When providing best-case and worst-case scenarios can be detrimental to judgmental demand forecasts and production decisions.

Paul Goodwin[a*], M. Sinan Gönül [b] & Dilek Önkal [c]

a. School of Management, University of Bath, Bath BA2 7AY, United Kingdom. p.goodwin@bath.ac.uk

b. Newcastle Business School, Northumbria University, 2 Ellison Place, Newcastle upon Tyne NE1 8ST, United Kingdom. sinan.gonul@northumbria.ac.uk

c. Newcastle Business School, Northumbria University, 2 Ellison Place, Newcastle upon Tyne NE1 8ST, United Kingdom. dilek.onkal@northumbria.ac.uk

* Corresponding author

When providing best-case and worst-case scenarios can be detrimental to judgmental demand forecasts and production decisions.

Abstract

This paper examines the accuracy of judgmental forecasts of product demand and the quality of subsequent production level decisions under two different conditions: (i) the availability of only time series information on past demand; (ii) the availability of time series information together with scenarios that outline possible prospects for the product in the forthcoming period. An experiment indicated that production level decisions made by participants had a greater deviation from optimality when they also received best-case and worst-case scenarios. This resulted from less accurate point forecasts made by these participants. Further analysis suggested that participants focussed on the scenario that was congruent with position of the latest observation relative to the series mean and discounted the opposing scenario. This led to greater weight being attached to this observation, thereby exacerbating the tendency of judgmental forecasters to see systematic changes in random movements in time series.

Keywords: Forecasting; Judgment; Extrapolation; Scenarios; Production planning

Highlights

- Providing scenarios to judgmental forecasters worsened production decisions.
- Scenarios did not reduce the tendency of forecasters to be overconfident.
- The scenario inconsistent with the latest demand seemed to be discounted.
- This exacerbated a tendency to see systematic change in random movements in demand.
- Point estimates of expected future demand were therefore less accurate.

Why providing best-case and worst-case scenarios can be detrimental to judgmental demand forecasts and production decisions.

1. Introduction

Forecasts are designed to support decisions. Of particular importance are decisions on how much to produce of each of a number of products when manufacturing capacity is constrained. Production decisions like these require estimates of both the central tendency and variation of the probability distribution of future demand so that factors such as customer service levels can be considered. Point forecasts will supply the former and probabilistic forecasts (in the form of density forecasts or prediction intervals) will supply the latter. In practice, such forecasts, and their associated production decisions, are often based on management judgment (Fildes and Goodwin, 2007) and, as a result, can be subject to a number of cognitive biases. This means that point forecasts are often inaccurate while probabilistic forecasts tend to underestimate the dispersion in probability distributions of demand (Lawrence et al., 2006).

In production planning, managers are likely to have access to time series information on past demand in addition to contextual information relating to demand (Fildes et al. 2009; Fildes et al. 2018). This contextual information may take the form of scenarios which suggest possible future events that may influence imminent demand. The forecaster will have the task of integrating these two sources and types of information. There has been relatively little research on the possible interaction between scenarios and time series information and its effects on forecasts and, in particular, the decisions that follow. This paper aims to fill that research gap by examining whether the availability of best and worst case scenarios alongside time series information enhances or reduces the accuracy of demand forecasts and the quality of ensuing production decisions.

The paper is structured as follows. After a review of the relevant literature, we describe an experiment that involved participants using their judgment to make demand forecasts for a series of products before deciding how available manufacturing capacity should be allocated between the products. The participants either received no scenarios or pairs of best case and worst case scenarios. In the latter case, the scenarios were designed to be either ‘weakly’ optimistic or pessimistic, or to express ‘strong’ optimism or pessimism. We then present the results of the experiment and discuss their practical implications.

2. Literature review

Accurate forecasts of future demand are crucial if production decisions are to be made efficiently. Although many organizations rely principally on point forecasts (e.g. Klassen and Flores, 2001; Danese and Kalchschmidt, 2011) ideally forecasts should also include estimates of the uncertainty in future demand, expressed either as density forecasts or prediction intervals. These estimates allow levels of safety stocks to be determined so that the probability of stock-outs can be reduced to acceptable levels and hence desired customer service levels can be achieved. It is common for demand forecasts in companies to be based on management judgment. Either forecasts are entirely based on judgment or judgmental adjustments are made to forecasts obtained from statistical algorithms (Sanders and Manrodt, 2003; Fildes and Goodwin, 2007).

Researchers have shown that judgmental forecasts are subject to a number of cognitive biases (e.g. see Lawrence et al. 2006). In particular, when extrapolating time series graphs to produce point forecasts, people overweight the most recent observation. For example, when series are untrended, their forecasts can be modelled as a weighted average of the mean of the series and the latest observation (Bolger and Harvey, 1993; Lawrence and O'Connor, 1992). This means that, when the latest observation deviates significantly from the series mean because of noise, the point forecast can be an inaccurate estimate of the mean of the probability distribution of future demand. In addition, when estimating the variation of this probability distribution –for example, by estimating a 95% prediction interval - judgmental forecasters tend to produce ranges that are too narrow, thereby underestimating the variation, a phenomenon known as overconfidence or hyperprecision (e.g. Önkal et al., 2009).

A number of different explanations have been put forward to account for overconfidence. These range from the idea that people anchor on, and under adjust from, the most likely value when estimating a prediction interval to the belief that they are unable to imagine why an outcome might lie outside a predicted range (Tversky and Kahneman, 1974; Arkes, 2001; Soll and Klayman, 2004; Moore et al., 2015). When estimating prediction intervals from time series information, Lawrence and Makridakis (1989) expected judgmental forecasters who saw untended series to use the maximum and minimum values from past observations as guides to where the bounds of the interval should be. Instead, they found that people made assessments of the most extreme slopes that the graph might follow in the future and used these to estimate the bounds of the interval. When the series were untrended, these estimated slopes tended to be small and the resulting intervals were far too narrow.

