EFFECTS OF FEEDBACK ON FINANCIAL FORECASTING

A THESIS

Serra DİRİMTEKİN

SEPTEMBER 1996

EFFECTS OF FEEDBACK

ON

FINANCIAL FORECASTING

A THESIS

SUBMITTED TO THE DEPARTMENT OF MANAGEMENT

AND

THE GRADUATE SCHOOL OF BUSINESS ADMINISTRATION

OF

BİLKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF MASTER OF BUSINESS ADMINISTRATION

BY

SERRA DİRİMTEKİN

September, 1996

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Assoc. Prof. Dilek ÖNKAL

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Assist. Prof. Can Şımga MUGAN

I certify that I have read this thesis and in my opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Business Administration.

Assist. Prof. Ayşe YÜCE

Approved for the Graduate School of Business Administration

Prof. Sübidey TOGAN

ABSTRACT

EFFECTS OF FEEDBACK

ON

FINANCIAL FORECASTING

SERRA DİRİMTEKİN

Master of Business Administration

Supervisor: Assoc. Prof. Dilek ÖNKAL

September, 1996

56 pages

The objective of this study is to examine the effects of feedback on financial forecasting. In particular, the effects of simple outcome feedback and calibration feedback as a type of performance feedback on the accuracy of probabilistic forecasts of stock prices and market indices in dichotomous format are analyzed. The study is conducted on subjects comprised of undergraduate and graduate students from the Faculty of Business Administration at Bilkent University. The results indicate that feedback, especially calibration feedback, has a considerable effect on the performance of forecasters. Implications of these findings for financial forecasting are discussed and directions for future research are given.

Key Words: Judgment, judgmental forecasting, probabilistic forecasting, stock price forecasting, financial forecasting, feedback, calibration feedback.

ÖZET

FİNANSAL TAHMİNLERİNDE GERİ BESLEMENİN ETKİSİ

SERRA DİRİMTEKİN

Yüksek Lisans Tezi, İşletme Enstitüsü

Tez Yöneticisi: Doç. Dr. Dilek ÖNKAL

Eylül, 1996

56 sayfa

Bu çalışmanın amacı, finansal tahminlerde geri beslemenin etkisini incelemektir. Bu baglamda, basit sonuç geri beslemesi ile başarı geri beslemesinin bir çeşidi olan ayar geri beslemesinin, hisse senedi fiyatlarının ve borsa endekslerinin iki sonuçlu format şeklindeki olasılıksal tahminleri üzerindeki etkisi incelenmiştir. Çalışma, Bilkent Üniversitesi İşletme Fakültesi lisans ve lisansüstü ögrencilerinden oluşan bir gruba uygulanmıştır. Sonuçlar, geri beslemenin; özellikle ayar geri beslemesinin tahminde bulunanlar üzerinde önemli etkisi oldugunu göstermiştir. Finansal tahminlerle ilgili sonuçlar tartışılmış ve gelecek çalışmalar için konular önerilmiştir.

Anahtar Kelimeler: olasılıksal tahmin, hisse senedi fiyat tahmini, finansal tahmin, geri besleme, ayar geri beslemesi.

ACKNOWLEDGEMENTS

This thesis has benefited greatly from the supervision and the contribution of Assoc. Prof. Dilek Önkal.

I am grateful to my colleagues for their participation to the thesis, and my friends for their support during the preparation of this thesis.

I owe special thanks to my parents for their life time support and encouragement.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Judgment in Forecasting

Judgment has been studied for many years by psychologists interested in human decision-making (Wright and Ayton, 1987). The research was undertaken from the perspective of *subjective expected utility theory* -decision theory. Decision theory depends on statistics and economics and proposes that two independent types of information are important in making good decisions: *subjective probabilities* attached to events occurring and *subjective values* or *utilities* attached to the outcomes of those events in the future.

Judgment plays a major role in the forecasting process. This role was emphasized in the studies of Batchelor and Dua (1990), Bunn and Wright (1991), Flores, Olson and Wolfe (1992), Goodwin and Wright (1991), Philips (1987), Turner (1990), Wolfe and Flores (1990), Zarnowitz and Lambros (1987). McNees (1990) observed that, with some significant exceptions, experts' judgmental adjustments of economic forecasts generated by models improved accuracy. Clemen and Murphy (1986) found out that weather forecasters have an advantage over model forecasters for short lead times; the former are able to adopt more easily to rapidly changing conditions. Yaniv and Hogarth (1993) proposed that given their different strengths, human and statistical predictions can be profitably combined to improve prediction.

1.1.1. Statistical Techniques versus Judgmental Forecasting

There are two reasons why human judgment might be better than statistical forecasting models in times of change (Remus, O'Connor and Griggs, 1995). Human judgment could be superior to the forecasting models in recognizing changes in the pattern of the data or it might be able to better integrate outside information about the change into the forecasting process.

Managers feel more comfortable dealing with their own or colleagues' estimates than with statistical models. The use of judgment in forecasting has been supported by both field and laboratory studies. Lawrence, Edmundson and O'Connor (1985) found that partly structured eyeballing by unsophisticated subjects was as accurate as the best statistical models. The variance of the forecast errors was significantly less using human judgment than when using statistical models.

The statistical techniques used for forecasting require a series of historical data. However, it may be hard to find such data; for instance, forecasting the sales of a new product. Then the manager can apply the concept of probability based on subjective judgments rather than historical frequencies. Nevertheless, Makridakis and Wheelwright (1979) noted that "forecasters tend to concentrate on well-behaved situations that can be forecasted with standard methodologies and to ignore the rapidly changing situation for which management may most want forecasts" (p. 339).

Other researchers argued that judgmental forecasts are used when there is insufficient time to obtain and use a statistical forecast or when situations are changing so rapidly that a statistically based forecast would be no use. Makridakis and Wheelwright (1979) concluded that "application of quantitative approaches will continue to increase and replace many of the applications now handled, through purely judgmental approaches" (p. 348).

However, Makridakis and Wheelwright (1979) also note that, "Of course it must be remembered that just as it is impossible to say which methodology is the best, it is always impossible to conclude that quantitative methods are always better than subjective or judgmentally based methods. Human forecasters can process much more information than most of the formalized quantitative methods, and such forecasters are more likely to have knowledge of specific near-term events that need to be reflected in current forecasts" (p. 348).

Additional studies are required in forecasting, since the generalizability of results from general-knowledge tasks to forecasting tasks is questionable. There exists a large amount of evidence that *overconfidence* is a prevalent feature of human intuitive judgment (Kahneman, Slovic and Tversky, 1992). For example, if people are given a general knowledge test and asked to estimate the likelihood that their answers are correct, then their estimates are consistently overconfident when compared with the objective probability of success. This overconfidence in intuitive judgment applies equally to judgments about future events, i.e., forecasts.

Fischhoff and MacGregor (1982) argued that the results from studies using almanac questions are generalizable. They asked the subjects to predict events that would be completed within 30 days of the experiment, e.g., results of local elections and popular sporting events. The proportion of correct predictions was 0.618, whereas the mean confidence in predictions was 0.722. However, Wright and Ayton (1986) and Ronis and Yates (1987) disputed their arguments.

One would expect people to learn from mistakes made in the past and realize their limitations as forecasters. In fact, related research reveals that people are quite poor at learning from past mistakes and display a phenomenon known as 'knew-it-all-along-effect' (Fischoff, 1982). It was demonstrated in a number of studies that people will improve their estimates if they are provided with outcome knowledge.

1.2. The Role of Feedback in Probability Assessment

The role of feedback in probability assessment tasks was emphasized in some studies. Hogarth (1975), in his study on subjective probability assessments and related cognitive processes, pointed out that "..substantive experts can make meaningful assessments in situations where they make forecasts over a period of trials and receive feedback as to the accuracy of their judgments" (p. 278). Moriarity (1985) studied the provision of feedback regarding the correspondence of forecasts with actual occurrences as an important design characteristic of forecasting systems that involve management judgment.

In spite of the emphasis on feedback in forecasting, not many empirical studies were conducted. Fischer (1982) suggested that outcome feedback is ineffective in improving the overall accuracy of probability forecasts. Outcome feedback is the information about the realization of a previously predicted event. Following Fischer's suggestion, studies tackled with scoring-rule feedback and calibration feedback.

Scoring-rule assigns an overall score to a forecaster based on a function of the forecaster's reported probability forecasts and the outcomes that actually occur computed over a set of probability forecasts (Winkler, 1969; Friedman, 1983). Staël von Holstein (1972) performed an experiment concerning the stock market. He focused on the accuracy of stock price predictions. For each of 12 stocks, subjects (bankers, stock market experts, teachers, statisticians, and students) made probabilistic forecasts that price changes over successive 2-week periods would fall into five specified intervals that partitioned the continuum. His primary aim was training. Every two week, he gave his subjects scoring rule feedback about their accuracy. However, all the training was found to be ineffective. Fischer (1983) also concluded that the provision of scores from such rules had no effect on the performances of their forecasters; Kidd (1973) showed that scoring-rule could be effective in improving forecasters' accuracy levels.

