European Journal of Operational Research 205 (2010) 195–201

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Do forecasts expressed as prediction intervals improve production planning decisions?

Paul Goodwin [a,*], Dilek Önkal [b], Mary Thomson [c]

[a] School of Management, University of Bath, Bath BA2 7AY, UK
[b] Brunel Business School, Brunel University, Uxbridge, Middlesex UB8 3PH, UK
[c] Division of Psychology, Glasgow Caledonian University, Cowcaddens Road, Glasgow G4 0BA, UK

ARTICLE INFO

Article history:
Received 17 April 2009
Accepted 17 December 2009
Available online 28 December 2009

Keywords:
Forecasting
Prediction intervals
Decision making
Asymmetric loss

ABSTRACT

A number of studies have shown that providing point forecasts to decision makers can lead to improved production planning decisions. However, point forecasts do not convey information about the level of uncertainty that is associated with forecasts. In theory, the provision of prediction intervals, in addition to point forecasts, should therefore lead to further enhancements in decision quality. To test whether this is the case in practice, participants in an experiment were asked to decide on the production levels that were needed to meet the following week's demand for a series of products. Either underproduction cost twice as much per unit as overproduction or vice versa. The participants were supplied with either a point forecast, a 50% prediction interval, or a 95% prediction interval for the following week's demand. The prediction intervals did not improve the quality of the decisions and also reduced the propensity of the decision makers to respond appropriately to the asymmetry in the loss function. A simple heuristic is suggested to allow people to make more effective use of prediction intervals. It is found that applying this heuristic to 85% prediction intervals would lead to nearly optimal decisions.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

The prime purpose of forecasting is to support decision making, but how effective is the support that forecasts provide? One decision which is commonly based on forecasting is the production planning decision where managers have to decide upon the level of production to meet demand in a future period. Overproduction can lead to excessive inventory costs while underproduction can lead to loss of business or the need to pay for emergency supplies. An analogous decision is faced by managers who have to determine the inventory levels that are required to meet demand in a future period. In many companies both of these decisions are made within the framework of sales and operations planning (S&OP) which is a cross-functional process designed to balance demand and supply while integrating operational and financial plans (e.g. see Slone et al., 2007). The most commonly used forecasting format in companies is the point forecast (e.g. Klassen and Flores, 2001; Dalrymple, 1987). This merely indicates a single possible future value for the variable of interest and provides no information on the amount of uncertainty surrounding that indication. Because of this, the use of more informative forecasting formats have been advocated, including fan charts, which display bands showing the probabilities that particular ranges will contain the realised value of the forecast variable (e.g., Wallis, 2003), density forecasts, which provide information on the entire probability distribution for this variable (e.g., Ericsson, 2003; Hua and Zhang, 2008), and probabilistic directional forecasts that convey the expected direction of change as well as its probability of occurrence (e.g., Önkal et al., 2003).

However, the prediction interval is the alternative to point forecasts that is most widely available in forecasting software and referred to most often in the literature, its use is also recommended by researchers (e.g. Armstrong, 2001) and a considerable amount of research effort has been devoted to obtaining reliable prediction intervals in different situations (e.g. see De Gooijer and Hyndman, 2006). Prediction intervals have the clear potential to be helpful to decision makers in a variety of situations. For example, a 99% prediction interval which has a positive lower limit will indicate that negative profits or negative growth in sales is unlikely and that preventive action may be unwarranted. Alternatively, when choosing between courses of action, comparisons between the prediction intervals of the outcomes for each course of action may indicate that one option stochastically dominates others or that one option is considerably less risky despite offering only a slightly lower expected payoff than others (see also Finch and Gavirneni, 2006). Even very wide prediction intervals may be helpful when they signal to a decision maker that more data needs to be obtained or more research needs to be carried out before an effective decision can be made.

* Corresponding author. Tel.: +44 122 538 3594; fax: +44 122 582 6473.
E-mail addresses: mnspg@management.bath.ac.uk (P. Goodwin), dilekon@gmail.com (D. Önkal), M.Thomson@gcal.ac.uk (M. Thomson).

