On the Links Between Forecasting Performance and Statistical Features of Time Series Applied to the Cash Flow of Self-Employed Workers*

Luis Palomero[1,2][0000−0002−9144−1017], Vicente García[3][0000−0003−2820−2918], and J. Salvador Sánchez[2][0000−0003−1053−4658]

[1] Declarando Asesores 3.0 S.L., Plaza Tetuán 38 3b 12001 Castelló de la Plana, Spain
[2] Institute of New Imaging Technologies, Dept. Llenguatges i Sistemes Informàtics, Universitat Jaume I, Av. Sos Baynat s/n, 12071 Castelló de la Plana, Spain
{al063893,sanchez}@uji.es
[3] División Multidisciplinaria en Ciudad Universitaria, Universidad Autónoma de Ciudad Juárez, Av. José de Jesús Macías, 32000, Juárez, México
vicente.jimenez@uacj.mx

Abstract. Proper cash flow forecasting is a complex task that can be done by modelling the cash flow data as a time series. Although parametric methods have been widely used to accomplish this task, they require some assumptions about the data that are difficult to hold. A well-founded alternative is the use of fuzzy inference systems, which have proven to be competitive in many practical problems. This paper presents a statistical study that compares the performance of fuzzy inference forecasting systems with that of a traditional parametric approach, in a cash flow forecasting problem based on the weekly income and expense data of 340 self-employed workers over a period of 338 weeks with four different time horizons (1, 4, 9 and 13 weeks). We also check for significant links between several statistical characteristics and observed performance, to determine which features might most affect the quality of the predictions. After finding that kurtosis is the most correlated feature, a more detailed exploration is performed on it.

Keywords: Cash flow · Self-employed workers · Time series · Forecasting · Fuzzy inference systems · Statistical features · Kurtosis

1 Introduction

Cash flow refers to the overall inflows and outflows of cash or cash equivalents from a company over a given time period due to its operating, investing, and financial activity. Cash flow analysis helps to understand the situation of a business and anticipate possible scenarios. In fact, the use of cash flow analysis as a key factor in risk assessment has become widespread [1]. Econometric models for cash flow analysis are based on a detailed study of a company's financial

* Supported by Universitat Jaume I.

statements. These studies are usually too complex for companies with few resources, as is the case of small and medium-sized enterprises, and especially for the approximately 26.9 million self-employed workers in the European Union, as of 2021 [2].

Another approach to predicting cash flow is to model it as a time series forecasting problem (i.e. predicting future data using knowledge of the past observations). Time series forecasting has proven to be an effective tool for decision-making processes in many economic and financial applications [3–5]. A time series Z of size n can be formulated as an ordered sequence of observations $Z = (z_1, z_2, \ldots, z_n)$ distributed over time, where $z_t \in \mathbb{R}$ denotes the value for the t-th time period. One of the simplest models in time series forecasting is the random walk (RW) model [6], where the prediction of the next value z_t is based on the observation at the previous time step (z_{t-1}) plus a random error term e_t at time period t. Other models are based on inferring a parametric distribution that might fit the data. For instance, a well-established procedure is the autoregressive integrated moving average (ARIMA) method [7], which is often used as a baseline. However, although parametric models are popular, they require some data assumptions such as normality, stationarity and lack of correlation that are often violated [8].

An alternative to parametric methods for time series forecasting is the computational approach, which is based on the extensive use of various machine learning algorithms such as regression trees, neural networks or support vector machines [9, 10]. These methods do not require modelling the data under any statistical assumption and offer similar or even better performance than traditional probabilistic models [11]. However, their main drawbacks are their conceptual complexity, the hyper-parametrization of the model and high computational training costs. Another research line focuses on the nature of the data itself through the concept of fuzzy logic [12] applied to time series forecasting. Oancea *et al.* [13] suggested that fuzzy-based systems can be a plausible alternative to parametric methods in some scenarios, although the limits of these scenarios are not well-defined.

To explore these limits, in this paper we compare the forecast performance of the commonly used ARIMA model and some fuzzy-based methods. After this general comparison, we also analyse some statistical features included in the paper by Salas-Molina *et al.* [8] together with four time series autocorrelation indices to verify any association between them and the performance of the models. Both studies are carried out over time series based on the weekly cash flow data of 340 Spanish self-employed workers in a period of 338 weeks. The models are first compared using the unscaled mean bounded relative absolute error (UMBRAE) [14], the values of which are then used with the statistical features of the time series to search for possible correlations. Second, we explore the differences of UMBRAE with the correlation indices to identify changes, or even degradation (inferred by comparing the results with RW), in forecast performance.

