

# Provider–user differences in perceived usefulness of forecasting formats

Dilek Önkal\*, Fergus Bolger

*Faculty of Business Administration, Bilkent University, 06800, Ankara, Turkey*

Received 11 September 2002; accepted 12 September 2003

## Abstract

This paper aims to examine potential differences in perceived usefulness of various forecasting formats from the perspectives of providers and users of predictions. Experimental procedure consists of asking participants to assume the role of forecast providers and to construct forecasts using different formats, followed by requesting usefulness ratings for these formats (Phase 1). Usefulness of the formats are rated again in hindsight after receiving individualized performance feedback (Phase 2). In the ensuing role switch exercise, given new series and external predictions, participants are required to assign usefulness ratings as forecast users (Phase 3). In the last phase, participants are given performance feedback and asked to rate the usefulness in hindsight as users of predictions (Phase 4). Results reveal that regardless of the forecasting role, 95% prediction intervals are considered to be the most useful format, followed by directional predictions, 50% interval forecasts, and lastly, point forecasts. Finally, for all formats and for both roles, usefulness in hindsight is found to be lower than usefulness prior to performance feedback presentation.

© 2003 Elsevier Ltd. All rights reserved.

*Keywords:* Judgment; Forecasting; Forecast format; Forecast provider; Forecast user

## 1. Introduction

The value of a forecast is a direct function of its provider, its user, and the interaction of these two sides in the prediction process. Disparities between the perceptions of users and preparers of forecasts have been only briefly addressed by previous research, with an emphasis on the lack of communication between the two parties [1–3]. Since discrepancies in expectations and interpretation of data may lead to undesirable forecasting consequences, it is important to understand and ameliorate any potential communication gaps. As asserted by Moon et al., “islands of analyses are detrimental to corporate performance” [4, p. 48], as signaled via communication problems between forecasters in different departments as well as communication gaps between forecast providers and users—if users are not engaged in the

forecasting process, they may discount the value of forecasts given by the providers, or may spuriously adjust the predictions, potentially deteriorating predictive performance.

A critical component of forecast communication is the format with which to convey the predictions [5]. Forecasting format represents an explicit choice to express the provider’s degree of uncertainty to be revealed to the user. In particular, when predictions are communicated as point forecasts (e.g., “value of USD/DM exchange rate will be  $X$ ”), users are presented with a usually deceptive sense of certainty regarding the conveyed number. In contrast, interval forecasts (e.g., “there is an  $XX\%$  probability that the earnings will be between  $Y$  and  $Z$ ”) and probabilistic directional forecasts (e.g., “there is an  $XX\%$  probability that the interest rate will increase”) convey relatively explicit statements of uncertainties surrounding the prediction.

Providers and users may consider some formats more/less useful than others, which may in turn affect how the forecasts are meant to be received versus how they are actually received. For example, several studies on earnings forecasts

\* Corresponding author. Tel.: +90-312-290-1596; fax: +90-312-266-4958.

*E-mail address:* [onkal@bilkent.edu.tr](mailto:onkal@bilkent.edu.tr) (D. Önkal).

have reported that the majority of managers' predictions are expressed in interval form (as conveyed via ranges or the use of conditioning phrases like "at least", "no more than", "greater than" etc.), rather than the point forecast format [6,7]. Such use may be viewed as stemming from a preference to communicate varying degrees of uncertainty or precision. That is, it may be that, rather than committing themselves to a single number and conveying an inflated sense of confidence in forecasts (as may be the case with point predictions), managers may prefer to share their uncertainty via intervals. This remains an untested explanation however, since the studies mentioned above have not systematically offered different formats to producers of predictions, nor have they examined the users' perceptions of given forecasts.

An exception in examining the forecast consumers' perspective is provided by Yates et al. [8] study, which employed only probabilistic forecasts. Their findings indicate that the forecast users may consider judgment extremeness (i.e., how far away the probabilities are from the least-desired probability of 50%) as a critical factor, treating it "as an indicator of a positive quality, confidence in one's convictions" [8, p. 54]. Consequently, forecast consumers may have a different perspective than forecast providers for evaluating predictions. In a similar vein, Yaniv and Foster [9,10] have asserted that in constructing and evaluating prediction intervals, there may exist an accuracy-informativeness trade-off. That is, as the width of interval forecasts increases, accuracy (as expressed by the normalized error, i.e.,  $\{ | \text{realized value} - \text{interval midpoint} | / \text{interval width} \}$ ) may increase. However, this increase in accuracy carries a cost of reduction in informativeness (as measured via a monotone function of the interval width). The researchers argue that narrow intervals, on the other hand, may be considered as informative, albeit at a cost of eroding accuracy.

