Evaluating probabilistic forecasts of stock prices in a developing stock market

Dilek Önkal and Gülnur Muradoğlu

Faculty of Business Administration, Bilkent University, 06533 Ankara, Turkey

Abstract: Recent literature on the accuracy of forecasting in financial markets reveals contradictory results. These discrepancies can be attributed to the differences in forecasting environments as well as the differences in forecaster expertise that are employed by the researchers. Since the use of point and interval predictions by themselves do not aid in explaining the various aspects of forecaster performance, probabilistic forecasting provides a better alternative that can be used to gain insight into forecasting accuracy in such settings. This study aims to test the effects of forecaster expertise and forecasting environment on forecasting accuracy. Accordingly, various aspects of forecasting performance are studied in a developing stock-market framework.

Keywords: Forecasting; Probability; Financial; Stock price

1. Introduction

Recent literature on the accuracy of earnings per share and stock price forecasts reveals contradictory results. In a recent review on methods for predicting stock market prices, Granger [10] suggests that stock prices may be forecastable quantities. Liljeblom [14] also provides evidence regarding the predictive power of financial analysts' forecasts of earnings per share for short-term forecast horizons of less than nine months in the Scandinavian securities market. De Bondt [7], on the other hand, concludes that "... economists very much lack predictive power for direction and magnitude of stock market changes" and suggests that "... finance should attempt to model the behavior of representative investors and the nature of their errors" [7, p.90].

The above-mentioned studies use point predictions provided by forecasters. The use of point

Correspondence to: Prof. Dilek Önkal, Faculty of Business Administration, Bilkent University, 06533 Ankara, Turkey.

forecasts does not allow a detailed analysis of the various performance characteristics of forecasters. In order to be able to explain the nature of forecasting errors, a design framework that utilizes probabilistic forecasts may be considered more appropriate.

Probabilistic forecasting tasks require assessors to report subjective probabilities as expressions of their degrees of belief in the occurrences of future outcomes. Accordingly, probability forecasting offers several distinct advantages over nonprobabilistic forecasting (where only point or categorical forecasts are requested). First, probability forecasts present quantitative descriptions of the forecaster's uncertainty [20]. Second, they provide a basic channel of communication (via subjective probabilities) for the transmission of this uncertainty to the users of forecasts. Third, as a direct result of the preceding arguments, probability forecasts enable their potential users to make more informed decisions by providing them with information regarding the forecaster's degrees of belief. That is, the users not only have

point/categorical forecasts, but are also given measures of the uncertainty associated with the forecasts, which in turn should enable them to better 'interpret' the forecasts. From the forecaster's point of view, an important advantage is "... that it provides a means for forecasters to express their true judgments, and, thereby, to reduce or eliminate biases such as overforecasting" [6, p.111].

Assessments of probability distributions for future stock prices were first used by Bartos [4] and Stael von Holstein [22] with different purposes and techniques. Bartos [4] attempted to replace historical inputs in the mean–variance portfolio method developed by Markowitz [15] by density functions and cumulative distribution functions assessed by security analysts with a forecast horizon of one month. Stael von Holstein [22], on the other hand, tried to explore the effectiveness of feedback on forecasting performance by using probability assessments for five intervals rather than continuous distributions and a forecast lead time of two weeks. In both studies, however, uniform distributions generally outperformed the forecasters' distributions. It was argued that a proper choice of the definition of variables (such as 'price') and the choice of the decision horizon could potentially improve the experimental findings [23].

In a recent study, Yates, McDaniel and Brown [25] conducted an experiment to explore three fundamental issues: (1) the relationship between the overall accuracy of forecasts and the length of the forecast horizon, (2) the relationship between forecast accuracy and forecast difficulty (i.e., forecasting stock prices may be more difficult than forecasting earnings per share), (3) the existence of the inverse expertise effect (i.e., inverse relationship between expertise and forecast accuracy). Their results revealed that, even for a three-month forecast horizon, "... forecasts of stock prices were surprisingly inaccurate" [25, p.75]. Forecasts of earnings were found to be better than forecasts of stock prices. Finally, the existence of the inverse expertise effect was observed. Yates, McDaniel and Brown [25] study leads us to conclude that the accuracy of forecasting can be explained by the complexity of the decision environment and the decision parameters.

