Forecasting Performance and M-Competition. Does the accuracy measure matter?

López, Ana Jesús
University of Oviedo, Department of Applied Economics
Campus del Cristo s/n
Oviedo (33006),Spain
E-mail: anaj@uniovi.es

Pérez, Rigoberto
University of Oviedo, Department of Applied Economics
Campus del Cristo s/n
Oviedo (33006),Spain
E-mail: rigo@uniovi.es

Moreno, Blanca
University of Oviedo, Department of Applied Economics
Campus del Cristo s/n
Oviedo (33006),Spain
E-mail: morenob@uniovi.es

From its very beginning, the Information Theory has been considered as a useful tool for economic analysis, since it provides suitable measures of inequality, industrial concentration or goodness of fit. One of the most widely used information measures is the Entropy proposed by Shannon. Given a variable X with probabilities: $p_1, p_2, \ldots, p_n; p_i > 0, \sum_{i=1}^{n} p_i = 1$, C.E. Shannon (1948) defined the Entropy Measure $H(X) = -\sum_{i=1}^{n} p_i \log p_i$. The result of this expression is null when $p_i = 1$ for a given i and $p_j = 0$, $\forall j \neq i$, and adopts its maximum value when all the results are equally likely $p_i = \frac{1}{n}$, $\forall i$.

Since Shannon's measure has some limitations, mainly related to its estimation, β-type uncertainty measures have been defined by some authors as J. Havrda and F. Charvat (1967), leading for the case $\beta = 2$ to the quadratic uncertainty proposed by R. Pérez (1985): $H^2(X) = 2\sum_{i=1}^{n} p_i(1 - p_i)$.

This author also defined measures of entropy involving utilities which show some advantages with regard to Shannon's measures. More specifically, given a variable X with related probabilities $p_1, p_2, \ldots, p_n$ and utilities $u_1, u_2, \ldots, u_n, u_i > 0$, the quadratic unquietness is defined as the difference between the quadratic uncertainty involving utilities and the quadratic uncertainty, leading to the expression:

$$HU^2(X) = 2\sum_{i=1}^{n} p_i \left(\frac{E(u)}{u_i} - 1 \right)$$

whose result can be interpreted as the uncertainty level specifically related to utilities and therefore is called "unquietness".

This kind of measures provides a suitable framework for the evaluation of forecasts. Thus, given a variable Y with observed values $Y_1, \ldots, Y_T$, the quadratic unquietness of Y can be computed as:

$$HU^2(Y) = \frac{2}{T}\sum_{i=1}^{n} \left(\frac{E(Y)}{Y_i} - 1 \right)$$

Within this framework, the quality of a set of forecasts $\hat{Y}_1, \ldots, \hat{Y}_T$ can be measured by comparing the previous expression with the quadratic unquietness of Y conditioned to its forecasts $\hat{Y}$

$$HU^2\left(Y/[\hat{Y}]\right) = \sum_j p_{[\hat{Y}_j]} \frac{E(Y)}{E_{[\hat{Y}_j]}(Y)} H\left(Y/[\hat{Y}_j]\right)$$

where we denote by $E(Y)$ the expected value of the observed variable while $E_{[\hat{Y}_j]}(Y) = \sum_t Y_t p_{Y/[\hat{Y}_j]}$ is the expected value of Y conditioned to the forecasting interval $[\hat{Y}_j]$.

Since the quadratic unquietness of Y conditioned to its forecasts $HU^2(Y/[\hat{Y}])$ is expected to be lower than the quadratic unquietness of Y, $HU^2(Y)$ the difference between both results can be interpreted as the information provided by forecasts.

Nevertheless, 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 we define the quadratic information provided by forecasts as the following expression:

$$IC\left(Y, [\hat{Y}]\right) = HU^2(Y) - HU^2\left(Y/[\hat{Y}]\right)\left(1 - r_{Y,\hat{Y}}\right)$$

whose result increases both with the reduction of unquietness related to the forecasts and with the correlation between actual and forecasted values.

