Published in "Proceedings of International Conference on Time Series and Forecasting (ITISE 2018)", 2018, 19-21 September, Granada, Spain, which should be cited to refer to this work.

Evaluation of forecasting approaches using hybrid multi-criteria decision-making models

Yvonne Badulescu[1,2]✉ and Naoufel Cheikhrouhou[1]✉

[1]Geneva School of Business Administration, University of Applied Sciences Western Switzerland (HES-SO), 1227 Carouge, Switzerland
[2]Faculty of Business and Economics (HEC), University of Lausanne (UNIL), Switzerland
✉Yvonne.Badulescu@hesge.ch
Yvonne.Badulescu@unil.ch
✉Naoufel.Cheikhrouhou@hesge.ch

Abstract. The demand forecast is the most influential input into enterprise activities planning thus creating a challenging issue for Demand Planning experts in model selection. Models including quantitative, qualitative and hybrid forecasting methods have been developed and are widely used. The literature reveals the use of several case-dependent error measures to evaluate forecasting accuracy, however, these performance indicators may at times differentiate in results making it more difficult in determining the most appropriate forecasting model for the users' needs. This paper presents the development of two hybrid multi-criteria decision making approaches, AHP-TOPSIS and ANP-TOPSIS, to evaluate and rank the relative performances based on error measures of alternate forecasting models. Validation is provided through an industrial application using empirical data from a plastic bag manufacturer based on five models; three regression forecast models and two hybrid demand forecast models using expert judgement. Results illustrate that subjective adjustment by experts of mathematical forecasts consistently gives a higher ranking due to proximity to the ideal solution, and that collaborative adjustment limits the risk of outliers due to forecasting errors that could be done by a single decision maker.

1 Introduction

There has been much research on the development of forecasting models to accurately determine the real future demand. Quantitative methods rely on mathematical approaches that can lead to reliable forecasts by extrapolating regular patterns in time-series. In addition to regression models, key actors in enterprises frequently obtain knowledge and insights about future non-periodic events that are expected to strongly influence the demand. Consequently, hybrid forecasting methods are employed, using single and collaborative judgmental approaches. Within this context, researchers and practitioners alike are confronted to the issue of selecting the most appropriate forecasting model and conventionally use goodness-of-fit performance criteria in determining the model which produces a forecast with the smallest error versus the actual demand

[1]. These error measures should meet some criteria identified in [2] and show the importance of variety, reliability, ease of interpretation, clarity of presentation, and support of statistical evaluation. However, a single error measure is insufficient in evaluating different forecasting techniques as each measurement may provide diverging results as well as only illustrate specific aspects of the forecast, such as, the impact of an outlier. Moreover, the hybrid approach used may provide a focus on improving only a particular pattern, which leads to substantial changes of some error measures and negligible improvements of some others. Thus, there is a need for a method to allow decision makers to select the most appropriate forecasting model within a group of alternatives based on the varying weights of importance they place on a range of error measures.

Therefore, the question arises: which method is the most appropriate to select in order to have the most reliable forecast for future demand? In an attempt to find an answer to this question, this work proposes the development of a multi-criteria decision making (MCDM) approach to evaluate different forecasting models based on a selected number of widely used error measures in forecasting, their interdependencies and their influence on model selection.

This paper is divided into 6 sections: The next section provides an overview of the state of the art in MCDM application in forecast selection. Section 3 describes the proposed framework to answer our research question followed by an empirical analysis using industrial data from a plastic bag manufacturer in Section 4. Section 5 presents the results and discussion, followed by the conclusion and suggestions for future work in Section 6.