One approach to reducing, or overcoming, overconfidence involves drawing the forecaster's attention to the possibility of extreme outcomes. Given that people tend naturally to engage with information in a narrative form (e.g. Satterfield et al., 2000), providing them with scenarios that describe alternative extreme futures may act as an antidote to overconfidence.
Scenarios are not forecasts. Instead they are flowing narrative accounts of plausible futures that might unfold. They allow qualitative factors to be considered when making decisions that may be impacted by future events. When multiple scenarios are supplied, it is argued that they sensitize decision makers to the degree of uncertainty that they face by suggesting alternative ways in which the future may play out (Goodwin and Wright, 2001).

Evidence for the effectiveness of this approach in domains, such as personal decision making (Griffin et al., 1990), strategic decision making (Schoemaker, 1993) and the forecasting of future values of assorted variables (e.g. the murder rate in Chicago and number of major league baseball teams) (Kuhn and Sniezek, 1996), has produced mixed results (Newby-Clark et al., 2000). However, a key feature of demand forecasting in many situations is the presence of time series information alongside the scenarios. This means that a forecaster will have to integrate the two sources of information. There is also the need to produce a point forecast in addition to an estimate of the variation of the probability distribution. These characteristics of the task lead to a number of possibilities.

First, the provision of scenarios might work as intended - reducing overconfidence by encouraging people to see a wider range of possible levels of demand than would be suggested by estimating plausible extreme slopes as in the Lawrence and Makridakis (1989) study. An alternative outcome is that overconfidence might increase. It is known that providing people with more information increases their confidence in their forecasts at a rate that is disproportional to the true relevance and value of this information (Davis et al., 1994). Schnaars and Topol (1987) investigated the possible benefits of providing 'optimistic', 'pessimistic' and 'middle ground' scenarios to people asked to produce point forecasts from time series information. They found that the scenarios increased the forecasters' confidence in their point forecasts even though their accuracy did not increase. Similarly, Önkal et al. (2013) found that the provision of scenarios also increased confidence when time-series based forecasts were expressed as prediction intervals.

A third possibility is that, when presented with a set of scenarios, people might focus on a single favoured scenario and discount the others (Schnaars and Topol, 1987). For example, when a 'middle ground' scenario is presented, this may receive unwarranted attention and divert the forecaster's focus

from the extreme possibilities. Alternatively, forecasters may have a preference for optimistic scenarios and ignore any pessimistic scenarios presented to them (Newby-Clark et al., 2000). One unexplored possibility is that there might be an interaction between the time series information and the provision of scenarios. Forecasters supplied with multiple cues, in this case the time series observations and the scenarios, are known to discount cues that are inconsistent with the other cues. A subset of cues is said to be consistent if a judge feels that they all agree in their implications for what is being judged (Slovic, 1966). As Yaniv (2004) argues, people have a need to form and maintain consonance and harmony and one way of dealing with inconsistent cues is to ignore the 'dissenting' information. In addition, there is evidence from the consumer behaviour literature (e.g. Wu and Cheng, 2011; Xu et al., 2013) that consistent cues have synergistic effect on judgments. Recall that people tend to focus on the most recent observation in a time series. If this observation represents an increase in demand from the previous observation or is above the series mean, then this is likely to be seen as consistent with a best case scenario, possibly leading to a discounting of the worst case scenario and an even greater tendency to overweight the most recent behaviour of the series. Similarly, if the most recent change in the series is a decrease in demand or the observation is below the series mean, this may interact with the worst case scenario so that the possibility of a future decline in sales is overestimated. Because the most recent movement in a series is likely, at least in part, to reflect noise, the likely effect of providing scenarios would be to decrease the accuracy of the point forecasts. Finally, the requirement to produce a point forecast is also likely to increase overconfidence (Russo & Schoemaker, 1992), possibly because people anchor on this forecast when estimating the bounds of the distribution or prediction interval.

When it comes to using forecasts to make production decisions for multiple products which share the same constrained manufacturing capacity, managers relying on judgment are likely to face a number of challenges. Possible objectives include maximising customer service level, maximising sales or minimising inventory costs. However, in practice these objectives may be ill defined or there may be an absence of appropriate metrics to measure the extent to which they are being achieved or balanced against one another (Gunasekaran et al., 2001). Decision makers tend to have difficulties in making decisions with multiple objectives optimally (Goodwin and Wright, 2014) so that they tend to resort to heuristics, such as focussing on a single, most important objective (Payne et al., 1993). There is limited evidence in the literature on which objectives are commonly prioritised, but a survey of Brazilian companies found that maximising customer service levels was considered to be the main objective of practitioners (Tomotani and de Mesquita, 2018). However, even focusing on a single objective like this is likely to be difficult, given the uncertainty involved, while the requirement to assign limited capacity amongst different products compounds the problem.

Given these challenges, we next describe an experiment that was designed to assess whether providing worst and best case scenarios led to improvements in judgmental demand forecasts and their associated production decisions.

3. The experiment

The participants were business students from Bilkent University who were completing a forecasting course. The experimental task was presented in a paper-and-pencil format where the setting involved making demand forecasts for different mobile phones and smart tablets produced by a mobile telecommunications company. Sixty-eight students completed the experiment.