0377-2217/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejor.2009.12.020

However, none of these situations corresponds directly to the production planner's decision on how much should be produced to meet future demand. Here, the potential usefulness of prediction intervals is less evident. Perhaps because of this, a study by Dalrymple (1987) found that only about 10% of companies regularly use prediction intervals in their sales forecasting. Similarly, Klassen and Flores (2001), in discussing the results of a survey of Canadian forecasting practices, stated "It seems, thus, that academics have not convinced the practitioners to use [prediction] intervals for the forecasts. Or the practitioners have found no real value added in the range" (p. 172). The basis of the recommendation to use prediction intervals by Armstrong (2001) is "received wisdom" and we know of no empirical study which has assessed the value of prediction intervals in business decision making. This paper aims partly to fill that gap by reporting the results of an experiment that was designed to address the important question of whether the provision of prediction intervals leads to improvements in production planning decisions when compared to decisions based on point forecasts. The paper also considers how the usefulness of prediction intervals to production planners might be enhanced.

The next section of the paper provides a review of the relevant literature. This is followed by a discussion of the experiment and its results. Then a simple heuristic is developed which is designed to increase the usefulness of prediction intervals in production planning. Finally, a case is made for the wider use of 85% prediction intervals as opposed to the 50%, 95% or 99% intervals that are commonly presented.

2. Literature review

The key role of forecasting is to help decision makers who face uncertainty about the future (Armstrong, 2001). Production planners usually face uncertainty when they have to decide how much to produce to meet demand in a given future period. This decision can be a cognitively demanding task. If carried out without support, the decision maker needs to: (i) make a prediction of the expected level of demand which will occur in the future period, (ii) assess the uncertainty associated with this prediction, and then (iii) combine this assessment with utility or cost functions, which are usually asymmetric, to determine the optimum level of production.

Supplying a point forecast to the decision maker should either remove the need to carry out the first of these stages or at least provide guidance on what should be predicted, thereby reducing the demands of the decision making task. Consistent with this, studies by Moskowitz (1972), Moskowitz and Miller (1975) and Goodwin (2005) found that providing point sales forecasts did lead to improved production planning decisions.[1]

However, point forecasts, in themselves, provide no information relating to the second component of the decision maker's task: assessing the amount of uncertainty associated with the prediction. If a display of the past history of a demand and its associated point forecasts is available, the decision maker may be able to make a rough visual assessment of how reliable the forecast is likely to be, but Goodwin (2005) and Lawrence and Makridakis (1989) found that people were poor at estimating this degree of reliability and were influenced by irrelevant factors such as the amount of space appearing above and below the most recent observation on the graphical display.

Prediction intervals do provide information on the degree of uncertainty associated with point forecasts and many people have therefore advocated their widespread provision and use by decision makers. For example, Principle 14.1 from the *Forecasting Principles Handbook* (Armstrong, 2001) states that one should "Estimate prediction intervals ... to improve use of forecasts...when decisions are affected by uncertainty" (pp. 718–719). In addition, participants in a study by Önkal and Bolger (2004) rated prediction intervals as being more useful than point forecasts. This was the case for participants in both the forecast provider and the forecast user roles.

The potential usefulness of prediction intervals has motivated several researchers to develop methods designed to ensure that well-calibrated intervals can be calculated in particular circumstances. For example, Gardner (1988) presented a simple method, based on Chebyshev's inequality which required few assumptions about the nature of the time series while Hyndman et al. (2005) have developed formulae for the calculation of prediction intervals for forecasts based on exponential smoothing.

However, despite the elegance of these techniques, there is no guarantee that decision makers will make appropriate use of the resulting intervals. Prospect theory (Kahneman and Tversky, 1979) suggests that probabilities provided to decision makers are commonly distorted. For example, relative to certainty, probabilities, such as 50%, 95% and 99%, which are typically used in prediction intervals, tend to be underweighted in decisions. This implies that the information provided by the interval may receive less attention than would be suggested by a normative decision model. Indeed, there is no guarantee that decision makers will take any account of prediction intervals at all. Fischhoff (1994) investigated why people sometimes pay scant attention to forecasts expressed in probabilistic terms. One reason was the perceived irrelevance of the forecasts. In contrast to point forecasts, wide prediction intervals may be seen as so lacking in definitiveness that they are seen as irrelevant and uninformative (Rush and Page, 1979; Yaniv and Foster, 1995, 1997) particularly by decision makers who have an intolerance of ambiguity (e.g. Lal and Hassell, 1998). Indeed, Granger (1996) has described 95% prediction intervals as often being 'embarrassingly wide' and argued that 50% intervals are more likely to be believable.

