GENERATING FORECAST COMBINATION WEIGHTS WITH NEURAL NETWORKS

Rafael de O. Valle dos Santos
PETROBRAS
Av. República do Chile 330, Centro, Rio de Janeiro/RJ, 20031-170
rvsantos@petrobras.com.br

Marley M. B. R. Vellasco
PUC-Rio
Rua Marquês de São Vicente 225, Gávea, Rio de Janeiro/RJ, 22451-900
marley@ele.puc-rio.br

ABSTRACT

Empirical evidence in the literature on time series suggests that combining forecasts is less risky than selecting a single winning forecasting model. Focusing on convex combinations – linear combinations with forecast weights constrained to be non-negative and to sum to unity – this paper proposes a new weight generation framework called Neural Expert Weighting (NEW). The framework generates dynamic weighting models based on neural networks. Assessed with 4 petroleum products sales time series, it presented promising results.

KEYWORDS. Forecast combination, Time series, Neural networks.

EST

1. Introduction

During the decision-making process, multiple forecasts for the same variable may be available to the planning team. In this context, as Timmerman (2006) observes, a natural question arises: what is the best way to exploit information in individual forecasts? As this paper describes, having many forecasting models should not be seen as a weakness. It allows constructing multi-forecaster systems that unite, in some manner, all available forecasting information. If properly designed, those systems lead to consensual decisions that outperform individual ones. Bates and Granger (1969) are often referred as a classical study on that matter.

From the theoretical perspective, a single forecaster *selection* can be considered a special case of a weighted *combination* scheme, with one forecasting model (expert) receiving all the weight (Sánchez, 2008). Conversely, from a practical viewpoint, these procedures (selection or combination) lead to diverse computational methods, and the practitioner must decide if it is better to select a dominant forecaster or perform a forecast combination, where some established mean weights the contribution of each individual model.

The result of a forecast combination does not always outperform the best individual forecast but is less risky (Hibon and Evgeniou, 2005). One factor that counts for the selection risk is that this process relies on experienced forecasting professionals. This is not attainable in many real-world cases due to a lack of time, money or human resources. Also, empirical conclusions show that “even if the best model could be identified at each point in time, combination may still be an attractive strategy due to diversification gains, although its success will depend on how well the combination weights can be determined” (Timmermann, 2006). To *diversify* here means to form a forecasting ensemble wherein each forecaster acts in a complementary manner, i.e., each combining individual compensates for their counterparts’ errors.

To weight a forecaster is to cast a value (weight) indicating its contribution to a combination scheme. The simplest type of weighting procedure is the *static* type: based on historical data, a unique weighting vector is estimated and applied over the entire forecast horizon. This procedure may be enhanced by considering *dynamic* weight generation, i.e., weight generation schemes where the weighting vectors vary over the forecast horizon. (In this paper, the term “dynamic” does not necessarily mean that the weight generation is made explicitly, modeled by some parametric model.) Practically speaking, there is no guarantee that a dynamic generation scheme always outperforms a static one, but “some time-variation or adaptive adjustment in the combination weights (or perhaps in the underlying models being combined) can often improve forecasting performance” (Timmermann, 2006).

Most traditional weighting methods are based on in-sample performances of the assembled experts. (A remarkable exception is the combination by simple average, wherein the weights are always the same regardless of performance measures: $1/N$, where N is the number of combined forecasts.) Weight generation schemes solely based on historical forecast errors (or in some error derived measure) are much more cited in the literature and present “hard to beat” results; the lack of an auxiliary model to update information along the forecast horizon may, however, be considered a limitation.

A straightforward alternative to the traditional approach is (to try) to explicitly model the weight time-variation process. As explicit modeling is in turn an involved task without any performance increase guarantee, we thus formed the following proposal, which constitutes the main objective of the paper: to develop weighting models that can *enhance* traditional weighting procedures, *relaxing* historical forecast performance dependence and *abstracting* model complexity. (To abstract or encapsulate complexity is to allow for a less involved use of some technology or tool, making it more accessible for the final users.) We here then propose a weight generation framework, called Neural Expert Weighting (NEW). It delivers robust dynamic weighting models focusing on convex forecast combinations (section 2).

