International Journal of Forecasting 7 (1991) 141–149
North-Holland

The effect of group interaction processes on performance in time series extrapolation

Soon Ang
Carlson School of Management, University of Minnesota, Minneapolis, MN 55455, USA

Marcus O'Connor
School of Information Systems, University of New South Wales, Kensington, NSW 2033, Australia

Abstract: This study explores the ability of groups to forecast and establish judgmental confidence intervals in time series extrapolation. Thirty-six three-person groups were used to evaluate four different group interaction processes. In addition to staticized, nominal group technique and consensus processes, the study utilizes a modified consensus process, where a selected group member completes the task prior to group discussion and interaction. Using real life time series, subjects produced forecasts and related confidence intervals for six periods.

Groups in the modified-consensus structuring process exhibited significantly greater forecast accuracy than all other experimental conditions ($p < 0.001$). The superiority was most pronounced for series of high forecast difficulty. These results are discussed in relation to the contribution of the initial estimates as an anchor on which the modified-consensus group can focus.

Keywords: Group judgment, Forecast accuracy.

1. Introduction

Past surveys of actual forecasting practice in business indicate that judgmental forecasting is the most common forecasting approach in organizations (Dalrymple, 1987, Conference Board, 1978), and is usually done in interacting groups (Mentzer and Cox, 1984; PoKempner and Bailey, 1970). A common scenario will have corporate executives sitting around a table, deciding as a group what their best estimate is for the item to be forecast (Edmundson et al., 1988).

The purpose of this research is to assess both the accuracy and confidence of groups in time series extrapolation of varying levels of forecast difficulty under four different group interaction processes. This study focuses on two dimensions of performance outcomes: (a) forecast accuracy, and (b) calibration of confidence intervals in time series.

Empirical evidence comparing *individuals vs. staticized* groups has shown that mathematically combining individual forecasts to form a staticized group forecast generally improves accuracy over the separate individual forecasts (Einhorn, Hogarth and Klempner, 1977). [1] In time series extrapolation, Lawrence et al. (1986) and Edmundson et al. (1988) found that averaging separate

[1] It should be emphasised that if a judgment differs from a true value by a random amount, staticized groups are necessarily better than individuals acting alone.

judgmental forecasts always reflected an improvement in accuracy over its individual forecasts. In comparing *individuals vs. interacting* groups, prior research has revealed that interacting groups generally led to poorer performance. This occurred in a variety of conditions (Einhorn et al., 1977).

With *staticized and interacting* groups, prior research has revealed a confusing picture. Hackman and Morris (1975) concluded that, "for many tasks", staticized groups outperform interacting groups. However, in probability forecasting, Fischer (1981) found "little or no difference" in performance. Yet others, such as Rohrbaugh (1979), asserted that interacting groups were better than staticized groups. Inconsistent results were also reported in the accounting literature. Trotman, Yetton and Zimmer (1983) found staticized groups outperformed interacting groups; Trotman and Yetton (1985), Libby et al. (1987) and Chalos and Pickard (1985) found no significant differences, while Trotman (1985) found interacting groups were better than staticized groups. Many reasons have been advanced to explain the poor performance of interacting groups. These include the "groupthink" phenomena (Janis, 1972), polarization, pluralistic ignorance and premature closure (Shaw, 1981).

Despite weighty arguments against the usefulness of interacting groups in forecasting, Armstrong (1985) appeals to researchers to devise and examine ways of improving group meetings and making them more effective for forecasting. This is because forecasting by groups is still a pervasive phenomenon, and many organizations continue to forecast in groups. Traditional group meetings are popular for many reasons. People generally need not expend much energy preparing meetings, so meetings are easy (Armstrong, 1985). Meetings also satisfy social needs, for example participants reported higher satisfaction with groups (Van de Ven and Delbecq, 1974; Boje and Murnighan, 1982). Finally, forecasting in interacting groups improves the commitment of the individual members to the group forecast – a result from their public involvement in the processes leading to the group forecast. A way of improving group meetings is to adopt an appropriate group interaction process to enhance the effectiveness of groups. This paper reports on an empirical study assessing the effectiveness of a new group interaction process, the modified consensus process.

