8628363

McEwen, Ruth Ann

AN EMPIRICAL ASSESSMENT OF ERROR METRICS APPLIED TO ANALYSTS' FORECASTS OF EARNINGS

Georgia Institute of Technology PH.D. 1986

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

AN EMPIRICAL ASSESSMENT OF
ERROR METRICS APPLIED TO
ANALYSTS' FORECASTS OF EARNINGS

A THESIS

Presented to

The Faculty of the Division of Graduate Studies

By

Ruth Ann McEwen

In Partial Fulfillment

Of The Requirements for the Degree

Doctor of Philosophy in the School of Management

Georgia Institute of Technology

July 1986

AN EMPIRICAL ASSESSMENT OF

ERROR METRICS APPLIED TO

ANALYSTS' FORECASTS OF EARNINGS

By

Ruth Ann McEwen

Approved:

Eugene E. Comiskey

Eugene E. Comiskey, Chairman

Charles W. Mulford

Charles W. Mulford

Bernell K. Stone

Bernell K. Stone

Russell G. Heikes

Russell G. Heikes

Leonard Jon Parsons

Leonard J. Parsons

Date Approved By Chairman:

July 24, 1986

ACKNOWLEDGEMENTS

I would like to thank the members of my committee for their support and encouragment during this project. Two members of this committee, Dr. Eugene Comiskey and Dr. Charles Mulford, deserve special mention for their guidance during every phase of this project, in addition to their support during the doctoral program.

TABLE OF CONTENTS

LIST OF TABLES AND EXHIBITS

SUMMARY

Certain choices confront researchers and other users of analysts' forecasts of earnings when measuring error. First, the computational form of the error metric may express either a general linear or nonlinear relationship between the forecast error and user loss. The second option is closely related and concerns the definition of the internal forecast parameter which is incorporated in the error metric. The purpose of this study is to empirically examine the effects of these choices on the measurement of analysts' forecast errors, and to analyze the effects of error metric choice in a variety of circumstances.

This purpose may be stated as four research objectives. These objectives are:

1) To analyze the effects of error metric selection on conclusions drawn from previous studies of comparative forecast accuracy of analysts with mechanical models.

2) To determine if forecast errors produced from error metrics which employ forecast median differ significantly from those which employ forecast mean.

3) To determine if forecast errors resulting from alternative error metrics provide significantly different estimates of risk.

4) In the event that alternative error metrics are shown to

provide significantly different estimates of risk, to determine if a particular error metric produces forecast errors which are more highly correlated with systematic risk.

The results of hypotheses designed to test the objectives offer the following conclusions.

1) Error metric selection affects the conclusions of previous studies in which analysts were compared with a naive, no-change model. This result indicates that conclusions of some previous studies must be viewed as tentative, since their results are error-metric dependent.

2) Analysis of the mean/median differences fail to yield significantly different error metrics, supporting the view that the distribution of analysts' forecasts is approximately symmetrical.

3) Alternative error metrics change the rank ordering of firms (ranked on forecast error). This result suggests that different risk estimates are provided using alternative error metrics.

4) An analysis of the relationship of forecast error with systematic risk implies that: a) investor loss functions may be asymmetric; b) nonlinear error metrics exhibit higher correlation with systematic risk for overestimates; and, c) investor loss functions may be described as linear for firms which were underestimated.

The implications of these results suggest that error metric form is an important consideration in assessing analysts' forecasts of earnings. Error metric selection affects both an analyst/model comparative analysis, and risk prediction. This study provides evidence which supports nonlinear error forms in risk prediction for firms which have been overestimated, and linear forms for firms which have been underestimated. These results emphasize the need to determine measures of error which may be consistently applied in a comparative analysis, and in assessments of security risk.

CHAPTER I

INTRODUCTION

Analysts' forecasts of earnings are employed by a variety of users. Earnings forecasts are: 1) sold as products by analysts; 2) applied in share valuation models by investors; 3) utilized in loan decisions by creditors; and, 4) incorporated into models which yield managements' forecasts of earnings.

As suggested by numerous research efforts, error metrics which are often used to evaluate the accuracy of analysts' forecasts of earnings should incorporate the losses incurred by these users. For example, Barefield and Comiskey [1975] suggested that user loss should be a major determinant in selection of an error metric. Brandon and Jarrett [1977], reiterated this position, suggesting that the user loss function could be expressed in general linear or nonlinear terms, and error metrics should correspond to the appropriate functional relationship.

The accuracy of analysts' forecasts of earnings and the properties of resultant forecast errors have been the subjects of numerous research efforts. Previous studies include those in which: 1) comparisons of analysts' forecast accuracy with other forecast agents (management and mechanical models) were performed; 2) the effects of accounting changes on analysts' forecast accuracy were estimated; 3) the informational content of analysts' forecast accuracy was inferred; and, 4) the contemporaneous relationship between analysts' forecast accuracy and capital market risk was exhibited.

Various forecast error metrics were employed in these studies. Yet, little theoretical or empirical support was offered to justify the use of the error metrics selected for empirical analysis. Moreover, none of these studies presented evidence that alternative error metrics would provide similar results.

As noted by Brandon and Jarrett [1977], alternative error metrics are not interchangeable. The authors provided preliminary evidence which indicated that for a limited number of firms, in a single empirical setting, alternative error metrics could produce different results.

The purpose of this study is to empirically examine the effects of error metric selection in a variety of circumstances. The results of this study will offer insights into: 1) the consistency of results from previously identified empirical settings based upon alternative error metrics; 2) the effects of user loss function assumptions on error metric definitions; and, 3) the appropriate metric definition in assessments of security risk.

Certain choices confront researchers in selection of a forecast error metric. First, the computational form of the metric may express a general linear or nonlinear relationship between the forecast error and user loss. The choice between linear and nonlinear expression has repeatedly been linked to the concept of a user loss function.

For example, Barefield and Comiskey [1975] were among the first to suggest that the assumption of a specific user loss functional relationship should be a major determinant of metric choice. That is, if the assumption is made that user loss is a linear (nonlinear) function of forecast accuracy, then forecast error should be expressed using a consistent linear (nonlinear) computational form. Therefore, the first alternative to be considered is that of linear versus nonlinear assumptions of user loss, and the effects of this assumption on the computational forms of forecast error metrics.

