Forecasting Performance and Information Measures. Revisiting the M-Competition [*]

ANA JESÚS LÓPEZ MENÉNDEZ [a], RIGOBERTO PÉREZ SUÁREZ [a]

[a] *Universidad de Oviedo, Facultad Economía y Empresa, Avda. del Cristo, s/n, 33006 Oviedo, España. E-mail:* anaj@uniovi.es, rigo@uniovi.es

ABSTRACT

Economic and financial time series are widely considered as one of the most challenging applications of modeling and forecasting. The increasing in forecasting availability and the controversial debate about the advantages of alternative forecasting procedures suggest the need of further research on the forecasting evaluation metrics.

In this context, this paper focuses on two information-based accuracy measures: Theil´s U Index and the Quadratic Information Accuracy Measure (QIAM), and aims to re-examine the empirical results of the M3-Competition by Makridakis and Hibon (2000), and specifically those referred to the subset of macroeconomic and financial series.

The computation of the proposed accuracy indicators leads to new rankings of forecasting techniques, showing some similarities and disagreements with the main conclusions by Makridakis & Hibon (2000), found on five error based accuracy measures. The obtained results also allow a complexity–accuracy analysis.

Keywords: Forecasting, Accuracy, M3-Competition, Theil´s U Index, Quadratic Information Accuracy Measure (QIAM).

Evaluación de Predicciones y Medidas de Información. Reexamen de la M-Competición

RESUMEN

La obtención de predicciones para series temporales económicas y financieras es una tarea de gran dificultad. En un contexto de disponibilidad creciente de predicciones y debate sobre las alternativas metodológicas para su obtención, resulta recomendable dedicar nuevos esfuerzos a las medidas utilizadas para su evaluación.

Este trabajo analiza dos indicadores de precisión basados en medidas de información: el índice U de Theil y la Medida de Información Cuadrática de Precisión (QIAM), cuya aplicación a la M-Competición de Makridakis and Hibon (2000) permite reexaminar los resultados empíricos obtenidos por estos autores para el conjunto de la base de datos y más concretamente para las series macroeconómicas y financieras.

El cálculo de las medidas propuestas proporciona un nuevo ranking de técnicas predictivas, que muestra coincidencias y diferencias con las conclusiones obtenidas por Makridakis & Hibon a partir de cinco medidas de precisión basadas en errores. Los resultados obtenidos permiten también un análisis de complejidad versus precisión.

Palabras clave: Predicción, Precisión, M-Competición, Indice de Theil, Medida de Información Cuadrática de Precisión (QIAM).

Clasificación JEL: C53, C52, C22

---

1. INTRODUCTION

Forecasting availability has significantly increased during the last decades, stressing the need of analysing the adequacy of alternative methods. One of the main empirical researches in this field is the M-Competition developed by Makridakis and Hibon, whose last edition (M3) is referred to year 2000. This investigation involves a total of 3003 time series with different characteristics (microeconomic, industry, macroeconomic, finance, demographic and others) and frequencies (yearly, quarterly, monthly and others), thus providing an excellent empirical ground.

The M3-Competition includes 24 forecasting procedures whose results are ranked according to five different accuracy measures: Symmetric Mean Absolute Percentage Error, Average Ranking, Percentage Better, Median Symmetric Absolute Percentage Error and Median Relative Absolute Error. The obtained conclusions confirm that -although the performance of the considered methods varies according to the accuracy measure being used- in general terms complex forecasting methods do not necessarily lead to more accurate forecasts than simpler ones.

Since the theory of information provides a good framework for a wide variety of fields, including the evaluation of econometric models and their applications, in this paper we propose the use of information-based measures for the evaluation of forecasting accuracy. More specifically, two information-based accuracy measures are considered: the U Index proposed by H. Theil (1966) and the Quadratic Information Accuracy Measure (QIAM) by López and Pérez (2012), which is based on the amount of information provided by forecasts.

With the aim of analysing the behaviour of both accuracy measures, an empirical application is developed on the M3-Competition results, allowing the identification of some similarities and divergences with the conclusions by Makridakis & Hibon (2000).

The paper is organised as follows: in the next section we introduce the proposed information-based accuracy measures while section three briefly describes the M3-Competition and its main findings.

The results of our empirical application are presented in section four, which summarises the accuracy rankings obtained with Theil´s U Index and the Quadratic Information Accuracy Measure for the whole M3-Competition and also for the subsets of macroeconomic and financial series.