1.1. Group interaction processes

Group interaction processes vary in their degree of control over group interaction and their manner of aggregating individual opinions to form a group judgment. There are two basic kinds of aggregating individual opinions – the mathematical approach and the behavioral approach (Ferrell, 1985). In the mathematical approach, individuals' judgments are averaged using some statistical composition models. In behavioral aggregation, individuals discuss the matter and come to a consensus on the estimate. Mixtures of the two are possible. For example, the Delphi or the nominal group technique (NGT) aim to improve behavioral aggregation by substituting the dysfunctional aspects of achieving consensus with a mathematical process of achieving the final group judgment.

Some studies comparing group interaction processes found NGT to enhance performance for a series of judgmental estimations of probability (Gustafson et al., 1973; Gough, 1975; Seaver, 1979), although for some of the studies the effect was not significant. However, in a judgmental forecasting task, Sniezek (1988) found little differential impact on the quality of group judgmental forecasts across staticized, Delphi, and consensus interaction processes. Both groups and individuals appear underconfident about their forecast in the easy tasks and overconfident in difficult tasks. This finding on judgmental confidence is consistent with research in both abstract tasks (Lichtenstein et al., 1982) and time series forecasting (O'Connor and Lawrence, 1989).

This study compares the performance of groups under four group interaction processes:

(a) a *staticized* group which averages individual forecasts in the group;

(b) a group based on the *nominal group technique* (NGT) (Van de Ven and Delbecq, 1971), where subjects make individual forecasts both prior to and after group interaction;

(c) a group that reaches a *consensus* after interaction; each group is presented with a time series forecasting task and is then requested to forecast judgmentally as a group;

(d) a group that reaches a *modified-consensus* after interaction; the group, as in the consensus process, reaches a consensus after interaction, but in this case, one member of the group has made some private estimate prior to group interaction.

Diagrams of these group structures are depicted in Exhibit 1.

The staticized, NGT, and the consensus group processes were included in the study because they are well documented in the literature. The modified-consensus group process, a variation of both the NGT and the consensus process is included after observing the dynamics of group sales forecasting in a large retailing organization (see Edmundson et al., 1988). In that investigation, it was hypothesized that the superior accuracy of consensus forecasts over statistical and individual judgmental extrapolation methods could have been partly due to the fact that one of the group members (the sales manager) had made estimates prior to the meeting. However, other participants also brought non-time series information to the meeting, which could have contributed to the superior accuracy of consensus forecasts over statistical or individual judgmental extrapolative methods. By eliminating this non time series information, this study seeks to investigate whether this particular group interaction process, modified-consensus, had any effect on accuracy.

