

An Exploratory Analysis of Portfolio Managers' Probabilistic Forecasts of Stock Prices

GÜLNUR MURADOĞLU AND DILEK ÖNKAL

Bilkent University, Turkey

ABSTRACT

This study reports the results of an experiment that examines (1) the effects of forecast horizon on the performance of probability forecasters, and (2) the alleged existence of an inverse expertise effect, i.e., an inverse relationship between expertise and probabilistic forecasting performance. Portfolio managers are used as forecasters with substantive expertise. Performance of this 'expert' group is compared to the performance of a 'semi-expert' group composed of other banking professionals trained in portfolio management. It is found that while both groups attain their best discrimination performances in the four-week forecast horizon, they show their worst calibration and skill performances in the 12-week forecast horizon. Also, while experts perform better in all performance measures for the one-week horizon, semi-experts achieve better calibration for the four-week horizon. It is concluded that these results may signal the existence of an inverse expertise effect that is contingent on the selected forecast horizon.

KEY WORDS Probabilistic forecasting Stock price forecasts
Calibration Inverse expertise effect

Forecasting accuracy of financial variables attracts considerable research attention, with predominantly conflicting findings. Liljeblom (1989) shows empirical evidence displaying the predictive power of financial analysts' forecasts of earnings per share for short-term forecast horizons in the Scandinavian securities market. DeBondt and Thaler (1990), on the other hand, conclude that the earnings forecasts provided by the security analysts are basically overreactive, i.e., when large earnings increases are forecasted actual earnings are lower than predictions and vice versa. Keane and Runkle (1990) use the ASA–NBER survey of economic forecasters and show that the price forecasts for GNP deflator by expert forecasters are rational demonstrating improved performance. Zarnovitz (1985) uses the same data set and rejects the rational expectations hypothesis for inflation forecasts. Summarizing the relevant work on financial forecasting, DeBondt (1991) concludes that 'finance should attempt to model the behavior of representative investors and the nature of their errors' (p. 90).

Contradictory results obtained in these studies may be viewed as resulting from data revisions, biases in aggregating data, and the definitions of forecast error that are employed.

Also, all the studies mentioned above focus on investigating the accuracy of point forecasts. (For discussions on enhancing point forecasts with prediction intervals and the associated complexities, see Klein, 1971; Granger and Newbold, 1986.) Probabilistic forecasting offers an alternative approach to forecasting in financial settings. Whereas the accuracy evaluations of point forecasts are quite limited, the use of probabilistic forecasts enables a detailed analysis of the various performance characteristics of forecasters.

Probabilistic forecasting tasks demand the assessment of subjective probabilities as articulations of the forecaster's degrees of belief in the occurrences of future outcomes. When viewed in this framework, probabilistic forecasts accord quantitative descriptions of forecasters' uncertainty (Murphy and Winkler, 1974). They also provide a channel for communicating this uncertainty to the users of the forecasts, who can, in turn, better 'interpret' the forecasts and make more informed decisions by assimilating the uncertainty associated with the forecasts. In addition to providing the information transmitted by point or categorical forecasts, probabilistic forecasts offer mechanisms that forecasters could utilize to convey their true judgments, hence reducing any biasing tendencies (Daan and Murphy, 1982). In short, it is argued that probabilistic forecasts are more useful than point or categorical forecasts, since they provide more detailed information to the users (Murphy and Winkler, 1992).

Within the realm of forecasting stock prices, subjective probability distributions were initially used by Bartos (1969) and Stael von Holstein (1972), with the common finding that the uniform distribution outperforms the distributions provided by various forecasters. Bartos (1969) replaced historical prices in the Markowitz (1959) mean-variance portfolio method by density functions and cumulative distribution functions assessed by security analysts, but could not delineate a better portfolio performance. Stael von Holstein (1972) investigated the effects of feedback on forecasting performance by using non-dichotomous distributions, but the forecasts could not outperform those based on past frequencies. Both studies concluded that further research on selection of appropriate forecast horizons is required.

More recently, Yates *et al.* (1991) reported experimental results on the overall inaccuracy of probabilistic forecasts of earnings per share and stock prices. Their work was especially important for their conjecture of the 'inverse expertise effect'. Yates *et al.* found an inverse relationship between expertise and forecast accuracy. Linking their findings to Hammond's (1966) work on probabilistic functionalism, the authors suggested that the inverse expertise effect is a byproduct of the experts' cue utilizations. That is, it is asserted that, in their decision-making processes, experts use a significantly larger number of cues than non-experts. This in turn, is believed to lead them to consolidate irrelevant cues into their judgment processes, thereby decreasing the accuracy of experts' forecasts.

The alleged inverse expertise effect was further found in a similar experiment utilizing only price forecasts (Önkal and Muradoğlu, 1994). Although both studies used students as forecasters, Yates *et al.* (1991) specified graduate students as experts, while Önkal and Muradoğlu (1994) defined experts as students who have previously made stock investment decisions. Also, while Yates *et al.* employed a three-month forecast horizon in a developed securities market, Önkal and Muradoğlu used a one-week horizon in an emerging market. Yates *et al.* used quarterly forecasts assuming that the two-week forecast horizon employed by Stael von Holstein (1972) was shorter than customary for professional forecasters. Bartos (1969) had used one-month, three-month, and six-month forecast horizons, but he found it difficult to make generalizations due to sample size restrictions. In order to reduce task complexity as well as to mimic the current practices of financial media, Önkal and Muradoğlu (1994) used a forecast horizon of one week.