Clearly, to ensure the relevance of a given prediction interval in a particular context, the nature of the task faced by the decision maker needs to be addressed. Rather than relying on standard intervals such as 50% and 95% , another way of achieving relevance is to tailor the coverage probability of the interval to meet the specific needs of the decision maker. Landon and Singpurwalla (2008) have proposed a prototypical approach to determining this probability. The appropriate coverage probability is the one which minimises the decision maker's expected total disutility that arises from: (i) the width of the interval – wider intervals, being less precise are associated with greater disutility, and (ii) the outcome falling outside the interval – a larger distance between the outcome and the nearest limit will incur greater disutility.

However, many forecasting software products used by production planners offer prediction intervals with a fixed coverage probability: most commonly 95%. Even where the 95% interval is only the default and other coverage probabilities are available, it is likely to be unclear to the planner why an alternative value should be chosen or what this value should be. While Landon and Singpurwalla (2008) indicate that the motivation for their approach is to support decision making (including decisions relating to inventory control and production planning), they do not articulate the connection between the prediction interval's width and the decision. Indeed the mathematics they use to optimise the coverage probability could, arguably, be better used directly to identify the optimum decision itself. For example, if a decision is to be based on the relative costs (or disutilities) of over and under stocking, their approach would focus on the relative 'costs' of having an interval that is too wide or too narrow –

[1] An alternative explored by Flores et al. (1993) involved the use of statistical forecasts that were designed to minimize asymmetric loss. This would effectively automate all three of the above elements of the decision, but such methods are not generally available in commercial software products.

the connection of these costs to the costs associated with the decision is unclear. Also, the implicit assumption is that, once an 'appropriate' interval has been identified, the decision maker will somehow be able to use it to arrive at the best decision. In addition, their approach may be too complex to apply in many practical contexts, particularly where a large number of different decisions need to be made quickly and frequently.

Given the difficulties of determining an appropriate coverage probability for a given decision and given the fact that 'fixed' intervals with standard coverage probabilities are usually automatically provided by software, an important practical question arises. Does the provision of prediction intervals with standard coverage probabilities lead to improvements in the quality of production planning decisions over and above that achieved by providing point forecasts? Formally, we test the following hypothesis:

H1: Decision makers supplied with point forecasts and either 50% or 95% prediction intervals do not achieve lower costs than those supplied only with point forecasts.

3. Details of the experiment

The design of the experiment was based on the one used by Goodwin (2005), although that study did not investigate the usefulness of prediction intervals. Participants were required to make a decision on the level of production that was required to meet next week's demand for each of ten products. 56 business students at Bilkent University in Ankara, Turkey participated in the experiment. A study by Remus (1986) found that students can act as reliable proxies for managers in decision making tasks. Similar results have been reported in a study by Mowen and Mowen (1986) while Ruchala (1999) found that student participants are less likely to be influenced by organizational and external issues, which are not relevant to the questions being investigated in the experiment.

All of the participants in the current experiment were familiar with the prediction interval concept through the course that they were studying and they had experience of generating prediction intervals using statistical software and interpreting the results. The participants were randomly assigned to one of six treatments which involved either a 'shortage worse' or a 'surplus worse' pair of asymmetric loss functions crossed with the availability of one of three types of demand forecast for the following week.

For the *"shortage worse"* loss functions, each unit that was surplus to requirements cost \$10, while producing too few units cost \$20 for each unit that was not available to meet demand. For the *"surplus worse"* loss functions, the unit costs of surplus and shortage were \$20 and \$10, respectively. Depending on the treatment, the one-week ahead forecast supplied to the participants was either (i) a point forecast, (ii) a point forecast accompanied by a 50% prediction interval, or (iii) a point forecast accompanied by a 95% prediction interval.

All of the participants ran a computer program which showed graphs of the demand for each of 10 products in turn in the previous 20 weeks (i.e. weeks 0–19). Superimposed on this was a graph of statistical point forecasts for these weeks and for week 20. These forecasts were obtained using the expert system in the *Forecast-Pro*