The second option is closely related and concerns the definition of the forecast parameter which is incorporated in the error metric. Definition of this internal forecast statistic has been forecast mean in all previous efforts which used data bases that included multiple forecasts for each firm. Yet, the general linear case, which assumes a corresponding linear user loss function, may be more appropriately defined by use of the forecast median. An analysis of minimum error cost in the general linear case presents compelling theoretical support for using the forecast median in this situation [Hamburg, 1983].

This study focuses on an analysis of user loss functions, and evaluates the effects of these options on alternative error metrics. The purposes of this study are summarized by the following research objectives:

1) To analyze the effects of error metric selection on conclusions drawn from previous studies of comparative forecast accuracy of analysts with mechanical models.

2) To determine if forecast errors produced from error metrics which employ forecast median differ significantly from those which employ forecast mean.

3) To determine if forecast errors resulting from alternative error metrics provide significantly different estimates of risk.

4) In the event that alternative error metrics are shown to provide significantly different estimates of risk, to determine if a particular error metric produces forecast errors which are more highly correlated with systematic risk.

Chapter II provides a literature review based upon previous research efforts which pertain to analyst forecast accuracy. The conclusions drawn from these studies will be employed in the definition of error metrics for the current study, and in the rationale for hypotheses used to test the research objectives.

Chapter III presents a definition of user loss functions and provides a theoretical basis for definition of two error metrics which correspond to general linear and nonlinear investor loss. In addition, seven other error metrics, which have been employed in previous studies, are presented.

Chapter IV presents a detailed discussion of each objective. In addition, this chapter states the hypotheses which are used to test the objectives, and provides an appropriate statistical design to test each hypothesis.

Chapter V reports the results of the empirical tests. In addition, descriptive statistics are presented for all error metrics.

Chapter VI concludes the study, and identifies its limitations. Suggestions for further research are also presented.

CHAPTER II

REVIEW OF THE LITERATURE

A diverse grouping of empirical studies form the basis of research efforts pertaining to analysts' forecast accuracy. Early studies focused on the accuracy of analysts' forecasts compared with those generated from mathematical models, with those generated by management, and with combinations of the alternate sources of forecasts. More recent studies have examined the properties of analysts' forecast errors, and the relationship of forecast error with systematic risk.

The current study focuses on these general areas. The insights gained from comparative accuracy studies, in addition to the studies which tested the properties of forecast error, are employed in defining error metrics and proposing hypotheses tests concerning the effects of alternative error metrics on comparative accuracy. The results of previous studies in which forecast error was viewed as a surrogate for security risk are used to identify an error metric which may best represent security risk (in the sense that one form may exhibit higher correlation with systematic risk). A review of revelant studies is presented in the paragraphs which follow.

Relative and Absolute Accuracy of Analysts' Forecasts

In previous studies, relative accuracy was defined in the context of a comparative analysis; accuracy measures were computed for alternative forecast agents, and the agent with the lowest forecast error was judged to be the superior agent. Accuracy measures were also presented in an absolute sense, whereby forecast error metrics were employed as one means of comprehensive assessment of analysts' abilities to forecast. Neither the relative nor the absolute accuracy measures were defined in a systematic manner which provided cross-study comparisons. Thus, a review of these research efforts provides a basis for analysis of error metrics and the underlying assumptions which limit comparability and interpretation.

Relative and Absolute Accuracy of Analysts' Forecasts Compared with Mathematical Models

Early efforts centered around the relative accuracy of the forecast. These studies compared the performance of analysts' forecasts to a variety of naive and mechanical models.

Givoly and Lakonishok [1984] suggested this preoccupation with accuracy was understandable. They provided the following motivation for emphasis on this area of research:

> Next to stock recommendations, earnings forecasts are perhaps the most prominent output of the financial analysts' industry. If FAF (financial analysts' forecasts of earnings), which are costly both socially and privately, do not outperform the much less expensive naive predictions, then their

> very existence becomes questionable; and because earnings predictions are used for stock valuation and selection, inaccurate predictions may lead to wrong investment decisions [1984, p 118].

The authors offered, as an additional impetus for research concerning relative accuracy, the increased interest in proposed mandatory disclosure of management forecasts. Citing the scarcity of management forecasts and the potentially biased nature of those available, the authors contended analysts' forecasts could be viewed as a "test ground" for evaluating management forecasts [1984, p. 119].

Empirical studies in this area provided mixed conclusions. Several studies suggested analysts were superior forecasters when compared with mathematical models, while other work concluded that analysts provided predictions with the same relative accuracy.

In one of the first empirical efforts, Cragg and Malkiel [1968] examined the degree of agreement among analysts' forecasts and four naive models. The authors also examined the association between past and forecasted earnings growth rates and the correlation between earnings growth forecasts and price/earnings ratios. Comparison of the forecast variable, a five-year analyst growth prediction, with a simple naive model based on no change in past growth rates, implied that analysts only slightly outperformed the mechanical model.

Additional conclusions included the assertion that price/earnings ratios did not predict future earnings growth any better than analysts' forecasts or past growth rates [1968, p. 83]. Absolute accuracy measures were based on estimates of normalized earnings. Therefore, summary statistics provided only estimates of ex post absolute accuracy.

The authors also stated:

> Similar analysis was performed to determine the extent to which errors in predictions were related to 1) errors in predicting the average over-all earnings growth of the sample firms; 2) errors in predicting the average growth rate of particular industries; and 3) errors in predicting the growth rates of firms within industries [1968, p. 76].

Results were difficult to interpret, and, in general, the authors were unable to associate accuracy with industry or company characteristics [1968, p. 80].

Elton and Gruber [1972] confirmed the results presented by Cragg and Malkiel [1968]. They tested the annual earnings forecasts for a group of analysts representing a large pension fund, a brokerage house, and an investment advisory service. They found no significant difference in the accuracy of the analysts' forecasts and a forecast generated by an exponential smoothing model. This empirical study also tested eight other mechanical models. Additionally, the time periods incorporated in the forecast included forecasts for two- and three-year periods. Absolute accuracy measures were reported only on a comparative basis for the analysts.

Both of the previously cited studies have been criticized as containing several limitations and biases. Interestingly, the research in the area of relative accuracy is divided between these earlier studies which concluded that analysts perform no better than mechanical models, and subsequent research which concluded that analysts were superior forecasters. Givoly and Lakonishok [1984] suggested this inconsistency may have been due to the inherent limitations of the earlier work.

The first limitation was that Cragg and Malkiel [1968] used predictions of five-year growth rates rather than yearly forecasts. Analysts may be capable of predicting short term changes in earnings to which naive models are "blind"; thus, five-year analyst forecasts were inappropriate for assessing analyst accuracy. In addition, Cragg and Malkiel [1968] did not define the earnings variable in a uniform manner across forecasters. Thus, resultant forecast errors in both a relative and an absolute sense were difficult to interpret.