As outlined above, there are few studies in this domain and they have not involved comparative prediction formats (studying only probabilistic forecasts in Yates et al. [8], and only interval predictions in Yaniv and Foster [9,10]), nor have they examined the impact of multiple forecasting roles. In fact, the lack of systematic investigation of prediction format effects from the perspective of a provider-user axis presents a focal gap in research on forecast communication and use [11]. In addressing this gap, the current study aims to explore potential differences in perceived usefulness of various forecasting formats from the viewpoints of the providers and users of predictions. Incongruent preferences may highlight different channels of vulnerability in miscommunicating the predictions between the two parties. Accordingly, understanding comparative perceptions of predictive-format usefulness may provide gateways to research on enhancing the value of forecasts.

In investigating reactions to predictive format, the current study utilizes judgmental forecasts since various surveys investigating the use of forecasting methods have concluded

that judgmental methods dominate the practitioners' predictive processes [12–20], even when their accuracy may be improved by integrating them with statistical methods [21,22]. Employing judgmental forecasts, this paper addresses issues on the forecast providers' and forecast users' understanding of predictions expressed via different formats as well as their relationship to the provision of relevant performance feedback. Accordingly, the paper is organized as follows: the research questions of interest are delineated in the next section. Section 3 explains the method used to explore these issues in the current study. Section 4 presents the results, while Section 5 summarizes the conclusions and presents future research avenues.

## 2. Research questions

Gaps in communication between users and producers of forecasts present major obstacles to better decision-making through the improved use of forecasting techniques in organizations. Prediction format constitutes a fundamental component of provider-user communication and provides the focus of this study. In particular, four prediction formats are investigated: point forecasts, 95% interval predictions, 50% interval predictions, and probabilistic directional predictions. Forecast providers and users may have different preferences in using these formats, which in turn, may be affected by feedback regarding predictive accuracy. Hence, research questions of interest are:

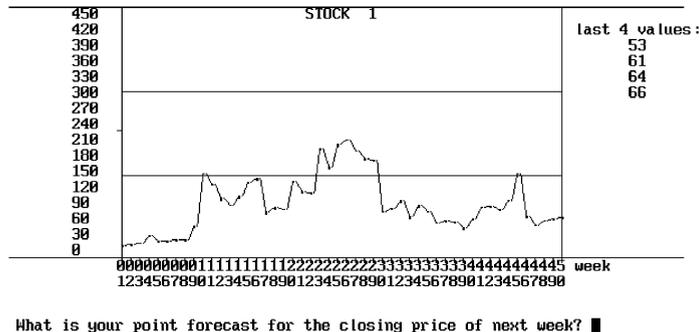
- (1) Do providers and users of forecasts differ in their perceived usefulness of various prediction formats?
- (2) If there are differences, do they realize this, and can they role switch?
- (3) Is the accuracy-informativeness trade-off a significant factor in users' evaluations of interval predictions?
- (4) How is perceived usefulness of forecasts affected by performance feedback (with performance defined as provider and user performance on given forecasting tasks)? In other words, is "usefulness in hindsight" (after the provision of performance feedback) different than "usefulness in foresight" (before the provision of performance feedback) and is it affected by provider versus user roles assumed?

## 3. Research design

### 3.1. Participants

A total of 102 third-year business students at Bilkent University, Turkey, completed the experiment towards extra credit in a forecasting course. Throughout the course, participants already had experience with various forecasting formats and accuracy measures.

(a) Time Series Example 1



(b) Time Series Example 2

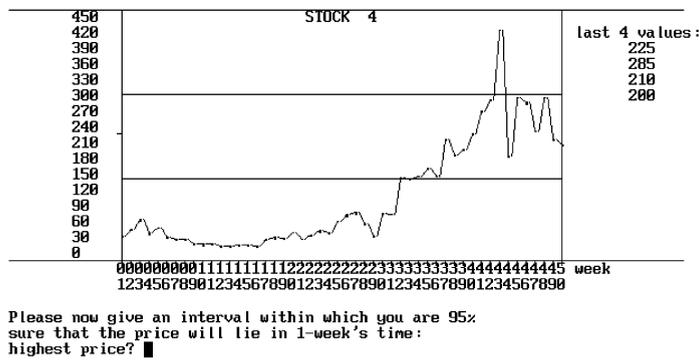


Fig. 1. Sample time series given to participants: (a) time series example 1; and (b) time series example 2.