The discussion of the obtained results is presented in section five, describing the similarities and differences between the information-based accuracy rankings and those initially provided by the M3-Competition. Finally, section six summarises the main conclusions.

2. INFORMATION MEASURES AND FORECASTING ACCURACY

From its very beginning with the seminal work "A Mathematical Theory of Communication" by A. Shannon (1948) the Information Theory has provided a suitable framework for the analysis of income inequality, industrial concentration and goodness of fit among other fields. However little attention has been paid to its applications to forecasting accuracy, with the exception of Theil´s Index.

Given a variable X with probabilities $p_1, p_2, ..., p_n; p_i > 0; \sum_{i=1}^{n} p_i = 1$, Shannon defined an Entropy Measure given by the expression $H(X) = -\sum_{i=1}^{n} p_i \, log \, p_i$, whose maximum value is achieved when all the results are equally likely $\left(p_i = \frac{1}{n}, \forall i = 1,...,n \right)$ and leads to a null result in the opposite case $\left(\exists i \, / \, p_i = 1, \, p_j = 0 \, \forall j \neq i \right)$.

Although Shannon´s Index is one of the most widely used information measures, this expression also has some limitations and therefore some alternative informational measures have been developed, including parametric expressions as in Havrda & Charvat (1967) and involving not only the probabilities but also the utilities associated to the considered variable as proposed by Pérez (1985) and Gil, Pérez & Gil (1989).

Focusing on the forecasting context, H. Theil (1966) proposed the use of the following expression: $U = \sqrt{\dfrac{\sum_t \left(\dfrac{\hat{Y}_t - Y_{t-1}}{Y_{t-1}} - \dfrac{Y_t - Y_{t-1}}{Y_{t-1}} \right)^2}{\sum_t \left(\dfrac{Y_t - Y_{t-1}}{Y_{t-1}} \right)^2}}$ where $\hat{Y}_t$ and Y_t respectively denote the forecasted and the actual value of the considered variable Y for a given period t.

Theil´s U Index reaches its lower boundary (U=0) at perfect forecasts ($\hat{Y}_t = Y_t \; \forall t$) increasing its value as forecasts become less accurate. Furthermore, it can be easily proved that Theil´s Index adopts the value of 1 for naïve no-change forecasts, since in this case $\hat{Y}_t = Y_{t-1} \; \forall t$. Therefore, although Theil´s Index does not have an upper bound, a psychological bound is often assumed:

when U is larger than 1 the corresponding forecasting method is to be rejected, as it cannot beat the simple no-change extrapolation.

Besides its interpretation, one of the main advantages of Theil´s U Index refers to its decomposability, allowing the additive disaggregation of U^2 in three terms, which are respectively related to bias, variance and covariance.

The need of improving the perceived quality of economic forecasts was stressed among others by Granger (1996), Granger & Jeon (2004) and Makridakis & Bakas (2016), and the information measures result to be a suitable framework for this goal. More specifically, the development of uncertainty measures involving utilities provides the possibility of assigning a specific utility level to each economic forecast depending on the corresponding quality.

Thus, if we focus on an economic variable Y, the quality of forecasts can be approached by computing the amount of information (or the reduction of uncertainty) achieved once these forecasts are available.

The quadratic uncertainty related to Y can then be computed as:

$$HU^2(Y) = \frac{2}{T} \sum_t \left(\frac{E(Y)}{Y_t} - 1 \right)$$ where $E(Y)$ denotes the expected value of *Y*.

Since the access to forecasts $\hat{Y}$ provides information about Y, the quadratic uncertainty conditioned to forecasts $HU^2(Y / \hat{Y})$ is expected to be lower than the initial value. Therefore, the difference between both expressions can be interpreted as the amount of information provided by forecasts.

Furthermore, since it is advisable to consider the existing relationship between observed and forecasted values, we also compute the linear correlation coefficient $r_{Y,\hat{Y}}$ and thus the Quadratic Information Accuracy Measure (QIAM) is given by the following expression: $QIAM(Y, \hat{Y}) = HU^2(Y) - HU^2(Y / \hat{Y})(1 - r_{Y,\hat{Y}})$, whose result increases with the forecasting accuracy, leading to a null value when forecasts are useless ($\hat{Y}$ does not provide any information about *Y*).

The two proposed accuracy measures, Theil´s U Index and Quadratic Information Accuracy Measure, show some similarities (mainly referred to their informational content) and also differences regarding to their interpretation (more accurate forecasts lead to lower values of Theil´s U Index, and higher values of the QIAM). Nevertheless, this is not a serious drawback since, following the M3-Competition practice, the comparison of different forecasting techniques is based on rankings, as we will explain in a further section.