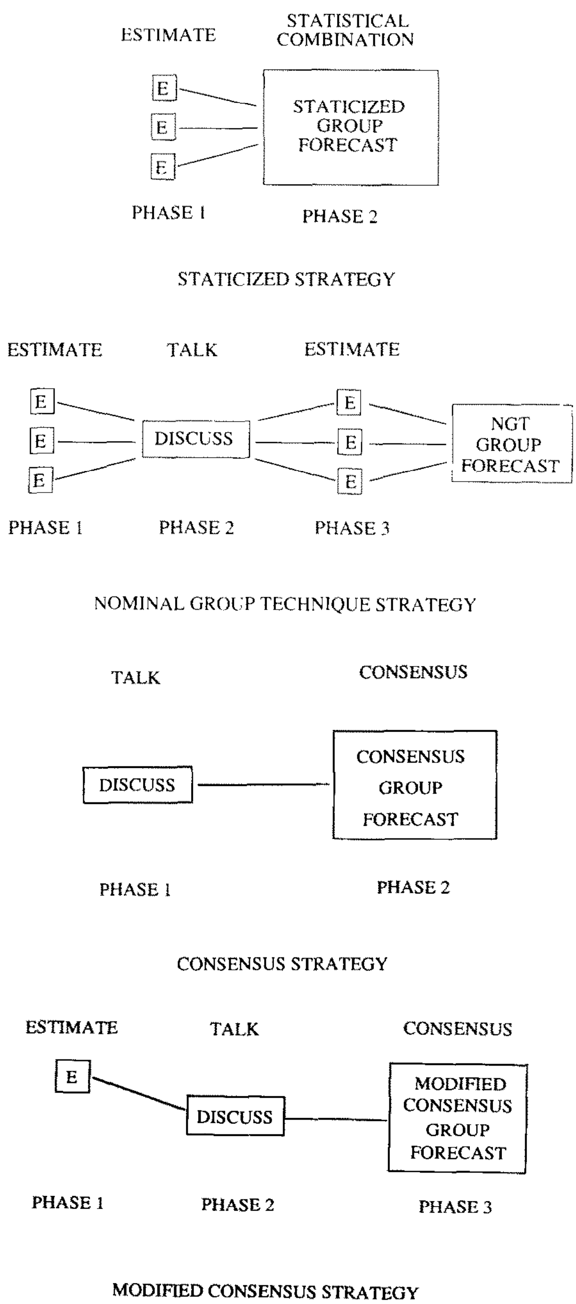

Exhibit 1. Diagrammatic depiction of group interaction processes.

1.2. Task difficulty

The influence of variations in task characteristics has generally been recognized as a major contributor to variation in performance (see Payne, 1982). Hackman (1968), for example, found that task differences accounted for about 50% of the performance variation across tasks ranging from discussion of values to problem solving. Morris (1966) and Golembiewski (1962) found similar task effects. Amongst task characteristics, Shaw (1981) concluded that task difficulty was the strongest task dimension. Accordingly, this study seeks to investigate the differences in performance of the different group interaction processes across forecasts of varying difficulty.

In this study, forecasts are classified into high and low forecast difficulty using the methodology developed in Lawrence et al. (1986). This methodology classified time series used in the *M*-competition (Makridakis et al., 1982) into varying levels of forecast difficulty based on forecast errors derived from previous experiments. Two series from the *M*-competition database were chosen to represent low and high difficulty forecasts. [2]

[2] These series were selected because they had:
(a) at least 3.5 years of monthly data each,
(b) roughly the same level of seasonality, and
(c) historical and validation data which did not involve any awkward turning points.

2. Research hypotheses

In relation to the literature reviewed above, we wish to examine the effect of group interaction processes on the ability of groups to assess both the accuracy and confidence of groups in time series extrapolation. The (null) hypotheses tested in this study were:

Hypothesis 1 (H1). Different group interaction processes do not lead to different levels of forecast accuracy in a judgmental time series extrapolation task.

Hypothesis 2 (H2). Different group interaction processes do not lead to different levels of calibration of the confidence intervals in a judgmental time series extrapolation task.

3. Experimental design

Four group interaction processes and two levels of forecast difficulty were examined in a 4×2 factorial design. Combinations of the two factors defined eight major, equal frequency cells. There were six judgmental predictions in each cell, making a total of 48 group predictions (with three individuals in each group).

3.1. Dependent measures

Group performance was assessed in terms of forecast accuracy and calibration of confidence intervals (CIs). Forecast accuracy was measured using Mean Absolute Percentage Error (MAPE). MAPE was chosen because it is both a common measure to assess relative accuracy (Carbone and Armstrong, 1982) and is more robust against extreme errors than squared measures (Lawrence et al., 1986). MAPE is also "dimensionless" (Armstrong, 1985) and enables one to compare time series of different measurement units. The metric chosen to measure the accuracy of calibration established by groups is the relative frequency with which the actual value falls inside the CIs. This is referred to as *calibration*.

4. Method

4.1. Subjects

Subjects were 108 part-time master's students at the University of New South Wales, most of whom were employed full-time in business related work, and were considered representative of personnel engaged in business forecasting. Subjects ages ranged from 23 to 40 years and there was approximately the same representation from both sexes. Using students in a laboratory experiment with restricted information on time series has generally been criticized to reduce the external validity of the study (Winkler and Murphy, 1973). However, given the research problem – i.e. the contribution of interaction and the extent to which different group interaction processes can improve groups in time series extrapolation – the use of a tightly controlled experiment is deemed most appropriate in investigating this issue.

Subjects participated on a voluntary basis, and received no course credit for their involvement. However, as an incentive, the best performing individual and group received A$25 and A$75, respectively. Students seemed to respond positively to this "tax-free" incentive.