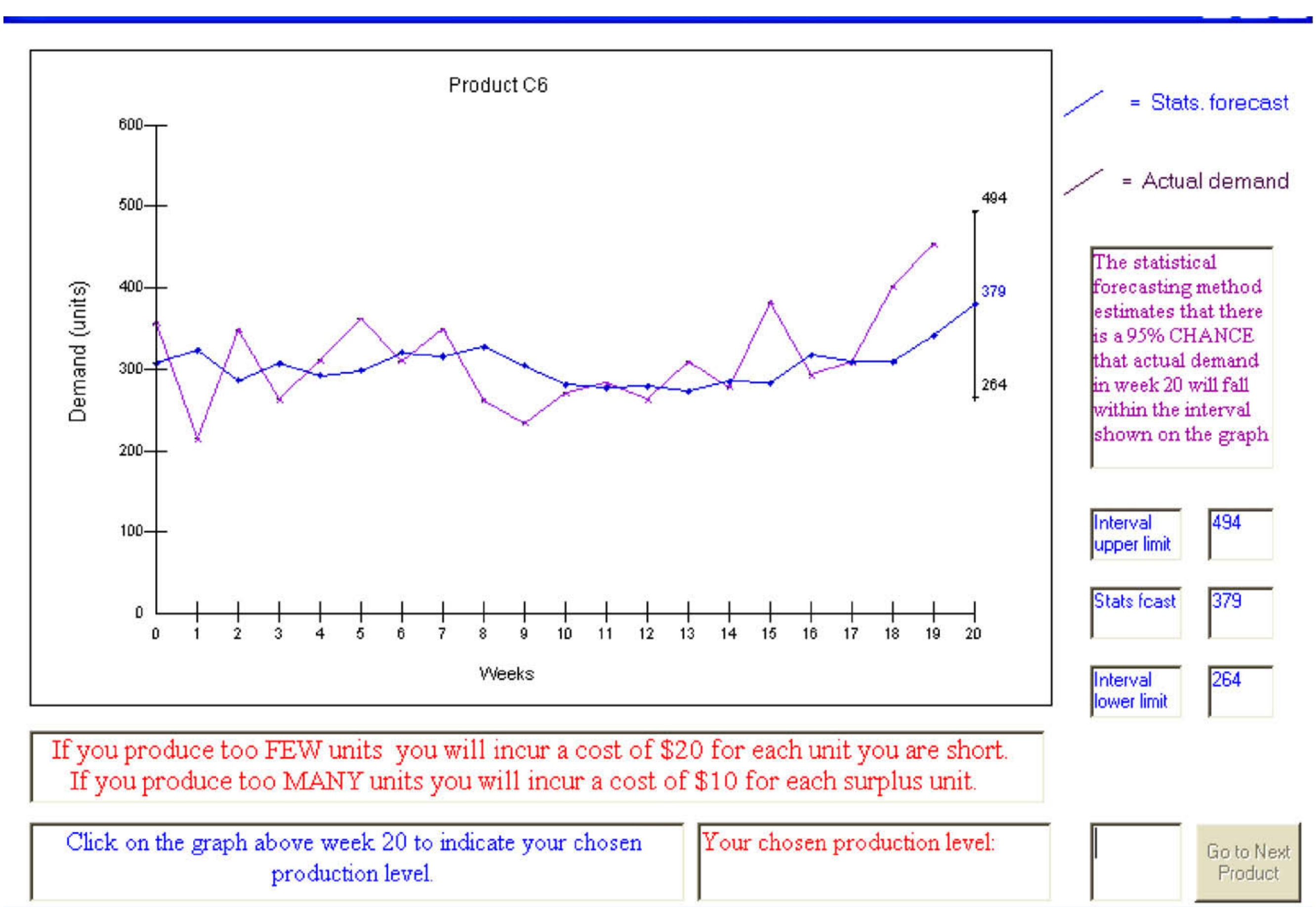

Fig. 1. A typical display.

package (Stellwagen and Goodrich, 1994). Where appropriate, a prediction interval was also displayed. The order of display of the demand for each product was randomised. Fig. 1 shows a typical display for the surplus worst, 95% interval treatment.

Participants were asked to decide on the level of production that was needed to cater for demand in week 20. In all cases, they input their decision by using the mouse to click on the graph at the appropriate point and, until they had confirmed their judgments by clicking on a command button, they were free to make changes.

Ten simulated demand series were generated for the products. These were intended to reflect a range of typical patterns found in product demand time series. The use of simulated series not only allowed participants' decisions to be investigated under a variety of controlled conditions, but it also had the advantage that the true probability distribution of next period's demand was known. This enabled the expected costs of participants' decisions to be assessed. The ten series included two AR(1) series, two series exhibiting a linear upward trend, two exhibiting a linear downward trend and four white noise series which exhibited a step change in the mean in the last two periods displayed (upwards in two cases and downwards in the other two). In all cases, noise was added to the underlying time series pattern. This was independently sampled from a normal distribution with a mean of zero and a standard deviation of 45. Details of the series generating functions are given in Appendix A. The same graph scales was used to display all the series to avoid the possible confounding effect of different scales.