Henceforward, the rest of the paper is organized as follows: Section 2 introduces the forecasting models included in this study. Section 3 presents the data and describes the experimental set-up. The results are discussed in Section 4. Finally, Section 5 summarizes the main findings and limitations, and suggests possible avenues for future research.

2 The Forecasting Models

In the ARIMA parametric model, it is necessary to define a set of parameters (p, d, q), which refer to the number of autoregressive coefficients, the differences and the moving average coefficients. These three parameters can be calculated using the Box-Jenkins method [7], an iterative procedure based on three steps: identification of a candidate model, estimation of its parameters, and evaluation of the candidate model.

The remaining methods in this study are fuzzy inference systems (FIS), which are based on fuzzy logic [12]. This logic assumes the vagueness of data instead of exact information represented by crisp data. The FIS algorithmic process consists of four steps: (i) fuzzification, (ii) definition of logical relations, (iii) making the appropriate inference, and (iv) defuzzification. The fuzzification process transforms the crisp input data into fuzzy sets according to the data structure and the fuzzification rules. Next, the association between the inputs and the outputs is described in form of fuzzy logical relations (FLRs), which are IF-THEN rules. The IF part is the antecedent and considers the input data, while the THEN part or consequent describes the output values found when some input occurs. Once the rules are generated, the new input data is processed in the third step according to the FLRs and then the inferences are made. Finally, the defuzzification process consists of the inverse transformation of the inferred fuzzy values into crisp data and the generation of the final forecast.

This study considers two major approaches to how the FLRs are defined: Mamdani [15, 16] and Takagi-Sugeno-Kang (TSK) [17, 18]. In both approaches, the antecedent part of the FLRs is determined by fuzzy sets, with the definition of the consequent part being the main difference between Mamdani and TSK. While in Mamdani-based models it is defined by fuzzy sets, in TSK-based models the THEN part is defined as a polynomial combination of the crisp input variables. Consequently, the FLRs rules generated under the Mamdani approach are more interpretable and understandable, while the TSK-based rules allow more accurate forecasts but are more difficult to understand.

Two Mamdani-type models have been used in this study: the Wang-Mendel (WM) model [19] and the hybrid neural fuzzy inference system (HYFIS) [20]. On the other hand, the adaptive-network-based fuzzy inference system (ANFIS) [21] and the dynamic evolving neural-fuzzy inference system (DENFIS) [22] are the TSK-based approaches included.

The WM algorithm is based on five steps: (i) The crisp input data is partitioned into equally-spaced fuzzy regions, (ii) this data is used to define the Mamdani-based FLRs, (iii) weights are assigned to the rules to avoid conflicts,

(iv) the fuzzy rules are combined to create a fuzzy rule base, and (v) the input data is mapped to the output using the defined FLRs and finally defuzzified. It is a simple and fast one-pass construction procedure, although this simplicity can be a drawback in accuracy when trying to capture complex data interactions.

The WM method is the basis of HYFIS. The idea behind this algorithm is to generate a fuzzy-based inference system where the FLRs are optimized using a neural network. A WM model is first generated and then a five-layer back-propagation network inspired by the FIS architecture is used to fit the parameters of the fuzzy model using the gradient descent method. In the first layer, the input data is fed into the network. The second layer calculates the degrees of membership using Gaussian membership functions. The third layer generates the first antecedent part of the FLR, whereas the consequent part is computed in the fourth layer. Finally, the crisp output signal is calculated in the fifth layer.

The TSK-based ANFIS model is also based on a five-layer network architecture. The fuzzification process is done on the first layer. The second layer implements the inference stage using a *T-norm* operator, where each node in this layer represents the fire strength of a rule. Then all the fire strengths are normalized in the third layer and used in the fourth to calculate the consequent part of the FLRs. The final layer aggregates the results of the previous layer forming the overall output. Each epoch in the training process comprises a forward and a backward pass. The functional signals go forward to the fourth layer in the forward pass, and the least-squares estimate identifies the consequent parameters. The error rates are propagated backwards in the backward step, where the premise parameters are updated using a gradient descent method.