2 State of the Art

Hybrid forecasts, combining both quantitative results and qualitative information, have showed substantial improvements in forecasting accuracy [3]. A systematic approach is developed by [4] to structure and integrate human knowledge of contextual factors into demand forecasting. The statistical forecast based on ARIMA method is considered as a basis and the structured knowledge of the experts is provided to adjust the initial forecasts based on four factors that represent most types of events: transient, transfer, trend change, and quantum jump factors. Further, as sequel to improve the forecasting accuracy, these factors corresponding to the human judgments forecast adjustment are grouped and structured to determine the forecast adjustment due to collaborative human judgement and are subsequently evaluated through a fuzzy inference system which show substantial benefits in the improvement of forecasting accuracy [3]. The collaborative process consists of integrating the judgements of several forecasters and structuring the information using complementarities of the different perceptions. The results of both [3, 4] illustrate an improvement in forecast accuracy based on the Mean Absolute Error (MAE) and the Mean Absolute Percentage Error (MAPE). However, the choice of a forecasting model cannot be solely selected based on any one specific error measure [5]. The MAE is shown to be weak to outliers as a strong error at one point of the data series will skew the mean [6]. In addition, the MAPE is not a symmetrical error measurement in that equal errors above the actual value result in a

greater average percentage error than those below the actual value. The MAPE is therefore best used when dealing with positive actual observations [7]. The literature highlights the importance of considering several error criteria in evaluating the performance of a forecasting model.

In [1], the authors propose a framework using MCDM methods in selecting the most appropriate regression forecasting models based on a tradeoff between several error measures as their criteria, and apply ELECTRE III, PROMETHEE I, and PROMETHEE II to evaluate the most appropriate forecast for oil prices. They find that the best performing models, a linear regression and exponential smoothing model, are not sensitive to the criteria weighting or the different MCDM methods used. The authors in [8] investigate the use of the PROMETHEE approach for selection of the most appropriate classification forecast algorithms. Although easier to use than ELECTRE, the PROMETHEE method lacks clarity in weight determination for the criteria and ignores potential interdependencies between them [9].

Often hybrid MCDM methods are employed, such as AHP and ANP paired with TOPSIS, which allows for easier ranking between alternatives [10]. A fuzzy AHP-TOPSIS method, including expert weighting, is used in [11] to determine the best alternative between a selection of collaborative software available on the market. Raut et al. [12] integrates the Balanced Scorecard (BSC) with a fuzzy AHP and fuzzy TOPSIS framework to determine the degree of sustainability in the banking industry that was implemented in the six largest commercial banks in India. Wang [13] proposed the hybrid method of decision making trial and evaluation laboratory technique (DEMATEL) with ANP (DANP) to decide upon interactive trade strategy to be adopted for Taiwan. Both [13, 14] employ DEMATEL in order to build a network relations map by investigating the interrelation among aspects and criteria. Chiu et al. [14] assess e-store strategies, with respect to Marketing and Customer Service, by also employing DANP with VIKOR methods to rank the alternatives from a set of often conflicting criteria. The hybrid DANP-VIKOR model aided in determining the effect of e-store management on sales and resulted in recommendations for strategy improvement. The main difference between VIKOR and TOPSIS methods is the aggregating function for determining the ranking of the alternatives. TOPSIS ranks the alternative based on the additive combination of the best and worst distances from the ideal solution whereas the VIKOR method takes into account the relative importance of the distances from the ideal solution by balancing the total and individual satisfaction. The use of n-dimensional Euclidean distance in the TOPSIS method accounts for this balance between the individual and total satisfaction [15].

Contrary to other MCDM methods such as the Multi-Attribute Utility Theory (MAUT) and Multi-Attribute Value Theory (MAVT), AHP and ANP apply pairwise comparison to compare alternatives as well as estimate weighting to the criteria and priority scales [16], and are therefore easy to use [9]. The MAUT/MAVT methods require high precision on the specific criteria weights which prove difficult in real-life circumstances [9]. On the other hand, the AHP and ANP methods are susceptible to rank reversal at the end of the process which could result in the final ranking to be reversed in order. Using TOPSIS addresses the issue of rank reversal when a non-optimal alternative is introduced [17], however the TOPSIS method alone does not consider criteria interrelationships nor provide an easy method of determining criteria weights, often being paired with the AHP and ANP methods [18]. The authors in [19] use a fuzzy AHP alongside

the TOPSIS method to rank fifteen Turkish cement firms based on their financial performance as well as subjective judgmental information. In addition, [20] also propose a hybrid approach using AHP and Fuzzy TOPSIS to rank banking performance in Iran. The hybrid methods AHP-TOPSIS and ANP-TOPSIS have also been applied to support the decision making process for personnel selection in a manufacturing firm [21]. The research concludes that the hybrid approaches of AHP-TOPSIS and ANP-TOPSIS are robust MCDM techniques to evaluate performance based on a set of defined criteria. These hybrid techniques have not yet been used in the evaluation of demand forecasting models based on error measurement criteria. This paper aims to bridge this gap in the literature by proposing the AHP-TOPSIS and ANP-TOPSIS models, to evaluate and rank forecasting models based on multiple error measures as the decision criteria.