The remainder of this paper is organized as follows: Section 2 provides basic nomenclature and theory. Section 3 introduces the NEW framework. Section 4 presents a case study, assessing the framework with 4 time series. Conclusions are presented in Section 5.

2. Basics

We here define some basic variables considered in the text:

1) *Training* series composed of τ observations (other names are *in-sample* or *historical* series):

$$\mathbf{y}^{\tau} = [y_1 \quad y_2 \quad \dots \quad y_{\tau}]' \tag{1}$$

2) *Test* series with maximum horizon H (another name is *out-of-sample* series):

$$\mathbf{y}^{\tau+H|\tau} = [y_{\tau+1} \quad y_{\tau+2} \quad \dots \quad y_{\tau+H}]' \tag{2}$$

3) Forecast vector at time t+h ($h \leq H$), estimated with data gathered at t, for N forecasters:

$$\hat{\mathbf{y}}_{t+h|t} = [\hat{y}_{t+h|t,1} \quad \hat{y}_{t+h|t,2} \quad \dots \quad \hat{y}_{t+h|t,N}]' \tag{3}$$

4) Weighting vector for the convex combination at time t+h, estimated with data gathered at t:

$$\hat{\mathbf{w}}_{t+h|t} = [\hat{w}_{t+h|t,1} \quad \hat{w}_{t+h|t,2} \quad \dots \quad \hat{w}_{t+h|t,N}]' \tag{4}$$

5) Convex combination of forecasts at time t+h, estimated using data gathered at t:

$$y^{C}_{t+h|t} = \sum_{k=1}^{N} \hat{w}_{t+h|t,k}\,\hat{y}_{t+h|t,k} \ , \tag{5}$$

where

$$\sum_{k=1}^{N} \hat{w}_{t+h|t,k} = 1 \quad \text{and} \quad \hat{w}_{t+h|t,k} \geq 0 \tag{6}$$

The constraints in (6) turn the linear combination (5) into a *convex* combination. Convex combinations have great practical interest for two reasons: they (i) guarantee that the combined forecast is unbiased if the underlying forecasters are unbiased (Timmermann, 2006) and (ii) make weight interpretations straightforward, as weights can be seen as ordinary percentages. This paper focuses on such combinations.

3. Neural Expert Weighting

Neural Expert Weighting (NEW) is a framework for generating forecast weighting models, based on Multilayer Perceptron (MLP) neural networks (Bishop, 1995; Reed and Marks, 1999). It delivers robust dynamic weighting models for convex combinations, both relaxing in-sample performance dependence and abstracting statistical complexity. Additionally, an important concept introduced within the NEW framework is that of *limiting forecasters*. This concept may be seen in parallel with the framework idea *per se*, but was here introduced to take advantage of the neural networks non-linear capabilities, as an attempt to enhance convex combinations.

3.1. Framework Description

Once we have a group of forecasters (experts) to be weighted and combined, the following sequence of steps defines the framework usage:

1) Build the training pairs;

2) Train several NEW weighting models;

3) Select the "best" model;

4) Test the selected model for performance.

In traditional weighting methods, the weighting vector at time $\tau + h$ almost always depends on in-sample forecast errors, measured over the sample $\mathbf{y}^\tau$. This dependence may be represented by a function f comprising the series realization ($\mathbf{y}^\tau$) and the available in-sample forecast vectors ($\hat{\mathbf{y}}$):

$$\hat{\mathbf{w}}_{\tau+h|\tau} = f(\mathbf{y}^\tau, \hat{\mathbf{y}}_{t|t-h}), \tag{7}$$

where

$$h \leq H < t \leq \tau \tag{8}$$

It is worth saying that the restriction $H < t$ guarantees that the forecast origin t-h is always positive. For example, to deliver 12-steps ahead weighting vectors, we should start the generation process at time $t = 13$, so that we have the first in-sample 12-steps ahead forecast made at time $t = 1$.