Current study aims to further investigate the research questions addressed by Yates, McDaniel and Brown [25] in an inherently less-complicated decision environment. The aim is to explore the determinants of forecasting accuracy and the potential existence of the inverse expertise effect within the framework of a developing economy with an emerging securities market. Contrary to the three-month forecasting period used in Yates, McDaniel and Brown's [25] work, our research focuses on a one-week forecasting horizon. The selection of a shorter forecasting period is based on the customary practices of the financial media and the inflationary structure of the economy in which this experiment took place. The choice of stock prices (in contrast to earnings per share) as the forecast items is based on the ease of accessibility to price information. Such information is available to non-professional investors on a daily basis through the mass media, whereas comparable information for earnings is limited in the mass media.

Accordingly, the setting of the experiment is portrayed in Section 2, while the experimental procedure is presented in Section 3. The various performance measures used to evaluate the probability assessment performance of the forecasters are reviewed in Section 4. Finally, the empirical results are given in Section 5 and conclusions are offered in Section 6.

2. The setting: An emerging securities market

Attempts for the liberalization of highly inefficient and strictly regulated financial markets started in Turkey at the beginning of the 1980's. Although establishment of the legal framework and regulatory agencies for the stock market were completed in 1982, Istanbul Securities Exchange (ISE), which is the only stock exchange in Turkey, was established in 1986. Until 1987, the employees of the stock exchange could hold stock portfolios without notification. It was in 1990 that a legislation against insider trading was passed for the first time. At the end of 1991, when this study was conducted, 123 stocks were traded at ISE and the annual volume of trade was approximately US$ 8.6 billion. Hence, the ISE provides a good exemplar of an emerging securities market.

Stock prices, when used as decision parame-

ters, are shown to yield inaccurate forecasts compared to earnings [25]. This may be attributed to the efficiency of the stock market in the United States. If the market is efficient, all the relevant information including price [8], public announcements [3], and even monopolistic information [13] is fully reflected in the stock prices so that no investor can beat the market continuously. Therefore, in efficient markets no investment method is assumed to be superior to the random selection of investment portfolios. (For extensive discussions of contradictory empirical evidence, please see [2], [9], and [10]). However, efficiency is not guaranteed in small, developing markets [5,12] where the market behavior may exhibit different characteristics [16].

The decision-making environment used in this study is an emerging stock market in a developing country and the only decision parameter employed is the stock price (which is assumed to be a somewhat less difficult quantity when compared to a developed-country setting). The relatively few number of stocks in the ISE in comparison to the ones in developed countries may be viewed as reducing the complexity of decision parameters for the decision maker – i.e., the investor. Also, there may be an increased awareness of sectoral performance and/or individual performance of stocks due to the smaller market size and volume. Hence, this study focuses on diminishing the complexity embodied in probabilistic forecasts in investment decisions by reducing the complexity of the decision making environment. Within such a framework, the goal is to explore the various dimensions of accuracy of stock price forecasts and to investigate the existence of a potential inverse expertise effect.

3. Procedure

All the subjects in the study were students recruited from the graduate and undergraduate classes at the Faculty of Business Administration of Bilkent University. Purpose of the study was briefly described in the announcements to students with the following inducements: (1) the opportunity to evaluate possible investment alternatives in a real stock market setting, (2) the opportunity to evaluate their forecasting skills,

and (3) the opportunity to learn about probabilistic forecasting. No other monetary or non-monetary bonuses were offered to the participants.

Of the 68 students who participated and completed the study, 25 were MBA students, 43 were undergraduates; 19 were exposed to probabilistic forecasting in decision analysis courses, and 48 had prior exposure to financial markets and financial forecasting in finance courses. There were 27 participants who had previous trading experience or were actively trading in the stock market with maximum duration of trading being 48 months. There were 7 students who had work experience in the stock market without being brokers or active traders. The resulting groups of 'novices' and 'semi-experts' were comparable in size and structure to Yates, McDaniel and Brown's [25] subject groups.