These early studies incorporated, at most, three annual forecasts. Relative and absolute accuracy of analysts' forecasts may vary over time. Later studies compared longer series of forecasts, thereby isolating the effects of time dependency on forecast errors. Later studies attempted to reduce the effects of these limitations on inferences drawn. The conclusions of this later body of work consistently provided support for analysts' superiority in forecasting when compared with mechanical models.

One of the first studies to incorporate a longer time period was performed by Barefield and Comiskey [1975]. The authors further altered the research methodology by selecting analysts' forecasts from Standard and Poor's Earnings Forecaster rather than nonpublic sources.

The authors compared six years of forecasts (1967-1972) provided for 100 New York Stock Exchange firms with December 31 reporting dates. The benchmark for comparison was a naive no-change model. Using Theil's Inequality Coefficient [Theil, 1966], the authors documented the superiority of analysts in 68 of the 100 cases tested [Barefield and Comiskey, 1975, p. 247]. Similarly, analysts were better predictors of turning points, accurately predicting 132 of 197 turning points [1975, p. 249].

Evidence relating to absolute accuracy was also provided. Analysts predicted earnings with an average forecast error of 16.07% across the six year period; however, analysts tended to overestimate earnings rather than underestimate earnings, indicating a potential bias may have been present [1975,pp. 247, 249]. This study also provided an analysis of the proposed determinants of analysts' forecast errors.

Brandon and Jarrett [1977] compared analysts' forecasts with a variety of sophisticated univariate and multivariate models, using both linear and nonlinear extrapolation. Of special interest was a quantitative comparison of error metrics. The authors confirmed the conclusions drawn by Barefield and Comiskey [1975] that analysts provided more accurate forecasts than were generated from mathematical models.

A limitation of Brandon and Jarrett [1977] is the method by which the test sample was chosen. Sample firms were selected by Standard and Poor's, the forecast publishers, at the request of the authors. Possible bias was introduced by this nonrandom, voluntary choice of sample firms.

An analysis of error metrics was provided by Brandon and Jarrett [1977]. A comparison of error metrics indicated that the choice of metric used for empirical analysis in previous efforts may have affected the conclusions drawn. Brandon and Jarrett, [1977, p. 45], noted that, "...measures of accuracy are not necessarily interchangeable."

Brandon and Jarrett's conclusions provide the motivation for the current study. Error metrics were used in previous efforts with little justification, and the impact of alternative error metrics was not examined. Subsequent efforts also failed to address this issue.

The current study examines the effects of error metric selection in a variety of circumstances. This study differs from Brandon and Jarrett by: 1) randomly selecting the sample from the Institutional Brokers Estimate System data source; 2) increasing the sample size to approximately 750 firms; 3) providing two error metrics which have been theoretically justified; and, 4) presenting the analysis in terms of user loss and corresponding loss functions.

A review of those studies presented subsequent to Brandon and Jarrett [1977] indicates that error metric selection was not a primary consideration. Instead, the authors focused on mechanical model selection, using a rational markets paradigm to infer that analysts must be superior, or their forecasts would not be purchased.

Richards, Benjamin, and Strawser [1977] confirmed previous conclusions by providing evidence that supported analysts' superior performance compared with mechanical models. While the authors stated their initial objective was to extend length of the forecast horizon of previous studies, data availability limited the analysis to four years (1972-1976). The authors compared forecasts from Standard and Poor's Earnings Forecaster with three naive models, and reported mean absolute relative errors for analysts and models as 24.1% and 28.9%, respectively [1977, p.82].

In an analysis of forecast error classified by industry, Richards, Benjamin and Strawser, [1977, p. 84], confirmed a conclusion drawn by Barefield and Comiskey [1975], among others, that, "The forecast errors generally reflect the variability in earnings across industries studied." The authors also noted that, "Mechanical models are more reliable for forecasting earnings of firms in stable industries." Yet, these conclusions may have been affected by the use of the error selected for empirical analysis.

A more sophisticated analysis and comparison of forecast errors was presented by Brown and Rozeff [1978]. The authors were among the first to use nonparametric statistical tests to compare analysts' forecasts with forecasts generated by three firm-fitted models. The authors tested two error metrics and multiple forecast horizons to compare forecasts for the years 1972-1975 generated from Value Line with those generated from: 1) a seasonal martingale model; 2) a seasonal submartingale model; and, 3) a Box-Jenkins autoregressive model. Brown and Rozeff, [1978 p. 1], contended that, "In contrast with other studies, the results overwhelmingly favor the superiority of analysts over time-series models."

This study differed from previous work in two respects. First, the authors used nonparametric tests of differences of means, asserting that the parametric paired t-test was inappropriate for testing mean error differences of forecast methods applied to cross-sectional earnings data. Second, the authors compared forecasts of analysts with firm-fitted mathematical models, using both annual and quarterly data.

While Brown and Rozeff [1978] did provide alternative definitions of error in their analysis, they did not offer theoretical justification for the metrics selected. Additionally, no evidence was provided which suggested that other error metrics could produce the same results.

Armstrong and Beuchert [1979] hypothesized that differential advantages arise for analysts when forecasting earnings. Summarizing evidence presented in previous research, the authors provided empirical analysis which suggested that analysts perform better than naive or sophisticated models, and management provided better forecasts than either analysts or models.

The superiority of analysts and managers, when compared with naive and sophisticated models, was theorized to be a function of three factors. First, they were assumed to have better knowledge of current EPS. Second, management had access to inside information and used it in their forecasts. Third, management had some control (presumably through accounting policy choices) over the actual earnings number reported.

While substantial empirical support was provided that analysts and managers performed better than mechanical models, Armstrong and Beuchert, [1979 p. 14], asserted that a combination forecast technique was superior to forecasts provided by any single source. This combination forecast was called an amalgamated forecast by the authors, and they stated that, "An amalgamated forecast based 80% on analyst forecast and 20% on extrapolations provided the optimal forecasts." The authors concluded that considerable research remained to be performed in this area.