4.2. Procedure

Each subject was given a booklet containing instructions for the task, and the two time series (illustrated in Exhibit 2). The time series were presented as line graphs. The nature of the series and the period of history the series referred to were masked. Subjects thus could only rely on time series information for their judgmental extrapolation. Each subject was required to forecast the two time series. The series were chosen from the *M*-competition database (Makridakis et al., 1982) and were classified as "easy" and "difficult" to forecast based on an analysis of judgmental forecasts in a previous study (Lawrence et al., 1985).

Subjects were told that the object of the experiment was to assess the effectiveness of groups in time series forecasting. The experimenter then demonstrated a time series extrapolation, showing them techniques in spotting trends and seasonality in a series. The demonstration lasted about 15 minutes.

Subjects were then required to forecast and estimate 50% confidence intervals for the next six periods. Each subject completed the first series individually. Data collected from these individual forecasts provided the basis for the staticized group forecasts. They were then randomly allocated to

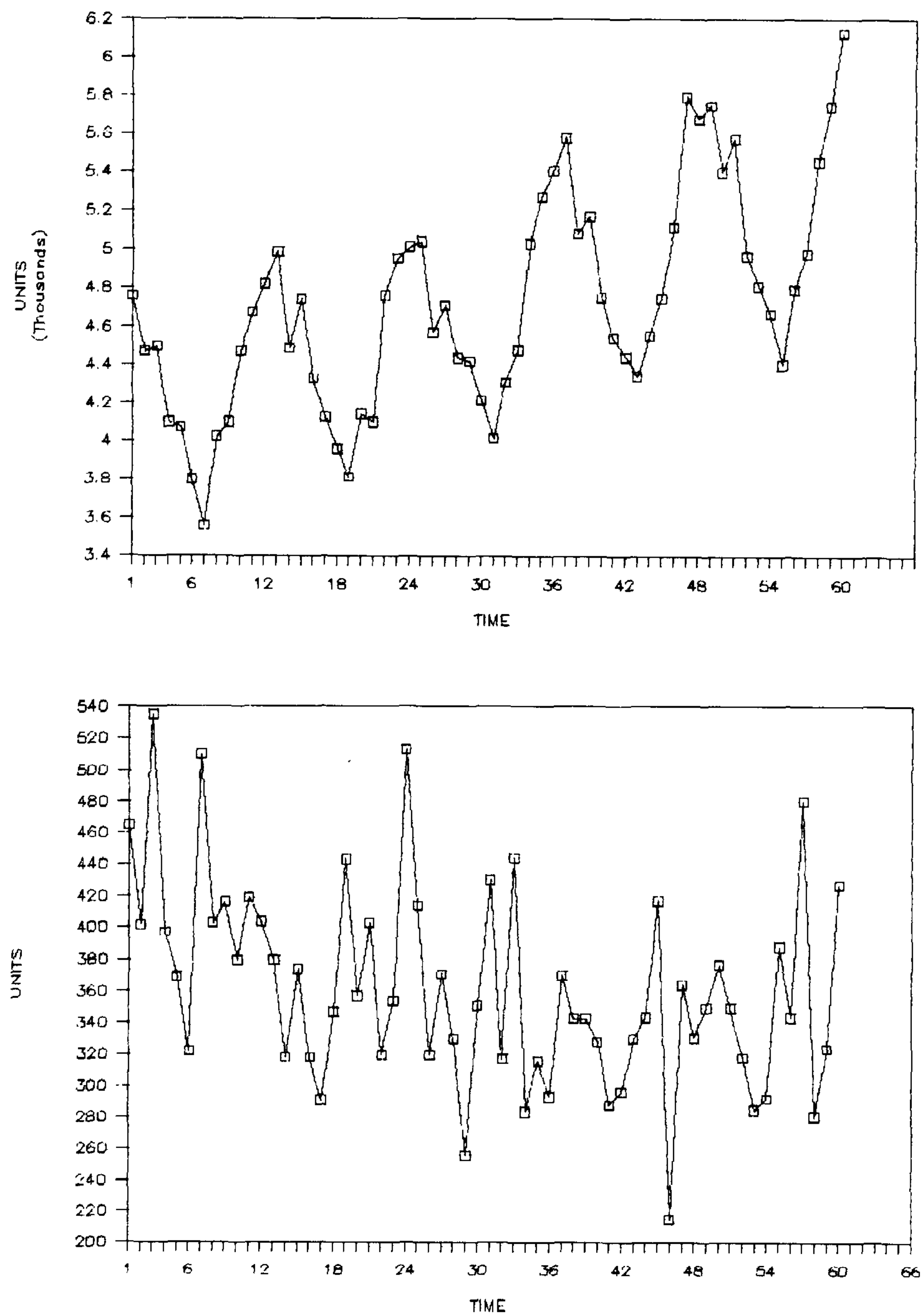

Exhibit 2. (above) The "easy" series. (below) The "difficult" series.

one of the group interaction processes – NGT, modified-consensus, or consensus. Documentation given to subjects explained the structuring method under which they were to work. For the modified consensus groups, one member was randomly selected to complete the task prior to the others. After he had finished the task, the forecasts and confidence intervals were then offered to the other members of the group for discussion and final estimates. All groups were given 30 minutes to complete the task. There were no differences between the interacting groups in the time taken to complete the forecasts.

5. Results

5.1. Manipulation check

A two-factor ANOVA design was adopted to assess the overall effects of different group interaction processes and the levels of forecast difficulty on forecast accuracy. Subjects working on the high difficulty series were more accurate than those working on the low difficulty series ($F = 167.6$, $df = 1{,}280$, $p < 0.000$). This confirms our manipulation check on varying forecast difficulty.

5.2. Accuracy of forecasts

An analysis of the mean MAPES presented in Exhibit 3 led to a rejection of H1 since performance significantly varies across group interaction processes ($F = 5.4$, $df = 3{,}280$, $p < 6.003$).