3 Proposed Framework

The objective of this paper is to develop an MCDM approach to evaluate and rank quantitative and hybrid (quantitative and qualitative) demand forecasting models with a particular focus on performance measures. Since different forecasting methods usually lead to different error measurements, the criteria selected in the MCDM method are:

- The Mean Error (ME), which indicates whether a forecast is biased, however it is possible that a negative error on one data point would counterbalance a positive error on another data point. [22] suggests that to counter this last effect, the Median can be used over the mean;
- The Mean Absolute Error (MAE), which measures the absolute error however but may skew the mean when confronted with large outliers [6];
- The Mean Percentage Error (MPE) which is based on the actual values rather than the absolute values and therefore is a good measure of the relative size and direction of the bias.
- The Root Mean Squared Percentage Error (RMSPE) takes only positive values due to the squaring function and therefore provides an average relative size of the error. On the other hand, large outliers will dramatically impact the measurement [23].
- The Root Mean Squared Error (RMSE), which is representative of the size of a "typical" error. RMSE gives extra weight to large errors due to the squaring function and is sensitive to scale [6]. In addition, it has been observed that the results frequently differ when applied to various sets of data [23].
- The Mean Absolute Percentage Error (MAPE) is one of the most common error measurements in demand forecasting. MAPE is pulled upward by asymmetrical distributions and outliers. It has a minimum of 0 but no upper boundary. Additionally, MAPE is unit-free and the result is given as a percentage and is best used when dealing with positive actual observations [7].
- The R-squared (R^2) is the coefficient of determination that represents the proportion of variability in a data set that is accounted for by the forecasting model. It provides a measure of how well future outcomes are likely to be predicted by the model. An R^2 measure of 1.0 can be understood in that the regression line perfectly fits the data. Negative values of R^2 may occur when fitting non-linear trends to

data. In seasonal time-series with non-linear trends, we can expect to see negative values of R^2.

These criteria have several key aspects to take into account. First of all, the criteria are for most cases incomparable, in the sense that they are not in the same units (for example, RMSPE and MAPE are in percentages, and R^2 is a ratio and therefore has no unit). Secondly, a common point is that the overall goal is to minimize the set of criteria, as a higher value means a larger error in general. R^2 differs from the others as it does not represent a mean computed using errors however it measures how well future outcomes are likely to be predicted by the model. The weights and importance corresponding to each criteria depends upon the chosen MCDM method.

In this paper, two hybrid MCDM techniques are proposed and developed: AHP-TOPSIS and ANP-TOPSIS, selected due to their robustness (based on the constant number of calculation steps for TOPSIS regardless of the number of attributes), scalability and possibility to integrate interdependencies (in the case of ANP) [9], as well as ability to consider both quantitative and qualitative information [11]. They are used to evaluate and compare five different forecasting techniques. Recognition of the dependence among criteria is also considered to calculate the criteria weights in using ANP as the same data is used to calculate the error measures. In AHP-TOPSIS, the interdependence is omitted. These alternatives are then evaluated with respect to criteria and then ranked.

4 Case study using industrial data

The case study is based on data and additional information collected from Company X, a plastic bag manufacturer in the south of Spain. The forecast experts are asked to analyze the polyethylene bag market to provide a demand forecast. The time-series used are composed of the aggregated monthly demand collected over a period of 36 months (2004 to 2006). The objective is to plan the demand for the year 2007. The three main customers of Company X are all supermarkets. Each expert is invited to analyze the historical data and the influencing factors in the plastic bag demand and forecast the specific events relying on their knowledge.

