The impact of different guidance types on judgemental demand forecasting: A laboratory experiment

Meysam Arvan
(meysam.arvan@sydney.edu.au)
The University of Sydney Business School

Behnam Fahimnia
The University of Sydney Business School

Mohsen Reisi
The University of Sydney Business School

Enno Siemsen
Wisconsin School of Business

Abstract

Product demand forecasting (PDF) is a critical task in supply chain management. Various quantitative methods are developed for PDF. However, previous surveys confirm a heavy reliance on human judgement in practice. This is mainly to incorporate information not considered by quantitative models. The literature suggests that a Forecasting Support System (FSS) that systematically guides the forecasters in applying judgement is a likely solution to improve forecast accuracy. Guidance is the core component of this FSS. Using a laboratory experiment, this paper examines how different guidance types can help improve forecast accuracy in various task complexity settings (adding promotions and noise).

Keywords: Forecasting, Forecasting support systems (FSS), Promotions

Introduction

Various quantitative methods have been developed and applied to achieve better forecasts. Nonetheless, most studies and surveys confirm a heavy reliance on human judgement in practice (Sanders & Manrodt 2003). When time series do not appropriately address the dynamics of the business context caused by activities such as sales promotions and competitors' activities, the extrapolation of quantitative methods is likely to fail. In practice, not all causes of variability in time series are easy to systematically identify and address; hence, the use of human judgement and 'gut feeling' brings some value-add to the forecasts produced by quantitative models. This has made human expertise an inseparable asset of PDF which has also engendered the subject area of *judgemental demand forecasting*.

Human mind is limited in capacity and processing power. To apply their judgement in complex tasks such as demand forecasting, people employ mental shortcuts and simplifying tools called mind heuristics (Tversky & Kahneman 1974). Heuristics are essential for being able to deal with a high load of information. Employing these heuristics, however, is not free of bias and inconsistency. For instance, when promotional activities are performed, the forecasters tend to completely ignore the system forecast (Goodwin & Fildes 1999). A number of heuristics are exercised by forecasters and broadly addressed in the literature (Gino & Pisano 2008). Using these heuristics imposes biases such as over optimism, anchoring, overconfidence, and confusion of the signal with the noise (Robert Fildes et al. 2009; Blattberg & Hoch 1990). In other words, forecasters' intervention to incorporate contextual information can come with some detrimental effects (Eroglu & Croxton 2010; Goodwin & Wright 1994). This topic has created controversy in the PDF literature: *to employ or not to employ human judgement*.

Despite all the deficiencies associated with employing human judgement in forecasting, the use of quantitative methods together with human judgement is perhaps the most common forecasting approach in practice. These approaches integrate the managerial judgement into the *system forecast* (i.e., forecasts generated by a quantitative methods). These methods are known as what is termed *integrating methods*. Previous studies however indicate that if not aided and structured, human interventions could be counterproductive (Lee et al. 2007). Besides, the integrating methods are not able to utilise the experts' intuition without exposing the forecasting outcome to experts' biases in applying their knowledge. Therefore, a Decision Support System (DSS) (known as FSS in a forecasting context) capable of synthesising the benefits of quantitative models and informed experts' judgement whilst avoiding unnecessary complications for its use by typical forecasters seems to be a promising approach.

An FSS is defined as "*any system that provides support to the forecasting function within a company*" (Boylan & Syntetos 2010). An FSS can be a documented structured process and does not necessarily need to be a software solution (Boylan & Syntetos 2010). An FSS can be more precisely defined as software, a framework, or a structured procedure; all of which can utilise the expert judgement and quantitative techniques in integrating historical data and contextual information with the aim of providing meaningful support to the forecaster for generating more accurate forecasts and analysing the outcome.

The literature suggests if there is no guidance and structured path to follow, forecasters' interventions can often damage the accuracy of final forecasts. The role of FSS is therefore to systematically provide guidance to forecasters and inform their judgement. Nonetheless, the extant literature has not thoroughly explored the provision of different guidance type and the human capability in employing them.

