





# Portfolio Investments in the Forex Market

Przemysław Juszczuk<sup>(✉)</sup>  and Jan Kozak 

Department of Machine Learning, University of Economics in Katowice,  
1 Maja, 40-287 Katowice, Poland  
{przemyslaw.juszczuk,jan.kozak}@ue.katowice.pl

**Abstract.** Investing in the forex market seems to be an especially challenging task due to the large variety of dependencies related to instruments. Among the crucial aspects that should be considered is the correlation between the currency pairs. In this article, we derive a general investing schema considering the signal generation based on the well-known classification methods and verify the quality of these signals with the idea of portfolio building. To do so, we derive a two-stage process, where the first stage is devoted to deriving the classifier capable of generating the trading signals on the forex market. We use the set of the most popular market indicators, and the decision about the potential buy (or sell) signal is dependent on the values of these indicators. Eventually, we derive the classifier in which quality is measured on the basis of accuracy, recall, and precision. Further, we use signals generated by the classifier to adjust the account balance of the decision-maker and estimate the relation between the quality of classification and the final account balance.

Experiments are performed using the trading system implemented by the authors on the real-world data covering several years.

**Keywords:** Financial data · Portfolio · Classification

## 1 Introduction

Investing in the forex market seems to be an especially challenging task – a decentralized global market with currency pairs as instruments open a wide range of opportunities for investors. Significant volatility, a variety of tools like technical analysis, fundamental analysis, and sentiment analysis with the possibility of trade 24 h/day, five days a week seems to be a competitive investment opportunity compared to the well-known stock market or bonds market. At the same time, it is still considered a relatively safe option compared to the cryptocurrency market, which is often related to a very high correlation between instruments.

The possibility of trading and earning profits on the market is based mainly on the tools emerging from technical and fundamental analyses. Significantly, technical analysis is still considered a critical trading tool among decision-makers [7]. Trading systems based on fuzzy sets [10], neural networks [22], text mining

[17], tend to generate trading opportunities for a single currency pair. Even in the case where multiple instruments are taken into account simultaneously, it is still a relatively new field with high risk related to the strong correlation among the instruments. Examples of works extending the portfolio idea on the forex market are still difficult to find.

It is because classical financial models like Markowitz [15] or Merton [16] consider both expected mean and risk measured based on the correlation among instruments. Such correlation on the forex market is visibly higher than on the stock market, which can be a severe drawback. However, it is not clear how much impact such correlation could have on the overall results on the portfolio derived based on the set of currency pairs.

In this article, we try to fill the gap related to the portfolio problem on the forex market and move towards the investing process based on the set of instruments simultaneously, rather than consider a set of signals independently. To do so, we present the investing approach involving the classification methods used to generate signals. Further, these signals are considered simultaneously, leading to the situation that the decision-maker portfolio could include several currency pairs for a given time. Profit/loss from these currency pairs is adjusted to the decision-maker portfolio. However, we assume that no additional information about the correlation among instruments is considered.

Our main goal is to investigate whether the quality of signals classification performed on the data is directly related to the profits achieved by the decision-maker at the end of the investing period. Therefore, we implement a trading system including the signals classification module and investing module to verify that. Furthermore, we compare the classification quality performed on the well-known algorithms with the final account balance measured in dollars.

The article is organized as follows: in the next section, we briefly describe the works related to trading systems and the portfolio selection problem, mainly focusing on the forex market. The third section includes the proposed solution in which the overall flowchart of the proposed system and the investing process are described. The fourth section presents the experiments performed on the real-world data, while the last section concludes.

## 2 Related Works

Forex market (Foreign currency exchange) is the high-volatility market, where currencies are traded. There are numerous factors related to the present situation for the single currency, economic, political, and psychological factors that affect the current value of the currency pair [6]. Thus it is challenging to find an effective way to predict the future value and direction of the instruments on the forex market. There were numerous attempts related to the rule-based trading systems based on various market indicators like moving average [14] or Bollinger Bands [2]. However, due to the chaotic nature and high data volatility, many approaches combining the classical rules and technical indicators are combined with machine learning methods and optimization techniques. Examples of such

works using the genetic algorithms can be found in [5]. Complex hybrid models combining the market indicators and machine learning techniques can be found in [18]. Comparative study for both: genetic algorithms and various methods from machine learning adapted for several different trading rules was presented in [3].