In the NEW framework, (7) is replaced by (9), where G represents a properly trained MLP neural network. From the statistical viewpoint, MLPs are non-linear regression models with great function approximation capabilities. From the computational (artificial) intelligence viewpoint, MLPs are data models with *learning* and *generalization* characteristics, allowing for inference based on training examples. Theoretically, both views allows for the construction of presumed complex forecast weighting functions.

$$\hat{\mathbf{w}}_{\tau+h|\tau} = G(\hat{\mathbf{y}}_{\tau+h|\tau}, h), \tag{9}$$

where

$$h \leq H \tag{10}$$

Forming training pairs to feed the NEW MLPs constitutes this methodology core. The process considers the following variables: in-sample *forecast vector* ($\hat{\mathbf{y}}^*$), *forecast horizon* (h) and *target weighting vector* ($\hat{\mathbf{w}}^*$). (We here use the asterisk (*) symbol to distinguish between the *training* and *test* phases for the framework.) When in-sample, for each time $t > H$, for each forecast horizon $h \leq H$, a given input vector $\{\hat{\mathbf{y}}^*_{t|t-h}, h\}$ can be associated with a target weighting vector $\hat{\mathbf{w}}^*_{t|t-h}$, considered to be optimal for the point realization of the series at hand (y_t) (Fig.1). The most straightforward approach to form target weighting vectors is by constrained least squares optimization (Timmermann, 2006; Gill et al., 1984).

Once the training pairs are set, the training phase may be conducted. This is done with statistical care, as the available algorithms are non-exact and depends on the initialization procedure. For that matter, we here use the training policy known as *repeated holdout* (Witten and Frank, 2005): a portion of the training sample is separated, forming a new (sub)sample known as *validation* set. The remaining portion, known as *estimation* set, is used to train several neural network candidate models, each of which differing in the number of hidden neurons (p) and initial parameters (β) values. The best model is selected using some performance criterion over the validation set.

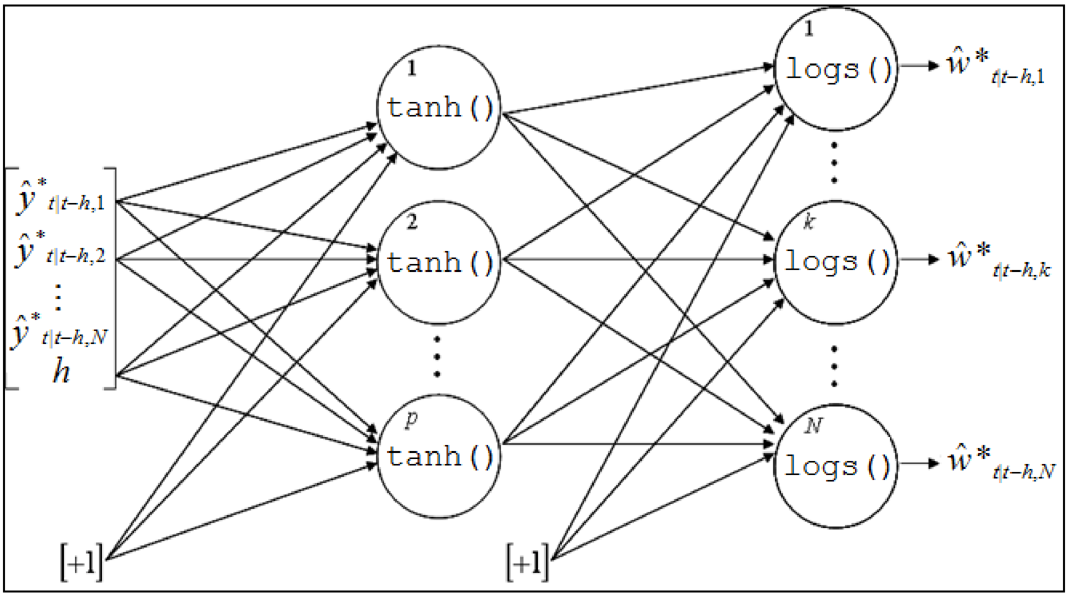

Fig.1. MLP neural network model for the NEW framework: each connection (link) represents a synaptic weight $\beta_{i,j}$. Given two consecutive layers, the indexes *i,j* define the link from processor *j* in the first layer to processor *i* in the second layer. The index *j=0* stands for a fixed *bias* input [+1]. *Tanh*() is the hyperbolic tangent function and *logs*() is the logistic sigmoid function.