To control the levels of uncertainty and trend, artificial time series were constructed to represent past demand. A total of six untrended series, three with high noise and three with low noise were used. Participants received these series in a randomized order. The procedure for constructing these artificial series was similar to previous studies on judgmental forecasting (e.g. Gönül, Önkal & Lawrence, 2006; Önkal, Gönül & Lawrence, 2008; Önkal, Sayım & Gönül, 2013). The formula used for generating the time series was:

$$y(t) = 125 + error(t) \qquad t = 0,1,....., 20$$

where *error(t)* was normally distributed with zero mean and a standard deviation of either 10% (*i.e.*, 0.1 x 125) for low noise or 20% (*i.e.*, 0.2 x 125) for high noise.

The participants received time-series plots showing past demand over the previous 20 weeks for the six products in random order (these series are displayed in Appendix A). For each product, participants were asked to:

(i) make a point forecast;
(ii) give their confidence (probabilistic estimate) that the realized value would be within $\pm$ 5% of their point forecast; and
(iii) make a production decision (i.e., decide on how many units they would order for production for this particular product). This represented an important decision that required them to translate their forecasts (and confidence in these forecasts) into actual action given that the total production capacity was set to 750 (set at number of products x baseline demand, i.e. 6 x 125) units.

Participants were randomly assigned to one of three following experimental groups and they received information accordingly:

Group 1 – No scenarios (23 participants): participants in this group were given the time series information only and were asked to complete their forecasts, state their confidence and make production decisions based only on the past demand as summarized in the plots (see Appendix A for a sample form given to participants in G1).

Group 2 –Both weak optimistic and weak pessimistic scenarios (23 participants): in addition to the time series information, participants received both weakly optimistic and weakly pessimistic scenarios (entitled as "Scenario A" and "Scenario B") (see Appendix B for a sample form given to participants in G2).

Group 3 –Both strong optimistic and strong pessimistic scenarios (22 participants) : these participants received the time series as well as both strongly optimistic and strongly pessimistic scenarios (entitled as "Scenario A" and "Scenario B") (see Appendix C for a sample form given to participants in G3).

The pairs of scenarios were constructed to be of roughly equal length and to be concise as is recommended in practice (Schnaars, 1987). A 'middle ground' scenario was not provided as the intention was to draw attention to factors that might lead to extreme levels of demand. Upon completing their forecasts, confidence assessment and production decisions, participants were asked to fill out a checklist showing their total production plan for all six products. The experiment ended with the participants answering an exit questionnaire.

4. The Results

4.1 Accuracy of point forecasts

Because the demand time series were artificially constructed the true value of the signal for all series for week 21 was known to be 125 units. The mean absolute errors (MAE) of the forecasts made by participants in each of the three groups are shown in Table 1 where the absolute error of a forecast = |125 – point forecast|. Also shown are the standard deviations of the MAEs

Group	No. of participants	MAE	SD
No scenarios	23	16.88	7.60
Weak scenarios	23	24.27	6.98
Strong scenarios	22	23.37	8.81

Table 1. Accuracy of point forecasts

To see if the 20 observations provided for each series were sufficient to produce reliable point forecasts, we used the expert system in *Forecast Pro*, a widely used commercial statistical forecasting

package, to produce forecasts for each of the series. These had an MAE of 10.67, so the software's forecasts were substantially more accurate than those of the human judges.

A one-way ANOVA revealed that there were significant differences between the MAEs of the judgmental forecasts ($F_{2, 65} = 6.09$, p=0.004). The group who did not receive scenarios, on average, made significantly more accurate point forecasts than those who received scenarios. However, the difference between those receiving weak and strong scenarios was not significant. An ANOVA showed that there was also no significant difference between the three groups in the mean bias (measured by the mean directional error) of their forecasts (p =0.785).

4.2 Calibration of confidence assessments

There are a number of ways of assessing the calibration of the confidence assessments made by the participants. For example, it is possible to assess the true probability that the demand would be within 5% of their point forecast. However, this measure would confound the accuracy of their point forecast with their ability to assess the variation of demand. The primary interest here is whether their ability make assessments of the variation in demand was improved by the access to scenarios. From their assessments, and given that demand was normally distributed, it is possible to infer an estimate of the standard deviation in demand and to compare this with the true standard deviation for a given series. For example, consider a point forecast of 140 and a confidence assessment of 80%. The range that lies within 5% of the point forecast is 133 to 147 units (i.e. a width of 14 units). Given a normal distribution with a mean equal to the point forecast, and the 80% assessment, this range covers 2.56 standard deviations so the implied estimate of the standard deviation is 14/2.56 = 5.47 units. Table 2 shows the mean bias in estimating the standard deviation (where: Bias = True SD – Implied estimate of SD) for the three groups and for the low and high noise series. The bracketed figures show the standard deviations of the mean biases. Almost all of the estimates were too low –only 6 out of 408 implied estimates were too high suggesting overconfidence.

Group	Low noise	SD	High noise	SD
No scenarios	7.66	(1.40)	19.60	(1.43)
Weak scenarios	6.96	(3.43)	18.66	(4.24)
Strong scenarios	7.11	(2.00)	19.66	(1.77)

Table 2. Mean bias in implied estimates of standard deviation of demand

One way ANOVAs indicated that there were no significant differences in the biases in estimating the standard deviation between the three groups either for low noise series ($p = 0.593$) or high noise series ($p = 0.400$). Accordingly, on this measure, there is no evidence that the provision of scenarios encouraged people to anticipate greater variation in demand. Interestingly, the differences between the biases for high and low noise series are typically close to 12.5 units (the difference between the true standard deviations) suggesting that people did not respond with higher implied estimates of variation when presented with series having higher levels of noise.