Considering the utilization of different time horizons in the studies reviewed above, it may

be claimed that the choice of forecast horizon is a critical variable that may have a powerful impact on experimental results. This study aims to investigate the effects of forecast horizon on the performance of probability forecasters. In doing so, it seeks to examine the alleged existence of the inverse expertise effect by using probabilistic forecasts provided for different time horizons by portfolio managers. Portfolio managers are used as forecasters with substantive expertise. Performance of this 'expert' group is compared to the performance of a 'semi-expert' group entailing other banking professionals trained in portfolio management. To construct these comparisons, the experimental framework of Yates *et al.* (1991) is adapted to an inherently less complicated decision environment, and only forecasts of stock prices are used.

One of the primary contributions of this research is that it employs professional fund managers as forecasters with substantive expertise. Second, various performance aspects of these expert and corresponding semi-expert groups are compared with respect to differing forecast horizons. Third, the particular setting used (i.e., a developing economy setting with an inefficient emerging security market) avoids any misinterpretations of forecast accuracy through assumptions regarding market efficiency. Whereas previous research has attempted to explain the relatively poor forecasting performance of experts via market behaviour (i.e., conditions of market efficiency), our research aims to analyse forecasting performance via experts' and semi-experts' expressions of uncertainty.

MARKET EFFICIENCY AND PORTRAYAL OF THE EXPERIMENTAL SETTING

A market is defined to be efficient with respect to an information set if it is not possible to make abnormal profits by adopting trading strategies using this information set (Jensen, 1978). The semi-strong form of market efficiency uses publicly available information in addition to past prices as the relevant information set. Therefore, if a market is efficient in the semi-strong form, expert-managed portfolios, for example, are not expected to beat the market since the information set is equally accessible to all (Jensen, 1968). In this case, 'the basic question most asked is—are price changes forecastable?' (Granger, 1992, p. 3). There is contradictory evidence from finance literature regarding the performance of expert-managed funds (Elton *et al.*, 1991; Ippolito, 1989). When expert-managed funds do not beat the market, this is typically attributed to the efficiency of the market, i.e., an explanation is provided via market behaviour rather than forecaster performance.

It may be argued that the existence of an inverse expertise effect can be used as a behavioural explanation to the inconclusive results of semi-strong form efficiency tests using professionally managed portfolios. The implication is that the setting used in the current study (i.e., a developing economy with an inefficient emerging security market (Şengül and Önkal, 1992; Unal, 1992)) prohibits any misinterpretations of forecast accuracy via market-efficiency assumptions.

In particular, the financial markets in Turkey were strictly regulated until 1980, when the IMF and the World Bank supported the introduction of a liberalization package. In 1982 the required legal framework and regulatory agencies for the stock market were established. Istanbul Securities Exchange, the only stock exchange in Turkey, started operations in 1986. Employees of the stock exchange could hold stock portfolios without notification until 1988, and there was no legislation against insider trading until 1990. When this study was conducted (i.e. February 1992), 143 stocks were traded at the Istanbul Securities Exchange, and the average daily volume of trade was US\$55 million. There was a total of 162 intermediaries and brokerage houses, 60 of which were affiliated with companies traded in the exchange.

PROCEDURE

All the subjects in the study were professionals in the stock market. No monetary or non-monetary bonuses were offered to the participants. The current study was depicted as one that suggests an alternative method of presenting stock price forecasts. Concerns on the evaluation of forecasting performance and reliability of forecasts were discussed as primary issues involving portfolio management. Probabilistic forecasts were presented as pertinent channels of communication between forecasters and users of forecasts. It was argued that if forecasters could employ probabilistic forecasts as quantitative descriptions of their uncertainty, it may be possible to reduce biases such as overforecasting.

Participants of the study were reached at two different locations at the same date. The first group, called 'semi-experts', was composed of two internal auditors and eight managers who had completed a company-paid 40-hour training programme on portfolio management. The internal auditors were trained in portfolio management since they were going to be specializing in the auditing of bank-affiliated brokerage houses. All the managers in this group were employed in banks. They attended the training programme because they were expected to work for bank-affiliated brokerage houses upon completion of the programme.

The second group, called 'experts', was composed of seven portfolio managers working for a bank-affiliated brokerage house. All the experts had licences as brokers and their job descriptions included managing investment funds and giving investment advice to customers with investments above US\$50,000.

The task was defined as the preparation of probabilistic forecasts of the closing stock prices for 34 companies listed on the Istanbul Stock Exchange. Companies had been selected on the basis of their volume of trade on the preceding 52-week period. This particular selection was made to minimize task complexity as the stocks with relatively high volumes of trade could be followed with minimum effort by all experts and semi-experts.