Hence, this paper aims to examine whether the provision of guidance through an FSS will improve forecasters' adjustments. We try to examine forecasters' capability in utilising task specific guidance in two static and dynamic forms. It is firstly explored whether subjects can adapt themselves to changes in the forecasting environment using a dynamic performance guidance. Second, we examine whether people can utilise predefined static guidance (in form of interval guidance) to improve their forecasts. This takes place various task complexity levels. Hence, we also examine the moderating impact of task complexity on individuals' performance in utilising guidance. An experimental approach is adopted in order to investigate the developed research questions.

Research questions

Guidance can be categorised into several forms (Silver 1991). In one definition, guidance is provided either in an informative (i.e., the system provides recommendation) or a suggestive form (i.e., the system provides detailed information without any recommendations) (Montazemi et al. 1996). In another definition, the provided guidance is either dynamic or predefined (Parikh, Fazlollahi & Verma 2001). In the predefined guidance type, the guidance changes according to the forecasting environment, however, dynamic guidance changes based on forecasters' performance and forecasting environment.

The first type of guidance tested in this paper is predefined but not entirely static. It alters based on the presence or absence of promotions. Our predefined guidance (called guidance type 1) basically informs the forecasters about the possible magnitude of a promotion's impact on sales by giving them an interval. *Guidance type 1* comes in two arrangements depending on the guidance reliability: *guidance type 1-1:* guidance with 90% confidence interval which gives a wider but more reliable interval to the forecaster; and *guidance type 1-2:* guidance with 80% confidence interval which gives a narrower but less reliable interval to the forecasters.

The second type of guidance (*guidance type 2*) is dynamic and changes based on the individual's performance. According to (Silver 1991), "If a system is to provide meaningful assistance to users deciding what to do next, that assistance must reflect what they have done already". Guidance type 2 takes into account the factors considered by the forecaster and how he/she performed in the non-guided forecasting attempts. We want to know how effective guidance type 1 and 2 are in a controlled setting. Therefore, the first research question is:

Question 1: Do *guidance type 1* and *guidance type 2* improve the forecasters' performance in making forecasts?

Several factors contribute to the complexity of a forecasting task. Forecasting inherits a dynamic complexity dimension (i.e., unpredictability) that requires individuals to frequently adapt themselves to changes in the available information. Further, size complexity exists in a forecasting task where the task gets more complex as the number of information cues increase (Rasmussen, Standal & Laumann 2015).

It is widely acknowledged that promotional activities are one of the main contributors to demand variations in PDF and forecast complexity (Fildes & Goodwin 2007; Trapero, Kourentzes & Fildes 2015). However, in practice, promotional activities are not usually fully captured and factored into quantitative models. This is mainly because the data relevant to promotions can be unmodelable, unstructured, not properly recorded (e.g., advertisement type), or inaccessible (e.g., competitors' activities). Consequently, promotional activities can be one of the main reasons for manual adjustment of quantitative forecasts (Fildes & Goodwin 2013; Fildes & Goodwin 2007). Nonetheless, little research has been done to examine forecasters' ability to forecast demand when promotional activities are present (Trapero, Kourentzes & Fildes 2015). The influence of the magnitude of a promotion on forecasters' performance is not known. We are also unaware how forecasters react when facing multiple promotions with different impact levels.

Another contributor to series complexity is the level of noise. Much of the noise is in fact caused by the unknown, unexplored and disregarded information that will be categorised as noise. Noise is higher when promotions are present. There is ample evidence that forecasters mistake the time series noise for systematic patterns (e.g., trend or seasonality) and make unnecessary adjustments in an attempt to capture the influence of these patterns (Goodwin & Fildes 1999). This false reaction amplifies as the noise increases in a series (O'Connor, Remus & Griggs 1993). Even statistical methods perform poorly when noise level is high.

Many other features can contribute to the complexity of the series. Variations of features such as trend, seasonality, length of the series and so on can make a series simpler or more complex. However, in the PDF context, not all these features are present in every series. The presence of different features in a series mainly depends on the nature of the product and consumer purchase behaviour. Arguably, noise and promotions could be considered as the most common complexity attributes. We are more interested in promotional activities as they are literally part of any marketing and forecasting activity in the PDF context. Hence, the second research question asks:

Question 2: How time series complexity influence forecasters' performance and moderates the impact of guidance?