In general, a lot of articles are devoted to the use of neural networks as an element of trading systems on the forex. Examples including classical neural networks [4] and self-organizing maps [18]. A detailed survey of articles used for financial forecasting related to deep learning, in general, can be found in [20].

Relatively small number of articles is devoted to the fundamental analysis [11], text mining [17], and news analysis in general [12]. The last element is also connected with the sentiment analysis on the market. At the same time, a little place is devoted to the portfolio analysis on the forex market [1, 19].

Numerous studies have described the profitability of technical analysis across several financial markets. There is no consensus related to the overall profitability of market indicators. An example of work pointing out the advantages of technical analysis and market indicators is [9]. Negative opinions about the efficiency of these tools were presented in papers like [21]. Despite diverse opinions, there is no doubt that the systems based on technical analysis and market indicators are very popular tools for practitioners.

### 3 Proposed Methodology

In this section, we describe the idea of our trading system based on the flowchart presented in Fig. 1.

Please note that the classification method and the investing algorithm described in this section can be freely modified and work independently from each other. Thus, the signal generation mechanism can be selected from the well-known methods from the literature or can be a simple approach based on a single technical indicator. Therefore, the overall system flowchart can be divided into four separate fragments:

- derive the data divided into learning and testing data;
- build a classifier based on the learning data;
- invest in the instruments from the testing data according to signals derived by the decision tree;
- update account balance according to the portfolio.

Our approach used the real-world data as a decision table with currency pair values and technical indicator values as conditional attributes. The decision class is represented by a discrete set including one of the following values: *STRONG BUY*, *BUY*, *WAIT*, *SELL*, *STRONG SELL*, where *BUY* (*STRONG BUY*) is the signal to open the position on the given currency pair, *SELL* (*STRONG SELL*) is the signal of closing the position (if the signal for this instrument was previously generated), while the *WAIT* is just a skip for the given instrument

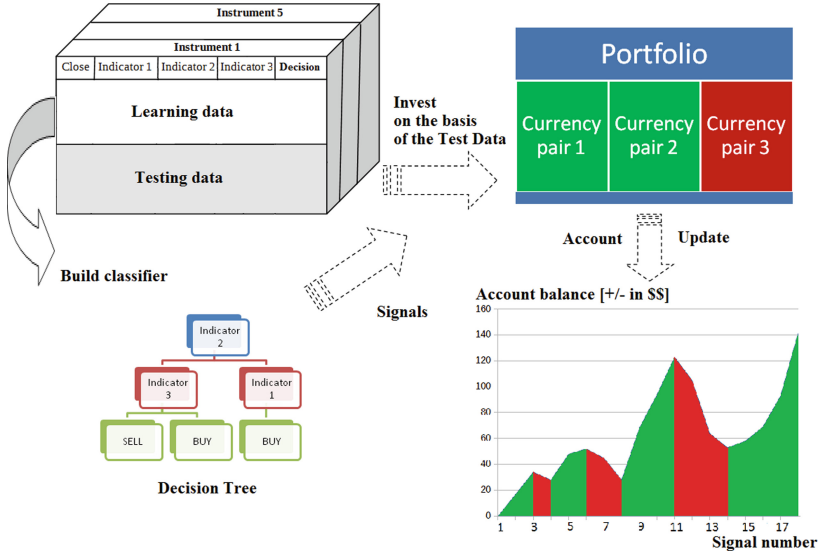


Fig. 1. The flowchart of the proposed system

at a time  $t$ . More details about the estimating signal value will be given further in this section.