4.1 Forecasting using empirical case study

The historical data is plotted as a time-series and shows a strong linear trend and seasonality (with peaks in summer and winter - due to demand increases in plastic bags during the months before summer and Christmas holidays, cf. [4]). The forecasts (alternative solutions) are determined using five different techniques: the Holt-Winter decomposition method with multiplicative seasonality [24], the ARIMA and SARIMA Methods, and two Hybrid Forecasting processes that include human judgement: ARIMA integrating single judgmental adjustment, and ARIMA integrating collaborative judgmental adjustment using a team of three experts.

With comparison to the actual data, Holt-Winters gives results that follow the demand seasonality, but has a lower amplitude than the actual data. The Holt-Winters method provides a forecast with few outliers shown by the close values of the MAE and RMSE, 75.9 and 92.5 respectively, in Table 1. Figure 1 illustrates that this forecast has values

that are relatively close (both above and below) to the actual results. The error measures in Table 1 quantify these discrepancies and show the ME and MPE (both non-absolute measures) are relatively low at 36.9 (from a demand of >1200) and 3.27%, respectively. In addition, the R^2 measure of 0.61 shows us that the fit to the regression is not too low for a time-series presenting strong seasonality.

The second forecasting method analyzed is the ARIMA (5,0,4) method (based on [4]). Figure 1 shows that the ARIMA method follows the actual demand relatively well, including the seasonality peaks (albeit with a slight delay on the second peak). The fit is better at the beginning of the time-series than at the end where there is a more noticeable positive error. The visual analysis is supported by the error measures in Table 1, where the ME and MPE are very low (1.33 and 0.51% respectively) suggesting an equal distribution of positive and negative errors. The RMSE is marginally larger than the MAE, which can explain the larger positive error in the later months. The R^2 measure of 0.24 is very low attributing to the high variance of error along the time-series.

The third model is the SARIMA forecast which introduces seasonality to the ARIMA (5,0,4) model used as the basis. The three parameters representing the orders of the seasonal autoregressive and moving average parts of the model are determined by simulating various configurations, using the R programming language and free software, to yield the lowest ME and RMSE, which result in (1,1,1). The results for the SARIMA (5,0,4)(1,1,1) are illustrated in Figure 1 showing a strong adherence to the real data and its seasonality. As the time-series in this case study explicitly shows a seasonal character, it was expected that SARIMA provides a better forecast than ARIMA. The forecast initially has some difficulty following the curve at the beginning of the 12-month period, followed by a very good fit in the later months (June to October) and then another deviation in the final two months of the forecast.

Analysis of the error measures of the SARIMA forecast in Table 1 shows a relatively low ME and MPE of 17.50 and 1.70%, respectively, showing a near-equal positive and negative error distribution. In addition, the closeness in the values of MAE and RMSE indicates the absence of outliers.

Table 1. Error Measures per forecasting model

Error Measures	Holt-Winter	ARIMA	SARIMA	Single Adjustment + ARIMA	Collab Adjustment + ARIMA
ME	36.94	1.33	17.5	-52.42	5.58
MAE	75.99	112.17	92	97.42	33.25
MSE	8562.4	16821	11473.33	29564.92	1836.08
RMSE	92.53	129.7	107.11	171.94	42.85
MPE	3.27 %	0.51 %	1.70 %	-3.29 %	0.47 %
MSPE	0.54 %	0.95 %	0.63 %	1.22 %	0.09 %
RMSPE	7.35 %	9.74 %	7.97 %	11.06 %	3.04 %
MAPE	5.78 %	8.19 %	6.68 %	6.66 %	2.36 %
R2	0.61	0.24	0.48	-0.34	0.92