From the mean MAPES presented in Exhibit 3, the modified-consensus groups appear to be more accurate than groups under the other interaction processes. Post hoc analysis using the Scheffe decision rule confirmed this proposition ($F = 4.7$, $df = 3{,}280$, $p < 0.001$). A significant interaction effect between group interaction process and forecast difficulty was also observed ($F = 4.9$, $df = 3{,}280$, $p < 0.005$). As Exhibit 3 depicts, accuracy for the different group interaction processes tended to converge for the low difficulty series. All groups hovered around the 5% MAPE region. However, for the high difficulty series, the modified-consensus groups were considerably more accurate, with about half the error rate. Subsequent post hoc analysis found a significant difference between the modified-consensus treatment and the average of the other three group interaction processes for the high difficulty series ($F = 4.7$, $df = 3{,}140$, $p < 0.001$).

Exhibit 3
MAPES for various group interaction processes over high and low task difficulty.

Group structuring process	Task difficulty	
	Low	High
Staticized	4.9	35.80
NGT	5.03	33.13
Modified consensus	4.7	16.87
Consensus	5.57	32.10

Exhibit 4
Pre- and post interaction MAPES (and standard deviations) for NGT and modified consensus groups over task difficulty.

Task difficulty	Group structuring process	Pre-interaction	Post-interaction
Low	Modified consensus	8.73 (4.56)	4.70 (1.98)
	NGT	6.0 (7.8)	5.1 (8.8)
High	Modified consensus	28.83 (18.01)	16.87 (7.29)
	NGT	33.72 (16.1)	33.0 (16.5)

Benefits appeared to have been derived from the initial estimate made by one randomly selected member of the modified-consensus group. Subjects using this process were observed by the researchers to use the initial forecast as their focal point for subsequent discussion. They then seemed to anchor and adjust on the initial forecast to accommodate each other's viewpoints. A corollary research issue then is to investigate if there were any significant improvements made between the initial private forecast and the final forecast derived by consensus under the modified-consensus process. This issue is vital because if the group consensus forecast was inferior to the initial private forecast, it can be argued that the quality of the modified-consensus groups may be attributable more to the ability of the member of the group who made the initial forecast than to the virtue of the interaction process. An analysis of the initial and post-interaction forecasts for the modified-consensus group revealed a significant difference in MAPES ($F = 6.70$, $df = 1{,}701$ $p < 0.003$) observed between the initial forecast and the final consensus forecast. The MAPES of the final consensus forecast were significantly lower than the

MAPES of the solitary initial private forecast (see Exhibit 4).

