An interval evolving fuzzy model for Bitcoin range-based volatility forecasting

***Abstract*—A fuzzy rule-based model to process interval-valued stream data is proposed in this paper. The model comprises interval adaptive fuzzy rules with affine consequents whose structure and functionality are updated as data are input. Adaptation of the antecedents of the rules uses participatory learning to cluster the interval-valued data. The rule consequents are found using the center and range of the intervals, and the weighted least squares. Computational experiments consider Bitcoin high and low prices forecasting and their application in range-volatility modeling. Results indicate that the fuzzy model forecasts appear as a potential tool for Bitcoin highs and lows prediction and in risk management applications.**

***Index Terms*—Evolving fuzzy systems, interval-valued data, Bitcoin, volatility, forecasting.**

I. Introduction

Symbolic Data Analysis (SDA) [1] belongs to the domain of multivariate analysis, pattern recognition, machine learning and artificial intelligence by extending classical methodologies of exploratory data analysis and statistical models for processing symbolic data. Symbolic data consists of the aggregation of information through lists, intervals, histograms, probability distributions, among others. In the current context of increasing development of autonomous systems, sensor networks, high-powered small-scale computing devices, and data capture and storage technologies, methods of summarizing and extracting relevant information from large databases play a fundamental role in data management and in many decision-making situations [2].

Interval-valued data are of particular interest in the domain of SDA because in many cases information can be naturally recorded as intervals due to imprecision and/or aggregation [3], [4]. Typical examples of interval data are daily temperatures [5], financial data [6]–[8], pulse pressure [9], air quality measures [10], [11], traffic speed monitoring [12], agricultural and livestock prices [13]. Methodologies for the analysis of interval-valued data have been extensively considered in the literature of SDA during the past three decades. The advantage of treating intervals is to consider the uncertainty and variability associated with the data [14].

Finance is an area that naturally produces interval-valued data indexed by time to form interval-valued time series (ITS). Asset prices during a negotiation period are recorded and can be summarized by the corresponding open, closing, low and high values. Many works in the SDA literature have investigated modeling and forecasting of interval-valued asset prices whose interval bounds are the highest and lowest prices [3], [6]–[8], [15]–[17]. For example, if only the daily closing (or opening) prices of a particular asset are collected, then the resulting time series neglects intraday variability and important information is missed, resulting in a deterioration in the predictive ability of models that considered such series [18].

Particularly, high and low prices are directly related to the concept of volatility, a key variable in finance. Interestingly, [19] suggests the difference between the logarithm of the maximum and minimum prices of an asset over time as an efficient measure of volatility, known as the volatility range[1]. [21] highlighted that the volatility range is a robust estimator for market microstructures, and its use overcomes some of the traditional volatility models limitations when calculated using closing prices, resulting in less accurate predictions.

Several studies have addressed the evaluation of range-based volatility models. For instance, [22] examined different range volatility estimators to analyze the channels of variance transmission across emerging and developed markets in Asia. The effect of non-trading periods on the volatility modeling of S&P 500 using range-based volatility models is investigated by [23]. Realized range-based volatility methods were considered by [24], [25], [26], and [27]. Generally, researchers have advocated the use of range-based volatility models, constructed by asset prices ITS, as a powerful risk measure.

The aim of this paper is to suggest an evolving fuzzy modeling approach designed to process interval-valued data. An application concerning volatility range modeling and forecasting for the cryptocurrency market, particularly Bitcoin, is developed. The approach addresses an evolving interval fuzzy model that processes interval-valued input data as a stream, updating dynamically its structure and functionality as new data arrives. The model structure accounts for the nonlinear, time-varying, and uncertain dynamics of the data, characteristics which are typical in financial time series. The model is composed by a collection of fuzzy functional rules whose antecedents (cluster structure) are identified through participatory learning clustering, and the consequent updated using weighted recursive least squares translated to handle ITS. Finally, Bitcoin highs and lows predictions are used to compute volatility through a range-based method. Volatility is also evaluated in terms of accuracy.

The contributions of this paper can be summarized as follows. First, it extends the fuzzy model of [28] in which the rule consequent parameters are computed using low and up interval bounds. Instead, this paper considers the center

[1]Literature on the "volatility range" as a proxy for volatility has been observed since the 1980s with work of [20].

and range as the evidence of a higher accuracy achieved when these interval characteristics are considered for ITS modeling [8], [29]. Second, interval forecasts are not evaluated in terms of accuracy solely, as most of the literature does, but proceeds by constructing a range-based volatility model. This provides an economic evaluation of the interval model and brings to the volatility literature the issue of effectiveness of range-based models when constructed from forecasts of interval-valued time series.

After this introduction, this work proceeds as follows. The interval evolving fuzzy model proposed is described in Section II. Section III presents an empirical analysis for Bitcoin high and low prices forecasting, in addition to the use of the forecasts to construct range-based volatility models. Finally, Section IV concludes the paper and lists topics for future investigation.