In our system we use the following notations and concepts:

- $t$  will be a discrete moment of time (reading) in which the instrument (currency pair) value and the market indicator values are calculated;
- $T$  will be a time period, for which the whole investing process occurs.  $T$  consists of large number of  $t$ ;
- $CP$  will be a full set of currency pairs available in the system with  $cp_i$  as  $i$ -th currency pair;
- time period – will be a time, which has to pass between two successive readings;
- $I$  – will be a set of indicators available in the system;
- $PT$  – will be the portfolio of the decision-maker. This set is initially empty, however, the currency pairs  $cp$  are added to the portfolio, while the signals are observed on the market;
- $c$  – will be the counter indicating the number of readings, for which the given  $cp$  is already in the portfolio;
- $max$  – is the maximal number of readings, for which the  $cp_i$  can be present in the portfolio;
- $p$  – is the number of readings which must pass to evaluate the decision for a given signal.

The general idea of the single indicator (example for a CCI indicator) is presented in Fig. 2. The trading rule for this particular indicator CCI, and currency pair  $cp_i$  in time  $t$  calculated on the past  $n$  readings is defined as follows:

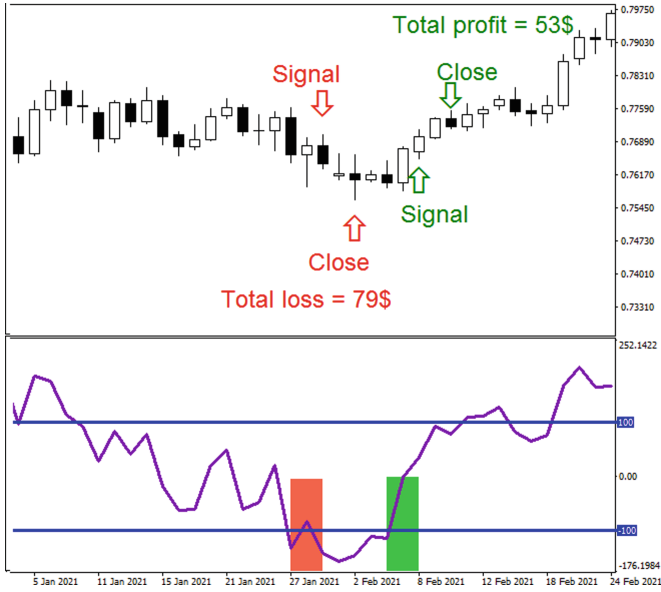


Fig. 2. Trading rule example – schema

$$f(CCI, c_{p_i}, t) = true \text{ if } CCI_n^l(t - 1) < l_1 \text{ AND } CCI_n^l(t) > l_1, \quad (1)$$

where  $l_1$  will be the predefined level for the particular indicator. In Fig. 2 this level can be observed as the blue bottom line in the lower part of the chart. The signal is generated when the indicator value (in two successive readings) crosses this predefined level. All indicators used in the research are based on the same idea where the signal is observed only if the value of the indicator crosses the predefined level. Whereas the decision for the data is calculated based on the formula:

$$decision = \begin{cases} STRONG BUY & \text{if } (price(t + p) - price(t)) > 3 \cdot \epsilon, \\ STRONG SELL & \text{if } (price(t + p) - price(t)) < 3 \cdot (-\epsilon), \\ BUY & \text{if } (price(t + p) - price(t)) > \epsilon, \\ SELL & \text{if } (price(t + p) - price(t)) < -\epsilon, \\ WAIT & \text{otherwise.} \end{cases} \quad (2)$$

This considers the difference between the instrument’s value for the time  $t$ , where the signal occurred, and the time after  $p$  readings. The difference between these two values indicates the decision class (*BUY*, if the difference was positive, *SELL*, if the difference was negative, and *WAIT* if the price was within the  $\epsilon$  range). Additionally, to initially clear the data for the classification process, the simple preprocessing schema was adapted. It covered the following steps:

- no empty values were observed in the data;
- all conditional attributes (except the Price) were discretized (classical interval discretization with the maximum of 20 intervals was performed);
- number of readings for all currency pairs was the same.

Especially the last condition will be crucial further in adapting the investing schema.

The preprocessed data acquired from the raw data is used as an input for the classification algorithm used to derive the decision for the system; we used the classical algorithm known from the literature – CART algorithm. One should know that the algorithm selection could visibly impact the results. Our initial assumption was to estimate the overall quality of the approach with the well-known approach from literature (without any additional modification). However, the selection of the algorithm should be investigated in further experiments.