3.2. Materials

A total of 32 50-week time-series graphs were used in the study. The last four values of the displayed series were also presented in tabular form next to each graph. Constructed series were used, and the participants were told that they showed the values of real Turkish stocks with undisclosed stock names and time periods.

The series varied in terms of the degree of first-order autocorrelation (4 levels: approximately 0.6, 0.3, 0, or -0.3), amount of noise (2 levels: low and high), and trend (2 levels: positive trend and no trend) (see Fig. 1 for two example graphs). The parameters were selected to reflect the behavior of actual Turkish stock price series at the time the study was conducted. For instance, given the high inflation rate, Turkish stocks tended not to display any long-term negative trends, and were more likely to show positive than negative autocorrelation. (Note that all the autocorrelation coefficients were computed for untrended series, then a positive linear trend was added where appropriate).

3.3. Procedure

The study was conducted in a single session in a computer lab. Subjects were given general instructions about

the study, detailing the various forecasting formats and the accuracy measures with illustrative examples. Each participant was then given a diskette that led the subjects through the various phases of the experiment as explained below, giving specific task instructions and saving their individual data. The experimental procedure consisted of four main phases, enabling the participants to experience first the role of a forecast provider (in Phases 1 and 2), followed by the role of a forecast user (in Phases 3 and 4).

In Phase 1, subjects were given 16 time series labeled as stock prices. For each time series, participants were asked to provide one-step-ahead forecasts by means of each of four prediction formats (i.e., point forecast, directional probability forecast, 50% prediction interval, and 95% prediction interval). They were then requested to rate each format in terms of *usefulness as a provider* of these forecasts. All ratings were to be made on a 7-point scale with 1 = not useful at all, 7 = extremely useful.

In Phase 2, participants received personalized feedback on their forecasting performance for each of the prediction formats. Subjects were then requested to rate each format again in terms of *usefulness in hindsight as a provider* of these predictions.

In Phase 3, participants were presented with 16 different time series plus external one-step-ahead forecasts in the same four formats. Each participant was asked to construct

a portfolio by selecting four stocks to invest (based on the time-series information and the externally provided forecasts), and then rate each format in terms of *usefulness as a user* of these predictions.

In Phase 4, subjects received outcome feedback for each stock, along with performance feedback for the external forecasts provided, as well as personalized feedback on the performance of their constructed portfolios (i.e., percentage return earned by the constructed portfolio). The participants were then requested to rate each format again, this time in terms of *usefulness in hindsight as a user* of these forecasts.

In sum, within-subjects factors used in the experimental design were: (1) role (provider versus user); (2) prediction format (point forecasts, directional forecasts, 50% prediction intervals, 95% prediction intervals), and (3) feedback (ratings before feedback versus after feedback). In addition, the size of external interval forecasts was manipulated as the between-subjects factor. Specifically, subjects were randomly divided into three groups that received different external interval forecasts, with all subjects receiving the same external point and directional forecasts. The external point forecasts were computed using trend and autocorrelation coefficients appropriate for each series. Directional forecasts were obtained by comparing the point predictions with the last realized value for each series, and judgmentally assigning confidence percentages of either 50% or 95% (resulting in an equal number of 50% and 95% directional predictions given to participants). For interval predictions, theoretical intervals ( $[\text{point forecast}] \pm [Z_{\alpha/2}][\text{standard deviation of observations}]$ ) were computed. One of the three groups received these computed intervals (Regular External Intervals Group), while a second group was given intervals that were reduced by 50% of their computed interval width (Narrow External Intervals Group), with a third group receiving intervals that were enlarged by 50% of their computed interval width (Wide External Intervals Group).

3.4. Performance measures

The following measures were utilized in assessing forecaster performance and providing feedback to participants in their forecast provider roles:

(1) Point forecasts were evaluated via the Mean Absolute Percentage Error (i.e.,  $MAPE = \{\sum |(\text{forecast error}/\text{realized value}) \times 100\} / \{\text{number of forecasts given}\}$ ).

(2) Directional forecasts were evaluated through the “percentage of directions correctly predicted” as well as the “average probability forecast”.

(3) 50% and 95% interval forecasts were assessed via the corresponding “hit rates” (i.e., percentages of intervals containing the realized values).