The body of research cited to this point has compared the accuracy of consensus forecasts with mechanical models. Brown and Rozeff [1980] tested the abilities of individual forecasters to outperform mathematical models. The authors compared forecasts submitted by <u>Value Line</u> analysts with forecasts generated by the Box-Jenkins autoregressive-integrated-moving-average class of models. Comparisons with a sixteen quarter series of forecasts (1973-1976) revealed that 10 of 11 analysts produced superior forecasts when superior abilities were defined as a smaller average of forecast errors. The sole exception was an analyst whose performance was virtually indistinguishable from the Box-Jenkins model [1980, p. 33].

This exploratory study provided only preliminary evidence concerning the abilities of individual analysts. The authors suggested their preliminary results warranted further research in the area of individual analyst performance.

Consensus forecast accuracy was the subject of a study performed by Collins and Hopwood [1980]. The authors used a multivariate analysis-of-variance technique to address the limitations of previous empirical research. Univariate models with past earnings identified as the sole parameter were the mechanical models used in previous studies. Collins and Hopwood [1980] summarized these models and presented compelling evidence that univariate model choice should remain an area of active interest, since a superior model which used only past earnings as a parameter had not been identified. The authors, however, also criticized these models, stating that the univariate models neglected additional public information that might have been potentially useful [Collins and Hopwood, 1980, p. 392].

The authors criticized previous research efforts for using univariate analysis when the multiple model and multiple time period factors indicated that a multivariate hypothesis was being considered [1980, p. 393]. The use of multiple time periods violated the independence of earnings assumed by univariate analysis. The authors asserted that the problems of combined reliability and statistical dependence may have affected the empirical findings and the resultant conclusions [1980, p. 394].

The current study provides evidence of one additional confounding factor: choice of error metric may produce inconsistent results in comparative studies.

Collins and Hopwood chose to overcome the problems of reliability and dependence by comparing analysts' forecasts of earnings with forecasts generated by mechanical models in a multivariate analysis-of-variance (MANOVA) design. The results of the MANOVA analysis indicated a statistical difference existed between time-series models and analysts' forecasts of earnings. Confirming earlier work, the authors concluded that analysts were superior forecasters. Preliminary evidence was also presented indicating that analysts' forecasts improved over time.

Branch and Berkowitz [1981] criticized earlier efforts for sample selection techniques which included only widely followed companies with relatively continuous histories of operation [1981, p. 215]. The authors tested Business Week annual earnings forecasts, citing the practice of this publication of using Standard and Poor's Earnings Forecaster predictions as the source of raw data. Thus, the Business Week forecasts implicitly included a greater number of forecasters for each firm, although the authors conceded their sample was also weighted toward larger, more widely followed firms [1981, p. 216].

This study confirmed the results and conclusions of previous research efforts. The authors also asserted that forecasts explained substantially less than 10 percent of the interfirm variation in per-share earnings changes [1981, p. 218]. While analysts' forecasts explained only a small proportion of the variation in year-to-year earnings, they were generally superior to time-series extrapolations [1981, p. 219].

Bhaskar and Morris [1984] performed the comparison of analysts' forecast accuracy with naive models for a group of firms operating primarily in the United Kingdom. Their conclusions were similar to those reported in previous studies. The authors noted that while analysts tended to outperform naive models, they also tended to underestimate future profits in the United Kingdom. One explanation for this finding is that profit forecasts are required for stock offerings in the United Kingdom, and conservative profit estimates would be less likely to adversely affect share prices in the event that error was large.

In summary, this body of research provided considerable evidence that analysts outperformed mechanical models. Yet, in every instance in which forecast agents were compared, selection of the error metric may have affected the results. None of these studies provided evidence which suggested that alternative error metrics produced consistent results. Studies comparing analysts' accuracy to that of management provided similar insights and are reviewed in the next section.

Relative and Absolute Accuracy of Analysts' Forecasts Compared with Management's Forecasts

A review of analysts' accuracy relative to that of management provides few additional insights into the rationale for error metric selection. Included in this review is one of the first examinations of the reliability of management's forecasts of earnings, performed by McDonald [1973].

Of interest to this study, as well as subsequent empirical studies of management forecasts, is the question of self-selection bias. Since management's forecasts of earnings are not mandatory disclosures, it is possible that only those firms with an above-average ability to predict earnings made their predictions public. Thus, the results of these studies must all be viewed as potentially biased and not generalizable to the population of all firms.

McDonald tested the absolute accuracy of forecasts of management and noted a tendency of managers to overpredict rather than underpredict earnings. Additional evidence was provided to support the hypothesis that the utility industry's managers were more accurate in predicting earnings than the other industries tested.

While McDonald [1973] did not address the question of relative accuracy, Copeland and Marioni [1972] did test management forecasts relative to six naive models. Their results indicated that management forecasts were superior to naive models. Other comparative studies included Lorek, McDonald and Patz [1976], who compared management forecasts to those generated by Box-Jenkins techniques. Their results indicated that managers did not outperform the firm-fitted time-series models. However, the authors suggested that sample selection and the publication date of the forecast may have affected the results; that is, self-selection bias was present, and a consistent time period for managerial forecasts, relative to quarterly earnings publication, was not apparent.

Basi, Carey and Twark [1976] tested the relative accuracy of management forecasts compared with the accuracy of financial analysts' forecasts. They suggested forecasts would be more accurate when: 1) earnings were more stable; 2) firms were larger, older and less risky; 3) information provided was more detailed; and, 4) time until the actual announcement date was shorter. The authors used pair-wise tests to compare proxies of the variables which were hypothesized to affect forecast accuracy. Basi, Carey and Twark, [1976, p. 253], also included measures of absolute accuracy such as, "... more than 70 percent of the estimates by both analysts and executives were within 10 percent of actual figures." Conclusions concerning relative accuracy included "... forecast accuracy does not appear to be highly impressive for either group."

In a criticism of this study, Albrecht, Johnson, Lookabill and Watson [1977] suggested that pair-wise comparison of hypothesized variables that affect accuracy was inappropriate. Interactions among variables were ignored. For example, pair-wise comparison could be confounded if analysts' forecasts were published before those of management and were incorporated into those of management.

This issue was also addressed by Ruland [1978], who investigated the relative accuracy of management and analysts, and concluded that no significant differences existed between management and analysts' forecasts. Additional evidence was presented supporting the superiority of both management and analysts when compared with a simple extrapolation model. Ruland controlled the publication dates of forecasts and noted that:

> Both management forecasts and analysts' forecasts prepared subsequent to the release of these management forecasts are superior to those developed using the simple extrapolation models. Analyst forecasts reported prior to the announcement of management forecasts were not significantly more accurate than those of the simple naive model [1978, p. 439].