Participants were requested to make probabilistic forecasts regarding the price changes for the 34 stocks for forecast horizons of one week, two weeks, four weeks, and twelve weeks. In particular, the forecasts were to be made regarding the percentage change between the previous Friday's closing stock price and (1) the closing stock price that will be realized on the Friday due in one week, (2) the closing stock price that will be realized on the Friday due in two weeks, (3) the closing stock price that will be realized on the Friday due in four weeks, and (4) the closing stock price that will be realized on the Friday due in 12 weeks.

Subjects were requested to provide these forecasts in the form of subjective probabilities conveying their degrees of belief in the actual price change falling into the designated percentage change categories. Specifically, they were asked to complete the following response form for each stock for each forecast horizon:

	PRICE CHANGE INTERVAL (in percentages, Friday to Friday)	PROBABILITY
(8)	INCREASE 15% or more	_____ %
(7)	INCREASE 10%—up to 15%	_____ %
(6)	INCREASE 5%—up to 10%	_____ %
(5)	INCREASE up to 5%	_____ %
(4)	DECREASE 0%—up to 5%	_____ %
(3)	DECREASE 5%—up to 10%	_____ %
(2)	DECREASE 10%—up to 15%	_____ %
(1)	DECREASE 15% or more	_____ %
		100 %

The range of stock price changes in the response form were formulated by considering the average weekly, bi-weekly, monthly, and quarterly price changes of the composite stock index during the previous 52-week period. For the past 52 weeks, weekly price changes were 3% on average with the maximum increase being 8% and the maximum decrease 5%. During the same period, the average quarterly price change was 13% with the maximum increase being 65% and the maximum decrease 21%. The first 5% increase range (Interval 5) contained the average weekly increase during the previous year, the second (Interval 6) contained the maximum average weekly change, the third (Interval 7) contained the average quarterly change, and the fourth range (Interval 8) was designed for stocks whose quarterly volatility could be higher than average. Intervals 1 to 4 were designed to be symmetric to intervals 5 to 8 for cognitive purposes. Subjects were not given different intervals for different forecasting periods in order not to convey and confound the personal opinions of the researchers regarding the expected changes for differing forecast horizons.

In the beginning of the experiment, concepts related to subjective probabilities and probabilistic forecasting tasks were discussed with detailed examples. Design and goals of the study were depicted. Participants were not given any information about the results of similar studies on inverse-expertise effect (e.g., Önkal and Muradoğlu, 1994; Yates *et al.*, 1991) so that their motivation would not be affected.

Subjects were informed that certain scores of probabilistic forecasting performance would be computed from their individual forecasts. They were advised that, due to the computational characteristics of these scores (i.e., proper scoring rules), participants could earn their best potential scores by expressing their true opinions and thereby avoiding hedging or bluffing. Subjects were told that they would be notified about their forecasting performance on a personal basis, and no information about their direct or implied individual performances would be provided to their managers or co-workers.

Each subject was provided with background folders for each of the 34 companies, delineating company name, industry, net profits as of the end of the third quarter of 1991, earnings per share, and price-earnings ratios as of the last day of the preceding week. Also provided were the weekly closing stock prices (i.e., the closing stock prices for each Friday) of the preceding 52 weeks in graphical form. Weekly closing stock prices for the last 3 months (12 weeks) were also presented in tabular form. Subjects were permitted to utilize any source of information other than the remaining participants of the study.

In order to duplicate real forecasting settings, subjects were allowed to take the background folders home. They were given the background folders and response sheets on Friday afternoon (after the session has closed and the closing prices were known) and were requested to submit the completed response sheets by Monday 9 a.m. (before the opening of the next session at the stock exchange). Participants were to render their forecasts using the response forms illustrated previously.

RESULTS

Forecasting performances of experts and semi-experts for the various forecast horizons are investigated using the following measures (see the Appendix for a detailed discussion of these performance indices):

- (1) Mean probability score (\overline{PSM})—presents a measure of overall forecasting accuracy;
- (2) Calibration—provides an index of the forecaster's ability to match the probabilistic forecasts with the relative frequencies of occurrence;

- (3) Mean slope—gives a measure of the forecaster's ability to discriminate between instances when the realized price change will or will not fall into the specified intervals;
- (4) Scatter—conveys a measure of excessive forecast dispersion;
- (5) Forecast profile variance—compares how close the forecaster's probability profile is to a flat profile that shows no variability across intervals; and
- (6) Skill—presents an index of the overall effect of those accuracy components that are under the forecaster's control.

The analyses utilizing these measures are conducted at three levels. First, performances across different forecast horizons are tracked within each group by using median tests. Next, for each of the forecast horizons, comparisons of experts and semi-experts are made using Mann-Whitney *U*-tests for each of the performance measures. Finally, scores of experts and semi-experts are compared to those that would be obtained by the uniform, historical, and base-rate forecasters (i.e., the three comparison standards for forecasters as suggested by Yates *et al.*, 1991). The uniform forecaster presents a comparison basis for forecasters, since it makes no discriminations among intervals, and hence, assigns equal probabilities to all the intervals (i.e., probability of 1/8 to each of the eight intervals for each stock). Second standard of comparison is given by a historical forecaster, who provides forecasts identical to the historical relative frequencies. Given the volatility of the stock market under consideration, the historical forecaster's probability forecasts are set equal to the relative frequencies realized in the previous week. The third comparison standard is provided by the base-rate forecaster. This forecaster is analogous to a clairvoyant who can perfectly foresee the relative frequencies (i.e., base rates) with which the price changes will occur for that forecast horizon.