1.3. Calibration Feedback

Under a frequentist interpretation of 'probability', a probability assessment is said to be 'good' if the assigned probability equals (in the long run) the relative frequency of occurrence (O'Connor, 1989). Thus, if a probability of 0.6 is assigned to each of 100 independent events, that assessment is 'good' if the event occurs on 60 occasions. This does not mean that the 'goodness' of any single event can be determined, only the assessment of many events. This interpretation of 'goodness' is termed *calibration*. Specifically, O'Connor calls a person 'perfectly calibrated' if the proportion of true events is equal to the designated probability, in the long run.

Calibration feedback involves giving forecasters information about their ability the assign appropriate probabilities to future outcomes. A forecaster is said to be *well calibrated* if for all predicted outcomes assigned a given probability, the probability of those that occur (proportion correct) is equal to the probability that is assigned by the subject (Önkal and Muradoglu, 1995). For example, if it actually rained on 40% of the days that a weather forecaster predicts a 0.4 chance of rain, the forecaster's 0.4 probability forecasts are well calibrated. Calibration feedback has not yet been standardized. It may consist numerical summaries and/or graphical displays of the reported probabilities, the proportion correct (the proportion of the outcomes that occur) associated with each probability value, and the number of assessments of each value (Benson and Önkal, 1992).

Calibration feedback is a promising means of improving the performance of probability forecasters. Murphy and Daan (1984) and Murphy and Brown (1985) found both individualized and group calibration feedback to be effective in field studies of weather forecasters even though only one feedback session was employed.

The official forecasts issued by the National Weather Service in the United States are subjective probability forecasts. Murphy and Brown (1985) evaluated these subjective forecasts and found that, for certain predicted categories of weather, they were more accurate than the available objective statistical techniques. In this case the forecasters have a very large amount of information available, including the output from statistical techniques. They also receive detailed feedback and have the opportunity to gain experience of making forecasts under wide range of meteorological conditions. Furthermore, they have considerable practice in quantifying their internal state of uncertainty. These circumstances may well be ideal for the relatively successful application of judgmental, as compared with quantitative, forecasting. They are certainly not the conditions available in most situations where judgment is obtained and utilized.

Benson and Önkal (1992) concluded that the provision of calibration feedback was effective in improving both the calibration and the overforecasting of probabilities of the forecasters, but the improvement was not progressive; it occurred in one step, between the second and third sessions. Simple outcome feedback had very little effect on forecasting performance. Unlike outcome feedback, the provision of performance feedback caused subjects to manage their use of probability scale. Subjects switched from two-digit probabilities to one-digit probabilities and those receiving calibration feedback also reduced the number of different probabilities they used.

The provision of frequent feedback would improve calibration (O'Connor, 1989). Experts in horse racing and weather forecasting are well calibrated because immediate feedback is provided for them to immediately assess the 'goodness' of their estimates. For those who are unfamiliar with a topic, training via extensive feedback will improve calibration.

Extensive reviews by Lichtenstein, Fischhoff and Philips (1982), Fischhoff and MacGregor (1982) and Wright and Ayton (1989) concluded that people are typically overconfident in their judgment and predictions. O'Connor (1989) suggested that people adjust their calibration to meet the demands of the task and its context.

1.3.1. The Conditions Under Which Good Calibration Can Be Expected

No definite answers can be found in research to date, but several conditions can be identified in the studies of good calibration (Philips, 1987). First, Wright and Ayton (1986) concluded that calibration provided better results for future events than for general-knowledge questions. Second, most of the studies showing good calibration were done with experts. Lichtenstein, Fischhoff and Philips (1982) conducted an experiment using general-knowledge questions and figured out that there existed no difference in calibration between experts and novices. However, no studies comparing the calibration of experts with that of non-experts were done using future-event questions.

Third, several studies were conducted with groups of assessors. Philips (1987) obtained probability assessments from various groups of people who had differing perspectives on the certain quantity or event in question. For all of these groups, individuals used their own experience to influence others. In general, the practitioner has the 'hands on' experience that makes the assessment process meaningful, the researcher with field experience extends the practitioner's knowledge, while the scientist (who is sometimes reluctant to assess probabilities) identifies and questions assumptions that others may be making.

Fourth, Lichtenstein and Fischhoff (1980) showed that feedback improves calibration, and that most improvement occurs in the first few training sessions. General knowledge questions were given to the forecasters; but extensive feedback via training was provided over 11 sessions. Weather forecasters in the Netherlands began making probability forecasts in October 1980, and by the end of the second year, calibration had improved substantially. Murphy and Daan (1984) attribute this to feedback given to the forecasters in October 1981 about their calibration during the first year, and to experience in probability forecasting gained during the first year.

1.4. Financial Forecasting

It is still being questioned how to harmonize judgment with financial decision-making process. The use of subjective probabilities opens the door for an answer. Probability forecasts supply efficient channels of communication between the providers and the users of financial information, considering the quantitative measures of uncertainty (Önkal and Muradoglu, 1996).

Bartos (1969) and Staël von Holstein (1972) were the first ones using subjective probability distributions. In both studies, uniform distributions outperformed the forecasters' distributions. In the studies of Yates, McDaniel and Brown (1991) and Önkal and Muradoglu (1994) probabilistic forecasts of stock prices displayed low levels of accuracy. Furthermore, historical forecasters (giving forecasts identical to the historical relative frequencies) outperformed the participants' probabilistic forecasts.

Stock price forecasts in the USA were shown to be relatively inaccurate when compared to earnings forecasts (Yates, McDaniel and Brown, 1991). This may be due to the efficiency of the stock market in US. If the market is efficient, all relevant information including knowledge of previous prices (Fama, 1965), public announcements (Ball and Brown, 1968) and even monopolistic information (Jensen, 1968) is fully reflected by the stock prices, so that no investor can beat the market continuously.

1.5. An Overview on Stocks and Stock Prices

Corporations use separate owners' equity accounts (Capital Stock and Retained Earnings) to represent (1) the capital invested by the stockholders (called *paid-in capital*) and (2) the capital acquired and retained through profitable operations (*earned capital*). All paid-in capital may be recorded in a single ledger account entitled Capital Stock. A corporation may issue several different types of capital stock.

Ordinary shares represent equal ownership in a corporation embodying such rights as the receipts of dividends subscription to bonus and rights issues and the liquidation of assets, including voting rights. Almost all shares quoted on the Istanbul Stock Exchange belong to this category. Preferred shares carry preferential rights as to voting rights or dividends in contrast to ordinary shares. In the founders' shares, the owner has special benefits in case of distribution of profits.

The articles of incorporation specify the number of shares of each type of capital stock which a corporation is authorized to issue and the *par value*, if any, per share. Large issues of capital stock to be offered for sale to the general public must be approved by the Securities Exchange Commission (SEC) as well as by the state officials.

Par value (or stated value) represents the legal capital per share -the amount below which stockholders' equity cannot be reduced except by losses from business operations. It can be regarded as a minimum cushion of equity capital existing for the protection of creditors.

If the stock is issued in exchange for other assets other than cash, the transaction is recorded at either the fair market value of the shares issued or the fair market value of the assets received, whichever can be determined more objectively.

Because the equity of each stockholder in a corporation is determined by the number of shares he or she owns, an accounting measurement of interest to many stockholders is *book value* per share of common stock. It is equal to the net assets (total assets minus total liabilities) represented by one share of stock. To some extent book value is used in evaluating the reasonableness of the *market price* of a stock.

Market value is the current price at which shares of stock may be bought or sold. When a stock is traded on an organized stock exchange, the market is quoted daily in the financial press. Market price is based upon a combination of factors, including investors' expectations of future earnings, dividend yield, interest rates, and alternative investment opportunities (Meigs et al. 1992).

1.5.1. Stock Market in Turkey

Securities trading in Turkey date back to the Crimean War in the middle of the 19th century. The first securities market was established immediately after the Crimean War under the name of the "Imperial Securities Exchange" in 1866 when the Ottoman sultan issued sovereign bonds to finance the war campaign. The Turkish and foreign securities were traded by means of telegram connections with the European stock exchanges. Although this bourse emerged as one of the leading financial centers in Europe, the market fell victim to a succession of wars. After the Turkish Republic was proclaimed in 1923, a new attempt was made to launch a stock exchange. However, this effort was averted by the Depression. After the Depression, as the pace of change in the political environment gained momentum throughout the world, the number of joint stock companies rose sharply. The environment was already matured for a revival of a stock market as far-reaching and extensive economic measures were exposed in 1980. In 1981, the Capital Market Board (CMB) was established. Subsequently, the "General Regulations" for the exchanges were legislated, and in 1986, the Istanbul Stock Exchange (ISE) was opened.