Jaggi [1978] performed similar research and concluded that management forecasts were more accurate than analysts', especially when analysts' forecasts were released before those of management. An analysis of forecast error by industry and by firm size indicated that industry was a significant factor in the accuracy of management forecasts but that firm size was not.

Somewhat contrary results were presented by Barefield, Comiskey and McDonald [1979], who replicated and extended the studies performed by Basi, Carey and Twark [1976] and Ruland [1978]. The authors extended the period of analysis and controlled the length of the forecast interval for both analysts and management. In most cases, management did not perform better than analysts. However, when the sample data were pooled, management performance appeared to be superior to that of analysts' [1979, p. 111].

Jaggi [1980] investigated the impact of firm size and industry classification on two measures of analysts' forecasts, Value Line Investment Service and Standard and Poor's Earnings Forecaster, in a replication and extension of Jaggi [1978]. The same conclusions were drawn, and only industry was considered a factor in the accuracy of management forecasts.

Additional studies included Porter's [1982] investigation of the determinants of management forecast errors. Also, Imhoff and Pare [1982] compared forecasts provided by management, analysts, and four firm-fitted Box-Jenkins models. No significant differences were noted among forecast agents.

A more recent test of management forecast accuracy was performed by Schreuder and Klassen [1984]. This effort differed from previous studies in one important area: management forecasts were provided from confidential sources rather than voluntarily published. While the self-selection bias noted in previous works may still have been a factor in this study, the authors postulated that the research methodology they employed had effectively reduced this bias.

The results of comparative accuracy tests indicated that management and analysts were not significantly different forecasters, confirming results noted by previous studies.

Brown, Foster and Noreen [1985, p. 150] provided an excellent summary of the reasons for differences between the forecasts of these groups. First, the information sets utilized by each of the forecast groups may be considered to be ordered sets. Mechanical models incorporated historical or annual series of earnings only. Analysts are presumed to have this same information in addition to a broader information set addressing macro-economic forecasts, the competitive structure of a firm's industry, and other factors in the public domain. Managers can incorporate all information in the public domain and have access to internal information. The failure to empirically support superiority of management forecasts in a conclusive manner may have been due to the interaction between management and analysts and the interdependencies of these two groups.

The time at which forecasts were made also may have affected the comparative analysis. While the later studies seemed to control for analysts' use of management forecasts and management's use of analysts' forecasts, complete control of this variable would seem to be impossible. Frequent interactions between these groups may or may not have occurred [1985, p. 150].

In summary, these research efforts have provided a comparative analysis of forecast agents based on different definitions of forecast error. The results from these studies, as well as additional efforts which have employed specific error metrics, may not have provided an appropriate assessment of analysts' abilities to forecast.

For example, the assumption of either a linear or nonlinear loss function, underlying the selection of a linear or nonlinear error metric, may have affected empirical analysis. This factor may explain the inconsistent ability of management to outperform analysts.

The limitations of these efforts may be summarized by one important factor. Error metrics which were neither theoretically supported nor empirically tested were selected. Thus, the results of these studies, and the conclusions drawn from these efforts may have been affected by the choice of error metric form. Chapter IV provides one method by which this assertion may be tested, and Chapter V presents the results of these tests which support the view that metric form selection affects the results of comparative studies.

A review of the error metric forms which have been employed in previous studies is provided in the following section. The limitations of error metric selection are also discussed. The last subsection of this paper presents a summary of those studies in which the relationship between forecast error and capital market risk has been tested.

Properties of Forecast Error and Limitations of Previous Research Efforts

The results and conclusions of these early research efforts would, in some cases, be difficult to replicate, compare, and interpret. Extensions of these studies and subsequent research of the properties of forecast error have revealed several limitations and potential biases. For example, most of the previous studies seemed to implicitly assume that all analysts were forecasting the same earnings variable, that is, primary earnings per share (PEPS). Yet, few studies explicitly confirmed this as the variable being forecast in a consistent manner across analysts. Additional analysis has focused on potential bias of forecast errors, and the time-series behavior of these errors. However, of major concern to this study is the choice of error metric utilized in empirical analysis.

Choice of Error Metric

Barefield and Comiskey [1975] were among the first to address error metric choice. Noting the choice of error metric should depend upon user loss functions, they stated:

> Since little is known about the nature of the loss function associated with earnings forecast errors, the mean absolute error has been selected due to its simplicity and also to its use in previous studies of earnings forecast errors [1975, p. 243].

The issue of investor loss function was further discussed by Brandon and Jarrett [1977], who suggested that the measure of absolute accuracy should reflect the consequences associated with forecasting errors which result from predictions that are not equal to their realizations [1977, p. 39]. In the investment setting, these consequences may be related to a gain or loss achieved by an investor who used the predictions to purchase securities.

In an efficient market [Fama, 1970], investors are price protected. They are rewarded for bearing risk which cannot be eliminated through formation of a diversified portfolio. Forecast error may be viewed as useful in predicting future security risk. However, alternative error metrics may produce risk predictions of varying degrees of accuracy. In this setting, choice of an error metric may cause users to assume either more or less risk than was intended.

Investor loss may be a linear function proportional to the size of the forecast error. Alternatively, the loss function may be nonlinear, implying that large errors have proportionally more serious consequences than moderate or small errors. Chapter III provides an analysis of user loss, and defines alternate error metrics which correspond to user loss.

Brandon and Jarrett [1977] included an analysis of five common linear and nonlinear error metrics. Their results led the authors to conclude:

> The measures were not consistent in the ranking of accuracy. The choice among measures of accuracy could influence the determination of the degree of accuracy. Thus, these measures are not necessarily interchangeable [1977, p. 43].

The current study provides an assessment of error metric selection in terms of risk, and considers underlying loss functions in the analysis of error metrics. Ranking of accuracy is one method by which error metrics can be tested for significant differences. If ranking is altered, due solely to error metric selection, then risk assessments will be affected by the error metric selected. Thus, the current study extends the analysis performed by Brandon and Jarrett, and suggests that one form of error metric may be most highly associated with security risk.

Additional problems associated with error metric choice, and an analysis of error metrics were provided by Brown, Foster and Noreen [1985]. The authors noted interpretational difficulties associated with the effects of sample outliers, and the effects of error metric definitions that allowed denominators to be negative [1985, p. 52].

Forecast outliers may be the result of factors which are specific to one single analyst. If one analyst provides a forecast which is extreme, forecast distributions will be skewed, thus, the consensus mean may not provide the best estimate of earnings for a firm.