Table I presents the medians for the six performance measures for experts and semi-experts, along with the scores that would be obtained by the uniform, historical, and base-rate forecasters, for the four forecast horizons of interest. As can be observed from this table, experts clearly outperform semi-experts for the one-week forecast horizon. In particular, experts attain better scores in *PSM* ($p = 0.026$), calibration ($p = 0.016$), mean slope ($p = 0.016$), scatter ($p = 0.049$), and skill ($p = 0.026$). Forecast profile variances of experts and semi-experts for this forecast horizon are very similar.

Experts and semi-experts also show very similar performances in all measures for the two- and 12-week forecast horizons. It is in the four-week forecast horizon that an inverse expertise effect with respect to calibration is observed. Calibration performance of semi-experts is found to be significantly better than that of the experts for the four-week horizon ($p = 0.012$). This indicates that semi-experts' probabilistic forecasts corresponded more closely with the realized relative frequencies for that particular forecast horizon.

In fact, the four-week forecast horizon appears to be the optimal time specification for the forecasters' discrimination performances, as indexed by their mean slope scores. This is true regardless of the forecasters' levels of expertise. For the expert group, mean slope scores obtained in the four-week horizon are better than the scores obtained in the one-, two-, and 12-week horizons (all $p = 0.015$). Also, calibration and skill performances of our expert participants are found to be at their worst level for the 12-week horizon. Calibration and skill performances for the 12-week period is worse than the corresponding performances in one, two, and four weeks (all $p = 0.0003$ for calibration; and all $p = 0.015$ for skill). As a result of these findings, forecasting accuracy (as indexed by the mean probability score) for the four-week horizon is found to be better than that of the 12-week horizon ($p = 0.015$).

It is interesting to note that the identical conclusions prevail for our semi-experts. Mean slope for the four-week horizon is found to be superior for semi-experts than that for the one-

week ($p = 0.0006$), two-week ($p = 0.089$), and 12-week ($p = 0.012$) horizons. Calibration and skill scores are again found to be the worst for the 12-week horizon. Specifically, the calibration performance in 12 weeks is worse than the performance displayed in one week ($p = 0.0006$), two weeks ($p = 0.0000$), and four weeks ($p = 0.0000$). Likewise, this group's skill performance for the 12-week horizon is surpassed by its performance in the one-week

Table I. Median values for various performance measures across different forecast horizons for experts and semi-experts, with corresponding measures for uniform (U), historical (H), and base-rate (B) forecasters

Performance measure	Forecast horizon			
	1 week	2 weeks	4 weeks	12 weeks
<i>PSM</i>				
Experts	0.884 ^{Ea}	1.035	1.038	1.298
Semi-experts	1.071	1.084	0.944	1.144
U	0.875	0.875	0.875	0.875
H	0.781	0.787	0.889	0.863
B	0.752	0.768	0.740	0.570
Calibration				
Experts	0.052 ^{Eb}	0.134	0.224	0.553
Semi-experts	0.218	0.100	0.130 ^{Sb}	0.498
U	0.123	0.107	0.135	0.305
H	0.029	0.019	0.148	0.293
B	0.000	0.000	0.000	0.000
Mean slope				
Experts	-0.015 ^{Eb}	-0.014	-0.002	-0.060
Semi-experts	-0.043	-0.026	0.017	-0.036
U	0.000	0.000	0.000	0.000
H	0.000	0.000	0.000	0.000
B	0.000	0.000	0.000	0.000
Scatter				
Experts	0.058 ^{Ea}	0.046	0.060	0.053
Semi-experts	0.115	0.115	0.088	0.111
U	0.000	0.000	0.000	0.000
H	0.000	0.000	0.000	0.000
B	0.000	0.000	0.000	0.000
Forecast profile variance				
Experts	0.032	0.032	0.039	0.038
Semi-experts	0.031	0.030	0.019	0.020
U	0.000	0.000	0.000	0.000
H	0.010	0.010	0.010	0.010
B	0.015	0.013	0.017	0.038
Skill				
Experts	0.132 ^{Ea}	0.220	0.298	0.728
Semi-experts	0.319	0.268	0.204	0.574
U	0.123	0.107	0.135	0.305
H	0.029	0.019	0.148	0.293
B	0.000	0.000	0.000	0.000

^E Experts better than semi-experts for the given forecast horizon.

^S Semi-experts better than experts for the given forecast horizon.

^a $p < 0.05$.

^b $p < 0.02$.

($p = 0.089$), two-week ($p = 0.012$), and four-week ($p = 0.012$) horizons. As a result of these performances, the mean probability scores for semi-experts for the four-week period are better than the 12-week period ($p = 0.089$).