Our goal in this trading system module is to initially estimate the quality of the classification based on the measures known from the literature. However, the classical confusion matrix used for the binary decision class is replaced here in such a way that a decision class including objects belonging to more than two classes is divided into subsets, and the following notation is introduced:  $TP$  – denotes all cases adequately classified to the selected class;  $TN$  – denoted all cases for which the proper assignment to the classes besides the selected class was made;  $FP$  – all cases incorrectly classified to the selected class;  $FN$  – all cases incorrectly classified to the classes besides the selected class. In addition, we used the following formulas:

$$acc(S) = \frac{TP + TN}{TP + TN + FP + FN}, \quad (3)$$

where  $S$  is the selected dataset. For the precision measure we used the equation:

$$prec(S) = \frac{TP}{TP + FP}, \quad (4)$$

while for the recall the following equation was used:

$$rec(S) = \frac{TP}{TP + FN}. \quad (5)$$

### 3.1 The Investing Process

Initially, acquired data is divided into two subsets. The first subset is used to derive the classifier, while the second subset measures classification quality. At the same time, the testing data will be used in the investing process (which will be performed simultaneously with the process of measuring the quality of classification). The schema for the investing process is presented in Algorithm 1.

The currency pairs can be added to the portfolio  $PT$  during the whole  $T$ . However, two different conditions should be satisfied – the decision for a given currency pair in reading  $t$  is  $BUY$  (or  $STRONG BUY$ ), and the currency pair

**Algorithm 1:** The investing process

---

```

begin
1  Create the empty set  $PT = \emptyset$ 
2  Select set of currency pairs  $CP$  and number of readings  $T$ 
3  Set additional parameters (number of decision classes and counter  $c$ )
4  for each reading  $t$  in  $T$  do
5      for each  $cp_i$  in  $CP$  do
6          Set decision class for  $cp_i$  according to formula (2)
7          if decision for  $cp_i$  is (BUY OR STRONG BUY) AND  $cp_i \notin PT$ 
            then
                add  $cp_i$  to  $PT$ 
                set counter  $c = 0$  for  $cp_i$ 
8          if decision for  $cp_i$  is (SELL OR STRONG SELL) AND  $cp_i \in PT$ 
            then
                remove  $cp_i$  from  $PT$ 
                update the value of account in $
9      for each  $cp_i$  in  $PT$  do
            increase counter  $c$ 
10     if  $c = \max$  for  $cp_i$  then
            remove  $cp_i$  from  $PT$ 
            update the value of account in $
11 close all opened positions
12 derive the final account balance to the decision-maker

```

---

is not in the portfolio at reading  $t$ . One should know that for any  $cp_i$  added to the  $PT$ , a counter  $c$  is set to 0. The counter  $c$  indicates the length (in a number of readings) for which the currency pair is already in the  $PT$ . We assumed that the instrument could be removed from the portfolio under the following circumstances:

- the decision for a currency pair in the portfolio is set to *STRONG SELL*;
- the maximal number of readings for the currency pair was achieved.

Eventually, after moving through all readings in  $T$ , all remaining transactions in the set  $PT$  are closed, the account balance is updated, and the final account balance is presented to the decision-maker.

In both cases, the current account balance measured in dollars is updated according to the following formula:

$$balance(t) = balance(t - 1) + K \cdot (price(t + p) - price(t)), \quad (6)$$

where  $balance(t)$  is account value in dollars in reading  $t$ ,  $price(t + p)$  is the instrument value after  $p$  readings, while  $price(t)$  is the instrument value in reading, where instrument was added to the portfolio  $PT$ . Finally, the  $K$  value is the constant value used to measure changes in dollars. For currency pairs with Japanese yen,  $K$  is set to 100, while for remaining currency pairs, it is set to 1000.