Experiment design

A laboratory experiment is designed to explore responses to the research questions articulated above. Laboratory experiments are a well-stablished methodology to examine human behaviour in the context of operations management (Bendoly, Donohue & Schultz 2006). Several instances of using laboratory experiments can be found in the judgemental forecasting literature (Goodwin & Lawton 1999; Kremer, Moritz & Siemsen 2011).

A 5×5 (guidance type × complexity level) mixed within-between experimental design is employed for this experiment. The between-subject variable was treatment or type of guidance and the complexity was treated as a within-subject variable. The main advantage of a mixed within between design is that it takes subjects as their own control. Consequently, their personal factors such as education and experience (that could possibly influence their forecasting capability) are factored in the analysis. Secondly, mixed within-between designs are known to be more powerful. Further, there can be less sample size yet higher degrees of freedom when a within-between design is adopted (List, Sadoff & MathisWagner 2013). The sequence of exposing the subjects to complexity levels in this experiment was fully randomised to minimise the chance of possible interactions between the levels (Hyndman & Embrey 2017).

A computerised FSS was developed using z-tree toolbox (Fischbacher 2007) in order to interact with the subjects. Subjects were randomly assigned to one of the five treatments. Each subject in each treatment was exposed to all complexity levels in a random order. Subjects were postgraduate students and alumni of the University of Sydney with adequate forecasting and data analytics background. According to (Remus 1986), management students are decent proxies for real managers in an experiment when it comes to making decisions in a supply chain context.

Before starting the experiment, subjects were given a cover story which familiarised them with their role in the experiment and the experiment procedure. The document contained information about the company, the product, retailors, promotions, times series data, and the type of guidance subjects receive. A step-by-step guideline was included in the document to describe what exactly need to be done over the course of the experiment. This was followed by collecting demographics information and past experiences. To make sure the subjects are fully familiar with the experiment's interface and the task; a practice session was run before the actual experiment began. Each subject made 10 forecasts (except from the practice forecasts) in the experiment.

The experiment consisted of five treatments. The treatments were designed according to the type of guidance that subjects received. In the control treatment, no guidance was provided to the subjects. Guidance type 1-1 and Guidance type 1-2 were provided to the subjects in treatments 1 and 2, respectively. Treatment 3 provided the subjects with guidance type 2. Subjects in the last treatment were given guidance type 1-2 and guidance type 2 concurrently. Table 1 summarises the information provided to the subjects in each treatment.

Table 1- Information shared with subjects in each treatment.

Treatment	**Feature**					
	Time series	**Statistical baseline**	**Promotional info**	**Guidance type 1 (80% confidence)**	**Guidance type 1 (90% confidence)**	**Guidance type 2**
Control	√	√	√	-	-	-
Treatment 1	√	√	√	√	-	-
Treatment 2	√	√	√	-	√	-
Treatment 3	√	√	√	-	-	√
Treatment 4	√	√	√	√	-	√

The complexity level of the forecasting task altered within each treatment. The simplest form of the forecasting task in treatments was to forecast for a low noise time series (noise level was 5% of the base sale) without promotions (O'Connor, Remus & Griggs 1993; Goodwin, Gönül & Önkal 2012). Complexity level 2 was high noise (15% of the base sale) without promotions. Complexity level 3 was minor promotions (lower in the impact scale), complexity level 4 was major promotions, and finally complexity level 5 was a mix of major and minor promotions. When promotions were present, promotional periods were exposed to additional noise to better reflect what takes place in practice.

A statistical baseline forecast was provided to all subjects. The baseline was generated according to a conventional exponential smoothing approach with alpha equal to 0.2. The subjects were informed about the approach used for generating the baseline and the fact that promotional impact is not considered in the baseline.