We also accumulate evidence for an inverse expertise effect that is contingent on the forecast horizon, when we compare the performances of experts and semi-experts with those of the

Table II. Percentage of experts and semi-experts attaining better scores in various performance measures when compared with the uniform (U), historical (H), and base-rate (B) forecasters for different forecast horizons

Performance measure	Forecast horizon			
	1 week	2 weeks	4 weeks	12 weeks
<i>PSM</i>				
Experts				
U	43%	0%	0%	0%
H	0%	0%	0%	0%
B	0%	0%	0%	0%
Semi-experts				
U	0%	10%	30%	10%
H	0%	0%	30%	0%
B	0%	0%	0%	0%
Calibration				
Experts				
U	57%	29%	14%	0%
H	29%	0%	14%	0%
B	0%	0%	0%	0%
Semi-experts				
U	30%	60%	60%	0%
H	0%	0%	60%	0%
B	0%	0%	0%	0%
Mean slope				
Experts				
U	14%	0%	43%	14%
H	14%	0%	43%	14%
B	14%	0%	43%	14%
Semi-experts				
U	0%	30%	60%	0%
H	0%	30%	60%	0%
B	0%	30%	60%	0%
Skill				
Experts				
U	43%	0%	0%	0%
H	0%	0%	0%	0%
B	0%	0%	0%	0%
Semi-experts				
U	0%	20%	30%	10%
H	0%	0%	30%	0%
B	0%	0%	0%	0%

Notes:

- (1) Scatter scores were not included since there were no experts/semi-experts achieving better scatter scores than U, H, B.
- (2) Comparisons with respect to the forecast profile variance were not given since this particular measure conveys the difference between a forecaster's probability profile and a flat profile given by a uniform forecaster.

uniform, historical, and base-rate forecasters. As can be observed from Table II, the overall performance of 43% of experts are better than the uniform forecaster only for the one-week forecast horizon, while semi-experts demonstrate improved overall performance in comparison to the uniform forecaster for longer forecast horizons. Also, 30% of semi-experts show an overall accuracy level better than that of the historical forecaster for the four-week forecast horizon. This may be viewed as signalling that the use of subjective probabilities instead of the customary use of historical data could result in higher profit opportunities.

Regarding the calibration performance, it can be deduced from Table II that experts' ability to assign probabilities matching the actual relative frequencies of future outcomes deteriorate for longer forecast horizons. In contrast, the semi-experts seem to show improvement for longer horizons, with the most pronounced difference surfacing in the four-week forecast horizon.

Mean slope scores indicate that the semi-experts' ability to discriminate between instances when the actual price change will and will not fall in the specified intervals ameliorates more than that of the experts for the four-week forecast horizon. This finding persists for all benchmarks, including the uniform, historical, and base-rate forecasters.

It can be seen from Table II that, while 43% of the experts show better skill than uniform forecaster for the one-week forecast horizon only, semi-experts attain better skill scores as the forecast horizon is extended, with better performances than the historical forecaster for the four-week horizon.

DISCUSSION AND CONCLUSION

This research was focused on analysing the probabilistic forecasts of stock prices given by portfolio managers and other banking professionals in an emerging securities market setting. Goals of the study included (1) exploring the potential effects of varying forecast horizons on different dimensions of forecasting accuracy, and (2) investigating the alleged existence of the so-called inverse expertise effect. A significant contribution of this work involved the critical finding that the previous assertions regarding the so-called inverse expertise effect (Stael von Holstein, 1972; Yates *et al.*, 1991; Önkal and Muradoğlu, 1994) could be expanded to incorporate a dependence on the selected forecast horizon. As amplified above, the portfolio managers and the other banking professionals participating in the current study show distinctly different performances for the varying forecast horizons of interest.

Our findings suggest that the forecast horizon yielding the best discrimination performance for both experts and semi-experts is four weeks under the current emerging securities market setting. All the participants show superior performance (as displayed by the mean slope scores) in discriminating between occasions when the actual price change will or will not fall into the specified intervals for the four-week horizon in comparison to the one-, two-, and 12-week horizons. Also, regardless of the level of expertise, calibration and skill scores for all participants reach their least desirable values for the 12-week forecast horizon. These results may be viewed as implying that the forecasters' ability to assign appropriate probabilities to actual outcomes as well as their overall forecasting skill may be better for shorter forecast horizons. It may also be that the poorer performance observed for 12 weeks may relate to caution or 'mean reversion' in forecasting which could be vindicated over a longer time series.

Comparisons of the portfolio managers' performances with those of other banking professionals suggest that the performance of experts becomes worse than that of semi-experts as the forecast horizon is extended. For the one-week forecast horizon, the performance of experts are significantly better than that of semi-experts. For the two-week forecast horizon

there is no difference between the forecast performance of portfolio managers and semi-experts. However, for the four-week forecast horizon, which appears to be the optimal period for discrimination performances, it is observed that the calibration scores of semi-experts are better than that of experts; i.e., banking professionals' skills in assigning probabilities that match the realized relative frequencies are better than that of portfolio managers.

