IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 25 May 2023, accepted 2 June 2023, date of publication 9 June 2023, date of current version 10 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3285082

RESEARCH ARTICLE

A Novel Linear-Model-Based Methodology for Predicting the Directional Movement of the Euro-Dollar Exchange Rate

MAURICIO ARGOTTY-ERAZO[1,2], ANTONIO BLÁZQUEZ-ZABALLOS[1], CARLOS A. ARGOTY-ERASO[3], LEANDRO L. LORENTE-LEYVA[2,4], NADIA N. SÁNCHEZ-POZO[5], AND DIEGO H. PELUFFO-ORDÓÑEZ[2,6]

[1]Department of Statistics, University of Salamanca, Campus Miguel de Unamuno, 37007 Salamanca, Spain
[2]Smart Data Analysis Systems Group (SDAS) Research Group, Ben Guerir 43150, Morocco
[3]Facultad de Ciencias Económicas y Administrativas, Universidad de Nariño, Torobajo, San Juan de Pasto 52001, Colombia
[4]Facultad de Derecho, Ciencias Administrativas y Sociales, Universidad UTE, Quito 170147, Ecuador
[5]Centro de Posgrado, Universidad Politécnica Estatal del Carchi, Tulcán 040101, Ecuador
[6]College of Computing, Mohammed VI Polytechnic University, Hay Moulay Rachid, Ben Guerir 43150, Morocco

Corresponding author: Mauricio Argotty-Erazo (mauricio.argotti@usal.es)

This work was supported in part by the College of Computing, Mohammed VI Polytechnic University; and in part by the Smart Data Analysis Systems Group (SDAS) Research Group (https://sdas-group.com/).

ABSTRACT Predicting the price and trends of financial instruments is a major challenge in the financial industry, impacting investment decision-making efficiency for various stakeholders. Although numerous and effective artificial intelligence techniques have been applied to time series analysis, the prediction of exchange rate movements in the Forex market still necessitates parsimonious, interpretable, and accurate solutions. This paper presents a novel methodology for predicting the short-term directional movement of the euro-dollar exchange rate using market data, specifically by measuring price action. The proposed methodology prioritizes using market inflection points and the multidimensional nature of the differences between uptrends and downtrends to construct a linear discriminant function (LDA). The core of our methodology is our novel Linear Classifier Configurator (LCC) which includes stages for data preparation, feature selection, and detection of underlying structures. We validate the results and interpretations using the statistical power of parametric tests. The experiments use market data of the euro-dollar exchange rate in 15-minute and 1-week time frames. Additionally, we incorporate a collection of intraday winning trades provided by an algorithmic trading model applied between January 1999 and April 2023. The proposed LCC methodology achieves an out-of-sample classification accuracy of 98.77%, outperforming other methodologies based on sophisticated approaches such as Long Short-Term Memory (LSTM), Deep reinforcement learning (DRL), Wavelet analysis (WA), Sentiment analysis of textual content, Support Vector Machines (SVM), and Genetic Algorithms (GA). Furthermore, our methodology improves financial performance and reduces risk exposure in trading strategies, as well as it is useful in selecting variables and transferable to other financial assets.

INDEX TERMS Linear discriminant analysis (LDA), foreign exchange market (FOREX), machine learning (ML), supervised learning (SL), time series forecasting (TSF), trading systems.

I. INTRODUCTION

The FOREX or FX market, also known as the ''Over the Counter'' (OTC) market, is the world's largest financial market with a wide range of players, from financial institutions to individual investors, participating in it. In this market, currency values are defined and currency exchange is encouraged. Recently, the market has experienced rapid growth, making it easier to access currency trading, and introducing new technological challenges in the current trading

The associate editor coordinating the review of this manuscript and approving it for publication was Dost Muhammad Khan.

and regulatory models. This has awakened a growing interest among investors, practitioners, and researchers to predict future price behavior. The euro-US dollar cross (EUR/USD) is one of the most traded financial instruments, and is particularly interesting to investors because the strength of currencies is determined by the level of economic activity in the economies they represent [1].

Currently, exchange rate volatility and the effects generated on the value of currencies are strongly influenced by market liquidity and the interaction of economic, political, psychological, and non-fundamental variables [2]. The 2008 financial crisis and the Covid-19 pandemic have emphasized the importance of monitoring market dynamics to identify, anticipate, and mitigate adverse movements, reinforcing both financial and operational resiliencies. Market participants now require the ability to detect and monitor inflection points and the directional movement of exchange rates to manage their utilization effectively. Thus, informed decision-making is crucial in enhancing trading, minimizing risk exposure, and enabling the timely development of hedging operations to safeguard investment value.

Traditionally, asset value prediction revolves around two main approaches: technical analysis and fundamental analysis [3]. The former relies on studying historical market data and visually identifying patterns in price charts to predict their trends [4]. On the other hand, from a fundamental analysis perspective, proponents of the efficient markets hypothesis [5] argue that the price of a financial asset reflects all market information, asserting that price changes are random and, consequently, future behavior is unpredictable. Despite initial research supporting these postulates, subsequent studies by [6], [7], [8], [9], [10] demonstrate that historical data can predict returns and prices. As a result, the predictability of asset prices and trends remains a debated research topic.

In artificial intelligence, researchers have tackled price forecasting as both a supervised and unsupervised classification problem. A literature review reveals that this complex issue has been examined through statistical, machine learning, and deep learning techniques [11], as well as hybrid models [11], [12]. Additionally, alternative approaches have utilized sentiment analysis of news and text mining in social networks [13], [14], [15]. In recent years, the authors in [16] have observed a strong preference for deep learning techniques, especially in economic and financial applications. These techniques are increasingly used due to their high accuracy in classification and prediction tasks [17], [18]. Although these approaches can identify patterns or trends in historical data, they cannot directly explain the influence of predictor variables on price developments. Due to their black-box nature, in addition to model complexity and lack of readability, they pose problems of overfitting, generalization, and hyperparameter definition [19], [20], [21], [22], the solution of which entails significant efforts. Unlike the methods above, Discriminant analysis does not suffer from the same issue. It is less susceptible to parameter instability and offers competitive interpretability and precision when prior knowledge exists regarding the discriminant variables that effectively distinguish between groups [23].

In particular, linear discriminant analysis (LDA) is the most widely used classification method since it was proposed by Fisher [24], [25], [26]. Although new variants appear in the specialized literature, among the most diverse approaches, the following methods stand out: Quadratic (QDA) [27], Heteroscedastic (HDA) [28], Regularized (RDA) [29], Sparse Linear Discriminant Analysis (S-LDA) [30] and High-Dimensional Discriminant Analysis (HDDA) [31]. LDA is a multivariate parametric statistical method particularly relevant in the financial field for classification and prediction tasks. Besides enhancing the discriminative power of the predictor variables when they are useful for generating significant differences between classes, the main advantage of LDA over other techniques is its ability to incorporate multiple quantitative variables in the analysis. Thus, classification can be approached as a dimensionality reduction problem or, ideally, by selecting the best subset of predictor variables to ensure model interpretability while achieving accurate classification between groups.

Empirical evidence has shown that discriminant analysis is effective in classification when predictor variables produce significant differences between groups. Several noteworthy studies use discriminant analysis and novel information to predict value-at-risk (VaR) and financial instrument prices. These studies use data from investor information [32], descriptive statistics [33], financial ratios [34], news headlines [35], [36], and fundamental data extracted from financial news (twits) [37]. The classification results show that using valuable information with high discriminant power improves the models' predictive power and classification accuracy. Discriminant analysis, as shown in [23], [37], [38], and [39], outperforms certain machine learning algorithms. This supports the notion that LDA is well-suited for forecasting asset price trends. By effectively capturing patterns in extensive datasets and leveraging the multidimensional distinctions between groups, LDA improves accuracy while maintaining interpretability in classification. However, discriminant analysis is less common than other artificial intelligence techniques in predicting price trends and logarithmic asset returns. One contributing factor to its limited usage is meeting the statistical assumptions necessary for its application. This challenge arises when dealing with asymmetrically and leptokurtically distributed data, preventing the variables from assuming a multivariate normal distribution [40], [41]. Although other econometric approaches have also explored this issue [42], [43], [44], [45], [46], [47], [48], [49], [50], the problem of non-normal data distribution in parametric models continues to be a topic of ongoing research and debate. This limitation hampers the use of specific techniques due to the data structure.

Given this context, there is no definitive solution to the problem of price trend prediction. While various approaches have been proposed, current research has not yielded a conclusive answer, leaving room for further investigation and potential breakthroughs. Consequently, there is a significant demand for the development of parsimonious, simple, and highly accurate forecasting models that can effectively identify patterns or trends in historical data for predicting future price movements. Despite the complexity of the relationship between predictor variables and price developments, formulating classification models that elucidate their influence remains a critical challenge, as it enables more precise, well-informed, and value-driven investment decisions [51].

Research papers thus far have predominantly explored this phenomenon with an emphasis on technique. These studies use a wide variety of data and variables (market data, technical and fundamental indicators) to discover valuable insights that improve forecasting accuracy. However, most of these studies overlook the context and the moments leading up to trend changes. Taking into account these observations, our approach departs from the methodologies employed in the cited articles.

The main contribution of this work, from the technical and statistical analysis field, is to provide a valuable methodological tool to build interpretable, parsimonious, and accurate classification models.

Our approach aims to improve investment decision-making by predicting short-term trends in the euro-dollar exchange rate while maintaining interpretability.

This study introduces a novel framework for understanding and predicting price direction by incorporating market turning points. It also introduces two new variables that effectively discriminate between upward (bullish) and downward (bearish) trends. Additionally, the study applies discriminant analysis to multivariate normal data drawn from an asymmetric and leptokurtic distribution, contributing to its originality.

The novelty of the proposed methodology lies in the systematic use of a well-defined set of procedures, as shown in Figure 1. These processes include data preparation, selection of discriminant and independent features, structure detection to validate the discriminant power of the selected variables, and integration of these partial solutions to construct a linear discriminant function. This function provides a more precise depiction of price action to predict future trends. Moreover, statistical significance tests duly support the quality and reliability of the results and interpretations in each stage of the proposed methodology.

Our proposed methodology achieves a remarkable out-of-sample classification accuracy of 98.77% using 15-minute trading sessions of the euro-dollar exchange rate market data from January 1999 to April 2023. This performance surpasses even the most sophisticated methods discussed in recent works, including [52], [53], [54], [55], [56], [57], and [58].

Consequently, the findings suggest that the change in direction is influenced by the average yield of closing prices and the slope of the regression line of these variations. Thus, at inflection points **ET** (where the change in trend takes place), an bullish/bearish movement is more likely to occur when the average yield decreases/increases above/below its historical average value and the slope of the regression line of the variations between closing prices is negative/positive. In addition, predictions can confirm buy or sell orders before placing them, which increases the probability of successful execution of orders. In predictor variables, our approach helps to identify support and resistance levels at which the price changes its trend. In addition, this method is suitable for trading strategies that work in different time frames.

This paper is structured in the following sections: Section II provides a synthesis of the state of the art of the field of study. Section III describes the structures of the data collections and the proposed methodology. Section IV presents the experiment setup and performance measures used. Section V provides the results and discussion of the executed experiments. Finally, section VI summarizes the study's conclusions.

II. RELATED WORKS

Accurate prediction of price movements in the foreign exchange market is crucial for making informed investment decisions. However, the lack of interpretability in predictive models can hinder the formulation of judgments before making investment decisions. While model development has traditionally emphasized technique and accuracy, there is a growing need to develop predictive approaches that strike a balance between accuracy and interpretability [59].

Ge et al. [60] and Aziz et al. [61] highlight numerous studies in the scientific literature that use statistical techniques, machine learning, and artificial intelligence algorithms to address the problem of price trend prediction. Commonly used approaches include artificial neural networks (ANNs), support vector machines (SVMs), wavelet analysis, textual content sentiment analysis, and genetic algorithms [61]. Deep learning techniques, such as LSTM and deep reinforcement learning, have been utilized to enhance prediction accuracy [62]. Additionally, several studies use SVM and genetic algorithms to reduce dimensionality and improve prediction accuracy [63]. Moreover, textual content sentiment analysis has been leveraged to understand the impact of news on market behavior and achieve a higher prediction accuracy [64].

Although these studies have improved the accuracy of predictions, the importance of the interpretability of the results has often been underestimated. In finance, the explanatory power of predictive models is essential in making investment decisions [65]. There remains a need to develop predictive approaches that balance accuracy with interpretability, generating more compelling results and supporting investment decisions with clear and understandable information [59].

In this context, discriminant analysis is a useful statistical technique that has proven its effectiveness in terms of accuracy and explanatory power in different financial applications [66]. Although less widely used in price prediction than machine learning algorithms, several studies use discriminant analysis to predict the direction of price movements in different financial assets, such as cryptocurrencies, exchange rates, stocks, investment portfolios, and stock market indices.

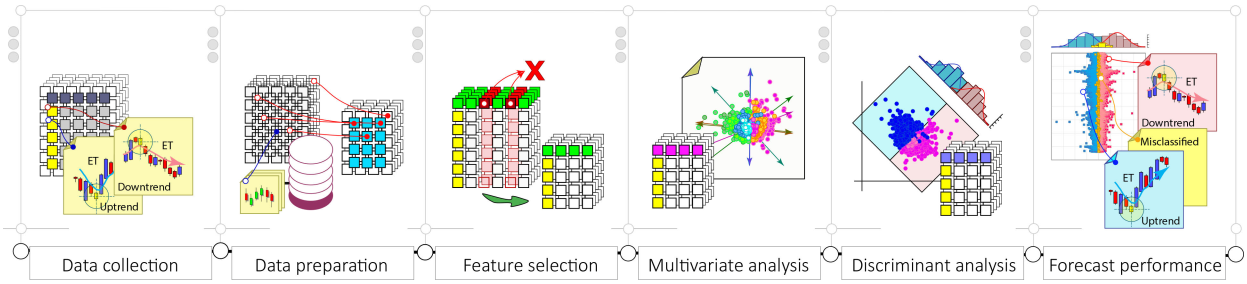


FIGURE 1. High-level diagram of the proposed methodology, Outlines the stages of data preparation, feature selection, and structure detection to predict the directional movement of the euro-dollar exchange rate. Additionally, it highlights the benefits of using the multidimensional nature of the differences between trend movements to construct a linear discriminant function.

A. CRYPTOCURRENCIES

Several studies have demonstrated the effectiveness of discriminant analysis in predicting digital assets. Grobys and Sapkota [32] achieve an 87% accuracy in classifying 146 digital assets using LDA and discriminant variables. The results show that reducing data dimensionality and selecting only the most informative variables improves model performance and allows investors to make informed decisions on high-default-risk assets. Gurrib and Kamalov [35] use sentiment analysis and LDA to predict the direction of bitcoin (BTC). With an accuracy of 58.5%, lower than that achieved by Grobys and Sapkota [32], they demonstrated the potential of Natural Language Processing (NLP) in extracting valuable information from heterogeneous data to design classification models for decision-making. Chen et al. [38] demonstrate the superiority of LDA and LogR over various machine learning algorithms (Quadratic Discriminant Analysis, Support Vector Machine, Random Forest, XGBoost, and Long Short Term Memory) in daily Bitcoin prediction, with an accuracy of 66%. These studies suggest that prediction accuracy does not necessarily depend on the complexity of the technique employed. However, using linear combinations of the original data as input variables may result in a loss of interpretability in predictions. In contrast, our proposed methodology overcomes this limitation, improving the transparency of predictions and facilitating informed decision-making.

B. EXCHANGE RATES

Several authors stress the importance of selecting the appropriate model and transforming the data to improve the accuracy of predictions. Zlicar et al. [67] achieved a 53.25% accuracy in predicting the direction of the most traded exchange rates using a Parzen sequential windowing (PW) algorithm for mapping price data. The results of this approach suggest that proper selection of the learning algorithm (among several methods evaluated, such as SVM, PW, and FDA) is critical to improving classification accuracy. On the other hand, Steurer et al. [68] focus on predicting daily exchange rate movements using machine learning techniques such as ANN, LDA, and LR. Their study also underscores the importance of selecting a suitable model and transforming the data to improve the accuracy of the predictions. Although the appropriate modeling technique choice depends on the problem and the data, it is worth noting that the most complex and attractive methods are not always the most effective. Therefore, it is essential to establish a coherent connection between the objective, the method, and the data used to obtain robust and reliable results. In the proposed methodology, this connection ensures the results' robustness and applicability in informed decision-making.

C. EQUITY INSTRUMENTS

Recent stock market studies have demonstrated the effectiveness of discriminant analysis in building predictive models and improving forecast accuracy. Mndawe et al. [37] use sentiment analysis to predict the stock price trend of South African companies. Their experiments reveal that LDA outperforms Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF) with 94% and 96% accuracy. The results highlight the relevance of using fundamental data from news headlines and social media (tweets). Meanwhile, Iworiso and Vrontos [23] propose an approach to predict the direction of the U.S. equity risk premium as a function of 14 financial performance variables. The experiments demonstrate the superiority of high-dimensional discriminant analysis (HDDA), achieving a 67% classification accuracy, while quadratic discriminant analysis (QDA) generates the highest cumulative returns. Bernardi et al. [36] use discriminant analysis to create sentiment indicators and reduce expected VaR violations. Their study focused on examining the influence of news on financial return prediction. The authors also demonstrated that incorporating exogenous quantitative variables from market news improves prediction accuracy. However, despite the accurate results obtained from these studies using fundamental data and discriminative exogenous information, the interpretability and transparency of the models are constrained due to feature extraction. In the proposed methodology, we handle these limitations through appropriate feature selection techniques.

Dao and Ahn [69] use a support vector machine (SVM) fuzzy logic model to predict stock market movements. Among other statistical methods, the authors use LDA to compare machine learning models (multiple discriminant analysis MDA, LogR logistic regression, CART regression tree, ANN artificial neural network, SVM, and fuzzy SVM models). Gorenc and Dejan [70] and Sharma et al. [71] focus on the use of statistical classifiers to improve the accuracy of predicting the movements of S&P 500 and ICICI Bank stock prices, respectively. Although these approaches use discriminant analysis as a benchmark to compare the effectiveness of other classifiers, feature extraction compromises the interpretability of the predictions of the best approaches. The proposed methodology ensures model transparency and interpretability by selecting readable features, which facilitates, facilitating decision-making processes.

D. INVESTMENT PORTFOLIOS

The use of discriminant analysis has proven effective in structuring investment portfolios with potentially profitable, low-risk assets.

Weiss [33] studies a portfolio classification approach that predicts the VaR and expected loss of 1500 bivariate portfolios containing stock, commodity, and currency futures data. The approach uses LDA to generate a parametric VaR and expected shortfall (ES) risk model. This model uses the descriptive statistics of the bivariate distribution of the portfolios as attributes, categorizing them according to the VaR and ES categories of the dependent variable. The results show a classification accuracy of 67.53% for VaR and 53.80% for ES. Although the interpretability of predictions and the model transparency are compromised, using descriptive statistics as predictor variables is remarkable. These statistics capture key aspects of the data distribution, leading to improved classification accuracy. By adopting this approach, the proposed methodology introduces two discriminant variables based on measuring average closing price returns and the slope of these returns at inflection points.

Okicic et al. [72] present an approach to stock selection and analysis based on LDA. This method seeks to facilitate the configuration and diversification of investment portfolios. The results emphasize the importance of identifying variables that can effectively predict the relevant categories for selecting suitable stocks. For their part, Zopounidis et al. [73] propose an approach using LDA to select financial ratios and compare its discriminant ability with that of other approaches. Studies show that using critical and relevant information improves classification accuracy by capturing distinctive characteristics of the dependent variable categories. The proposed methodology capitalizes on these findings by using discriminative measures to differentiate the categories of the dependent variable.

On the other hand, combining multiple techniques leads to substantial improvements in classification performance. Hwang et al. [74] develop a linear discriminant function to classify stocks into superior and inferior according to their financial performance. They confirm that the combination of data envelopment analysis (DEA) and discriminant analysis (LDA) significantly improves classification accuracy, reaching 85%. This outcome indicates that employing more robust approaches can benefit from incorporating pre-modeling procedures and complementary techniques.

Kwag et al. [34] introduced an approach that uses a logistic regression model to predict stock price direction. The model's prediction parameters are determined using LDA. Furthermore, their study reveals that financial ratios have a greater impact than macroeconomic indicators in explaining and predicting stock price variations.

Ogut et al. [75] propose a model for detecting price manipulation in the Turkish stock market. The study compares several machine learning algorithms, such as discriminant analysis, logistic regression, artificial neural network, and support vector machine. The results show that discriminant analysis helps detect price manipulation in the stock market to prevent fraudulent practices. These studies demonstrate the versatility and effectiveness of discriminant analysis in classification when combined with complementary approaches. The results improve with data preparation, feature extraction, or feature selection. The proposed methodology considers other complementary approaches to improve classification accuracy, such as underlying structure analysis.

E. STOCK MARKET INDEXES

Discriminant analysis has been used in notable studies to predict the direction of stock market indexes.

Leung et al. [39] evaluate the effectiveness of several classification methods (LDA, logit, probit, and probabilistic neural network) in predicting the direction of stock indexes. The study highlights the performance of the discriminant function, which achieved the highest accuracy of 68% when applied to the NIKKEI 225 index. These results indicate that using methods that capture and exploit discriminant relationships and patterns in the data can increase prediction accuracy.

Huang et al. [76] propose a combined approach that integrates multiple classification methods to improve prediction accuracy. Specifically, the model combines support vector machine (SVM), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and Elman backpropagation neural networks. By leveraging this combined approach, they achieved improved accuracy in predicting the movement of the NIKKEI 225 index, with a hit rate of 75%. These results suggest that combining different approaches capitalizes on their strengths and compensates for their individual limitations, resulting in more robust and accurate prediction models.

Azis et al. [61] presents a remarkable study summarizing the machine learning approaches used in the investigation of financial phenomena. This article details the most commonly used machine learning techniques in financial asset price prediction. It is worth highlighting that harnessing discriminant information leads to a significant improvement in prediction accuracy.

This brief literature review provides a clear and detailed rationale for why the proposed methodology can achieve more compelling results in predicting the evolution of the euro-dollar exchange rate. We also identify the issues and challenges these approaches face that remain to be solved.

The literature review supports the selection of discriminant analysis and the criteria defined in the proposed methodology. It offers insights into the strengths and limitations of the cited approaches and identifies best practices that have demonstrated high predictive ability in their respective fields.

We develop the proposed methodology with a comprehensive approach that addresses common challenges in current predictive approaches. We also carefully consider relevant factors such as market inflection points and the introduction of new variables that influence the prediction of financial asset prices. In addition, we address the need for interpretability, stability, generalization, scalability, accuracy, complexity, and transparency of the proposed model.

An exceptional aspect of this methodology is its emphasis on preserving the interpretability of the input variables. This feature allows investors to gain a deeper understanding of how investment decisions are made and instills confidence in the model's outcomes. As a result, the main contribution of this methodology is to offer a transparent and informed understanding of the price action of the traded exchange rate. By leveraging such information, this approach generates more compelling and valuable predictions, thereby empowering investors to make well-informed investment decisions.