Before starting the experiment, participants made decisions for a trial run involving five series and received feedback on the costs associated with their decisions. For the experiment itself, no feedback was provided until all ten decisions had been made. No time limit was imposed on the participants, although the time taken to make each decision was recorded.

4. Results

The expected costs of each decision were calculated on the basis that the demand for the next period followed the normal probability distribution implied by the series generating function. A table of partial expectations in the normal distribution (e.g.Brown, 1967; Chase and Aquilano, 1981) was used to calculate the expected shortage and the expected surplus for a given decision and hence the expected total costs. For example, consider the probability distribution of demand for the tenth product in week 20. This was a normal distribution with a mean of 420 units and a standard deviation of 45 units. If a participant decided to produce 447 units (0.6 standard deviations above the mean) the tables show that the expected shortage would be 7.6 units and the expected surplus would be 34.6 units. Depending on the direction of the asymmetry of the loss functions, the expected costs for this decision could then be calculated. Expected costs were considered to be a better measure of the decision makers' performances than the alternative of generating a single demand value from the distribution and determining the associated costs. This is because the alternative would have introduced an unnecessary element of randomness into the performance measure.

The experiment was a 2 (type of loss) $\times$ 3 (type of forecast) factorial design with expected costs over the ten series, as the dependent variable. When ANOVA was applied to the results no significant interaction was found between type of loss and type of forecast ($p = 0.805$). The main effect for type of loss was highly significant ($p = 0.001$) with 'surplus worse' decisions incurring more than double the expected costs of 'shortage worse' decisions. However, the main effect for type of forecasts was not significant ($p = 0.330$). Thus, there was no evidence that providing either a 50% or a 95% prediction interval improved the decisions made by the participants compared to those who only received a point forecast.

5. Discussion

There are two possible reasons why the decisions based on prediction intervals failed to improve significantly on the decisions based on point forecasts. Either the prediction intervals were ignored or they influenced the decisions, but did so in a way that was neither damaging nor beneficial.

To investigate which of these explanations seemed more plausible, the decisions made by the participants were compared with the optimum decisions for each product. Given the linear nature of the loss functions it is easy to show (Winston, 1987, p. 710) that expected costs will be minimised by producing q^* units where q^* is the smallest value of q which satisfies:

$$p(x\ exceeds\ q^*) \geqslant C_s/(C_s + C_u)$$

where x the level of demand, C_s the cost of each surplus unit and C_u the cost incurred for each shortage unit. For the "shortage worse" treatment this indicates that the production level should be set at a level of demand that has a 1/3 probability of being exceeded. The corresponding probability for the "surplus worse" treatment is 2/3.

Two attributes can be discerned to assess the decisions based on the three types of forecast: (i) bias, and (ii) discrimination. Bias is the extent to which a decision tended to be too high or too low compared to the optimum. We define discrimination as the extent to which the decisions for the shortage worse situation differed from those where surplus was worse. Goodwin (2005) found that decision makers supplied with point forecasts *were* able to respond in the direction that was appropriate for the asymmetric loss functions that they faced. For example, in a 'shortage worse' situation they would, correctly, tend to set the decision above the point forecast.

Table 1 shows the mean percentage bias (MPB) of the decisions made compared to the optimum decision, where:

$$\text{MPB} = 100 \times \frac{\text{Actual decision} - \text{Optimum decision}}{\text{Optimum decision}}$$

It can be seen that, irrespective of the type of forecast provided, there was a tendency to set the production level too high and that this was particularly the case when a surplus was worse than a shortage (this replicates the finding in Goodwin (2005) and a discussion of the possible reasons for this can be found in that paper). It can be seen that shortage worst decisions were, on average, set at just about the correct level when a 95% interval was provided. It is not clear why this was the case, though as we argue below, the longer bars of the 95% interval (they were nearly three times longer than those of the 50% interval) may have made the uncertainty associated with demand more salient and hence led to a hedging strategy so that decisions were set closer to the middle of the

Table 1
Mean percentage bias (MPB) compared to optimum decisions.

Loss function	Point	50%	95%
	Forecast	Interval	Interval
Shortage worst	4.8	4.7	0.7
Surplus worst	10.9	12.9	12.9

range. The effect of this would be to mitigate the upward bias. Apart from this, the bias of decisions based on the three types of forecasts is quite similar.