3.2. Limiting Forecasters

The NEW framework focuses on convex combinations. One aspect of this combination type is that the magnitude of the consensual forecast is limited by the most extreme individual forecasts, as in Fig.2. Looking at this aspect as a potential weakness, we propose a convex combination paradigm where each individual forecaster is replaced by two new ones, each of which accounting for the bounds of the 95% confidence interval encompassing the original forecast. Those new forecasters are called *limiting forecasters* and are labeled with the original forecaster name – e.g. Multiplicative Holt-Winter (HW) or ARIMA Box & Jenkins (BJ) (section 4) – followed by a "+" (upper bound) or a "-" (lower bound) suffix, as in HW+, HW-, BJ+ and BJ- (Fig.3). The 95% confidence interval is (approximately) defined as the interval of ± 2 standard deviations from the original forecasts. The standard deviation is here said to be constant, computed as the square root of the mean squared error ($MSE^{1/2}$), taken in-sample. From the practical perspective, when forming limiting forecasters, care should be taken to avoid negative prediction values (when this is non-desirable, like in sales series).

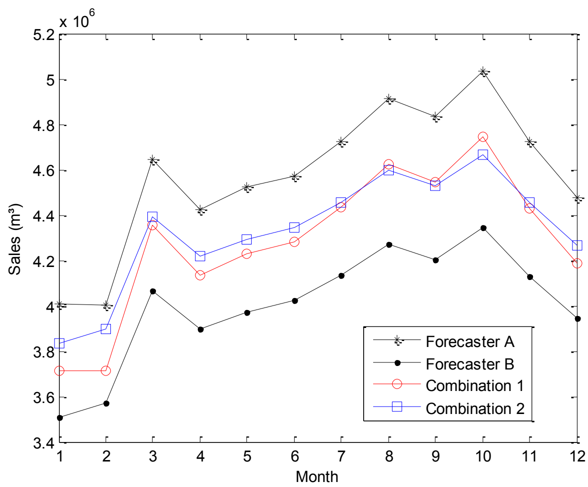

Fig.2. Convex combinations between forecasters A and B generate forecasts constrained to the region they (A and B) delimit. Two possible combinations are shown.

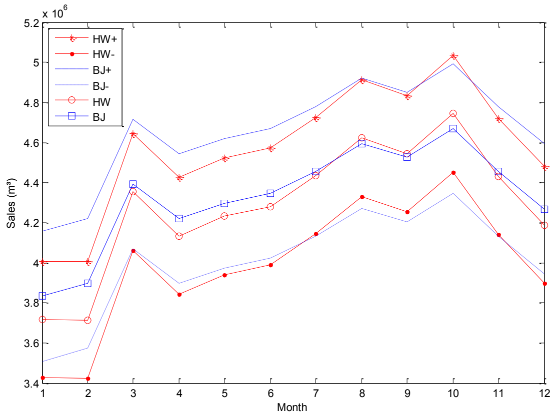

Fig.3. *Limiting forecasters*' formation: the original forecasters HW and BJ are replaced by four new ones: HW+, HW-, BJ+ and BJ-. They represent the 95% confidence interval for the original forecasts. This procedure tries to enhance convex combinations.

The idea of using limiting forecasters aims at enhancing the resultant convex combination. Although there is no practical restriction in employing this concept – other than avoiding negative values – there can be some theoretical harm when considering (linear) *diversification*: two limiting forecasters from the same source are perfectly correlated, i.e., the correlation coefficient (Kachigan, 1986) for the forecasting errors of HW+ and HW- (or BJ+ and BJ-) equals one. Nevertheless, this restriction loses power within the NEW models, due to the flexible non-linear nature of the neural networks.

4. Case Study

We here analyze Brazil's monthly sales of diesel fuel (DIESEL), regular motor gasoline (GASOLINE), liquefied petroleum gas (LPG) and kerosene type jet fuel (JET), as published by the local government agency "Agência Nacional do Petróleo, Gás Natural e Biocombustíveis" (ANP, 2012). Fig.4 shows these series.