Each subject in the experiment received a unique database in each forecasting attempt. The subjects were provided with 30 weeks of data and were asked to forecast sales for week 31 considering the available information. We used the following formulation to generate the sales figures (Kremer, Moritz & Siemsen 2011):

$$S_t = \mu_t + \varepsilon_t + x_t(\beta + \delta_t) + y_t(\theta + \partial_t) \quad \text{Equation (1)}$$

$$\mu_t = \mu_{t-1} + C \quad \text{Equation (2)}$$

$$S.t: \sum_{t}^{N} x_t = f \quad \text{Equation (3)}$$

$$\sum_{t}^{N} y_t = k \quad \text{Equation (4)}$$

$$y_t + x_t \leq 1 \quad \text{Equation (5)}$$

$$y_t, x_t = \{0, 1\} \quad \text{Equation (6)}$$

Where S_t is the sales in the week t, x_t is the binary variable for major promotions, and y_t is the binary variable for minor promotions in the week t. Furthermore, ε_t ~Normal (0, r^2), δ_t ~Normal (0, h^2) and ∂_t ~Normal (0, p^2) are independent random variables resembling noise in the baseline sales, major promotions, and minor promotions impacts, respectively. Promotions' impact is shown by β for major promotions and θ for minor promotions. Equation (3) calculates the seed for generating random sales figures. Since the level change is not considered in the experiment, the seed remains constant throughout the forecasting horizon meaning $c = 0$. The first constraint makes sure that only f~Poisson(z, λ) major promotions happen in the entire forecasting horizon. Similarly, the number of minor promotions are constrained by Equation (5). Lastly, Equation (6) assures that either a minor promotion or a major promotion takes place at a period. All the parameters in this model are estimated according to the real data collected from a major FMCG company in Australia. Clearly, x and y will be set to zero depending on the complexity level.

A normative benchmark was employed in order to assess the effectiveness of produced forecasts (Kremer, Moritz & Siemsen 2011). To calculate the benchmark for above generated sales when promotions are present, an exponential smoothing with lift adjustment is used (Ali et al. 2009). The process is similar to decomposing the series into promotional vs non-promotional periods. In this approach, if there is no promotion the benchmark forecast is equal to the outcome of an exponential smoothing method for previous non-promotional periods. In the presence of promotions, the average uplift amount is added to the exponential smoothing figure calculated for non-promotional periods.

Results

Before analysing the collected data, the extreme outliers were removed. A systematic approach was adopted to remove the outliers from the data. A total of 151 subjects participated in the experiment (30 subjects per treatment, in average). Across all treatments, 2328 forecasts were made by 151 subjects. Four measures of mean absolute error (MAE), mean absolute percentage error (MAPE), symmetric mean absolute percentage error (sMAPE), and Root Mean Square Error (RMSE) were used to evaluate the forecasts.

Running a mixed ANOVA (table 2) indicates that the effects of complexity level and treatment are both significant. This is persistent regardless of the measure used.

Table 2 – ANOVA for treatments and complexity effects on MAPE, sMAPE, RMSE and MAE.

	MAPE		**sMAPE**		**RMSE**		**MAE**	
Effect	**F**	**P-value**	**F**	**P-value**	**F**	**P-value**	**F**	**P-value**
Treatment	7.72	**<0.0001**	7.22	**<0.0001**	4.25	**0.002**	17.59	**0.0014**
Complexity	55.73	**<0.0001**	41.92	**<0.0001**	145.26	**<0.0001**	535.31	**<0.0001**
Treatment: Complexity	1.83	**0.024**	1.77	**0.031**	1.96	**0.013**	30.38	**0.016**