1.5.2. Istanbul Stock Exchange (ISE)

The ISE is a semi-professional organization. Its revenues come from the fees charged for the transactions, the listing procedures and miscellaneous services. The profit of the stock exchange is retained to meet future expenses and investments and is not distributed to any third parties. The ISE provides markets for trading the following instruments to their members; stocks and right coupons, 'A type' mutual funds, treasury bills, government bonds, repo/reverse repo transactions, corporate bonds and revenue sharing certificates.

There are three categories of members in the ISE. They are banks which are investment and development banks, commercial banks and non-bank intermediary institutions which are brokerage houses. All of the ISE members are allowed to trade for their own account. As of 1995, the ISE had a total of 165 members: 11 investment and development banks, 50 commercial banks and 104 brokerage houses.

Beginning from 1994, the stock market was divided into Regional Stock Market and a National Stock Market. In Regional Stock Market 12 companies' shares are traded. Whereas, in the National Stock Market, there are 196 companies.

The ISE was computing and publishing a stock price index (the ISE index) as a comprehensive measure of the market's performance since its introduction in January 1986. This index was weighted by market value. However, since the beginning of 1991, the ISE restructured its existing index with minor changes in the method applied in calculating the index and two new sub-indices were introduced. The new index was called the 'The ISE Composite Index'. Composite index is weighted by the proportion of the product of the company's number of stocks, multiplied by the market price of the stocks offered to the public. Therefore, any price change in the stocks of companies in the First Market with a large market value and widely held by the public will have greater impact on the Composite Index.

According to previous studies, financial markets in Turkey were found to be inefficient and strictly regulated until 1980. Attempts to liberalize financial markets started at the beginning of the 1980s with the introduction of a liberalization package encouraged by the World Bank and the International Monetary Fund. Establishment of the legal framework and regulatory agencies for the stock market were completed in 1982, but in 1986 the ISE, the only stock exchange in Turkey was established (Önkal and Muradoglu, 1996). Turkish Stock Exchange has been attracting attention since its establishment. With its growing trading volume, it has got an important place in the international stock exchange markets.

1.5.3. Effect of Market Efficiency on Stock Price Forecasting

Roberts (1967) defined three levels of market efficiency according to the judgments of these researchers. The first is the case in which prices reflect all information contained in the record of past prices; called as the *weak form of efficiency.* The second level of efficiency is the case in which prices reflect not only past prices but all other published information; called as the *semi strong form of efficiency.* Finally, *strong form of efficiency* is the case in which prices effectively impound all available information.

The efficient-market hypothesis is frequently misinterpreted. One common error is to think it implies perfect forecasting ability. In fact, it implies only that prices reflect all available information (Brealey and Myers, 1991). Therefore, in efficient markets, no investment method is assumed to be superior to the random selection of investment portfolios (Önkal and Muradoglu, 1996).

1.5.4. The Place of the ISE in Stock Price Forecasting

The ISE serves as a better medium than a developed market for predicting stock prices due to the inefficiency of the market. The ISE is known to be weak form (Muradoglu and Oktay, 1993; Muradoglu and Ünal, 1993) and semi-strong form (Muradoglu and Önkal, 1992) efficient. What is more, since the ISE contains fewer number of stocks than the exchanges in the developed countries, the investor will cope with less complexity. In the ISE, there may be a potential for improving stock price forecasting performance (Önkal and Muradoglu, 1995). In this study, the objective is to determine if feedback can achieve this potential.

1.6. An Overview on the Study

In this study, the effects of outcome and calibration feedback on the accuracy of probabilistic forecasts regarding stock prices are examined. The experimental framework of Yates, McDaniel and Brown (1991) is taken as a basis. In their study, undergraduate and graduate students in finance courses made probabilistic forecasts of the quarterly changes in the stock prices and earnings of publicly traded companies. They aimed to re-examine previous results (Staël von Holstein, 1972) on accuracy of probability judgments on stocks, and test the existence of an inverse relationship between expertise and accuracy. The overall accuracy of both price and earnings forecasts was very modest. Also, undergraduate subjects were more accurate than graduate subjects, implying an inverse-expertise effect.

Following Önkal and Muradoglu (1994), Yates, McDaniel and Brown's (1991) procedure is adapted to the Turkish stock market and extended to examine the effects of feedback on probabilistic forecasts of stock prices. In this study two types of feedback are put to use:

(1) simple outcome feedback,

(2) performance feedback in the form of calibration feedback.

This thesis is organized as follows: In Chapter 2, the procedure of the study is presented. In Chapter 3, the performance measures used in measuring the accuracy of probabilistic forecasting of stock prices are discussed. Chapter 4 presents findings and Chapter 5 offers some concluding comments.

2. PROCEDURE

Subjects of the study were recruited from graduate and undergraduate classes from the Faculty of Business Administration of Bilkent University. The purpose of the study was described in preposted announcements. Subjects participated in this study on a voluntary basis. No monetary nor non-monetary bonuses were offered apart from the opportunity to evaluate possible investment alternatives in a real stock market setting and improve probabilistic forecasting skills.

The subjects were randomly assigned to two feedback groups:

(1) simple outcome feedback group (control group)

(2) calibration feedback group.

Feedback groups consisted of 14 and 17 subjects respectively. A total of 31 subjects completed the three-week-long experiment.

The experiment involved three weekly forecasting sessions and the task was to provide probability forecasts of closing stock prices of 30 companies listed in the ISE and 6 market indices -for a general overview (Appendix 1). The choice of stocks was made among the stocks that are included in the ISE composite index, since subjects are expected to make probabilistic forecasts also on the ISE composite index, in addition to five foreign stock exchange indices that are presumed to be better known. The data is gathered from the ISE Weekly Bulletin and the ISE itself.

Subjects were asked to make forecasts regarding the weekly price changes for each of 30 stocks and 6 market indices using a dichotomous format. The name of the stocks and the market indices were not provided for the subjects.

2.1. Response Sheets

Forecasts with the dichotomous format required the forecaster to state whether he/she believed the closing price for the current Friday would (a)increase, or (b)decrease/or stay the same with respect to the previous Friday's closing stock price. Then they were asked to state their degree of belief with a subjective probability for the forecasted direction of price change. They were asked to complete the following response form for each stock:

WHEN COMPARED TO THE PREVIOUS FRIDAY'S CLOSING STOCK PRICE, THIS FRIDAY'S CLOSING PRICE WILL

A. INCREASE
B. STAY THE SAME or DECREASE

YOUR FORECAST (A or B) :
PROBABILITY THAT YOUR FORECAST
WILL INDEED OCCUR
(I.E., PROBABILITY THAT THE WEEKLY
PRICE CHANGE WILL ACTUALLY FALL
IN THE DIRECTION YOU PREDICTED)
(BETWEEN 50% and 100%) : ....

It is preferred to use dichotomous format in the forecasts throughout the study, but not multiple interval format, because the period that the study was conducted, was very volatile due to the instabilities in the economy and upcoming elections. This way, subjects could make more proper forecasts. Furthermore, the dichotomous scale may be viewed as providing a preferable medium of representation for expressing forecasts based on the limited knowledge of novices, supporting the argument of Murphy and Wright (1984), that rich presentations (e.g. multiple-interval scale) are a function of the level of expertise (Önkal and Muradoglu, 1996).

At the beginning of the first session, all subjects were given detailed information about the design and goals of the study. Afterward, they were presented with folders containing response sheets illustrated previously (see Appendix 2 also) and instructions about the forecasting task. Folders provided graphical plots of the weekly closing prices for each Friday from October 1994 until December 1995 and the preceding 15 weeks' data in tabular form. Graphs were used, since figures are more meaningful for observing changes in prices.

Both groups were provided with the same data sets. This supported consistency across the judgmental forecasts, since research shows that judgmental accuracy depends on the method of data presentation (Angus-Leppan and Fatseas, 1986).

Participants were told that certain scores of probability forecasting performance would be computed from their individual forecasts and their performance would be reported on a personal basis.

To duplicate real forecasting settings, the subjects were allowed to take the folders home. They were given the folders on Mondays and expected to bring them back with their forecasts on Tuesdays, so that they could observe only Monday closings, and be less affected for forecasting Friday closing prices. They were allowed to utilize any information source they preferred, other than other participants of the study.

After the folders had been collected from the subjects in the first week, their predicted outcomes were analyzed in Minitab and their performance measures; mean probability score, calibration, scatter, slope and bias scores were computed. In the second and third sessions, control group (simple outcome-feedback group) was provided simple outcome feedback only, while the other group was additionally given calibration feedback derived from their previous forecasts, with an explanation of how they would interpret that score.

2.2. Feedback

2.2.1. Simple Outcome Feedback Group

This group served as a control group for the experiment. They received previous Friday's closing price marked in their graphical and tabular information for each of the 30 stocks and 6 market indices. The ready-made format helped the simple outcome feedback group decrease their perceived task difficulty with respect to the calibration feedback group.

2.2.2. Calibration Feedback Group

Subjects in this group received feedback given to the simple outcome feedback group and their calibration scored computed from the previous week's forecasts.