II. Interval evolving fuzzy modeling

A. Interval-valued time series

Let $[x] = [x^L, x^U] \in \mathcal{K}_c(\mathbb{R})$ be an interval-valued datum, where $\mathcal{K}_c(\mathbb{R}) = \{[x^L, x^U] : x^L, x^U \in \mathbb{R}, x^L \leq x^U\}$ is the set of closed intervals of the real line $\mathbb{R}$, and x^L and x^U are the lower and upper bounds of the interval $[x]$. The interval $[x]$ can also be denoted by a two-dimensional vector $[x] = [x^L, x^U]^T$.

This paper uses interval-valued data computation of [30], respectively:

$$\begin{aligned} [x] + [y] &= [x^L + y^L, x^U + y^U] \\ [x] - [y] &= [x^L - y^U, x^U - y^L] \\ [x] \cdot [y] &= [\min\{x^L y^L, x^L y^U, x^U y^L, x^U y^U\}, \\ &\quad \max\{x^L y^L, x^L y^U, x^U y^L, x^U y^U\}] \\ [x] \,/\, [y] &= [x] \cdot (1/[y]),\ 1/[y] = [1/y^U, 1/y^L] \end{aligned} \quad (1)$$

The center or mid-point of $[x]$, denoted as x^c, is calculated as $x^c = \dfrac{x^L + x^U}{2}$, and the range (or radii/radius) of $[x]$, x^r is $x^r = \dfrac{x^U - x^L}{2}$. Using the center and the range of an interval-valued datum $[x]$, its lower and upper bounds can be recovered as $x^L = x^c - x^r$ and $x^U = x^c + x^r$, respectively.

The definition of the distance between two intervals is required, as it is needed to measure their (dis)similarity, a key feature when clustering interval-valued data using participatory learning algorithm. Here we consider the Hausdorff-Pompeiu distance for intervals [31], [32]. The distance $dH([x], [y])$ between $[x]$ and $[y]$ is:

$$dH([x], [y]) = \max\left\{|x^L - y^L|, |x^U - y^U|\right\}. \quad (2)$$

A sequence of interval-valued variables in consecutive time steps $t = 1, 2, \ldots, N$, denoted by $\{[x]^t\} = \{[x]^1, [x]^2, \ldots, [x]^N\}$, is an interval-valued time series (ITS), where N is the sample size or the number of intervals in the time series, and $[x]^t = [x^{L,t}, x^{U,t}]$. Evolving fuzzy system for interval-valued data (eFS$_I$) processes ITS as streams using the interval modeling framework described in the next section.

B. eFS$_I$ structure

The eFS$_I$ for interval-valued data modeling consists of a rule base of interval fuzzy rules. Each fuzzy rule has an antecedent part associated with the state of the interval-valued input variable, and a consequent part specifying the corresponding interval-valued output variable. More precisely, the model is a collection of functional interval fuzzy rules, as in [33], but extended to process interval data. More precisely, the interval fuzzy rule base in composed by c fuzzy rules of the form:

$$\mathcal{R}_i:\ \text{IF } [\mathbf{x}] \text{ is } \mathcal{M}_i \text{ THEN } [y_i] = [\beta_{i,0}] + [\beta_{i,1}][x_1] + \ldots + [\beta_{i,n}][x_n], \quad (3)$$

where $\mathcal{R}_i$ is the i-th fuzzy rule, $i = 1, 2, \ldots, c$, c is the number of fuzzy rules in the rule base, $[\mathbf{x}] = ([x_1], [x_2], \ldots, [x_n])^T$ the n-dimensional vector of inputs, $[x_j] = [x_j^L, x_j^U] \in \mathcal{K}_c(\mathbb{R})$, $j = 1, \ldots, n$, and $[\beta_{i,l}] = [\beta_{i,l}^L, \beta_{i,l}^U] \in \mathcal{K}_c(\mathbb{R})$ are interval-valued parameters of the rule consequent, $l = 0, 1, \ldots, n$. $\mathcal{M}_i$ is the fuzzy set of the antecedent whose membership function is $\mu_i([\mathbf{x}]) : \mathcal{K}_c(\mathbb{R}^n) \to [0, 1]$, and $[y_i] = [y_i^L, y_i^U] \in \mathcal{K}_c(\mathbb{R})$ is the output of the i-th rule.

The output of the model, $[y]$, is computed as the weighted average of the individual output intervals of the rules (3):

$$[y] = \sum_{i=1}^{c} \left(\frac{\mu_i([\mathbf{x}])[y_i]}{\sum_{j=1}^{c} \mu_j([\mathbf{x}])} \right) = \sum_{i=1}^{c} \lambda_i [y_i], \quad (4)$$

where $\lambda_i = \dfrac{\mu_i([\mathbf{x}])}{\sum_{j=1}^{c} \mu_j([\mathbf{x}])}$ is the normalized degree of activation of the i-th rule. The membership degree $\mu_i([\mathbf{x}])$ of datum $[\mathbf{x}]$ is given by:

$$\mu_i([\mathbf{x}]) = \left[\sum_{h=1}^{c} \left(\frac{\sum_{j=1}^{n} \left(\max\left\{ |x_j^L - v_{i,j}^L|, |x_j^U - v_{i,j}^U| \right\} \right)}{\sum_{j=1}^{n} \left(\max\left\{ |x_j^L - v_{h,j}^L|, |x_j^U - v_{h,j}^U| \right\} \right)} \right)^{\frac{2}{(m-1)}} \right]^{-1}, \quad (5)$$

where m is a fuzzification parameter ($m = 2$ in this paper), and $[v_{i,j}] = [v_{i,j}^L, v_{i,j}^U] \in \mathcal{K}_c(\mathbb{R})$ is the cluster center of the i-th cluster/rule, $j = 1, \ldots, n$ and $i = 1, \ldots, c$.