The final two hybrid forecasting methods utilize both time-series and qualitative information to determine the forecast. The first uses ARIMA(5,0,4) single judgement adjustment in which the expert judgement is integrated in a structured manner as complementary information to the ARIMA(5,0,4) forecast. It can be seen from Figure 2 that this forecast has a very good fit with the actual data, including month-over-month trend, seasonality, and peaks. Conversely, this forecast has a very strong negative outlier error (-549) due to human error when the decision maker was adjusting the mathematical forecast. In addition, there are very large differences observed between the MAE and RMSE of 97.42 and 171.94, respectively (Table 1) which shows the strong impact of the outlier. However, as the remaining fit is very good, other error measures mitigate the outlier's impact. The R^2 measure is negative at -0.34, which is expected due to the extremely high variation in errors due to the outlier and thus non-linear error trend.
The second hybrid forecasting method utilizes the collaborative expert judgement of several people [3] and is based on ARIMA(5,0,4). Figure 2 shows a very strong overall fit between the forecast and the actual data. The only minor differences observed are slight positive errors at the beginning of each seasonal rising slope (March and July). Both the ME and MPE are small showing quasi-equal distribution of positive and negative errors in Table 1, and the MAE and RMSE are also very low due to the impact of the two small peak errors. The R^2 measure of 0.92 shows that there is an extremely good fit between forecast and data: there is very little difference in error trend along the time-series.

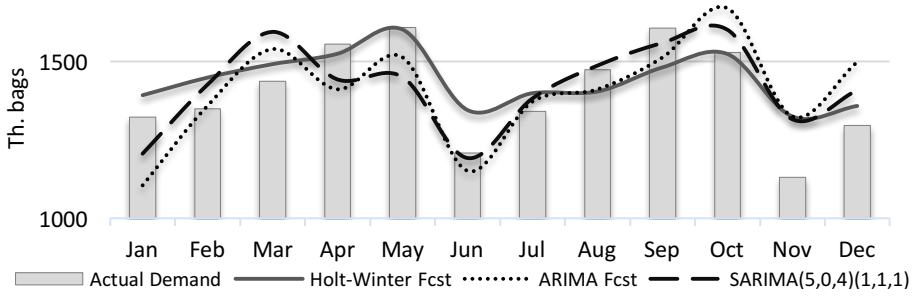

Fig. 1. Demand vs Quantitative Forecasts for 12 months horizon

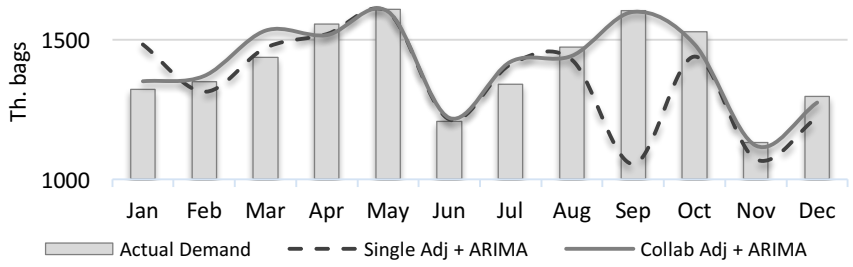

Fig. 2. Demand vs Hybrid Forecasts for 12 months horizon

4.2 Solving the MCDM Problem

We develop two hybrid decision-making models: AHP-TOPSIS and ANP-TOPSIS and apply them to compare the forecasting models.
For both AHP and ANP, a pairwise comparison matrix is formed by experts using Saaty's 1-9 scale [25] on the identified error measures criteria. Initially, no interdependencies are considered. Their weights in terms of priority are determined and presented in Table 2. The consistency ratio (CR) for this case is 0.0935, therefore considered acceptable as it is < 0.1 [26]. Priorities are selected based on the amount of useful information conveyed by the criteria. The RMSE has a little higher overall priority to MAE due to its use in identifying the presence of outliers (square power). Following the pairwise comparisons, a normalized weight vector w is determined by calculating the eigenvectors of the priority matrix [21].

Table 2. Criteria priorities & weights using AHP

Criteria	ME	MAE	RMSE	MPE	RMSPE	MAPE	R^2	Weights (w)
ME	1	1/5	1/4	1/3	1/5	1/5	1	0.0464
MAE	5	1	1/3	1	1/3	1/3	1	0.0975
RMSE	4	3	1	3	1	1/3	1	0.1666
MPE	3	1	1/3	1	1/3	1/5	1	0.0802
RMSPE	5	3	1	3	1	1/3	1	0.1726
MAPE	5	3	3	5	3	1	1	0.3075
R^2	1	1	1	1	1	1	1	0.1292

In order to consider the interdependencies between criteria for the ANP, another pairwise comparison matrix is created where the decision makers examine the impact of all the criteria on each other. In this case, interdependencies are chosen based on the way the criteria are calculated. For example, R^2 is not calculated in the same way as any other criteria, but the MAE and MPE both show components of how MAPE is calculated. The links are shown in the Figure 3.