## 4 Numerical Experiments

In the experiments, we used 7 different currency pairs: AUDCAD, AUDCHF, AUDJPY, AUDNZD, AUDUSD, USDCAD, and USDCHF. For these instruments, we calculated values of six different market indicators: Bulls indicator, CCI (Commodity Channel Index), DM (DeMarker), OSMA (Oscillator of Moving Average), RSI (Relative Strength Index), and Stoch (Stochastic Oscillator). In Table 1 we can see summary including number of readings (size of  $T$ ) for each dataset. One should know that there is no consensus about the length of the financial data, which should be considered in the literature. However, we selected the data covering a large period. Thus the obtained results could be (in some limited way) extrapolated to other periods. At the same time, we selected the D1 and H4 time windows, limiting the noise's impact on the obtained results. The time window selection is closely related to the presented approach, and it should be rather changed in the case with the small time windows (like 5 of 1 min per reading).

**Table 1.** Number of readings for datasets used in the experiments

Data set	All		Rising		Side	
	D1	H4	D1	H4	D1	H4
No. of readings	2940	17666	827	4968	1499	8993

The whole experiment was divided into two separate phases. In the first phase, we perform the classification. However, we assumed that the training data would cover only 50% of all data; thus, it can be expected that overall results could be slightly worse than in the common division 70% training and 30% learning data. In the following section, we identify set 1 as the All data for the D1 time window, set 2 as the All data for the H4 time window. The number 3 and 4 are related to the data in the rising trend (Rising D1 and Rising H4, respectively), while the two last numbers 5 and 6 are included data for the consolidation (side trend) D1 and H4.

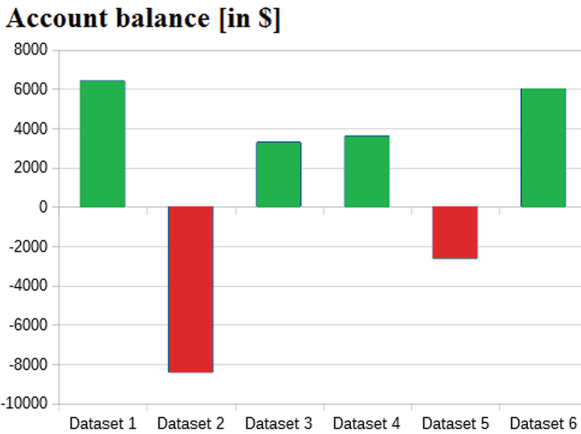
We performed the classification with the use of the CART algorithm, and the results for the aggregated data can be seen in Table 2. Thus both: *BUY* and *STRONG BUY* (the same occurs for the *SELL* and *STRONG SELL*) classes indicate the same price direction; we decided to aggregate this data. Please note that the results in the table represent the average classification efficiency for all instruments in the given dataset. One should remember that we used 50% of the data in the experiments as a testing set. The most important observation is the classification quality for the time window D1 (sets 1, 3, and 5), which are visibly higher than the remaining sets. It can be related to the fact that it is relatively easier to correctly classify the objects for a large time window (each new reading is derived every day), where the market noise is limited. However, since we aggregated the *BUY* and *STRONG BUY* (and the *SELL* and *STRONG SELL*)



classes, presented results leave visible room for improvements. It is essential when we differentiate the results according to decision classes. In actual results, we focused instead on driving the decision about the general instrument value direction (*BUY* or *SELL*); however, from the point of view of the decision-maker, it could be essential to know the strength of the movement as well. Moreover, the number of decision classes was set arbitrarily. Still, it is possible to include more decision classes corresponding to the trend’s strength (in such an example, the decision class depends directly on the range of the instrument value movement). In border cases, it is even possible to move towards the regression task, where instead of predicting the decision class, we rather expect the exact instrument value.