The task was to provide probability forecasts regarding the closing stock prices of 34 companies listed in the Istanbul Stock Exchange. The companies had been selected on the basis of their volume of trade on the preceding 52-week period and belonged to various sectors such as consumer goods, durables, utilities, pharmaceuticals, automotive, banking and insurance sectors. The choice was made deliberately to minimize task complexity as the stocks with the highest volume of trade could easily be followed through the media.

For each of the companies, subjects were asked to make forecasts regarding the weekly price changes of the associated stocks. That is, forecasts were to be made regarding the percentage change between the previous Friday's closing stock price and the closing stock price that would be realized on the current Friday. The selection of one week as forecast period may appear too short for forecasting in developed markets. However, in Turkey, where this experiment was conducted, the financial media report all evaluations as well as buy–hold recommendations on a weekly basis. This mainly stems from the high inflation rate, which results in a shorter decision-making horizon.

Subjects were requested to provide their forecasts in the form of subjective probabilities conveying their degree of belief in the actual price change falling into the designated percentage change categories. Specifically, subjects were asked to complete the following response form for each company:

WEEKLY PRICE CHANGE INTERVAL (in percentages, Friday to Friday)		PROBA-BILITY
(6)	INCREASE more than 10%	____ %
(5)	INCREASE 5–10%	____ %
(4)	INCREASE up to 5%	____ %
(3)	DECREASE 0–5%	____ %
(2)	DECREASE 5–10%	____ %
(1)	DECREASE more than 10%	____ %
		____ %

The range of stock price changes in the response form were prepared by considering the average weekly price changes of the composite stock index of the ISE during the previous 52-week period. For the past 52 weeks, weekly price changes were 3% on the average with the maximum increase being 8% and the maximum decrease being 5%. The first 5% increase range contained the average increase during the previous year, the second contained the maximum average weekly change, and the third range was designed for stocks whose volatility was higher than average. Intervals 1 through 3 were designed to be symmetric to intervals 4 through 6 for cognitive purposes. Although these intervals may seem too broad for developed markets, as witnessed in the Yates, McDaniel and Brown [25] study which used the same 5% increments for a three-month forecasting period, they are realistic for a highly volatile emerging market experiencing an annual inflation rate around 70%. The choice of our 5% intervals were also justified by the distribution of weekly price changes that actually occurred in the forecasting period. During the week of the forecasts, (1) 3% of the stocks' prices decreased more than 10%; (2) 3% of the stocks' prices decreased 5–10%; (3) 21% of the stocks' prices decreased 0–5%; (4) 24% of the stocks' prices increased up to 5%; (5) 15% of the stocks' prices increased 5–10%; and (6) 34% of the stocks' prices increased more than 10%.

In the beginning of the experiment, subjects were given detailed instructions describing the design and goals of the study. Definitions of 'subjective probability' and 'probability forecasting tasks' were stressed and examples were discussed. Subjects were told that certain scores of probability forecasting performance would be computed from their individual forecasts. They were informed that, due to the computational

characteristics of these scores (i.e., proper scoring rules), each subject could earn the best potential score by expressing his/her true opinion and thereby avoiding hedging of the forecasts.

Each subject was also presented with a folder that contained three separate forms: background forms, response sheets, and questionnaires. Background forms contained information on the name of the company, its industry, its net profits as of the end of the third quarter of 1991, earnings per share, and price–earnings ratio as of the last day of the preceding week. The folders also provided the weekly closing stock prices (i.e., the closing stock prices for each Friday) of the preceding 3 months (12 weeks) in tabular form as well as the weekly closing prices for the last 52 weeks in graphical form.

Response sheets were comprised of the response forms illustrated previously and instructions about the forecasting task. The questionnaire was to be filled after the forecasting task was finished. The instrument was designed to learn the participants' field of study, year in school, names of previous and current finance and decision analysis courses, previous and current experience and duration in the stock market, previous and current experience in trading and the duration. Subjects were also asked to delineate the sources they utilized in making their forecasts and to give a ranking with respect to the frequency of usage.