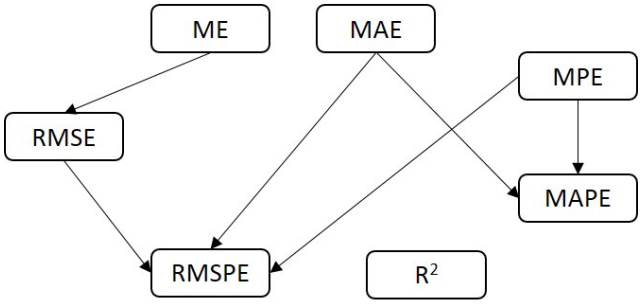

Fig. 3. Interdependencies between MCDM criteria

Table 3. Normalized interdependences and weights (ANP-TOPSIS)

Criteria	ME	MAE	RMSE	MPE	RMSPE	MAPE	R^2	Weight (w_n)
ME	1	0	0.106	0	0	0	0	0.0642
MAE	0	1	0.149	0	0	0.200	0	0.1838
RMSE	0	0	0.745	0	0.138	0	0	0.1479
MPE	0	0	0	1	0.172	0.200	0	0.1714
RMSPE	0	0	0	0	0.690	0	0	0.1190
MAPE	0	0	0	0	0	0.600	0	0.1845
R^2	0	0	0	0	0	0	1	0.1292

Using the hybrid methodology in [21], a decision matrix for the ranking of alternatives is created on the basis of the criteria per alternative. The decision table is then normalized, to allow for comparison between values, shown in Table 3.

The final part of both AHP-TOPSIS and ANP-TOPSIS hybrid models consists of determining the positive-ideal and negative-ideal solutions, then calculating each alternative's distance to them and establishing a ranking based on these distances. To establish the positive-ideal solution vector (noted V+) and the negative-ideal solution vector (noted V-), the minimum value is determined from the alternatives for each criteria. In the same way, we determine which value is maximum for V- from the alternatives for each criteria (Table 5).

Subsequently the separation measures are calculated using the Euclidian distance [21]. The separation of each alternative to the positive-ideal V+ is noted D+. Similarly, the separation of each alternative from the negative-ideal solution V- is noted D-. The relative closeness to the ideal solution is calculated as $C_i = \frac{D_i^-}{D_i^- + D_i^+}$. and the performance order is ranked (Table 6 for AHP-TOPSIS and 7 for ANP-TOPSIS). A larger index value means the better the performance of the alternative.

Table 5. Positive-ideal and negative-ideal solutions for AHP-TOPSIS and ANP-TOPSIS

Error Measures		ME	MAE	RMSE	MPE	RMSPE	MAPE	R^2
AHP-TOPSIS	V+	-0.0364	0.01676	0.00828	0.01315	-0.1139	0.0161	0.07702
	V-	0.02571	0.05653	0.13331	0.05278	0.11303	0.21353	0.02115
ANP-TOPSIS	V+	-0.0504	0.03159	0.02425	-0.1131	0.01949	0.03124	0.09315
	V-	0.03553	0.10658	0.09733	0.11227	0.07098	0.10819	-0.035

Table 6. Distances and final ranking (AHP-TOPSIS)

	Holt-Winter	ARIMA	Single Adj	Collab Adj	SARIMA
D+	0.252601085	0.21921914	0.23921742	0.14738874	0.15769996
D-	0.15807352	0.13286061	0.24201139	0.25998246	0.148354036
C	0.38491185	0.37735942	0.50290295	0.63819548	0.48473157
Rank	**4**	**5**	**2**	**1**	**3**

Table 7. Distances and final ranking (ANP-TOPSIS)

	Holt-Winter	ARIMA	Single Adj	Collab Adj	SARIMA
D+	0.253813597	0.200892915	0.177048763	0.140907385	0.211287706
D-	0.11918427	0.11944873	0.24253111	0.21469460	0.11289793
C	0.31953071	0.37287917	0.57803324	0.60374973	0.34825088
Rank	**5**	**3**	**2**	**1**	**4**