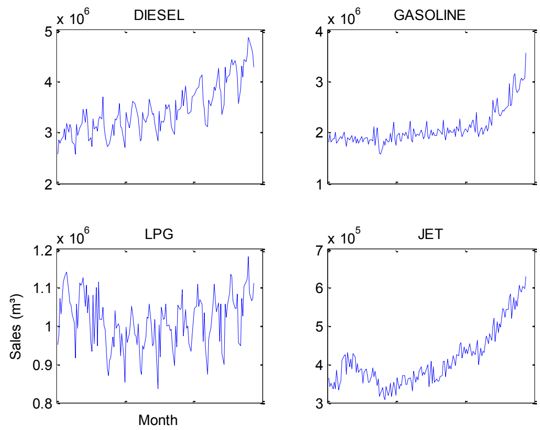

Fig.4. Petroleum products: monthly sales in Brazil, between Jan/2000 and Dec/2011.

4.1. Available Forecasters

Any combination scheme relies on the availability of individual forecasters. These forecasters are here chosen to derive from the methods/methodologies listed below (Kachigan, 1986; Makridakis et al., 1997; Harvey, 1991).

1) Multiplicative Holt-Winters (HW);

2) ARIMA Box & Jenkins (BJ).

Our choosing criterion for individual forecasters was three-fold: (i) to have standard, well known and easy to implement models, (ii) to be able to handle growth and seasonality, recurring aspects over real-world time series and (iii) to have different statistical natures, allowing for diversification.

4.2. Performance Measure

Our experiments use the Symmetric Mean Absolute Percentage Error (SMAPE) (11) as main performance measure. Time series-related papers often use the SMAPE, measured in percentages (%), mostly due to its usual application as an official performance measure in forecast competitions (Makridakis and Hibon, 2000).

$$SMAPE = \frac{1}{H}\sum_{h=1}^{H}\frac{\left|y_{\tau+h}-\hat{y}_{\tau+h|\tau}\right|}{\left(\frac{\left|y_{\tau+h}\right|+\left|\hat{y}_{\tau+h|\tau}\right|}{2}\right)}.100\% \tag{11}$$

4.3. Hypothesis Testing

To deepen our analysis, hypothesis tests are applied over part of the gathered results. The selected tests – *t* (Kachigan, 1986), *sign* and *Wilcoxon* (Gibbons, 1992) – verify the following null hypothesis (H_0): "the *mean* (for the t-test) or the *median* (for the sign and Wilcoxon tests) of the performance difference between two forecasting methods equals zero". The t-test is parametric and assumes that the performance difference follows a normal distribution, which sometimes is not the case, mostly for small samples (< 30). To check the validity of the normal assumption, Q-Q graphs (Johnson and Wichern, 2007) or specific tests, such as the *Jarque-Bera* (Cromwell et al., 1994), may be employed. Both sign and Wilcoxon tests are nonparametric, dismissing the normal assumption. Those tests have already been used to compare forecasting methods (Flores, 1986 and 1989).

We here propose a hypothesis testing architecture in which performance differences are measured "step-by-step" along the forecast horizon. In other words, as the horizon ranges from 1 to H steps ahead, H performance differences are taken.

4.4. Results

Aided by the *Forecast Pro* software (BFS, 2013), a well-known commercial forecasting tool, the best possible HW and BJ models (section 4.1) are automatically defined for the presented dataset. For each series, the last 12 months are hold-out for testing and the fitted models generate forecasts up to 12 steps ahead (with no parameter re-estimation at each step).

Once the individual forecasters are set, we can proceed to the combination experiments with the NEW framework. The assessed combinations act on the following groups of individual forecasters: (i) HW/BJ and (ii) HW+/HW-/BJ+/BJ- (limiting forecasters' combination, section 3.2). These experiments are implemented using the MATLAB software (MATHWORKS, 2013).