III. MATERIALS AND METHODS

A. DATABASE

This research work uses the collection of intraday winning trades generated by an algorithmic trading model. These data are obtained from a trend-following trading strategy based on a momentum trading approach (Momentum Trading Strategy). The model captures the beginning of the direction of price movement and identifies it as an **ET** entry momentum. These reference points are useful for identifying the variables that best predict the onset of the direction of the future movement of the euro-dollar exchange rate. Thus, trades made on **ET** entry points are made in favor of the trend and produce positive profit margins.

Table 1, shows the performance of the trading strategy. It should be noted that positions with negative financial performance are excluded from the data preparation and analysis as they negatively impact the results of the study. The opening and closing of each contract is done on CP closing prices on a 15-minute time frame. This time frame is selected based on the trading strategy's financial performance and risk exposure. Each contract is executed at the **ET** instant at which the trend change occurs. Commission costs of $45/1000000 USD trade are charged after each transaction. The commission charged for holding a position open overnight is applied according to the type of position, Swap long $= -4.96$ and Swap short $= -0.96$. Slippage is not deducted from trades because delays in execution negatively affect the definition of **ET** study times and the definition of microtrends.

Consequently, the trades have a winning probability of 72% and generate net profits of USD 13162.85, during the sample period, from January 2006 to December 2020. Thus, the winning trades represent 75.34% of the 20894 trades made and take an average of 7 hours. Seventy-five percent of these trades achieved gross returns in the first two-thirds of the holding period. The Win/Loss Ratio with a value of 2.37 shows that the number of winning positions is 2 times higher than the number of losing positions. Profit Factor of 1.68 shows that $ 1.68 is earned for every dollar lost.

Statistically, the test of spurts and the Z-score of the trading system determines the existence of dependence between trades' results. Thus, a Z-score of -47.68 confirms a positive dependence between trades within a 99.7% confidence limit. This means, a smaller number of spurts than the normal probability function would imply, so that winning trades generate more winners and losing trades more losers. It is certainly possible to exploit the dependence relationships that exist between positions and improve the performance of the system, but the use of capital management measures and confirmation signals is beyond the scope of this paper.

1) MARKET DATA

The study uses two data sets with market information on the euro-dollar exchange rate. The quotes for the euro-dollar pair are obtained from the *Alpari International* trading platform. These quotes come from 15-minute and 1-week time frames. The first analysis sample spans 15 years, running from January 1, 2006 to December 31, 2020. The second sample is 18 years, from January 4, 2004 to June 27, 2021. The first data set is compiled into a $\mathbf{Q}\ \langle 375000 \times 7\rangle$ matrix consisting of 375000 observations and 7 metrics. The second set is compiled into a matrix $\mathbf{W}\ \langle 913 \times 7\rangle$ consisting of 913 observations and 7 metrics. Each observation, in both data sets, consists of Record Number (id), Date and Time, Opening Price (op_i), High Price (hp_i), Low Price (lp_i), Closing Price (cp_i), and Volume (Vo_i).

Figure 2 shows the price history (a) and log returns (b) of the euro-dollar exchange rate. The price and return series correspond to 15-minute and 1-week time frames. These values come from the data matrices **Q** and **W**. Figure 2a shows that the exchange rate, while exhibiting a range movement, tends to maintain a bear market regime during the analysis period. The trend effect of the historical price series is suppressed by the logarithmic returns. The series of returns are obtained from the following expression $\vartheta_i = ln\,(cp_i/cp_{i-1})$, where ϑ_i is the inter-price return calculated for each time frame and cp_i is the closing price associated to the time instant i.

The mean $\bar{\vartheta}$ of these returns is higher in the 15-minute returns than in the 1-week returns ($\bar{\vartheta}_{15m} = 8.42 \times 10^{-08}$, $\bar{\vartheta}_{1w} = -6.56 \times 10^{-05}$). Analysis of Figure 2b reveals that volatility, in both time frames, is highest in late 2008 and early 2009. Subsequently, volatility in the 15-minute time frame becomes strong in early 2014 and late 2015, while in the weekly time frame it becomes strong in mid-2010 and early 2020. The skewness coefficients (γ) of these

TABLE 1. Trading strategy performance report.

Specifications					
Symbol:	EUR/USD	Base Currency:	USD	Bars:	370903
Initial Deposit ($):	1000	Leverage:	1:100	Contract Size:	100000
Ending Balance ($):	14162.85	Traded Volume in Lots:	0.010	Pip Value ($):	0.10
Closed Trade P/L ($):	15800.68	Commissions ($):	-2370.53	Swaps ($):	-267.30
Minimum Margin Level (%):	6570.47	Risk Free Rate (%):	0.00	Market Risk (%):	0.063
Total Net Profit ($):	13162.85	Gross Profit ($):	32496.94	Gross Loss ($):	-19334.09
Profit Factor:	1.68	Expected Payoff:	0.63	Probability of Success:	0.72
Recovery Factor:	38.39	Sharpe Ratio:	0.13	Probability of Failure:	0.28
AHPR:	1.0003	GHPR:	1.0003	Risk per Trade (%):	0.14
Total Trades:	20894	Short Trades Won:	10502	Long Trades Won:	10392
Null Trades:	88	Winning Trades:	15026	Losing Trades:	5780
ROI:	13.16	Largest profit trade ($):	74.36	Largest loss trade ($):	-39.79
Analysis period (Years):	15.21	Average profit trade ($):	2.21	Average loss trade ($):	-3.12
Profit Trades (%):	70.32	Max consecutive wins:	33	Max consecutive losses:	21
Loss Trades (%):	29.68	Max consecutive profit ($):	278.76	Max consecutive loss ($):	-342.01
(Win / Loss) Ratio:	2.37:1	Average consecutive wins:	5	Average consecutive losses:	2
Mathematical Expectation:	$ 0.76	(Profit / Loss) Ratio:	0.71:1	Trading Advantage:	0.41
Standard Deviation Profits:	$ 4.10	LR Correlation:	0.99	The Runs Test Z-Score:	-47.68
Coefficient of Variation:	5.43	LR Standard Error:	397.52	Z Score Confidence Limit:	0.997
Balance Drawdown Absolute ($):	34.06	Balance Drawdown Maximal ($):	342.84	Balance Drawdown Relative (%):	2.42

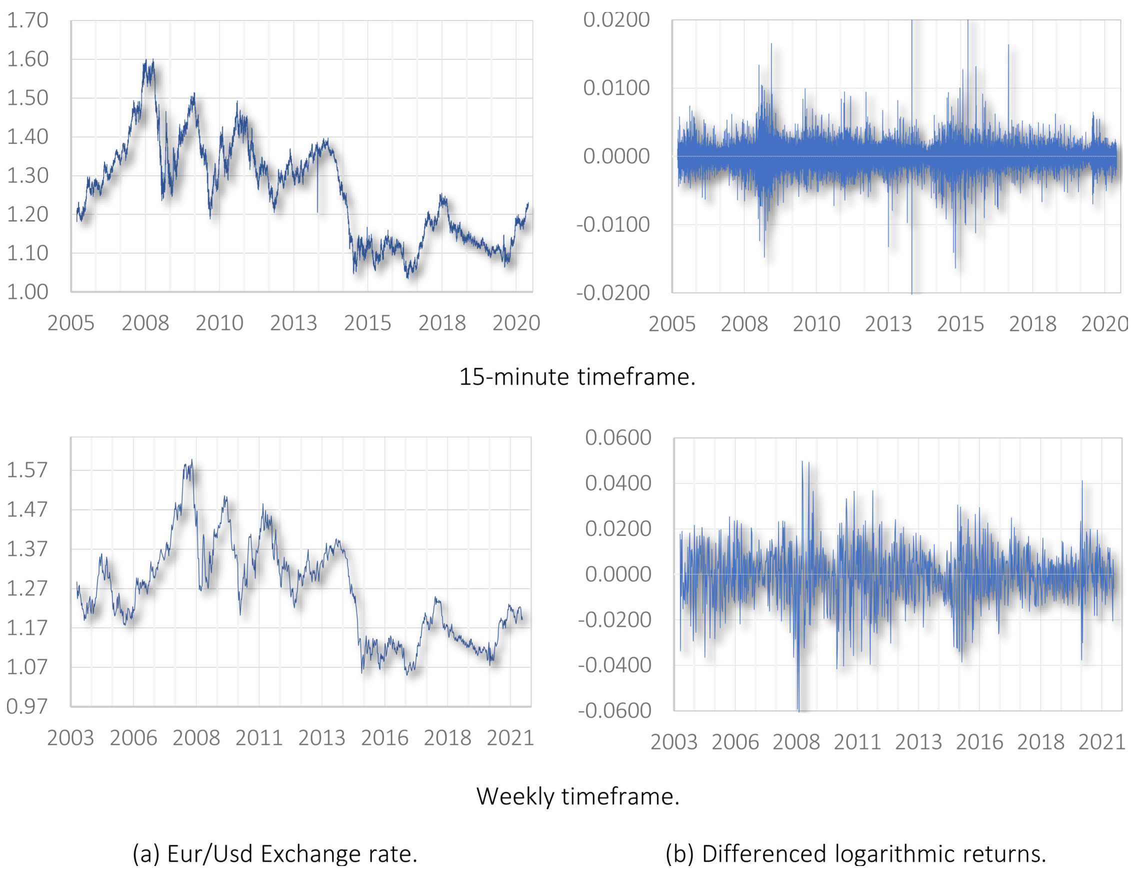


FIGURE 2. Historical closing prices (a) and logarithmic returns (b) of the euro-dollar exchange rate. Data is measured on a 15-minute time frame sampled over a 15-year time horizon, ranging from January 1, 2006, to December 31, 2020. Weekly data is sampled over an 18-year period, from January 4, 2004, through June 27, 2021.

returns are negative. The negative skewness is strongest in the 15-minute ($\gamma_{15m} = -1.02$, $\gamma_{1w} = -0.30$). The kurtosis coefficient (κ) is larger in the 15-minute returns distribution ($\kappa_{15m} = 7570.83$, $\kappa_{1w} = 1.60$). Thus, the distributions of log returns, of the euro-dollar exchange rate for both time frames, are asymmetric and leptokurtic. Furthermore, this suggests preparing the data to make the use of multivariate forecasting techniques feasible.

2) INTRADAY WINNING TRADE DATA

Intraday winning trades are obtained from a trend-following trading model. This dataset $\mathbf{A}\,\langle 14770 \times 18\rangle$ constitutes the starting unit for the detection of micro-trends in favor of the long-term trend movements to be forecasted. This collection consists of 14770 contracts and 18 metrics that measure the market behavior before and after the opening of each contract at the **ET** time. The units of analysis consist of 7380 long trades and 7390 intraday short trades. The analysis period is 15 years and runs from January 2006 to December 2020.

- *Indexes and variables.*

Trades are measured by 18 indicators. The first fifteen indices, referenced in the table 2, measure market performance 18 periods prior to the **ET** moment (on a 15-minute time frame). These indicators are calculated with a number of periods (18) defined by the best performance achieved with the ranking model. These indicators are grouped into three areas of analysis, price, volume and divergence between price and volume dimension variables. In addition, these indicators are used as candidate variables in the feature selection process. The best indices are used to predict the onset of the directional movement of the euro-dollar exchange rate.

The following three metrics (GPL_$, RP% M[1], and Sp.XP-EP) measure the performance obtained by the trade after the end of the contract's holding time. These indicators are used in the detection of underlying structures and the identification of the determinants of the differences between trend movements. The indicators (M.Sp.P.ET, LTT and, M.Sp) help to integrate the micro-trends of intraday winning trades (**A**) into a larger time frame (**W**). Consequently, the instances that are in favor of long-term trend movements are extracted and form the study population **O** (See Section III-B1).

Finally, the categorical variable "Trend" labels at time **ET** the direction of the exchange rate trend. The class labels used are "Upward Trend Movement" and "Downward Trend Movement". Sideways price movements are not considered in the present work.

- *Descriptive statistics of the indexes.*

Table 3 summarizes the descriptive statistics of the multivariate data matrix **A**. The mean of the data at the **ET** moments (where the change in direction of the preceding trend takes place) is higher for the variables Sp.Vo.18P, GPL_$, Dv.CP.RSI.18P, and s.Vr.RSI.18P. In addition, the first two indices (Sp.Vo.18P, GPL_$) report a volatile effect, while the last two have one of the lowest standard deviations. The mean and median of the data are different. The skewness and kurtosis statistics are far from zero. The indicators s.Vr.Vo.18P, x.Vr.Vo.18P, and Sp.Vr.Vo.18P provide evidence of disproportionate values in the tails of the distributions. Since none of these measures are close to the mean, this is strong evidence that the data set, for some indicators, exhibits asymmetric and leptokurtic behavior.

B. PROPOSED METHODOLOGY

This approach has practical applications in real-world scenarios, particularly in intraday trading of financial instruments like foreign exchange rates. Making informed decisions under uncertainty is crucial in this context. Our methodology allows analysts to leverage market inflection points and use data preparation, feature selection, and structure detection to develop effective classification models. These models facilitate the interpretation of price behavior and offer accurate short-term trend predictions for the traded asset. Incorporating these predictions into intraday trading frameworks improves the effectiveness of value judgments in investment decision-making and minimizes risk exposure. As the core of our methodology is the configuration and validation of the model so that it becomes suitable for linear classification, we name it Linear Classifier Configurator - LCC in short.

The proposed methodology allows predicting, in the short term, the beginning of the direction of the future movement of the euro-dollar exchange rate. The methodological design is based on five stages and four collections of data $\Xi\,\langle i \times j\rangle$ of size i rows by j attributes. Where i is the number of study moments **ET**, at which the change in direction of the preceding trend occurs and j are the numerical observable variables that measure the behavior of the closing prices before and after the moment **ET**. Figure 3 summarizes the procedure followed at each stage. First, the data collections are prepared, in order to make them analyzable and relevant, in fulfillment of the purposes established in each of the proposed phases. Secondly, from the technical analysis point of view, the predictors with the highest discriminatory power are selected on the basis of their statistical merit. Thirdly, the underlying structures are detected and analyzed to identify the determinants of the differences between trend movements. The usefulness of the selected predictors and the adequacy of the discriminant analysis is validated with the results of this stage. Fourth, the classification of the direction of the exchange rate movement is performed using discriminant analysis. Finally, the predictive power of the classification model is evaluated out-of-sample. The effectiveness of the proposed methodology is confirmed by the results of this stage. It is important to consider a suggestion when setting up new data sets before applying the proposed methodology (Figure 3). The benchmarks used for prediction should be measured using the variables specified in Table 2. This recommendation ensures that the features are defined based on measurements that effectively distinguish one observation from another. These benchmarks can be represented by turning points or the opening moments of successful buy or sell trades. These market entry signals must have been previously validated technically and financially through a trading strategy.

1) DATA PREPARATION

To maximize the usefulness of the data, make them analyzable and increase their potential in the processes of feature

TABLE 2. Indexes and variables.

Dimension	Name in short	Features	Formulas	
Price action	s.V.PT	Dispersion of returns from closing prices.	$\sqrt{\frac{\sum(\ddot{x}_i-\bar{\ddot{x}})^2}{(n-1)}}$,	$\ddot{x}_i = \frac{cp_i-cp_{i-1}}{cp_{i-1}}$
	x.V*PT	Average Return.	$\frac{\sum_{i=1}^{n}\check{x}_i}{n}$,	$\check{x}_i = ln\frac{cp_i}{cp_{i-1}}$
	Sp.CP.PT	Close-Price Slope.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = cp_i$
	Sp.Vr*CP.PT	Slope of Close-Price Variation.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = ln\frac{cp_i}{cp_{i-1}}$
	Sp.Vr.HL.PT	Slope of High-Low Price variation.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \frac{(hp_i-lp_i)}{(hp_{i-1}-lp_{i-1})} - 1$
	Sp.Vr.CO.PT	Slope of Close-Open Price variation.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \frac{(cp_i-op_i)}{(cp_{i-1}-op_{i-1})} - 1$
	s.Vr.RSI.PT	RSI variation.	$\sqrt{\frac{\sum(\dot{x}_i-\bar{\dot{x}})^2}{(n-1)}}$,	$\dot{x}_i = \frac{RSI(9p)_i}{RSI(9p)_{i-1}} - 1$
	Sp.Vr.RSI.PT	RSI variation Slope.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \frac{RSI(9p)_i}{RSI(9p)_{i-1}} - 1$
Volume	s.Vr.Vo.PT	Variation in Volume.	$\sqrt{\frac{\sum(x_i^{\circ}-\bar{x^{\circ}})^2}{(n-1)}}$,	$x_i^{\circ} = \frac{Vo_i-Vo_{i-1}}{Vo_{i-1}}$
	x.Vr.Vo.PT	Average Variation in Volume.	$\frac{\sum_{i=1}^{n}\hat{x}_i}{n}$,	$\hat{x}_i = \frac{Vo_i}{Vo_{i-1}} - 1$
	Sp.Vr.Vo.PT	Variation slope in Volume.	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \frac{Vo_i-Vo_{i-1}}{Vo_{i-1}}$
	Sp.Vo.PT	Market Volume Slope.	$\frac{\sum(x_i-\bar{x})(Vo_i-\bar{Vo})}{\sum(x_i-\bar{x})^2}$	
Divergences	Dv.CP-RSI.PT	Close-Price and RSI divergence.	$\frac{\sum(cp_i-\bar{cp})(y_i-\bar{y})}{\sqrt{\sum(cp_i-\bar{cp})^2\sum(y_i-\bar{y})^2}}$,	$y_i = RSI(9p)_i$
	Dv.Vo-CP.PT	Close-Price and Volume divergence.	$\frac{\sum(Vo_i-\bar{Vo})(cp_i-\bar{cp})}{\sqrt{\sum(Vo_i-\bar{Vo})^2\sum(cp_i-\bar{cp})^2}}$	
	Dv.Vo-RSI.PT	RSI and Volume divergence.	$\frac{\sum(Vo_i-\bar{Vo})(y_i-\bar{y})}{\sqrt{\sum(Vo_i-\bar{Vo})^2\sum(y_i-\bar{y})^2}}$,	$y_i = RSI(9p)_i$
After holding time	* GPL_$	Gross profit or loss for Long and Short trade.	$V_{\varphi} \cdot cs\,(xp - ep)$,	$V_{\varphi} \cdot cs\,(ep - xp)$
	* RP% M[1]	Money Risk by long or short trade.	$\frac{V_{\varphi}\cdot cs(lp-ep)}{sb}$,	$\frac{V_{\varphi}\cdot cs(ep-hp)}{sb}$
	* Sp.XP-EP	Open and close price trend line slope.	Slope $(\Delta p/\Delta t)$,	Long = $\frac{xp-sp}{xt-st}$, Short = $\frac{sp-xp}{xt-st}$
Long-term trend	M.Sp.P.ET	Market slope prior to ET moment (W1).	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \{cp_{i-96}, ..., cp_{i=\mathtt{ET}}\}$
	* LTT	Long-term trendline (W1).	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \{cp_{i=\mathtt{ET}}, ..., cp_{i=96}\}$
	* M.Sp	Market slope (W1).	$\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sum(x_i-\bar{x})^2}$,	$y_i = \{cp_{i-96}, \ldots, cp_{i=\mathtt{ET}}, \ldots, cp_{i=96}\}$
N/A	RSI$(9p)_{Smooth}$	Relative Strength Index.	$100 - \left(100/\left(1 + \bar{U}/\bar{D}\right)\right)$	
$^{\nabla}$Trend	Trend	0 = Uptrend: $(GPL\_\$ > 0,$ $Slope\,(\Delta p/\Delta t) > 0,$ $Long\ \ Trade)$ 1 = Downtrend: $(GPL\_\$ > 0,$ $Slope\,(\Delta p/\Delta t) < 0,$ $Short\ \ Trade)$		

*Notes: All variables are measured in 15-minute time frames, except (W1), which is measured in a 1-week time frame. (W1) = Formulas used for the definition of long-term trend movements. * = Measurements after the **ET** turning point (not involved in the feature selection process). Vo_i = Market volume in the period i or Number of times the price changes in each period i, in a 15-minute time frame. i = Analysis period, where $i \in \{1, \ldots, n\}$. cp_i = Closing price for the period i. hp_i = Maximum price of the period i. lp_i = Minimum price of the period i. op_i = Opening price in the period i. $RSI(9p)_i$ = Relative strength index for the period i, measured from the last 9 historical closing prices. $x_i = \{x_1, \ldots, x_n\}$ Time series, $x_i, i \in \{1, \ldots, n\}$. **ET** = Turning point in study time et. V_{φ} = Volume traded in lots. cs = Contract size 100000 units of base currency (Euro). sb = Initial account balance in US dollars. xp = Exit Price. sp = Entry Price. xt = Exit time. st = Start time. $\bar{U}$ = Average of upward price differences. $\bar{D}$ = Average of falling price differences. $^{\nabla}$Trend = Categorical variable.

selection, structure detection and model formulation, they are prepared and arranged in four data collections $\Xi\,\langle i \times j\rangle$. Where i is the number of **ET** times at which the change in direction of the preceding trend occurs, in 15-minute trading sessions, and j are the variables that measure their behavior before and after the trend shift. According to Figure 4, data preparation is summarized in four phases. First, the bullish and bearish micro-trends are categorized and arranged for the feature selection process. Second, the study population is extracted and prepared based on the micro-trends that are in favor of the long-term trend movements. The instances that meet this requirement are used to predict, in the short run, the

TABLE 3. Descriptive statistics of indices.

Measure	Mean	Std.Dev	Median	Min	Max	Skew	Kurtosis
Price action:							
s.V.18P	0.00051011	0.00034042	0.00042776	0.000044579	0.00439635	2.3730	11.7879
x.V*18P	-0.0000002	0.00019849	0.00000143	-0.00209697	0.00164601	0.0050	5.15348
Sp.CP.18P	-0.0000008	0.00025970	0.00000289	-0.00233251	0.00192982	-0.028	6.19084
Sp.Vr*CP.18P	-0.0000004	0.00003325	-0.0000002	-0.00040831	0.00033698	-0.055	6.48955
Sp.Vr.HL.18P	0.00630912	0.03272341	0.00538741	-0.58322522	0.72498519	0.6652	38.0406
Sp.Vr.CO.18P	0.00800454	0.46597394	0.01403601	-6.46007878	8.77988644	0.0522	31.9152
s.Vr.RSI.18P	**0.16739490**	0.07751317	0.15280451	0.031895221	1.33360205	**2.6943**	18.7865
Sp.Vr.RSI.18P	-0.0005816	0.00824177	-0.0004332	-0.04381362	0.08385499	0.2604	3.14648
Volume:							
s.Vr.Vo.18P	0.47597319	1.24018626	0.35617189	0.047287739	69.8530494	**40.416**	**1937.19**
x.Vr.Vo.18P	0.10736958	0.31392351	0.07387271	-0.12090153	16.9855561	**35.890**	**1639.09**
Sp.Vr.Vo.18P	0.00231944	0.05945863	0.00230310	-3.32210627	2.84634278	-1.024	**1542.24**
Sp.Vo.18P	**10.7184210**	**41.9964534**	5.06553148	-295.573787	357.643962	0.5038	6.78630
Divergences:							
Dv.CP-RSI.18P	**0.85151545**	**0.20628068**	0.93089873	-0.65551962	0.99928244	-2.903	9.93598
Dv.Vo-CP.18P	-0.01375776	0.51348876	-0.0276017	-0.97874454	0.9763092	0.0360	-1.1499
Dv.Vo-RSI.18P	-0.01580351	0.49230421	-0.0234383	-0.97742819	0.96606119	0.0382	-1.1572
After holding time:							
GPL_$	**2.174395396**	**2.51882690**	1.47000000	0.010000000	44.3000000	4.4641	34.7811
RP% M[1]	-0.00079658	0.01564524	-0.0017000	-0.22980000	0.17700000	0.8083	14.9109
Sp.XP-EP	0.000009683	0.03751680	-0.0000302	-0.54240000	1.03680000	1.1255	57.2277

direction of the future movement of the euro-dollar exchange rate. Third, the best set of observable variables is prepared and arranged for the detection of underlying structures. The relevance of the selected discriminant analysis and predictors is determined by the adequacy of the data. Fourth, a multivariate normally distributed data sample (*MND*) is drawn from the study population to overcome the asymmetric and leptokurtic nature of the observations and meet the requirements of the discriminant analysis.