This confirmed our belief that the interaction process in the modified-consensus process contributed to the accuracy of the group forecast. It provided additional assurance that the superiority of the modified-consensus groups was not solely attributed to the forecast ability of the member of the group who made the initial forecast.

We also investigated the pre- and post-interaction forecasts for the NGT groups. Recall that the NGT process starts with all members of the group undergoing a private, individual generation phase. They then come together, discuss their initial forecasts, and redo their individual estimates. These final estimates were then averaged to form the group's forecast. For purposes of comparing the pre- and post-interaction forecast accuracy, a simple average of the initial private estimates is computed to represent the pre-interaction forecast of the NGT group. This is then compared with the group forecast derived from a simple average of the final individual estimates. We found no significant difference in MAPE between the average of the initial forecasts and the final group forecast. Unlike the MAPES for the modified-consensus groups, the MAPE levels for the NGT groups remained relatively constant before and after NGT interaction. The NGT interaction session appeared to contribute little to the accuracy of the final group forecast. Thus, the interaction process was beneficial for the modified-consensus groups, but not for the NGT groups. These issues are discussed later.

Exhibit 4 also indicates the standard deviation of the forecast errors for the NGT groups did not change after interaction for both the high and low difficulty series. However, the standard deviations of forecast errors for the modified-consensus groups decreased substantially, especially for the high difficulty series. This result has important implications for business forecasting since a reduction in the variability of errors reduces uncertainty and greater confidence can be placed in decisions using the forecast.

5.3. *Calibration of the Confidence Intervals (CIs)*

As mentioned, calibration of the CIs refers to the number of actual values falling within the intervals. Thus, 50% CIs are said to be perfectly calibrated when 50% of the actual values fall within the intervals. If more than 50% of the values fall within the intervals, they are said to be under-confident. If too few fall within the intervals, they are overconfident.

Exhibit 5
Calibration of CIs over task difficulty.

Group structuring process	Task difficulty	
	Low	High
Staticized	37.96	13.89
NGT	27.78	14.82
Modified consensus	41.67	47.22
Consensus	36.11	16.67

Exhibit 5 indicates all groups were over-confident. However, the modified-consensus groups seemed less over-confident, especially for the high difficulty series. However, the effect of group interaction processes and task difficulty on calibration scores was not significant. We thus do not reject H2.

6. Discussion

In terms of forecast accuracy, not one group interaction process was clearly preferable for an easy series. However, for the high difficulty series, modified-consensus groups were significantly more accurate than the staticized, NGT or the consensus groups.

We believe that the initial private forecast performed by *only one member* of the group was the essence to the modified-consensus' superior performance. The initial forecast seemed to provide a focus for discussion. Groups under this process were observed to concentrate better on the task. Groups seemed to anchor their debate on the initial forecast and adjust this forecast to derive a final forecast which accommodated each other's viewpoints.

Whilst the presence of an initial private estimate seemed a vital factor in the success of the modified-consensus groups, the presence of three initial private estimates did not increase the performances of the NGT groups. One possible explanation may be that members in the NGT groups experienced and suffered *cognitive overload*. During the interaction phase, each member in a NGT group was faced with three-forecasts (his own plus the other two members' individual forecasts). Each forecast may be further decomposed into three information cues: the trend, seasonality, and noise of the series. This means that modified-consensus groups were required to assimilate and deliberate over three information cues, whilst NGT groups had to digest and absorb nine bits of information each. In many cases, lapses of silence were observed as each NGT member tried to wrestle with the forecasts of his counterparts. Members were also observed to devote a larger proportion of the time rationalizing their own initial estimates, and reconciling any differing viewpoints about the behavior of the series. This resulted in a post-interaction forecast which was not significantly different from the pre-interaction forecast. Accuracy was not statistically different from that obtained using the staticized group process where interaction of any form was prohibited.

Forecast accuracy of the consensus groups was also found to be inferior to that of the modified-consensus groups. However, the consensus group did not differ significantly from the staticized or the NGT groups. Consensus groups had minimal group structure levied: members may freely interact and discuss the problem at hand. However, without prior individual forecasts to steer the discussion, members were observed to spend an initial period engrossed in silent idea-generation. Group members also appeared to incline towards *premature closure*, which is the tendency to adopt the first feasible or apparently satisfactory alternative, rather than fully explore the problem.