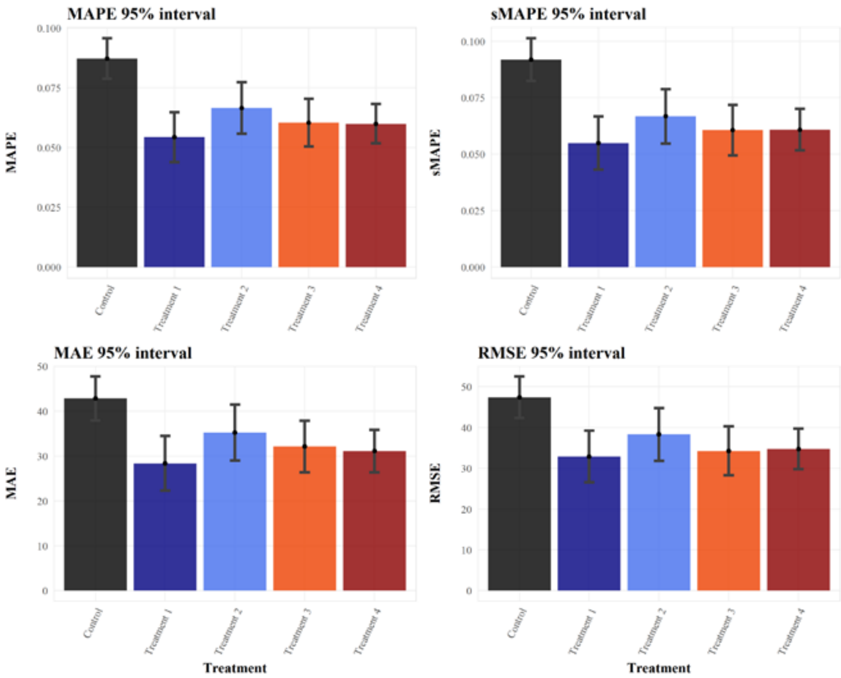

Figure 1 –Treatments' effect on forecast accuracy using MAPE, sMAPE, RMSE and MAE.

To further scrutinise the impact of guidance types, treatments' effect against forecast accuracy is illustrated in Figure 1. We observe that all guidance types improve forecasts to some extent. This improvement is more sensible when relative forecast accuracy measures are used (MAPE and sMAPE). The guidance type that has created the greatest improvement is 80% interval guidance (p-value of <.0001 for all accuracy measures when compared to the control group). However, Tukey test indicates that guidance types do not significantly differ in terms of forecast accuracy improvement.

As discussed earlier, five levels of complexity were introduced into the experiment. Figure 2 shows the average forecast accuracy in various complexity levels. The effect of treatment is averaged out in this figure. The figure indicates that as expected, there is a major difference between forecasting for low noise time series and other complexity levels. This is independent of the accuracy measure used. However, when relative measures (MAPE and sMAPE) are employed, the difference between high noise series forecasts accuracy and promotional forecast accuracy is less substantial. Relative measures are not sensitive to the scale of error and do not fully reflect the possible repercussions triggered by inaccuracy. Therefore, for promotional forecasts, we rely on scale sensitive measures. Using RMSE and MAE indicates that the difference between forecast accuracy of minor promotions and the other two promotional periods is significant (*p-value=0.0482*, *p-value=0.0006* when compared to major promotions and mixed promotions, respectively – using RMSE). However, major and mixed promotions are not significantly different (*p-value=0.1313* – using RMSE). The consistency of subjects in making forecasts for more complex series significantly increases as compared to less complex series.

It is also worthwhile to explore how guidance types aid forecasters in various complexity levels. The heat map in Figure 3 indicates that the provision of guidance has made greater improvement in more complex forecasting tasks. Comparing guided treatments with the unguided treatment indicates that mixed promotions complexity level has undergone the most improvement by provision of guidance. Statistical tests confirm this proposition (*p-value of 0.0106* and *0.0023* when comparing forecasts in treatment 1 and the control – using RMSE and MAE, respectively).