- *Short-term trend movements.*

In this phase, the units of analysis are prepared for discretization into bullish and bearish micro-trends. All prepared data matrices employ these class labels as a factor or grouping variable. In addition, measurements performed on these instances are used for feature selection. Thus, the multivariate data collection $\mathbf{A}\,\langle 14770 \times 18\rangle$, consisting of intraday winning trades, from a trend-following trading strategy, are categorized into two class groups: micro uptrend and micro downtrend. The information used to perform this procedure is reported in the table 4. The discretization process assigns a class label based on two quantitative variables (Gross profit or loss and Trade slope) and confirms this assignment based on a true categorical variable (Trade type). Accordingly, an upward micro-movement occurs if a long trade generates a positive gross profit on a positive slope. A bearish micro-trend is defined by a short trade with positive gross profit on a negative market slope.

- *Long-term trend movements.*

At this stage the study population is identified and compiled into the **O** data matrix. These data are used to predict, in the short term, the onset of the future direction of the euro-dollar exchange rate. The matrix $\mathbf{O}\,\langle 1873 \times 2\rangle$ is formed by **i** moments of analysis $\mathtt{ET}$, at which the change of the preceding trend occurs, on which **p** chosen predictor variables $\hat{\mathbf{x}}_j$ are measured. This collection is the result of confirming each micro-trend, from matrix **A**, with a long-term trend movement, according to market information from data matrix $\mathbf{W}\,\langle 913 \times 7\rangle$. Consequently each change of direction of the preceding trend is confirmed 96 weekly periods before and after the time $\mathtt{ET}$. The results obtained during data preparation show that this number of periods is sufficient to confirm the dominance of the newly formed trend on a weekly time frame. Thus, the study time $\mathtt{ET}$, is selected as a valid instance, when the change of direction of the preceding trend, in 15-minute trading sessions, is confirmed with a continuation or a change of trend in favor of the dominant trend, in weekly trading sessions, otherwise the observation is discarded. It should be noted that short-term trend changes are part of long-term trend movements, but are not determinative of their formation. Note that the purpose of this study is limited to predicting the beginning of the trend direction in 15-minute trading sessions and, therefore, does not necessarily mean that such prediction is determinant of the beginning of the long-term trend movement.

Figure 5 shows a graphical explanation of the analysis unit selection process described in the table 5. The various long-term bullish and bearish market regimes, according to the study moments $\mathtt{ET}$, are formed by movements with continuity in the trend and movement with change in the direction of the preceding trend. It should be noted that the study

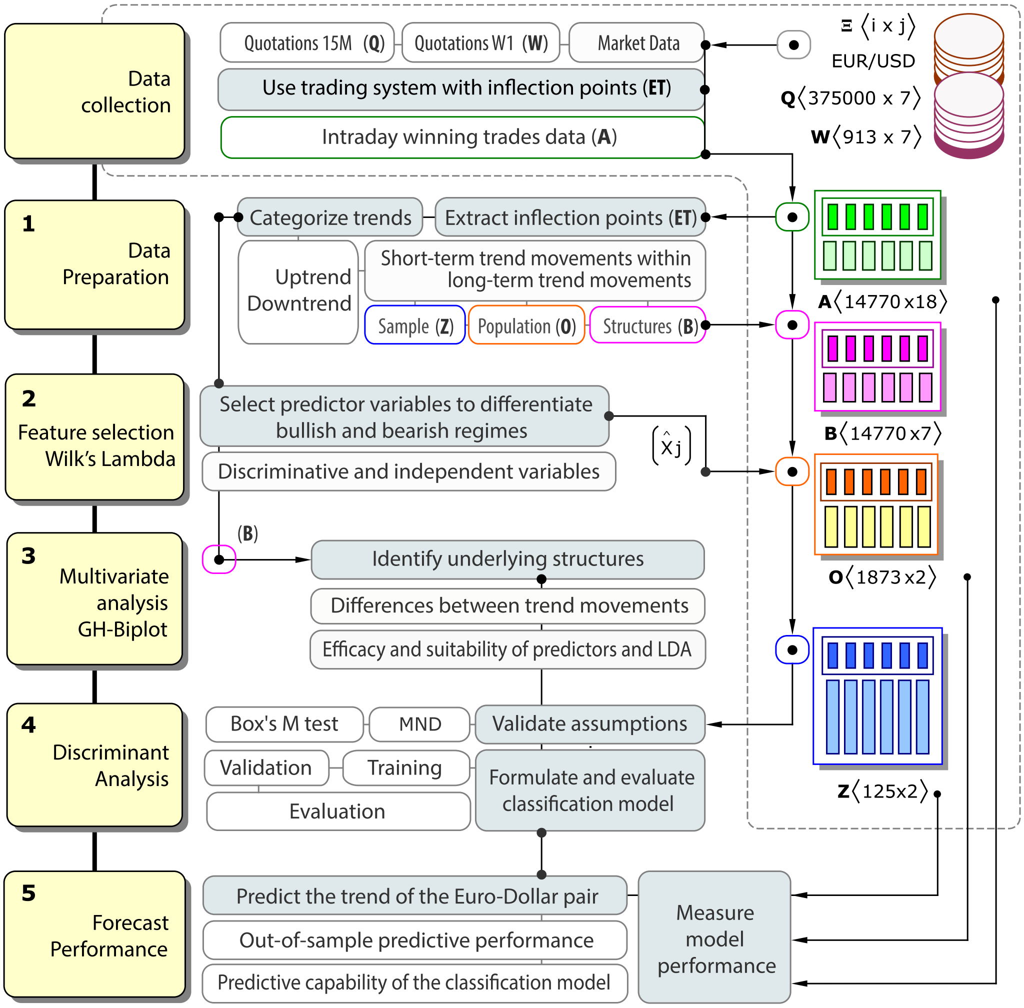


FIGURE 3. Proposed methodology to predict, in the short term, the onset of the future movement of the euro-dollar exchange rate. Each phase (in yellow) is divided into actions (in blue) and their corresponding results (in white). This methodology is summarized in data preparation procedures, selection of features with discriminant power, multidimensional analysis of differences between trend movements, discriminant analysis, and evaluation of the predictive power of the classification model.

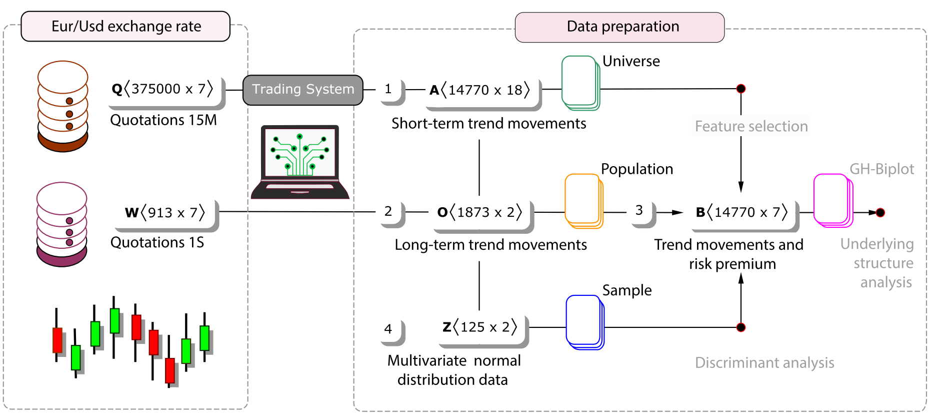


FIGURE 4. Data preparation for short-term prediction of the beginning of the direction of the future movement of the euro-dollar exchange rate. (1) Definition of micro-trends. (2) Definition of micro-trends as a function of long-term trend movements. (3) Preparation of data for underlying structure analysis. (4) Extraction of micro-trends with multivariate normal distribution.

moments **ET** in which the change of direction of the short-term trend occurs, according to the 15-minute trading session, are a function of the dominant market trend, according to the weekly trading session.

TABLE 4. Categorization of micro movements.

Gross profit or loss	* Trade slope	Trade type	Micro movement label
Winning trade	Slope $(\Delta p/\Delta t) > 0$	Long trade (Buy)	Upward Micro movement
Winning trade	Slope $(\Delta p/\Delta t) < 0$	Short trade (Sell)	Downward Micro movement

Winning trade = GPL_$ > 0. *Sp.XP-EP = Slope $(\Delta p/\Delta t)$. The formulas for the calculation of slopes, for long and short positions, are referenced in Table 2.

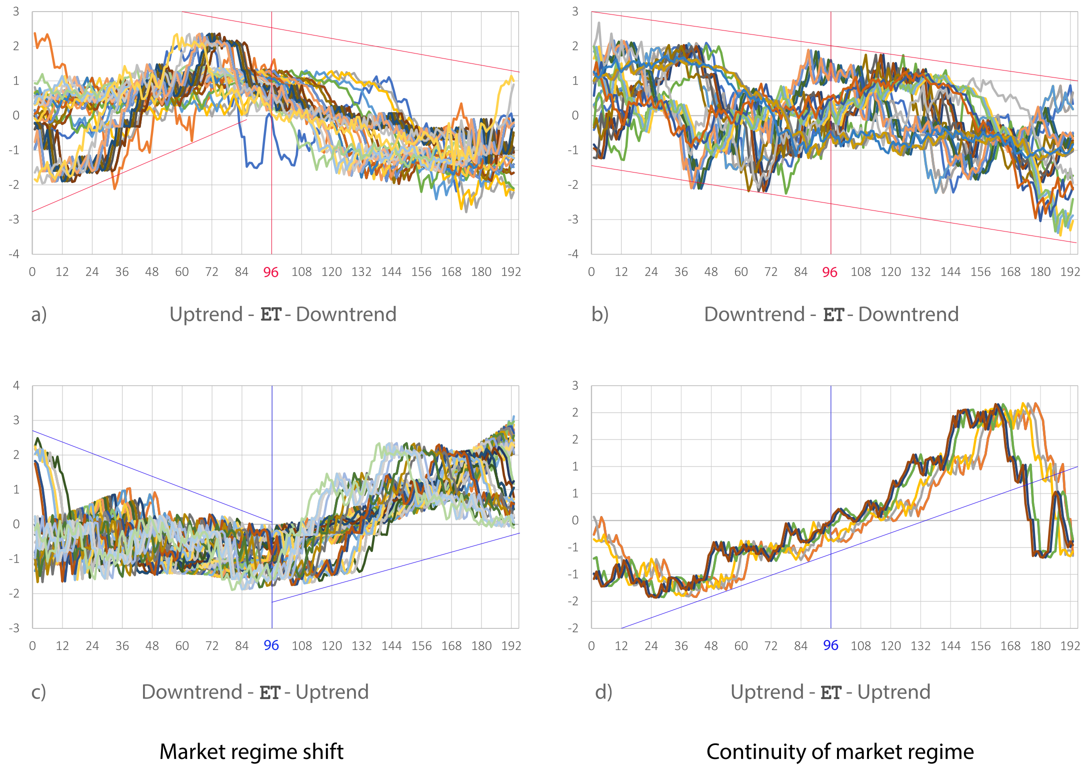


FIGURE 5. Market movements of the euro-dollar exchange rate based on weekly closing prices with the integration of intraday data at time ET. Row-normalized data. (a) Regime shift, upward to downward directional movement. (b) Regime continuity, downward directional movement. (c) Regime shift, downward to upward directional movement. (d) Regime continuity, upward directional movement.

For purposes of selection of the **ET** study units, Table 5 details in the first column the process of formation of the long-term trend movements. The study moments **ET** of micro-trends and long-term trend movements overlap. Thus, the moment of study **ET** is the reference point used to confirm, in weekly trading sessions, the continuity of the preceding trend (FT - Following the trend) or to identify the change in the direction of the preceding trend (CT - Shift trend). Thus, two trend formations are recognized, upward movements and downward movements, with two forms of origin, change of trend or continuity in the trend. The second and third columns define the assumptions used to identify the type of regime before and after the **ET** moment. The fourth column confirms the dominant slope of the market. Finally, the last column details the number of observations obtained for the study.

In summary, the study population **O** consists of 1873 micro-trends that are in favor of the direction of the dominant trend, but are not determinants of the initial long-term trend formation. Consequently, the study population consists of 414 and 228 bullish micro-trends, the former with trend reversal in favor of the dominant uptrend and the latter supporting the long-term uptrend. 556 and 675 bearish micro-trends, the former with change of direction of the preceding trend in favor of the dominant downtrend and the latter supporting the continuation of the long-term downward movement.

- *Trend movements and Risk-award.*

At this stage, the data collection **B** is prepared. According to Figure 4, the best observable variables are evaluated and selected for structure detection. The data are prepared in such a way that the analysis of the underlying structures reveals the multidimensional nature of the differences between trend movements. Consequently, the results of this study help to determine the relevance of the discriminant analysis and the appropriateness of the selected predictors. The data collection **B** $\langle 14770 \times 7 \rangle$ is made up of i analysis moments **ET** and

TABLE 5. Micro movements into long-term trend movements.

Regime Market Before - ET - After	**Market slope prior to ET** SRL $[cp_{i-96}, ..., cp_{i=\mathrm{ET}}]$	**Long-term trendline** SRL $[cp_{i=\mathrm{ET}}, ..., cp_{i=96}]$	**Market Slope** SRL $[cp_{i-96}, ..., cp_{i=\mathrm{ET}}, ..., cp_{i=96}]$	**O** 1873
Downtrend - CT - Uptrend	Slope RL $(y_i/x_i) > 0$	Slope RL $(y_i/x_i) > 0$	Slope RL $(y_i/x_i) > 0$	414
Uptrend - FT - Uptrend	Slope RL $(y_i/x_i) > 0$	Slope RL $(y_i/x_i) > 0$	Slope RL $(y_i/x_i) > 0$	228
Uptrend - CT - Downtrend	Slope RL $(y_i/x_i) > 0$	Slope RL $(y_i/x_i) < 0$	Slope RL $(y_i/x_i) < 0$	556
Downtrend - FT - Downtrend	Slope RL $(y_i/x_i) < 0$	Slope RL $(y_i/x_i) < 0$	Slope RL $(y_i/x_i) < 0$	675

ET = Turning point and Entry Time. CT = Trend shift, FT = Following trend. *M.Sp.P.ET = Market slope prior to **ET**. SRL = Slope RL = Slope of the time-series regression line. cp_i = Closing price of period i. *LTT = Long-term trendline. *M.Sp = Market Slope. Market Slope is the slope of the regression line of 96 closing prices before and after the **ET** turning point. The trend formed after the **ET** turning point, which is detected on a 15-minute time frame, is validated on a 1-week time frame with Market Slope. The equations for their computation are referenced in Table 2.

j variables measuring price action before and after the **ET** moment. There are two groups of noteworthy variables, the first group corresponding to the selected predictor variables and the second group consisting of the risk premium analysis (Sp.XP-EP, RP%M[1], GPL$). The latter group validates the direction of microtrends after earning positive gross profits following trading at **ET** inflection points. Variables that do not explain the differences between the onset of an upward and downward movement are excluded from the data structure.

- *Normally distributed multivariate data.*

Finally, in this phase, the study sample **Z** is prepared. These data are used to formulate, validate and evaluate the performance of the classification model. To overcome the asymmetric and leptokurtic effect of the study population and to ensure that the assumptions validating the application of discriminant analysis are met, a small sample size with multivariate normal distribution is used. To solve this problem, we draw from the study population $\mathbf{O}\langle 1873 \times 2\rangle$ a sample of observations that follows a multivariate normal distribution. Outliers are excluded from the study sample. This procedure ensures that the assumptions of multivariate normality and equality of variance and covariance matrices within groups are met for predictors and groups. Consequently, the data collection $\mathbf{Z}\langle 125 \times 2\rangle$ consists of 125 **ET** observations on which 2 chosen predictor variables $\hat{\mathbf{x}}_j$ are measured.

2) FEATURE SELECTION

The predictive power of the classification model is determined by the quality of the selected predictor variables and the adequacy of the selection process. Furthermore, the use of irrelevant variables, which do not contribute to the differentiation between classes, negatively affects the performance of the classification model. To address this problem, based on the multidimensional nature of the differences between bullish and bearish movements, the feature selection process is conceived under an exploratory-confirmatory statistical approach. This procedure helps to identify and confirm the effectiveness of the variables responsible for the separation of these clustering structures.

In this way, the proposed methodology helps to identify and select, from a multivariate data matrix $\mathbf{A}\langle i \times j\rangle$ formed by $\mathbf{i}$ trend movements on which p observable numerical variables $\mathbf{x}_j = (\mathrm{x}_1, \ldots, \mathrm{x}_p)$, the determinants $\hat{\mathbf{x}}_j$ of the movement of the euro-dollar exchange rate. The distinction between groups is made with a single categorical variable defined by the data vector $\dot{\mathbf{y}} = \{0, 1\}$, where 0 = "Upward Trend Movement" and 1 = "Downward Trend Movement". Thus, the vector $\dot{\mathbf{y}}$ is divided into two groups $\check{y}_g = \{\check{y}_1, \check{y}_2\}$. Consequently, the selection process yields the subset of eligible predictors $\hat{\mathbf{x}}_j$ which is formed by $\mathbf{p}$ independent variables $\hat{\mathbf{x}}_j = (\hat{\mathrm{x}}_1, \ldots, \hat{\mathrm{x}}_p)$, and is better than $\mathbf{x}_j$, if $(\hat{\mathbf{x}}_j \subset \mathbf{x_j})$ and $(\mathbf{x}_j > \hat{\mathbf{x}}_j)$, so that, $\mathbf{A}' \in \mathbb{R}^{i \times \hat{j}}$, with $\hat{j} < j$.

The function max $\left(\beta_j^\top L \beta_j\right) / \left(\beta_j^\top T \beta_j\right)$ evaluates the variables $\mathbf{x}_j$ using the data matrix **A**. So, it maximizes the between-group variability as a function of the total variability from a subset of $\mathbf{x}_j$ variables. The total covariance $\mathbf{T} = \mathbf{K} + \mathbf{L}$, is composed of the within-group and between-group covariances. β_j is the vector of coefficients of the discriminant function that evaluates the observable variables $\mathbf{x}_j$. Thus, the subset of eligible predictors $\hat{\mathbf{x}}_j$ is the one that best contributes to the differentiation between groups, has the highest discriminant power, and offers the best predictive performance. Additionally, the selection of the most plausible features is validated by performing statistical significance tests. These predictors are used in the construction of the classification model.

- *Variable selection process.*

It is composed of three stages, as shown in Figure 6. First, the ability to differentiate between groups is the potential that an eligible candidate variable has. This criterion is detected by one-way Anova analysis. The number of candidate variables is reduced with this initial screening. Second, the predictors that best contribute to the separation between groups are identified using the Wilks' Lambda statistic. The discriminant potential of each candidate variable is measured by this criterion. Levene's test of homogeneity of variances validates the test of equality of group means and reduces the number of eligible variables detected with the Wilks' Lambda statistic. Finally, the association between eligible variables is measured with Pearson's correlation coefficient. Predictors that may cause multicollinearity problems are discarded from the model formulation. Variables that are left out of the analysis, do not improve the separation of the groups, do not contribute information to the model and are therefore excluded from the subsequent procedures.

- *Discriminant potential of eligible variables.*

Referring to the process shown in Figure 6, the second step is based on the stepwise selection method. This helps to identify, statistically, explanatory variables that have a strong discriminative power and provide the best performance in the construction of classification models. This statistical procedure is an iterative method in which the study variables are analyzed individually until a subset of eligible variables $\hat{\mathbf{x}}_j$ is obtained. The Wilks' Lambda statistic and the value of the associated statistic, Snedecor's F, are used to determine the discriminant potential of each variable and of the subset of eligible variables $\hat{\mathbf{x}}_j$. Attributes with large Snedecor F-values contribute to the separation of group means and therefore discriminate better. In contrast, lower values do not discriminate due to attributes with closely spaced groups and widely scattered data. The procedure starts by calculating, for each analysis variable, the Snedecor's F-Values of inclusion F_ε and removal F_r given by equations (1) and (2), as follows:

$$F_\varepsilon = \frac{\left(\check{f} - \check{e}\right)(n-q-g)}{\check{e}\,(g-1)} \tag{1}$$

$$F_r = \frac{\left(\check{e} - \check{f}\right)(n-q-g+1)}{\check{f}\,(g-1)} \tag{2}$$

where $\check{e}$ is Within-groups Sums of Squares and Cross-product Matrix, $\check{f}$ is Total Sums of Squares and Cross-product Matrix, q corresponds to the group of variables incorporated in the analysis, g is the number of groups, n is the number of observations.

Variables with inclusion levels higher than the F-value to enter ($F_\varepsilon = 3.84$) enter the subset, of eligible variables $\hat{\mathbf{x}}_j$, and variables with removal values lower than the elimination F-value ($F_r = 2.71$) exit the analysis. The process is repeated until there are no candidate variables to remove. If the tolerance of the analysis variable falls below the specified tolerance (0.001), the variable is ineligible. The tolerance ψ_i can be written as $\psi_i = \left(1 - R^2\right)$. The tolerance is understood as the percentage of unexplained variance between the analyzed variable and the included variables. The value decreases drastically if the variables are correlated, while it will produce small variations when the variables are independent.