Finally, the DENFIS model rests upon the idea that, depending on the position of the input vector in the input space, a FIS can be dynamically formed based on the fuzzy rules created during the previous learning process. These dynamic models are built by implementing an evolving clustering algorithm to automatically determine both the input data partition and the consequent part of the FLRs. The number of fuzzy sets used to partition the data in the fuzzification process is defined according to a threshold parameter (Dthr). This method can be executed both online and offline, the main difference being how the learning process is performed. In the online DENFIS model, the rules are created and updated simultaneously along with the input space partition using the clustering algorithm. Each time new data is fed into the system, a new iteration of the clustering algorithm is run with the established clusters and FLRs, which are updated accordingly. In the offline version, all these dynamic processes are executed once using a tailored version of the clustering algorithm.

3 Data and Experimental Set-up

Declarando, a Spanish online accounting consultancy specialized in self-employed workers, provided the data for this study. The information on the profit of each self-employed worker is used as the basis for the construction of each time series.

Here, profit consists of the difference between accounting incomes and expenses. The reason for using profit data instead of direct cash flow data are: (i) these data are validated by the Spanish Treasury Agency as they are used to calculate taxes, (ii) the accounting of self-employed companies is relatively simple and similar to cash flow data and, (iii) direct cash flow data is often unreliable as it tends to mix personal and professional information, which can cause distortions. We collected weekly information on income and expenses from 340 self-employed workers, where all records were gathered in a range of 338 weeks between the first week of 2016 and the 26th week of 2022. The weeks without income and without expenses were considered inactive and imputed with 0. The statistical indices used in this study are the length of the time series, percentage of weeks without income and expenses (ratio of zeros), mean, median, standard deviation, kurtosis, skewness, and four autocorrelation indices, from the second to the fifth.

We carried out an independent forecasting process by time series, predictive method and four different time horizons ($h = \{1, 4, 9, 13\}$). Each experiment consisted of two parts: the parametrization of the model (training/validation phase) and the forecast itself (test). The time series were divided into training, validation and test groups according to the number of weeks (horizons, h) to forecast. The parameters of the forecasting models are given in Table 1.

Table 1: Parameters of the models (l_Z denotes the number of elements in a time series Z).

Method	**Parameter**	**Variation range** (Initial: Final; Step)
ARIMA	p (AR order)	0: $\sqrt{log_2(l_Z)}$; 2
	d (degree of differencing)	0: 2; 1
	q (MA order)	0: $\sqrt{log_2(l_Z)}$; 2
WM; HYFIS; ANFIS	n (number of linguistic variables)	7: $integer(\frac{l_Z}{2} - 1)$; 6
	Number of lags	2: 4; 1
DENFIS	Dthr (dynamic threshold for clustering)	0.05: 5; 0.05
	Number of lags	2: 4; 1

Except for the DENFIS model, all implementations of FIS require prior normalization of the data in the range $[0, 1]$. Therefore, we normalized all input data using the min-max method as indicated in Eq. 1. The forecasts were then denormalized using Eq. 2 and compared with the true test values. The root mean square error (RMSE) was taken as the selection criteria.

$$z' = \frac{z - min(z)}{max(z) - min(z)} \tag{1}$$

$$z = z' \times (\max(z) - \min(z)) + \min(z) \tag{2}$$

where z denotes the input data and z' the normalized data.

The UMBRAE measure [14] was used to estimate the performance of the models. This is based on a modification of the mean bounded relative absolute error and describes a forecasting error in terms of that obtained from a benchmark model. The UMBRAE metric is symmetric, scale-independent, and easy to interpret [14]. The meaning of the UMBRAE values is as follows:

- UMBRAE = 1 means that the model performs the same as the benchmark method (in our experiments, RW was the benchmark algorithm),
- UMBRAE < 1 means that the model performs approximately $(1-UMBRAE)*100\%$ better than RW, and
- UMBRAE > 1 means that the model performs $(UMBRAE-1)*100\%$ worse than RW.

The statistical significance of differences between the forecasting models was assessed using the Friedman rank sum test [23] at the 5% significance level. When the null hypothesis (i.e. all algorithms are equivalent) in the Friedman test was rejected, the pairwise post-hoc Nemenyi test was run. This test is based on the pairwise comparison of the ranked values and considering significant differences when the average of these ranks is greater than an overall critical difference value.