Previous research has adopted two main explanations for the inverse expertise effect. The first explanation stems from the assertion that experts use more abstract representations of the problem situation, hence causing novices to outperform the experts (Adelson, 1984). The second suggests that since experts use richer representations (Murphy and Wright, 1984), their use of additional cues make the judgment task more difficult for them, thus distorting the accuracy of their judgments (Yates *et al.*, 1991). Current results may be viewed as implying that the representations used by portfolio managers, along with their cue utilizations, may change with the forecast horizons. It may be argued that, since portfolio managers predominantly make short-term forecasts in emerging markets with higher volatilities, shorter forecast horizons provide a better fit to these experts' natural environments. This leads to a superior forecasting performance for the one-week forecast horizon. However, as the forecast horizon becomes longer, there emerges a discrepancy with the experts' natural domain. Portfolio managers may be viewed as responding to this discrepancy by showing a deteriorated calibration performance, thus further confirming the arguments of the ecological approach that probability assessors are well calibrated to their natural environments (Gigerenzer *et al.*, 1991; Juslin, 1993). Relatedly, portfolio managers are not found to be successful in dispersing their uncertainty through intervals for longer forecast horizons. Instead, they choose to concentrate on a few intervals, and their calibration scores are distorted.

Financial forecasts are generally reported in the form of point or categorical forecasts that do not disclose how firmly the forecasters believe in their expectations. Griffin and Tversky (1992) have asserted that experts could be expected to become more overconfident than novices in unpredictable domains like the stock market, since they are more likely to give unwarranted credibility to their fallible expert knowledge. Portfolio managers' potential miscalibration and the resulting overconfidence may mislead the users of financial forecasts in several ways. First, investment advice supplied by these experts may result in investors forming less diversified portfolios. Second, the long-run performance of portfolios managed by these experts may not be different from randomly selected others or benchmarks utilizing some normal return criteria. Inefficiency of the market in strong form may in fact be a signal of the expert's behavior.

Previous research has repeatedly emphasized the critical role played by judgment in economic and financial forecasting (Wright and Ayton, 1987; Batchelor and Dua, 1990; McNees, 1990; Turner, 1990; Wolfe and Flores, 1990; Bunn and Wright, 1991; Flores, Olson and Wolfe, 1992; Donihue, 1993). The use of probabilistic forecasts provides a means for consolidating the uncertainty inherent in the stock market into financial forecasts. The users of financial information are already accustomed to probability distributions, which may in turn be effectively utilized to convey the forecasters' degrees of uncertainty. The providers of financial forecasts can employ probabilities as mechanisms for assessing and evaluating their uncertainty, hence furnishing more detailed information to the users. In short, it is our belief that probabilistic forecasting constitutes an important channel of communication between the providers and the users of forecasts, and that it demands efficient utilization in financial settings.

We would expect future work in this area to concentrate on the following research avenues. First, the differential performances observed for the forecast horizons in this study may in part

be a product of the specific conditions of the financial environment. Therefore, comparisons of portfolio managers' forecasts in different settings need to be made by controlling for the volatility of the stock markets and the efficiency conditions. Second, similar experiments need to be conducted in bull and bear markets to control for the effects of cyclical movements in stock markets. Finally, feedback studies with portfolio managers are required to examine whether the forecast-horizon-dependent inverse expertise effect found in the current study will disappear with relevant training and feedback.

Following the framework of O'Connor (1989), performance in probabilistic forecasting tasks may be viewed as a function of the participants' familiarity with the topic, experience with subjective probabilities, availability of feedback, and the environmental context. Accordingly, delineating the implications of these factors constitutes a critical threshold for effective applications of judgmental forecasting in financial settings.

APPENDIX: PERFORMANCE MEASURES USED

1. Probability score for multiple events

The forecast vector given by a subject for each stock is defined as $\mathbf{f} = (f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8)$, where f_k denotes the probabilistic forecast that the stock's price change will fall into interval k , $k = 1, 2, \dots, 8$. The outcome index vector is defined as $\mathbf{d} = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)$, where d_k assumes the value of 1 if the realized price change falls within interval k and the value of 0 if it does not. The scalar product of the difference between the forecast vector and the outcome index vector gives the probability score for multiple events (PSM). That is:

$$PSM = (\mathbf{f} - \mathbf{d})(\mathbf{f} - \mathbf{d})^T = \sum (f_k - d_k)^2$$

As defined above, the range of PSM is $[0, 2]$. The lower this value, the better is the forecaster's accuracy with respect to the particular stock in question. Accordingly, a forecaster's overall accuracy level can be indexed by taking the mean of the probability scores (PSM) over a specified number of forecasting instances (i.e., over a given number of stocks). Components emanating from the Yates decomposition of PSM (as explained below) are employed as indices of various performance attributes of forecasters (Yates, 1988).

2. Calibration

Calibration provides information about the forecaster's ability to assign appropriate probabilities to outcomes. A forecaster is well calibrated if for all predicted outcomes assigned a given probability, the proportion of those outcomes that occur (i.e., proportion correct) is equal to the probability. For example, suppose that, over 100 predictions, a forecaster assesses a probability of 0.3 that the given stock's price will increase by more than 15%. This forecaster's 0.3 assessments are well calibrated if an increase of more than 15% is actually observed on 30 of the 100 predictions. If the forecaster's other probability forecasts similarly match event frequencies, the forecaster is said to be well calibrated. Accordingly, a calibration score can be computed as a function of \bar{f}_k , representing the mean probability forecast for interval k , and \bar{d}_k , representing the proportion correct for interval k :

$$\text{Calibration} = \sum (\bar{f}_k - \bar{d}_k)^2$$

Lower scores refer to better calibration performance.