eFS$_I$ learning has two steps, antecedents and consequents learning, respectively. Learning antecedents concerns clustering the interval-valued data space, whereas estimation of local models parameters corresponds to consequents learning.

C. eFS$_I$ antecedents learning

eFS$_I$ antecedent rules learning is done using the participatory learning fuzzy clustering algorithm extended to handle interval-valued data proposed in [34]. The interval data set $[\mathbf{X}] = \{[\mathbf{x}_1], \ldots, [\mathbf{x}_N]\}$ is partitioned in c fuzzy subsets, $2 \leq c \leq N$, where N is the number of samples. Intervals $[\mathbf{x}_j]$ bounds are assumed to be normalized using min-max operator:

$$x_{norm}^B = \frac{x^B - \min\{x^L, x^U\}}{\max\{x^L, x^U\} - \min\{x^L, x^U\}}, \quad (6)$$

where B denotes either the lower bound L, or the upper bound U of the interval.

Clustering, using participatory learning in the interval-valued data space, assumes that the current cluster structure plays a key role in the learning process whenever new data

are input. The compatibility of the data with the current structure directly influences the cluster structure organization, which suggests the need for model revision based on the compatibility of the new data with the current cluster structure.

The cluster structure is defined by the cluster centers $[\mathbf{V}] = [[\mathbf{v}_1], \ldots, [\mathbf{v}_c]]$, $[\mathbf{v}_i] = ([v_{i,1}], \ldots, [v_{i,n}])^T$, $[v_{i,j}] = [v_{i,j}^L, v_{i,j}^U] \in \mathcal{K}_c([0,1])$, $j = 1, \ldots, n$ with $i = 1, \ldots, c$. Participatory learning constructs the model structure, i.e. defines $[\mathbf{V}]$ from input data $[\mathbf{x}]^t \in [0,1]^n$, $t = 1, \ldots$. Notice that $[\mathbf{x}]^t$ is used as a vehicle to learn about $[\mathbf{V}]$.

The contribution of $[\mathbf{x}]^t$ to the learning process at step t is evaluated based on its compatibility to the current cluster structure $[\mathbf{V}]^{t-1}$, i.e. the learning process is participatory. The compatibility $\rho_i^t \in [0,1]$ of input $[\mathbf{x}]^t$ with the cluster center $[\mathbf{v}_i]^{t-1}$ of $[\mathbf{V}]^{t-1}$, $i = 1, \ldots, c$ is computed using (2) as follows:

$$\rho_i^t = 1 - \frac{1}{n}\sum_{j=1}^{n}\left(\max\left\{|x_j^{L,t} - v_{i,j}^{L,t-1}|, |x_j^{U,t} - v_{i,j}^{U,t-1}|\right\}\right). \quad (7)$$

If cluster center $[\mathbf{v}_i]^{t-1}$ is the one most compatible with $[\mathbf{x}]^t$, that is, $i = \arg\max_{j=1,\ldots,c}\{\rho_j^t\}$, then it is updated using:

$$[\mathbf{v}_i]^t = [\mathbf{v}_i]^{t-1} + G_i^t([\mathbf{x}]^t - [\mathbf{v}_i]^{t-1}), G_i^t = \alpha\rho_i^t \quad (8)$$

where $\alpha \in [0,1]$ is the basic learning rate. Notice that (8) can be viewed as a form of exponential smoothing modulated by the compatibility of data with the model structure, which is the nature of participatory learning.

A sequence of input data with low compatibility with the current cluster structure indicates that the current model should be revised in front of new information. Participatory learning uses an arousal mechanism to monitor the need to revise the cluster structure. A high arousal value indicates less confidence in how the current model fits recent data, a complement of confidence [34]. A way to compute the arousal $a_i^t \in [0,1]$ at step t is:

$$a_i^t = a_i^{t-1} + \beta(1 - \rho_i^t - a_i^{t-1}), \quad (9)$$

where $\beta \in [0,1]$ controls the rate of change of arousal. The closer β is to one, the faster the learning process senses compatibility variations.