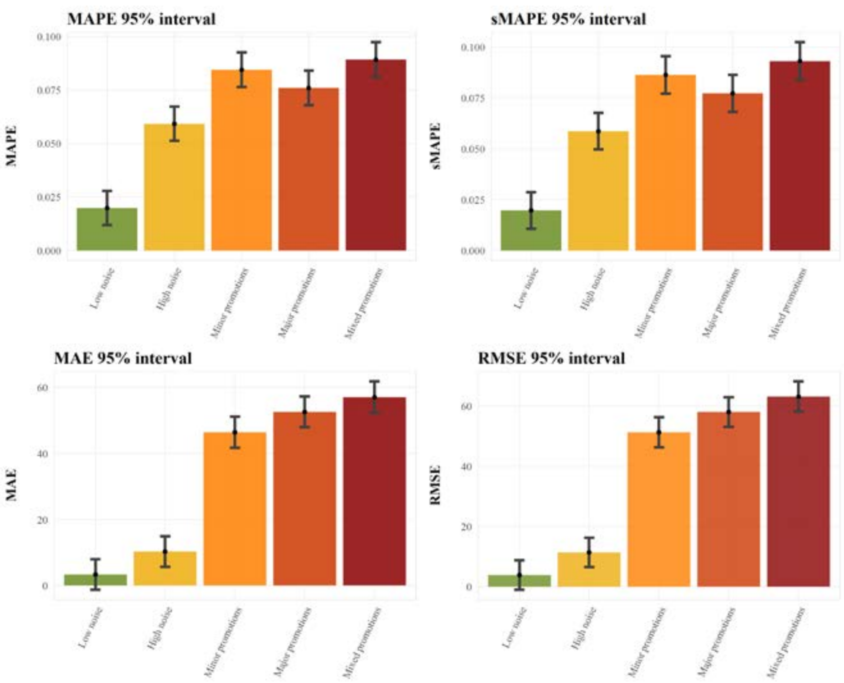

Figure 2 – Complexity effect on forecast accuracy using MAPE, sMAPE, RMSE and MAE.

The story is different for promotional forecasting. Surprisingly, the forecasters' performance when forecasting for minor promotions does not significantly improve by introducing guidance. However, in major promotions (treatments 1 and 4), forecast accuracy improves when compared to the control group.

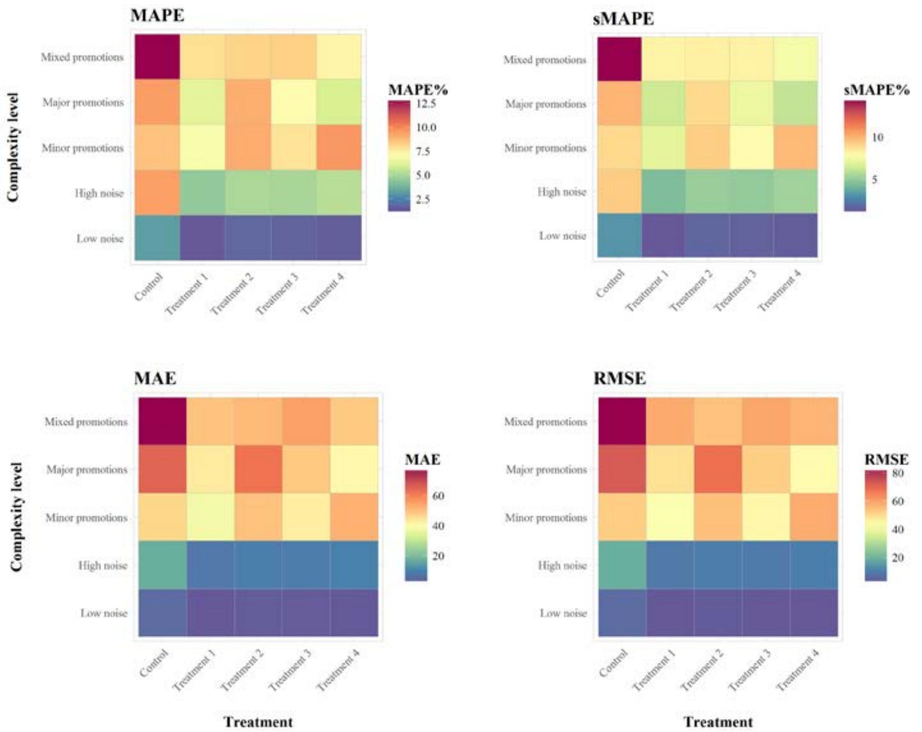

Figure 3 – The effect of guidance types on forecast accuracy under different complexity levels.