3. Mean slope

Mean slope provides an indication of the forecaster's ability to discriminate between occasions when the actual price change will and will not fall into the specified intervals. Higher values of the mean slope imply finer discrimination. The computation of mean slope is as follows:

$$\text{Mean slope} = (1/K) \Sigma \text{Slope}_k = (1/K) \Sigma (\bar{f}_{1k} - \bar{f}_{0k})$$

where \bar{f}_{1k} is the mean of probability forecasts for a price change falling into interval k computed over all the cases where the realized price change actually fell into interval k . Similarly, \bar{f}_{0k} is the mean of probability forecasts for a price change falling into interval k computed over all the times when the realized price change did not fall into the specified interval.

4. Scatter

Scatter is that part of the overall forecast variance that is not attributable to the forecaster's ability to discriminate between occasions when the actual price change will or will not fall into the specified intervals. The best value of the scatter would be zero, since it signals excessive variance due basically to the forecaster's reaction to nonpredictive environmental cues. A scatter index could be computed as follows:

$$\begin{aligned} \text{Scatter} &= \Sigma \text{Scatter}_k \\ &= \Sigma (1/N) [(N_{1k} * \text{Var}(f_{1k})) + (N_{0k} * \text{Var}(f_{0k}))] \end{aligned}$$

where $\text{Var}(f_{1k})$ is the conditional variance of the N_{1k} forecasts given for a price change falling into interval k when it actually occurred. Similarly, $\text{Var}(f_{0k})$ is the conditional variance of the N_{0k} forecasts given for a price change falling into interval k when it did not. Of course, $N = N_{1k} + N_{0k}$.

5. Forecast profile variance

Forecast profile variance measures the discrepancy between a forecaster's set of probabilities and a uniform set of probabilities. Hence, the forecast profile variance compares the forecaster's probability profile with a flat profile that shows no variability across intervals. An index of the forecast profile variance (for our study employing eight intervals) could be computed as follows:

$$\text{Forecast profile variance} = (1/N) \sum_n \left[\left(\sum_k (f_k - 0.125)^2 \right) / 8 \right]$$

This index provides a measure of how different the forecaster's probabilities are from the non-discriminating probabilities of the uniform forecaster.

6. Skill

The overall effect of those \overline{PSM} components under the forecaster's control can be indexed via a skill score, which can be computed as follows:

$$\text{Skill} = \overline{PSM} - \Sigma \text{Var}(d_k) = \overline{PSM} - \Sigma [(\bar{d}_k) * (1 - \bar{d}_k)]$$

where $\text{Var}(d_k)$ is the variance of the outcome index d_k for interval k . Since d_k is determined by what happens in the forecasting environment (i.e. the realized price change), $\Sigma \text{Var}(d_k)$ indexes an uncontrollable element of \overline{PSM} . Subtracting this 'base-rate' component from \overline{PSM} , we have the overall effect of those components that are under the forecaster's control. Hence,

lower skill scores imply better overall forecasting quality as displayed by the probability forecasts.

ACKNOWLEDGEMENTS

We would like to thank Derek Bunn and two anonymous reviewers for their helpful comments and valuable suggestions.

REFERENCES

- Adelson, B., 'When novices surpass experts: The difficulty of a task may increase by expertise', *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10** (1984), 483–495.
- Bartos, J. A., *The Assessment of Probability Distributions for Future Security Prices*, Unpublished Ph.D. thesis, Indiana University, Graduate School of Business, 1969.
- Batchelor, R. and Dua, P., 'Forecaster ideology, forecasting technique, and the accuracy of economic forecasts', *International Journal of Forecasting*, **6** (1990), 3–10.
- Bunn, D. and Wright, G., 'Interaction of judgemental and statistical forecasting methods: issues and analysis', *Management Science*, **37** (1991), 501–518.
- Daan, H. and Murphy, A. H., 'Subjective probability forecasting in the Netherlands: some operational and experimental results', *Meteorologische Rundschau*, **35** (1982), 99–112.
- DeBondt, W. F. M., 'What do economists know about the stock market?' *Journal of Portfolio Management*, **4** (1991), 84–91.
- DeBondt, W. F. M. and Thaler, R. M., 'Do security analysts overreact?' *American Economic Review*, **80** (1990), 52–57.
- Donihue, M. R., 'Evaluating the role judgment plays in forecast accuracy', *Journal of Forecasting*, **12** (1993), 81–92.
- Elton, E. J., Gruber, M. J., Das, S. and Hklarka, M., 'Efficiency with costly information: a reinterpretation of evidence from managed portfolios', Unpublished manuscript, New York University, 1991.
- Flores, B. E., Olson, D. L. and Wolfe, C., 'Judgmental adjustment of forecasts: a comparison of methods', *International Journal of Forecasting*, **7** (1992), 421–433.
- Gigerenzer, G., Hoffrage, U. and Kleinbölting, H., 'Probabilistic mental models: a Brunswikian theory of confidence', *Psychological Review*, **98** (1991), 506–528.
- Granger, C. W. J., 'Forecasting stock market prices: lessons for forecasters', *International Journal of Forecasting*, **8** (1992), 3–13.
- Granger, C. W. J. and Newbold, P., *Forecasting Economic Time Series*, 2nd edition, San Diego: Academic Press, 1986.
- Griffin, D. and Tversky, A., 'The weighting of evidence and the determinants of confidence', *Cognitive Psychology*, **24** (1992), 411–435.
- Hammond, K. R., 'Probabilistic functionalism: Egon Brunswick's integration of the history, theory and method of psychology', in Hammond, K. R. (ed.), *The Psychology of Egon Brunswick*, New York: Holt, Rinehart and Winston, 1966.
- Ippolito, R. A., 'Efficiency with costly information: a study of mutual fund performance, 1965–1984', *Quarterly Journal of Economics*, **104** (1989), 1–23.
- Jensen, M. C., 'The performance of mutual funds in the period 1945–64', *Journal of Finance*, **2** (1968), 389–416.
- Jensen, M. C., 'Some anomalous evidence regarding market efficiency', *Journal of Financial Economics*, **6** (1978), 95–101.
- Juslin, P., 'An explanation of the hard–easy effect in studies of realism of confidence in one's general knowledge', *European Journal of Cognitive Psychology*, **5** (1993), 55–71.
- Keane, M. P. and Runkle, D. E., 'Testing the rationality of price forecasts: new evidence from panel data', *American Economic Review*, **80** (1990), 714–735.
- Klein, L. R., *An Essay on the Theory of Economic Prediction*. Chicago: Markham, 1971.