When the 95% prediction intervals obtained from *Forecast Pro* were used to infer estimates of the standard deviations of demand, they had a mean bias of only -0.60 units for low noise and -2.72 for high noise series, suggesting that the estimates had a very slight tendency to overestimate the standard deviations, in marked contrast to the implied estimates of the human judges.

4.3 Decision quality

We assessed the production decisions on two dimensions: the mean customer service level achieved across the products and the expected level of total sales that would be attained in week 21. These were calculated using the known normal probability distribution of demand. In the case of expected sales, partial expectations of the normal distribution were used. No information about inventory costs was provided to participants and there was no indication of the relative importance of the products so all products were considered to have equal priority.

Given the production capacity constraint of 750 units, a grid search was used to compare around 2.4 million combinations of production for the six products. This revealed that the maximum average customer service level that could be achieved is 76.9% (meaning that 76.9% of demand could be fulfilled). To achieve this service level, 150 units should be produced for five of the products with one 'high noise' product having zero production. This strategy would lead to expected total sales of 620.5 units. The alternative of producing 125 units of all six products would, of course, only lead to a mean customer service level of 50% given that demand for all products was a normal distribution with a mean of 125 units. However, a grid search showed that this production plan would achieve the maximum possible expected total sales of 695.8 units. Alternatively, trade-offs can be made between performances on these dimensions to achieve a balance between them and Figure 1 shows the efficient frontier containing non-dominated combinations of mean service level and expected total sales.

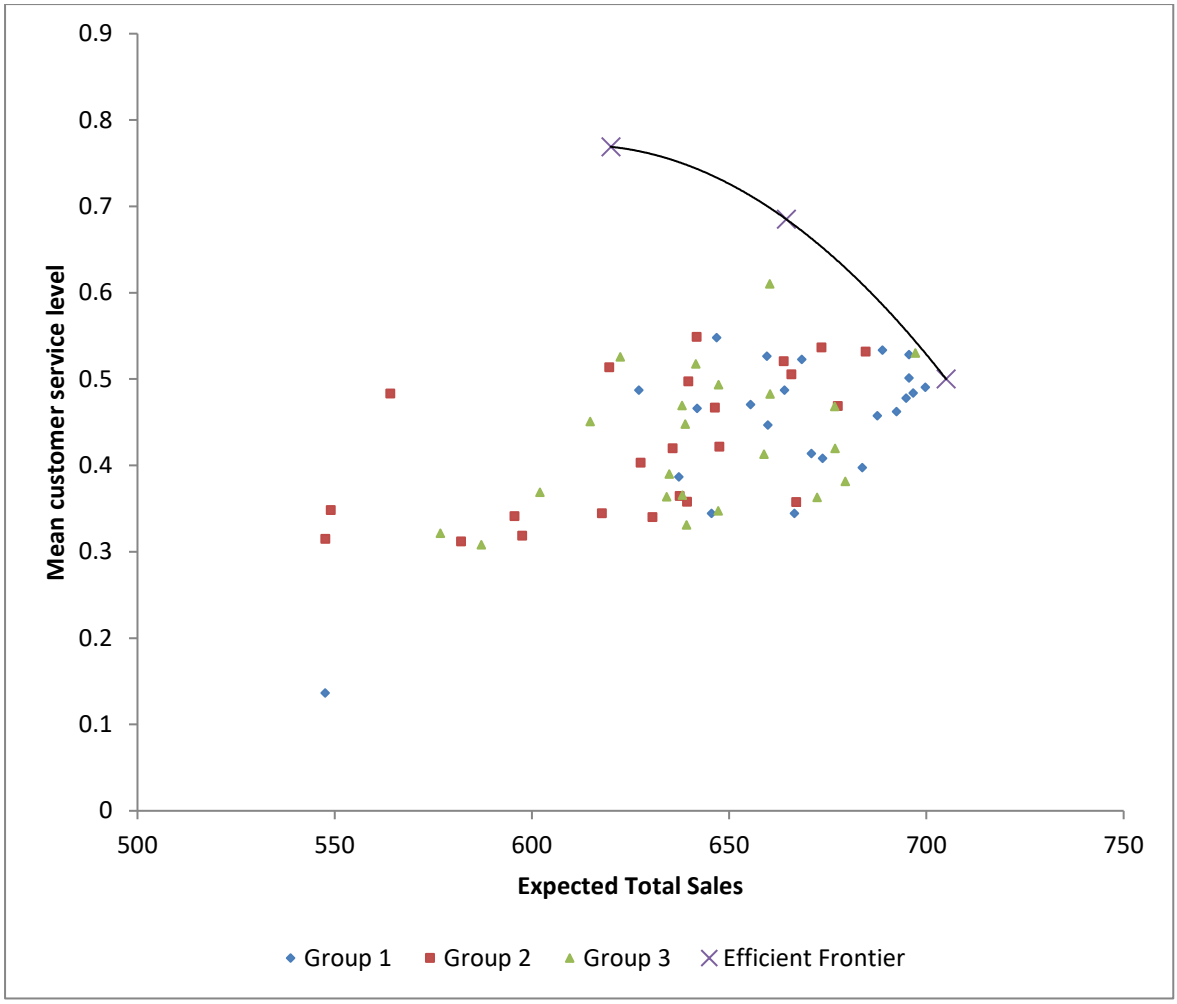


Figure 1. Mean customer service levels and expected total sales achieved by participants

Figure 1 also shows how well the participants' production decisions performed on the two dimensions. Of course, the participants did not have knowledge of the true probability distributions of demand or access to a grid search algorithm and it can be seen that many of their production plans were far from the efficient frontier. Nevertheless, with the exception of an outlier, the decisions of people who did not receive scenarios (Group 1) tended to be closer to the frontier. This can be seen more clearly in Table 3, which shows the mean number of times each group's decisions dominated those of each of the other groups. For example, decisions made by Group 1 (no scenarios) participants, on average, dominated 12.57 Group 2 decisions and 11.39 Group 3 decisions. It can be seen that dominance of Group 1 decisions (see the first column of the table) by the other groups was relatively rare. Table 4 shows the mean customer service levels and mean expected total sales achieved by the groups. A MANOVA indicated that the means for Groups 1 were significantly higher than those of the other groups (Wilks' lambda statistic had $F_{4, 128} = 3.44$ with $p = 0.0105$).