Exploring the M-Competition. New evidence from information accuracy measures

Forecasting availability has widely increased, suggesting the need of analyzing the adequacy of different alternative methods. One of the main empirical researches in this field is the M-Competition developed by Makridakis and Hibon [6], whose last edition (M3) is referred to year 2000 and includes 3003 time series classified in micro (828), industry (519), macro (731), finance (308), demographic (413) and other (204).

The M-Competition considers 24 different forecasting procedures as it is summarized in table 1:

Table 1: Categories and Methods included in the M-Competition

Single techniques	Explicit trend models	Decomposition	ARIMA Models	Expert System	Neural Networks
Nave	Holt	Theta	B-J	Forecast Pro	Automat ANN
Single	Robust-Trend		Authomatic	SmartFcs	
	Winter		Autobox1	RBF	
	Dampen		Autobox2	Flores/Pearce1	
	PP-autocast		Autobox3	Flores/Pearce2	
	Theta-sm		AAM1	ForecastX	
	Comb S-H-D		AAM2		
			ARARMA		

Regarding the evaluation of forecasts, the M-Competition includes five different accuracy measures: the *symmetric mean absolute percentage error*, defined as $sMAPE = \sum_t \frac{|Y_t - \hat{Y}_t|}{\frac{Y_t + \hat{Y}_t}{2}} 100$, 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 of 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 time that a given method has a smaller forecasting error than another method) and finally the *relative absolute error* (RAE, computed as the absolute error for the considered method relative to the absolute error for naive models).

According to these criteria, the results of the M-Competition suggest some interesting facts:

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 individual methods being combined.

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

With the aim of analyzing the robustness of these conclusions, we have evaluated the M-competition results through new forecasting accuracy criteria, based on information measures: *U-Theil's Index* and the previously described *quadratic information provided by forecasts*.

The so-called U-Index has been proposed by H. Theil in 1966 and it is given by the expression:

$$U = \sqrt{\frac{\sum_t \left(\frac{\hat{Y}_t - Y_{t-1}}{Y_{t-1}} - \frac{Y_t - Y_{t-1}}{Y_{t-1}}\right)^2}{\sum_t \left(\frac{Y_t - Y_{t-1}}{Y_{t-1}}\right)^2}}$$

whose value decreases as forecasts become more accurate, leading to null values when actual and forecasted rates of growth are coincident.

One of the main advantages of Theil's index is the decomposability, allowing the additive disaggregation of U^2 in three terms, respectively related to bias, variance and covariance factors. Furthermore, since the information theory provides a suitable framework for several economic applications including the evaluation of forecasts we propose the use of the previously defined *Quadratic information provided by forecasts*:

$$IC\left(Y, [\hat{Y}]\right) = HU^2(Y) - HU^2\left(Y/[\hat{Y}]\right)\left(1 - r_{Y,\hat{Y}}\right)$$

which can be applied over the M-Competition forecasting quartiles for each of the considered series and procedures.

The application of these two information measures to the M-Competition results allows the identification of the most accurate forecasting method for each of the considered series. More specifically, given the previously described interpretations, U Theil's results are ordered from the largest to the smallest while Quadratic information figures are ordered from the smallest to the largest, thus leading to the corresponding rankings of forecasting techniques.

Following these criteria, we summarize the results referred to the whole M-Competition and to the macroeconomic series.