The modeling of the discriminant power of the linear function is performed as a function of some or all of the **p** eligible variables $\hat{\mathbf{x}}_j$. The Wilks' Lambda statistic is the benchmark used to assess the discriminant power of the variables used in the model. This criterion contrasts the dispersion of the instances within groups and the dispersion of the data without distinguishing between groups. Its computation is given in the equation (3). The closer the value is to zero, the greater the discriminant power of the variables analyzed, whereas if the value is close to one, the lower the discriminant power.

$$\Lambda = \frac{|\mathbf{K}|}{|\mathbf{T}|} \tag{3}$$

In matrix notation $\mathbf{T} = \mathbf{K} + \mathbf{L}$ is the total covariance, with $\mathbf{K}$ and $\mathbf{L}$ as the within-group and between-group covariances. **Lambda** is the Wilks' Lambda statistic. Finally, the use of statistical significance tests facilitates the identification of the subset of eligible variables $\hat{\mathbf{x}}_j$, which, because of their statistical validity and from a technical analysis point of view, help to explain the emergence of a new trend as a function of the culmination process of the previous movement. In addition to this connotation, the eligible predictors $\hat{\mathbf{x}}_j$ are more useful because they provide explanatory value to the results obtained and contribute to the formulation of more accurate prediction models.

3) MULTIVARIATE ANALYSIS

$\mathbf{GH-Biplot}$ is a multivariate analysis method useful for representing and inspecting big data matrices [77]. This analysis technique is an alternative approach to the original method introduced by Gabriel [78] that improves the quality of representation of observations (points) and variables (vectors) in the same reduced-dimensional reference system. In this work, the $\mathbf{GH-Biplot}$ analysis is used to detect structures and to describe the underlying relationships between the selected observable variables $\hat{\mathbf{x}}_j$ and the factors that explain them. Thus, the analysis allows us to suggest, according to the structure of the data and the discriminant capacity of the selected predictor variables, the appropriateness of formulating a discriminant model.

In Figure 7a, the multivariate data matrix $\mathbf{B}\langle i \times j\rangle$ after singular value decomposition (*SVD*) approximates the matrix $\mathbf{B} \cong \mathbf{UDV}^T$, where, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, and $\mathbf{D} = (\lambda_1, \ldots, \lambda_J)$ contains the singular values. Consequently, $\mathbf{G}$ is the matrix $\mathbf{U}$, and $\mathbf{H}$ the matrix of the first two columns of the product $\mathbf{VD}$. Thus, this technique with the appropriate selection of markers, $\eth_i = (\eth_i, \ldots, \eth_n)$ for the units of analysis and $\mathrm{h_j} = \left(h_j, \ldots, h_n\right)$ for the variables, allows to represent simultaneously in the same space observations (points) and variables (vectors).

The graphical representation and interpretation rules of a biplot analysis are shown in Figure 7b. Although the $\mathbf{GH-Biplot}$ analysis is specifically used for the detection and study of patterns and covariation relationships between variables, this analysis uses the following interpretation rules: the length of each vector represents the variability of the measurements; the angles formed between variables determine the level of covariation or correlation; and the angles formed between variables and factors are understood as the relationship or contribution of each variable to the factor. However, although this type of analysis does not focus on the study of the relationships between observations and variables, the distance between units of analysis is interpreted as a measure of similarity and the orthogonal projection of each observation on the body of the vector determines the order of the measurement in the original matrix.

4) DISCRIMINANT ANALYSIS

Discriminant analysis is a multivariate statistical technique introduced by Fisher [79]. It allows modeling and predicting,

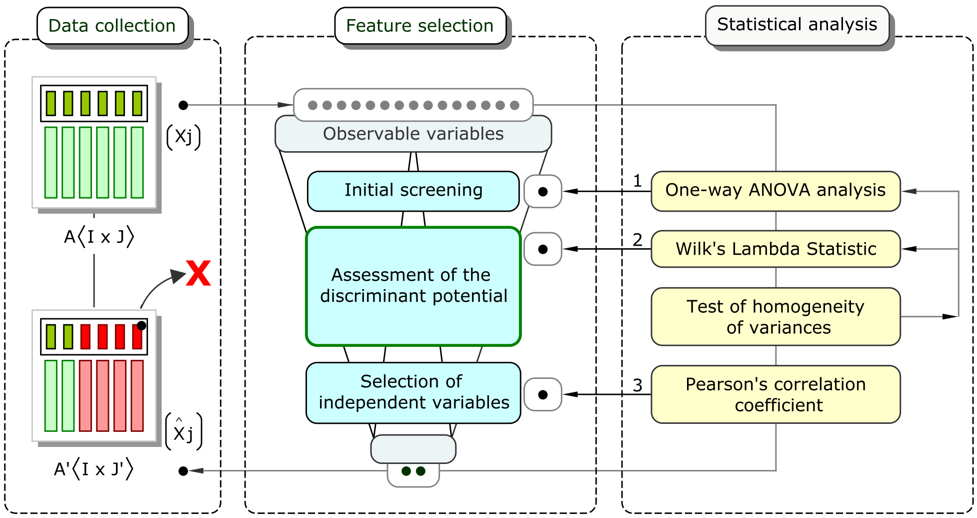
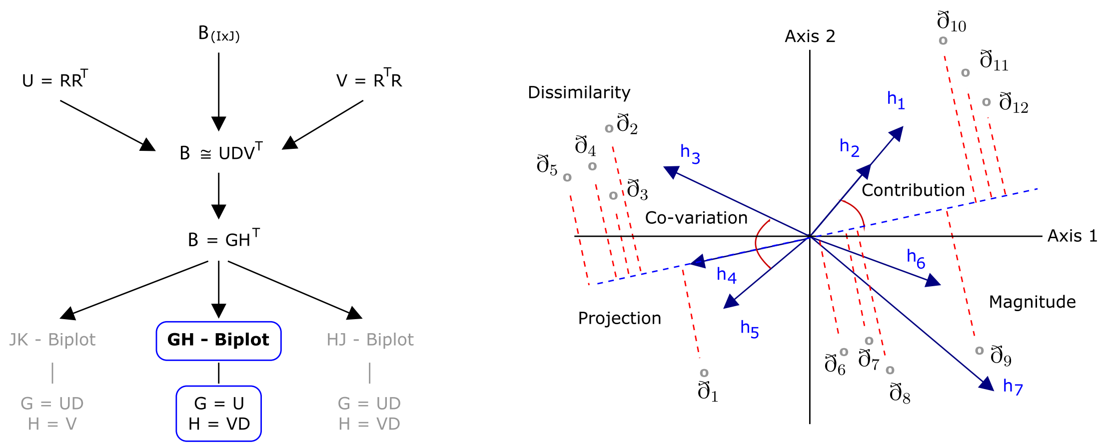


FIGURE 6. **Proposed procedure to identify and select the predictor variables with the greatest discriminant power. This procedure is summarized in three stages, initial screening of observable variables, evaluation of the discriminant potential, and selection of independent predictor variables.**



(a) Biplot typologies. (b) Biplot analysis.

FIGURE 7. **GH – Biplot analysis: A new tool to diagnose the discriminatory power of selected predictor variables and the suitability of classification models in the underlying structures of the variables. (a) Typologies of biplot analysis. (b) Graphical representation and rules of interpretation of a biplot analysis.**

on a categorical dependent variable, group membership as a function of a set of independent quantitative variables. Violation of assumptions and limitation in the fulfillment of requirements can distort results and interpretations. To overcome these difficulties, the linear discriminant function is constructed, validated and evaluated from the data matrix $\mathbf{Z}\langle i \times j\rangle$ formed by **i** study moments **ET** (at which the direction of the preceding trend changes) on which **p** chosen predictor variables $\hat{\mathbf{x}}_j = (\hat{x}_1, \ldots, \hat{x}_p)$ are measured. The membership group corresponds to the beginning of the future movement direction and is defined by the label vector $\dot{\mathbf{y}} = \{0, 1\}$, where 0 = Upward Trend Movement "$\mathrm{C_U}$" and 1 = Downward Trend Movement "$\mathrm{C_D}$". The data vector $\dot{\mathbf{y}}$ is divided into $\check{y}_g$ groups with g $= \{1, \ldots G\}$, where, $\check{y}_g$ denotes the g^{th} group of size n_g, such that $\check{y}_g = \{\check{y}_1, \check{y}_2\}$ is the a priori probability of membership of each case to the group $\check{y}_g$. Consequently, the linear function expressed in equation (4) is the one that best discriminates between bullish and bearish movements.

$$y(\hat{\mathbf{X}}) = \beta_j^{\top}\hat{\mathbf{X}}_j + \beta_0, \tag{4}$$

where $y(\hat{\mathbf{X}})$ is the discriminant score. β_j is the vector of discriminant weights or coefficients. $\hat{\mathbf{X}}_j$ are the selected independent variables and β_0 is a constant. The function (4) can be written as the equation (5):

$$y(\hat{\mathbf{x}}) = \beta_0 + \beta_1\hat{x}_1 + \beta_2\hat{x}_2. \tag{5}$$

The classification criteria available for case assignment are multiple, however, in this work the Fisher classification functions $y(\hat{x}_g)$ derived from the main equation (5) are employed.

For each group g a function is extracted, thus $y(\hat{x}_U)$ and $y(\hat{x}_D)$ discriminate between upward "C_U" and downward "C_D" movements respectively. The coefficients of these functions are obtained assuming bivariate normality for the subset of predictors $\hat{\mathbf{x}}_j$, maximum likelihood and equal a priori probabilities for $\check{y}_g$. The vector of coefficients of each ranking function is determined by the equation (6).

$$y(\hat{x_g}) = \bar{\hat{\mathbf{x}}}_{jg}^{\top} S^{-1} x - \frac{1}{2} \bar{\hat{\mathbf{x}}}_{jg}^{\top} S^{-1} \bar{\hat{\mathbf{x}}}_{jg} + \ln\left(\pi_g\right), \tag{6}$$

where $\bar{\hat{\mathbf{x}}}_{jg}^{\top}$ is the transposed vector of the means of the predictor variables $\hat{\mathbf{x}}_j$ of the group g. S is the combined intra-group covariance matrix, x is the name of the chosen independent variables, π_g is the a priori probability of belonging to the group g. The classification of each observation i (turning point **ET**) is made according to the obtained discriminant score $y(\hat{\mathbf{x}})$. The model assigns the case to the group that achieves the highest discriminant score. The rule that assigns the cases to the membership group is summarized below:

If $y(\hat{x}_U) \geq y(\hat{x}_D)$ then $y(\hat{x}_U)$ belongs in C_U.
Else if $y(\hat{x}_U) < y(\hat{x}_D)$ then $y(\hat{x}_D)$ belongs in C_D,

where $y(\hat{x}_U)$ is the bullish linear function, $y(\hat{x}_D)$ is the bearish linear function, C_U y C_D are the group identification labels according to bullish and bearish movements.

Goodness-of-fit measures indicate the model's ability to fit the data set. Analysis of these results can reveal discrepancies between observed and expected values of the model. The explained variance, the canonical correlation and the function test are the most commonly used goodness-of-fit measures. The total variance explained is due to the first eigenvalues that are associated with each of the extracted discriminant functions. Thus, the m discriminant functions $y_i = (y_1, \ldots, y_m)$, where $m = Min\left(g-1, \hat{x_j}\right)$ are linearly independent and have an associated eigenvalue λ indicating the proportion of total variance explained by that function. For a single discriminant function $m = 1$, there is only one nonzero eigenvalue λ_1. The ratio of variance explained by the function y_i to the total variance explained by the extractable functions y_m is determined by the expression $\varpi(y_i)$ of the equation (7):

$$\varpi(\mathrm{y}_i) = \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i}. \tag{7}$$

Note that in the denominator of the equation (7) is the sum of all the eigenvalues λ_i, corresponding to the total variance explained by all the discriminant functions y_m. In the numerator is the eigenvalue λ_i associated to the discriminant function of analysis y_i.

The canonical correlation is a measure of the effectiveness of the discriminant power of a function and provides valuable information when differentiation is made between two groups. Again, the effectiveness of the function is known from the eigenvalues extracted. The canonical correlation measures the percentage of the total variance that in the $i-th$ function y_i is being explained by the difference between groups. Consequently, the $i-th$ function y_i is more discriminant the closer the canonical correlation is to one. This value is obtained from (8):

$$C\left(\mathrm{y}_i\right) = \sqrt{\frac{\lambda_i}{1+\lambda_i}}, \tag{8}$$

where $C\left(y_i\right)$ is the canonical correlation of the discriminant function y_i, λ_i is the eigenvalue associated to the discriminant function y_i.

Finally, the function test evaluates in the constructed model the ability to differentiate between groups. The overall significance of the discriminant function is obtained using the Wilks' Lambda contrast statistic of equation (9):

$$\Lambda = \frac{1}{1+\lambda_i}, \tag{9}$$

where λ_i is the eigenvalue of the linear discriminant function. The linear function discriminates more between groups when the Wilks' Lambda statistic is closer to zero.

5) FORECAST PERFORMANCE

At this stage, the predictive power of the classification model is evaluated. After training the linear discriminant function, the model is used to classify the direction of the future movement of the euro-dollar exchange rate. By replacing the measures of the predictor variables $\hat{\mathbf{x}}_j$ in the discriminant function $y(\hat{\mathbf{x}})$, it predicts, in the short term, the direction of the movement of the euro-dollar exchange rate. The performance of the model is evaluated using four additional samples, OOS_1, OOS_2, OOS_3 and OOS_4, which capture information from different market conditions and time horizons over a 15-minute time frame. The performance indices used to measure the predictive power of the model are described in more detail in the section (IV-B). The use of performance measures and additional data allows an objective and comparative evaluation of the model with other approaches in terms of generalization, accuracy, and interpretability.

6) LCC ALGORITHM

Algorithm 1 is a pseudocode that gathers all the steps of the proposed LCC methodology.

IV. EXPERIMENTAL SETUP

Considering the nature of the data and their asymmetric and leptokurtic arrangement, the proposed methodology is designed to reduce the natural bias of the data and to give reliability to the interpretations. To this end, a subset of the data is extracted and prepared. The cases are independent. Predictors and groups assume a multivariate normal distribution. Intra-group variance-covariance matrices are equal and the membership of each case to a particular group is exclusive.

The identification of the determinants of the directional movement of the euro-dollar exchange rate due to the change of "market regime" is approached as a problem of selection of observable characteristics [80] and the distinction between

Algorithm 1 Linear Classifier Configurator (LCC) for Predicting the Directional Movement of the Euro-Dollar Exchange Rate

Input: Market Data (OHLCV): $\mathbf{Q}\,\langle 375000 \times 7\rangle$ 15-minutes, $\mathbf{W}\,\langle 913 \times 7\rangle$ 1-week, Trading system, IWT data: $\mathbf{A}\,\langle 14770 \times 18\rangle$.

1: Data preparation:
- Extract **ET** inflection points in $\mathbf{A}\,\langle 14770 \times 18\rangle$ from $\mathbf{Q}\,\langle 375000 \times 7\rangle$ using Trading system.
- Measure **ET** inflection points using Table 2 variables.
- Assign labels to trends using Table 4.
- Extract population data $\mathbf{O}\,\langle 1873 \times 2\rangle$ from $\mathbf{A}\,\langle 14770 \times 18\rangle$ based on weekly market data $\mathbf{W}\,\langle 913 \times 7\rangle$ using Table 5.
- Extract the normally distributed sample $\mathbf{Z}\,\langle 125 \times 2\rangle$ from the population $\mathbf{O}\,\langle 1873 \times 2\rangle$.
- Extract selected subset features $\mathbf{B}\,\langle 14770 \times 7\rangle$ with underlying structures from $\mathbf{A}\,\langle 14770 \times 18\rangle$.

2: Feature selection:
- Use candidate variables from $\mathbf{A}\,\langle 14770 \times 18\rangle$ with specified parameters from Table 6.
- Conduct one-way ANOVA analysis to assess group differences among candidate variables.
- Evaluate candidate variables' discriminant potential via stepwise variable selection (Equations 1, 2, and 3).
- Validate the equality of means among groups using the Levene's test.
- Measure the linear relationship between variables with Pearson's correlation coefficient.
- Extract the subset $\hat{\mathbf{X}}$ of discriminative and independent variables $\mathbf{A}'\,\langle 14770 \times 2\rangle$.

3: Multivariate analysis:
- Use the subset of features $\mathbf{B}\,\langle 14770 \times 7\rangle$ for structural analysis with GH-Biplot (Section III-B3)
- Assess data adequacy for structure detection with the KMO measure and Barlett's test of sphericity.
- Validate:
  - Individual differences in trend movements from a multivariate perspective.
  - The discriminative efficacy and independence of the selected predictors.
  - The effectiveness of discriminant analysis.

4: Discriminant analysis:
- Use the data matrix $\mathbf{Z}\,\langle 125 \times 2\rangle$ and the parameters in Table 6.
- Analyze descriptive statistics of $\mathbf{Z}\,\langle 125 \times 2\rangle$ as described in Table 13.
- Verify the fulfillment of statistical assumptions (Section V-C2).
- Formulate, validate and evaluate the classification model ($y(\hat{\mathbf{X}}) = \beta_0 + \beta_1\hat{X}_1 + \beta_2\hat{X}_2$) following Equations (4), (5) and (6).
- Assess the model fit to the data structure using goodness-of-fit tests based on Equations (7), (8), and (9).

Output: Linear discriminant model $y(\hat{\mathbf{X}}) = \beta_0 + \beta_1\hat{X}_1 + \beta_2\hat{X}_2$.

groups is made with a single categorical variable consisting of two classes (uptrend and downtrend).

The prediction of the direction of the exchange rate is approached as a supervised learning classification problem. The formulation of the classification model is done based on the sampled data and the selected variables. The performance evaluation of the classification model is done with four additional data samples. Therefore, to validate the effectiveness of the results obtained in all experiments the proposed methodology uses an interpretive approach based on the statistical power of parametric tests and on the reliability of the results.

During the experimental phase, six key technical considerations significantly impact the quality of results achieved with the proposed methodology: (i) Ensure data quality by maintaining accuracy and consistency, particularly in critical activities during the preparation phase. (ii) Use discriminative variables that effectively differentiate between observations. (iii) The categories of the dependent variable must be relevant in number, independent, and significantly discriminative. (iv) The effectiveness of the selection process of predictor variables must be observable through GH-Biplot structure analysis. (v) The integration of these partial solutions in a classification model must be reflected with high-performance values. Finally, (vi) statistical significance tests must support the generalization and validity of interpretations and conclusions.

The open source, machine learning PyCaret library was used in Python (3.9.5) notebook environment for data preparation and model implementation. IBM SPSS Statistics for Windows, Version 27.0. (IBM Corp. Released, 2019) [81] and RStudio Team (2021) software were used for statistical analysis. MultBiplot software [82] was used for GH-Biplot analysis and multivariate data visualization. The MVN [83] package in R code was used to evaluate multivariate and univariate normality assumptions and provide the graphical evidence. The experiments were performed on an Intel(R) Xeon(R) Silver 4110 CPU @ 2.10Ghz (2 Processors) with 32 Gb in 64-Bit RAM.

A. TRAINING, VALIDATION AND TESTING

The analysis sample consists of 125 observations divided into two subsets, one for training and one for evaluation. The size of these subsets should be large and representative enough for the model to accurately learn, generalize, and assess its performance. However, there is no standard rule defining their specific size. For the analysis, the cases are randomly selected, approximately 70% of the upward and downward movements are used to create the model (training set) and the same cases are used to cross-validate the model. The remaining trend movements, approximately 30%, are used to validate the performance of the model. Four additional data samples are used to evaluate the predictive power of the model. Set OOS_3 consists of 1748 confirmed trend movements on a long-term weekly time frame and sets OOS_1, OOS_2 and OOS_4 consist of 6581, 14645 and 2336 trend movements measured on a 15-minute time frame. The table, 6, provides the parameter values used to conduct all experiments, including technical specifications of the criteria used in the feature selection process, the discriminant analysis and the analysis sample sizes. In the feature selection process, the default values assigned for Tolerance (ψ_i), input levels ($F_\varepsilon = 3.84$) and removal ($F_r = 2.71$), in the attribute evaluation, are the ones that produce the best performance.

B. PERFORMANCE MEASUREMENTS

The performance of the classification model is measured based on the results given in the table 7. The confusion matrix consists of four categories. In the main diagonal are placed the observations correctly classified by the model. There are the True Positives T_P which is the number of observations correctly classified as bullish movements and the True Negatives T_N which is the number of cases correctly classified as bearish movements. On the opposite diagonal are the cases incorrectly classified by the model. The False Positive F_P is the number of bearish moves incorrectly classified as bullish and the False Negative F_N is the number of bullish moves that are misclassified as bearish.

Although there is no general consensus on the effective use of a particular performance measure, *Accuracy* is one of the most commonly used metrics to evaluate the performance of models in classifying new observations. The equation (10) measures the proportion of observations that the model classifies correctly:

$$Accuracy = \frac{T_P + T_N}{F_P + T_P + T_N + F_N} \tag{10}$$

The geometric mean $G - mean$ is a performance measure that is used when groups are unbalanced [84]. Equation (13) measures the geometric mean of the positive hit ratio (Recall or Sensitivity) and the negative hit ratio (Specificity). *Recall* and *Specificity* are measured from equations (11) and (12) respectively.

$$Recall = \frac{T_P}{T_P + F_N} \tag{11}$$

$$Specificity = \frac{T_N}{F_P + T_N} \tag{12}$$

$$G - mean \quad Accuracy = \sqrt{Recall \cdot Specificity} \tag{13}$$

The positive predictive value, also called *Precision*, is the ratio computed between the number of cases classified as true positives and the total number of positive cases. The equation (14) defines its calculation.

$$Precision = \frac{T_P}{F_P + T_P} \tag{14}$$

$F_1 - Score$ is the harmonic mean of *Recall* and *Precision*. This measure is useful for comparing prediction quality between models and can be interpreted as a function of counts of true positives, false negatives, and false positives [85]. Equation (15) shows how it is calculated.

$$F_1 = \frac{2 \cdot Recall \cdot Precisin}{Recall + Precision} \tag{15}$$

The *Kappa* statistic is calculated from the equation (16). It is useful in model building and is used to find out whether the formulated model is better, equal or worse than a random model [86]. Note that the *Kappa* statistic is computed as a function of a prime rate called *Baseline*. This rate defines the baseline or benchmark for model comparison and is computed with the expression (17):

$$Kappa = \frac{Accuracy - Baseline}{1 - Baseline}, \tag{16}$$

where

$$Baseline = \frac{(PP \cdot AP)}{(PP + PN)^2} + \frac{(PN \cdot AN)}{(AP + AN)^2}. \tag{17}$$

Matthews correlation coefficient MCC is a robust rate statistic that provides a more reliable measure of the performance of the classification model. It produces higher scores only when the predictions are correct in all four categories of the confusion matrix [87]. It is calculated as shown in equation (18):

$$MCC = \frac{T_P \cdot T_N - F_P \cdot F_N}{\sqrt{(T_P + F_P) \cdot (T_P + F_N) \cdot (T_N + F_P) \cdot (T_N + F_N)}}. \tag{18}$$

V. RESULTS AND DISCUSSION

This section presents the results of the methodology outlined in Figure 3. The results are discussed in four distinct areas, highlighting key findings and insights. These areas include: A. Selection of attributes with high discriminant power that best explain the price action; B. Multidimensional analysis of differences between trend movements; C. Construction, validation, and evaluation of a classification model for predicting the short-term trend of the euro-dollar pair; and D. Evaluation of the model's performance using out-of-sample data and comparison with other forecasting approaches.