4 Results

This section was divided into three blocks. First, we compared the performance of all forecasting models considered in this study. In the second block, we applied Spearman's rank correlation test to verify any association between each model and some statistical characteristics of the time series. In the last block, we concentrated our experiments on exploring any possible relationship between the UMBRAE values of ARIMA and WM with kurtosis.

4.1 Performance Comparison

In this block of experiments, we used the Friedman test to compare the difference in the UMBRAE values obtained by each model. Here we also included the results of RW as this was the benchmarking algorithm used to calculate UMBRAE. Note that the UMBRAE values for RW are always equal to 1 due to the definition of this metric (see Section 3).

In all cases, the Friedman test revealed high statistical evidence against the null hypothesis of no significant differences between the models at a significance level of 0.05 (p-value $\ll 1 \times 10^{-10}$). Therefore, we applied the pairwise post-hoc Nemenyi test to compare each pair of models for each horizon and the results were visually represented with a critical distance diagram in Figure 1. The critical distance (CD) is shown above the graph, the top line represents the axis on which the average ranks are plotted from the lowest (best) on the left side to the highest (worst) on the right side of the diagram. The thick line connecting models reflects non-significant differences.

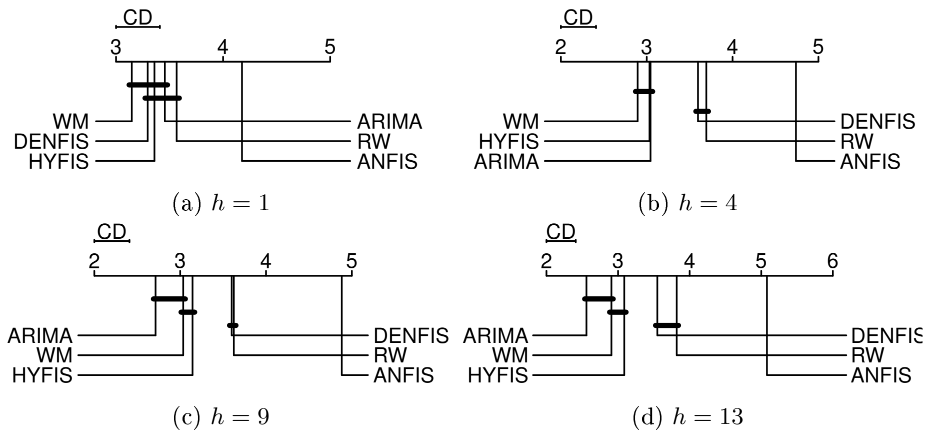


Fig. 1: Critical distance diagrams for the pairwise comparison of models.

The results of the Nemenyi test suggest that the Mamdani-based models performed similarly to ARIMA, while the TSK-based methods were consistently the worst in all cases. In fact, regarding the TSK-based models, DENFIS performed well only for $h = 1$, while ANFIS performed significantly worse than any other model (including the RW benchmarking method as well) for all horizons. When focusing on the Mamdani-based approaches, it appears that the WM results were better than those of HYFIS, but the differences were not statistically significant.

4.2 Correlation Analysis Between Performance and Statistical Features

Here we applied Spearman's rank correlation coefficient to test for any association between the performance of the forecasting models and the statistical characteristics of the time series at a significance level of 5%. Table 2 summarizes this analysis, where each value indicates how many times the UMBRAE derived from the model in the column was significantly correlated with the attribute in the row for the four forecasting time horizons h considered in the experiments. Note that standard deviation and time series length were omitted because no statistically significant correlation with model performance was found.

The results in Table 2 reveal that the statistical features of the time series most correlated with the UMBRAE values were kurtosis and autocorrelation lags. For kurtosis and autocorrelation lags 2, 3 and 4, the correlations were positive, while for autocorrelation lag 5 they were negative. Although not shown in this table, it should be noted that the correlation values were similar in all cases (i.e. all model-attribute pairs) in which the association was statistically significant.

Kurtosis (K) measures the tailedness of a distribution, that is, it represents the probability or frequency of values that are extremely high or low compared to

Table 2: Summary of Spearman's correlation analysis.