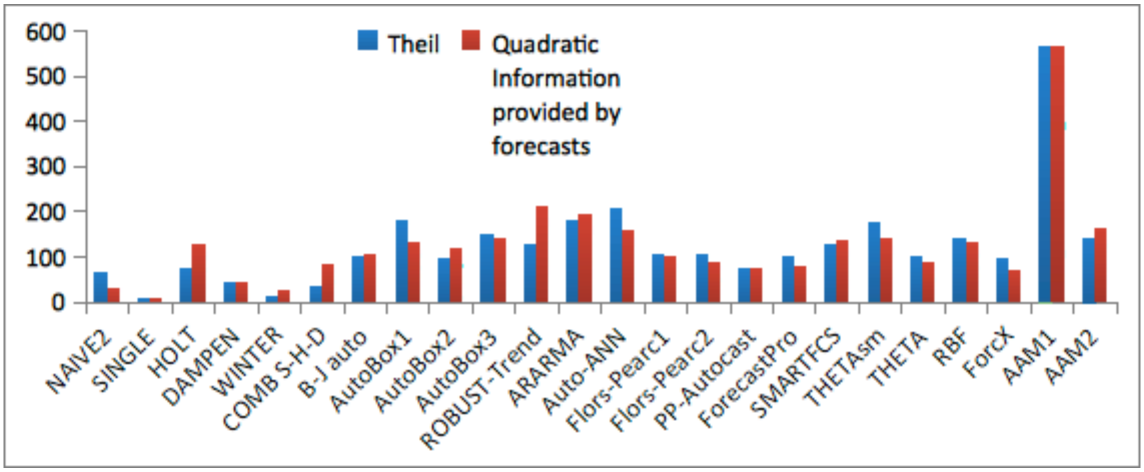
Main findings for the 3003 M-Competition series

The computation of U Theil's Index and the *quadratic information provided by forecasts* IC to the 3003 M-Competition series leads to similar results. In fact, if we rank the 24 considered procedures according to their forecasting accuracy, in 32% of the series the most accurate technique is coincident for both Theil and quadratic information measures.

Furthermore, if we extend this analysis to the top-five forecasting techniques, partial coincidences are found in almost 90% of the series.

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

Figure 1: Number of top positions in the ranking of forecasting procedures



According to this analysis, automatic ARIMA modeling with intervention analysis (AAM1) heads the ranking of most accurate techniques, followed by another sophisticated procedures such as Automated Artificial Neural Networks (Auto-ANN), Automated Parzens methodology with auto regressive filter (ARARMA) and Robust ARIMA univariate Box-Jenkins (AutoBox3). These results suggest that more sophisticated techniques provide more accurate forecasts, thus differing from the M-Competition analysis carried out by Makridakis and Hibon (2000).

Nevertheless, several coincidences are also found with the M-competition results, regarding two explicit trend models: Robust-Trend procedure (which is a non-parametric version of Holts linear model with median based estimate of trend) is located in an outstanding position of the ranking according to the quadratic information while Theta-sm (successive smoothing plus a set of rules for dampening the trend) performs well according to Theil's measure. Furthermore, both techniques perform particularly well in yearly series, thus agreeing with the conclusions of Makridakis and Hibon.

Main findings for macroeconomic series

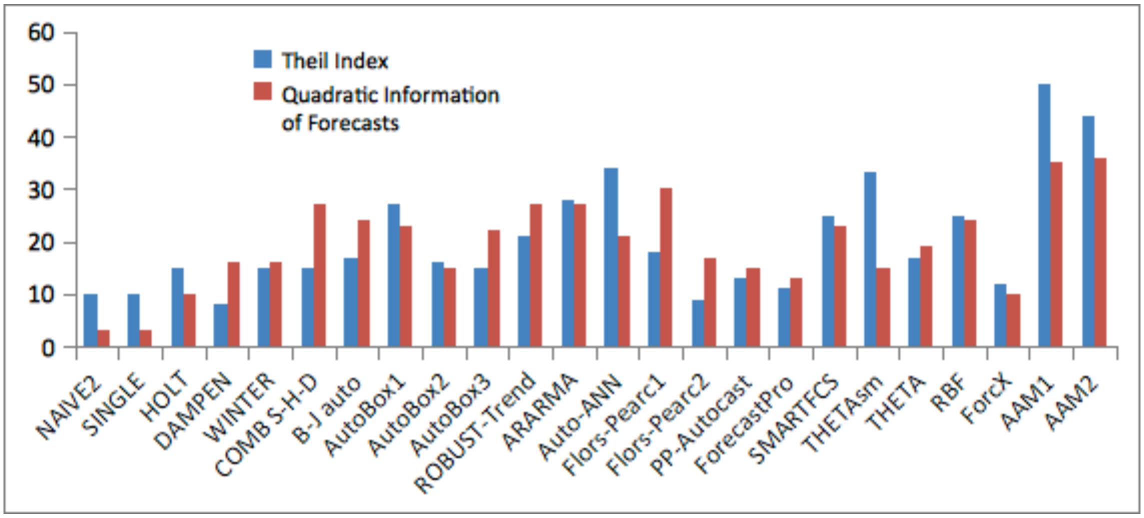
Since the behavior of the forecasting techniques could depend of the different types of series (micro, industry, macro, finance, demographic and others) in this section we focus on the M-Competition macroeconomic series whose total amount is 731, including yearly data (83), quarterly data (336) and monthly data (312). The considered forecasting horizons depend on the frequency of the respective series, and thus 6 forecasts are requested for yearly data, 8 for quarterly data and 18 for monthly data.