In order to duplicate real forecasting settings, subjects were allowed to take the background folders home. They were given the experimental material on Monday morning and were requested to submit the completed response sheets by Tuesday morning 9 a.m. before the opening of the session at the stock exchange. They were to complete their forecasts by the next day, since each additional day would give them more information about the current week's closing stock prices. They were also permitted to utilize any information source they would like in making their forecasts, excluding the other participants of the study.

4. Performance measures used

The evaluation of probability forecasts was made using the components of the probability

score for multiple events [24]. This scoring rule and its decomposition is outlined next.

We may define the forecast vector given by a subject for each stock as $f = (f_1, f_2, f_3, f_4, f_5, f_6)$, where f_k denotes a probability forecast that the stock's price change will fall into interval k, $k = 1, 2, \ldots, 6$. (Note that the intervals were defined in the previous section.) We may also define an outcome index vector $d = (d_1, d_2, d_3, d_4, d_5, d_6)$, where d_k takes on the value 1 if the realized price change falls within interval k, and takes on the value 0 if it does not fall within interval k.

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:

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

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 ($\overline{\text{PSM}}$) over a specified number of forecasting instances (i.e., over a given number of stocks).

The mean probability score has certain partitioning properties that facilitate the identification of various aspects of forecaster performance. Details of these decompositions have been provided elsewhere [17–19,24]. The performance measures utilized in this study present the most widely adopted subset of the evaluation criteria emanating from these decompositions [25]. The specific performance measures employed and their brief descriptions are given next.

4.1. 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 given over 100 predictions, a forecaster assesses a probability of 0.2 that the given stock's price will increase more than 10%. This forecaster's 0.2 assessments are well-calibrated if an increase of more than

10% was actually observed on 20 of the 100 predictions (i.e., 0.2 of the cases). 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 follows:

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

where $\bar{f}_k$ represents the mean probability forecast for interval k and $\bar{d}_k$ is the proportion correct for interval k. The lower the calibration score, the better calibrated is the forecaster.

4.2. Mean slope

Mean slope indexes another measure of forecaster performance that gives an indication of the forecaster's ability to discriminate between instances when the actual price change will and will not fall into the specified intervals. The higher the mean slope, the better the forecaster is able to discriminate. The computation of mean slope is as follows:

$$\text{Mean slope} = \frac{1}{K} \sum \text{Slope}_k = \frac{1}{K} \sum \left(\bar{f}_{1k} - \bar{f}_{0k} \right)$$

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.

4.3. 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 and will not fall into the specified intervals. Ideally, scatter would be zero since it is excessive variance due basically to the forecaster's reaction to nonpredictive environmental cues. A scatter index could be computed as follows:

$$\begin{aligned}
\text{Scatter} &= \sum \text{Scatter}_k \\
&= \sum \frac{1}{N} \left[\left(N_{1k} * \text{Var}(f_{1k}) \right) \right. \\
&\qquad \left. + \left(N_{0k} * \text{Var}(f_{0k}) \right) \right]
\end{aligned}$$

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 occur. As obvious,

$$N = N_{1k} + N_{0k}.$$

4.4. Forecast profile variance

Another criterion that can be used to evaluate forecaster performance is the forecast profile variance. This measure looks at the discrepancy between a forecaster's set of probabilities (i.e., $f = (f_1, f_2, f_3, f_4, f_5, f_6)$) and a uniform set of probabilities (i.e., $f = (0.167, 0.167, 0.167, 0.167, 0.167, 0.167)$). 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 could be computed as follows:

Forecast profile variance

$$= \frac{1}{N} \sum_n \left[\frac{1}{6} \sum_k (f_k - 0.167)^2 \right].$$

This index provides an opportunity to examine the profiles of probability forecasts from an 'across-interval variance' point of view.