	ANFIS	**ARIMA**	**DENFIS**	**HYFIS**	**WM**	**Total**
Kurtosis	1	3	4	3	4	15
Autocorrelation (lag 4)	3	2	3	3	3	14
Autocorrelation (lag 2)	1	3	3	3	3	13
Autocorrelation (lag 3)	2	1	3	4	3	13
Autocorrelation (lag 5)	3	3	2	3	2	13
Ratio of zeroes	1	0	3	1	1	6
Median	0	1	0	1	1	3
Mean	0	1	0	1	0	2
Skewness	2	0	0	0	0	2

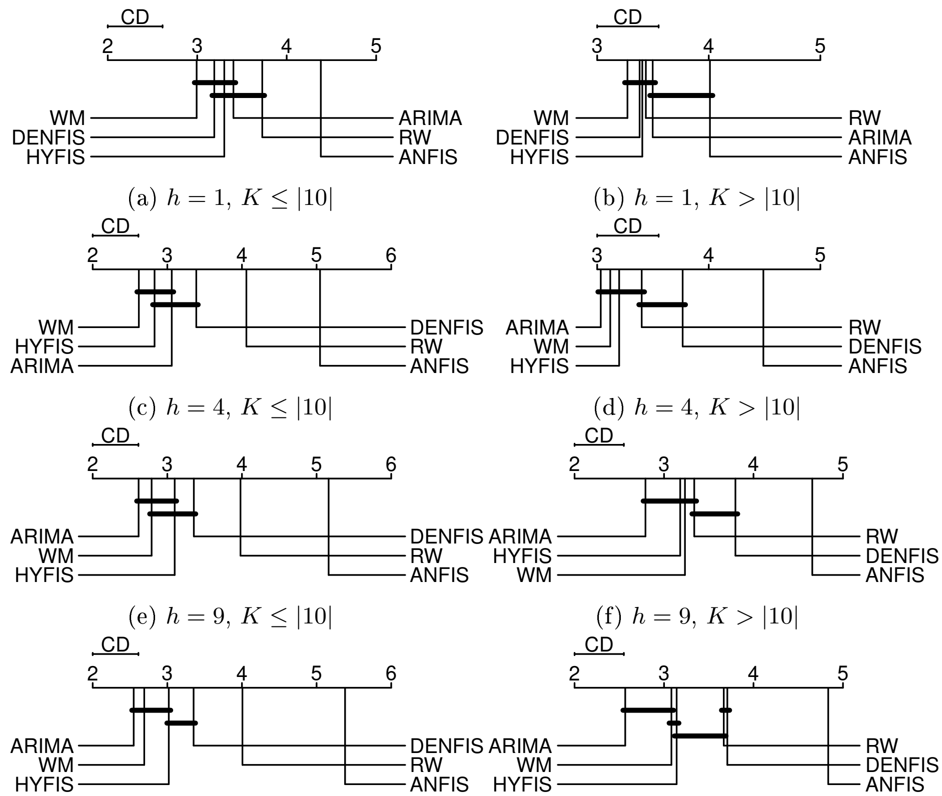
the mean. Therefore, kurtosis measures the frequency of outliers. The K values of our data were in the range $[-0.13, 218]$, with a mean and median of 23.77 and 11.79, respectively. Although there is no consensus on what an acceptable K would be, a general rule of thumb is to consider $K > |10|$ as severe kurtosis [24]. We adopted this criterion to divide the 340 experimental time series into two groups according to the severity of kurtosis: data with non-severe kurtosis ($n = 152$) and data with severe kurtosis ($n = 188$).

As in the analysis carried out in the previous section, here we applied the post-hoc Nemenyi test to compare each pair of models for each forecasting time horizon and each group of time series according to the severity of kurtosis. Figure 2 shows the eight critical distance diagrams for visualizing the test results. As can be seen, the results for the time series with non-severe kurtosis ($K \leq |10|$) were quite similar to those represented in Figure 1, that is, the Mamdani-based models performed the same as ARIMA and significantly better than the TSK-based methods and the benchmarking algorithm. In general, for the case of non-severe kurtosis (plots on the right in the figure), the best forecast models were WM, ARIMA and HYFIS, regardless of the time horizon.

For the time series with severe kurtosis ($k > 10$), the plots on the left in Figure 2 indicate that the performance of the RW algorithm was similar to that of ARIMA, WM, HYFIS and DENFIS and even better than the performance of ANFIS for the time horizons $h = 1$, $h = 4$ and $h = 9$. It is also worth noting that differences between WM and HYFIS were much smaller than in the time series with the non-severe kurtosis. However, it should be pointed out that again the best models were WM and ARIMA, as was the case with the non-severe kurtosis data.