5 Results and Discussion

In both AHP-TOPSIS (Table 6) and ANP-TOPSIS (Table 7), a recurring aspect is that the hybrid forecasts combining mathematical forecasts with expert adjustments are ranked higher than pure mathematical forecasts. In addition, the collaborative expert adjustment method is ranked highest using both MCDM hybrid models. In fact, its relative closeness C to the ideal solution is much higher than the other methods, especially using AHP-TOPSIS. Single adjustment comes closer to collaborative adjustment in ANP-TOPSIS (C=0.60 to C=0.58) due to the interdependence between error measures. The results indicate that the judgmentally adjusted forecasts, whether done by a single expert or as a collaboration between a number of experts, is better than the purely statistical methods.

On the other hand, the purely mathematical models ARIMA, SARIMA and Holt-Winter models, all changed in ranking between the AHP-TOPSIS and ANP-TOPSIS methods. In particular, the models that integrate seasonality, SARIMA and Holt-Winter, indicate a stronger sensitivity to the weights of the criteria as well as their interdependencies. Conversely, ARIMA seemed to be least sensitive to the change in weights in the criteria between the two MCDM methods, even though it had the largest change in ranking relative to the other four forecasting models. Surprisingly, the Holt-Winter method is consistently ranked in the bottom two, as two of its resulting weighted criteria are chosen for the negative-ideal solution and none for the positive (particularly due to the high ME). Based on the comparison with SARIMA, it could be argued that SARIMA provides a better mathematical basis for the judgmental adjustment methods than ARIMA, due to a much higher value of C. Though SARIMA is ranked 4th in ANP-TOPSIS, after ARIMA, their respective values of C are very close (0.37 to 0.35 respectively).

The AHP-TOPSIS and ANP-TOPSIS methods, which produce a final ranking in terms of relative distance to ideals, the actual error measures in respect to weights for each alternative are less decisive than their proximity to the positive-ideal/negative-ideal solutions.

6 Conclusion and Future Work

This paper develops two hybrid MCDM approaches, namely AHP-TOPSIS and ANP-TOPSIS, to support the selection of the most appropriate demand forecasting method applied to a plastic bag manufacturer case. The alternative forecasting methods include typical quantitative regression and hybrid demand forecasting models. The criteria on

which the forecasts are evaluated are their goodness-to-fit, measured as the error between the forecasts and actual sales. Several error measures are considered for the selection in a first step as independent for which AHP-TOPSIS is developed, and in a second step, as interdependent, for which ANP-TOPSIS is developed. The results show that in both cases, the judgmental adjusted forecasts are ranked highest, and that collaborative adjustment provides better results than single adjustment since it takes into account complementary judgments from several experts. However, the rankings are strongly influenced by the chosen criteria weights and nature (both priority and interdependence).

Although it shows high sensitivity to the interdependencies between criteria, it is recommended to replace the ARIMA model with SARIMA as the mathematical basis in the judgmentally adjusted forecasts for the industrial case study. Even though SARIMA changed rank between the two MCDM methods, the relative closeness C is still very close to the result for ARIMA (in the case of ANP-TOPSIS).

As it is demonstrated that collaborative judgment adjustments improve the forecasts and reduce the risk of significant outliers, future research will investigate the impact of the number of experts taking part to the collaborative consensus process, taking into account SARIMA as the mathematical basis for the approach.