4.5. Skill

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

$$\text{Skill} = \overline{\text{PSM}} - \sum \text{Var}(d_k)$$

$$= \overline{\text{PSM}} - \sum \left[(\bar{d}_k) * (1 - \bar{d}_k) \right]$$

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), $\sum \text{Var}(d_k)$ indexes an uncontrollable element of $\overline{\text{PSM}}$. Subtracting this 'base-rate' component from $\overline{\text{PSM}}$, we have the global effect of all the performance components that are under the forecaster's control. Hence, the lower the skill score, the better is the overall forecasting quality as displayed by the probability forecasts.

Table 1
Median values for various performance measures for active-traders (AT) and non-traders (NT), with corresponding measures for uniform (U), historical (H), and base-rate (B) forecasters

Performance measure	AT	NT [a]	U	H	B
$\overline{\text{PSM}}$	0.927	0.873 **	0.833	0.778	0.754
Calibration	0.078	0.077	0.079	0.024	0.000
Mean slope	0.004	0.022 ***	0.000	0.000	0.000
Scatter	0.089	0.074	0.000	0.000	0.000
F. profile variance	0.035	0.031 *	0.000	0.023	0.013
Skill	0.173	0.120 **	0.079	0.024	0.000

[a] *: p-value < 0.10;
 **: p-value < 0.05;
 ***: p-value < 0.02.

5. Results

The six performance measures used were the mean probability score ($\overline{\text{PSM}}$), calibration score, mean slope, scatter score, forecast profile variance, and the skill score. Table 1 shows the various scores attained by students involved in active trading and by students with no stock-market experience. To provide standards for comparison, same performance criteria are used to compute the scores that would be attained by a (1) uniform forecaster, (2) historical forecaster, and (3) base-rate forecaster.

A 'uniform forecaster' always provides 'equally likely' assessments to all the possible occurrences. Since our task involves six intervals, a uniform forecaster would give

$$f = (0.167, 0.167, 0.167, 0.167, 0.167, 0.167)$$

to all the stocks in question.

A 'historical forecaster' provides another standard for comparison since (s)he gives forecasts identical to the historical relative frequencies. Given the volatility of the stock market under consideration, we set the historical forecaster's probability forecasts equal to the relative frequencies realized in the previous week.

Finally, the 'base-rate 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 week.

A comparison of subjects with and without experience in the stock market reveals that non-traders are more accurate (i.e., have lower $\overline{\text{PSM}}$)

in general (p-value = 0.018). (Comparisons are made using the Mann–Whitney U test since the sampling distributions of the accuracy measures are unknown.) Their overall accuracy is better than that of the uniform forecaster, but worse than the historical and base-rate forecasters.

Subjects who are not actively involved in the stock market seem to achieve this superior accuracy mainly through their higher (i.e., better) mean slopes (p-value = 0.015). This indicates that the active traders' ability to discriminate between occasions when the actual price change will and will not fall into the specified intervals is worse than that of the non-traders. This finding can be linked to the forecast profile variances of the two groups. It can be hypothesized that the non-traders know their limits in discrimination better than the active-traders, and hence, give forecasts with smaller across-interval variances. That is, the forecast profile variances of non-traders more closely resemble that of a uniform forecaster. This is in fact observed from Table 1 (p-value = 0.089).

The overall effect of the mean slope is also observed in the skill scores of the two groups. Although there are no significant differences in the calibration and scatter of the two groups, the forecasting skills of non-traders are better (i.e., lower) than that of active-traders (p-value = 0.032).

More insight into these findings can be gained by looking at the correlations between the length of active trading and the performance measures. Table 2 shows that the overall accuracy deteriorates ($\overline{PSM}$ increases) as length of active trading is increased (p-value = 0.049). The primary contributing factor of this worsening of overall accu-

racy is the mean slope. As length of active trading goes up, the mean slope deteriorates (i.e., goes down; p-value = 0.025). This causes a degenerated (i.e., increased) skill score (p-value = 0.075). It is also interesting to note that an increase in the length of active trading is accompanied by an increased forecast profile variance (p-value = 0.042). This implies that, as subjects get more involved in active-trading, their forecasts deviate more from that of a uniform forecaster.