For the forecast user role, percentage return of constructed portfolios (i.e., average percentage return computed over the four stocks chosen by the participant, where the percentage return for stock  $i = \{(\text{realized value} - \text{last stock price}$

$\text{given to subjects}) / (\text{last stock price given to subjects})\}$ ) were computed and utilized as performance feedback for each of the participants. All the performance measures had been explained to the participants in the beginning of the experiment.

4. Results

4.1. Performance in forecast provider and user roles

Performance scores of participants in forecast provider and forecast user roles are summarized in Table 1. As providers of predictions, participants’ point forecasts revealed an average MAPE of 7.1%; while the average probability assessment used for directional predictions was 70%, with 61% of directions correctly predicted, indicating some overconfidence. Around 73% of stated 95% prediction intervals tended to include the realized value, indicating the participants’ general overconfidence in these interval assessments. No such overconfidence was observed for the 50% intervals, however, as revealed by their average hit rate of 51%.

In the user role, participants’ portfolios showed the highest average percentage return (6.3%) for the group given the narrow external intervals, followed by the group given regular external intervals (i.e., mean percentage return of 5.7%). Subjects in the group receiving wide external intervals earned the lowest average percentage return (4.9%) on their portfolios. However, these differences were overshadowed by the wide ranges in percentage returns attained by the three groups, leading to no statistically significant differences based on informativeness/width of the external prediction intervals ( $F_{2,99} = 0.85; p > 0.10$ ).

Table 1  
Performance scores

	Mean (%)	St.Dev. (%)	Range (%)
(A) Provider Role: MAPE, percentage directions correctly predicted, mean probability, hit rate			
MAPE	7.1	2.6	[3.7–22.7]
% directions correct	61	17	[25–100]
Mean probability	70	8	[50–95]
Hit rate (95% PI)	73	20	[13–100]
Hit rate (50% PI)	51	23	[6–100]
(B) User Role: percentage return of constructed portfolios			
Regular external intervals group	5.7	3.9	[–2.6–11.5]
Narrow external intervals group	6.3	4.5	[–4.7–12.1]
Wide external intervals group	4.9	4.3	[–3.0–10.3]

Table 2  
Mean (standard deviation) usefulness ratings given in the provider and user roles

Forecast format	Forecast provider		Forecast user	
	Before feedback	After feedback	Before feedback	After feedback
95% Interval	5.55 (1.13)	5.49 (1.30)	5.52 (1.31)	5.33 (1.29)
Directional	4.98 (1.31)	4.69 (1.23)	5.21 (1.16)	5.06 (1.20)
50% Interval	4.78 (1.19)	4.71 (1.12)	4.93 (1.15)	4.82 (1.11)
Point	4.68 (1.25)	4.45 (1.38)	4.74 (1.25)	4.68 (1.34)

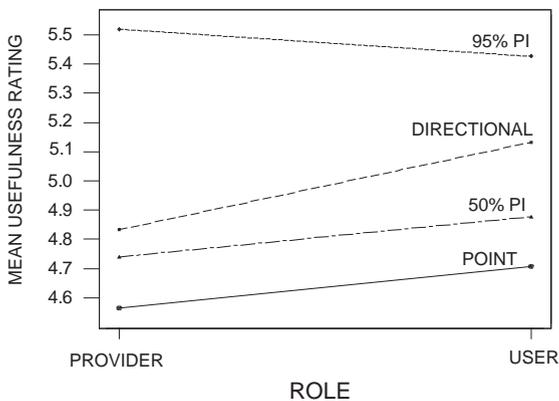
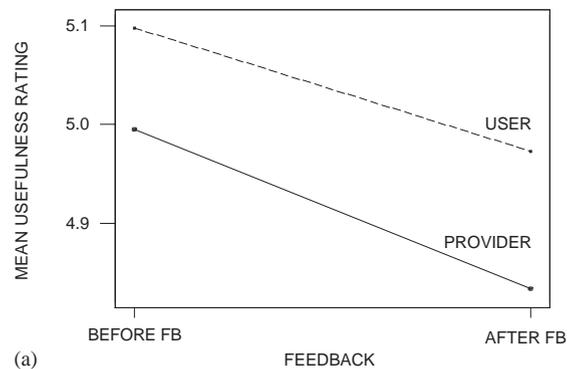


Fig. 2. Usefulness ratings given in provider and user roles.

