described in section 4.1 and spans from 9-Jan-2013 ($t = 1,767$) until 2-May-2013 ($t^* = 1,845$).

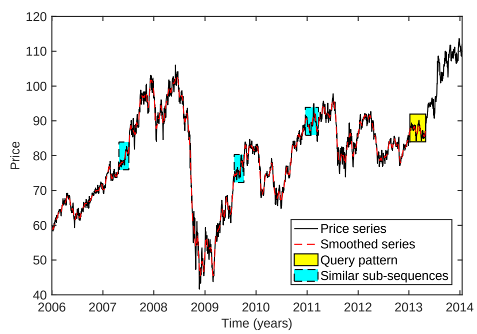


Fig. A.1. Example of three identified historical subsequences similar to a query pattern on a stock price series.

Fig. A.2 presents the subsequence DTW algorithm. The longer sequence, $\boldsymbol{Y}_{1\times 1,845}$ (Fig. A.2 (c)), results by standardising the derivative (Keogh and Pazzani, 2001) of the smoothed initial price series. The query pattern $\boldsymbol{Q}_{1\times 79}$ corresponds to the last 79 observations of $\boldsymbol{Y}$ (Fig. A.2 (a)). In order to identify the three historical subsequences which are more similar to the $\boldsymbol{Q}$ pattern, the cost matrix is calculated and presented in Fig. A.2 (b). Subsequently the accumulated cost matrix, $\widetilde{\boldsymbol{D}}$, is calculated by adopting Eq. (1) which is presented in Fig. A.2 (e). Fig. A.2 (d) presents the total costs of all optimal warping paths,

[5] Bloomberg Ticker: APD US Equity

Table A.1. Starting, ending points and warping costs for the three target patterns of Figs. A.1 and A.2.

l	α_l^*	ω_l^*	$\tilde{d}(79, \omega_l^*)$
1	331 (27-Apr-2007)	380 (9-Jul-2007)	13.51
2	900 (30-Jul-2009)	948 (7-Oct-2009)	10.74
3	1,251 (20-Dec-2010)	1,316 (24-Mar-2011)	13.28

$\tilde{\boldsymbol{d}}_{79\times m}$, where $m = 1, 2, \cdots, 1,845$. Observations ω_1^*, ω_2^* and ω_3^* realise the minimum warping costs with values 13.51, 10.74 and 13.28 respectively. These points signify the ending times of the three most similar subsequences to the query pattern. In other words, if we had set a marginally higher value than 13.51 for the threshold $\tau_{\rm p}$ the algorithm would still have returned these specific subsequences. The warping paths for these three ending points are defined by Eq. (3) and are illustrated in Fig. A.2 (e) on the accumulated cost matrix. This paths are used to identify the starting points α_1^*, α_2^* and α_3^* for the three subsequences. Table A.2 presents the aforementioned information regarding the three identified target patterns. Finally, Fig. A.2 (f) duplicates Fig. A.2 (c) with the difference that the similar subsequences are highlighted.

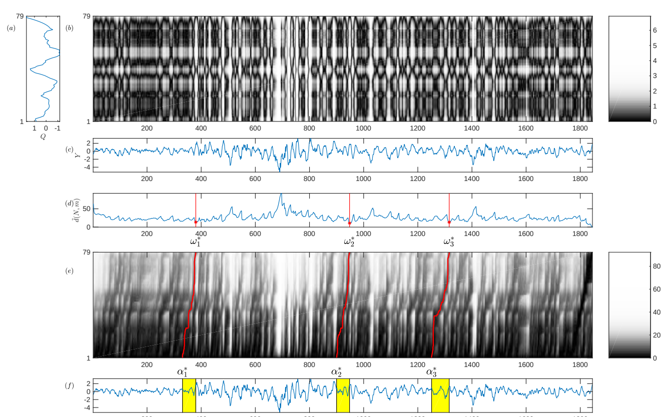


Fig. A.2. The subsequence DTW algorithm.