Negative denominators affect interpretation only in the event that the numerator is also negative. For example, if forecast error were defined as Actual EPS less the Forecast EPS expressed relative to the Actual EPS, then instances could arise such as the following. If the Actual EPS was equal to -$1.00 and the Forecast EPS was equal to $.50, then, forecast error would be computed as follows:

Forecast error= (-1) - (.5) / (-1) = 150%

In this situation, an underestimate has occurred. Yet, the forecast error is positive, implying that an overestimate occurred. Thus, problems of interpretation are noted with error metrics which do not constrain the denominator to be positive.

Negative error metrics introduce a confounding factor into analysis of grouped data. For example, two analysts may provide forecasts which are inaccurate. Analyst A overestimates earnings per share by 100%, while analyst B underestimates earnings per share by 100%. An analysis of grouped forecast error would indicate that mean error was equal to 0% (e.g., (-1) + (1) / 2 = 0). If the error metrics were constrained to yield positive results, 100% error would result (e.g., (1) + (1) / 2 = 100%).

These problems have been addressed in previous studies using a variety of analytical techniques. Outlier effects on inferences drawn about forecast errors may be reduced by using an analysis technique that does not totally rely on the distribution mean, or by truncating the sample.

The effects of negative denominators have frequently been addressed by eliminating those data points from the sample, assuming their effects to be inconsequential in the analysis, or using a metric incorporating absolute value in the denominator.

The current study presents results for a randomly selected sample, and, to reduce the effects of outliers, a truncated sample. The truncated sample was defined in a manner which reduced the skewness of the distribution of error metrics without reducing the sample size below 500 firms. Chapter IV provides the technique which was employed.

In response to problems associated with negative denominators and negative error metrics, all error metrics in this study were defined using the absolute value operator in both the numerator and the denominator. Chapter III provides mathematical definition of each error metric used in hypotheses tests.

A summary of the more commonly used error metrics, and an analysis of limitations concerning each, is presented in the equations and paragraphs which follow. (Note that each error metric was computed for each year of the designated time period, then averaged over the time periods involved).

<u>Mean Absolute Percent Error (MAPE)</u>

$$MAPE = | P_i - A_i | \ / \ P_i \qquad (2.1)$$

where:

P_i = Mean forecast EPS

A_i = Actual EPS

<u>Total Mean Error (ME)</u>

$$ME = (P_i - A_i) \tag{2.2}$$

<u>Mean Absolute Error (MAE)</u>

$$MAE = | P_i - A_i | \tag{2.3}$$

<u>Relative Mean Error (RME)</u>

$$RME = (P_i - A_i) / A_i \tag{2.4}$$

<u>Relative Mean Absolute Error (RMAE)</u>

$$RMAE = |P_i - A_i| / | A_i | \tag{2.5}$$

<u>Quadratic Mean Absolute Error (QMAE)</u>

$$QMAE = (P_i - A_i)^2 \tag{2.6}$$

where in all cases, the metric was computed for each year of the designated time period, then averaged over the time periods involved.

It is interesting to note the differences in error metrics, in view of the criticisms offered by Brandon and Jarrett [1977] and Brown, Foster and Noreen [1985]. All of the metrics are defined in terms of the distribution mean;

thus, they are subject to the effects of outliers. Only RMAE defines the denominator in absolute value terms; thus, the other measures are subject to the effects of negative denominators. Most of these measures assess accuracy relative to actual EPS; yet, each could have been measured relative to forecast EPS. Finally, all of the metrics implicitly assumed a linear investor loss function, with the sole exception of QMAE.

While this series of metrics is not exhaustive, it does represent the more common measures used in previous studies. In fact, RMAE was the most common metric used in the studies cited. A review of the metrics used, and the empirical results and conclusions of these studies, provides evidence that comparability and interpretation of results and conclusions may be difficult.

Barefield and Comiskey [1975] defined forecast error using MAPE, and reported an average forecast error of 16.07 percent, relative to the forecast, for the years 1967-1972. Brandon and Jarrett [1977] reported error statistics for ME, MAE, RME, RMAE, and QMAE of 9.4%, 24.7%, 13.4%, 20.3%, and 26.5%, respectively. These results are not directly comparable to those provided by Barefield and Comiskey [1975] due to the differing definitions of the denominators. Further, the time periods utilized in the two studies span the early 1970s, a time period in which other economic factors may have affected forecast error.

RMAE, the most commonly used metric, was reported by Richards, Benjamin and Strawser [1977] as 24.1% for the years 1972-1976. Brown and Rozeff [1978] also utilized this metric in empirical analysis, but did not report these results separately. Jaggi [1978] used this metric as the comparative statistic for analysts' versus managements' estimates, reporting 28.3% analyst error for the years 1971-1974. Bhaskar and Morris [1984] reported similar results (16.3% for the years 1970-1974), as did Crichfield, Dyckman and Lakonishok [1978], and Collins and Hopwood [1980].

In a more recent study, Elton, Gruber and Gultekin (EGG) [1984] investigated the size and pattern of analysts' errors. The authors analyzed 414 December 31 year-end firms collected from the Institutional Brokers Estimate System (IBES) data base for each of the years 1976-1978 [1984, p. 2]. (The IBES data includes forecast information for approximately 2800 firms. While the underlying distribution of analysts' forecasts for each firm is not available, summary information about each firm is provided. For example, forecast mean and median statistics are provided, in addition to the highest and lowest estimates, and the standard deviation of estimates. Chapter IV presents a complete description of this data base.)

Three error metrics were employed by EGG. The first was simply the absolute dollar value difference between actual and forecast EPS. The second measured the error in estimated growth, while the final metric was Theil's Inequality Coefficient [1966]. Similar results were noted for each measure.

As is evident from the preceding summary, choice of error metric affected comparability of the results of previous studies. Additionally, assumptions concerning the form of the user loss function affected interpretation. In the next subsection, issues of alternate inputs to error metrics are discussed.

Definition of the Internal Forecast Parameter

The choice of the forecast parameter used in the error metric has not been addressed in previous studies. Yet, the assumption of either a linear or a nonlinear user loss function affects, in concept, this choice (although, in practice the choice may have no impact). If loss functions are assumed to be linear, minimum cost of error analysis argues against use of the forecast mean (the most common statistic used in the studies reviewed) as the internal forecast parameter. Chapter III will discuss this issue further, and will provide a basis for this position.

Use of an observed empirical relationship between forecast error and systematic risk provides one method by which alternative error metrics can be evaluated. Tests of the degree of association between forecast errors and market beta impound the perceptions of market participants regarding loss functions and error metrics.