4.3 Effect of Kurtosis on the Performance of ARIMA and WM

The third block of the experiments focused on a deeper analysis of the possible links between kurtosis and the UMBRAE values of ARIMA and WM, which correspond to the best forecasting models, as shown in Figure 1. This analysis used the Wilcoxon signed-rank sum test, running an independent comparison for each forecast time horizon. The box plots in Figure 3 illustrate the range of



(a) $h = 1,\ K \leq |10|$ (b) $h = 1,\ K > |10|$

(c) $h = 4,\ K \leq |10|$ (d) $h = 4,\ K > |10|$

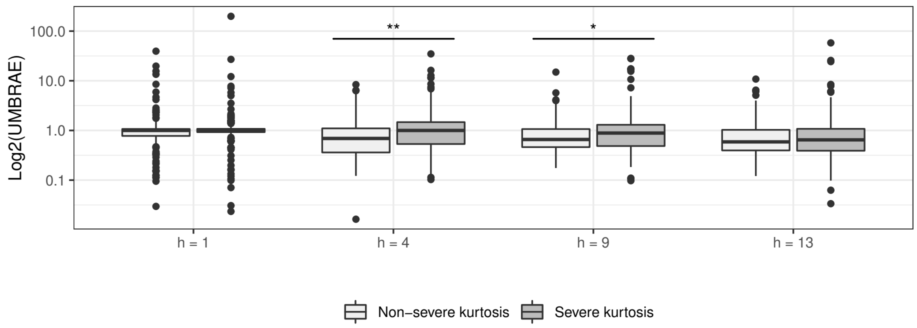
(e) $h = 9,\ K \leq |10|$ (f) $h = 9,\ K > |10|$

(g) $h = 13,\ K \leq |10|$ (h) $h = 13,\ K > |10|$

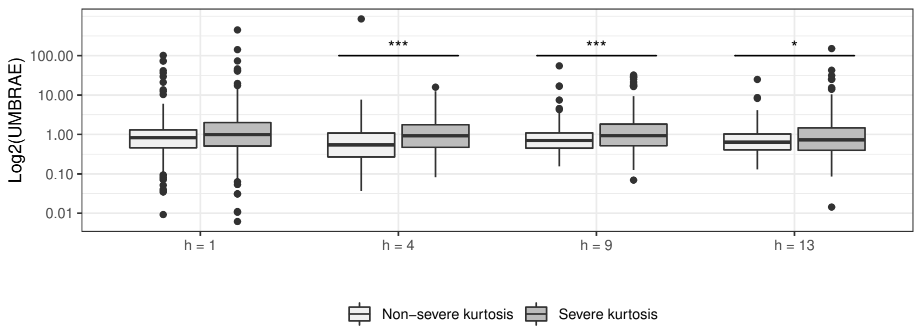
Fig. 2: Critical distance diagram for the pairwise comparison of models based on the time horizon and the severity of kurtosis.

UMBRAE values in log2 scale obtained by ARIMA and WM for each time series group based on the severity of kurtosis. Significant differences are highlighted above each box plot as follows according to the level of significance: p-value $<$ 0.05 (*), p-value $<$ 0.01 (**) and p-value $<$ 0.001 (***).

In summary, we found significant differences in both models for the time horizons $h = 4$ and $h = 9$, while no significant difference was detected for $h = 1$. In the case of $h = 13$, only significant differences were found when applying WM. On the other hand, these plots also show that the time series with severe kurtosis generally reached higher UMBRAE values than those with non-severe kurtosis. The Wilcoxon signed-rank sum test suggests that the association between kurtosis and prediction performance is stronger with WM than with ARIMA.



(a) ARIMA



(b) WM

Fig. 3: Box plot of the UMBRAE values in log2 scale based on the severity of kurtosis for each time horizon.

5 Conclusions and Future Work

This study compares the performance of four fuzzy inference systems (two Mamdani-based models and two TSK-based models) for the weekly cash flow forecasting of self-employed workers, including ARIMA and RW as baseline methods. Next, the possible associations of the forecast values in terms of UMBRAE with 11 statistical characteristics of the time series have been tested, finding significant correlations with kurtosis and autocorrelation indices. A deeper exploration of the differences between time series grouped according to the severity of kurtosis has also been carried out.