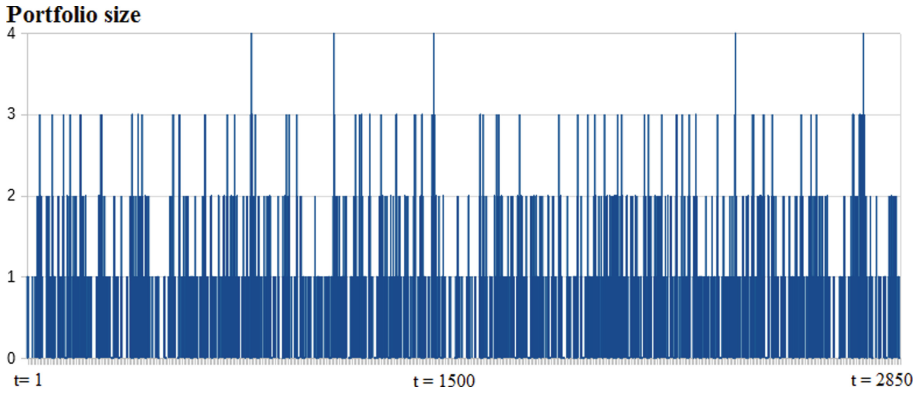
**Table 2.** Classification measures for all datasets (– denotes the situation, for which no object was classified to this class). *S. BUY* denotes the *STRONG BUY* class, while *S. SELL* denotes the *STRONG SELL* class

Set	Precision			Recall			Accuracy
	BUY/S. BUY	SELL/S. SELL	WAIT	BUY/S. BUY	SELL/S. SELL	WAIT	
1	43.83%	46.50%	-	56.00%	50.67%	-	41.67%
2	25.75%	28.25%	24.75%	75.00%	10.25%	24.50%	25.50%
3	47.75%	47.75%	-	71.00%	40.25%	-	46.25%
4	28.67%	25.00%	20.67%	23.67%	44.50%	12.00%	27.00%
5	43.80%	41.00%	-	51.00%	56.50%	-	41.80%
6	27.00%	43.25%	19.25%	26.75%	35.33%	32.50%	25.00%



**Fig. 3.** Preliminary results for the account balance calculation

The second part of the experiments is focused on the aspect related to the overall profit achieved in the decision-maker’s account balance at the end of



**Fig. 4.** Example portfolio composition variability over time

the investing period. These results are presented in Fig. 3. One should know that despite the overall good results (positive profit in four out of six) datasets, these results are ambiguous in connection to the classification results. First of all, the experiment design is somewhat limited, and some strict assumptions have been made. The most important observation is the fact that we do not have a clear answer about the impact of the classification quality on the account balance. It seems that the correlation between these two aspects is limited, and overall good classification quality does not positively impact the account balance. Moreover, presented classification results are here somehow averaged (average classification value overall used currency pairs). Thus there can be cases where a single currency pair draws the whole portfolio towards the positive account balance.

Additionally, we checked the variability of the portfolio (in terms of the size of the portfolio) for the example dataset. We selected four currency pairs. Thus the maximal portfolio size for this example is four. These results are presented in Fig. 4. It can be seen that the maximal portfolio size equal to 4 was observed only for several readings. Moreover, it lasted relatively short. Thus it can be assumed that the portfolio composition variability is very high.

## 5 Conclusions

In this article, we proposed the idea of the trading system, in which the trading signals were derived based on the classifier (decision tree). At the same time, instruments were added to the portfolio according to these signals. It leads to introducing the portfolio on the forex market. First, we presented the flowchart of the trading system, in which the data was divided into two sets. The first set derived the classifier, while the second test is used as an input for the portfolio module. Eventually, the account balance (measures in US dollars) was derived at the end of the investing process.

Our main goal was to investigate if there is a strict relation between the quality of classification introduced by the decision tree and the final account balance presented to the decision-maker. We used a large set of real-world data, including data in rising trends (bull market, where instruments values are rising) and the side trend (consolidation, where no dynamic moves are observed). In addition, two different time windows were used: H4 and D1. Despite the diversity of data used in the experiments, results were ambiguous, and no direct relation between the quality of classification and the account balance was observed. It concludes that these two problems (data classification and portfolio management on the forex market) should be considered independently. Thus the relatively good classification of the data does not imply the overall positive account balance at the end of the investing period.

As stated in the experimental section, the analyzed classification problem could be easily transformed into a regression case. Instead of predicting the decision classes, we rather focus on deriving the exact value of the instrument. The common assumption in the investing field is to know the general direction of the instrument value; however, for ongoing transactions learning the exact instrument value range could be vital.