In the task of confidence interval calibration, our finding concurred with the major reviews in the forecasting literature which showed that in general, over-confidence prevails in situations where subjects were unfamiliar with the tasks (see O'Connor, 1989). Contrary to the results of Ferrell (1985), our results indicate different group interaction processes did not effect the calibration of the CIs. All groups were consistently *over*confident, given that the subjects were unaccustomed to CI estimation. This implies that it may not be necessary to use interacting groups, since staticized groups (with total restriction on interaction) performed no worse than groups interacting under any of the other three group interaction processes.

7. Conclusion

For an easy time series, no one group interaction process was clearly preferable. However, for a difficult series, the modified-consensus groups were found to be significantly more accurate that the staticized, NGT or the consensus groups. We believe the modified-consensus groups performed better because the initial private forecast made by one member in the group provided a nucleus for discussion, and consequently, the group was observed to better focus on the task. With regards to calibrating judgmental confidence intervals, all groups tended to be *over*confident. This may result from lack of familiarity with the task.

The above findings on group performances in a forecasting task provided only a quantitative skeletal structure which may be subjected to a number of alternative explanations. The discussions provided in the paper were predominantly based on the observations of the researchers on the subjects during the experiment. Future research should attempt to enrich this analysis by juxtaposing it with more reliable, qualitative data obtained from group participants, and administrative diary records maintained by the researcher. By systematically observing the interaction process, one may be in a better position to discover the underlying reasons for differences in performance across the group interaction processes. Moreover, since this study is (we believe) the first to focus on the modified consensus approach, further research needs to establish the efficacy of such an interaction process in similar and different task settings.

References

Armstrong, J.S., 1985, *Long Range Forecasting*, 2nd ed. (Wiley–Interscience, New York, NY).

Boje, D.M. and J.K. Murnighan, 1982, "Group confidence pressures in iterative decisions", *Management Science*, 28, 1187–1196.

Carbone, R. and J.S. Armstrong, 1982, "Evaluation of extrapolative forecasting methods: results for academicians and practitioners", *Journal of Forecasting*, 1, 215–217.

Chalos, P. and S. Pickard, 1985, "Information choice and cue use: An experiment in group information processing", *Journal of Applied Psychology*, 10, 634–641.

Conference Board, 1978, *Sales Forecasting* (New York).

Dalrymple, D.J., 1987, "Sales forecasting practices", *International Journal of Forecasting*, 3, 379–391.

Edmundson, R.H., M.J. Lawrence and M. O'Connor, 1988, "The use of non time-series information in sales forecasting: A case study", *Journal of Forecasting*, 7, 201–211.

Einhorn, H.J., R.M. Hogarth and E. Klempner, 1977, "Quality of Group Judgement", *Psychological Bulletin*, 84, 158–172.

Ferrell, W.R., 1985, "Combining individual judgments", G. Wright, ed., *Behavioural Decision Making* G. Wright; (Plenum, Press, New York, NY), 111–143.

Fischer, G.W., 1981, "When oracles fail – A comparison of four procedures for aggregating subjective probability forecasts", *Organizational Behaviour and Human Performance*, 28, 96–110.

Golembiewski, R.T., 1962, *The Small Group: An Analysis of Research Concepts and Operations* (University of Chicago Press, Chicago, IL).

Gough G., 1975, "The effect of group format on aggregate subjective probability distributions", in: D. Wendt and C. Vlek, eds., *Utility, Probability, and Human Decision Making* (Kluwer, Dordrecht, Holland).

Gustafson, D.H., R.K. Shukla, A. Delbecq and G.W. Walster, 1973, "A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups", *Organizational Behaviour and Human Performance*, 9, 280–291.

Hackman, J.R., 1968, "Effects of task characteristics on group products", *Journal of Experimental Social Psychology*, 4, 162–187.

Hackman, J.R. and C.G. Morris, 1975, "Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration", in: L. Berkowitz, ed., (Academic Press, New York, NY), 45–99.

Janis, I.T., 1972, *Victims of Groupthink* (Houghton Mifflin, Boston, MA).