Table I shows the per series test set errors – SMAPEs over 12 months out-of-sample – considering the individual forecasters, HW and BJ, and the "best" found NEW combinations. The "best" NEW scheme for a given series is chosen *via* least SMAPE, considering a validation set mounted with the last in-sample 24 months. From the table, it can be seen that the NEW framework outperforms the individual models in 4 out of 4 cases, delivering the least mean error (3.13) and standard deviation (2.48). In fact, the lower the mean error, the higher the method's average accuracy. The lower the standard deviation, the lower the method's risk (uncertainty).

Table II depicts the "best" found NEW combination models. The table shows that, considering the present case study, the best combinations are always based on limiting forecasters. Fig.5 shows how the weighting vectors generated by these winning combinations evolve out-of-sample.

TABLE I
PER SERIES PERFORMANCE

series	*NEW*	*HW*	*BJ*
DIESEL	**2.73**	3.02	2.82
GASOLINE	**6.74**	12.03	10.46
LPG	**1.64**	1.66	1.97
JET	**1.40**	3.57	2.25
Mean	**3.13**	5.07	4.37
SD	**2.48**	4.71	4.07

Per series out-of-sample SMAPEs (%): bold-faced numerals indicate the best observed results.
The final rows contain general means and standard deviations (SD).

TABLE II
"BEST" NEW MODELS

series	*COMBINED EXPERTS*	*NEURAL NETWORK DETAILS*
DIESEL	HW+/HW-/BJ+/BJ-	20 hidden neurons 1500 candidate models 1.71h
GASOLINE	HW+/HW-/BJ+/BJ-	20 hidden neurons 1500 candidate models 1.93h
LPG	HW+/HW-/BJ+/BJ-	5 hidden neurons 1500 candidate models 1.53h
JET	HW+/HW-/BJ+/BJ-	5 hidden neurons 1500 candidate models 1.87h

"Best" NEW model per series: the *neural network details* column brings the selected number of hidden neurons, the number of candidate models trained and the total training time. Experiments were carried out with the MATLAB software, using an Intel i5 processor with Windows 7 and 4GB of RAM.

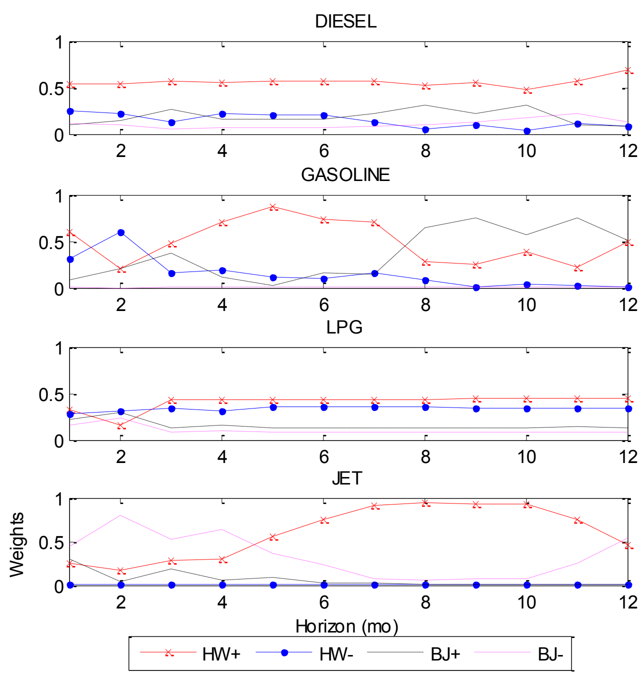

Fig.5. Weighting vectors evolution along the forecast horizon.

Table III and Fig.6 show the SMAPEs' averages' evolution along the test horizon (h). As each SMAPE is an accumulated mean, the table's last row accounts for the whole forecast horizon.

For comparison purposes, Table IV shows the step-by-step average performance differences for the NEW methodology, each difference being calculated as "NEW's SMAPE *minus* other's SMAPE". As difference values become more negative, the average NEW performance gets better. Here, the average NEW's performance is clearly distinguished for $h \geq 5$.