## References

1. Amiri, M., Zandieh, M., Vahdani, B., Soltani, R., Roshanaei, V.: An integrated eigenvector-DEA-TOPSIS methodology for portfolio risk evaluation in the FOREX spot market. *Expert Syst. App.* **37**, 509–516 (2010)
2. Bollinger, J.: *Bollinger on Bollinger Bands*, McGraw-Hill, (2002)
3. de Brito, R. F. B., Oliveira, A. L. I.: Comparative study of forex trading systems builtwith SVR+GHSOM and genetic algorithms optimization of technical indicators, in: *Proceedings of the 2012 24th IEEE International Conference on Tools withArtificial Intelligence*, IEEE, pp. 351–358, (2012)
4. Carapuco, J., Neves, R., Horta, N.: Reinforcement learning applied to Forex trading. *Appl. Soft Comput.* **73**, 783–794 (2018)
5. Deng, S., Yoshiyama, K., Mitsubuchi, T., Sakurai, A.: Hybrid method of multiple kernel learning and genetic algorithm for forecasting short-term foreign exchanges. *Comput. Econ.* **45**, 49–89 (2015)
6. Edwards, D.: *Risk Management in Trading: Techniques to Drive Profitability of Hedge Funds And Trading Desks*, Wiley Finance (2014)
7. Gehrig, T., Menhoffs, L.: Extended evidence on the usage of technical analysis in foreign exchange. *Int. J. Fin. Econ.* **11**(4), 327–338 (2006)
8. Hryshko, A., Downs, T.: System for foreign exchange trading using genetic algorithms and reinforcement learning. *Int. J. Syst. Sci.* **35**(13–14), 763–774 (2004)
9. Hsu, P.-H., Taylor, M.P., Wang, Z.: Technical trading: is it still beating the foreign exchange market? *J. Int. Econ.* **102**, 188–208 (2016)
10. Juszczuk, P., Kruś, L.: Soft multicriteria computing supporting decisions on the Forex market. *Appl. Soft Comput.* **96**, 106654 (2020)
11. Kaltwasser, P.R.: Uncertainty about fundamentals and herding behavior in the FOREX market. *Phys. A Statist. Mech. App.* **389**(6), 1215–1222 (2010)
12. Kocenda, E., Moravcova, M.: Intraday effect of news on emerging European forex markets: an event study analysis. *Econ. Syst.* **42**(4), 597–615 (2018)

13. Korczak, J., Lipinski, P.: Evolutionary building of stock trading experts in a real-time system. In: Proceedings, 2004 Congress on Evolutionary Computation, pp. 940–947. IEEE (2004)
14. Larsen, F.: Automatic stock market trading based on technical analysis. Master's thesis, Norwegian University of Science and Technology (2007)
15. Markowitz, H.: Portfolio selection. *J. Finan.* **7**(1), 77–91 (1952)
16. Merton, R.: An analytic derivation of the efficient portfolio frontier. *J. Finan. Quant. Anal.* **7**(4), 1851–1872 (1972)
17. Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y., Ngo, D.C.L.: Text mining of news-headlines for FOREX market prediction: a multi-layer dimension reduction algorithm with semantics and sentiment. *Expert Syst. App.* **42**(1), 306–324 (2015)
18. Ni, H., Yin, H.: Exchange rate prediction using hybrid neural networks and trading indicators. *Neurocomputing* **72**, 2815–2823 (2009)
19. Petropoulos, A., Chatzis, S.P., Siakoulis, V., Vlachogiannakis, N.: A stacked generalization system for automated FOREX portfolio trading. *Expert Syst. App.* **90**, 290–302 (2017)
20. Sezer, O.B., Gudelek, M.Y., Ozbayoglu, A.M.: Financial time series forecasting with deep learning?: a systematic literature review: 2005–2019. *App. Soft Comput.* **90**, 106181 (2020)
21. Shynkevich, A.: Predictability in bond returns using technical trading rules. *J. Bank. Finan.* **70**, 55–69 (2016)
22. Yao, J., Tan, C.L.: A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing* **34**, 79–98 (2000)