Lawrence, M.J., R.H. Edmundson and M.J. O'Connor, 1985, "An examination of the accuracy of judgmental extrapolation of time series", *International Journal of Forecasting*, 1, 25–35.

Lawrence, M.J., R.H. Edmundson and M.J. O'Connor, 1986, "The accuracy of combining judgmental and statistical forecasts", *Management Science*, 32, 1521–1532.

Libby, R., K.T. Trotman and I. Zimmer, 1987, "Member variation, recognition of expertise, and group performance", *Journal of Applied Psychology*, 72, 81–87.

Lichtenstein, S., B. Fischhoff and L. Phillips, 1982, "Calibration of probabilities: The state of the art to 1980", in: Kahneman, Slovic, and Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, UK).

Makridakis, S., A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen and R. Winkler, 1982, "The accuracy of extrapolation (time series) methods: Results of a forecasting competition", *Journal of Forecasting*, 1, 111–1153.

Mentzer, J.T. and J.E. Cox, 1984, "Familiarity, application and performance of sales forecasting technique", *Journal of Forecasting*, 3, 27–36.

Morris, C.G., 1966, "Task effects on group interaction", *Journal of Personality and Social Psychology*, 4, 545–554.

O'Connor, M.J., 1989, "Models of human behaviour and confidence in judgement: A review", *International Journal of Forecasting*, 5, 159–169.

O'Connor, M.J. and M.J. Lawrence, 1988, "An examination of the accuracy of judgemental confidence intervals in time series forecasting", *Journal of Forecasting*, 8, 141–155.

Payne, J.W., 1982, "Contingent decision behaviour", *Psychological Bulletin*, 92, 382–402.

PoKempner, S.J. and E. Bailey, 1970, *Sales Forecasting Practices* (The Conference Board, New York, NY).

Rohrbaugh, J., 1979, "Improving the quality of group judgment: Social judgement analysis and the Delphi technique; *Organizational Behaviour and Human Performance*, 24, 73–92.

Seaver, D.A., 1979, "Assessing probability with multiple individuals", Unpublished Doctoral Dissertation, University of Southern California.

Shaw, M.E., 1981, *Group Dynamic: The Psychology of Small Group Behaviour*, 3rd ed. (McGraw-Hill, New York, NY).

Sniezek, J.A., 1988, "A comparison of group techniques in judgmental forecasting with time series data", Unpublished paper, Department of Psychology, University of Illinois at Urbana-Champaign.

Trotman, K.T., 1985, "The review process and the accuracy of auditor judgments", *Journal of Accounting Research*, 23, 740–752.

Trotman, K.T. and P.W. Yetton, 1985, "The effect of the review process on auditor judgments", *Journal of Accounting Research*, 23, 256–267.

Trotman, K.T., P.W. Yetton and I.R. Zimmer, 1983, "Individual versus group judgments of internal control systems", *Journal of Accounting Research*, 21, 286–292.

Van de Ven, A. and A.L. Delbecq, 1971, "Nominal versus interacting group processes for committee decision-making effectiveness", *Academy of Management Journal*, 14, 203–212.

Van de Ven, A. and A.L. Delbecq, 1974, "The effectiveness of nominal, Delphi and interacting group decision making processes", *Academy of Management Journal*, 17, 605–621.

Winkler, R.L. and A.H. Murphy, 1973, "Experiments in the laboratory and real world", *Organizational Behaviour and Human Performance, 10, 252–270.*

Biographies: Soon ANG is a Senior Tutor with the School of Accountancy and Business at Nanyang Technological Institute, Singapore. She holds a Master of Commerce (Honours) degree in Information Systems from the University of New South Wales, Australia and is pursuing a Ph.D. in MIS at the University of Minnesota. Her current research interests include outsourcing, feedback seeking through electronic communication and valuing Information Systems investments using human resource accounting.

Marcus O'CONNOR is a Senior Lecturer in Information Systems at the University of New South Wales, Australia. His research interests centre on the role of human judgment in managerial decision making and the development of decision support systens to support people. He is particularly interested in the capacities and attitudes of people in forecasting tasks. He has published in many journals including *International Journal of Forecasting, Journal of Forecasting* and *Management Science*.