Calibration is the most widely used performance criterion (Lichtenstein, Fischhoff and Philips, 1982). Calibration provides information about the forecaster's ability to assign appropriate probabilities to outcomes. Computational formula will be explained in detail in the next chapter, which provides a review of the performance measures used in this study.

3. PERFORMANCE MEASURES UTILIZED

When probabilistic forecasts are expressed in dichotomous format, there are two possible codings that can be utilized (Önkal and Muradoglu, 1996). The first coding, *external coding,* involves deriving forecasts for a given target event (e.g. stock price increases). These forecasts are then evaluated with the use of an *outcome index* that is defined with respect to the occurrence of the prespecified target event. The second coding, *internal coding,* requires that the forecaster first chooses one of the two possible outcomes and then assesses the probability that his/her predicted outcome will occur. This is the type of coding employed in this study. These forecasts are then evaluated with the use of an outcome index that is defined with respect to the occurrence of the predicted outcome. Ronis and Yates (1987) discussed that their interpretation vary substantially, even though the codings share the same performance measures.

3.1. Mean Probability Score

The dichotomous format requires the forecaster to first choose from two outcomes (i.e., whether the stock price will (a) increase, or (b) decrease or stay the same). Then he/she is requested to state his/her degree of belief in the occurrence of the chosen outcome by assessing subjective probabilities associated with the forecasted direction of price change.

F_i denotes the forecaster's probability that his/her chosen outcome will occur for stock i. Correspondingly, $0.5 \leq F_i \leq 1.0$.

D_i denotes the outcome index, assuming a value of 1 if the chosen outcome indeed occurs for stock i, and takes a value of 0 if the chosen outcome does not occur for stock i.

Hence, PS_i denotes the probability score for stock i ; $PS_i = (F_i - D_i)^2$

The mean of probability scores $(\overline{PS})$ over a given number of stocks gives an index of a forecaster's probability judgment accuracy. The lower the score, the better the overall accuracy with respect to the stocks in question.

3.2. Calibration

Calibration provides information about the forecaster's ability to match the probability assessments with the mean outcome indices (i.e., proportions of correct forecasts). If a forecaster attains 50% correct forecast for all her 0.5 assessments, 60% correct forecast for all her 0.6 assessments, etc., then the forecaster is said to be perfectly calibrated. Lower the calibration score, better the performance in assigning probabilities that match the proportions correct.

Accordingly, a calibration score can be computed as follows:

$$\text{Calibration} = (1/N) \sum N_p (\overline{F_p} - \overline{D_p})^2$$

$\overline{F_p}$: mean of probability forecast categories (e.g. each forecast can be rounded to the nearest tenth, resulting in 0, .1, .2, ..., 1.0)

$\overline{D_p}$: mean outcome index (i.e. the proportion of times the predicted outcome actually occurs) corresponding to forecast F_p

N : total number of stocks

N_p : number of instances in which a forecast of F_p is used.

3.3. Scatter

Scatter gives a weighted average of the variability in the instances when the predicted outcome actually occurs in addition to the variability in the instances when the predicted outcome does not occur. In fact, scatter is an index of the useless variability in the probabilistic forecasts, with lower the scatter value, better the performance is.

Scatter index is computed as:

$$\text{Scatter} = \left[\ (N_1 * Var(F_1)) + (N_0 * Var(F_0)) \ \right] / N$$

$Var(F_1)$: variance of probabilities for all the N_1 cases when the stock price increases

$Var(F_0)$: variance of probabilities for all the N_0 cases when the stock price does not increase

Hence, $N = N_0 + N_1$

3.4. Slope

Slope provides an indication of the forecaster's performance in assigning higher probabilities to instances when his/her chosen outcome occurs than when it does not occur. Higher the slope, better the forecaster is able to discriminate cases where the stock price will or will not increase.

Slope is computed as:

$$\text{Slope} = (\overline{F_1} - \overline{F_0})$$

$\overline{F_1}$: mean of probability forecasts for all the cases when the stock price increases

$\overline{F_0}$: mean of probability forecasts for all the cases when the stock price does not increase

3.5. Bias -- Over/Underconfidence

Bias reflects the forecaster's performance in matching his/her probability assignments ($\overline{F}$) to the overall proportion of correct forecasts ($\overline{D}$). If the mean of the probabilistic forecasts exceed the overall proportion of correct forecasts, than the forecaster is said to be "overconfident". Else, if the overall proportion of correct forecasts exceed the mean of the probabilistic forecasts, then the forecaster is said to be "underconfident" (Lischtenstein and Fischhoff, 1977).

Bias is computed as:

$$\text{Bias} = \overline{F} - \overline{D}$$

Bias gives an indication of tendency to judge the actual occurrence of the predicted outcome as being more likely or less likely than it really is.

4. FINDINGS

Performance measures used to explore the effects of two types of feedback on probabilistic forecasts of stock prices and market indices were : the mean probability score, calibration, scatter, slope and bias.

Performances of two groups were compared session by session using Wilcoxon Matched-Pairs Signed-Ranks Test for each of the performance measures (Appendices 4a-4f). An evaluation of the probabilistic forecasts of both groups is made using an outcome index that is defined in terms of the correctness of the forecaster's predicted outcome.

Descriptive statistics for the scores mentioned above, given by SPSS, including the median, mean, standard deviation, minimum and maximum values are presented in Appendices 3a and 3b for a general idea on both groups in each session.

The median values of the performance measures for the dichotomous forecasts of outcome feedback and calibration feedback groups are as follows:

Median Values for Performance Measures for Dichotomous Forecasts of Simple Outcome Feedback Group and Calibration Feedback Group

Outcome Feedback Group

VARIABLE	SESSION 1	SESSION 2	SESSION 3
$\overline{\mathrm{PS}}$ ↓	.257	.256	.294
$\overline{\mathrm{F}}$	.667	.656	.652
$\overline{\mathrm{D}}$	.544	.528	.458
BIAS 0	.128	.085	.210
CALIBRATION↓	.037	.039	.066
SLOPE ↑	.013	-.001	.000
SCATTER ↓	.004	.003	.004

↓ : smaller values better

↑ : larger values better

0 : values near zero better

Calibration Feedback Group

VARIABLE	SESSION 1	SESSION 2	SESSION 3
$\overline{\mathrm{PS}}$ ↓	.297	.231*[B]	.275*[L]
$\overline{\mathrm{F}}$	.667	.650	.647
$\overline{\mathrm{D}}$	.471	.667*[B]	.500*[W]
BIAS 0	.202	-.085*[B]	.132*[W]
CALIBRATION↓	.078	.032*[B]	.056*[L]
SLOPE ↑	-.018	.004	.014
SCATTER ↓	.008	.006**[B]	.006*[L]

* : $p < 0.05$

** : $p < 0.01$

[B] : Better than previous session

[W]: Worse than previous session

[F] : First session better than last session

[L] : Last session better than first session

Simple Outcome Feedback Group

Simple outcome feedback group, starting with a mean probability score of 0.257 in the first session, sustained their performance in the second session, but had a deterioration in their forecast accuracy in the last session and increased it up to 0.294. In the mean time, their calibration scores staying the same in the first two sessions at a low value, increased to 0.066 in the third session, decreasing their ability to assign probabilities that match the proportions correct. An analysis of scatter scores indicated that, outcome feedback group remained constant in three sessions in the variability in their probabilistic forecasts. This group's ability to discriminate cases whether the stock price increase would or would not occur, depreciated between the first and second sessions, and slope became zero in the last session. Outcome feedback group having lower values in the first two sessions, could not get rid of overconfidence and came up with a higher value in the last session (far from zero) . Therefore, they seemed to display inferior achievement in matching their mean probability assignment to the overall proportion of correct forecasts. However, none of these improvements or deteriorations were found to be statistically significant (all $p>.05$).

Calibration Feedback Group

Calibration feedback group starting with a high mean probability score in the first session, in the second session, after receiving feedback, demonstrated superior results and decreased their score ($p=.0495$). In the third session, calibration feedback group's mean probability score was again found to be better than the first session. When calibration scores were analyzed, it was observed that, starting with a poor performance in assigning probabilities that match the proportions correct, after acquiring calibration feedback, resulted with a lower calibration score in the last session than the first session (p=.0352). An analysis of scatter scores indicated that, having a scatter score of 0.008 in the first session, calibration feedback group accomplished to decrease it to 0.006 (p=.0086), that is, decreasing useless variability and keeping it consistently in the last session (p=.0312). A study of the mean slopes denoted that, calibration feedback group's ability to discriminate cases whether the stock price increase would or would not occur, improved between the first and second sessions (p=.0392), and increased up to 0.014 in the last session, but the increase was not statistically significant. Calibration feedback group initiating with a high positive bias (overconfidence), attained a negative value (underconfidence) nearer to zero in the second session (p=.0352), but could not maintain it and eventuated in overconfidence, being in a better position than the first session. Their improvement in expressing their forecasts may be attributed to their effective use of calibration feedback.