TABLE III
AVERAGE PERFORMANCES

h	*NEW*	*HW*	*BJ*
1	3.80	2.48	3.74
2	2.96	2.74	2.57
3	3.45	3.64	3.16
4	4.01	4.28	3.83
5	3.63	4.42	3.76
6	3.33	4.36	3.55
7	3.18	4.34	3.51
8	3.21	4.66	3.91
9	3.12	4.86	4.14
10	3.10	4.84	4.10
11	3.04	4.97	4.24
12	**3.13**	**5.07**	**4.37**

Average SMAPE (%) along the forecast horizon *h* (mo): each result is a cumulative average for the petroleum products series. The final row (bold-faced) represents the complete test set.

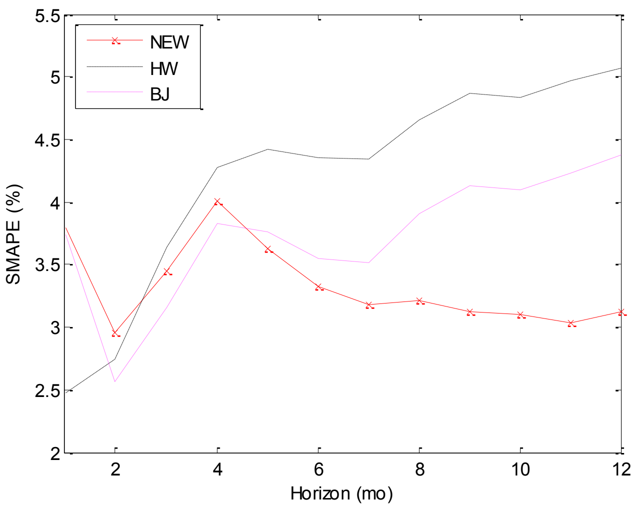

Fig.6. SMAPEs' averages' evolution along the forecast horizon.

Considering significance level of 5%, Table V shows the conclusions for the hypothesis tests that verify whether the performance differences' means/medians equal zero (section 4.3). On that matter, we define *positive conclusion* as a test result pointing that "the NEW combination is better". There are 4 out of 6 positive conclusions favoring the NEW framework. The remaining 2 test responses show equivalence.

TABLE IV
AVERAGE PERFORMANCE DIFFERENCES

h	*HW*	*BJ*
1	1.32	0.05
2	0.22	0.39
3	-0.19	0.29
4	-0.28	0.17
5	-0.79	-0.13
6	-1.03	-0.23
7	-1.16	-0.33
8	-1.45	-0.69
9	-1.75	-1.02
10	-1.74	-1.00
11	-1.93	-1.20
12	-1.95	-1.25
Mean	**-0.89**	**-0.41**
Median	**-1.10**	**-0.28**

Average performance differences relative to the NEW framework, along the forecast horizon *h* (mo): a more negative value indicates a better NEW average. The final rows show means and medians. If one mean/median (statistically) equals zero, the comparing methodologies are said to be equivalent.

TABLE V
HYPOTHESIS TESTING

reference model	*p-value*	*low*	*upp*	*normal?*	*conclusion*
			t-test		
HW	0.0103	-1.53	-0.26	YES	NEW is better
BJ	0.0361	-0.79	-0.03	YES	NEW is better
			sign test		
HW	0.0386	-1.75	-0.19	YES	NEW is better
BJ	0.3877	-1.02	0.17	YES	Equivalent
			Wilcoxon test		
HW	0.0161	-1.59	-0.21	YES	NEW is better
BJ	0.0771	-0.84	0.03	YES	Equivalent
			***positive conclusions*: 4 (out of 6)**		

For each test, the following information is shown: *p-value*, lower (*low*) and upper (*upp*) confidence limits for the observed means/medians, the Jarque-Bera normality test status for the SMAPE differences distribution (*normal?*) and the final *conclusion*. The last row shows the number of *positive conclusions* (at 5% significance level).

5. Conclusion

Focusing on convex forecast combinations, this paper proposes a new weight generation framework, called Neural Expert Weighting (NEW). The framework generates dynamic weighting models based on neural networks. During the NEW development, we also present the idea of limiting forecasters as a possible way to enhance combination performance.

The present work comprises forecasting experiments with 4 petroleum products sales time series. The gathered results sustain the following conclusions:

1) There are advantages in using the NEW framework;

2) Using limiting forecasters can deliver good results.