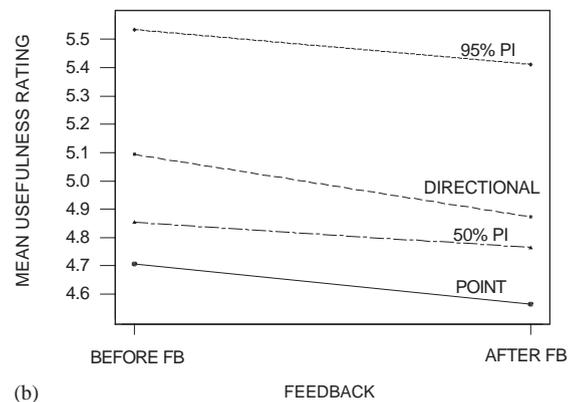
#### 4.2. Usefulness ratings

Table 2 presents a summary of the usefulness ratings given by participants for various forecasting formats. It can be clearly observed that the participants consistently rated “95% prediction intervals” as the most useful format, followed by the “directional probability forecasting” format. “50% prediction intervals” represented the third most useful format, with “point forecasts” representing the least useful format. As can be gleaned from Fig. 2, this ordering did not change when the presenter assumed the role of a forecast provider versus the role of a forecast user, or before or after feedback.

Except for the 95% prediction interval format where provider and user ratings were quite similar, participants appeared to assign higher ratings overall when assuming the user role than when assuming the provider role. It is also worth noting that the perceived discrepancy in usefulness between 95% intervals over other formats was emphasized more in the provider than in the user role. In short, the results demonstrate a significant interaction between prediction format and role ( $F_{3,297} = 4.06$ ;  $p = 0.008$ ),



(a)



(b)

Fig. 3. Usefulness ratings given before and after feedback: (a) role  $\times$  feedback interaction; and (b) format  $\times$  feedback interaction.

accompanied by the significant main effect of forecasting format ( $F_{3,297} = 23.66$ ;  $p < 0.001$ ).

#### 4.3. The effects of feedback on perceived usefulness

Usefulness ratings before and after feedback are profiled graphically in Fig. 3. When assessing usefulness

after feedback, participants gave lower ratings than prior to feedback regardless of their assumed roles as provider versus user of forecasts (as shown in Fig. 3a), and irrespective of predictive formats utilized (as displayed in Fig. 3b). That is, “usefulness in hindsight” was lower than perceived usefulness prior to performance feedback ( $F_{1,297} = 7.37$ ;  $p = 0.008$ ).

Given the significant lowering effect of feedback on usefulness ratings, we next analyzed the factors that could potentially affect perceived usefulness in hindsight in both the forecast provider and user roles using regression analysis. In the provider role, providing hit rate information was found to significantly affect (i.e., lower) usefulness ratings for both the 95% prediction interval format ( $t_{99} = 3.52$ ,  $p = 0.001$ ) and the 50% prediction interval format ( $t_{99} = 2.88$ ,  $p = 0.005$ ). Similarly, participants seemed to lower their ratings of usefulness for directional forecasting format upon being presented with their attained percentage of correctly predicted directions ( $t_{98} = 2.64$ ,  $p = 0.010$ ), but with no corresponding response to feedback on their average probability assessments ( $t_{98} = 0.27$ ,  $p > 0.10$ ). Interestingly, usefulness ratings given to point predictions appeared to be unresponsive to feedback regarding the participants’ point forecasting performance (i.e., MAPE did not seem to affect the ratings;  $t_{99} = 0.86$ ,  $p > 0.10$ ).

When participants assume user roles, percentage returns earned by their constructed portfolios becomes the relevant feedback item to be investigated (i.e., since subjects are not constructing forecasts under the user frame, measures of forecasting performance are no longer pertinent). Results reveal that information on portfolio performance (percentage return) effectively alters the usefulness ratings for directional forecasts ( $t_{97} = 2.27$ ,  $p = 0.025$ ) and 50% prediction intervals ( $t_{97} = 2.69$ ,  $p = 0.008$ ), while the ratings for point forecasts and 95% prediction intervals are comparatively unaffected ( $t_{97} = 1.73$ ,  $p = 0.086$ ; and  $t_{97} = 1.84$ ,  $p = 0.069$ , respectively). Furthermore, there seem to be no significant differences in usefulness ratings for the three groups receiving external interval forecasts of differing widths ( $F_{2,99} = 0.03$ ;  $p > 0.10$ ), as depicted in Table 3.