A different picture emerges when one considers the tendency of the decisions to discriminate between 'shortage worst' and 'surplus worst' situations. The mean discrimination for decisions based on the two prediction intervals was compared with those based on point forecasts using the following measure:

$$\text{Relative discrimination} = 100 \times \frac{\text{Mean shortage worst decision} - \text{Mean surplus worst decision})\text{based on interval}}{(\text{Mean shortage worst decision} - \text{Mean surplus worst decision})\text{based on point forecast}}$$

For the 50% interval the relative discrimination was 83.8%, but for the 95% interval it was only 44.1%. Thus, the effect of providing prediction intervals appears to have reduced the tendency of the participants to respond to the asymmetry of the loss functions when they made their decision, particularly when they were provided with a 95% interval. It thus appears that, the 95% intervals (at least) were not ignored. However, while they reduced bias for the shortage worse condition, they also impeded the decision makers' propensity to respond to asymmetric loss. Further work will be needed to establish why the intervals had this effect. It is possible that the longer bars of the 95% interval which appeared either side of the point forecast on the graph (see Fig. 1) may have drawn participants' attention to the fact that actual demand could occur anywhere within a range that included demand that was well above or well below the point forecast. As a result, they may have 'hedged their bets' and kept their decisions closer to the middle of the interval.

6. Improving decision makers' ability to make effective use of prediction intervals

While most researchers recommend the provision of prediction intervals to decision makers, our results suggest that the availability of an interval will not lead to a better production planning decision. This raises the question: is it possible to provide simple, approximate, guidance to production planners, who are presented with prediction intervals having predetermined coverage probabilities? Assuming that linear loss functions apply (as in the above experiment), a plausible heuristic might be to calculate the ratio $C_u/(C_s + C_u)$ and to set the decision at this fraction between the lower and upper limits of the interval (note that this ratio is one minus the ratio given earlier). For example, if each shortage unit costs twice as much as each surplus unit then the ratio will be 2/3 and the decision could be set 2/3 of the way between the lower and upper limits.

Fig. 2 shows how well this heuristic would work, for a range of different loss asymmetries, if it was applied to a perfectly calibrated 95% prediction interval in a 'surplus worst' situation when demand was normally distributed. Because of symmetry, an analogous result would apply for 'shortage worst'. The graph shows where the optimum decision would be positioned between the lower and upper limits of the interval and where the decision based on the heuristic would be positioned. For example a value of 0.4 on the y-axis indicates that the decision is 40% of the way between the lower and upper limits. It can be seen that the heuristic works best when the asymmetry is relatively mild, but also that there is a convergence between the heuristic and the optimum

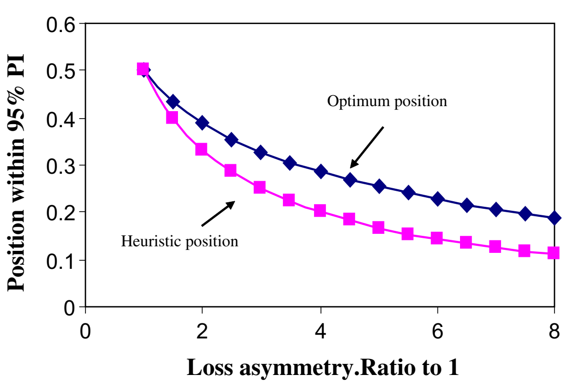

Fig. 2. Performance of heuristic for different loss asymmetries.

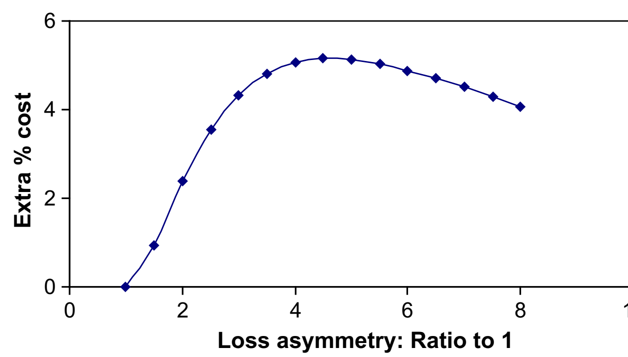

Fig. 3. Extra percentage cost incurred by using heuristic for different loss asymmetries.

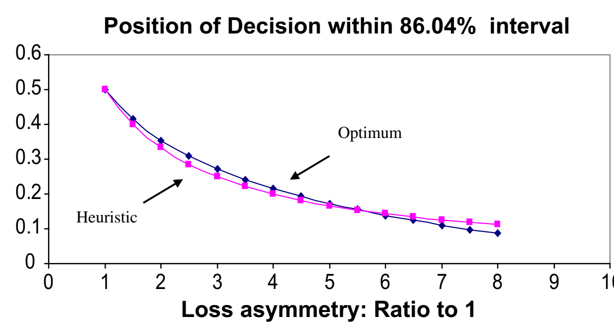

Fig. 4. Accuracy of heuristic when used with 86.04% interval.