The first conclusion – "there are advantages in using the NEW framework" – meets our main objective: to develop weighting models that can enhance traditional weighting procedures, leading to better convex forecast combinations. The NEW framework works as a combination methodology in that, in many aspects, it outperforms the individual assembled forecasters (experts). When considering the 4 sales time series with full forecast horizon (12 months), the NEW models deliver the least mean error (3.13) and standard deviation (2.48). The lower the mean error, the higher the method's average accuracy; the lower the standard deviation, the lower the method's risk. Moreover, accuracy superiority holds not only for the overall analysis, but for each series alone.

When comparing results in a step-by-step manner along the test samples' horizons (12 months), the NEW framework is still attractive, as accounted by the total positive conclusions from the hypothesis tests. In addition, there seems to be some evidence that the average NEW's performance is clearly distinguished for higher portions of the forecast horizon ($h \geq 5$ in this case). This fact may be further explored in future works.

The second conclusion – "using limiting forecasters can deliver good results" – arises from the fact that the winning combination schemes used *limiting* versions of the original forecasters. Despite the reasonable evidence, this fact – alongside with the aforementioned forecast horizon matter – may also be further explored in future works. Other extensions can follow these paths: (i) testing different time series frequencies (other than monthly) and (ii) combining other types of forecasting models.

References

ANP (2012), Dados estatísticos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (in portuguese). Available: http://www.anp.gov.br/
Bates, J. M. and Granger, C. W. J. (1969), The combination of forecasts, *Operations Research quarterly*, vol. 20, pp. 451-468.
BFS (2013). Forecast Pro software. Available: http://www.forecastpro.com/
Bishop, C. M., *Neural networks for pattern recognition*, Oxford University Press, UK, 1995.
Flores, B. E. (1986), Use of the sign test to supplement the percentage better statistic, *International Journal of Forecasting,* vol. 2, pp. 477-489.
Flores, B. E. (1989), The utilization of the Wilcoxon test to compare forecasting methods: A note, *International Journal of Forecasting,* vol. 5, pp. 529-535.
Gibbons, J. D., *Nonparametric statistics: an introduction*, Sage Publications, [S.l.], 1992.
Gill, P. E. *et al.* (1984), Procedures for optimization problems with a mixture of bounds and general linear constraints, *ACM Transactions on Mathematical Software*, vol. 10, pp. 282-298.
Harvey, A. C., *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, [S.l.], 1991.
Hibon, M. and Evgeniou, T. (2005), To combine or not to combine: selecting among forecasts and their combinations, *International Journal of Forecasting*, vol. 21, pp. 5-24.
J. B. Cromwell, J. B., Labys, W. C. and Terraza, M., *Univariate tests for time series models*, Sage Publications, [S.l.], 1994.
Johnson, R.A. and Wichern, D. W., *Applied multivariate statistical analysis* 6th ed., Prentice Hall, New Jersey, 2007.

Kachigan, S. K., *Statistical analysis: an interdisciplinary introduction to univariate & multivariate methods*, Radius Press, [S.l.], 1986.
Makridakis, S. G. and Hibon, M. (2000), The M3-competition: results, conclusions and implications, *International Journal of Forecasting* (special issue), vol. 16, pp. 451-476.
Makridakis, S. G., Wheelwright, S. C. and Hyndman, R. J., *Forecasting: methods and applications* 3rd ed., Wiley, [S.l.], 1998.
MATHWORKS (2013). MATLAB software. Available:
http://www.mathworks.com/products/matlab/
Reed, R. and Marks, R. J., *Neural smithing: supervised learning in feedforward artificial neural networks*, The MIT Press, [S.l.], 1999.
Sánchez, I. (2008), Adaptive combination of forecasts with application to wind energy, *International Journal of Forecasting*, vol. 24, pp. 679-693.
Timmermann, A., Forecast combinations, in G. Elliott, C. W. J. Granger and A. Timmermann (Eds.), *Handbook of economic forecasting* vol. 1,North-Holland, [S.l.], pp. 135-196, 2006.
Witten, I. H. and Frank, E., *Data mining: practical machine learning tools and techniques* 2nd ed., Morgan Kaufmann, [S.l.], 2005.