3. THE M-COMPETITION

The M-Competition is one of the main empirical investigations in the field of forecasting, since it compares the performance of a large number of time series methods, provided by recognised experts. More specifically, its third edition (M3-Competition, https://forecasters.org/resources/time-series-data/m3-competition/), was developed by S. Makridakis & M. Hibon (2000) including 3003 time series classified in micro (828), industry (519), macro (731), finance (308), demographic (413) and other (204). Regarding the forecasting methods, Table 1 summarises the 24 considered techniques which vary from simple to quite sophisticated procedures.

Table 1
Categories and Methods included in the M-Competition

Single techniques	Explicit trend models	Decomposition	ARIMA Models	Expert Systems	Neural Networks
Naïve Single	Holt Robust-Trend Winter Dampen PP-Autocast Theta-Sm Comb S-H-D	Theta	BJ Authomatic Autobox 1 Autobox 2 Autobox 3 AAM1 AAM2 ARARMA	Forecast Pro SmartFcs RBF Flores/Pearce 1 Flores/Pearce 2 Forecast X	Automat ANN

Source: Makridakis & Hibon (2000).

The number of required forecasts in the M3-Competition depends on the frequencies of the considered series. Thus, six forecasts are required for yearly data, while eight forecasts are asked for quarterly data and eighteen forecasts for monthly data.

Once these forecasts are provided, the previously described competing methods are ranked according to their accuracy measured by five expressions:

The *symmetric Mean Absolute Percentage Error*, defined as:

$$ sMAPE = \sum_t \frac{\left| Y_t - \hat{Y}_t \right|}{\dfrac{Y_t + \hat{Y}_t}{2}} 100 . $$

The *Median symmetric Absolute Percentage Error*, where the mean is replaced by the median which is not influenced by extreme values and therefore leads to more robust results.

The *Average Ranking*, computed by sorting, for each forecasting horizon, the symmetric absolute percentage error for each method from the smallest to the largest, and then computing the mean ranking for each forecasting horizon.

The *Percentage Better*, reporting the percentage of times that a give method has a smaller forecasting error than another method.

The *Relative Absolute Error* (RAE) computed as the absolute error of the considered method relative to the absolute error of a naïve model.

According to these criteria, the results of the M3-Competition lead to the following conclusions:

a) Statistically sophisticated methods do not necessarily provide more accurate forecasts than simpler ones.

b) The relative ranking of the performance of the various methods varies according to the accuracy measure being used.

c) The accuracy when various methods are being combined outperforms, on average, the accuracy of the individual methods being combined.

d) The accuracy of the various methods depends upon the length of the forecasting horizon.

With the aim of analysing if these conclusions are also valid for the information-based accuracy measures, we describe the empirical findings obtained when Theil´s U Index and the Quadratic Information Accuracy Measure are applied to the M3 Competition results and to the subsets of macroeconomic and financial time series.

4. REVISITING THE M-3 COMPETITION. EMPIRICAL FINDINGS WITH INFORMATION-BASED ACCURACY MEASURES

In the previously described framework, Theil´s U Index and QIAM have been computed for the M3-Competition series, leading to the corresponding forecasting techniques accuracy rankings. More specifically, given the interpretation of each of the proposed measures, Theil' s U index results are ordered from the smallest to the largest value, while QIAM figures are ordered from the largest to the smaller.

Following this procedure, in this section we first summarise the global conclusions referred to the M3-Competition. Afterwords, since the performance of forecasting techniques could depend on the type of variables, we pay attention to the subsets of macroeconomic and finance series.

4.1. M-3 Competition Global findings

The comparison of Theil´s U Index and QIAM results for 3003 M3-Competition series shows outstanding similarities according to the ranking analysis. In fact, in 32% of the series the most accurate technique results to be the same for both Theil´s U Index and QIAM, and if the analysis is extended to the

"top-five" forecasting techniques, partial coincidences between the proposed information measures are found in almost 90% of the series.

This ranking analysis also allows the comparison of the considered forecasting procedures, as it is summarised in Figure 1.