Comparisons of subjects who have and have not taken previous finance courses revealed no significant differences for any of the performance measures (all p-values > 0.10). Likewise, no significant differences could be found in the performances of undergraduate and graduate subjects. The only other significant differences were found in contrasting subjects who have been exposed to the 'subjective probability' concepts with the remaining subjects. It was observed that, familiarity with the subjective probability notion led to an increased scatter in the subjects' forecasts (p-value = 0.019). This could be due to the efforts of subjects with previous subjective probability training to disperse their uncertainty as much as possible, instead of concentrating it on few intervals. In fact, those subjects who have taken Decision Analysis courses commented that they tried to give probabilities to all the intervals, since this would better reflect the uncertainty they feel towards the highly volatile stock market. Evidently, their efforts resulted in highly scattered forecasts. The remaining accuracy dimensions showed no significant differences in performances of the two groups. That is, the mere knowledge of subjective probability concepts did not make any differences in skill, calibration, slope, forecast profile variance, and $\overline{PSM}$ attained by the subjects.

Table 2
Correlations between length of active trading and the performance measures

Performance Measure	Correlation [a]
$\overline{PSM}$	0.202 **
Calibration	0.059
Mean slope	−0.239 **
Scatter	0.050
F. profile variance	0.212 **
Skill	0.176 *

[a] * p-value < 0.10;
  ** p-value < 0.05.

6. Discussion and conclusion

This study was aimed at exploring the determinants of forecast accuracy and the existence of inverse expertise effect in forecasting stock prices in an inherently less complex environment. Confirming the conclusions reached by Stael von Holstein [22] and Yates, McDaniel and Brown [25], the so-called inverse expertise effect was observed. Inverse expertise effect was also found to

be positively correlated with the length of active-trading.

The results of the present study reveal that superior overall accuracy (as reflected by $\overline{PSM}$) of non-traders is mainly due to their ability to better discriminate between occasions when the actual price change will and will not fall into the specified intervals (i.e., their mean slope scores). Smaller across-interval variance (profile variance) and skill scores of non-traders also support the explanation that non-traders convey their forecasts using more dispersed distributions.

Correlations between the length of active trading and the performance measures also provide support for the existence of the inverse-expertise effect by showing that as the length of active trading increases, all the performance measures deteriorate (p-value < 0.10).

Previous research has utilized two main arguments for expert reasoning in such tasks. First, the more abstract representation of the problem situation by experts has been thought to lead novices to outperform the experts [1]. Secondly, experts' representations may be viewed to be richer than novices' representations [21], with additional cues making the judgment task more difficult or their misuse potentially distorting the accuracy of judgments [25]. We prefer to explain the findings of this research by overconfidence of experts relative to that of non-experts. Overconfidence of experts is revealed through less dispersed distributions of expected price changes. There were some experts in the study who forecasted the price changes of the stocks they followed with 100% certainty systematically. The only group of subjects among the graduates and the undergraduates that had a significantly different performance score were the subjects who were exposed to subjective probability concepts previously in Decision Analysis courses. The significantly higher scatter scores observed in those subjects could be due to their previous training to disperse uncertainty through intervals instead of concentrating on a few intervals. In the post-experimental interviews, these subjects presented their strategy, in fact, as one that gives probabilities to all intervals to reflect the uncertainty inherent in the stock market.

The financial community does not use probabilistic forecasts to represent predictions. Financial forecasts are frequently reported as point estimates or forecasts of ranges which do not reveal how firmly the forecaster believes in his/her expectations. This study suggests that the overconfidence of experts may mislead the users of non-probabilistic financial forecasts. Using probabilistic forecasts, on the other hand, it is possible to incorporate the uncertainty inherent in the stock market into forecasts and the degree of uncertainty may be represented in the form of probability distributions, with which the users of financial information are already familiar.