The Relationship Between Forecast Error and Risk

Use of accounting information to predict and assess systematic risk has been the subject of many studies. Beaver, Kettler and Scholes [1970] were among the first to test for a relationship between earnings variability and risk, and this relationship has been identified numerous times. Yet, a prevailing view has been that risk is created by an inability to predict earnings, not earnings variability *per se*.

Comiskey, Mulford and Porter [1986, p. 261] suggested the pertinent theoretical concept:

> The fundamental theoretical concept which motivates the work is that shareholders are rewarded only for bearing risk which cannot be eliminated through formation of a diversified portfolio. In terms of accounting earnings, risk should be determined by forecasting difficulty and not simply historic (or prospective) variability. Further, it should only be forecasting difficulty (or forecast error) that cannot be eliminated through diversification (systematic forecast error) which should be rewarded with higher return and hence be associated with systematic security risk (i.e., market beta).

The causal link suggested by the authors implied that systematic, and not total, forecast error should be the measure of risk for which investors are to be rewarded.

The concept that total forecast error may be viewed as a measure of risk has been the subject of other reseach efforts. Earlier, Barefield and Comiskey [1975a] suggested that theory posits an inverse relationship between earnings variability and share prices; yet, empirical support for this relationship had been weak. One explanation proposed by the authors was that earnings variability did not measure risk, but was simply a surrogate for the underlying events which constitute risk [1975a, p. 315].

Elton and Gruber [1972a, p. 316] also postulated that risk is more closely associated with forecast error. Citing previous research, they stated:

> Most authors have defined risk in terms of earnings instability. Yet a company with a regular and predictable pattern of cyclical earnings is not risky *per se*. The formulation of risk in terms of inability to predict is much more in keeping with the postulates of subjective risk assessment.

Thus, as proposed by Barefield and Comiskey [1974, 1975a], and others, the risk of a firm is related to an earnings surprise; that is, risk may be viewed as being partially composed of factors which contribute to forecast inaccuracy (and vice versa). Forecast error may be viewed as a surrogate for risk.

Barefield and Comiskey [1974] tested the association of earnings variability and forecast error with market beta. Their results indicated a high degree of association between forecast error and earnings variability (with a lower degree of association noted for forecast error and beta). Further, it was revealed to be more difficult to forecast earnings of firms with greater historical earnings variability [Barefield and Comiskey, 1974, p. 321]. A more surprising result was that forecast error exhibited about the same degree of association with systematic risk (market beta) as did historic earnings variability (Spearman Rank Correlation Coefficients of 0.513 and 0.519, respectively).

This finding served as the basis for two subsequent efforts in the area. In the first, Barefield and Comiskey [1979] refined the methodology used; in the second, Comiskey, Mulford and Porter [1986] decomposed forecast errors into systematic and nonsystematic components, thus providing an estimate of forecast error variability more conceptually analogous to systematic risk.

Barefield and Comiskey [1979] proposed a divergence measure to overcome the circumstance where forecast error was effectively a surrogate for earnings variability (and vice versa) due to extremely high cross-sectional correlation between the two measures. Divergence in rank was used to partition the sample into groups in which forecast error was not a surrogate for historic earnings variability (high divergence). For high divergence firms, the degree of association between forecast error and market beta should exceed the degree of association noted in lower divergence firms. This hypothesis was empirically supported. The authors concluded:

> When a firm's earnings are either more or less forecastable than they are variable (either positive or negative divergence measures of relatively larger size), then the systematic risk of the company's common stock tracks forecastability more closely than variability [1979, p. 7].

Yet, the authors expressed concern that forecast error was a total risk measure, while market beta measures only a systematic component. This issue was addressed by Comiskey, Mulford and Porter [1986]. In this study, the authors tested the degree of association between systematic risk (market beta) and systematic forecast error (forecast error beta), and confirmed their hypothesis that forecast-error beta should exhibit a significantly higher degree of association with market beta than accounting beta. Other studies have provided similar results (e.g., Beaver, Clark and Wright [1979]).

In summary, the results of these studies suggest that forecast error may be viewed as a surrogate for security risk, and offer one method by which error metrics may be evaluated. If one form produces forecast errors which are more highly associated with market beta, then that form may best represent security risk.

Chapter Summary

Chapter II provided a review of the literature which provides insights into and suggestions for the current study. The first general group of research efforts which were reviewed compared the accuracy of analysts to mechanical models or to management. From these studies, limitations, such as error metric definition, were noted. The second general group of studies provided the foundation for subsequent hypotheses tests.

Comparisons of analysts' accuracy with the accuracy of mechanical models, or with the accuracy of management, was the focus of many previous research efforts. Yet, few efforts provided justification for the error metrics selected in their analysis. No previous effort provided either a theoretical basis for use of an error metric, or an analysis of results and conclusions using a variety of error metrics. The current study provides both a theoretical justification for two forms of error (see Chapter III), and an analysis of the effects of alternative error metrics on the results and conclusions of previous efforts.

The current study employs the results of the second general area of this literature review (in which the relatioship of forecast error to risk was established) in an analysis of alternative forms of error. If forecast error may be viewed as a surrogate for security risk, then the effects of alternative definitions of error should be assessed. If significant differences are noted among alternative error metrics in risk estimation, then the relationship identified between forecast error and systematic risk is to be employed to determine if one form of error exhibits a higher association with systematic risk.

Chapter III provides the theoretical basis for two forms of error, and defines all error metrics which will be employed in hypotheses tests. Chapter IV presents the hypotheses tests, and related research methodology. Chapter V reports the results of the hypotheses tests, and Chapter

VI suggests interpretation of and limitations pertaining to this study.

CHAPTER III

ERROR METRIC DEFINITION

Each of the previous studies presented some form of error metric which was used to compare forecast agents. Additionally, many of the efforts suggested that error metric choice should correspond to a concept of user loss. Yet, this concept was not carefully linked to the metric(s) actually employed, and metrics appropriate for use under alternative circumstances were not considered.

This chapter presents a general explanation of the concept of the user. Also, a basis for error metric definition under alternative user loss function circumstances is provided. Additionally, definitions of the internal forecast variable are addressed. The chapter concludes with the definitions of alternative error metrics which are empirically tested.

Error Metric Definitional Form

Previous efforts have suggested that the form of the error metric should correspond to a concept of a user loss function. A user loss function describes the relationship between forecast error and loss associated with use of a forecast which is typically subject to error.