When $a_i^t \geq \tau \in [0,1]$ for $i = 1, \ldots, c$, where τ is a threshold chosen by the user, a new cluster should be created, assigning the current data as its cluster center, that is, $[\mathbf{v}_{c+1}]^t = [\mathbf{x}]^t$. Otherwise, the center with the highest compatibility is updated to accommodate input data using (8). The arousal mechanism (9) becomes part of the learning process by converting G_i^t of (8) in an effective learning rate:

$$G_i^t = \alpha(\rho_i^t)^{1-a_i^t}. \quad (10)$$

The clustering process using participatory learning also verifies if redundant clusters are formed by (8). A cluster i is redundant if its similarity with any other cluster h, $\rho_{i,h}^t$, is greater than or equal to a threshold $\lambda \in [0,1]$. The similarity between cluster centers i and h is found based on the distance between their centers $dH([\mathbf{v}_i]^t, [\mathbf{v}_h]^t)$, that is:

$$\rho_{i,h}^t = 1 - \frac{1}{n}\sum_{j=1}^{n}\left(\max\left\{|v_{i,j}^{L,t} - v_{h,j}^{L,t}|, |v_{i,j}^{U,t} - v_{h,j}^{U,t}|\right\}\right). \quad (11)$$

If clusters i and h are declared redundant, then they are replaced by a cluster whose center is the average of their respective centers.

D. eFS$_I$ consequents learning

Learning of the rule consequents in eFS$_I$ has to do with the estimation of the interval-valued parameters $[\beta_{i,0}], [\beta_{i,1}], \ldots, [\beta_{i,n}]$. The weighted recursive least squares algorithm (wRLS) [35], [36] can be used for interval data. However, differently from [28], which uses the lower and upper bounds of the intervals, here the center and range based-estimation are because of the evidence of its higher accuracy when modeling interval-valued data [8], [29]. Therefore, centers $\beta_{i,l}^c$ and ranges $\beta_{i,l}^r$, $i = 1, \ldots, c$ and $l = 0, 1, \ldots, n$, are estimated separately. Recall that the interval-valued parameter $[\beta]$ can be represented as $\beta^L = \beta^c - \beta^r$ and $\beta^U = \beta^c + \beta^r$.

The output of eFS$_I$ (4) can be expressed in terms of the input interval data center and range individually, namely, $y^{\{c,r\}} = \Lambda^T\Theta$ where $\Lambda = \left[\lambda_1\mathbf{x}_e^T, \lambda_2\mathbf{x}_e^T, \ldots, \lambda_c\mathbf{x}_e^T\right]^T$ is the fuzzily weighted input data, $\mathbf{x}_e = \left[1, x_1^{\{c,r\}}, x_2^{\{c,r\}}, \ldots, x_n^{\{c,r\}}\right]^T$ is the extended input, $\Theta = \left[\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_c^T\right]^T$, and $\boldsymbol{\beta}_i = [\beta_{i,0}^{\{c,r\}}, \beta_{i,1}^{\{c,r\}}, \ldots, \beta_{i,n}^{\{c,r\}}]^T$. Superscript c denotes the center c, and r the range (radii) r of the interval. The locally optimal error criterion wRLS is:

$$\min E_i^t = \min\sum_{k=1}^{t}\lambda_i\left(y^{\{c,r\},k} - (\mathbf{x}_e^k)^T\boldsymbol{\beta}_i^k\right)^2, \quad (12)$$

whose solution can be expressed recursively by [35]:

$$\boldsymbol{\beta}_i^{t+1} = \boldsymbol{\beta}_i^t + \Sigma_i^t\mathbf{x}_e^t\lambda_i^t\left(y^{\{c,r\},t}(\mathbf{x}_e^t)^T\boldsymbol{\beta}_i^t\right),\ \boldsymbol{\beta}_i^0 = 0, \quad (13)$$

$$\Sigma_i^{t+1} = \Sigma_i^t - \frac{\lambda_i^t\Sigma_i^t\mathbf{x}_e^t(\mathbf{x}_e^t)^T\Sigma_i^t}{1 + \lambda_i^t(\mathbf{x}_e^t)^T\Sigma_i^t\mathbf{x}_e^t},\ \Sigma_i^0 = \Omega I, \quad (14)$$

where Ω is a large number (typically $\Omega = 1,000$), and Σ is the dispersion matrix.

Expressions (13)-(14) individually estimate the center $\beta_{i,l}^c$ and range $\beta_{i,l}^r$ of the interval-valued parameters $[\beta_{i,l}] = [\beta_{i,l}^L, \beta_{i,l}^U]$, $l = 0, 1, \ldots, n$, with $\beta^L = \beta^c - \beta^r$ and $\beta^U = \beta^c + \beta^r$.

The output of the $i-th$ fuzzy rule (3) at $t+1$ is:

$$[y_i]^{t+1} = [\beta_{i,0}]^t + [\beta_{i,1}]^t[x_1]^t + \ldots + [\beta_{i,n}]^t[x_n]^t. \quad (15)$$

$i = 0, 1, \ldots, c$, and the model output $[y]^{t+1}$ at $t+1$ is the weighted average of the outputs $[y_i]^{t+1}$ as in (4).