## 5. Discussion and conclusions

Communicating the relative perspectives of forecast providers and users remains a critical issue that has received limited research attention. Focusing on forecasting format as a first step, the current work yields promising findings on the perceived usefulness of predictions. Regardless of forecasting role, 95% prediction intervals are considered to be the most useful format (much more than the other formats, especially in the user mode), while point forecasts represent the least useful format. This result may signal that, to many users and producers of forecasts, a point prediction may seem quite incomplete, since it is known that it is very difficult for a specific value to occur. In contrast, a range (as given by interval predictions) may be more realistic and acceptable in that it serves a double purpose by simultaneously communicating uncertainty and providing a set of possible values with a reasonable chance of occurrence.

Our finding that both directional and interval predictions are perceived as more useful than point predictions further supports Fischhoff’s assertion that the users need an indication of the confidence in predictions, since greater confidence in forecasts “allows one to take more decisive action, to curtail information collection, to plan for a narrower range of possible contingencies, and to invest less in vigilance for surprises” [23, p. 391]. Overall, the results may be viewed as suggesting that the communication of uncertainty is considered critical from both the provider’s and the user’s perspective. That is, for both roles, communicating a single number as a forecast may be viewed as conveying partial information while inducing a false sense of completeness. Thus, both the forecasters and users appear to be requiring the declared prediction to be supplemented by a measure of uncertainty or confidence in the stated forecast.

Results also reveal that the 50% prediction intervals are perceived as being less useful than 95% intervals. This is quite interesting given that, in the provider role, forecasters were found to be overconfident with 95% intervals (i.e., average hit rate was 73%), while showing no such overconfidence for their 50% interval judgments (i.e., their intervals

Table 3  
Mean (standard deviation) usefulness ratings given for interval predictions by users from different external interval groups

	95% Prediction interval		50% Prediction interval	
	Before feedback	After feedback	Before feedback	After feedback
Regular external intervals group	5.59 (1.28)	5.35 (1.15)	4.85 (1.13)	4.88 (1.27)
Narrow external intervals group	5.50 (1.40)	5.35 (1.56)	4.85 (1.16)	4.79 (1.10)
Wide external intervals group	5.47 (1.29)	5.29 (1.17)	5.09 (1.16)	4.79 (0.98)

included the realized value on 51% of the occasions, on average). Stated preference of 95% prediction intervals over the 50% intervals could be due to clarity of communication concerns, i.e., concerns about the 50% intervals being interpreted as there existing a 50% chance of the realized value falling *outside* the given interval (yielding a 50% chance of being “incorrect”). In a similar vein, Yates et al. [8] had found a “disdain” for 50% probabilities by the consumers of forecasts, who “took such judgments as indications that the forecasters were either generally incompetent, ignorant of the facts in a given case, or lazy, unwilling to expend the effort required to gather information that would justify greater confidence” (p. 45).

Findings may also be related to expectations of relative informativeness and accuracy. In both roles, participants may have regarded 95% intervals as more meaningful and informative, since 50% intervals may have conveyed a meaning of “just as likely for the realized value to fall into the interval as to fall outside the interval”, potentially invoking non-informative connotations. However, 50% intervals should be narrower and thus more informative than 95% intervals, according to Yaniv and Foster [9,10]. A potential explanation of the discrepancy between our findings and the inferences from Yaniv and Foster’s work may lie in the probability values chosen. That is, it may be that, if we had specified 75% intervals instead of 50% (which may be the least preferred probability value due to an unclear knowledge implication), our participants could have found 75% intervals as more informative and useful than 95% intervals. Similarly, the finding that probabilistic directional forecasts were considered more useful than 50% prediction intervals may be explained by the former format perceived as being less ambiguous in information content.

In the user role, whether the intervals provided were inflated, deflated or unmodified did not seem to affect their perceived usefulness, again seemingly contradicting Yaniv and Foster’s inferences. The provisional conclusion that the informativeness of intervals appears to have no effect on usefulness ratings for especially the 50% and 95% interval forecasts may suggest a myopic view of evaluating the use of prediction intervals. In fact, a common explanation offered by the participants in post-experimental discussions focused on the design of the study. Since the experimental setup involved presenting external forecasts in all four formats, the participants emphasized that they did not solely rely on interval forecasts, but rather complemented the information received via interval predictions with information gleaned from other formats (directional and/or point predictions), effectively covering the informational deficiencies of intervals that appeared too narrow or extremely wide. It may also be that the adjustments to external interval widths were not sufficient to induce significant changes in informativeness expectations, hence not affecting usefulness considerations. In any case, future work is needed to systematically isolate the presented formats so that the missing information cannot be compensated using predictions from other formats.