5. CONCLUSION

Many studies were conducted concerning the reliability of financial services in forecasting the stock market and none of them were found to be particularly encouraging. In other words, their forecasts were little better than those that could be expected from pure chance. Therefore, researchers started investigating other ways to enhance forecasting accuracy. The idea of using judgmental forecasting instead of statistical forecasting emerged. Moreover, ways to improve accuracy of probabilistic forecasts became their main concern.

This study tested the effects of two different types of feedback on the accuracy of financial forecasts. The two types of feedback put to use were: (1) simple outcome feedback, and (2) calibration feedback. Like the results of previous studies (Murphy and Daan, 1984; Murphy, Hsu and Winkler, 1985, Benson and Önkal, 1992; Önkal and Muradoglu, 1995), calibration feedback is found to improve forecast accuracy. Önkal and Muradoglu (1995) suggested that feedback in all forms, improved the forecasters' ability to assign accurate probabilities to future outcomes that match actual relative frequencies (i.e., improved forecasters' calibration).

Önkal and Muradoglu (1995) concluded that, feedback, independent of its form, improves the ability of forecasters to assign meaningful probabilities to future outcomes in a financial setting. They argued that, in a dynamic environment like the stock market, the claim that rational expectations can be improved with the assistance of feedback, is important. This opens a way for the comparison of portfolio models for utilizing adaptive expectations (historical data) versus rational expectations (subjective forecasts as inputs).

The simple outcome feedback group which received realized stock prices as the only feedback could not give rise to improved calibration scores; in fact, there existed certain deteriorations in other scores (e.g. slope). Simple outcome feedback was not as successful as calibration feedback in improving forecasters' performance. As a start, simple outcome feedback group's median calibration score was better than calibration feedback group; they could not sustain this outperformance. This implies that, only with simple outcome feedback, investors cannot recover their ability to assign probabilities that match the actual relative frequencies of future outcomes. This inability of simple outcome feedback to improve calibration and overforecasting is consistent with findings of Benson and Önkal (1992).

For the calibration feedback group, a significant improvement is observed in calibration and overforecasting relative to the control group. Calibration feedback group shifted from using two-digit probabilities to one-digit probabilities in later sessions. In addition, they used fewer different probabilities. These suggest that calibration feedback and training led subjects to reduce the number of probability categories to better manage their forecasts. Subjects improved their mean slopes, along with their calibration, which indicated that they improved their ability to discriminate between occasions when the actual price change did or did not occur. Improved calibration and overforecasting are important to forecast users. The better calibrated the forecaster, the more his/her probability forecasts are like relative frequencies and the easier they are to interpret and use (Benson and Önkal, 1992). It is worth exploring whether forecasters' calibration performances would deteriorate if calibration feedback was cut off.

The consistent pattern observed in the calibration feedback group (more improvement in second session, but less in the third session) may be partially due to fluctuations in the market during the period that the study was conducted. An emerging market may be relatively more volatile than a developed market. The forecast horizon was chosen as one-week to guarantee that the forecast-period volatility of the study is comparable to the forecast-period of other studies conducted in developed markets. Due to exchange rates and volatility differences, weekly percentage changes of stock prices in Turkey can be comparable to quarterly percentage changes of stock prices in US (Önkal and Muradoglu, 1995). Future research may compensate the market volatility by running similar experiments for more iterations using different forecast horizons.

One can say that where a person is unfamiliar with a topic or task, where the task is difficult, where he/she is not accountable for the task, or where the task is not significant to the firm; then overconfidence can be expected (O'Connor, 1989). This may well be the typical situation of the use of probabilistic assessment in conjunction with decision analysis in a business environment. Therefore, the users of these probabilities should be aware of this potential problem, and, in future research the choice of subjects can be made according to such relevant categories.

This study suggests that, training may have an impact if it is supported with feedback. Provision of training with feedback may be regarded as an important step towards establishing an effective way of communication using subjective probabilities. Further research about the use of probabilistic forecasting and feedback in different financial settings will be helpful to financial markets.

In this study, calibration feedback is observed to be superior to simple outcome feedback in improving the accuracy of forecasts. This is meaningful for the training of forecasters in financial settings. If feedback improves forecasting abilities as suggested by the study of Önkal and Muradoglu (1995) and this study; this implies that, investors and analysts might be trained in using subjective probabilities for better decisions. The use of probability distributions in financial forecasting along with training on the subjective probabilities, will be helpful in improving the investors' and analysts' understanding and presentation of uncertainty in portfolio management.

REFERENCES

Angus-Leppan, P., Fatseas, V., 1985, "The forecasting accuracy of judgmental and extrapolative methods in forecasting annual earnings", *Accounting and Business Research*, 179-188.

Bachelor, R., Dua, P., 1990, "Forecasting ideology, forecasting technique, and the accuracy of economic forecasts", *International Journal of Forecasting*, 6, 3-10.

Ball, R., Brown, P., 1968, "An empirical evaluation of accounting income numbers", *Journal of Accounting Research*, 6, 159-178.

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

Benson, P.G., Önkal, D., 1992, "The effects of feedback and training on the performance of probability forecasters", *International Journal of Forecasting*, 8, 559-573.

Brealey, R.A., Myers, S.C., 1991, *Principles of Corporate Finance*, McGraw-Hill, Inc.

Bunn, D., Wright, G., 1991, "Interaction of judgmental and statistical forecasting methods: issues and analysis", *Management Science*, 37, 501-518.

Clemen, R.T., Murphy, A.H., 1986, "Objective and subjective precipitation probability forecasts: Statistical analyses of some interrelationships", *Weather and Forecasting*, 1,56-65.

Fama, E.F., 1965, "The behavior of stock market prices", *Journal of Business*, 38, 34-105.

Fischer, G.W., 1982, "Scoring-rule feedback and the overconfidence syndrome in subjective probability distributions", *Organizational Behavior and Human Performance*, 29, 352-369.

Fischhoff, B., 1982, "For those condemned to study the past: Heuristics and biases in hindsight", in Kahneman, D., Slovic, P., Tversky, A., *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Fischhoff, B., MacGregor, D., 1982, "Subjective confidence in forecasts", *Journal of Forecasting*, 1, 155-172.

Flores, B.E., Olson, D.L., Wolfe, C., 1992, "Judgmental adjustment of forecasts: A comparison of methods", *International Journal of Forecasting*, 7, 421-433.

Friedman, D., 1983, "Effective scoring rules for probabilistic forecasts", *Management Science*, 29, 447-454.

Goodwin, P., Wright, G., *Decision Analysis for Management Judgment*, Chichester: John Wiley & Sons.

Hogarth, R.M., 1975, " Cognitive processes and the assessment of subjective probability distributions", *Journal of the American Statistical Association*, 70, 271-289.

Istanbul Stock Exchange, ISE Publications, No:1, 1992.

Jensen, M.C., 1968, "The performance of mutual funds in the period 1945-64", *The Journal of Finance*, 2, 389-416.

Kahneman, D., Slovic, P., Tversky, A., 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge: Cambridge University Press.

Kidd, J.B.,1973, *Scoring Rules for Subjective Assessments*, A paper written for the Annual Conference of the Operational Research Society, Torbay, England.

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

Lichtenstein, S., Fischhoff, B., 1977, "Do those who know more also know about how much they know?: the calibration of probability judgments", *Organizational Behavior and Human Performance*, 20, 159-183.

Lichtenstein, S., Fischhoff, B., 1980, "Training for calibration", *Organizational Behavior and Human Performance*, 28, 149-171.

Lichtenstein, S., Fischhoff, B., Philips, L.D., 1982, "Calibration of probabilities: State of the art to 1980", in D. Kahneman, P. Slovic and A. Tversky (eds.), *Judgment under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Makridakis, S., Wheelwright, S.C., 1979, eds., "Forecasting the future and the future of forecasting", in *Studies in the Managerial Sciences*, 12, *Forecasting*, Amsterdam: North Holland.

McNees, S.K., 1990, "The role of judgment in macroeconomic forecasting accuracy", *International Journal of Forecasting*, 6, 287-299.

Meigs, Robert F., Meigs, Walter B., *Financial Accounting*, McGraw-Hill, Inc., 1992.

Moriarity, M.M., 1985, "Design features of forecasting systems involving management judgments", *Journal of Marketing Research*, 22, 353-364.

Muradoglu, G., Oktay, T., 1993, "Calendar anomalies in the Turkish Stock Market", Paper presented at the International Conference on Business and Economic Development in Middle Eastern and Mediterranean Countries, İstanbul.

Muradoglu, G., Önkal, D., 1992, "Semi-strong form efficiency in a thin market: A case study", Paper presented at the 19th European Finance Association Meeting, Lisbon.

Muradoglu, G., Ünal, M., 1993, "Weak-form efficiency in the thinly traded ISE", Paper presented at the International Conference on Business and Economic Development in Middle Eastern and Mediterranean Countries, İstanbul.