E. eFS$_I$ algorithm

The evolving fuzzy modeling approach for interval-valued data eFS$_I$ updates model structure and functionality recursively, being memory efficient when processing stream data. This is very appealing when real-time applications are of interest, as in performing risk management through volatility forecasting, a task considered in this work. The steps of eFS$_I$ modeling are summarized in the following algorithm:

eFS$_I$ algorithm

choose parameters α, β, τ, and λ.
set $a_1^1 = 0$, $\Sigma_i^0 = \Omega I$, $[\beta_0^1] = [0, 0]$.
start cluster structure: $c = 1$, $[\mathbf{v}_1]^1 = [\mathbf{x}]^1$.
for $t = 2, 3, \ldots$ **do**
  read $[\mathbf{x}]^t$
  for $i = 1, \ldots, c$
    compute $\rho_i^t = 1 - dH([x], [y])$
    update $a_i^t = a_i^{t-1} + \beta(1 - \rho_i^t - a_i^{t-1})$
  end for
  if $a_i^t \geq \tau$ for $i = 1, \ldots, c$
    create a new cluster: $c = c + 1$
    $[\mathbf{v}_c]^t = [\mathbf{x}]^t$
    reset $a_i^t = 0$
    else
      find cluster s, $s = \max_{j=1,\ldots,c} \{\rho_j^t\}$
      update $[\mathbf{v}_s]^t = [\mathbf{v}_s]^{t-1} + G_s^t([\mathbf{x}]^t - [\mathbf{v}_s]^{t-1})$,
  end if
  for $i = 1, \ldots, c-1$ and $h = i+1, \ldots, c$
    compute $\rho_{i,h}^t = 1 - \frac{1}{n} dH([\mathbf{v}_i]^t, [\mathbf{v}_h]^t)$
    if $\rho_{i,h}^t \geq \lambda$
      redefine $[\mathbf{v}_i]^t$ using i and h, delete $[\mathbf{v}_h]^t$
      $c = c - 1$
    end if
  end for
  compute consequent parameters with wRLS
  compute model output $[y]^{t+1}$
end for

III. Computational experiments

This paper evaluates the predictability of Bitcoin high and low prices using the evolving fuzzy modeling framework designed for interval-valued streams. Out-of-sample performance of eFS$_I$ is compared against Autoregressive Integrated Moving Average (ARIMA) and Multilayer Neural Networks (MLP) models for one-step-ahead predictions. In addition, the forecasts are used to compute a range-based volatility model. Results are evaluated in different out-of-sample data sets, comprising distinct crypto market dynamics to promote robust findings.

A. BitCoin data

Data is the daily high and low prices of BitCoin (BTC) that assemble the corresponding interval-valued time series (ITS). The sample covers the period from January 1, 2018 to December 31, 2022, a total of 1,826 observations[2]. The data was divided into three in-sample and out-of-sample sets. The in-sample data sets contain data of two consecutive years: 2018-2019; 2019-2020; and 2020-2021. Out-of-sample data sets contain a whole year of data, i.e. the corresponding years of 2020, 2021 and 2022. All evaluations reported here concern the out-of-sample data, as the in-sample data were used for modeling and validation. This approach is not necessary for eFS$_I$, as it continuously processes data as a stream, differently from the counterparts ARIMA and MLP. These methods produce interval forecasts from the individual lowest and highest daily BTC data, computed separately. On the other hand, eFS$_I$ do processes ITS naturally, considering the interval structure of the data.

[2]The data are available at https://coinmarketcap.com/. Sample begins in 2017 with the higher liquidity of BTC, and ends in 2022 to cope with different dynamics, including the impact of COVID-19 pandemic, and the Ukrainian-Russian war.

B. Forecasting evaluation

The performance of the models performance are evaluated using the mean absolute percentage error (MAPE) and the root mean squared error (RMSE), respectively:

$$\text{MAPE}^B = \frac{100}{N} \sum_{t=1}^{N} \frac{|y^{B,t} - \hat{y}^{B,t}|}{y^{B,t}}, \tag{16}$$

$$\text{RMSE}^B = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y^{B,t} - \hat{y}^{B,t})^2} \tag{17}$$

where the superscript B denotes either the lower bound L, or the upper bound U of the interval BTC prices, $[y]^t = [y^{L,t}, y^{U,t}]$ is the actual price interval, $[\hat{y}]^t = [\hat{y}^{L,t}, \hat{y}^{U,t}]$ is the forecast price interval at t, and N is the size of the out-of-sample data set.

C. Range-based volatility models and VaR

This section develops a range-based volatility measure using the high and low prices forecasts. This approach enables us to account for intraday variability through the use of highs and lows. Here, the range-based volatility estimator suggested in [20] is considered. The variance $\sigma_{P,t}^2$ is computed as:

$$\sigma_{P,t}^2 = \frac{1}{4ln(2)} \left(ln \left(\frac{y^{U,t}}{y^{L,t}} \right) \right)^2, \tag{18}$$

where $y^{U,t}$ and $y^{L,t}$ are the upper and lower bound of interval-value price series $[y]^t$, and $\sigma_{P,t}^2$ is the variance at t. Notice that $y^{U,t}$ and $y^{L,t}$ are the high and low price at time (day) t. This volatility estimator produces a more informative variability measure as it accounts for intraday price range, differently from close-to-close variance metrics.