For all formats, usefulness in hindsight was lower than usefulness prior to performance feedback presentation in both the provider and user roles. When confronted with their realized performance (in constructing forecasts of stock prices in the provider role; in using the provided forecasts to construct portfolios in the user role), participants modified the higher usefulness ratings they gave prior to feedback. Looking back given the actual performance information, participants may in hindsight have felt that the formats were not as useful as initially thought, thus dampening their ratings.

The only exception was provided by point forecasts (i.e., least useful format). This result may be a reflection of both the participants’ and the users’ comparatively lower expectations from point forecasts, leading to no significant adjustments on usefulness ratings when confronted with their realized performance. That is, the perceived usefulness of point predictions may simply have been so low that the performance feedback accenting the participants’ overall poor scores could not significantly reduce these ratings any further.

The findings further indicate that the ratings assigned to the formats considered to be the least useful (point prediction) and the most useful (95% prediction intervals) do not seem to be influenced by performance feedback as much as the formats given intermediate usefulness ratings (directional forecasts and 50% prediction intervals). Usefulness perceptions for formats in between appear to be strongly affected by feedback on participants’ actual performance in using the external forecasts to construct investment portfolios, potentially reflecting the surprise factor associated with their realized versus expected performance in those formats.

It is worth noting that the participants in this study were students. Repeating similar work with professional forecasters and forecast users is needed to enhance the generalizability of results. Likewise, we only studied perceived usefulness of forecast formats within the (stated) context of stock-market forecasting. It is well-known in the judgment and decision-making literature that the “frame” or context of a judgment or decision can have a substantial influence on it [24]. The same seems to be the case for judgmental forecasting where a number of studies have shown that immediate context, such as the scale of presentation of a time-series graph, or the labeling of its axes, affect forecasting performance (see, e.g. [25] for a review). We cannot therefore be certain that the results would be the same if transferred to another (stated) context, such as sales or cash-flow forecasting.

The current design also requested the same participants to assume both the provider and the user roles. Such role switches are realistic in many organizational settings, where individuals are expected to construct forecasts for certain variables, while using the provided predictions for other decisions. However, an extension of the current research could involve studying usefulness perceptions and needs of professionals who are *only* responsible for making forecasts versus those *solely* accountable for using given

predictions. Relatedly, future work where users are given different “user tasks” (e.g., choosing a single stock to invest versus constructing a portfolio with a small/large number of stocks) is needed to explore the users’ responses to provided forecasts.

A further limitation of our study is related to the design feature of using the same participants in both provider and user roles discussed in the previous paragraph. Since all participants first acted as forecast providers than later acted as forecast users, there may have been an order effect. Although the results for assessed usefulness were essentially the same in both roles, and it seems unlikely that order was a factor, it would be useful to verify this through future research. Similarly, the forecasts in the four formats were also always produced in the same order, which again raises the question of possible order effects. In particular, since the point forecast was made first, this may have acted as an anchor for subsequent forecasts, thereby biasing performance—these issues should also be addressed in future research.

Another methodological consideration is whether there is a confound between the roles assumed by the participants (user or provider) and the perceived source of the forecasts. For instance, it is possible that the externally provided forecasts in the user role may have been given greater credence than the self-generated forecasts of the provider role, thus influencing usefulness ratings, probably in an upward direction. Role itself may therefore have had little or no influence on the ratings, with the (non-significant) increase in usefulness ratings between the provider and user roles being due either partly, or entirely, to beliefs about the source of forecasts, and their consequent credibility. Again, this possibility does not detract from the main finding that there is a consistent hierarchy of formats in terms of their usefulness over both roles. Further, we would argue that a major difference between the roles of forecast provider and forecast user is exactly that of the source of forecasts—internal versus external—so it is difficult in practice to separate these two things. However, manipulating the credibility of the source independently of role would be possible, and it might be enlightening to do this in future studies.