Murphy, A.H., Brown, B.G. 1985, "A comparative evaluation of objective and subjective weather forecasts in the United States", in Wright, G.N. (ed.), *Behavioral Decision Making*, New York: Plenum.

Murphy, A.H., Daan, H., 1984, "Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zienkze experiment", *Monthly Weather Review*, 112, 413-423.

Murphy, A.H., Hsu, W., Winkler, D.S., 1985, "The use of probabilities in subjective quantitative precipitation forecasts: Some experimental results", *Monthly Weather Review*, 113, 2075-2089.

Murphy, A.H., 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, 144-155.

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

Önkal, D., Muradoglu, G., 1994, "Evaluating probabilistic forecasts of stock prices in a developing stock market", *European Journal of Operational Research*, 74, 350-358.

Önkal, D., Muradoglu, G., 1995, "Effects of feedback on probabilistic forecasts of stock prices", *International Journal of Forecasting*, 11, 307-319.

Önkal, D., Muradoglu, G., 1996, "Effects of task format on probabilistic forecasting of stock prices", *International Journal of Forecasting*, 12, 9-24.

Philips, L.D., 1987, "On the adequacy of judgmental forecasts", *Judgmental Forecasting*, in G. Wright, P. Ayton (Eds.), 11-30, John Wiley & Sons Ltd.

Remus, W., O'Connor, M., Griggs, K., 1995, "Does reliable information improve the accuracy of judgmental forecasts?", *International Journal of Forecasting*, 11, 285-293.

Roberts, H.V., 1967, "Statistical versus clinical prediction of the stock market", Paper presented to the Seminar on the Analysis of Security Prices, University of Chicago.

Ronis, D.L., Yates, J.F., 1987, "Components of probability judgment accuracy: Individual consistency and effects of subject matter and assessment method", *Organizational Behavior and Human Decision Processes*, 40, 193-218.

Staël von Holstein, C.A.S., 1972, "Probabilistic forecasting: An experiment related to the stock market", *Organizational Behavior and Human Performance*, 8, 139-158.

Turner, D.S., 1990, "The role of judgment in macroeconomic forecasting", *Journal of Forecasting*, 9, 315-345.

Winkler, R.L., 1969, "Scoring rules and the evaluation of probability assessors", *Journal of the American Statistical Association*, 64, 1073-1078.

Wolfe, C., Flores, B., 1990, "Judgmental adjustment of earnings forecasts", *Journal of Forecasting*, 9, 389-405.

Wright, G., Ayton, P., 1986, "Subjective confidence in forecasts: A response to Fishhoff and MacGregor", *Journal of Forecasting*, 5, 117-123.

Wright, G., Ayton, P., 1987, "The psychology of forecasting", in Wright, G., Ayton, P. (Eds.), *Judgmental Forecasting*, 83-105, John Wiley & Sons Ltd.

Wright, G., Ayton, P., 1989, "Judgmental probability forecasts for personal and impersonal events", *International Journal of Forecasting*, 5, 117-125.

Yaniv, I., Hogarth, R.M., 1993, "Judgmental versus statistical prediction: Information asymmetry and combination rules", *Psychological Science*, 4, 58-62.

Yates, J.F., McDaniel, L.S., Brown, E.S., 1991, "Probabilistic forecasts of stock prices: The hazards of nascent expertise", *Organizational Behavior and Human Decision Processes*, 40, 60-79.

Zarnowitz, V., Lambros, L.A., 1987, "Consensus and uncertainty in economic prediction", *Journal of Political Economy*, 95, 591-621.

APPENDIX 1

STOCKS

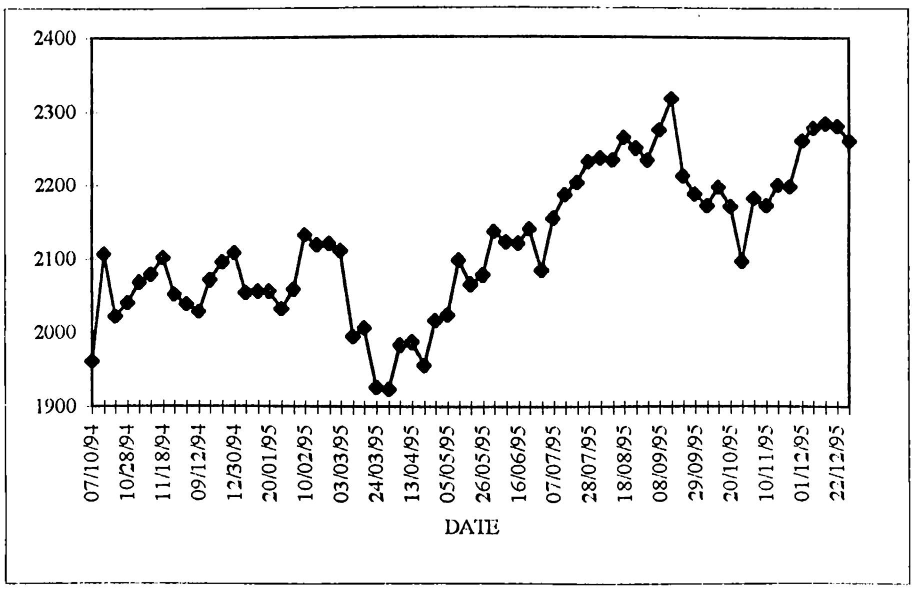
1. ADANA ÇİMENTO(A)
2. AKAL TEKSTİL
3. ALARKO SANAYİ
4. ARÇELİK
5. ASELSAN
6. BOLU ÇİMENTO
7. BRİSA
8. ÇUKUROVA ELEKTRİK
9. DEVA HOLDİNG
10. DIŞBANK
11. DOĞAN HOLDİNG
12. ECZACIBAŞI İLAÇ
13. EREĞLİ DEMİR-ÇELİK
14. GENTAŞ
15. İŞ BANKASI(C)
16. İZOCAM
17. KORDSA
18. KÜTAHYA PORSELEN
19. MİGROS
20. MİLLİYET GAZETECİLİK
21. NET HOLDİNG
22. PETROL OFİSİ
23. PINAR SÜT
24. RAKS ELEKTRONİK
25. SABAH YAYINCILIK
26. SARKUYSAN
27. TAT KONSERVE
28. TİRE KUTSAN
29. TRANSTÜRK HOLDİNG
30. USAŞ

MARKET INDICES

a. DAX
b. ISE COMPOSITE INDEX
c. FT-SE 100
d. DOW JONES INDUSTRIALS
e. CAC 40
f. NIKKEI 225

APPENDIX 2

SAMPLE PAGE FROM THE RESPONSE SHEETS PROVIDED FOR THE SUBJECTS



DATE	
22/09/95	2,212
29/09/95	2,187
06/10/95	2,171
13/10/95	2,197
20/10/95	2,170
27/10/95	2,096
03/11/95	2,182
10/11/95	2,172
17/11/95	2,201
24/11/95	2,198
01/12/95	2,261
08/12/95	2,278
15/12/95	2,284
22/12/95	2,280
29/12/95	2,261
05/01/96	

WHEN COMPARED TO THE PREVIOUS FRIDAY'S CLOSING STOCK PRICE, THIS FRIDAY'S CLOSING PRICE WILL

A. INCREASE
B. STAY THE SAME or DECREASE

YOUR FORECAST (A or B) :

PROBABILITY THAT YOUR FORECAST WILL INDEED OCCUR (I.E., PROBABILITY THAT THE WEEKLY PRICE CHANGE WILL ACTUALLY FALL IN THE DIRECTION YOU PREDICTED)

(BETWEEN 50% and 100%) :

APPENDIX 3a

DESCRIPTIVE STATISTICS

OUTCOME FEEDBACK GROUP

SESSION I

VARIABLE	MEDIAN	MEAN	STDDEV	MINIMUM	MAXIMUM
SLOPE	.013	.02	.03	-.0259815	.1058210
SCATTER	.004	.01	.01	.0002941	.0236056
CALIB	.037	.06	.07	.0074120	.2166330
BIAS	.128	.17	.15	-.0968954	.4310210
$\overline{PS}$	.257	.27	.06	.1747060	.4149410
$\overline{D}$	.544	.52	.16	.2352940	.7941180
$\overline{F}$	.667	.67	.08	.5710560	.8786110

SESSION II

VARIABLE	MEDIAN	MEAN	STD DEV	MINIMUM	MAXIMUM
SLOPE	-.001	.00	.04	-.0462338	.0802749
SCATTER	.003	.01	.00	.0002857	.0139700
CALIB	.039	.07	.07	.0073935	.2371060
BIAS	.085	.12	.19	-.1971430	.4569440
$\overline{PS}$	.256	.27	.08	.1597920	.4389580
$\overline{D}$	.528	.54	.19	.2777780	.8055560
$\overline{F}$	.656	.67	.06	.6028570	.7633330