The range-based volatility measure of [20], $\sigma_{P,t}^2$, is computed using ARIMA, MLP and eFS$_I$ forecasts. Accuracy is evaluated using the mean absolute percentage error (MAPE) and the root mean squared error (RMSE), respectively:

$$\text{MAPE} = \frac{100}{N} \sum_{t=1}^{N} \frac{|\sigma_t^2 - \hat{\sigma}_t^2|}{\sigma_t^2}, \tag{19}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (\sigma_t^2 - \hat{\sigma}_t^2)^2} \tag{20}$$

where $\hat{\sigma}_t^2$ is the volatility range calculated by [20] estimator using the forecasts, and σ_t^2 is the actual variance at t, i.e. calculated uisng actual high and low prices.

D. Results

The eFS$_I$ forecasting ability is compared with those of ARIMA and MLP models for BitCoin high and low prices prediction. It should be noted that ARIMA and MLP are univariate time series techniques, i.e. they use a model for each interval bound (lower and upper). Contrary, eFS$_I$ is an interval-valued method and produces interval (with low and high price bounds) forecasts. In this paper, eFS$_I$ uses as input the lagged

values of the interval time series $[y]^t = [y^{L,t}, y^{U,t}]$. One-step-ahead forecasts, $[y]^{t+1} = [y^{L,t+1}, y^{U,t+1}]$, are produced as $[\hat{y}]^{t+1} = f_{eFS_I}\left([y]^t, [y]^{t-1}, [y]^{t-2}, \ldots, [y]^{t-l}\right)$, where l is the number of lagged values used in the input, and $f_{eFS_I}(\cdot)$ encodes the eFS$_I$ modeling framework.

Models setting and training were performed concerning the three in-sample data sets: 2018-2019, 2019-2020, and 2020-2021. Simulations where performed to select eFS$_I$ control parameters associated with the lowest values of MAPE and RMSE, which are: $l = 6$, $\beta = 0.11$, $\tau = 0.58$, $\alpha = 0.035$ and $\lambda = 0.85$. The algorithm eFS$_I$ was implemented in MATLAB, and ARIMA and MLP in R and its *forecast* package [3].

Models accuracy performance in terms of RMSE and MAPE of low and high prices of BitCoin forecasting for the three out-of-sample sets, i.e. using data from the years of 2020, 2021 and 2022 are summarized in Table I The best values are highlighted in bold.

TABLE I
RMSE AND MAPE VALUES FOR ONE-STEP-AHEAD FORECASTING OF DAILY BITCOIN LOW AND HIGH PRICES FOR EACH OUT-OF-SAMPLE DATA SET.

Metric	ARIMA	MLP	eFS$_I$
Panel A: 2020			
RMSE^L	369.9284	372.8852	**344.1982**
RMSE^U	377.4879	391.7221	**373.1011**
MAPE^L	2.1074	2.2588	**1.9884**
MAPE^U	1.9422	2.0163	**1.8730**
Panel B: 2021			
RMSE^L	2085.6976	2258.1114	**1855.0777**
RMSE^U	1664.7557	1889.6623	**1557.0080**
MAPE^L	3.3873	3.4099	**3.0041**
MAPE^U	2.5408	2.5559	**2.1384**
Panel C: 2022			
RMSE^L	979.0947	1001.8854	**955.8333**
RMSE^U	923.0636	933.1030	**908.4528**
MAPE^L	2.1694	2.2555	**1.9906**
MAPE^U	2.0298	2.1241	**1.9054**

The RMSE and MAPE values for both BitCoin low and high prices forecasting indicate that the fuzzy model outperforms its competitors. ARIMA achieves the second best performance, superior than MLP. Concerning the out-of-sample periods, results of Table I clearly indicate a predictability pattern for Bitcoin. Error measures are considerably higher for 2021, a period of high volatility dynamics for the digital currency. In periods of bear markets, when BTC prices suffered a significant depreciation (2022), error measures are also high. Models were generally more accurate during 2020, indicating a higher predictability of BItcoin highs and lows in bull markets (price appreciation).

Nevertheless, literature shows that forecasting accuracy cannot be measured in terms of a criteria that depends on the magnitude of the forecasting error only [37], [38]. Thus, forecasts were used to compute range-based volatility using [20] estimator. Table II gives the RMSE and MAPE values for volatility forecasting. The results indicate that, again, the fuzzy model produces more accurate volatility forecasts. However, it is interesting to note that besides the out-of-sample set dynamics, the level of accuracy is quite similar for all models, with exception of 2022, the 'winter-crypto' period, where Bitcoin prices suffered a significant, revealing some sort of asymmetry in accuracy for bear and bull markets.

TABLE II
RMSE AND MAPE VALUES FOR ONE-STEP-AHEAD RANGE-VOLATILITY FORECASTING OF BITCOIN FOR EACH OUT-OF-SAMPLE DATA SET.