Figure 1
Number of "*top positions*" in the ranking of forecasting techniques. M3-Competition
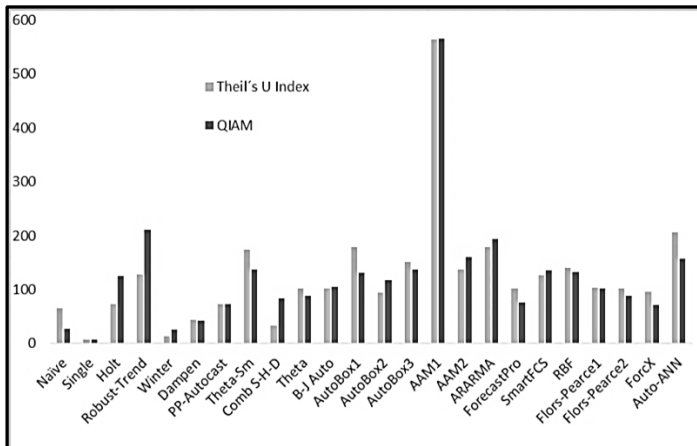
Source: Own elaboration.

According to the obtained rankings, automatic ARIMA modeling with intervention analysis (AAM1) is the most accurate technique, followed by other sophisticated procedures such as Automated Artificial Neural Networks (Auto-ANN), the Automated Parzen´s methodology with auto regressive filter (ARARMA) and Robust ARIMA univariate Box-Jenkins (AutoBox3). These results suggest that more sophisticated techniques provide more accurate forecasts, a conclusion that differs from the M3-Competition analysis carried out by Makridakis & Hibon (2000) and the recent work by Green & Armstrong (2015).

Nevertheless, several similarities are found with the M3-competition results regarding two explicit trend models: the Robust-Trend procedure, which is a non-parametric version of the Holt´s linear model with median based estimate of trend, is located in an outstanding position of the ranking according to the quadratic information while the Theta-sm (successive smoothing plus a set of rules for dampening the trend) performs well according to Theil´s U Index. Furthermore, both techniques perform particularly well in yearly series, thus agreeing with the conclusions of Makridakis & Hibon (2000) Fildes *et al.* (1998) and Fildes & Petropoulos (2015).

4.2. Findings for macroeconomic series

A total of 731 macroeconomic series have been included in the M3-Competition, comprising yearly data (83), quarterly data (336) and monthly data (312). Our empirical application to the subset of macroeconomic series has been carried out following the previously explained method, with the exception of the automatic ARIMA procedures (AAM1 and AAM2) since they do not provide the required monthly forecasts.

The obtained results, represented in Figure 2, show a more uniform ranking that the one obtained for the whole M3-Competition. According to Theil´s U Index the performance of the most sophisticated procedure (Auto-ANN) is similar to some other techniques with lower level of complexity (ARARMA, AutoBox1) and even to quite simple alternatives (Theta Sm and Robust-Trend). A similar behaviour is shown by the Quadratic Information Accuracy Measure (QIAM), with the exception of the Robust-Trend method, that clearly heads the ranking.

Figure 2
Number of "*top positions*" in the ranking of forecasting techniques.
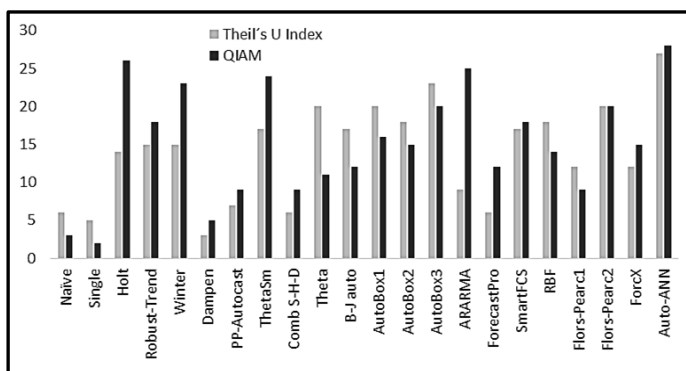Macroeconomic series
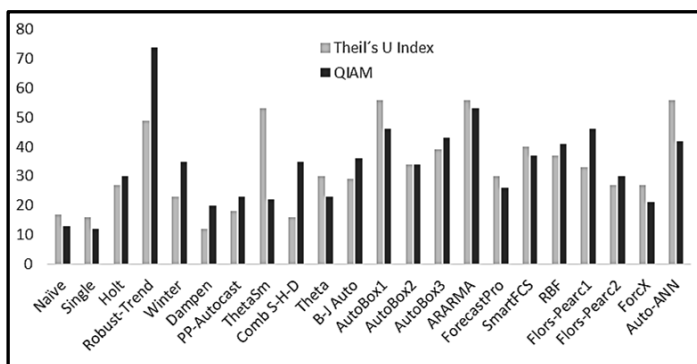
Source: Own elaboration.