SESSION III

VARIABLE	MEDIAN	MEAN	STD DEV	MINIMUM	MAXIMUM
SLOPE	.000	.00	.02	-.0303405	.0505874
SCATTER	.004	.01	.01	.0000000	.0203318
CALIB	.066	.07	.04	.0056787	.1708360
BIAS	.210	.18	.09	-.0416110	.3344440
$\overline{PS}$	.294	.29	.04	.2305560	.3830170
$\overline{D}$	.458	.48	.09	.3611110	.6388890
$\overline{F}$	.652	.67	.06	.5972780	.7788890

APPENDIX 3b

DESCRIPTIVE STATISTICS

CALIBRATION FEEDBACK GROUP

SESSION I

VARIABLE	MEDIAN	MEAN	STD DEV	MINIMUM	MAXIMUM
SLOPE	-.018	-.01	.04	-.0921031	.0811874
SCATTER	.008	.01	.01	.0009134	.0316197
CALIB	.202	.12	.10	.0108148	.3293830
BIAS	.078	.23	.24	-.1256780	.8291990
$\overline{PS}$	.297	.32	.08	.1958240	.5057060
$\overline{D}$	.471	.46	.18	.2058820	.7647060
$\overline{F}$	.667	.63	.16	.0639028	.7723610

SESSION II

VARIABLE	MEDIAN	MEAN	STD DEV	MINIMUM	MAXIMUM
SLOPE	.004	.00	.02	-.0330132	.0371853
SCATTER	.006	.01	.01	.0009983	.0223910
CALIB	.032	.06	.05	.0119213	.1491270
BIAS	-.085	.03	.21	-.2218890	.3459170
$\overline{PS}$	.231	.25	.07	.1656250	.3794450
$\overline{D}$	.667	.62	.19	.2222220	.8333330
$\overline{F}$	.650	.65	.06	.5553060	.7555560

SESSION III

VARIABLE	MEDIAN	MEAN	STD DEV	MINIMUM	MAXIMUM
SLOPE	.014	.01	.03	-.0719444	.0525037
SCATTER	.006	.01	.01	.0006815	.0208404
CALIB	.056	.05	.03	.0022224	.1063710
BIAS	.132	.14	.10	-.0331110	.3113050
$\overline{PS}$	.275	.27	.03	.1989000	.3367500
$\overline{D}$	.500	.51	.09	.3333330	.6857140
$\overline{F}$	.647	.65	.05	.5699450	.7527780

APPENDIX 4a

```
*****************************************
*       OUTCOME FEEDBACK GROUP          *
*       SESSION I VS. SESSION II        *
*****************************************
```

Wilcoxon Matched-Pairs Signed-Ranks Test

BIAS1
with BIAS2

Mean Rank	Cases
6.50	8 - Ranks (BIAS2 LT BIAS1)
7.80	5 + Ranks (BIAS2 GT BIAS1)
	0 Ties (BIAS2 EQ BIAS1)
	--
	13 Total

Z = -.4543 2-Tailed P = **.6496**

CALIB1
with CALIB2

Mean Rank	Cases
5.86	7 - Ranks (CALIB2 LT CALIB1)
8.33	6 + Ranks (CALIB2 GT CALIB1)
	0 Ties (CALIB2 EQ CALIB1)
	--
	13 Total

Z = -.3145 2-Tailed P = **.7532**

DBAR1
with DBAR2

Mean Rank	Cases
7.50	6 - Ranks (DBAR2 LT DBAR1)
6.57	7 + Ranks (DBAR2 GT DBAR1)
	0 Ties (DBAR2 EQ DBAR1)
	--
	13 Total

Z = -.0349 2-Tailed P = **.9721**

FBAR1
with FBAR2

Mean Rank	Cases
7.00	5 - Ranks (FBAR2 LT FBAR1)
7.00	8 + Ranks (FBAR2 GT FBAR1)
	0 Ties (FBAR2 EQ FBAR1)
	--
	13 Total

Z = -.7338 2-Tailed P = **.4631**

PSBAR1
with PSBAR2

Mean Rank	Cases
7.00	6 - Ranks (PSBAR3 LT PSBAR2)
7.00	7 + Ranks (PSBAR3 GT PSBAR2)
	0 Ties (PSBAR3 EQ PSBAR2)
	--
	13 Total

Z = -.2446 2-Tailed P = **.8068**

SCATTER1
with SCATTER2

Mean Rank	Cases
6.56	9 - Ranks (SCATTER2 LT SCATTER1)
8.00	4 + Ranks (SCATTER2 GT SCATTER1)
	0 Ties (SCATTER2 EQ SCATTER1)
	--
	13 Total

Z = -.9435 2-Tailed P = **.3454**

SLOPE1
with SLOPE2

Mean Rank	Cases
6.82	11 - Ranks (SLOPE2 LT SLOPE1)
8.00	2 + Ranks (SLOPE2 GT SLOPE1)
	0 Ties (SLOPE2 EQ SLOPE1)
	--
	13 Total

Z = -2.0616 2-Tailed P = **.0392**

APPENDIX 4b

```
*******************************************
*      OUTCOME FEEDBACK GROUP            *
*      SESSION I VS. SESSION III         *
*******************************************
```

<u>Wilcoxon Matched-Pairs Signed-Ranks Test</u>

BIAS1
with BIAS3

Mean Rank	Cases
8.60	5 - Ranks (BIAS3 LT BIAS1)
6.89	9 + Ranks (BIAS3 GT BIAS1)
	0 Ties (BIAS3 EQ BIAS1)
	--
	14 Total

Z = -.5964   2-Tailed P = **.5509**

CALIB1
with CALIB3

Mean Rank	Cases
9.75	4 - Ranks (CALIB3 LT CALIB1)
6.60	10 + Ranks (CALIB3 GT CALIB1)
	0 Ties (CALIB3 EQ CALIB1)
	--
	14 Total

Z = -.8475   2-Tailed P = **.3967**

DBAR1
with DBAR3

Mean Rank	Cases
6.80	10 - Ranks (DBAR3 LT DBAR1)
9.25	4 + Ranks (DBAR3 GT DBAR1)
	0 Ties (DBAR3 EQ DBAR1)
	--
	14 Total

Z = -.9730   2-Tailed P = **.3305**

FBAR1
with FBAR3

Mean Rank	Cases
6.94	8 - Ranks (FBAR3 LT FBAR1)
8.25	6 + Ranks (FBAR3 GT FBAR1)
	0 Ties (FBAR3 EQ FBAR1)
	--
	14 Total

Z = -.1883 2-Tailed P = **.8506**

PSBAR1
with PSBAR3

Mean Rank	Cases
7.25	4 - Ranks (PSBAR3 LT PSBAR1)
7.60	10 + Ranks (PSBAR3 GT PSBAR1)
	0 Ties (PSBAR3 EQ PSBAR1)
	--
	14 Total

Z = -1.4752 2-Tailed P = **.1401**

SCATTER1
with SCATTER3

Mean Rank	Cases
8.67	9 - Ranks (SCATTER3 LT SCATTER1)
5.40	5 + Ranks (SCATTER3 GT SCATTER1)
	0 Ties (SCATTER3 EQ SCATTER1)
	--
	14 Total

Z = -1.6008 2-Tailed P = **.1094**

SLOPE1
with SLOPE3

Mean Rank	Cases
7.70	10 - Ranks (SLOPE3 LT SLOPE1)
7.00	4 + Ranks (SLOPE3 GT SLOPE1)
	0 Ties (SLOPE3 EQ SLOPE1)
	--
	14 Total

Z = -1.5380 2-Tailed P = **.1240**

APPENDIX 4c

```
******************************************
*       OUTCOME FEEDBACK GROUP           *
*       SESSION II VS. SESSION III       *
******************************************
```

<u>Wilcoxon Matched-Pairs Signed-Ranks Test</u>

BIAS2
with BIAS3

Mean Rank	Cases
9.33	3 - Ranks (BIAS3 LT BIAS2)
6.30	10 + Ranks (BIAS3 GT BIAS2)
	0 Ties (BIAS3 EQ BIAS2)
	--
	13 Total

Z = -1.2230 2-Tailed P = **.2213**

CALIB2
with CALIB3

Mean Rank	Cases
7.80	5 - Ranks (CALIB3 LT CALIB2)
6.50	8 + Ranks (CALIB3 GT CALIB2)
	0 Ties (CALIB3 EQ CALIB2)
	--
	13 Total

Z = -.4543 2-Tailed P = **.6496**

DBAR2
with DBAR3

Mean Rank	Cases
6.88	8 - Ranks (DBAR3 LT DBAR2)
5.75	4 + Ranks (DBAR3 GT DBAR2)
	1 Ties (DBAR3 EQ DBAR2)
	--
	13 Total

Z = -1.2551 2-Tailed P = **.2094**

FBAR2
with FBAR3

Mean Rank	Cases
7.50	8 - Ranks (FBAR3 LT FBAR2)
6.20	5 + Ranks (FBAR3 GT FBAR2)
	0 Ties (FBAR3 EQ FBAR2)
	--
	13 Total