Metric	ARIMA	MLP	eFS$_I$
Panel A: 2020			
RMSE	195.5849	199.1007	**193.1224**
MAPE	0.0062	0.0066	**0.0059**
Panel B: 2021			
RMSE	183.3777	194.9138	**172.1982**
MAPE	0.0051	0.0068	**0.0045**
Panel C: 2022			
RMSE	232.2073	334.8451	**228.0441**
MAPE	0.0018	0.0020	**0.0016**

Figure 1 illustrates the eFS$_I$ in modeling Bitcoin range-based volatility during the years of 2020, 20221 and 2022. It shows the high capability of the eFS$_I$ model in forecasting the cryptocurrency variance over time, under different market dynamics.

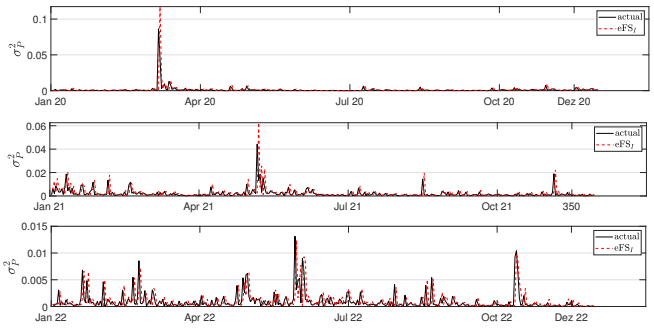

Fig. 1. Actual and eFS$_I$ forecasts for Bitcoin range-based volatility for different out-of-sample sets.

IV. CONCLUSION

An evolving fuzzy model and modeling method for interval-valued data is suggested in this paper. The method is based on a fuzzy rule-based system with affine consequents designed to process interval-valued data. Computational experiments concerning Bitcoin high and low prices one-step-ahead forecasting were done. The evolving fuzzy model accuracy was compared with statistical and machine learning alternatives using RMSE and MAPE. The results produced by the evolving fuzzy model were also evaluated in economic terms, using high and low prices forecasts to compute range-based volatility, a natural application of modeling interval time series in finance.

[3]ARIMA structures were selected based on information criteria such as the Bayesian Information Criteria (BIC). MLP topology was comprised by a single hidden layer with a different number of neurons for each in-sample set, whose selection was based on simulations in order to find the lowest values of the error metrics. All details concerning competitors setting are available under request.

Overall, the result suggests that the evolving fuzzy interval model is a promising candidate for volatility prediction. Future research shall consider volatility forecasts in risk management applications, such as in Value-at-Risk, a common risk measure in financial markets.

REFERENCES

[1] E. Diday and M. Noirhomme-Fraiture, *Symbolic data analysis and the SODAS software*. England: John Wiley & Sons, 2008.

[2] M. Alvo, *Symbolic Data Analysis*. Cham: Springer International Publishing, 2022, pp. 247–274.

[3] Y. Sun, X. Zhang, A. T. Wan, and S. Wang, "Model averaging for interval-valued data," *European Journal of Operational Research*, vol. 301, no. 2, pp. 772–784, 2022.

[4] P. Brito, "Symbolic data analysis: another look at the interaction of data mining and statistics," *WIREs Data Mining and Knowledge Discovery*, vol. 4, no. 4, pp. 281–295, 2014.

[5] E. A. Lima Neto, G. M. Cordeiro, and F. de A.T. de Carvalho, "Bivariate symbolic regression models for interval-valued variables," *Journal of Statistical Computation and Simulation*, vol. 81, no. 11, pp. 1727–1744, 2011.

[6] T. Xiong, C. Li, and Y. Bao, "Interval-valued time series forecasting using a novel hybrid Holt and MSVR model," *Economic Modelling*, vol. 60, pp. 11–23, 2017.

[7] G. González-Rivera and J. Arroyo, "Time series modeling of histogram-valued data: The daily histogram time series of s&p500 intradaily returns," *International Journal of Forecasting*, vol. 28, no. 1, pp. 20–33, 2012.

[8] J. Arroyo, R. Espínola, and C. Maté, "Different approaches to forecast interval time series: A comparison in finance," *Computational Economics*, vol. 27, no. 2, pp. 169–191, 2011.

[9] G. González-Rivera and W. Lin, "Constrained regression for interval-valued data," *Journal of Business & Economic Statistics*, vol. 31, no. 4, pp. 473–490, 2013.

[10] J. Wang, Y. Wang, Z. Li, H. Li, and H. Yang, "Design of a combined system based on multi-objective optimization for point and interval forecasting of air pollution," *Expert Systems with Applications*, vol. 191, p. 116345, 2022.

[11] Z. Yang, D. K. Lin, and A. Zhang, "Interval-valued data prediction via regularized artificial neural network," *Neurocomputing*, vol. 331, pp. 336–345, 2019.

[12] Z. Song, W. Feng, and W. Liu, "Interval prediction of short-term traffic speed with limited data input: Application of fuzzy-grey combined prediction model," *Expert Systems with Applications*, vol. 187, p. 115878, 2022.

[13] W. Lin and G. González-Rivera, "Interval-valued time series models: Estimation based on order statistics exploring the agriculture marketing service data," *Computational Statistics & Data Analysis*, vol. 100, pp. 694–711, 2016.