	Mean number of:		
	dominated Group 1 decisions	dominated Group 2 decisions	dominated Group 3 decisions
Group 1	xxxx	12.57	11.39
Group 2	3.39	xxxx	6.43
Group 3	3.68	9.14	xxxx

Table 3. The relative dominance of decisions made in the three groups

	Mean service level (%)	Expected total sales (units)
Group 1	44.90	665.2
Group 2	42.20	628.4
Group 3	42.60	642.9

Table 4. Mean service levels and expected totals sales achieved by the three groups.

5. Discussion

These results raise two main questions. Why were many of the decisions so far from the efficient frontier and why did those who did not receive scenarios make 'better' decisions? In relation to the first question, was this due to inaccurate point forecasts, underestimation of the variance of the probability distribution of demand, or an inability to handle the need to allocate the 750 units of production capacity?

There was evidence that the participants struggled with the production capacity constraint. The percentage of participants who did not use their full production capacity was 65%, 74% and 73%, respectively for Groups 1, 2 and 3. In addition, we saw earlier that all of the participants grossly underestimated the standard deviation of demand when their estimates were inferred from their question about confidence. However, it is by no means certain that this estimate was used when they were asked to make their decision on production levels. To investigate this, we calculated, for each product, the difference between each participant's production decision and their point forecast. We then determined what decision would have been made if the *Forecast Pro* point forecast had been

used instead: -i.e. Corrected decision = *Forecast Pro* point forecast + (Production decision - Original point forecast). The corrected decisions were adjusted to ensure that they summed to 750 by sharing any surplus or deficit equally between the products. The mean performances of the corrected decisions are shown in Table 5.

	Mean service level (%)	Expected total sales (units)
Group 1	49.60	694.1
Group 2	47.50	680.8
Group 3	49.20	698.5

Table 5. Mean performances of corrected decisions

When compared to the results in Table 4, those in Table 5 show that the correction led, on average, to improved performances on the two dimensions. Moreover, a MANOVA showed that there was no longer a significant difference between the groups (for Wilks' lambda, $F_{4, 128}=1.22$, $p = 0.31$). This shows that an inability to use the full 750 units of production capacity and relatively inaccurate judgmental point forecasts reduced the performance of the decisions. But why did the decisions made by the Group 1 participants perform better? A chi-squared contingency table test did not indicate a significant association between group membership and a tendency to underuse the 750 units capacity ($\chi^2 =0.49$ with 2 df, $p=0.783$). This suggests that it was the more accurate point forecasts of Group 1 participants that led to decisions that performed better.

To investigate why providing scenarios was associated with less accurate point forecasts, the method of Generalized Estimating Equations (GEE) was used to obtain a model that showed the relationship between the dependent variable: Point forecast – 125 (i.e. the error in estimating the signal multiplied by -1 for ease of interpretation) and the following independent variables:

(i) the last observation minus the mean of the series (recall that Bolger and Harvey, 1993 and Lawrence and O'Connor, 1992 found that people used this cue when extrapolating untrended series);

(ii) whether scenarios were provided (as a dummy variable: 1= yes, 0 = no);

(iii) whether the series had high noise (as a dummy variable: 1= yes, 0 = no);

(iv) two way interactions between (i), (ii) and (iii).

GEE was used because each participant made six forecasts so the data involved repeated measures. The method requires initial assumptions to be made about the correlations between the repeated measures. Three assumptions were tried (independent, unstructured and exchangeable) and all gave broadly similar results. Table 6 shows the results when the independent assumption was used. Significant parameters are shown in bold.

Variable	Parameter	Std error	Wald Chi-Square	*p*-value
Intercept	-2.28	2.02	1.27	0.259
Last obs - Series mean	**1.06**	0.07	256.06	**<0.0005**
Scenarios provided	3.27	2.72	1.44	0.230
High noise	2.00	2.44	0.67	0.410
(Last obs - Series mean) x Scenarios provided	**0.46**	0.13	12.02	**0.001**
Scenarios provided x High Noise	-0.52	3.35	0.02	0.880
(Last obs - Series mean) x High noise	**0.18**	0.08	4.46	**0.035**

Table 6. A GEE model of the error in forecasting the time series signal

Table 6 shows, as expected, that a latest observation above the series mean tends to be associated with a point forecast that is too high and vice versa. However, there is also a highly significant interaction between the latest observation minus the series mean and whether scenarios are provided. When scenarios were provided, this bias tended to be exaggerated.