From a global point of view, Mamdani-based models have performed better than TSK-based approaches. Specifically, the WM results were better than those of the other three fuzzy inference systems in all cases. However, the analyses have revealed that in no case did any fuzzy-based models outperform the ARIMA results, either globally or according to the severity of the kurtosis. In time series with severe kurtosis, we have also identified that RW behaves similarly to the best models for forecasting time horizons $h = 4$ and $h = 9$. These results suggest a potential relationship between the kurtosis index and the complexity of the forecasts, which will require a deeper and more exhaustive study.

Despite the interesting contributions of this study, some limitations that could be addressed in future work should also be taken into account. First, the nature of the data (the weekly balance of self-employed workers) has been extracted from a specific context made up of weekly time series. Secondly, a small number of forecasting models and statistical characteristics of time series have been included in the analyses carried out. Third, the study on the relationship between statistical characteristics and forecasting methods was limited to a comparison between the kurtosis index and the UMBRAE values obtained by ARIMA or WM. Therefore, it might be useful to extend this study to other contexts, forecasting models, and further exploration of the association between forecasts and statistical features.

References

1. Dorfleitner, G., Gleißner, W.: Valuing streams of risky cashflows with risk-value models. Journal of Risk **20**(3) (2018) 1–27
2. : Number of self-employed people in the European Union (EU27) from 2002 to 2021, Statista Research Department. `https://www.statista.com/statistics/946989/self-employed-persons-in-eu/` (2023) [Online; accessed 02-May-2023].
3. Hajizadeh, E., Mahootchi, M., Esfahanipour, A., Kh, M.M.: A new NN-PSO hybrid model for forecasting Euro/Dollar exchange rate volatility. Neural Computing & Applications **31** (2019) 2063–2071
4. Nyberg, H., Pönkä, H.: International sign predictability of stock returns: The role of the United States. Economic Modelling **58** (2016) 323–338
5. Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M.: Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing **90** (2020) 106181
6. Lawler, G.F., Limic, V.: Random Walk: A Modern Introduction. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2010)
7. Box, G.E.P., Jenkins, G.M., Reinsel, G.C., Ljung, G.M.: Time Series Analysis: Forecasting and Control. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ (2015)
8. Salas-Molina, F., Juan A., R.A., Joan, S., Montserrat, G., Francisco J., M.: Empirical analysis of daily cash flow time-series and its implications for forecasting. SORT-Statistics and Operations Research Transactions **42**(1) (1) 73–98
9. Ahmed, N.K., Atiya, A.F., El-Gayar, N., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. Econometric Reviews **29**(5–6) (2010) 594–621
10. Masini, R.P., Medeiros, M.C., Mendes, E.F.: Machine learning advances for time series forecasting. Journal of Economic Surveys **37**(1) (2023) 76–111
11. Parmezan, A.R.S., Souza, V.M., Batista, G.E.: Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. Information Sciences **484** (2019) 302–337
12. Zadeh, L.: Fuzzy sets. Information and Control **8**(3) (1965) 338–353

13. Oancea, B., Pospíšil, R., Jula, M.N., Imbrișcă, C.I.: Experiments with fuzzy methods for forecasting time series as alternatives to classical methods. Mathematics **9**(19) (2021) 1–17
14. Chen, C., Twycross, J., Garibaldi, J.M.: A new accuracy measure based on bounded relative error for time series forecasting. PLoS One **12**(3) (2017) e0174202
15. Mamdani, E.: Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of the Institution of Electrical Engineers **121**(12) (1974) 1585
16. Mamdani, E., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies **7**(1) (1975) 1–13
17. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its applications to modeling and control. IEEE Transactions on Systems, Man, and Cybernetics **SMC-15**(1) (1985) 116–132
18. Sugeno, M., Kang, G.: Structure identification of fuzzy model. Fuzzy Sets and Systems **28**(1) (1988) 15–33
19. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics **22**(6) (1992) 1414–1427
20. Kim, J., Kasabov, N.: Hyfis: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems. Neural Networks **12**(9) (1999) 1301–1319
21. Jang, J.S.: Anfis: adaptive-network-based fuzzy inference system. IEEE Transactions on Systems, Man, and Cybernetics **23**(3) (1993) 665–685
22. Kasabov, N., Song, Q.: Denfis: dynamic evolving neural-fuzzy inference system and its application for time-series prediction. IEEE Transactions on Fuzzy Systems **10**(2) (2002) 144–154
23. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research **7**(1) (2006) 1–30
24. Kline, R.: Principles and practice of structural equation modeling, fourth edition. Methodology in the Social Sciences. Guilford Publications (2015)