Z = -1.0133 2-Tailed P = **.3109**

PSBAR2
with PSBAR3

Mean Rank	Cases
6.00	5 - Ranks (PSBAR3 LT PSBAR2)
7.63	8 + Ranks (PSBAR3 GT PSBAR2)
	0 Ties (PSBAR3 EQ PSBAR2)
	--
	13 Total

Z = -1.0832 2-Tailed P = **.2787**

SCATTER2
with SCATTER3

Mean Rank	Cases
6.13	8 - Ranks (SCATTER3 LT SCATTER2)
8.40	5 + Ranks (SCATTER3 GT SCATTER2)
	0 Ties (SCATTER3 EQ SCATTER2)
	--
	13 Total

Z = -.2446 2-Tailed P = **.8068**

SLOPE2
with SLOPE3

Mean Rank	Cases
5.25	8 - Ranks (SLOPE3 LT SLOPE2)
9.80	5 + Ranks (SLOPE3 GT SLOPE2)
	0 Ties (SLOPE3 EQ SLOPE2)
	--
	13 Total

Z = -.2446 2-Tailed P = **.8068**

APPENDIX 4d

****
*   CALIBRATION FEEDBACK GROUP   *
*   SESSION I VS. SESSION II   *
****

Wilcoxon Matched-Pairs Signed-Ranks Test

BIAS1
with BIAS2

Mean Rank	Cases
11.00	11 - Ranks (BIAS2 LT BIAS1)
5.33	6 + Ranks (BIAS2 GT BIAS1)
	0 Ties (BIAS2 EQ BIAS1)
	--
	17 Total

Z = -2.1065   2-Tailed P = **.0352**

CALIB1
with CALIB2

Mean Rank	Cases
11.40	10 - Ranks (CALIB2 LT CALIB1)
5.57	7 + Ranks (CALIB2 GT CALIB1)
	0 Ties (CALIB2 EQ CALIB1)
	--
	17 Total

Z = -1.7752   2-Tailed P = **.0759**

DBAR1
with DBAR2

Mean Rank	Cases
6.33	6 - Ranks (DBAR2 LT DBAR1)
10.45	11 + Ranks (DBAR2 GT DBAR1)
	0 Ties (DBAR2 EQ DBAR1)
	--
	17 Total

Z = -1.8225   2-Tailed P = **.0684**

FBAR1
with FBAR2

Mean Rank	Cases
10.25	8 - Ranks (FBAR2 LT FBAR1)
7.89	9 + Ranks (FBAR2 GT FBAR1)
	0 Ties (FBAR2 EQ FBAR1)
	--
	17 Total

Z = -.2604   2-Tailed P = **.7946**

PSBAR1
with PSBAR2

Mean Rank	Cases
10.73	11 - Ranks (PSBAR2 LT PSBAR1)
5.83	6 + Ranks (PSBAR2 GT PSBAR1)
	0 Ties (PSBAR2 EQ PSBAR1)
	--
	17 Total

Z = -1.9645   2-Tailed P = **.0495**

SCATTER1
with SCATTER2

Mean Rank	Cases
10.15	13 - Ranks (SCATTER2 LT SCATTER1)
5.25	4 + Ranks (SCATTER2 GT SCATTER1)
	0 Ties (SCATTER2 EQ SCATTER1)
	--
	17 Total

Z = -2.6273   2-Tailed P = **.0086**

SLOPE1
with SLOPE2

Mean Rank	Cases
6.57	7 - Ranks (SLOPE2 LT SLOPE1)
10.70	10 + Ranks (SLOPE2 GT SLOPE1)
	0 Ties (SLOPE2 EQ SLOPE1)
	--
	17 Total

Z = -1.4438   2-Tailed P = **.1488**

APPENDIX 4e

****
*   CALIBRATION FEEDBACK GROUP   *
*   SESSION I VS. SESSION III   *
****

Wilcoxon Matched-Pairs Signed-Ranks Test

BIAS1
with **BIAS3**

Mean Rank	Cases
8.58	12 - Ranks (BIAS3 LT BIAS1)
10.00	5 + Ranks (BIAS3 GT BIAS1)
	0 Ties (BIAS3 EQ BIAS1)
	--
	17 Total

Z = -1.2545   2-Tailed P = **.2097**

CALIB1
with **CALIB3**

Mean Rank	Cases
11.00	11 - Ranks (CALIB3 LT CALIB1)
5.33	6 + Ranks (CALIB3 GT CALIB1)
	0 Ties (CALIB3 EQ CALIB1)
	--
	17 Total

Z = -2.1065   2-Tailed P = **.0352**

DBAR1
with **DBAR3**

Mean Rank	Cases
8.50	6 - Ranks (DBAR3 LT DBAR1)
8.50	10 + Ranks (DBAR3 GT DBAR1)
	1 Ties (DBAR3 EQ DBAR1)
	--
	17 Total

Z = -.8790   2-Tailed P = **.3794**

FBAR1
with FBAR3

Mean Rank	Cases
8.27	11 - Ranks (FBAR3 LT FBAR1)
10.33	6 + Ranks (FBAR3 GT FBAR1)
	0 Ties (FBAR3 EQ FBAR1)
	--
	17 Total

Z = -.6864 2-Tailed P = **.4925**

PSBAR1
with PSBAR3

Mean Rank	Cases
11.20	10 - Ranks (PSBAR3 LT PSBAR1)
5.86	7 + Ranks (PSBAR3 GT PSBAR1)
	0 Ties (PSBAR3 EQ PSBAR1)
	--
	17 Total

Z = -1.6805 2-Tailed P = **.0929**

SCATTER1
with SCATTER3

Mean Rank	Cases
8.71	14 - Ranks (SCATTER3 LT SCATTER1)
10.33	3 + Ranks (SCATTER3 GT SCATTER1)
	0 Ties (SCATTER3 EQ SCATTER1)
	--
	17 Total

Z = -2.1539 2-Tailed P = **.0312**

SLOPE1
with SLOPE3

Mean Rank	Cases
6.14	7 - Ranks (SLOPE3 LT SLOPE1)
11.00	10 + Ranks (SLOPE3 GT SLOPE1)
	0 Ties (SLOPE3 EQ SLOPE1)
	--
	17 Total

Z = -1.5858 2-Tailed P = **.1128**

APPENDIX 4f

****

* **CALIBRATION FEEDBACK GROUP** *
* **SESSION II VS. SESSION III** *

****

Wilcoxon Matched-Pairs Signed-Ranks Test

BIAS2
with **BIAS3**

Mean Rank	Cases
7.40	5 - Ranks (BIAS3 LT BIAS2)
9.67	12 + Ranks (BIAS3 GT BIAS2)
	0 Ties (BIAS3 EQ BIAS2)
	--
	17 Total

Z = -1.8699 2-Tailed P = **.0615**

CALIB2
with **CALIB3**

Mean Rank	Cases
9.75	8 - Ranks (CALIB3 LT CALIB2)
8.33	9 + Ranks (CALIB3 GT CALIB2)
	0 Ties (CALIB3 EQ CALIB2)
	--
	17 Total

Z = -.0710 2-Tailed P = **.9434**

DBAR2
with **DBAR3**

Mean Rank	Cases
9.50	12 - Ranks (DBAR3 LT DBAR2)
7.80	5 + Ranks (DBAR3 GT DBAR2)
	0 Ties (DBAR3 EQ DBAR2)
	--
	17 Total

Z = -1.7752 2-Tailed P = **.0759**

FBAR2
with FBAR3

Mean Rank	Cases
9.00	8 - Ranks (FBAR3 LT FBAR2)
9.00	9 + Ranks (FBAR3 GT FBAR2)
	0 Ties (FBAR3 EQ FBAR2)
	--
	17 Total

Z = -.2130 2-Tailed P = **.8313**

PSBAR2
with PSBAR3

Mean Rank	Cases
8.60	5 - Ranks (PSBAR3 LT PSBAR2)
9.17	12 + Ranks (PSBAR3 GT PSBAR2)
	0 Ties (PSBAR3 EQ PSBAR2)
	--
	17 Total

Z = -1.5858 2-Tailed P = **.1128**

SCATTER2
with SCATTER3

Mean Rank	Cases
7.00	11 - Ranks (SCATTER3 LT SCATTER2)
12.67	6 + Ranks (SCATTER3 GT SCATTER2)
	0 Ties (SCATTER3 EQ SCATTER2)
	--
	17 Total

Z = -.0237 2-Tailed P = **.9811**

SLOPE2
with SLOPE3

Mean Rank	Cases
8.83	6 - Ranks (SLOPE3 LT SLOPE2)
9.09	11 + Ranks (SLOPE3 GT SLOPE2)
	0 Ties (SLOPE3 EQ SLOPE2)
	--
	17 Total

Z = -1.1124 2-Tailed P = **.2659**