It can also be observed that, according to the QIAM, the accuracy of combined forecasts (as in Comb S-H-D) outperforms the accuracy of the individual methods being combined, thus agreeing with the M3-Competition third conclusion.

4.3. Findings for financial series

Since financial time series are inherently noisy and non-stationary, they are considered as particularly challenging for forecasting purposes. Therefore, a specific analysis of the subset of 308 financial time series has been carried out following the previously explained procedure.

Although the information-based accuracy measures show several coincidences with the previous sections, some differential aspects are found in the financial series. As it can be seen in Figure 3, one of our main finding is that the Automated Artificial Neural Networks (AutoAnn) results to be the most accurate procedure for forecasting financial series, according to both Theil´s U Index and the Quadratic Information Accuracy Measure. Thus, our conclusions agree with those obtained by Sohl &Venkatachalam (1995) and Balkin & Ord (2000) among others.

Figure 3
Number of "*top positions*" in the ranking of forecasting techniques. Financial Series

Source: Own elaboration.

Another complex methods such as AutoBox and FloresPearce2 also show a good behavior according to both informational accuracy measures. Nevertheless, the similarities between Theil´s U index and QIAM are lower for financial series, as we can see in Figure 3. More specifically, outstanding positions are found for some explicit-Trend models such as Holt, Winter, Theta-Sm and Robust Trend, that according to the QIAM, surpass the behavior of more sophisticated methods.

It can also be observed that, according to both Theil´s U and QIAM, when dealing with financial series the accuracy of the combined forecasts (as in Comb S-H-D) does not outperform the accuracy of the individual methods being combined, thus disagreeing with the third conclusion of the M3-Competition and also the more recent papers by Hibon & Evgeniou (2005) and Graefe (2015).

5. DISCUSSION

The application of the proposed information measures (Theil´s U Index and QIAM) to the M3-Competition developed by Makridakis & Hibon (2000) confirms that the performance ranking of the considered forecasting methods

depends on the accuracy measure being used. Nevertheless, high levels of similarity are found between Theil´s U Index and the QIAM, suggesting the influence of their informational content. Furthermore, the accuracy of the different methods depends upon the length of the forecasting horizon involved, thus agreeing with Makridakis and Hibon´s findings.

Nevertheless, as it is summarised in Table 2, we have also found some conclusions differing from the M3-Competition analysis. The main divergence is that, while these authors conclude that statistically sophisticated procedures do not lead to more accurate forecasts, the results provided by Theil's U index and the Quadratic Information Accuracy Measure suggest that the most accurate forecasts correspond to sophisticated methods. This is especially true in the case of financial time series, since automated artificial neural network is found to be the best forecasting technique. Nevertheless, the superiority of the more complex forecasting methods -such as ARIMA modeling with intervention analysis (AAM1), Automated Artificial Neural Networks (Auto-ANN), Automated Parzen´s methodology with auto-regresive filter (ARARMA) or Robust ARIMA univariate Box-Jenkins (AutoBox3)- can also be observed for the global M3-Competition.

Table 2
Informational Measures agreement with M3-Competition Conclusions

M3-Competition Conclusions	Agreement		
	3003 Series	Macro Series	Finance Series
Statistically sophisticated methods do not necessarily provide more accurate forecasts than simpler ones.	No	Yes	No
The relative ranking of the performance of the various methods varies according to the accuracy measure being used.	Yes	Yes	Yes
The accuracy when various methods are being combined outperforms, on average, the accuracy of the individual methods being combined.	No	Yes	No
The accuracy of the various methods depends upon the length of the forecasting horizon.	Yes	Yes	Yes

*Source:* Makridakis and Hibon (2000) and own elaboration.

This fact leads to the simplicity versus complexity debate, suggesting the need of an operational definition of both terms, not easy to establish. Although according to some authors as Green & Armstrong (2015) a forecasting method can be considered as simple if it is understood by forecasting users, in this paper we propose a more detailed scale, assigning different levels of complexity (denoted by C_i) to the different techniques as follows:

1. Simple techniques (Naïve and single)
2. Explicit trend models (Holt, Robust-Trend, Winter, Dampen, PP-Autocast, Theta-Sm, Comb S-H-D) and Decomposition (Theta)

3. ARIMA Models (BJ Authomatic, Autobox 1, Autobox 2, Autobox, AAM1 and AAM2)

4. Expert Systems (Forecast Pro, SmartFcs, RBF, Flores/Pearce 1, Flores/Pearce 2, Forecast X)

5. Neural Networks (Automated Artificial Neural Networks)

Regarding the accuracy level (denoted by A_i), in order to obtain a similar scale, an index –normalised to 5- is computed for each forecasting technique as $A_i = \frac{5N_i}{máx\ \{N_i\}}$, where N_i represents the respective number of top positions in the forecasting ranking.