Conclusion

This study explored how forecasters could utilise two types of guidance (interval guidance and performance guidance) embedded in an FSS when making forecasts for promotional and non-promotional periods. A laboratory experiment was designed to examine the forecasters' ability to use these guidance types to make forecasts at different complexity levels. The complexity level was adjusted using different promotions and noise levels.

The results confirm that providing guidance improves forecast accuracy. The use of less reliable but narrower interval guidance (80%) made more improvement in forecast accuracy. However, compared to other types, it was only significantly better than the 90% interval guidance. This may imply that the subjects had no trust issues with the use of the provided guidance but rather were concerned about applying the guidance (the narrower an interval is, the easier it might be to apply). This result is in contrary with that of some of the past research where confidence in use of the interval guidance was emphasised over its ease of use (Gönül, Önkal & Lawrence 2006). Joint provision of interval guidance and dynamic guidance did not make significant improvements compared to each guidance type alone. Perhaps information load might be a factor influencing the effective use of the guidance (Schroder, Driver & Streufert 1967).

Both guidance types were more efficient in more complex series. This is in line with the findings of (Silver 1991) in which a connection was made between the provision of guidance in a DSS and task complexity. The most improvement was achieved when the series comprised a mix of minor and major promotions. As expected, task complexity deteriorates forecasters' performance. Even introducing 10% more noise to the series significantly damaged forecasts accuracy (confirming the findings of Sanders & Ritzman (1990) & Hogarth (1975)). Minor and major promotion tasks were slightly different for forecasters; however, mixed promotions were considerably harder to forecast compared to minor promotions. Finally, there was no significant difference between major and mixed promotion tasks. Our results do not provide enough evidence for the worsening impact of promotions' scale but mixing different promotions in a series could confuse the forecasters.

These results provide important insights for building an FSS that interacts with forecasters through the provision of guidance; what we frequently refer to as a behaviourally-informed FSS. Our experiment used relatively inexperienced postgraduate students as subjects. Future studies can replicate this experiment in a practical setting using professional forecasters as subjects, and compare the results.

References

Ali, ÖG. Sayin, S. van Woensel, T. & Fransoo, J (2009), "SKU demand forecasting in the presence of promotions", *Expert Systems with Applications*, Vol. 36, No. 10, pp. 12340–12348.

Bendoly, E. Donohue, K & Schultz, KL (2006), "Behavior in operations management: Assessing recent findings and revisiting old assumptions", *Journal of Operations Management*, Vol. 24, No. 6, pp. 737–752.

Blattberg, RC & Hoch, SJ (1990), "Database Models and Managerial Intuition: 50% Model + 50% Manager", *Management Science*, Vol. 36, No. 8, pp. 887–899.

Boylan, JE & Syntetos, AA (2010), "Spare parts management: A review of forecasting research and extensions", *IMA Journal of Management Mathematics*, Vol. 21, No. 3, pp. 227–237.

Eroglu, C. & Croxton, KL (2010), "Biases in judgmental adjustments of statistical forecasts: The role of individual differences", *International Journal of Forecasting*, Vol. 26, No. 1, pp. 116–133.

Fildes, R. & Goodwin, P. (2007), "Against your better judgment? How organizations can improve their use of management judgment in forecasting", *Interfaces*, Vol. 37, No. 6, pp. 570–576.

Fildes, R. & Goodwin, P. (2013), "Forecasting support systems: What we know, what we need to know", *International Journal of Forecasting*, Vol. 29, No. 2, pp. 290–294.

Fildes, R. Goodwin, P. Lawrence, M & Nikolopoulos, K (2009), "Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning", *International Journal of Forecasting*, Vol. 25, No. 1, pp. 3–23.

Fischbacher, U (2007), "z-Tree : Zurich toolbox for ready-made economic experiments", , pp. 171–178.

Franses, PH & Legerstee, R. (2011), "Experts' adjustment to model-based SKU-level forecasts: Does the forecast horizon matter', *Journal of the Operational Research Society*, Vol. 62, No. 3, pp. 537–543.

Gino, F. & Pisano, G. (2008), "Toward a Theory of Behavioral Operations", *Manufacturing & Service Operations Management*, Vol. 10, No. 4, pp. 676–691.