- Liljebloom, E., 'An analysis of earnings per share forecasts for stocks listed on the Stockholm Stock Exchange', *Scandinavian Journal of Economics* (1989), 565–581.
- Markowitz, H., *Portfolio Selection: Efficient Diversification of Investments*, New York: John Wiley, 1959.
- McNees, S. K., 'The role of judgment in macroeconomic forecasting accuracy', *International Journal of Forecasting*, 6 (1990), 287–299.
- Murphy, A. H. and Winkler, R. L., 'Subjective probability experiments in meteorology: some preliminary results', *Bulletin of the American Meteorological Society*, 55 (1974), 1206–1216.
- Murphy, A. H. and Winkler, R. L., 'Diagnostic verification of probability forecasts', *International Journal of Forecasting*, 7 (1992), 435–455.
- Murphy, G. L. and Wright, J. C., 'Changes in conceptual structure with expertise: differences between real world experts and novices', *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10 (1984), 144–155.
- O'Connor, M., 'Models of human behaviour and confidence in judgement: a review', *International Journal of Forecasting*, 5 (1989), 159–169.
- Önkal, D. and Muradoğlu, G., 'Evaluating probabilistic forecasts of stock prices in a developing stock market,' *European Journal of Operational Research*, 74 (1994), 350–358.
- Şengül, G. M. and Önkal, D., 'Semistrong form efficiency in a thin market: a case study'. Paper presented at the 19th Annual Meeting of the European Finance Association, Lisbon, 1992.
- Stael von Holstein, C. A. S., 'Probabilistic forecasting: an experiment related to the stock market', *Organizational Behavior and Human Performance*, 8 (1972), 139–158.
- Turner, D. S., 'The role of judgment in macroeconomic forecasting', *Journal of Forecasting*, 9 (1990), 315–345.
- Unal, M., *Weak Form Efficiency Tests in Istanbul Stock Exchange*, Unpublished MBA Thesis, Bilkent University, Ankara, 1992.
- Wolfe, C. and Flores, B., 'Judgmental adjustment of earnings forecasts', *Journal of Forecasting*, 9 (1990), 83–105.
- Wright, G. and Ayton, P., 'The psychology of forecasting', in Wright, G. and Ayton, P. (eds), *Judgemental Forecasting*, Chichester: John Wiley, 1987, 83–105.
- Yates, J. F., 'Analyzing the accuracy of probability judgments for multiple events: an extension of the covariance decomposition', *Organizational Behavior and Human Decision Processes*, 42 (1988), 281–299.
- Yates, J. F., McDaniel, L. S. and Brown, E. S., 'Probabilistic forecasts of stock prices and earnings: the hazards of nascent expertise', *Organizational Behavior and Human Decision Processes*, 49 (1991), 60–79.
- Zarnovitz, V., 'Rational expectations and macroeconomic forecasts', *Journal of Business and Economic Statistics*, 3 (1985), 293–311.

Authors' biographies:

Dilek Önkal is an Assistant Professor of Decision Sciences at Bilkent University, Turkey. She received a Ph.D. in Decision Sciences from the University of Minnesota, and is doing research on decision analysis and probability forecasting.

Gülnur Muradoğlu is an Assistant Professor of Finance at Bilkent University, Turkey. She received a Ph.D. in Accounting and Finance from Bogazici University and is doing research on stock market efficiency and stock price forecasting.

Authors' addresses:

Gülnur Muradoğlu and **Dilek Önkal**, Faculty of Business Administration, Bilkent University, 06533 Ankara, Turkey.