Figure 4
Accuracy-Complexity relationship for Theil´s U Index
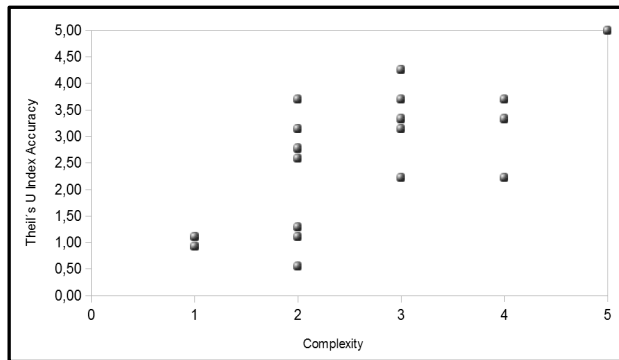
Source: Own elaboration.

Figure 5
Accuracy-Complexity relationship for the QIAM
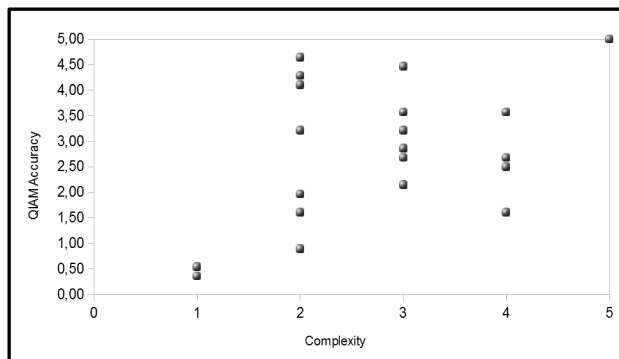
Source: Own elaboration.

The complexity-accuracy analysis, summarised in Figures 4 and 5, confirms the superiority of neural networks, especially when using Theil´s U Index, while the Quadratic Information Accuracy Measure provides weaker evidence, due to the good behavior of some explicit-Trend models whose level of complexity is quite low.

These scatter diagrams lead to the least-squares estimations summarised in Table 3, confirming that the level of complexity in the forecasting methods significantly increases the level of accuracy, especially when it is measured through Theil´s U Index.

Table 3
Accuracy-Complexity least-squares estimation results

Accuracy Measure	Constant	Complexity level
Theil´s U Index	1.1 (0.64)	0.55** (0.21)
QIAM	1.32 (0.49)	0.49* (0.25)

**Significant at 5% level, * Significant at 10% level

Source: Own elaboration.

6. CONCLUSIONS

In a context of increasing forecasting availability, this paper focuses on forecasting evaluation, suggesting the use of two informational measures: Theil´s U Index and Quadratic Information Accuracy Measure (QIAM). The application of these measures to the M3-Competition developed by Makridakis & Hibon (2000) confirms that, even if the performance ranking of the considered forecasting methods depends on the accuracy measure being used, high levels of similarity are found between Theil´s U Index and the QIAM, suggesting the influence of their informational content.

As expected, the accuracy of the different methods depends upon the length of the forecasting horizon involved, thus agreeing with Makridakis and Hibon´s findings.

Nevertheless, some conclusions differing from the M3-Competition analysis have also been found. The main divergence is that the results provided by Theil's U index and the Quadratic Information Accuracy Measure suggest that the most accurate forecasts correspond to sophisticated methods. This is especially true in the case of financial time series, since automated artificial neural network (AANN) is found to be the best forecasting technique.

The superiority of the more complex forecasting procedures -such as ARIMA modeling with intervention analysis (AAM1), Automated Artificial Neural Networks (Auto-ANN), Automated Parzen´s methodology with auto-

regresive filter (ARARMA) or Robust ARIMA univariate Box-Jenkins (AutoBox3)- can also be observed for the global M3-Competition, while the subset of macroeconomic series provides more uniform results.