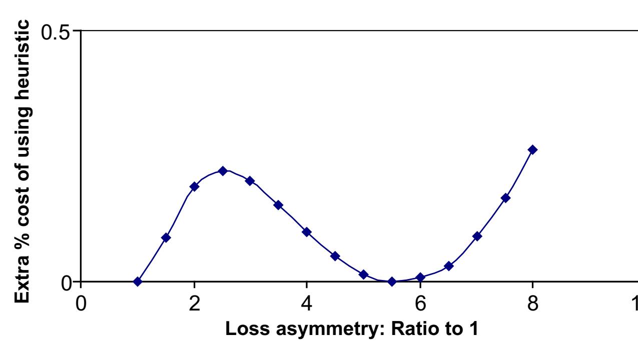

Fig. 5. Costs of using heuristic with 86.04% prediction interval.

for extreme asymmetries. Fig. 3 shows the extra percentage cost that would be incurred by using the heuristic compared to the optimally positioned decisions. It can be seen that this is, at worst, little more than 5% and is much less for moderate asymmetries. In practice, the significance of these extra costs would be obscured by the less-than-perfect calibration of the prediction interval, the fact that actual demand would be likely to depart from a normal distribution or the likelihood that the true loss functions may not be perfectly linear.

However, the encouraging performance of the heuristic raises another possibility. Is it possible to find a coverage probability for the interval which will enable the heuristic to be used to maximum effect? An optimisation algorithm was used to find the coverage probability that minimised the mean squared difference between the position of the heuristic in the interval and the optimum position for the loss asymmetries displayed in Fig. 2 (above). The resulting coverage probability was 86.04%. Figs. 4 and 5 show how close the positions of the heuristic decision and the optimum decision are within the interval and that the extra percentage cost of using the heuristic would be at most 0.22%. In practice, it would be 'neater' for software suppliers to provide 85% or 86% intervals – the effects of this rounding of the coverage probability would be negligible in most circumstances.

7. Conclusions

Forecasts are intended to be used by decision makers and their presentation should be designed with that in mind. Although widely recommended, providing 'standard' prediction intervals without guidance to decision makers may have no benefits and may even have negative effects such as reducing their responsiveness to the direction of asymmetric loss. In particular, the widespread provision of 95% intervals appears to be based purely on convention. However, the study suggests that there may be advantages, in a wide range of decision making contexts, of presenting intervals with alternative coverage probabilities, such as 85%, *together* with simple guidance on how to use the intervals.

Of course, the inferences that can be drawn from the experiment may be limited by the nature of the loss functions that applied, the simulated series used – in particular none of the series had a seasonal pattern – and the fact that the study was carried out in a laboratory, rather than a commercial environment where pressures to make the decisions quickly may apply. In addition, although business students have been found to act as good proxies for managers in experiments such as this, they do lack the domain knowledge that managers would possess. They will also be less familiar with a task that may be carried out regularly by practitioners. Thus their performance may underestimate that which would be achieved by practicing managers.

Appendix A. Time series generating functions

These series were also used in the study carried out by Goodwin (2005).

Y_t the number of units demanded at period t; e_t is the noise at period *t*.

For all series: $e_t \sim N(0, 45)$.

AR(1) series

$$Y_t = 150 + 0.5Y_{t-1} + e_t$$

Linear upward trend

$$Y_t = 150 + 10t + e_t$$

Linear downward trend

$$Y_t = 350 - 10t + e_t$$

White noise with downward step

$$Y_t \sim N\ (300, 45) \text{ for t = 0 to 17.} \quad \text{From t = 18 : } Y_t \sim N\ (180, 45)$$

White noise with upward step

$$Y_t \sim N\ (300, 45) \text{ for t = 0 to 17.} \quad \text{From t = 18 : } Y_t \sim N\ (420, 45)$$

References

Armstrong, J.S., 2001. Principles of Forecasting. Kluwer Academic Publishers, Boston.

Brown, R.G., 1967. Decision Rules for Inventory Management, Holt. Rinehart and Winston, New York.

Chase, R.B., Aquilano, N.J., 1981. Production and Operations Management: A Life Cycle Approach, third ed. Richard D. Irwin, Inc., Homewood, Illinois.