The avenues for further research suggested by the results of this study are mainly two-fold: first, the effects of feedback on probabilistic forecasts must be explored to test whether forecast accuracy can be improved via the learning of subjective probability concepts and the experience gained with the repeated use of these concepts. Secondly, the performance of portfolio models must be tested by comparing historical data versus subjective forecasts as inputs. We believe that probability forecasting provides an important channel of communication between the forecasters and the users of these forecasts. Hence, future research into the use of probability forecasting in financial settings will be of great benefit to all providers and users of financial information.

Acknowledgements

References

[1] 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.

[2] Ball, R., "What do we know about stock market efficiency", in: R.M.C. Guimares, B.G. Kingsman and S.J. Taylor (eds.), A Reappraisal of the Efficiency of Financial Markets, Springer-Verlag, Heidelberg; 1989, 25–55.

[3] Ball, R., and Brown, P., "An empirical evaluation of accounting income numbers", Journal of Accounting Research 6 (1968) 159–178.

[4] Bartos, J.A., "The assessment of probability distributions for future security prices", Unpublished Ph.D. Thesis, Graduate School of Business, Indiana University, 1969.

[5] Booth, G.G., Martikainen, T., Sarkar, S., Virtanen, I., and Yli-Olli, P., "Nonlinear dependence in Finnish stock returns", Paper presented at the EURO-TIMS Meeting, Helsinki, 1992.

[6] Daan, H., and Murphy, A.H., "Subjective probability forecasting in the Netherlands: Some operational and experimental results", Meteorologische Rundschau 35 (1982) 99–112.

[7] De Bondt, W.F.M., "What do economists know about the stock market?", Journal of Portfolio Management 4 (1991) 84–91.

[8] Fama, E.F., "The behavior of stock market prices", Journal of Business 38 (1965) 34–105.

[9] Foster, G., Olsen, C., and Shevlin, T., "Earnings releases, anomalies and the behavior of security returns", Accounting Review 59 (1984) 574–603.

[10] Granger, C.W.J., "Forecasting stock market prices: lessons for forecasters", International Journal of Forecasting 8 (1992) 3–13.

[11] Hammond, K.R., "Probabilistic functionalism: Egon Brunswick's integration of the history, theory and method of psychology", in: K.R. Hammond (ed.), The Psychology of Egon Brunswick, Holt, Rinehart & Winston, New York, 1966.

[12] Hawawini, G.A., and Michel, P.A., European Equity Markets: Risk, Return and Efficiency, Garland, New York, 1984.

[13] Jensen, M.C. "The performance of mutual funds in the period 1945–64", The Journal of Finance 2 (1968) 389–416.

[14] Liljeblom, E., "An analysis of earnings per share forecasts for stocks listed on the Stockholm Stock Exchange", Scandinavian Journal of Economics (1989) 565–581.

[15] Markowitz, H., Portfolio Selection: Efficient Diversification of Investments, Wiley, New York, 1959.

[16] Martikainen, T., and Perttunen, J., "Return intervals, systematic risk estimates and firm size: Empirical evidence from a thin security market", Economics Letters 36 (1991) 311–315.

[17] Murphy, A.H., "Scalar and vector partitions of the probability score: Part I. Two-state situation", Journal of Applied Meteorology 11 (1972) 273–282.

[18] Murphy, A.H., "Scalar and vector partitions of the probability score: Part II. N-state situation", Journal of Applied Meteorology 11 (1972) 1183–1192.

[19] Murphy, A.H., "A new vector partition of the probability score", Journal of Applied Meteorology 12 (1973) 595–600.

[20] 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.

[21] 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.

[22] Stael von Holstein, C.A.S., "Probabilistic forecasting: An experiment related to the stock market", Organizational Behavior and Human Performance 8 (1972) 139–158.

[23] Winkler, R.L., "The assessment of probability distributions for future security prices", in: J.L. Bicksler (ed.), Methodology in Finance-Investments, 1972.

[24] Yates, J.F., "Analyzing the accuracy of probability judgments for multiple events: An extension of the covariance decomposition", Organizational Behavior and Human Performance 30 (1988) 132–156.

[25] 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 40 (1991) 60–79.