The behaviour of macroeconomic series is also different to the M3-Competition as a whole with regard to the advantages of combining forecasts. In this case, and especially when Theil´s U Index is used, Comb-SHD method (a combination of Single, Holt and Dampen procedures) outperforms the individual methods, thus agreeing with Makridakis & Hibon (2000) and Hibon & Evgeniou (2005).

In summary, our empirical results confirm that the performance of forecasting techniques depends on both the considered accuracy measure and the type of variables. Furthermore, according to the information-based accuracy measures U and QIAM significant differences are found between macroeconomic and financial time series.

Finally, although this paper has provided some new evidence about the M3-Competition results based on informational accuracy measures, we strongly believe that, following the recommendations by Makridakis and Hibon "*more empirical research is needed to advance the field of forecasting and make it more practical for business and another organizations requiring predictions, since ignoring empirical findings is contrary to rational thinking and scientific inquiry*".

REFERENCES

ARMSTRONG, J.S.; GREEN, K.C. y GRAEFE, A. (2015). "Golden rule of Forecasting: Be conservative". *Journal of Business Research*, 68, pp. 1717-1731.

BALKIN, S.D. y ORD, J.K. (2000). "Automatic neural network modeling for univariate time series". *International Journal of Forecasting*, 16(4), pp. 509-515.

BIEMEL, F. y MACKAY, D.B. (1973). "Theil´s Forecast Accuracy Coefficient: A Clarification". *Journal of Marketing Research*, X, pp. 444-446.

BINNER, J.M.; BISSONDEEAL, R.K.; ELGER, T.; GAZELY, A.M. y MULINEUX, A.W. (2005). "A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia". *Applied Economics*, 37(6), pp. 665-680.

BRANDON, C.; FRITZ, R. y XANDER, J. (1983). "Econometric forecasts: evaluation and revision". *Applied Economics*, 15(2), pp.187-201.

CHATFIELD, C. (1988). "What is the 'best' method of forecasting?". *Journal of Applied Statistic*s, 15(1), pp.19-38.

CICCARELLI, M. y HUBRICH K. (2010). "Forecast uncertainty: sources, measurement and evaluation". *Journal of Applied Econometrics*, 25(4), pp. 509-513.

COUTIÑO, A. (2015). "Acerca del poder predictivo de Klein". *Estudios de Economía Aplicada*, 33(2), pp. 385-392.

DAGUM, E.B. (2010). "Business Cycles and Current Economic Analysis". *Estudios de Economía Aplicada*, 28(3), pp. 577-594.

FILDES, R.; HIBON, M.; MAKRIDAKIS, S. y MEADE, N. (1998). "Generalising about univariate forecasting methods: further empirical evidence". *International Journal of Forecasting*, 14, pp. 339-351.

FILDES, R. y PETROPOULOS, F. (2015). "Simple versus complex selection rules for forecasting many time series". *Journal of Business Research*, 68, pp. 1692-1701.

FILDES, R. y PETROPOULOS, F. (2015). "Is there a Golden rule?". *Journal of Business Research*, 68, pp. 1742-1745.

FOMBY, T.B. y SAMANTHA, S.K. (1993). "Average of forecasts: an improvement". *Applied Economics,* 25(5), pp .689-694.

GIL, M.A.; PÉREZ, R. y GIL, P. (1989). "A family of measures of uncertainty involving utilities: Definition, properties, applications and statistical inferences". *Metrika*, 36, pp.129-147.

GRAEFE, A. (2015). "Improving forecasts using equally weighted predictors". *Journal of Business Research*, 68, pp.1792-1799.

GRANGER, C.W.J. (1996). "Can we improve the perceived quality of economic forecasts?". *Journal of Applied Econometrics*, 11(5), pp. 455-473.

GRANGER, C.W.J. y JEON, Y. (2003). "Interactions between large macro models and time series analysis". *International Journal of Finance & Economics*, 8(1), pp.1-10.

GRANGER, C.W.J. y JEON, Y. (2004). "Forecasting Performance of Information Criteria with Many Macro Series". *Journal of Applied Statistics*, 31(10), pp.1227-1240.

GREEN, K.C. y ARMSTRONG, J.S. (2015). "Simple versus complex forecasting: The evidence". *Journal of Business Research*, 68, pp.1678-1685.

GREER, M.R. (2005). "Combination forecasting for directional accuracy: An application to survey interest rate forecasts". *Journal of Applied Statistics*, 32(6), pp. 607-615.

HARVDA, J. y CHARVAT, F. (1967). "Quantification method of classification Processes". *Kybernetica*, 3, pp. 129-147.