Dalrymple, D.J., 1987. Sales forecasting practices, results from a United States survey. International Journal of Forecasting 3, 379–391.

De Gooijer, J.G., Hyndman, R.J., 2006. 25 years of time series forecasting. International Journal of Forecasting 22, 443–473.

Ericsson, N.R., 2003. Forecast uncertainty in economic modelling. In: Hendry, D.F., Ericsson, N.R. (Eds.), Understanding Economic Forecasts. MIT Press, Cambridge, pp. 68–92.

Finch, B., Gavirneni, S., 2006. Confidence intervals for optimal selection amongst alternatives with stochastic variable costs. International Journal of Production Research 44, 4329–4342.

Fischhoff, B., 1994. What forecasts (seem to) mean. International Journal of Forecasting 10, 387–403.

Flores, B.E., Olson, D.L., Pearce, S.L., 1993. Use of cost and accuracy measures in forecasting method selection – a physical distribution example. International Journal of Production Research 31, 139–160.

Gardner, E.S., 1988. A simple method of computing prediction intervals for time series forecasts. Management Science 34, 541–546.

Goodwin, P., 2005. Providing support for decisions based on time series information under conditions of asymmetric loss. European Journal of Operational Research 163, 388–402.

Granger, C.W.J., 1996. Can we improve the perceived quality of economic forecasts? Journal of Applied Econometrics 11, 455–473.

Hua, Z., Zhang, B., 2008. Improving density forecast by modelling asymmetric features: An application to S&P500 returns. European Journal of Operational Research 185, 716–725.

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2005. Prediction intervals for exponential smoothing using two new classes of state space models. Journal of Forecasting 24, 17–37.

Kahneman, D., Tversky, A., 1979. Prospect theory – analysis of decision under risk. Econometrica 47, 263–291.

Klassen, R.D., Flores, B.E., 2001. Forecasting practices of Canadian firms: Survey results and comparisons. International Journal of Production Economics 70, 163–174.

Lal, M., Hassell, L., 1998. The joint impact of environmental uncertainty and tolerance of ambiguity on top managers' perceptions of usefulness of non-conventional management accounting information. Scandinavian Journal of Management 14, 259–271.

Landon, J., Singpurwalla, N.D., 2008. Choosing a coverage probability for prediction intervals. The American Statistician, 62 120–124.

Lawrence, M.J., Makridakis, S., 1989. Factors affecting judgmental forecasts and confidence intervals. Organizational Behavior and Human Decision Processes 42, 172–187.

Moskowitz, H., 1972. The value of information in aggregate production planning – A behavioral experiment. AIIE Transactions 4, 290–297.

Moskowitz, H., Miller, J.G., 1975. Information and decision systems for production planning. Management Science 22, 359–371.

Mowen, M.M., Mowen, J.C., 1986. An empirical examination of the biasing effects of framing on business decisions. Decision Sciences 17, 596–602.

Önkal, D., Bolger, F., 2004. Provider-user differences in perceived usefulness of forecasting formats. Omega International Journal of Management Science 32, 31–39.

Önkal, D., Yates, J.F., Simga-Mugan, C., Oztin, S., 2003. Professional vs amateur judgment accuracy: The case of foreign exchange rates. Organizational Behavior and Human Decision Processes 91, 169–185.

Remus, W., 1986. Graduate-students as surrogates for managers in experiments on business decision-making. Journal of Business Research 14, 19–25.

Ruchala, L.V., 1999. The influence of budget goal attainment on risk attitudes and escalation. Behavioral Research in Accounting 11, 161–191.

Rush, H., Page, W., 1979. Long-term metals forecasting: The track record 1910–1964. Futures 1, 321–337.

Slone, R.E., Mentzer, J.T., Dittmann, J.P., 2007. Are you the weakest link in your company's supply chain? Harvard Business Review 85, 116–127.

Stellwagen, E.A., Goodrich, R.L., 1994. Forecast Pro for Windows. Business Forecast Systems Inc., Belmont, MA.
Wallis, K.F., 2003. Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. International Journal of Forecasting 19, 165–175.
Winston, W.L., 1987. Operations Research: Applications and Algorithms. Duxbury Press, Boston.
Yaniv, I., Foster, D.P., 1995. Graininess of judgment under uncertainty: an accuracy-informativeness trade-off. Journal of Experimental Psychology: General 124, 424–432.
Yaniv, I., Foster, D.P., 1997. Precision and accuracy of judgmental estimation. Journal of Behavioral Decision Making 10, 21–32.