A. FEATURE SELECTION

According to the table 2, of the first 15 metrics that capture information prior to the market regime change, at the **ET** instant, the variables that are relevant to the study are not known a priori. Not all of them are useful and significant in the process of discriminating the direction of the future movement of the euro-dollar exchange rate. The importance of the variables lies not only in the statistical and individual significance that each one contributes to the differentiation between groups, but also in their explanatory capacity for the interpretation of the study problem. The identification and selection of the predictor variables that best contribute

TABLE 6. Parameter values.

Parameters	Criteria	Description
Feature selection		
F-Snedecor to enter	F_ε = 3.84	Threshold for input variable.
F-Snedecor to remove	F_ε = 2.71	Threshold for variable removal.
Minimum tolerance	ψ = 0.001	Minimum value to enter in the analysis.
Discriminant analysis		
Probabilities of equal groups	$\pi_g = 0.50$	Definition of function coefficients.
Covariance matrix	Within-groups equal	Classification of cases.
Cross-validation	Leave-one-out classification	Classification for cross validation.
ET Inflection point sets for analysis		
Training set:	(70%) 87/125 observations	Selected cases to create the model.
Cross-validation set:	(70%) 87/125 observations	Cross-validated cases.
Testing set:	(30%) 38/125 observations	Cases selected to evaluate the model.
Out-of-sample sets:		
OOS_1: 1999-2005	6581 observations	Price movements in bear and bull markets.
OOS_2: 2006-2020	14645 observations	Price movements in a bearish market channel.
OOS_3: 2006-2020	1748 observations	Price movements in a bearish market channel.
OOS_4: 2021-2023	2336 observations	Price movements in bear and bull markets.

TABLE 7. Confusion matrix.

Predictions / Actual	Actual Positives	Actual Negatives	Total
Predicted Positive PP	True Positives T_P	False Positives F_P	PP = $T_P + F_P$
Predicted Negative PN	False Negatives F_N	True Negatives T_N	PN = $F_N + T_N$
Total	AP = $T_P + F_N$	AN = $F_P + T_N$	AP+AN = PP+PN

to the differentiation between upward and downward trend movements is done in three steps, using the multivariate data matrix **A**.

1) EVALUATION OF THE CONTRIBUTION OF EACH PREDICTOR

The test for equality of group means, from an exploratory point of view, helps to identify the potential that each predictor variable has in group differentiation before participating in the formulation of the model. The evaluation of the contribution that each variable makes in the differentiation of groups allows to recognize the predictor variables that should be considered in the study. The test of equality of group means referred to in the table 8 contrasts the H_0 on equality of means between groups. According to one-way ANOVA analysis and with a significance value higher than 5%, the H_0 on equality of means between groups is accepted, so that the variables highlighted in gray do not have the potential to discriminate between upward and downward trend movements and, therefore, should not be considered in the study.

2) EVALUATION OF THE DISCRIMINANT POWER OF EACH PREDICTOR

The search for a simple, parsimonious and easy to explain prediction model leads to the identification and selection of predictors that best explain the price action and, consequently, best contribute to the separation between groups. The stepwise variable selection method is useful in achieving this purpose and thus ensures the identification and arrangement of the best candidate variables to be used in the model formulation.

Table 9 summarizes the statistics for the subset of significant candidate variables that could be incorporated into a single model and, taken together, are the predictor variables that best discriminate between upward and downward trend movement.

At each step, the predictor with the highest F-value for entry that exceeds the entry criterion (default, 3.84) is added to the model. Thus, the small values of the Wilks' Lambda statistic obtained at each step indicate that the subset of predictors chosen are the ones that best contribute to the separation between groups. In the last step the variables that are left out of the analysis all have F-values for entry smaller than 3.84, so no more are added to the subset.

According to the results of the table 9, s.V.18P does not minimize the value of the Wilks' Lambda statistic obtained in the previous step (7), consequently this variable (8) does not contribute information to the model and therefore is excluded from further analysis.

The potential that each predictor has in class differentiation and the ability to achieve a high level of classification performance are measured through the Wilk's Lambda statistic. Figure 8 shows in order of importance the discriminant power of the best predictor variables. x.V*18P, Sp.Vr*CP.18P and Sp.CP.18P are the variables that best explain the price action prior to the market regime change. The second, third and

TABLE 8. One-Way ANOVA.

Measure	Sum of squares	Df_1	Df_2	Mean square	F-Statistic	P-Value
x.V*18P	0.00	1	14768	0.00	16856.79	0.000
Sp.Vr*CP.18P	0.00	1	14768	0.00	5001.01	0.000
s.Vr.Vo.18P	1.80	1	14768	1.80	1.17	**0.279**
x.Vr.Vo.18P	0.07	1	14768	0.07	0.67	**0.412**
Sp.Vr.Vo.18P	0.00	1	14768	0.00	0.05	**0.830**
Sp.Vo.18P	24.89	1	14768	24.89	0.01	**0.905**
s.V.18P	0.00	1	14768	0.00	5.50	0.019
Sp.CP.18P	0.00	1	14768	0.00	11645.00	0.000
Sp.Vr.HL.18P	0.00	1	14768	0.00	0.16	**0.688**
Sp.Vr.CO.18P	0.04	1	14768	0.04	0.19	**0.667**
s.Vr.RSI.18P	10.41	1	14768	10.41	1962.97	0.000
Sp.Vr.RSI.18P	0.09	1	14768	0.09	1500.26	0.000
Dv.CP-RSI.18P	0.06	1	14768	0.06	1.38	**0.241**
Dv.Vo-CP.18P	872.86	1	14768	872.86	4266.50	0.000
Dv.Vo-RSI.18P	1129.54	1	14768	1129.54	6808.76	0.000

TABLE 9. Discriminant power of predictor variables.

Step	Selected Features	Wilks' Lambda	Df_1	Df_2	Df_3	F-Value	Df_1	Df_2	P-Value
1	x.V*18P	0.467	1	1	14768	16856.787	1	14768	0.000
2	Sp.Vr.RSI.18P	0.365	2	1	14768	12851.92	2	14767	0.000
3	Dv.Vo-RSI.18P	0.337	3	1	14768	9666.854	3	14766	0.000
4	s.Vr.RSI.18P	0.327	4	1	14768	7590.954	4	14765	0.000
5	Sp.Vr*CP.18P	0.325	5	1	14768	6137.861	5	14764	0.000
6	Sp.CP.18P	0.324	6	1	14768	5142.587	6	14763	0.000
7	**Dv.Vo-CP.18P**	**0.323**	7	1	14768	4410.364	7	14762	0.000
8	**s.V.18P**	**0.323**	8	1	14768	3861.226	8	14761	0.000

fourth variables measure market behavior as a function of the technical indicator $RSI(9P)$. Finally, the variable Dv.Vo-CP.18P measures the divergence between the closing price and the volume in the market.

Levene's test for homogeneity of variances allows us to establish whether the study groups come from populations with equal variances. The reliability of the test of equality of means, one-way ANOVA, is validated by compliance with the test of homogeneity of variances. Consequently, by verifying in the table 10 the fulfillment of the assumption of homogeneity of variances, with Levene's statistic, the rejection of the hypothesis Ho that the means of the groups of predictors 1, 2 and 6 are equal is reaffirmed. Therefore, now that it is known for the predictors highlighted in gray that the group means do indeed differ, it is possible through multivariate analysis to approach the knowledge of the differences that changes in the market regime generate in the price action.

3) INDEPENDENCE ANALYSIS BETWEEN PREDICTOR VARIABLES

According to the results presented in the table 11, there is a high correlation between the variables Sp.CP.18P and x.V*18P. The 95.6% correlation is significant at the 1% bilateral level. If all variables are used in the model formulation. The instability in the signs of the coefficients of the Fisher linear discriminant functions could negatively affect the classification results. Therefore, to avoid the multicollinearity problem, the variable Sp.CP.18P is excluded from the model formulation.

In summary, according to the proposed methodology, out of 15 candidate variables, the predictor and independent variables selected according to the exploratory confirmatory approach are two. Average yield of closing prices x.V*18P and Slope of the regression line of the variations between closing prices Sp.Vr*CP.18P, both calculated on the basis of the last 18 periods before the change of market regime at the **ET** time. According to statistical merit, the chosen subset of variables $\hat{\mathbf{x}}_j$ are the ones that best contribute to group separation and have the highest discriminant power and, from a technical analysis point of view, are the ones that best explain the price action before a market regime shift occurs at the **ET** instant.

B. MULTIVARIATE ANALYSIS

The purpose of the $\mathbf{GH-Biplot}$ analysis is to detect and describe, in the underlying structures of the data, the multidimensional nature of the differences between the onset of an upward and downward directional movement as a function of the chosen predictor variables $\hat{\mathbf{x}}_j$. The results of the analysis allow us to confirm the suitability of the selected predictor variables and the relevance of the discriminant analysis in the classification tasks.

The analysis is performed on the data collection $\mathbf{B}\langle 14770 \times 7\rangle$ consisting of 14770 trend movements on which 7 numerical observable variables and a categorical

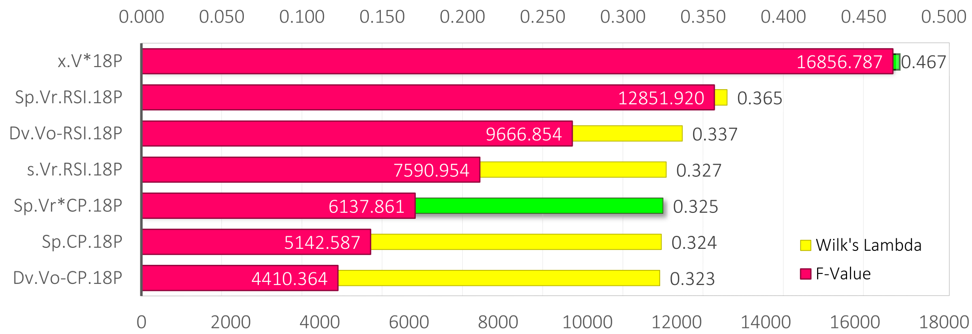


FIGURE 8. **Discriminant ability of the predictor variables. Order of importance of predictors according to discriminant capacity between groups. The Wilks' Lambda statistic measures the discriminant potential of each predictor to differentiate between bullish and bearish movements. Predictors with F-value greater than the entry criterion (default, 3.84) are suitable for differentiating between class averages.**

TABLE 10. **Variance Homogeneity Test.**

ID	Measure	Levene Statistic	Df_1	Df_2	P-Value
1	x.V*18P	0.578	1	14768	**0.446**
2	Sp.Vr*CP.18P	1.820	1	14768	**0.177**
3	Sp.Vr.RSI.18P	439.873	1	14768	0.000
4	Dv.Vo-RSI.18P	16.092	1	14768	0.000
5	s.Vr.RSI.18P	188.975	1	14768	0.000
6	Sp.CP.18P	3.165	1	14768	**0.075**
7	Dv.Vo-CP.18P	9.311	1	14768	0.002

TABLE 11. **Independence test of predictor variables.**

Measure	Sp.Vr*CP.18P	x.V*18P	Sp.CP.18P
Sp.Vr*CP.18P	1	0.362	0.312
x.V*18P		1	**0.956**
Sp.CP.18P			1

variable labeling the direction of the trend of the euro-dollar exchange rate are measured in two groups $\check{y}_g$. The groups, according to the label vector $\dot{\mathbf{y}} = \{0, 1\}$, are identified with the categories 0 = "Upward Trend Movement" and 1 = "Downward Trend Movement".

The effectiveness of the **GH − Biplot** analysis in detecting underlying structures is determined by the adequacy of the data. The relevance of the multivariate data matrix $\mathbf{B}\langle 14770 \times 7\rangle$ for structure detection is verified by the Kaiser-Meyer-Olkin KMO sampling adequacy measures and Bartlett's test of sphericity. The results of these statistical tests suggest that the analysis is appropriate. The KMO value of 0.639 close to 1 indicates that the proportion of variation in the variables used in the analysis are caused by underlying factors. The Chi-square value $\tilde{\chi}^2 = 47208.373$ associated with the KMO test statistic and a P-Value of less than 5% suggests rejection of the hypothesis H_0 that the correlation matrix is an identity matrix ($\tilde{\chi}^2 = 47208.373$, Df = 21, P-Value = 0.000). Hence, covariation between variables is suitable for the detection of underlying structures.

The manifest variables that best explain the factor axes are summarized in the table 12. The relative factor contributions to the column items can be interpreted as a measure of goodness-of-fit with which the observed variables best explain the factor axes. The reliability of the analysis and of the suggested classification model is verified by the accuracy achieved in the classification process.

The table 12 details, by groups, the variables involved in the analysis. The first group measures the behavior of the closing prices preceding the trend change at the time **ET**. This group of variables 1, 2, 4 and 5 are the ones that best discriminate between upward and downward movement directions. The proposed solution, by the feature selection method, is formed by variables 1 and 4 highlighted in gray color. The second group is formed by variables 3, 6 and 7 that financially validate the expected trend direction with the realization of profits after the **ET** moment. Other parameters available in the table 2 do not participate in the analysis because they do not help to explain the multidimensional nature of the differences between the beginning of a bullish and bearish movement.

Figure 9 shows the **GH − Biplot** representation of the **B** data collection. The interpretation of the proposed solution is made on the first two factors that explain 55.012% of the accumulated variability of the analyzed variables. This

TABLE 12. Relative contributions of the factor to the column elements.

ID	Variables	Group 1	Group 2	Axis 1	Axis 2
1	x.V*18P	1		**875**	2
2	Sp.CP.18P	1		**831**	5
3	Sp.XP-EP		2	411	27
4	Sp.Vr*CP.18P	1		301	75
5	s.Vr.RSI.18P	1		221	119
6	RP%M[1]		2	0	**461**
7	GPL$		2	0	**523**

suggests two latent influences associated with the change in direction of the preceding trends, with a significant unexplained margin of variation. Therefore, the proportion of variance absorbed is sufficient to explain the underlying relationship between the historical behavior of the exchange rate and gross realizable profit, when the direction of the future trend is known a priori.

The column markers or vectors represent the variable measurements (technical parameters before and after the onset of each microtrend after the **ET** moment). The row markers represented by dots define, in terms of values, the time prior to the trend change **ET**. The scale of measurement of the vectors defines the magnitude of the variance of the measurements. The angles formed between vectors represent the correlations between variables. The center of the GH-Biplot is the midpoint of the differences between upward and downward movements.

Consequently, two latent factors are identified and are associated with the multivariate nature of the differences between groups. The first factor axis (Axis 1) describes the historical behavior of the exchange rate preceding the trend change. This axis is highly correlated with the variables: average return x.V*18P and closing price trend Sp.CP.18P. In the first factor, there is greater variability in the left-hand side measurements than in the origin of the GH-Biplot. These negative variations, between closing prices, depreciate the exchange rate. Thus, the sustained accumulation of negative average returns is an indicator that precedes the beginning of a new upward movement. This effect experienced by the depreciation of the euro against the US dollar is highlighted during the period 2006 - 2020.

Towards the right side of the first factor, the variability of the measurements is lower. The sustained appreciation of the exchange rate is an effect of positive variations between closing prices and is an indicator that precedes the beginning of a new downward movement. Consequently, the first factorial axis is the one that best discriminates between group means. According to the covariation angles formed between the factor and the explaining variables, the mean return x.V*18P is a better predictor. Note that the high correlation with the closing price trend Sp.CP.18P may generate multicollinearity problems, in case it is also decided to include this variable in the construction of the model.

The second factorial axis (Axis 2) describes the risk premium. This axis is explained by the variables: Market Risk RP%M[1] and Gross Profit or Loss GPL$. Both vectors score in the opposite direction. Theoretically, gross profits subject to risk are achieved, with the opening and closing of trades, depending on the start and end of each micro-trend. Thus, on the second axis, the variability of the profit obtained as a function of risk exposure is greater above the midpoint of the biplot than below.

The plane variables: slope of the variations between closing prices Sp.Vr*CP.18P measured before the moment **ET** and slope of the micro-trend Sp.XP-EP measured after the event **ET**, between both vectors, register an inverse covariation. Thus, negative slopes of the variations between closing prices precede the formation of upward micro-trends (Points in quadrants II and III) and positive slopes of the variations between closing prices precede the formation of downward micro-trends (Points in quadrants I and IV). The independence between Sp.Vr*CP.18P ands.Vr.RSI.18P is confirmed by the 90° angle they form between them, note that this feature is not sufficient to discriminate between bullish and bearish movements.

Overall, taken together, these results suggest the formation of four quadrants that explain the multivariate nature of the differences between bullish and bearish moves. Quadrants II and III can be interpreted as the oversold zone, where each row marker represents, with respect to the first factor axis, the lowest depreciation experienced by the exchange rate prior to the trend change. Quadrants I and IV can be understood as the overbought zone, where the observations represent, with respect to the first factorial axis, the highest appreciation reached prior to the change in trend. Quadrants I and II can be interpreted as high risk exposure zone, where each row marker corresponds, with respect to the second factorial axis, to Micro trends generating positive gross profits with exposure to floating losses before reaching the target closing price. Quadrants III and IV can be understood as low risk exposure zone, where each row marker corresponds, with respect to the second factorial axis, to positive gross profit generating micro trends with exposure to floating gains before reaching the target closing price.

The evidence from this study suggests that the first factor axis best explains the differences between the means of the

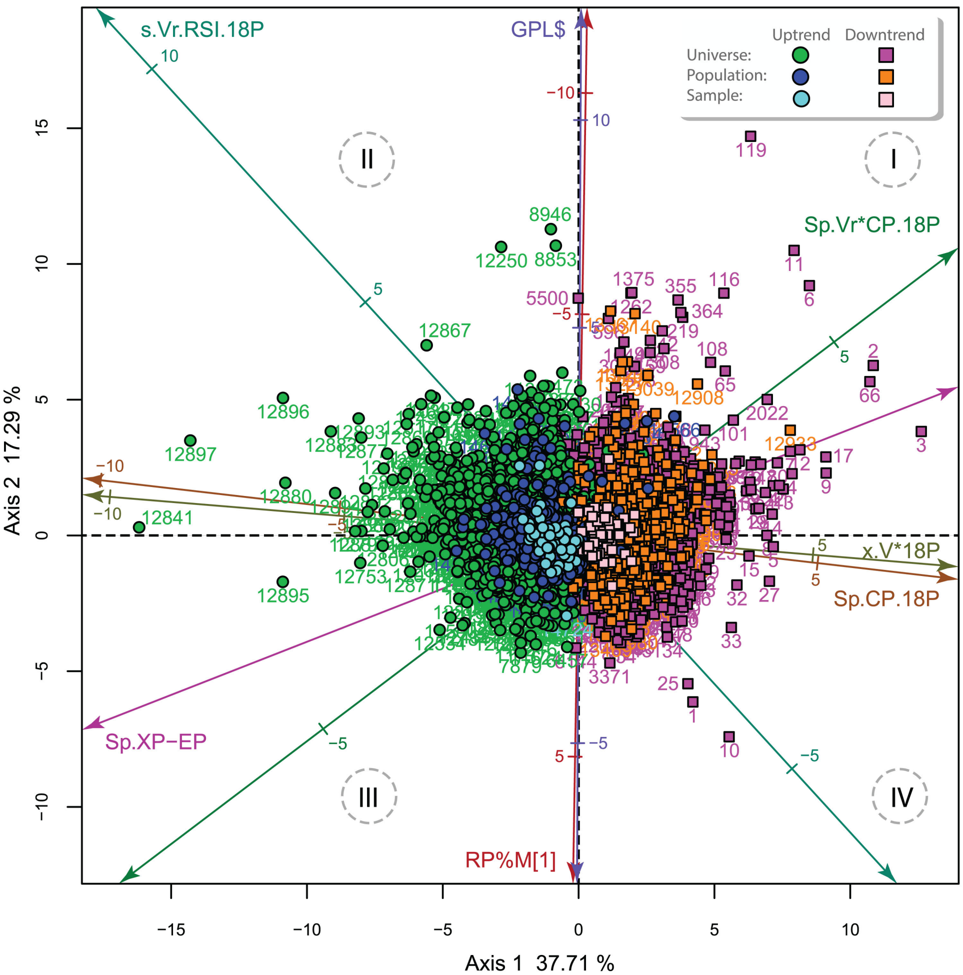


FIGURE 9. GH – Biplot of euro-dollar exchange rate movement determinants.Normalized data is standardized by column and partitioned according to *SVD*. Factorial plane 1-2. Explained variance of 55.012% (37.71% of axis 1 and 17.29% of axis 2). Detection of underlying relationships between variables (vectors) and identification of predictors that best discriminate between bullish and bearish micro-trends (row markers in quadrants II-III and quadrants I-IV, respectively).

upward and downward movements. Given the pattern of correlation with the observable variables, the predictor variable x.V*18P is the best contributor in differentiating between the onset of a new upward and downward trend movement. This predictor is the most influential to consider in building a classification model. Therefore, a discriminant analysis model is sufficient to distinguish, classify and predict the beginning of a trend movement.

C. LINEAR DISCRIMINANT ANALYSIS MODEL

To explore the advantages of the solutions provided by the proposed methodology, in terms of data, predictors and context, this paper evaluates the performance of the classification model based on three established measures: (**i**) sample with multivariate normality to overcome the asymmetric and leptokurtic nature of the data; (**ii**) predictor variables chosen according to technical and statistical merit; (**iii**) adequacy of the discriminant analysis with respect to the multidimensional nature of the differences between bullish and bearish movements. To this end, this section presents the results in five specific areas. 1) Descriptive statistics for the study subset (**Z** data matrix). 2) Verification of assumptions and requirements of the discriminant analysis. 3) Evaluation of model fit. 4) Classification of the direction of euro-dollar exchange rate movements and 5) Validation of the model.

Discriminant analysis is performed on the data set $\mathbf{Z}\langle 125 \times 2\rangle$ using two independent variables $\hat{\mathbf{x}}_j$ that collect information about the price action. These predictors are used to classify the observations **i** into two groups $\check{y}_g$. The $\hat{\mathbf{x}}_j$ predictor variables are: Average closing price returns x.V*18P and slope of the regression line of the interclosing price changes Sp.Vr*CP.18P, both predictors are calculated 18 periods prior to the **ET** study time, at which the change in direction of the preceding trend occurs. The clusters, according to the vector of labels $\dot{\mathbf{y}} = \{0, 1\}$, are trend movements described by the euro-dollar exchange rate, where 0 = "Upward Trend Movement" and 1 = "Downward Trend Movement". These class labels identify the beginning of a new short-term directional movement (measured on a 15-minute time frame) confirmed on a long-term (1-week) time frame.

1) DESCRIPTIVE STATISTICS

Descriptive statistics prior to the onset of the **ET** event, where the change in direction occurs relative to the onset of new directional motion, are reported in the table 13. On average, the mean return x.V*18P records, over the analysis period (2006-2020), a slight depreciation of the euro against the US dollar with a high variance in mean returns.