HIBON, M. y EVGENIOU, T. (2005). "To combine or not to combine: selecting among forecasts and their combinations". *International Journal of Forecasting*, 21, pp.15-24.

HYNDMAN, R. (2006). "Another look at forecast-accuracy metrics for intermittent demand". Foresight, 4, pp. 43-46.

INTERNATIONAL INSTITUTE OF FORECASTERS (2000). *M3-Competition*, https://forecasters.org/resources/time-series-data/m3-competition/ (last accessed on April 2017).

LÓPEZ, A.J. y PÉREZ, R. (2012). "Financial Forecasting Accuracy: Exploring the M3-Competition". *6th CSDA International Conference on Computational and Financial Econometrics*, Oviedo, Spain.

LÓPEZ, A.J.; PÉREZ, R. y MORENO, B. (2011). "Forecasting Performance and M-Competition. Does the accuracy measure matter?. *International Statistical Institute (ISI), Proceedings of the 58th World Statistical Congress*, Dublin.

MAKRIDAKIS, S. y BAKAS, N. (2016). "Forecasting and Uncertainty: A Survey". *Risk and Decision Analysis*, 6(1), pp. 37-64.

MAKRIDAKIS, S. y HIBON, M. (2000). "The M-3 competition: results, conclusions and implications*". International Journal of Forecasting*, 16, pp. 451-476.

MAKRIDAKIS, S. y TALEB, N. (2009). "Living in a world of low levels of predictability". *International Journal of Forecasting*, 25, pp. 840-844.

MORENO, B. y LÓPEZ, A.J. (2013). "Combining Economic Forecasts by using a Maximum Entropy Econometric Approach". *Journal of Forecasting*, 32(2), pp.124-136.

PÉREZ, R. (1985). "Estimación de la incertidumbre, la incertidumbre útil y la inquietud en poblaciones finitas". *Revista Real Academia de las Ciencias de Madrid*, LXXIX, 4: pp. 651-654.

PETROPOULOS, F.; MAKRIDAKIS, S.; ASSIMAKOPOULOS, V. y NIKOLOPOULOS, K. (2014). "Horses for Courses in Demand Forecasting". *European Journal of Operational Research*, 237, pp. 162-173.

PULIDO, A. (2015). "La herencia de Klein (1920-2013): Una visión de futuro". *Estudios de Economía Aplicada*, 33(2), pp. 359-384.

PULIDO, A. y PÉREZ, J. (2006). "Lawrence R. Klein y la Economía Aplicada". *Estudios de Economía Aplicada*, 24(1), pp. 43-94.

RAMOS, R.; CLAR, M. y SURIÑACH, J. (2000). "Comparación de la capacidad predictiva de los modelos de coeficientes fijos frente a variables en los modelos econométricos regionales: un análisis para Cataluña". *Estudios de Economía Aplicada* 15, pp. 125-162.

SCHERVISH, M.J.; SEIDENFELD, T. Y KADANE, J.B. (2009). "Proper Scoring Rules, Dominated Forecasts". and *Coherence Decision Analysis*, 6(4), pp. 202-221.

SHANNON, C. (1948). "A mathematical theory of communication". *Bell System Technical Journal*, 27, pp. 379-423.

SOHL, J.E. y VENKATACHALAM, A.R. (1995). "A neural network approach to forecasting model selection". *Information & Management*, 29, pp. 297-303.

SOYER, E. y HOGARTH, R.M. (2015). "The Golden Rule of forecasting: Objections, refinements and enhacements". *Journal of Business Research*, 68, pp. 1702-1704.

SYNTETOS. A.; NIKOLOPOULOS, K. y BOYLAN, J. (2010). "Judging the judges through accuracy-implication metrics: The case of inventory forecasting". *International Journal of Forecasting*, 26, pp. 134-143.

THEIL, H. (1955). "Who forecasts best?". *International Economic Papers,* 5, pp. 194-199.

THEIL, H. (1957). "A Statistical Appraisal of Postwar Macroeconomic Forecasts in the Netherlands and Scandinavia". *Bulletin of the International Statistical Institute,* 35, pp. 49-64.

THEIL, H. (1966). *Applied Economic Forecasting*. North Holland Publishing.

TSANGARI., H. (2007). "An Alternative Methodology for Combining Different Forecasting Models". *Journal of Applied Statistics*, 34(4), pp. 403-421.

WENZEL, T. (2001). "Hits-and-misses for the evaluation and combination of forecasts". *Journal of Applied Statistics*, 28(6), pp. 759-773.