Given that both the participants in Groups 2 and 3 saw both worst case and best case scenarios, this suggests that they only paid attention to the scenario that was consistent with the relative position of the latest observation, while discounting the account in the 'dissenting' scenario. Note that, for all of the series, the slope of the last segment (i.e., the difference between the last two observations) was always upwards when the latest observation was above the series mean and always downwards when it was below. So an upward sloping segment would also have been seen as being consistent with a best case scenario and vice versa. Thus it appears that the scenarios were damaging point forecast accuracy by increasing the tendency of the forecasters to overweight the most recent observation in the time series, even though two conflicting possible futures were suggested as possibilities. Yet, in the exit questionnaire, participants who received scenarios indicated that they perceived them as being helpful (as shown in Table 7).

	weak opt. & weak pess. (G2)	strong opt. & strong pess. (G3)
Scenarios enhanced future-focused thinking	4.24	4.20
Scenarios were useful for constructing forecasts	4.38	4.25
Scenarios were clear to understand	4.19	4.45
Scenarios were realistic	4.24	4.20
Scenarios provided important additional information	4.09	4.10

Table 7. Mean ratings regarding perceptions of scenarios [1=totally disagree to 5=totally agree]

Finally, surprisingly, there was no significant difference between participants who received 'weak' scenarios and those who received 'strong' scenarios, either in their perception of the usefulness of the scenarios or in the effect on the forecasts and decisions. This may reflect a failure in the experimental manipulation or it may suggest that people have difficulties in translating the strength of the verbal arguments in a scenario into quantitative estimates. Further research would be needed to investigate whether this was the case.

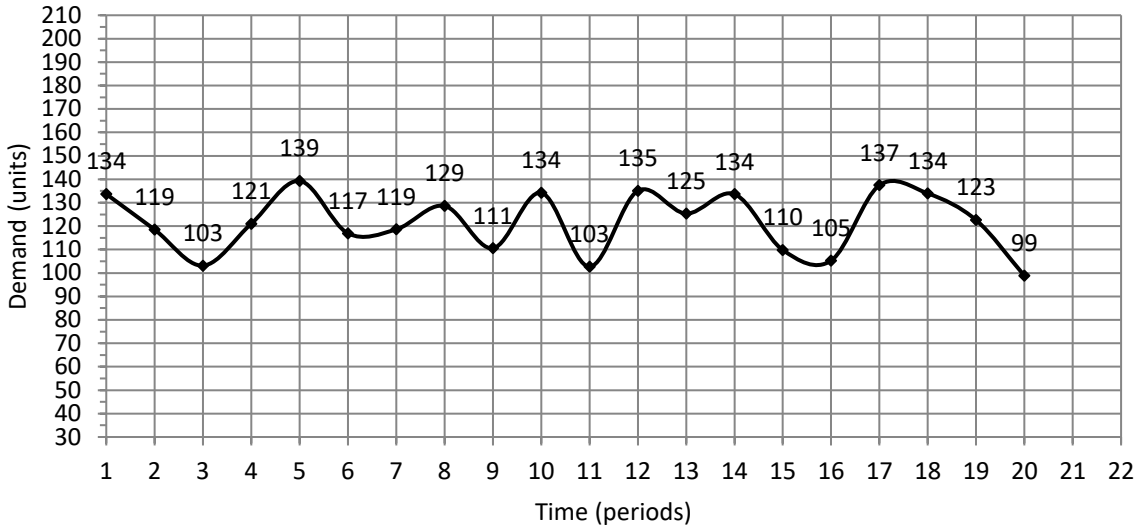
6. Conclusions

This paper has shown that providing forecasters with more information, in this case in the form of scenarios, can be inimical to the accuracy of point forecasts and the quality of decisions that are associated with them. In particular, the scenarios did not reduce overconfidence - a finding that is consistent with those of several other studies. The damaging effect of the scenarios appears to have occurred because, when estimating their point forecasts, the participants focussed on the scenario that was congruent with the most recent behaviour of the time series and discounted the alternative. The result was a reinforcement of the biases commonly reported in judgmental time series extrapolation, namely an overweighting of the last observation in the series and a belief that a random movement in time series represented a systematic change.

Further work could usefully investigate how such biases might be overcome so that value could be extracted from the scenarios. For example, in the experiment reported here, scenarios were supplied to participants. Some researchers have found that asking people to generate their own scenarios increases their plausibility and reduces overconfidence (Newby-Clark et al., 2000). In addition, many forecasts and production decisions are made by groups of people (Fildes et al., 2009) so the research could be extended to a group context, possibly involving 'devil's advocates' who ensure that their optimistic and pessimistic scenarios receive sufficient air time. Alternatively asking two independent groups to work on the basis of optimistic and pessimistic scenarios, respectively might ensure that the full range of uncertainty is addressed prior to making a production decision. Yet another alternative would be to allow the interaction of statistical methods and management judgment in such a way that their respective strengths are harnessed appropriately. Examining designs for effectively engaging scenarios in forecasting will contribute towards developing improved support systems for translating good forecasts to winning decisions.

Appendix B: Sample form given to G1 participants

PRODUCT K



YOUR FORECAST :

What is your *point forecast* for period 21 : ……

What is your confidence (probabilistic estimate) that

the realized value would be within ± 5% of your point forecast: …… (between 0% and 100%)