In order to carry out this empirical application we should keep in mind that Automatic ARIMA Modelling with and without intervention (AAM1 and AAM2) do not provide monthly forecasts. With this regard, two solutions can be considered in order to guarantee homogeneous comparisons.

A first option consists in the exclusion of monthly series, thus providing a ranking of the 24 forecasting procedures, as it is summarized in Figure 2.

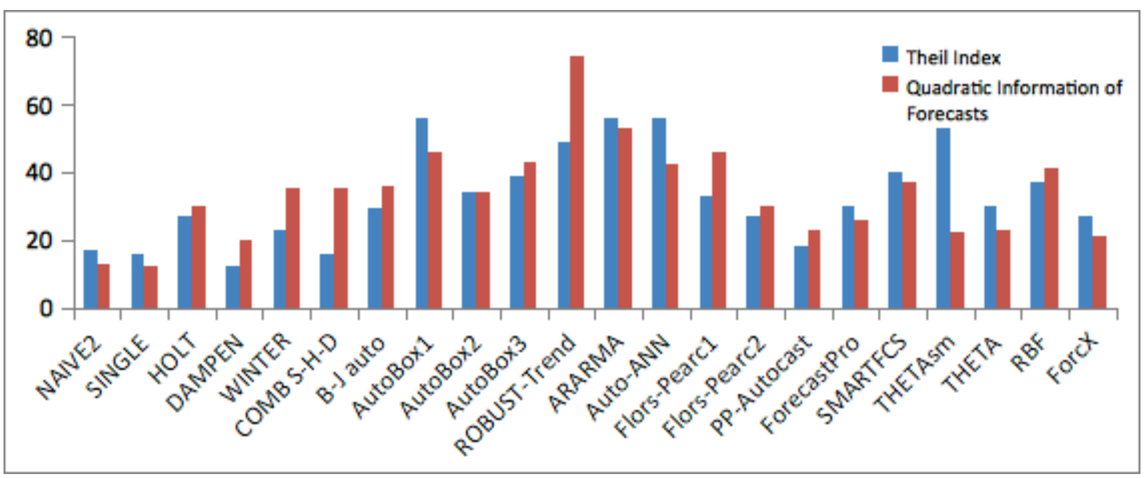
According to these results, the orderings related to Theil's Index and Quadratic Information Measures are quite similar, and both suggest that the Automatic ARIMA modeling techniques (AAMA1 and AAMA2) should be considered the most accurate methods, followed by another sophisticated procedures such as Auto-ANN (Automated Artificial Neural Networks), ARARMA (Automated Parzens methodology with auto regressive filter) and AutoBox1 (Robust ARIMA univariate Box-Jenkins) and also by some less sophisticated explicit trend models as Theta-sm (which is a successive smoothing) and Robust-Trend (a non-parametric version of Holts linear model).

Figure 2: Number of top positions in the ranking of forecasting procedures for yearly and quarterly macroeconomic M-competition series



A second alternative for the analysis of the M-Competition macroeconomic series is the exclusion of AAM1 and AAM2 methods, thus allowing the consideration of all the 732 available series. In this case the Theil Index provides similar results (with ARARMA, Auto-ANN and AutoBox1 heading the ranking) while the Quadratic information of forecasts leads to the election of the Robust-Trend procedure.