The slope of the variations between closing prices Sp.Vr*CP.18P preceding the formation of a new market regime (bullish/bearish) reports on average a slight positive trend with high variation in the measurements.

When analyzing the average values of the groups, the average of the mean returns x.V*18P that precede the formation of an uptrend reports on average a slightly higher loss of value (depreciation) relative to its appreciation process, and a high variation between closing prices.

The slope of the closing price-to-closing price variations Sp.Vr*CP.18P preceding the onset of a directional up/down move reports on average a slight negative/positive trend with high variance in the measurements.

While, the average x.V*18P returns preceding the formation of bullish/bearish moves report a slight negative/positive skewness, this asymmetric and leptokurtic effect is common in the returns generated by financial assets. However, the subset of data overcomes this problem, as confirmed in the next section.

2) VALIDATION OF ASSUMPTIONS

The multivariate normality assumption is one of the most important requirements of some multivariate parametric statistical procedures to ensure the reliability of interpretations. If significance tests and goodness-of-fit assessment are not satisfied, it will negatively affect the reliability of interpretations based on the results of these procedures [88], [89]. The results of tests validating compliance with these requirements are presented below.

- Multivariate normality tests for predictors.

There are numerous tests that evaluate multivariate normality assumptions but there is no single standard procedure. However, the Henze-Zirkler and Royston tests suggested in [90] are used in this study because of their good control of type I error and power. Table 14 shows the results of the Henze-Zirkler test for samples larger than 100 observations and its result is validated with the Royston test.

The predictors x.V*18P and Sp.Vr*CP.18P with a significance level greater than 0.05 assume a multivariate normal distribution. In this context, the Henze-Zirkler test statistic follows an approximately log-normal distribution and, therefore, it is more likely that the data, as a function of the variables and groups, also assume a univariate normal distribution [88].

Graphical methods also validate the results obtained. Figure 10 summarizes this analysis. The first graph (**a**) shows the clear agreement between the quantiles of the hypothesized and observed probability distributions. Values in Y tend to be equal in X. That is, the evaluated data do not deviate from multivariate normality.

The ellipsoidal contour line plot (**b**) is useful to verify the multivariate normality of the predictors. The top view of the contour plot of x.V*18P and Sp.Vr*CP.18P hints at a slight Gaussian bell figure with an obvious separation and differentiation between the negative and positive mean returns generated by the Depreciation / Appreciation of the exchange rate, prior to the change in trend. Looking at the histograms in Figure 11, this difference is more evident in x.V*18P than in Sp.Vr*CP.18P and this is because the logarithmic returns are leptokurtic and asymmetric. In addition, the contour lines confirm the positive correlation between groups, as detected in the **GH − Biplot** de la figura 9.

The perspective plot (**c**) is a bivariate representation on a three-dimensional probability distribution surface. It gives information about where the data tend to be concentrated and since the variables are correlated, its shape approximates a Gaussian distribution.

- Univariate normality tests for predictors.

The results of the tests in the table 15 indicate that the subset of data satisfies the assumptions of univariate normality at a significance level of 5%. $AD - Statistic$ is the value of the Anderson-Darling test statistic used to test the hypothesis H_0 that the sample is from a normally distributed population.

Figure 11 verifies through normality plots the structure of the analyzed data. The predictors x.V*18P and Sp.Vr*CP.18P have a univariate normal distribution at 5% of significance level.

- Multivariate normality tests for groups.

Since the significance value derived from the Henze-Zirkler and Royston tests are mathematically greater than 0.05 the results of the 16 table allow us to conclude that the data set of the directions of the upward and downward movements of the euro-dollar exchange rate conform to a multivariate normal distribution.

While the subsets of observations on uptrending and downtrending price directions fit a multivariate normal distribution, in Figure 12 in the contour (**b**) and outlook (**c**) plots it is observed that the distribution of data for the uptrending and downtrending groups have a slight negative and positive skewed distribution. As mentioned, this is due to the fact that the closing price returns, prior to the trend change, have an asymmetric and leptokurtic behavior. However, it is emphasized that these samples follow a multivariate normal distribution.

If the results of the table 16 indicate that the subset of data satisfies the assumptions of multivariate normality at a significance level of 5% and the graphical analysis of the

TABLE 13. Descriptive statistics of predictors and groups. Values prior to time ET.

Measure	n	Mean	Std.Dev	Median	Min	Max	Skew	Kurtosis
Predictor variables								
x.V*18P	125	-0.0000024	0.0001211	0.0000017	-0.0002938	0.000263	-0.0880643	-0.883730
Sp.Vr*CP.18P	125	0.0000011	0.0000204	0.0000013	-0.0000482	0.0000461	-0.021371	-0.3078016
Upward movement								
x.V*18P	62	-0.0001056	0.0000676	-0.0000967	-0.0002938	0.0000006	-0.3798269	-0.4650322
Sp.Vr*CP.18P	62	-0.000012	0.000016	-0.0000122	-0.0000482	0.0000286	-0.0526824	-0.0757937
Downward movement								
x.V*18P	63	0.0000992	0.0000607	0.0000929	0.0000017	0.000263	0.4674263	-0.4223248
Sp.Vr*CP.18P	63	0.0000139	0.0000154	0.0000124	-0.0000207	0.0000461	0.1728039	-0.5063708

TABLE 14. Multivariate normality tests for predictors.

Test	Measures		Statistic	P-Value
Henze-Zirkler	x.V*18P	Sp.Vr*CP.18P	0.987	0.051
Royston	x.V*18P	Sp.Vr*CP.18P	2.708	0.260

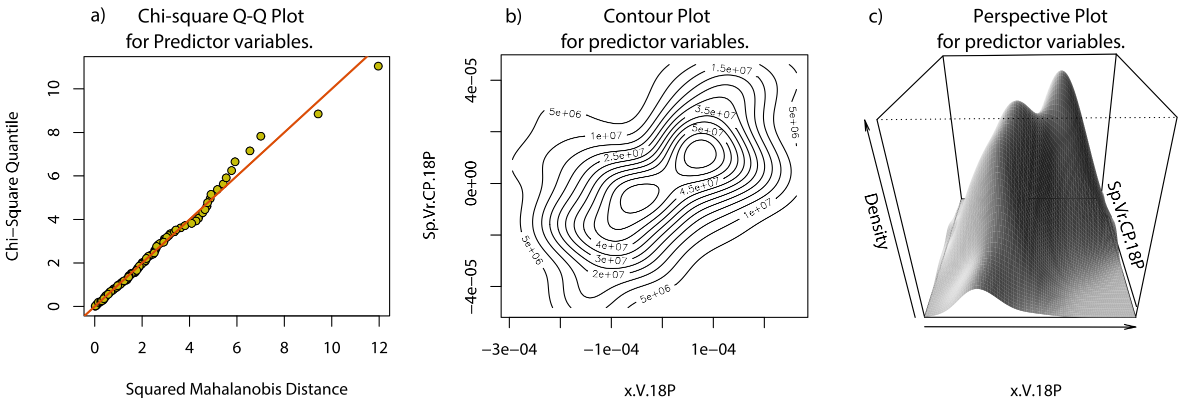


FIGURE 10. Multivariate charts of predictor variables. The structure of the data matrix $\mathbf{Z}\langle 125 \times 2\rangle$ approximates a multivariate normal distribution. (a) Chi-Square Q-Q Plot of long-term movements. (b) Contour Plot from x.V*18P and Sp.Vr*CP.18P. (c) Perspective plot from predictor variables.

TABLE 15. Univariate normality tests for predictors.

Test	Measure	Cases	AD-Statistic	P-Value
Anderson-Darling	x.V*18P	125	0.724	0.058
Anderson-Darling	Sp.Vr*CP.18P	125	0.182	0.912

TABLE 16. Multivariate normality tests for groups.

Test	Trend	Measures		n	Statistic	P-Value
Henze-Zirkler	Upward movement	x.V*18P	Sp.Vr*CP.18P	125	0.874	0.061
	Downward movement	x.V*18P	Sp.Vr*CP.18P	125	0.594	0.317
Royston	Upward movement	x.V*18P	Sp.Vr*CP.18P	125	1.737	0.419
	Downward movement	x.V*18P	Sp.Vr*CP.18P	125	2.705	0.259

figure 12 corroborates the results of the tests, then it is valid to state, as confirmed in the following section, that each predictor and hence each group within each predictor has a univariate normal distribution [88].

- Unvariate normality tests for groups.

According to results in table 17, in all analyses, the P-Value of the Anderson-Darling test statistic is greater than 0.05, consequently, the subsets of observations from each of

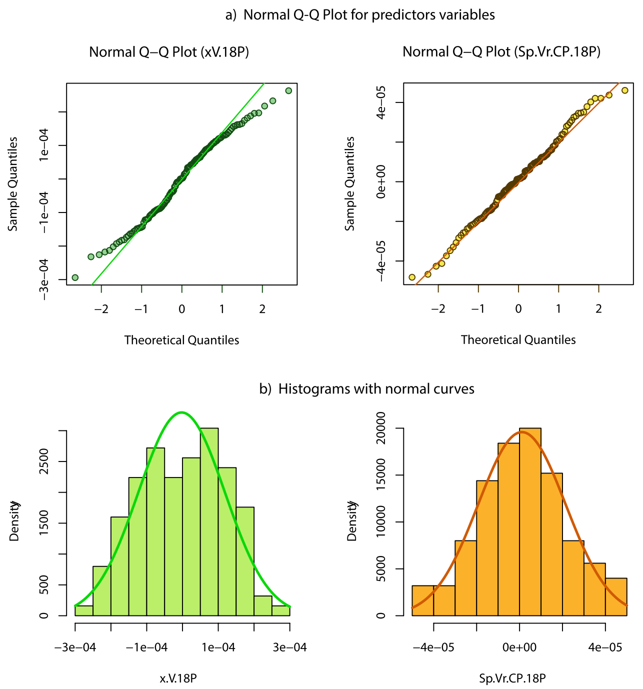


FIGURE 11. **Univariate charts of predictor variables. The structure of the data matrix Z⟨125 × 2⟩ approximates a normal distribution. (a) Normal Q-Q Plot for predictors. (b) Histograms with normal curves.**

the trend groups within each of the predictors x.V*18P y Sp.Vr*CP.18P come from a normally distributed population.

As seen in 13, the distributions of the data for the upward and downward trending groups follow a normal distribution within each of the predictors x.V*18P and Sp.Vr*CP.18P. It is concluded that the selection sample satisfies the multivariate and univariate normality assumptions, and the validity of the results obtained with the test statistics is confirmed by inspection of the multivariate and univariate normality plots.

- Assessment of the contribution of each predictor.

The test for equality of group means measures the potential that each independent variable has before becoming part of the model. The table 18 shows the results of the one-way ANOVA and at a significance value of less than 5% the predictor variables discriminate between group means.

The Wilks' Lambda statistic is another criterion that measures the discriminant potential of the variables. Small values indicate that the selected variable is better at discriminating between groups. The table 18 suggests that the mean return measured 18 periods prior to the regime shift x.V*18P is a better discriminant than the slope of the variances measured 18 moments prior to the trend change Sp.Vr*CP.18P.

- Assessment of the collinearity of the predictors.

The matrix of correlations within groups, from the table 19, shows the correlation between predictors x.V*18P and Sp.Vr*CP.18P. The value obtained suggests a clear independence between predictors. The correlation coefficient is not large enough for instability problems to occur in the signs of the coefficients of the model variables. Consequently, the proposed methodology through the selection of predictors helps to overcome the multicollinearity problem.

In the table 20, the structure matrix is defined by the correlations within groups combined between the predictors and the standardized canonical discriminant function. The order of the variables is a function of the absolute value of this correlation and the coefficients of the ranking functions. Moreover, this order is identical to that shown in the test for equality of group means in the 18 table. This concordance confirms the absence of multicollinearity between the selected independent variables.

The standardized coefficients facilitate the comparison of the independent variables measured on different scales. The coefficient with the highest absolute value corresponds to the predictor variable with the highest discriminant capacity. Thus, the order of the predictors, according to their standardized coefficients, shows the importance of the average return x.V*18P in predicting the future direction of the euro-dollar exchange rate.

Consequently, as the discriminant function is not affected by multicollinearity, it is safe to mention that the future direction of the price movement is mainly determined by the average return x.V*18P of the euro-dollar exchange rate and, therefore, the average return x.V*18P discriminates better between the directions of the upward and downward trend movements.

- Assessment of homogeneity of covariance matrices.

The table 21 presents the values of the ranks and logarithms of the determinants of the covariance matrices of the groups. The logarithms of the determinants are a measure of the variability of the groups. Small differences in the logarithms of the determinants indicate groups with equal covariance matrices. The $Box's\ M$ statistic of the table 22 evaluates the assumption of equality of covariance matrices between groups.

According to results from the table 22, the P-Value of the $Box's\ M$ test statistic is greater than 0.05, consequently, the subsets of observations of the groups in the predictors x.V*18P and Sp.Vr*CP.18P come from populations with equal variance-covariance matrices and, therefore, it can be implicitly inferred that the formulation of the classification functions do not require separate group covariance matrices.

3) ASSESSMENT OF MODEL FIT

Theoretically, the size of the groups in the analysis sample determines the a priori probability of belonging to each group. The definition of the coefficients of the classification functions and the performance of the classification process are determined by the prior probabilities used in the formulation of the models. According to the table 13, the study sample (Z) is formed by groups of equal size. Consequently, so that the calculation of the coefficients of the functions is not affected, equal prior probabilities are used for both groups ($\pi_g = 0.50$).

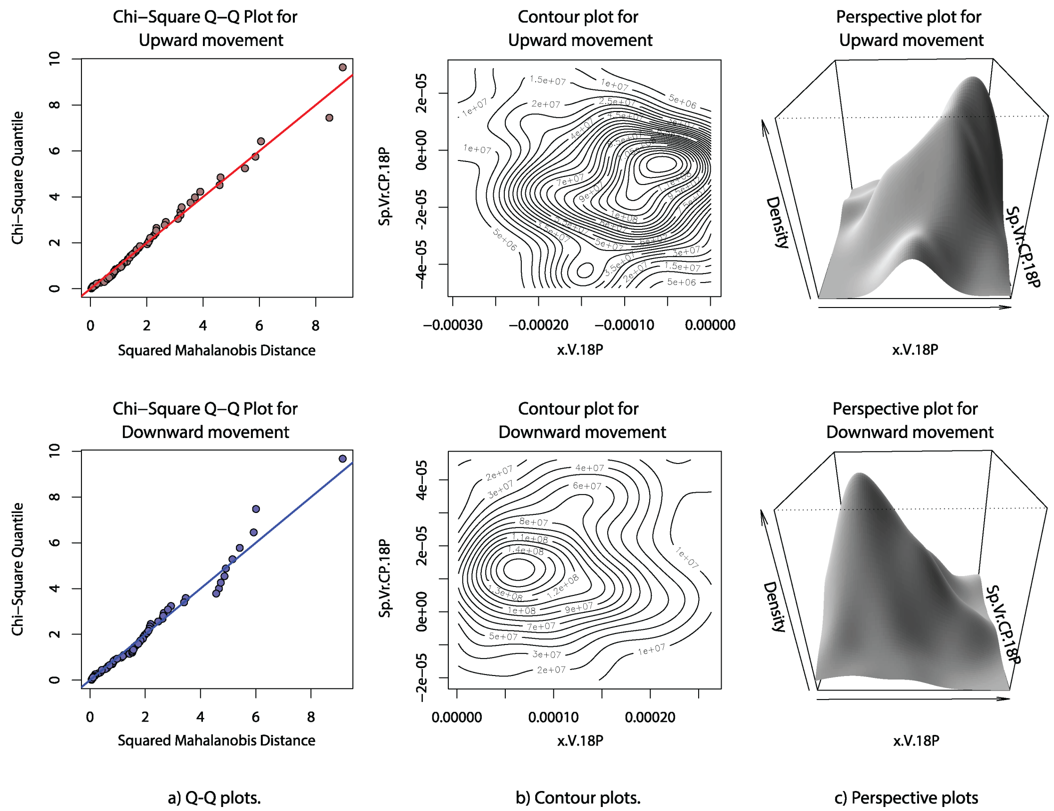


FIGURE 12. **Multivariate charts of groups. The group structures of the data matrix $\mathbf{Z}\langle 125 \times 2\rangle$, upward and downward trending movements, approximate a multivariate normal distribution.(a) Normal Q-Q Plot for groups. (b) Contour plots for groups. (c) Perspective plots for groups.**

TABLE 17. **Univariate normality tests for groups.**

Test	Measure	Trend	Cases	AD-Statistic	P-Value
Anderson-Darling	x.V*18P	Upward movement	62	0.343	0.480
		Downward movement	63	0.402	0.348
Anderson-Darling	Sp.Vr*CP.18P	Upward movement	62	0.212	0.851
		Downward movement	63	0.462	0.251

TABLE 18. **Tests of equality of group means.**

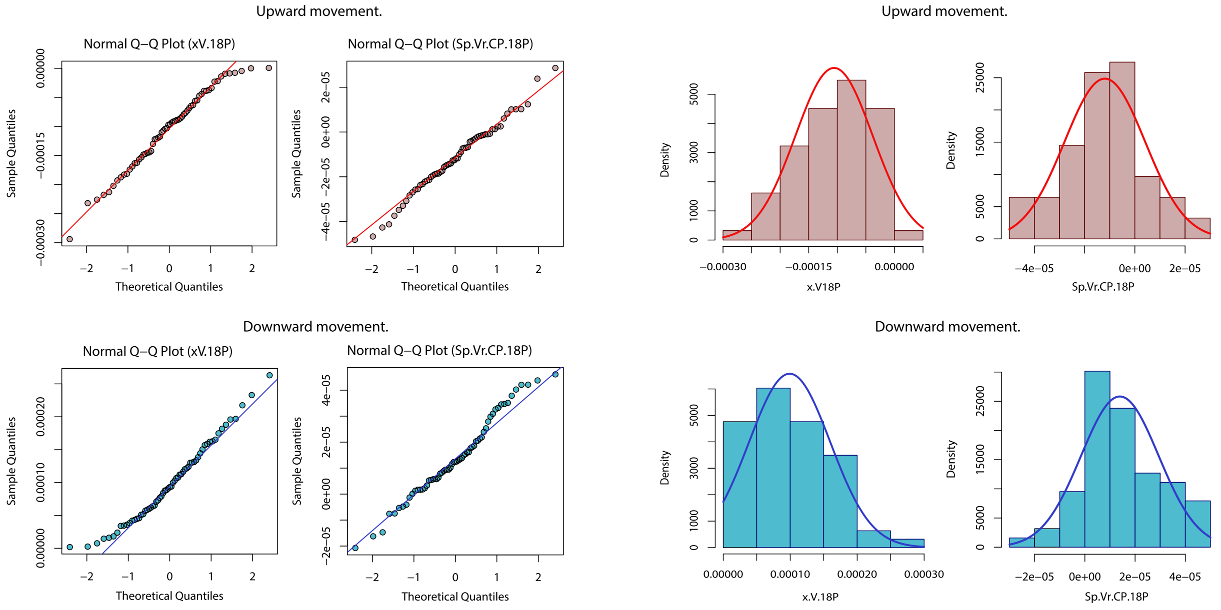
ID	Measure	Wilks' Lambda	F-Value	Df_1	Df_2	P-Value
1	x.V*18P	0.293	205.529	1	85	0.000
2	Sp.Vr*CP.18P	0.638	48.238	1	85	0.000

TABLE 19. **Within-groups correlation matrix.**

Measure	x.V*18P	Sp.Vr*CP.18P
x.V*18P	1	-0.026
Sp.Vr*CP.18P	-0.026	1

The table 23 provides information related to the efficacy measure of the linear discriminant function. For the case in question, since there are two groups to be discriminated, the canonical correlation is used as a measure of the function's effectiveness. This measure is similar to Pearson's correlation between the actual values of the observations in each group and the predicted discriminant scores.

As a result, with an appreciable eigenvalue of **3.048** the predictor variables involved in the model formulation explain 100% of the total cumulative variance, and the **87**% canonical



a) Q-Q plots. b) Histograms with normal curves

FIGURE 13. Univariate charts of groups. The group structures of the data matrix $\mathbf{Z}\langle 125 \times 2\rangle$, upward and downward trending movements, approximate a normal distribution.(a) Normal Q-Q Plot for groups. (b) Histograms with normal curves.

TABLE 20. Structure matrix and standardized function coefficients.

ID	Measure	Discriminant function	Standardized function coefficients
1	x.V*18P	0.891	0.902
2	Sp.Vr*CP.18P	0.432	0.455

TABLE 21. Measure of variability of the groups.

Trend	Matrix rank	Log Determinant
Upward movement	2	-41.261
Downward movement	2	-41.395
Pooled within-groups	2	-41.283

TABLE 22. Box's M-Test of equality of covariances across groups.

Box's M	F-Value	Df_1	Df_2	P-Value
3.092	1.004	3	1122087.83	0.390

correlation between the observations in the groups and the predicted discriminant scores gives a glimpse of a model that fits the behavior of the data and provides, in terms of efficiency, a high correlation between the observed values and the predicted data.

In the table 24 the Wilks' Lambda statistic is used as a measure that checks how well the linear discriminant function separates cases into groups. Small values, close to zero, of the Wilks' Lambda statistic are indicators of the discriminant ability of the function. Chi-square is the value of the associated statistic used to test the H_0 hypothesis. So with a P-value of less than 5%, the hypothesis H_0 that the means of Fisher linear functions are equal between groups is rejected and, therefore, the linear discriminant model contributes to the differentiation of cases between group means.

4) CLASSIFYING THE DIRECTION OF EXCHANGE RATE MOVEMENT

The Fisher linear functions help to assign the cases to a particular membership group. For each case a classification score is calculated and the discriminant analysis model assigns the case to the group whose classification function obtained the highest discriminant score. The linear Fisher functions that best discriminate between upward and downward trending

TABLE 23. Efficiency of the linear discriminant function.

Eigenvalue	% Variance	% Cumulative Variance	Canonical correlation
3.048	100	100	0.868

TABLE 24. Test of functions.

Wilks' Lambda	Chi-Square	Df_1	P-Value
0.247	117.443	2	0.000

movements are summarized in equation (19).

$$y(\hat{\mathbf{X}}) = \beta_0 + \beta_1 \hat{X}_1 + \beta_2 \hat{X}_2, \tag{19}$$

where, $y(\hat{X})$ is the discriminant score, $\hat{X}_1$ is the average return (x.V*18P), $\hat{X}_2$ is the slope of the regression line of the variations between closing prices (Sp.Vr*CP.18P), both criteria calculated over the last 18 periods. The coefficients of the ranking functions β_0, β_1 y β_2 are referenced in the table 25.