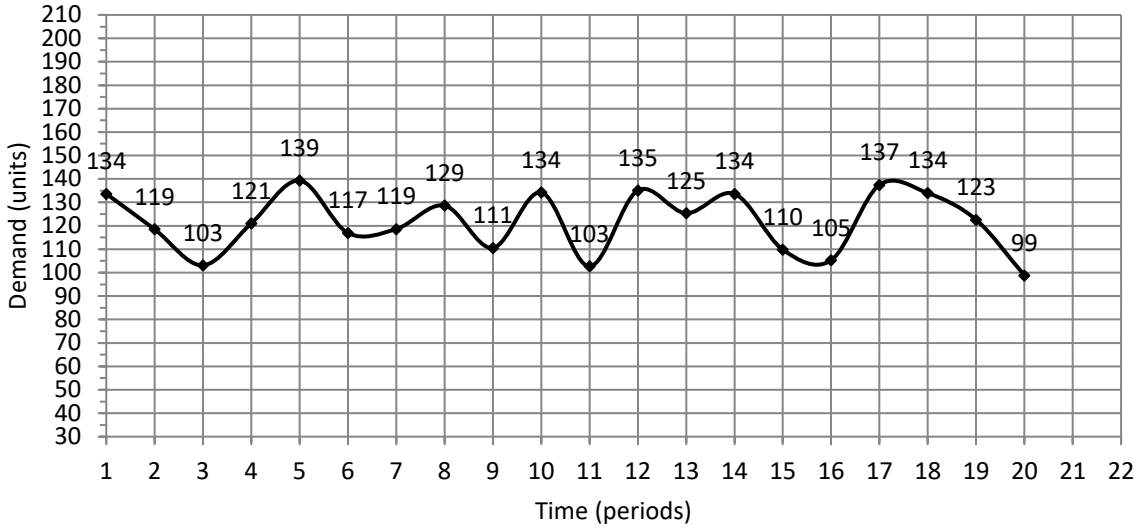
YOUR PRODUCTION DECISION

How many units will you order for production? …… (between 0 and 750)

(Please note that *total production capacity* for period 21 is 750 units. Therefore your production orders for all six products should add up to a maximum of 750. Please keep in mind that there are different costs associated with over-production vs. under-production and make your decisions accordingly. Please use the checklist in the end for production plans)

Appendix B: Sample form given to G2 participants

PRODUCT K



Scenario A:

Product K, a mobile phone with multifaceted functionality, has stable demand. It has a combination of necessary features to compete successfully in its target market. It is a nicely designed phone with full-fledged features, and comes with a fairly positioned price and promotion package. It receives positive comments in the industry magazines/websites and good feedback from customers. Given the recent economic conditions, it is possible that there may be a slightly higher demand for this product in the periods to come.

Scenario B:

This product has been serving its purpose and target market for a long time. While its customers are happy with it, its sales performance is stable within a reasonable band. It could have continued like this for a sufficiently long time. However, our company has been experiencing some problems with a major supplier, which happens to be the producer of a key part for this model. If the dispute cannot be solved shortly, we may not be able to produce Product K until we find another supplier with similar credentials. And it could take some time until (a) we find such a supplier, and (b) it starts delivering the required parts. If the customers learn about this problem, it is likely to face a somewhat lower demand in the next period.

YOUR FORECAST :

What is your *point forecast* for period 21 : ……

What is your confidence (probabilistic estimate) that

the realized value would be within ± 5% of your point forecast: …… (between 0% and 100%)

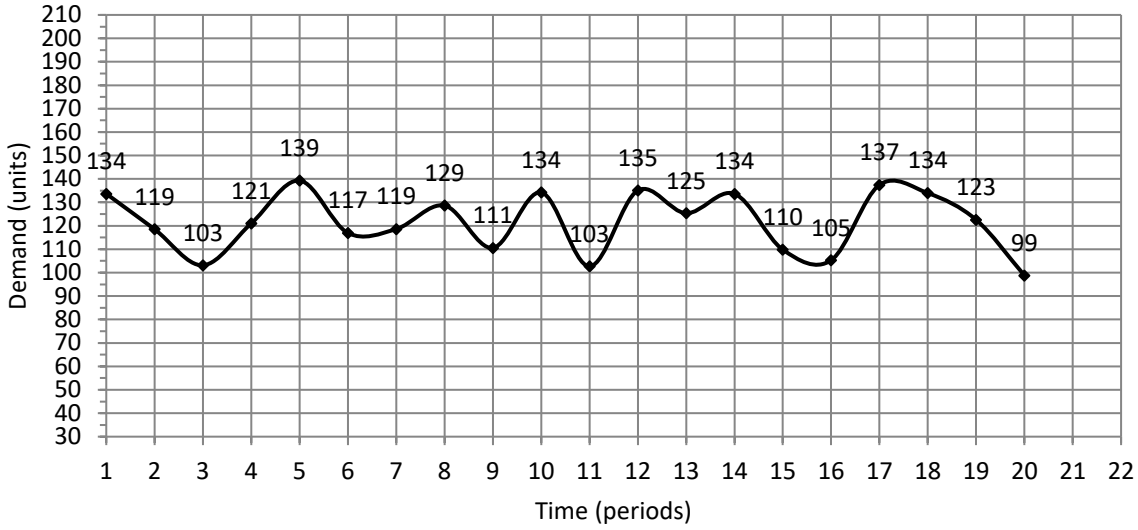
YOUR PRODUCTION DECISION

How many units will you order for production? …… (between 0 and 750)

(Please note that *total production capacity* for period 21 is 750 units. Therefore your production orders for all six products should add up to a maximum of 750. Please keep in mind that there are different costs associated with over-production vs. under-production and make your decisions accordingly. Please use the checklist in the end for production plans)

Appendix C: Sample form given to G3 participants

PRODUCT K



Scenario A:

Product K, a mobile phone with multifaceted functionality, has extremely stable demand. It has got all that is necessary to compete very successfully in its target market. It is an attractively designed phone with full-fledged features, and comes with a nicely positioned price and exceptionally encouraging promotion package. It regularly receives exceedingly positive comments in the industry magazines/websites and first-class feedback from customers. Given the recent economic conditions, we strongly expect even higher demand for this product in the periods to come.