Figure 3: Number of top positions in the ranking of forecasting procedures for the macroeconomic M-competition series



Concluding remarks

The increasing number of forecasting techniques raises questions about their accuracy, suggesting the need of evaluating alternative forecasting procedures. With this aim, this paper has focused on the last M-Competition developed by Makridakis and Hibon (2000), whose results have been compared with those obtained with two information measures: U Theil's Index and quadratic information provided by forecasts, allowing the identification of some similarities and differences.

Starting with the similarities, the obtained results confirm that the relative ranking of the performance of the considered methods depends on the accuracy measure being used. Nevertheless, a high level of similarity is found between Theil's Index and the quadratic information of forecasts, suggesting the influence of their informational content.

Furthermore, according to Makridakis and Hibons findings, the accuracy of the different methods depends upon the length of the forecasting horizon involved. Our results referred to macroeconomic series confirm this conclusion, since our first analysis (excluding monthly series) does no fully agree with the second (including the whole database).

Our results also show that, as expected, the Comb-SHD method, obtained as a combination of Single, Holt and Dampen procedures, usually outperforms the individual methods, especially when U Theil's Index is considered.

Nevertheless, we have also found some conclusions differing from the M-Competition analysis carried out by Makridakis and Hibon. These divergences are mainly related to the fact that, while these authors conclude that statistically sophisticated procedures do not provide more accurate forecasts, according to Theil's index and Quadratic information the most accurate forecasts correspond to sophisticated methods such as ARIMA modeling with intervention analysis (AAM1), Automated Artificial Neural Networks (Auto-ANN), Automated Parzens methodology with auto-regresive filter (ARARMA) or Robust ARIMA univariate Box-Jenkins (AutoBox3), although two explicit trend models (Robust-Trend and Theta-sm) are also located in outstanding positions of the accuracy ranking.

Finally, it must be stressed that the previously described results are referred to the whole forecasting horizon considered in each case (6 for yearly series, 18 for monthly series and 8 for quarterly and others) and, since they could vary with the length of the forecasting horizon, a more detailed analysis would be advisable.

In this way, following the considerations by Makridakis and Hibon *similar questions, if answered, can contribute to improving forecasting accuracy a great deal and make the field of forecasting more useful and relevant*.

REFERENCES

Alvargonzález, M.; López, A.J. and Pérez, R. (2004): "Growth-inequality relationship. An analytical approach and some evidence for Latin America", *Applied Econometrics and International Development*, vol 4, 2, pp. 91-108.

Billah, B.; Hyndman, R.J. and Koehler, A.B. (2005): "Empirical information criteria for time series forecasting model selection", *Journal of Statistical Computation and Simulation*, Volume 75, Issue 10, pp. 831 840.

Gil, M.A.; Pérez, R. and Gil, P. (1989): "A Family of Measures of Uncertainty Involving Utilities: Definition, Properties, Applications and Statistical Inferences", *Metrika*, 36, pp. 129-147.

Havrda, J. and Chavrat, F. (1967): "Quantification method of classification processes", *Kybernetica*, 3, pp. 30-35.

López, A.J. Pérez, R. and Moreno, B. (2010): "Forecasting Accuracy. New Evidences Based on the MCompetition", *International Conference on Trends and Perspectives in Linear Statistical Inference Linstat*, Tomar.

Makridakis, S.; Chatfield, C.; Hibon, M.; Lawrence, M.; Mills, T.; Ord, K. and Simmons, L.F. (1993): "The M-2 Competition: a real-time judgmentally based forecasting study", *International Journal of Forecasting*, 9, pp. 5-23.

Makridakis, S. and Hibon, M. (2000): "The M-3 Competition: results, conclusions and implications", *International Journal of Forecasting*, 16, pp. 451-476.

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

Pérez, R.; Gil, M.A. and Gil, P. (1986): "Estimating the uncertainty associated with a variable in a finite population", *Kybernetes*, 15, pp. 251-256.

Shannon, C.E. (1948): "A Mathematical Theory of Communication", *Bell System Technical Journal*, 27, pp. 379-423.

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