Now, starting from the main function (19), the discriminant analysis model works with two linear functions denoted as $y(\hat{X}_U)$ and $y(\hat{X}_D)$ to discriminate between upward trending movements C_U and downward trending movements C_D. The rule that assigns the cases to the membership group is:

$$\text{If } y(\hat{X}_U) \geq y(\hat{X}_D) \text{ then } y(\hat{X}_U) \text{ belongs in } C_U$$
$$\text{Else if } y(\hat{X}_U) < y(\hat{X}_D) \text{ then } y(\hat{X}_D) \text{ belongs in } C_D$$

The structure of the discriminant scores is affected by the signs and magnitude of the coefficients of the variables participating in the model. The upward movement ranking function, according to the table 25, is characterized by the fact that the coefficients of the predictors are less than zero. This means, according to the contribution that each variable makes in the model, that the smaller the average return x.V*18P with respect to the slope of these variations Sp.Vr*CP.18P, the less likely it is that the price of the euro against the US dollar will continue to depreciate. Consequently, after the fall of the exchange rate slows down and enters a phase of exhaustion and correction, at this level, the negative behavior of the price action and occurring before the change of market regime, is the value used by the discriminant function to predict the beginning of an upward movement.

Also in the table 25, in the bearish movement classification function, the change of sign in the coefficients indicates that the cases with positive mean returns x.V*18P greater than the slopes of these variations Sp.Vr*CP.18P, are more likely to have reached the end of an appreciation process, thus announcing the beginning of a bearish market regime.

In contrast to what was said, the box plot in Figure 14 shows the distribution of predictor measures by groups. The expected variances of the data are symmetrically arranged in two different regions bounded by a central zero level. Negative mean returns x.V*18P indicate that the exchange rate has depreciated and remains in the oversold zone. Conversely, when the mean returns x.V*18P are positive, the exchange rate is appreciating and is in the overbought zone. Theoretically, the Q_1, Q_2 and Q_3 quartiles of the box plots define the levels of support and resistance at which the direction of the trend has historically changed and implicitly indicate the number of moves generated at each level. The results show that for each group there is explicit agreement between the predictor measures. Note the 75% similarity between the signs of the values of Sp.Vr*CP.18P and x.V*18P.

The findings obtained seem to suggest that a change of direction in price movement is more likely when the following signals occur. An upward movement is more likely to occur when the negative mean return x.V*18P is greater than the mean of the negative mean returns x.V*18P and the slope of the variances Sp.Vr*CP.18P is negative. Otherwise, a downward movement is more likely to occur when the positive average return x.V*18P is less than the mean of the positive average returns x.V*18P and the slope of the variances Sp.Vr*CP.18P is positive.

In summary, because the mean return has the largest contribution in the discriminant function, a sustained decrease or increase in the mean return of the exchange rate above or below its historical average value makes it more likely that each case, according to the highest discriminant score, will be classified as the beginning of an upward or downward trend movement respectively.

5) MODEL VALIDATION

Table 26 presents the results of applying the model. The initial analysis achieved an impressive classification accuracy of **98.9**%. Out of 38 bearish moves, 37 were correctly classified, and all 49 bullish moves were accurately identified. These cases train and cross-validate the discriminant function, showcasing the model's effectiveness and ability to generalize during cross-validation. In the subsequent phase, a perfect classification accuracy of **100**% was attained by employing observations excluded during the model formulation. This outcome indicates that the model is efficient and capable of accurately detecting the differences between the onset of a bullish move and a bearish one. Importantly, it should be highlighted that the class imbalance within the data subsets does not have a detrimental impact on classification accuracy. The exceptional performance can be attributed to the representative nature of the data sample,

TABLE 25. Coefficients of the classification functions.

Measure	Symbol	Predicted trend	
		Upward movement $y(\hat{X}_U)$	Downward movement $y(\hat{X}_D)$
x.V*18P	β_1	-24598.42	22755.91
Sp.Vr*CP.18P	β_2	-44485.91	52112.42
Constant	β_0	-2.26	-2.16

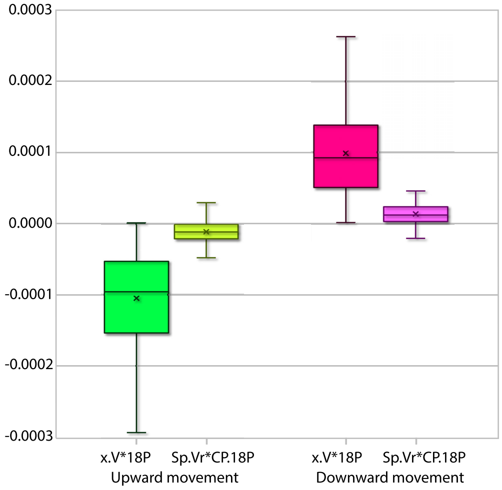


FIGURE 14. Box plots of predictors and groups. Values of x.V*18P and Sp.Vr*CP.18P at the inflection points ET before the trend shift, from the data matrix $\mathbf{Z}\langle 125 \times 2\rangle$. (a) Upward movement. (b) Downward movement.

the discriminative power of the chosen predictor variables, and the suitableness of the employed discriminant analysis methodology.

The obtained accuracy values support the model's usefulness and reliability in forecasting tasks, making it a valuable tool for market analysts and investors, empowering them to make well-informed decisions.

D. FORECAST PERFORMANCE

This section presents the evaluation results of the short-term predictive power of the model for predicting the direction of movement of the euro-dollar exchange rate. Additional experimental results are presented to assess the effectiveness of the proposed methodology. These results are based on out-of-sample data encompassing diverse market conditions and spanning various time horizons. In each instance, the model predicts the initiation of a new directional movement from an inflection point **ET** when the change in the direction of price movements has historically occurred.

Table 27 presents the classification results using out-of-sample data from Fisher's linear functions. The data are divided into two sections: The first corresponds to the training and cross-validation data, and the second to the test data: OOS_1, OOS_2, OOS_3, and OOS_4. Overall, the method ranks the bullish and bearish data exceptionally well, with values from 98.10% to 99.0% and 98.8% to 99.0%, respectively. The results suggest Fisher's linear functions fit the data structure and accurately discriminate between bullish and bearish trends. This finding demonstrates that discriminant analysis is an effective tool for classifying trend data. Furthermore, the relevance of these results and their relation to the objective of the proposed methodology demonstrate effectiveness in predicting short-term trends.

Additional experiments conducted with the OOS_1 and OOS_4 ensembles show classification performance consistent with earlier results. The model correctly classifies bullish and bearish movements with values ranging from 98.7% to 99.0%, and from 98.8% to 99.9%, respectively. It is worth noting that the class imbalance observed in OOS_3 does not have a detrimental impact on classification accuracy. This exceptional performance can be attributed to the representative nature of the training sample, the discriminative quality of the selected predictor variables, and the suitableness of the discriminant analysis methodology.

The experimental results demonstrate the model's goodness of fit and its generalizability and accuracy under different market conditions and time horizons. However, specific requirements must be met to maintain the consistency of the obtained results: (i) Ensure the quality of the results during feature preparation, selection, and validation through structural analysis. (ii) Avoid using inflection points when they are influenced by the release of fundamental data, as prediction accuracy may be affected by the resulting volatility. (iii) Use discriminant and independent characteristics that provide significant differences between trends to ensure accurate classification. (iv) Ensure compliance with the assumptions of multivariate normality in the training and validation samples when opting for discriminant analysis as a classification method. Failure to meet these conditions may compromise the reliability of interpretations, rendering the results merely descriptive rather than inferable. Additionally, it is crucial to perform statistical significance tests at each stage of the process to ensure the reliability and consistency of the obtained results. This strategy ensures the successful application of the proposed approach to other financial instruments.

- Predictive capability of the classification model

Table 28 provides an overview of the predictive power of the classification model obtained through the proposed methodology. Three sections with different performance measures have been included: the first details the performance measures for the training and validation data set; the second, the test data sets; and the third, the descriptive statistics of the mean performance values. The classification model undergoes training and validation on a single training dataset,

TABLE 26. **Classification results.**

Sample analysis cases Subset	125	Trend	Predicted group membership Upward movement	Downward movement	Total
Training set		Upward Movement (n)	49	0	49
		Downward Movement (n)	1	37	38
		Upward Movement (%)	100%	0%	100%
		Downward Movement (%)	2.60%	97.40%	100%
Cross-Validation		Upward Movement (n)	49	0	49
Training set		Downward Movement (n)	1	37	38
	87	Upward Movement (%)	100%	0%	100%
	70%	Downward Movement (%)	2.60%	97.40%	100%
Testing set		Upward Movement (n)	13	0	13
		Downward Movement(n)	0	25	25
	38	Upward Movement (%)	100%	0%	100%
	30%	Downward Movement (%)	0%	100%	100%

TABLE 27. **Out-of-sample classification results.**

Sample analysis cases Subset		Trend	Predicted group membership Upward movement	Downward movement	Total
Training and Cross-Validation Set	125	Upward Movement (n)	62	0	62
		Downward Movement (n)	1	62	63
		Upward Movement (%)	100%	0%	100%
		Downward Movement (%)	1.60%	98.40%	100%
Testing sets:					
OOS_1: 1999-2005	6581	Upward Movement (n)	3306	35	3341
		Downward Movement(n)	40	3200	3240
		Upward Movement (%)	99.00%	1.00%	100%
		Downward Movement (%)	1.20%	98.80%	100%
OOS_2: 2006-2020	14645	Upward Movement (n)	7227	91	7318
		Downward Movement(n)	84	7243	7327
		Upward Movement (%)	98.80%	1.20%	100%
		Downward Movement (%)	1.10%	98.90%	100%
OOS_3: 2006-2020	1748	Upward Movement (n)	569	11	580
		Downward Movement(n)	14	1154	1168
		Upward Movement (%)	98.10%	1.90%	100%
		Downward Movement (%)	1.20%	98.80%	100%
OOS_4: 2021-2023	2336	Upward Movement (n)	1172	16	1188
		Downward Movement(n)	11	1137	1148
		Upward Movement (%)	98.70%	1.30%	100%
		Downward Movement (%)	1.00%	99.00%	100%

and its performance is evaluated on four distinct test datasets: OOS_1, OOS_2, OOS_3, and OOS_4. Several performance metrics are measured on each dataset, including accuracy, sensitivity, specificity, AUC, precision, F1-score, baseline, Kappa coefficient, and Matthews correlation coefficient.

The results indicate that the model has an exceptionally high predictive ability on all test data sets, with an accuracy of over 98.57%. Additionally, the high values of the area under the curve (AUC) indicate a strong discriminatory ability to differentiate between bullish and bearish movements. The model demonstrates impressive sensitivity and specificity, exceeding 97.52% and 98.51%, respectively, across all test datasets. These figures indicate the model's capability to accurately detect the onset of bullish and bearish movements. Furthermore, the model achieves remarkable accuracy and F1-score values, surpassing 98.03% and 97.85%, respectively. These results highlight the model's exceptional ability to predict bullish movements and maintain a strong balance between precision and sensitivity, confirming its effectiveness in identifying both types of trends. Moreover, the Kappa and Matthews correlation coefficients (MCC) exhibit high values exceeding 96.75% across all test datasets. Therefore, the model possesses an exceptional ability to predict the correct class, further reinforcing its overall performance and reliability.

The outstanding performance achieved in this study confirms that the training sample effectively picks up the patterns that best differentiate the trends through the predictor variables. Consequently, the model has a high generalization capacity and produces very accurate predictions without compromising the readability and interpretability of the predictions.

TABLE 28. Model's predictive power.

Data	n	Accuracy	Recall	Specificity	AUC	Precision	F1	Baseline	Kappa	MCC
Training set	125	0.992	0.984	1.000	0.992	1.000	0.992	0.500	0.984	0.984
Cross-validation	125	0.992	0.984	1.000	0.992	1.000	0.992	0.500	0.984	0.984
Testing sets:										
OOS_1: 1999-2005	6581	**0.989**	0.988	0.989	**0.989**	**0.990**	**0.989**	0.500	**0.977**	**0.977**
OOS_2: 2006-2020	14645	0.988	0.989	0.988	0.988	0.988	0.988	0.500	0.976	0.976
OOS_3: 2006-2020	1748	0.986	0.976	**0.991**	0.983	0.981	0.979	**0.556**	0.968	0.968
OOS_4: 2021-2023	2336	0.988	**0.991**	0.986	0.988	0.987	**0.989**	0.500	**0.977**	**0.977**
Descriptive statistics:										
Mean performance		0.9877	0.9860	0.9885	0.9870	0.9865	0.9862	0.5140	0.9745	0.9745
95% confidence	UL	0.9897	0.9967	0.9918	0.9913	0.9926	0.9939	0.5585	0.9814	0.9814
interval for the mean	LL	0.9857	0.9752	0.9851	0.9826	0.9803	0.9785	0.4594	0.9675	0.9675
Standard error		0.0006	0.0033	0.0010	0.0013	0.0019	0.0024	0.0140	0.0021	0.0021
Standard deviation		0.0012	0.0067	0.0020	0.0027	0.0038	0.0048	0.0280	0.0043	0.0043

Finally, the proposed methodology outperforms the reference value (Baseline) across all performance evaluation metrics. Even if the best performance of the state of the art is considered as the benchmark (0.8955, see Table 29), the proposed methodology is superior in all performance evaluation metrics. Moreover, the 95% confidence limits for all performance metrics are closely aligned with their mean values, indicating high accuracy in the results. Additionally, both the standard error and standard deviation are minimal, suggesting low variability in model performance across different datasets. This consistency implies a high level of generalization, accuracy, and reliability of the model over various time horizons and market conditions.

- Comparison with the state of the art

Accurate and interpretable classification models are essential for predicting market behavior in the investment field. Evaluating these characteristics is crucial to determine the effectiveness of the chosen approach. In this regard, Table 29 comprehensively compares the proposed methodology with state-of-the-art techniques, summarizing the key characteristics of the evaluated methodologies. These methodologies include ensemble approaches [91], hybrid approaches [57], neural networks [92], [93], and trend prediction based on sentiment analysis [58]. These approaches are chosen not only for their effectiveness but also for their ability to predict the direction of the euro-dollar exchange rate.

The methodologies evaluated in the study rely on market data (MI) and technical indicators (TI) as inputs. However, these models have limitations due to the restricted nature of the study samples, which only consider specific time horizons. Consequently, their generalizability in the long term may be compromised. Moreover, feature extraction (FE) is widespread in these approaches, which enhances accuracy but often hinders interpretability. While the assessed approaches have demonstrated predictive power using both training and out-of-sample (OOS) data, particular studies, such as [91] and [93], were solely evaluated using in-sample data (TVT). This reliance on in-sample data may result in an overestimation of their predictive ability and a lack of generalizability to new data.

The definition of the hyperparameters in the referred methods is one problem that requires careful assignment to guarantee the reported accuracy. The incorrect assignment of hyperparameters can lead to overfitting issues and limit the model's generalization capacity to specific market conditions and time horizons. In contrast, the proposed methodology, which uses discriminant analysis, avoids the instability of hyperparameters by employing discriminant and independent predictor variables [23].

Table 29 presents the results of the accuracy (Acc) comparison, highlighting the exceptional performance of the proposed methodology (FE and FS) with OOS data, achieving accuracies of 0.975 and 0.987, respectively. The statistical analysis using the HSD Tukey test, with a significance level of less than 1%, confirms the superiority of the proposed methodology in terms of accuracy. Compared to the state-of-the-art approaches, which use deep neural networks [92] with a 0.8955 accuracy, the proposed methodology (FS) demonstrates remarkable improvement in classification accuracy while maintaining the interpretability of the results. This improvement is attributed to various factors, including the incorporation of inflection points, the introduction of new attributes, the discriminant and independent feature selection process, and the multivariate validation prior to constructing a linear discriminant function.

The proposed methodology leverages a comprehensive and representative data set spanning from 1999 to 2023, surpassing the scope of other approaches. This substantial data coverage enhances robustness and generalization when constructing classification models evaluated under different market conditions. Consequently, the resulting classification model is more accurate and reliable, instilling greater confidence in the results' validity. Using a larger data set reduces the risk of the results being influenced by specific sample artifacts, increasing the reliability of the proposed methodology.

Overall, classification models based on artificial intelligence and machine learning techniques significantly impact classification accuracy, requiring sufficient computational resources for training, validation, and deployment phases. However, the simplicity of the formulated model makes it

TABLE 29. Accuracy assessment of prediction approaches.

Methodology	Instrument	Data	FPM	Sample	Prediction	PP	Acc
Sadeghi, 2021 [91]	EUR/USD	MI, TI	FE	2014 - 2019	Daily, Up / Down / Sideway	TVT	0.808
Hybrid model [57]	EUR/USD	MI, TI	FS, FE	2010 - 2015	Daily, Up / Down	OOS	0.8865
Market sentiment ANN [58]	USD/EUR	MI, SN	FS, FSp	2013	Daily, Up / Down	OOS	0.6346
Deep networks CNN [92]	EUR/USD	MI	FEng	2010 - 2015	Daily, Up / Down 1-30 Periods	OOS	0.8955
ANN and DTW [93]	EUR/USD	MI	FS, FE	2011 - 2013	Daily, Up / Down	TVT	0.72
Proposed methodology – LCC (FE)	EUR/USD	MI	FE (15)	**1999 - 2023**	Intraday, Up / Down	OOS	**0.975** ‡
Proposed methodology – LCC (FS)	EUR/USD	MI	FS (2)	**1999 - 2023**	Intraday, Up / Down	OOS	**0.987** ‡

MI = Market Information (OHLCV). **TI** = Technical Indicators. **SN** = Social Networks (StockTwits posts). **FPM** = Feature Pre-processing Method. FE = Feature Extraction, FS = Feature Selection, FSp = Feature Space, FEng = Feature Engineering. **PP** = Model's ability to predict new data outside the training set. TVT = Training, Validation and Testing, OOS = Out-of-Sample Performance. **Acc** = Accuracy. FE(15) = Feature extraction from the first 15 variables in Table 2. FS(2) = Feature selection using discriminative and independent predictors x.V*PT and Sp.Vr*CP.PT measuring price action (See Table 2).
‡ HSD Tukey tests, at less than 1% significance level, reveal a significant difference in accuracy in favor of the control group.

easily interpretable (see Linear discriminant model), stable (see Assessment of the collinearity of the predictors), generalizable (see Out-of-sample results), scalable (see Extra-out-of-sample results) and accurate (see Model's predictive power).

VI. CONCLUSION

This paper demonstrates the effectiveness of the proposed methodology for the formulation of a simple, accurate, parsimonious and easy-to-explain prediction model [69]. Specifically, the here-introduced Linear Classifier Configurator (LCC) integrates data preparation for discriminant analysis, feature selection, and the formulation of a classification model that predicts, in the short term, the onset of the direction of the future movement of the euro-dollar exchange rate. This ingenious approach begins with the exploratory statistical analysis of candidate variables. The effectiveness of the selected variables is validated with the results of the ranking process. The prediction model is formulated from a representative sample following a multivariate normal distribution. Finally, the model is evaluated using out-of-sample data. On average, the proposed methodology provides an exceptionally satisfactory out-of-sample classification accuracy of 98.77%.

The suitability of the LCC methodology makes it particularly useful for identifying the best determinants of the direction of the euro-dollar exchange rate movement. The selected independent predictor variables, according to statistical merit, are the ones that best explain the price action before the change of market regime (object of prediction), have the highest discriminant capacity, the highest predictive power and are the variables that best contribute to the formulation of the classification model. A significant advantage of the proposed method is its ability to deal with and overcome the asymmetric and leptokurtic arrangement of the studied data. Thus, the preparation of data and the extraction of a representative study sample guarantees the requirements of multivariate normality between predictors and groups, overcomes the influence of imbalance in the groups and ensures the fulfillment of statistical assumptions that validate the interpretations made on the results obtained with the discriminant analysis.

The proposed methodology is applicable, from the point of view of technical analysis, to the study of the price behavior of financial instruments, especially those traded on the basis of a reliable and consistent trading strategy over time. In this context, the proposed method is appropriate to identify the determinants of the price movement of the traded asset. The quality of the selected variables is validated with the performance of the constructed classification model.

Consequently, the obtained results allow to conclude, that on the basis of a reliable and consistent trading strategy in time, it is predictable in the short term (a period in the future in 15-minute trading session) the future direction of the euro-dollar exchange rate movement (as a consequence of the change of the market regime). In these cases the results of the study can help to improve the financial performance of trading strategies. In addition, the prediction model can act as a decision support tool. Thus, the proposed model can confirm a buy or sell order prior to its issuance.

The obtained findings suggest that the change of direction is mainly influenced by the average return x.V*18P and the slope of the variations between closing prices Sp.Vr*CP.18P. Consequently, an upward or downward movement is more likely to occur when the average negative or positive return of the exchange rate decreases or increases above or below its historical average value and the slope of the inter-closing price changes is negative or positive.

The implementation of the predictive model in the algorithmic trading system is a future work in progress. The study focuses on systematically analyzing and evaluating, in real time, the performance of the trading system. The evaluation will comprehensively improve the effectiveness of the trading system and the forecasting model in the face of euro-dollar exchange rate volatility. In addition to the exploration of other supervised learning methods, future work planned by the authors also focuses on benchmarking the performance provided by other classification techniques.

An important problem to be addressed in future studies is the prediction of the duration of trending price movements. Although, the prediction model achieves, out of the analysis sample, **98.77%** classification accuracy, a problem that still needs to be investigated, from a technical analysis point of view, is the causality or covariation relationships between the identified metrics and trend price movement duration times.

ACKNOWLEDGMENT

The authors would like to thank the proofreading and valuable feedback by Juan Sebastián Mejía-Ordóñez.

REFERENCES

[1] E. Ilzetzki, C. M. Reinhart, and K. S. Rogoff, "Why is the euro punching below its weight?" *Econ. Policy*, vol. 35, no. 103, pp. 405–460, Apr. 2021.

[2] M. A. Ayadi, W. Ben Omrane, J. Wang, and R. Welch, "Senior official speech attributes and foreign exchange risk around business cycles," *Int. Rev. Financial Anal.*, vol. 80, Mar. 2022, Art. no. 102011.

[3] N. Eiamkanitchat, T. Moontuy, and S. Ramingwong, "Fundamental analysis and technical analysis integrated system for stock filtration," *Cluster Comput.*, vol. 20, no. 1, pp. 883–894, Mar. 2017.

[4] R. D. Edwards, J. Magee, and W. C. Bassetti, *Technical Analysis of Stock Trends*. Boca Raton, FL, USA: CRC Press, 1967.

[5] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, pp. 383–417, May 1970.

[6] K.-Y. Woo, C. Mai, M. McAleer, and W.-K. Wong, "Review on efficiency and anomalies in stock markets," *Economies*, vol. 8, no. 1, p. 20, Mar. 2020.

[7] J. M. Griffin, P. J. Kelly, and F. Nardari, "Do market efficiency measures yield correct inferences? A comparison of developed and emerging markets," *Rev. Financial Stud.*, vol. 23, no. 8, pp. 3225–3277, Aug. 2010.

[8] L. Menkhoff and M. P. Taylor, "The obstinate passion of foreign exchange professionals: Technical analysis," *J. Econ. Literature*, vol. 45, no. 4, pp. 936–972, Nov. 2007.