7 Acknowledgement

The research leading to these results was funded by the Swiss National Science Foundation under project n° [176349], “Inventory management of short life cycle products with demand forecasts using Big data and judgmental information”

8 References

1. Xu, B., Ouenniche, J.: Performance evaluation of competing forecasting models: A multidimensional framework based on MCDA. Expert Systems with Applications. 39, 8312–8324 (2012).
2. Tayman, J., Swanson, D.A.: On the validity of MAPE as a measure of population forecast accuracy. Population Research and Policy Review. 18, 299–322 (1999).
3. Cheikhrouhou, N., Marmier, F., Ayadi, O., Wieser, P.: A collaborative demand forecasting process with event-based fuzzy judgements. Computers & Industrial Engineering. 61, 409–421 (2011).
4. Marmier, F., Cheikhrouhou, N.: Structuring and integrating human knowledge in demand forecasting: a judgemental adjustment approach. Production Planning & Control. 21, 399–412 (2010).
5. Mahmoud, E.: Accuracy in forecasting: A survey. Journal of Forecasting. 3, 139–159.
6. Chatfield, C.: Apples, oranges and mean square error. International Journal of Forecasting. 4, 515–518 (1988).
7. Ren, L., Glasure, Y.: Applicability of the Revised Mean Absolute Percentage Errors (MAPE) Approach to Some Popular Normal and Non-normal Independent Time Series. International Advances in Economic Research. 15, 409–420 (2009).
8. Mehdiyev, N., Enke, D., Fettke, P., Loos, P.: Evaluating Forecasting Methods by Considering Different Accuracy Measures. Procedia Computer Science. 95, 264–271 (2016).

9. Velasquez, M., Hester, P.T.: An Analysis of Multi-Criteria Decision Making Methods. International Journal of Operations Research. 10, 11 (2013).
10. Sipahi, S., Timor, M.: The analytic hierarchy process and analytic network process: an overview of applications. Management Decision. 48, 775–808 (2010).
11. Kara, S.S., Cheikhrouhou, N.: A multi criteria group decision making approach for collaborative software selection problem. Journal of Intelligent & Fuzzy Systems. 26, 37–47 (2014).
12. Raut, R., Cheikhrouhou, N., Kharat, M.: Sustainability in The Banking Industry: A Strategic Multi-Criterion Analysis. Business Strategy and the Environment. 26, 550–568.
13. Wang, T.-C.: The interactive trade decision-making research: An application case of novel hybrid MCDM model. Economic Modelling. 29, 926–935 (2012).
14. Chiu, W.-Y., Tzeng, G.-H., Li, H.-L.: A new hybrid MCDM model combining DANP with VIKOR to improve e-store business. Knowledge-Based Systems. 37, 48–61 (2013).
15. Opricovic, S., Tzeng, G.-H.: Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research. 156, 445–455 (2004).
16. Saaty, T.L.: Decision Making with Dependence and Feedback: The Analytic Network Process : the Organization and Prioritization of Complexity. Rws Publications (2001).
17. Zanakis, S.H., Solomon, A., Wishart, N., Dublish, S.: Multi-attribute decision making: A simulation comparison of select methods. European Journal of Operational Research. 107, 507–529 (1998).
18. Tao, L., Chen, Y., Liu, X., Wang, X.: An integrated multiple criteria decision making model applying axiomatic fuzzy set theory. Applied Mathematical Modelling. 36, 5046–5058 (2012).
19. Ertuğrul, İ., Karakaşoğlu, N.: Performance evaluation of Turkish cement firms with fuzzy analytic hierarchy process and TOPSIS methods. Expert Systems with Applications. 36, 702–715 (2009).
20. Beheshtinia, M.A., Omidi, S.: A hybrid MCDM approach for performance evaluation in the banking industry. Kybernetes. 46, 1386–1407 (2017).
21. Dağdeviren, M.: A hybrid multi-criteria decision-making model for personnel selection in manufacturing systems. Journal of Intelligent Manufacturing. 21, 451–460 (2010).
22. Franses, P.H.: Averaging Model Forecasts and Expert Forecasts: Why Does It Work? Interfaces. 41, 177–181 (2011).
23. Armstrong, J.S., Collopy, F.: Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting. 8, 69–80 (1992).
24. Holt, C.C., Modigliani, F., Muth, J.F., Simon, H.A., Bonini, C.P., Winters, P.R.: Planning Production, Inventories, and Work Force. Prentice Hall (1960).
25. Saaty, T.L.: How to make a decision: The analytic hierarchy process. European Journal of Operational Research. 48, 9–26 (1990).
26. Chu, P., Liu, J.K.-H.: Note on Consistency Ratio. Mathematical and Computer Modelling. 35, 1077–1080 (2002).