Gönül, MS. Önkal, D. & Lawrence, M (2006), "The effects of structural characteristics of explanations on use of a DSS", *Decision Support Systems*, Vol. 42, No. 3, pp. 1481–1493.

Goodwin, P. & Fildes, R. (1999), "Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy?", *Journal of Behavioral Decision Making*, Vol. 12, No. 1, pp. 37–53.

Goodwin, P. Gönül, MS & Önkal, D (2012), "Antecedents and effects of trust in forecasting advice", *International Journal of Forecasting*, Vol. 29, No. 2, pp. 354–366.

Goodwin, P. & Lawton, R. (1999), "On the asymmetry of the symmetric MAPE", *International Journal of Forecasting*, Vol. 85, No. 3, 2, pp. 1059–1069.

Goodwin, P. & Wright, G. (1994), "Heuristics, biases and improvement strategies in judgmental time series forecasting", *Omega*, Vol. 22, No. 6, pp. 553–568.

Hogarth, RM (1975), "Cognitive Processes and the Assessment of Subjective Probability Distributions: Rejoinder", *Journal of the American Statistical Association*, Vol. 70, No. 350, p. 294.

Hyndman, K. & Embrey, M. (2017), "Econometrics For Experiments", in *Handbook of Behavioral Operations*, pp. 1–4.

Kremer, M. Moritz, B. & Siemsen, E. (2011), "Demand forecasting behavior: System neglect and change detection", *Management Science*, Vol. 57, No. 10, pp. 1827–1843.

Lee, WY. Goodwin, P. Fildes, R. Nikolopoulos, K & Lawrence, M (2007), "Providing support for the use of analogies in demand forecasting tasks", *International Journal of Forecasting*, Vol. 23, No. 3, pp. 377–390.

List, JA. Sadoff, S. & MathisWagner (2013), "So You Want To Run an Experiment, Now What?", *Journal of Chemical Information and Modeling*, Vol. 53, No. 9, pp. 1689–1699.

Montazemi, AR. Wang, F. Nainar, SMK & Bart, CK (1996), "On the effectiveness of decisional guidance", *Decision Support Systems*, Vol. 18, No. 2, pp. 181–198.

O'Connor, M. Remus, W. & Griggs, K. (1993), "Judgemental forecasting in times of change", *International Journal of Forecasting*, Vol. 9, No. 2, pp. 163–172.

Parikh, M. Fazlollahi, B. & Verma, S. (2001), "The effectiveness of decisional guidance: An empirical evaluation", *Decision Sciences*, Vol. 32, No. 2, pp. 303–332.

Rasmussen, M. Standal, MI & Laumann, K. (2015), "Task complexity as a performance shaping factor: A review and recommendations in Standardized Plant Analysis Risk-Human Reliability Analysis (SPAR-H) adaption", *Safety Science*, Vol. 76, pp. 228–238.

Remus, W. (1986), "Graduate students as surrogates for managers in experiments on business decision making", *Journal of Business Research*, Vol. 14, No. 1, pp. 19–25.

Sanders, NR & Manrodt, KB (2003), "Forecasting Software in Practice : Use, Satisfaction and Performance", *Interfaces*, Vol. 33, No. 5, pp. 90–93.

Sanders, NR & Ritzman, LP (1990), "Improving short-term forecasts", *Omega*, Vol. 18, No. 4, pp. 365–373.

Schroder, HM. Driver, MJ & Streufert, S. (1967), *Human information processing*,.

Silver, MS (1991), "Decisional Guidance for Computer-Based Decision Support", *MIS Quarterly*, Vol. 15, No. 1, pp. 105–122.

Trapero, JR. Kourentzes, N. & Fildes, R. (2015), "On the Identification of Sales Forecasting Models in the Presence of Promotions", *Journal of the Operational Research Society*, Vol. 66, No. 2, pp. 1–21.

Tversky, A. & Kahneman, D. (1974), "Judgment under Uncertainty: Heuristics and Biases", *Science (New York, N.Y.)*, Vol. 185, No. 4157, pp. 1124–1131.