This research stream is based on the premise that preparers and users of predictions may have complementary information and skills—it is through a coordinated effort to share such differences that improved forecasting performance, leading to better decisions, will result. In addition to potential information asymmetry, users and preparers may also have differing perceptions and expectations of forecasts depending on their organizational roles. For example, users may “want forecasts which will enable them to succeed in an environment which is increasingly complex, interdependent, and uncertain” [26, p. 242], while providers may need to trust that their predictions will not be “unnecessarily” altered. Future work to enhance our understanding of these potential differences may indeed prove quite critical for organizational interdependencies. The current research provides an exploratory step in this direction.

## References

- [1] Mahmoud E, DeRoeck R, Brown R, Rice G. Bridging the gap between theory and practice in forecasting. *International Journal of Forecasting* 1992;8:251–67.
- [2] Sanders NR. Managing the forecasting function. *Industrial Management Data Systems* 1995;95:12–8.
- [3] Wheelwright SC, Clarke DG. Corporate forecasting: promise and reality. *Harvard Business Review* 1976;54:40–7.
- [4] Moon MA, Mentzer JT, Smith CD, Garver MS. Seven keys to better forecasting. *Business Horizons* 1998;September–October:44–52.
- [5] Önkal-Atay D, Thomson ME, Pollock AC. Judgemental forecasting. In: Clements MP, Hendry DF, editors. *A companion to economic forecasting*. Oxford: Blackwell Publishers; 2002. p. 133–51.
- [6] Baginski SP, Conrad EJ, Hassell JM. The effects of management forecast precision on equity pricing and on the assessment of earnings uncertainty. *The Accounting Review* 1993;68:913–27.
- [7] Pownall G, Wasley C, Waymire G. The stock price effects of alternative types of management earnings forecasts. *The Accounting Review* 1993;68:896–912.
- [8] Yates JF, Price PC, Lee J-W, Ramirez J. Good probabilistic forecasters: the ‘consumer’s’ perspective. *International Journal of Forecasting* 1996;12:41–56.
- [9] Yaniv I, Foster DP. Graininess of judgment under uncertainty: an accuracy-informativeness trade-off. *Journal of Experimental Psychology: General* 1995;124:424–32.
- [10] Yaniv I, Foster DP. Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making* 1997;10:21–32.
- [11] Önkal-Atay D. Financial forecasting with judgment. In: Wright G, Goodwin P, editors. *Forecasting with judgment*. Chichester: Wiley; 1998. p. 139–67.
- [12] Dalrymple DJ. Sales forecasting methods and accuracy. *Business Horizons* 1975;18:69–73.
- [13] Dalrymple DJ. Sales forecasting practices—results from a United States survey. *International Journal of Forecasting* 1987;3:379–91.
- [14] Mentzer J, Cox J. Familiarity, application and performance of sales forecasting techniques. *Journal of Forecasting* 1984;3: 27–36.
- [15] Peterson RT. The role of experts’ judgment in sales forecasting. *The Journal of Business Forecasting* 1990;9: 16–21.
- [16] Rothe JT. Effectiveness of sales forecasting methods. *Industrial Marketing Management* 1978;April:114–8.
- [17] Sanders NR. The status of forecasting in manufacturing firms. *Production and Inventory Management Journal* 1997;38:32–5.
- [18] Sanders NR, Manrodt KB. Forecasting practices in US corporations: survey results. *Interfaces* 1994;24:92–100.
- [19] Sparkes JR, McHugh AK. Awareness and use of forecasting techniques in British industry. *Journal of Forecasting* 1984;3:37–42.
- [20] Winklhofer H, Diamantopoulos A, Witt SF. Forecasting practice: a review of the empirical literature and an agenda for future research. *International Journal of Forecasting* 1996;12:193–221.
- [21] Goodwin P. Statistical correction of judgmental point forecasts and decisions. *OMEGA: International Journal of Management Science* 1996;24:551–9.

- [22] Goodwin P. Integrating management judgment and statistical methods to improve short-term forecasts. *OMEGA: International Journal of Management Science* 2002;30: 127–35.
- [23] Fischhoff B. What forecast (seem to) mean. *International Journal of Forecasting* 1994;10:387–403.
- [24] Tversky A, Kahneman D. Rational choice and the framing of decisions. *Journal of Business* 1986;59:5251–78.
- [25] O'Connor M, Lawrence M. Judgmental forecasting and the use of available information. In: Wright G, Goodwin P, editors. *Forecasting with judgment*. Chichester: Wiley; 1998. p. 65–90.
- [26] Jones VD, Bretschneider S, Gorr WL. Organizational pressures on forecast evaluation: managerial, political, and procedural influences. *Journal of forecasting* 1997;16:241–54.