[9] A. W. Lo and A. C. MacKinlay, "Stock market prices do not follow random walks: Evidence from a simple specification test," *Rev. Financial Stud.*, vol. 1, no. 1, pp. 41–66, Jan. 1988.

[10] J. Conrad and G. Kaul, "Mean reversion in short-horizon expected returns," *Rev. Financial Stud.*, vol. 2, no. 2, pp. 225–240, Apr. 1989.

[11] D. Pradeepkumar and V. Ravi, "Soft computing hybrids for FOREX rate prediction: A comprehensive review," *Comput. Oper. Res.*, vol. 99, pp. 262–284, Nov. 2018.

[12] Z. Hajirahimi and M. Khashei, "Hybridization of hybrid structures for time series forecasting: A review," *Artif. Intell. Rev.*, vol. 56, pp. 1201–1261, May 2022.

[13] X. Li, P. Wu, and W. Wang, "Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong," *Inf. Process. Manage.*, vol. 57, no. 5, Sep. 2020, Art. no. 102212.

[14] G. Zhou, "Measuring investor sentiment," *Annu. Rev. Financial Econ.*, vol. 10, no. 1, pp. 239–259, Nov. 2018.

[15] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, "Text mining for market prediction: A systematic review," *Exp. Syst. Appl.*, vol. 41, no. 16, pp. 7653–7670, Nov. 2014.

[16] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106181.

[17] Z. Hu, Y. Zhao, and M. Khushi, "A survey of forex and stock price prediction using deep learning," *Appl. Syst. Innov.*, vol. 4, no. 1, p. 9, Feb. 2021.

[18] S. Nosratabadi, A. Mosavi, P. Duan, P. Ghamisi, F. Filip, S. Band, U. Reuter, J. Gama, and A. Gandomi, "Data science in economics: Comprehensive review of advanced machine learning and deep learning methods," *Mathematics*, vol. 8, no. 10, p. 1799, Oct. 2020.

[19] A. Thakkar and K. Chaudhari, "A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions," *Exp. Syst. Appl.*, vol. 177, Sep. 2021, Art. no. 114800.

[20] M.-L. Shen, C.-F. Lee, H.-H. Liu, P.-Y. Chang, and C.-H. Yang, "An effective hybrid approach for forecasting currency exchange rates," *Sustainability*, vol. 13, no. 5, p. 2761, Mar. 2021.

[21] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[22] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, May 2019.

[23] J. Iworiso and S. Vrontos, "On the directional predictability of equity premium using machine learning techniques," *J. Forecasting*, vol. 39, no. 3, pp. 449–469, Apr. 2020.

[24] J. W. Goodell, S. Kumar, W. M. Lim, and D. Pattnaik, "Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis," *J. Behav. Experim. Finance*, vol. 32, Dec. 2021, Art. no. 100577.

[25] G. P. Zhang, "Neural networks for classification: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 451–462, Nov. 2000.

[26] L. A. Berrueta, R. M. Alonso-Salces, and K. Héberger, "Supervised pattern recognition in food analysis," *J. Chromatography A*, vol. 1158, nos. 1–2, pp. 196–214, Jul. 2007.

[27] S. Bose, A. Pal, R. SahaRay, and J. Nayak, "Generalized quadratic discriminant analysis," *Pattern Recognit.*, vol. 48, no. 8, pp. 2676–2684, Aug. 2015.

[28] N. Kumar and A. G. Andreou, "Heteroscedastic discriminant analysis and reduced rank HMMs for improved speech recognition," *Speech Commun.*, vol. 26, no. 4, pp. 283–297, Dec. 1998.

[29] J. H. Friedman, "Regularized discriminant analysis," *J. Amer. Statist. Assoc.*, vol. 84, no. 405, pp. 165–175, 1989.

[30] L. Clemmensen, T. Hastie, D. Witten, and B. Ersboll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[31] C. Bouveyron, S. Girard, and C. Schmid, "High dimensional discriminant analysis," *Commun. Statistics-Theory Methods*, vol. 36, no. 14, pp. 2607–2623, 2007.

[32] K. Grobys and N. Sapkota, "Predicting cryptocurrency defaults," *Appl. Econ.*, vol. 52, no. 46, pp. 5060–5076, Oct. 2020.

[33] G. N. F. Weiß, "Copula-GARCH versus dynamic conditional correlation: An empirical study on VaR and ES forecasting accuracy," *Rev. Quant. Finance Accounting*, vol. 41, no. 2, pp. 179–202, Aug. 2013.

[34] S. W. Kwag and Y. S. Kim, "Stock price predictability of financial ratios and macroeconomic variables: A regulatory perspective," *Ind. Eng. Manage. Syst.*, vol. 12, no. 4, pp. 406–415, Dec. 2013.

[35] I. Gurrib and F. Kamalov, "Predicting Bitcoin price movements using sentiment analysis: A machine learning approach," *Stud. Econ. Finance*, vol. 39, no. 3, pp. 347–364, Apr. 2022.

[36] M. Bernardi, L. Catania, and L. Petrella, "Are news important to predict the value-at-risk?" *Eur. J. Finance*, vol. 23, no. 6, pp. 535–572, May 2017.

[37] S. T. Mndawe, B. S. Paul, and W. Doorsamy, "Development of a stock price prediction framework for intelligent media and technical analysis," *Appl. Sci.*, vol. 12, no. 2, p. 719, Jan. 2022.

[38] Z. Chen, C. Li, and W. Sun, "Bitcoin price prediction using machine learning: An approach to sample dimension engineering," *J. Comput. Appl. Math.*, vol. 365, Feb. 2020, Art. no. 112395.

[39] M. T. Leung, H. Daouk, and A.-S. Chen, "Forecasting stock indices: A comparison of classification and level estimation models," *Int. J. Forecasting*, vol. 16, no. 2, pp. 173–190, Apr. 2000.

[40] J. Pekár and M. Pcolár, "Empirical distribution of daily stock returns of selected developing and emerging markets with application to financial risk management," *Central Eur. J. Operations Res.*, vol. 30, no. 2, pp. 699–731, Jun. 2022.

[41] R. A. Eisenbeis, "Pitfalls in the application of discriminant analysis in business, finance, and economics," *J. Finance*, vol. 32, no. 3, pp. 875–900, Jun. 1977.

[42] S. Patra, "Revisiting value-at-risk and expected shortfall in oil markets under structural breaks: The role of fat-tailed distributions," *Energy Econ.*, vol. 101, Sep. 2021, Art. no. 105452.

[43] J. B. Horpestad, Š. Lyócsa, P. Molnár, and T. B. Olsen, "Asymmetric volatility in equity markets around the world," *North Amer. J. Econ. Finance*, vol. 48, pp. 540–554, Apr. 2019.

[44] C. G. Corlu and A. Corlu, "Modelling exchange rate returns: Which flexible distribution to use?" *Quant. Finance*, vol. 15, no. 11, pp. 1851–1864, Nov. 2015.

[45] H.-C. Liu and J.-C. Hung, "Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models," *Exp. Syst. Appl.*, vol. 37, no. 7, pp. 4928–4934, Jul. 2010.

[46] P. Choi and K. Nam, "Asymmetric and leptokurtic distribution for heteroscedastic asset returns: The SU-normal distribution," *J. Empirical Finance*, vol. 15, no. 1, pp. 41–63, Jan. 2008.

[47] J.-C. Hung, M.-C. Lee, and H.-C. Liu, "Estimation of value-at-risk for energy commodities via fat-tailed GARCH models," *Energy Econ.*, vol. 30, no. 3, pp. 1173–1191, May 2008.

[48] D. N. Politis, "A heavy-tailed distribution for arch residuals with application to volatility prediction," Dept. Econ., Univ. California San Diego, San Diego, CA, USA, Tech. Rep., 2004.

[49] P. Theodossiou, "Financial data and the skewed generalized T distribution," *Manage. Sci.*, vol. 44, no. 12, pp. 1650–1661, Dec. 1998.

[50] T. Bollerslev, "A conditionally heteroskedastic time series model for speculative prices and rates of return," *Rev. Econ. Statist.*, vol. 69, no. 3, pp. 542–547, 1987.

[51] R. P. Buckley, D. A. Zetzsche, D. W. Arner, and B. W. Tang, "Regulating artificial intelligence in finance: Putting the human in the loop," *Sydney Law Rev.*, vol. 43, no. 1, pp. 43–81, 2021.

[52] E. Y. Matsumoto, E. Del-Moral-Hernandez, C. E. Yoshinaga, and A. de Campos Pinto, "Forecasting US dollar exchange rate movement with computational models and human behavior," *Exp. Syst. Appl.*, vol. 194, May 2022, Art. no. 116521.

[53] N. Gabdrakhmanova, V. Fedin, B. Matsuta, and M. Pilgun, "The modeling of forecasting new situations in the dynamics of the economic system on the example of several financial indicators," *Proc. Comput. Sci.*, vol. 186, pp. 512–520, 2021.

[54] D. C. Yıldırım, I. H. Toroslu, and U. Fiore, "Forecasting directional movement of forex data using LSTM with technical and macroeconomic indicators," *Financial Innov.*, vol. 7, no. 1, pp. 1–36, Dec. 2021.

[55] S. A. S. Alzaeemi and S. Sathasivam, "Examining the forecasting movement of palm oil price using RBFNN-2SATRA metaheuristic algorithms for logic mining," *IEEE Access*, vol. 9, pp. 22542–22557, 2021.

[56] C. Lu, "Analysis of early warning of RMB exchange rate fluctuation and value at risk measurement based on deep learning," *Comput. Econ.*, vol. 59, pp. 1501–1524, Aug. 2021.

[57] P. P. Das, R. Bisoi, and P. K. Dash, "Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis," *Exp. Syst. Appl.*, vol. 96, pp. 427–449, Apr. 2018.

[58] V. Plakandaras, T. Papadimitriou, P. Gogas, and K. Diamantaras, "Market sentiment and exchange rate directional forecasting," *Algorithmic Finance*, vol. 4, nos. 1–2, pp. 69–79, Apr. 2015.

[59] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[60] W. Ge, P. Lalbakhsh, L. Isai, A. Lenskiy, and H. Suominen, "Neural network–based financial volatility forecasting: A systematic review," *ACM Comput. Surveys*, vol. 55, no. 1, pp. 1–30, Jan. 2022.

[61] S. Aziz, M. Dowling, H. Hammami, and A. Piepenbrink, "Machine learning in finance: A topic modeling approach," *Eur. Financial Manage.*, vol. 28, no. 3, pp. 744–770, Jun. 2022.

[62] Y. Qiu, Y. Qiu, Y. Yuan, Z. Chen, and R. Lee, "QF-TraderNet: Intraday trading via deep reinforcement with quantum price levels based profit-and-loss control," *Frontiers Artif. Intell.*, vol. 4, pp. 1–12, Oct. 2021.

[63] M. O. Özorhan, İ. H. Toroslu, and O. T. Şehitoğlu, "A strength-biased prediction model for forecasting exchange rates using support vector machines and genetic algorithms," *Soft Comput.*, vol. 21, no. 22, pp. 6653–6671, Nov. 2017.

[64] Y. Qiu, Z. Song, and Z. Chen, "Short-term stock trends prediction based on sentiment analysis and machine learning," *Soft Comput.*, vol. 26, pp. 2209–2224, Jan. 2022.

[65] S. Ahmed, M. M. Alshater, A. E. Ammari, and H. Hammami, "Artificial intelligence and machine learning in finance: A bibliometric review," *Res. Int. Bus. Finance*, vol. 61, Oct. 2022, Art. no. 101646.

[66] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106181.

[67] B. Žlicar and S. Cousins, "Discrete representation strategies for foreign exchange prediction," *J. Intell. Inf. Syst.*, vol. 50, no. 1, pp. 129–164, Feb. 2018.

[68] E. Steurer, M. Rothenhausler, and Y. Yeo, "Forecasting discretized daily USD/DEM exchange rate movements with quantitative models," in *Proc. 22nd Annu. Conf. German-Classification-Soc.*, 1999, pp. 467–472.

[69] T. Dao and H. Ahn, "An optimized combination of π-fuzzy logic and support vector machine for stock market prediction," *J. Intell. Inf. Syst.*, vol. 20, no. 4, pp. 43–58, Dec. 2014.

[70] M. G. Novak and D. Velušcek, "Prediction of stock price movement based on daily high prices," *Quant. Finance*, vol. 16, no. 5, pp. 793–826, May 2016.

[71] M. Sharma, S. Sharma, and G. Singh, "Performance analysis of statistical and supervised learning techniques in stock data mining," *Data*, vol. 3, no. 4, p. 54, Nov. 2018.

[72] J. Okicic, S. Remetic-Horvath, and B. Buyukdemir, "Stock selection based on discriminant analysis: Case of capital market of Bosnia and Herzegovina," *J. Econ. Social Stud.*, vol. 4, no. 2, pp. 5–30, 2014.

[73] C. Zopounidis, M. Doumpos, and S. Zanakis, "Stock evaluation using a preference disaggregation methodology," *Decis. Sci.*, vol. 30, no. 2, pp. 313–336, Mar. 1999.

[74] S.-N. Hwang, C.-T. Lin, and W.-C. Chuang, "Stock selection using data envelopment analysis-discriminant analysis," *J. Inf. Optim. Sci.*, vol. 28, no. 1, pp. 33–50, Jan. 2007.

[75] H. Ögüt, M. M. Doğanay, and R. Aktas, "Detecting stock-price manipulation in an emerging market: The case of Turkey," *Exp. Syst. Appl.*, vol. 36, no. 9, pp. 11944–11949, Nov. 2009.

[76] W. Huang, Y. Nakamori, and S.-Y. Wang, "Forecasting stock market movement direction with support vector machine," *Comput. Operations Res.*, vol. 32, no. 10, pp. 2513–2522, Oct. 2005.

[77] M. P. G. Villardón, "Una alternativa de representacion simultanea: HJ-BIPLOT," *Qüestiió, Quaderns d'estadística I Investigació Operativa*, vol. 10, pp. 13–23, Apr. 1986.

[78] K. R. Gabriel, "The biplot graphic display of matrices with application to principal component analysis," *Biometrika*, vol. 58, no. 3, pp. 453–467, 1971.

[79] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[80] Z. Bousbaa, O. Bencharef, and A. Nabaji, "Stock market speculation system development based on technico temporal indicators and data mining tools," in *Heuristics for Optimization and Learning*. Cham, Switzerland: Springer, 2021, pp. 239–251.

[81] *IBM SPSS Statistics for Windows, Version 26.0*, IBM Corp., Armonk, NY, USA, 2019.

[82] J. L. Vicente-Villardón, "MULTBIPLOT: A package for multivariate analysis using biplots," Mathlab Softw., Departamento de Estadística, Universidad de Salamanca, Salamanca, Spain, Tech. Rep., 2015.

[83] S. Korkmaz, D. Goksuluk, and G. Zararsiz, "MVN: An R package for assessing multivariate normality," *R J.*, vol. 6, no. 2, p. 151, 2014.

[84] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: One-sided selection," in *Proc. 4th Int. Conf. Mach. Learn.* Princeton, NJ, USA: Citeseer, vol. 97, 1997, pp. 179–186.

[85] Z. C. Lipton, C. Elkan, and B. Naryanaswamy, "Optimal thresholding of classifiers to maximize F1 measure," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Berlin, Germany: Springer, 2014, pp. 225–239.

[86] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[87] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, pp. 1–13, Dec. 2020.

[88] B. G. Tabachnick, L. S. Fidell, and J. B. Ullman, *Using Multivariate Statistics*. Boston, MA, USA: Pearson Education, 2007.

[89] H. Finch, "Identification of variables associated with group separation in descriptive discriminant analysis: Comparison of methods for interpreting structure coefficients," *J. Experim. Educ.*, vol. 78, no. 1, pp. 26–52, Sep. 2009.

[90] T. Svantesson and J. W. Wallace, "Tests for assessing multivariate normality and the covariance structure of MIMO data," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 4, Apr. 2003, p. 656.

[91] A. Sadeghi, A. Daneshvar, and M. M. Zaj, "Combined ensemble multi-class SVM and fuzzy NSGA-II for trend forecasting and trading in forex markets," *Exp. Syst. Appl.*, vol. 185, Dec. 2021, Art. no. 115566.

[92] S. Galeshchuk and S. Mukherjee, "Deep networks for predicting direction of change in foreign exchange rates," *Intell. Syst. Accounting, Finance Manage.*, vol. 24, no. 4, pp. 100–110, Oct. 2017.

[93] L. C. Tiong, D. C. Ngo, and Y. Lee, "Forex prediction engine: Framework, modelling techniques and implementations," *Int. J. Comput. Sci. Eng.*, vol. 13, no. 4, pp. 364–377, 2016.

MAURICIO ARGOTTY-ERAZO was born in La Unión, Nariño, Colombia, in 1976. He received the degree (Hons.) in industrial engineering and the Specialist degree in administration and management from Universidad Cooperativa de Colombia, Pasto, in 2000 and 2009, respectively, the master's degree (Hons.) in financial management and consulting from Universidad Mariana, in 2015, and the master's degree (Hons.) in advanced multivariate data analysis and big data from Universidad de Salamanca, Spain, in 2018, where he is currently pursuing the Ph.D. degree in applied multivariate statistics.

From 2002 to 2016, he was a Research Professor and a Coordinator of the Engineering Laboratories, Industrial Engineering Program, Universidad Cooperativa de Colombia, Pasto. In 2016, he was a Postgraduate Professor with the Senior Management, Accounting, and International Auditing Specialization Programs and the Administration and Competitiveness Master's Program, Universidad Mariana, Pasto, Colombia. Currently, he is a Master's Thesis Advisor with the Marketing Master's Program, Universidad de Nariño, Pasto. His research interests include quantitative analysis, simulation, financial modeling, multicriteria optimization, advanced multivariate data analysis and big data, statistical analysis, data visualization, machine learning, and artificial intelligence applied to financial markets.

ANTONIO BLÁZQUEZ-ZABALLOS was born in Macotera, Salamanca, Spain, in 1969. He received the B.S. degree in mathematics and the Ph.D. degree in mathematical sciences from Universidad de Salamanca, Spain, in 1992 and 1998, respectively.

He was a Research Professor with the Department of Statistics, Universidad de Salamanca, in the following positions: from 1994 to 1999, he was an Assistant University Professor. From 1999 to 2004, he was a full-time Associate Professor. From 2004 to 2009, he was an Assistant Professor (Ph.D.). Since 2009, he has been an Associate Professor (Ph.D.). His research interests include multivariate statistical methods applied to the analysis of economic and biological data and applied mathematics.

Dr. Blázquez-Zaballos awards and distinctions include the Extraordinary Doctorate Award (1997–1998).

CARLOS A. ARGOTY-ERASO was born in Pasto, Nariño, Colombia, in 1977. He received the degree in financial administration from Universidad Mariana, Pasto, in 2002. He is currently pursuing the master's degree in marketing with Universidad de Nariño, Colombia.

From 1998 to 2005, he was the Assistant Manager in the private sector. He promoted the launching of new products and the brand positioning with Lincoln Inversiones, Pasto. From 2002 to 2004, he worked in private banking, promoting access to new financial products and services. From 2005 to 2007, he was the Business Manager of GMAC Financiera de Colombia S.A., Pasto. Since 2008, he has been working for different local and national companies, both public and private, in researching market behavior and consumer needs. His research interests include the measurement, modeling, the management of financial risk, the formulation of credit scoring models, digital marketing and consumer psychology, predictive models of consumer behavior based on market data, and artificial intelligence methods.

LEANDRO L. LORENTE-LEYVA was born in Manzanillo, Cuba, in 1985. He received the degree in mechanical engineering and the master's degree in CAD/CAM from Universidad de Holguín, Cuba, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree in engineering and industrial production with Universitat Politécnica de Valéncia (UPV), Spain. From 2009 to 2013, he was a Specialist with Astilleros de Oriente, Manzanillo, Cuba. From 2014 to 2015, he was a Project Advisor with Ministerio de Industrias, Cuba—PDVSA Industrial, Caracas—Venezuela. In 2015, he was a Senior Engineer with Empresa de Ingeniería del Reciclaje, Ministerio de Industrias, Cuba. From 2015 to 2020, he was a Researcher/Professor with Universidad Técnica del Norte, Ecuador. Since 2021, he has been a Program Coordinator with the Postgraduate Center, Universidad Politécnica Estatal del Carchi, Ecuador. He is currently a Senior Researcher with the SDAS Research Group. His professional and research interests include discrete and computational optimization and artificial intelligence applications, manufacturing engineering, computer-aided design, planning and manufacturing (CAD/CAPP/CAM), and application of numerical methods in engineering. In addition, time series forecasting, data analysis, and the applications of heuristic methods as solutions to engineering problems.

NADIA N. SÁNCHEZ-POZO was born in Tulcán, Ecuador, in 1992. She received the degree in electronics and communication networks engineering from Universidad Técnica del Norte, Ibarra, Ecuador, in 2018, and the master's degree in data science from Universitat Oberta de Catalunya, Spain, in 2021. From 2018 to 2022, she was a DL/ML Engineer with the SDAS Research Group. Currently, she is a Master's Thesis Advisor and a Lecturer with the Applied Statistics Master's Program, Universidad Politécnica Estatal del Carchi (UPEC), Ecuador. Her main research interests include big data, machine learning, and natural language processing.

DIEGO H. PELUFFO-ORDÓÑEZ was born in Pasto, Colombia, in 1986. He received the degree in electronic engineering, the master's degree in industrial automation, and the Ph.D. degree in engineering from Universidad Nacional de Colombia, Manizales, Colombia, in 2008, 2010, and 2013, respectively. In 2012, he undertook his doctoral internship with KU Leuven, Belgium. From 2013 to 2014, he was a Postdoctoral Researcher with Université Catholique de Louvain, Louvain la-Neuve, Belgium. From 2014 to 2015, he was a Lecturer with Universidad Cooperativa de Colombia, Pasto. From 2015 to 2017, he was a Researcher/Professor with Universidad Técnica del Norte, Ecuador. From 2017 to 2020, he was a Researcher/Professor awith Yachay Tech. Ecuador. From 2020 to 2022, he was a Consultant/Curriculum Author with deeplearning.ai. Currently, he is an Assistant Professor with the College of Computing, Mohammed VI Polytechnic University, Morocco. He is also the Founder and the Head of the SDAS Research Group. He is also an invited Lecturer and an External Researcher with Corporación Universitaria Autónoma de Nariño, Pasto, and a member of the SEDMATEC Research Group. Also, he is an External Collaborator with the Writing Laboratory, Tecnológico de Monterrey, Mexico. He is also an External Supervisor of Ph.D. Programs with Universidad de Granada, Spain, Universitat Politécnica de Valéncia, Spain, and Universidad Nacional de La Plata, Argentina. His main research interests include kernel-based and spectral methods for data clustering and dimensionality reduction, complex high-dimensional data, signal, image, and video analysis for medical and industry applications. He has served as an organizing committee member (the general chair, the session chair, and the competitions chair) and a keynote speaker at several conferences. Also, he has served as a Guest Editor for the *Computers and Electrical Engineering Journal*.