Scenario B:

This product has been serving its purpose and target market for a long time. Its customers seem to be satisfied with it and its sales performance is stable within a band. It could have continued like this for some time. However, our company has been experiencing vital problems with a major supplier, which happens to be the producer of a key part for this model. If this dispute cannot be solved shortly, we certainly will not be able to produce Product K until we find another supplier with equally good credentials. While it is very difficult to replace the existing one, it will certainly take some time until (a) we find such a supplier, and (b) it starts delivering the required parts. If customers learn about this problem, there is a very high possibility that we will be faced with significantly lower demand in the next period.

YOUR FORECAST :

What is your *point forecast* for period 21 : ……

What is your confidence (probabilistic estimate) that

the realized value would be within ± 5% of your point forecast: …… (between 0% and 100%)

YOUR PRODUCTION DECISION

How many units will you order for production? …… (between 0 and 750)

(Please note that *total production capacity* for period 21 is 750 units. Therefore your production orders for all six products should add up to a maximum of 750. Please keep in mind that there are different costs associated with over-production vs. under-production and make your decisions accordingly. Please use the checklist in the end for production plans)

References

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In J. S. Armstrong (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 495-515). Boston, MA: Springer US.

Bolger, F., & Harvey, N. (1993). Context-sensitive heuristics in statistical reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 779-811.

Davis, F. D., Lohse, G. L., & Kottemann, J. E. (1994). Harmful effects of seemingly helpful information on forecasts of stock earnings. *Journal of Economic Psychology*, *15*(2), 253-267.

Danese, P., & Kalchschmidt, M. (2011). The role of the forecasting process in improving forecast accuracy and operational performance. *International Journal of Production Economics*, *131*(1), 204-214.

Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, *37*(6), 570-576.

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*, *25*(1), 3-23.

Fildes, R., Goodwin, P. & Önkal, D. (in press) Use and misuse of information in supply chain forecasting of promotion effects. *International Journal of Forecasting*.

Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, *42*(3), 1481-1493.

Goodwin, P., & Wright, G. (2001). Enhancing strategy evaluation in scenario planning: a role for decision analysis. *Journal of management studies*, *38*(1), 1-16.

Goodwin, P. and Wright, G. (2014). *Decision Analysis for Management Judgment*, 5th edition. Chichester; Wiley.

Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of personality and social psychology*, *59*(6), 1128.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, *21*(1/2), 71-87.

Klassen, R. D., & Flores, B. E. (2001). Forecasting practices of Canadian firms: Survey results and comparisons. *International Journal of Production Economics*, *70*(2), 163-174.

Kuhn, K. M., & Sniezek, J. A. (1996). Confidence and uncertainty in judgmental forecasting: Differential effects of scenario presentation. *Journal of Behavioral Decision Making*, *9*(4), 231-247.

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493-518.

Lawrence, M., & Makridakis, S. (1989). Factors affecting judgmental forecasts and confidence intervals. *Organizational Behavior and Human Decision Processes*, *43*(2), 172-187.

Lawrence, M., & O'Connor, M. (1992). Exploring judgemental forecasting. *International Journal of Forecasting*, *8*(1), 15-26.

Moore, D. A., Carter, A. B., & Yang, H. H. J. (2015). Wide of the mark: Evidence on the underlying causes of overprecision in judgment. *Organizational Behavior and Human Decision Processes*, *131*, 110-120.

Newby-Clark, I. R., Ross, M., Buehler, R., Koehler, D. J., & Griffin, D. (2000). People focus on optimistic scenarios and disregard pessimistic scenarios while predicting task completion times. *Journal of Experimental Psychology: Applied*, *6*(3), 171.

Önkal, D., Gönül, M. S., & Lawrence, M. (2008). Judgmental adjustments of previously adjusted forecasts. *Decision Sciences*, *39*(2), 213-238.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390-409.

Önkal, D., Sayım, K. Z., & Gönül, M. S. (2013). Scenarios as channels of forecast advice. *Technological Forecasting and Social Change*, *80*(4), 772-788.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision maker*. New York: Cambridge University Press.

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7.

Sanders, N. R., & Manrodt, K. B. (2003). Forecasting software in practice: Use, satisfaction, and performance. *Interfaces*, *33*(5), 90-93.

Satterfield, T., Slovic, P., & Gregory, R. (2000). Narrative valuation in a policy judgment context. *Ecological Economics*, *34*(3), 315-331.

Schnaars, S. P. (1987). How to develop and use scenarios. *Long range planning*, *20*(1), 105-114.

Schnaars, S. P., & Topol, M. T. (1987). The use of multiple scenarios in sales forecasting: An empirical test. *International Journal of Forecasting*, *3*(3-4), 405-419.

Schoemaker, P. J. (1993). Multiple scenario development: Its conceptual and behavioral foundation. *Strategic Management Journal*, *14*(3), 193-213.

Slovic, P. (1966). Cue-consistency and cue-utilization in judgment. *The American Journal of Psychology*, *79*(3), 427-434.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology. Learning, Memory & Cognition*, *30*(2), 299-314.

Tomotani, J. V., & de Mesquita, M. A. (2018). Lot sizing and scheduling: a survey of practices in Brazilian companies. *Production Planning & Control*, *29*(3), 236-246.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131.

Wu, C. S., & Cheng, F. F. (2011). The joint effect of framing and anchoring on internet buyers' decision-making. *Electronic Commerce Research and Applications*, *10*(3), 358-368.

Xu, Y. C., Cai, S., & Kim, H. W. (2013). Cue consistency and page value perception: Implications for web-based catalog design. *Information & Management*, *50*(1), 33-42.

Yaniv, I. (2004). The benefit of additional opinions. *Current Directions in Psychological Science*, *13*(2), 75-78.