[14] E. Diday, *Explanatory Tools for Machine Learning in the Symbolic Data Analysis Framework*. John Wiley & Sons, Ltd, 2020, ch. 1, pp. 1–30.

[15] C. Cappelli, R. Cerqueti, P. D'Urso, and F. Di Iorio, "Multiple breaks detection in financial interval-valued time series," *Expert Systems with Applications*, vol. 164, p. 113775, 2021.

[16] T. T. Buansing, A. Golan, and A. Ullah, "An information-theoretic approach for forecasting interval-valued sp500 daily returns," *International Journal of Forecasting*, vol. 36, no. 3, pp. 800–813, 2020.

[17] Y.-L. Cheung, Y.-W. Cheung, and A. T. K. Wan, "A high–low model of daily stock price ranges," *Journal of Forecasting*, vol. 28, no. 2, pp. 103–119, 2009.

[18] S. Degiannakis and C. Floros, "Modeling cac40 volatility using ultra-high frequency data," *Research in International Business and Finance*, vol. 28, pp. 68–81, 2013.

[19] S. Alizadeh, M. W. Brandt, and F. X. Diebold, "Range-based estimation of stochastic volatility," *The Journal of Finance*, vol. 57, no. 3, pp. 1047–1091, 2002.

[20] M. Parkinson, "The extreme value method for estimating the variance of the rate of return," *The Journal of Business*, vol. 53, no. 1, pp. 61–65, 1980.

[21] M. W. Brandt and F. X. Diebold, "A no-arbitrage approach to range-based estimation of return covariances and correlations," *The Journal of Business*, vol. 79, no. 1, pp. 61–74, 2006.

[22] L. Yarovaya, J. Brzeszczyński, and C. K. M. Lau, "Volatility spillovers across stock index futures in asian markets: Evidence from range volatility estimators," *Finance Research Letters*, vol. 17, pp. 158–166, 2016.

[23] A.-C. Díaz-Mendoza and A. Pardo, "Holidays, weekends and range-based volatility," *The North American Journal of Economics and Finance*, vol. 52, p. 101124, 2020.

[24] M. Caporin and G. G. Velo, "Realized range volatility forecasting: Dynamic features and predictive variables," *International Review of Economics & Finance*, vol. 40, pp. 98–112, 2015.

[25] F. Ma, Y. Zhang, D. Huang, and X. Lai, "Forecasting oil futures price volatility: New evidence from realized range-based volatility," *Energy Economics*, vol. 75, pp. 400–409, 2018.

[26] Y. Xu, D. Huang, F. Ma, and G. Qiao, "Liquidity and realized range-based volatility forecasting: Evidence from china," *Physica A: Statistical Mechanics and its Applications*, vol. 525, pp. 1102–1113, 2019.

[27] W. Xu, J. Wang, F. Ma, and X. Lu, "Forecast the realized range-based volatility: The role of investor sentiment and regime switching," *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121422, 2019.

[28] L. Maciel, R. Ballini, and F. Gomide, "Adaptive fuzzy modeling of interval-valued stream data and application in cryptocurrencies prediction," *Neural Computing and Applications*, vol. https://doi.org/10.1007/s00521-021-06263-5, 2021.

[29] P. Rodrigues and N. Salish, "Modeling and forecasting interval time series with threshold models," *Advances in Data Analysis and Classification*, vol. 9, no. 1, pp. 41–57, 2015.

[30] R. Moore, R. Kearfott, and M. Cloud, *Introduction to interval analysis*. Philadelphia: SIAM Press, 2009.

[31] D. Huttenlocher, G. Klanderman, and W. Rucklidge, "Comparing images using the hausdorff distance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850–863, 1993.

[32] B. Li, Y. Shen, and B. Li, "A new algorithm for computing the minimum hausdorff distance between two point sets on a line under translation," *Information Processing Letters*, vol. 106, no. 2, pp. 52–58, 2008.

[33] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems Man and Cybernetics*, vol. SMC-15, no. 1, pp. 116–132, 1985.

[34] L. Maciel, R. Ballini, F. Gomide, and R. R. Yager, "Participatory learning fuzzy clustering for interval-valued data," in *Proceedings of the 16th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2016)*, Eindhoven, The Netherlands, 2016, pp. 1–8.

[35] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[36] P. Angelov, "Evolving Takagi-Sugeno fuzzy systems from data streams (eTS+)," in *Evolving intelligent systems: Methodology and applications*, P. Angelov, D. Filev, and N. Kasabov, Eds. Hoboken, NJ, USA: Wiley & IEEE Press, 2010, pp. 21–50.

[37] I. Moosa and K. Burns, "The unbeatable random walk in exchange rate forecasting: Reality or myth?" *Journal of Macroeconomics*, vol. 40, pp. 69–81, 2014.

[38] K. Burns and I. Moosa, "Enhancing the forecasting power of exchange rate models by introducing nonlinearity: Does it work?" *Economic Modelling*, vol. 50, pp. 27–39, 2015.