In these studies, users were assumed to be investors or creditors. This assumption is appropriate in most cases; however, in the context of analysts' forecasts, at least two other groups of users should be considered.

Analysts are both producers and consumers of forecasts of earnings; forecasts are both products to be sold and are employed as inputs into models used to produce forecasts of stock prices. Managers may incorporate analysts' forecasts in assessments of their own forecasts, and may also use analysts' forecasts in real investment decisions.

For the analyst, if forecasts are considered to be products, gain or loss may be considered a function of forecast accuracy. That is to say, across time, increased accuracy generally results in a gain, while consistent over- or underestimates may be associated with losses. This loss is related to the reputation of the forecaster. (In a rational market, purchasers will discount the value of an analyst who consistently produces large errors.)

The loss function associated with inaccuracy for the analyst may be linear, in which loss is directly proportional to the size of the error, or nonlinear, in which larger errors result in more than proportional penalties. Additionally, the functional relationship may be symmetrical, in which losses resulting from overestimates equal losses resulting from underestimates, or asymmetrical, in which over- and underestimates result in different levels of losses. Of equal importance is the threshold at which loss is realized. Small errors may result in no loss, while larger errors result in some level of loss. Figure 3.1 provides an illustration of these types of functions.

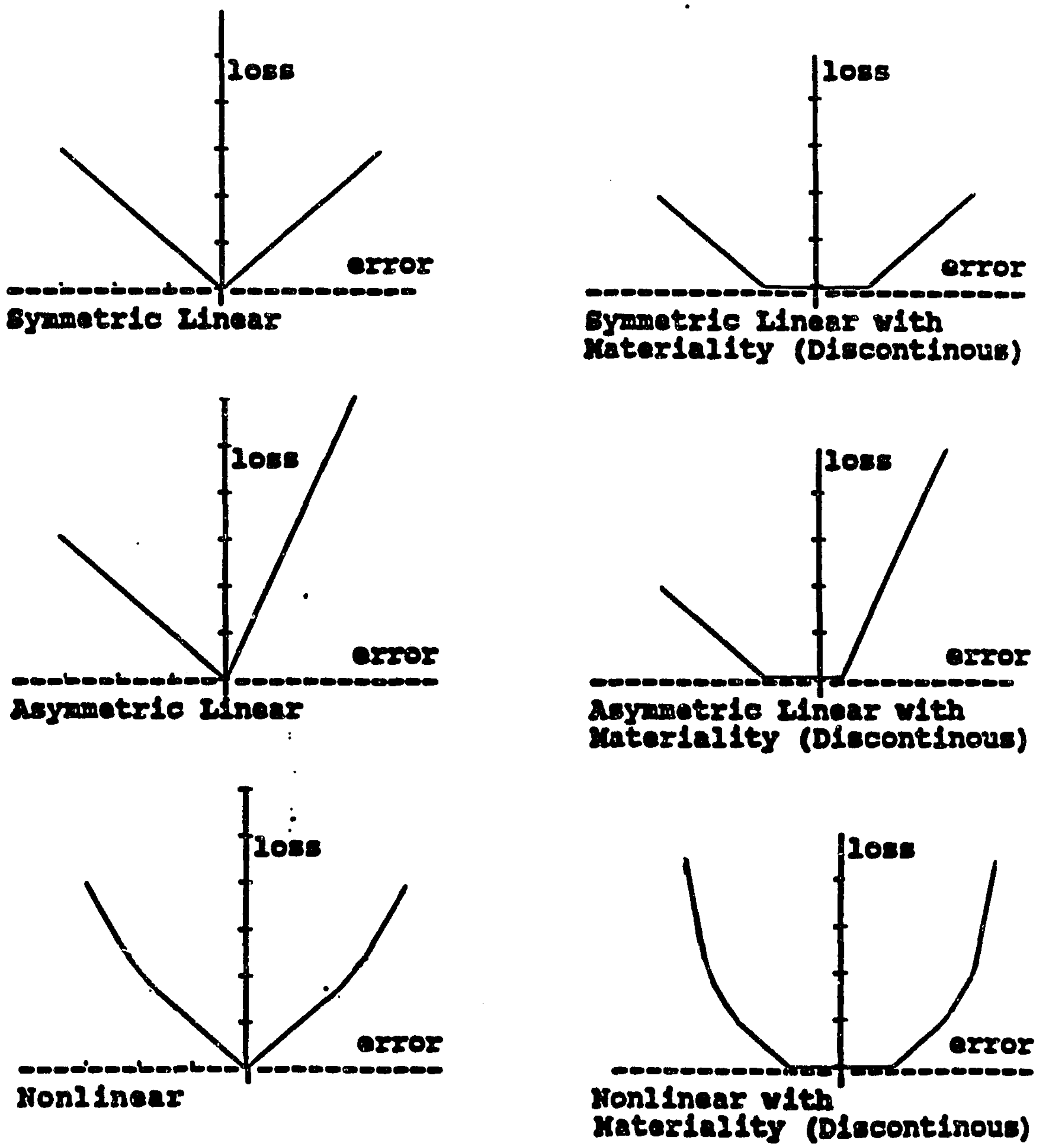

Figure 3.1 Representative Loss Functions

All of the representative loss functions express loss as a function of forecast inaccuracy. Both positive and negative forecast errors result in losses, and forecast accuracy does not result in a loss. Accuracy is not associated with a gain for the user.

Threshold levels suggest discontinous functions because forecast inaccuracy may not produce losses for small errors. Since inaccuracy is not associated with gains, the loss functions are discontinous.

A practical example of this type of function is one in which forecasts of sales are employed in an inventory ordering system. Overestimates of sales result in losses related to the carrying costs of unused or unsold inventory. Underestimates of sales result in losses related to stockouts and lost sales. Forecast accuracy or small errors do not result in losses for the user of the forecast.

Note that, in this setting, forecast accuracy does not result in a gain. Instead, accuracy is associated with minimized levels of loss.

This analysis extends to managers who may employ analysts' forecasts of earnings in real investment decisions, and the underlying financing decisions relating to the investment. Estimates of earnings may be employed in models which provide estimates of cash flows. Overestimates of earnings result in losses if the financing decision relied on these or related estimates of cash flows to repay loans or finance the investment. Underestimates produce losses, in the sense of an opportunity loss, since investment and financing may be limited by the estimates of cash flows. Forecast accuracy or small levels of inaccuracy do not produce significant loss for the user.

The specification of the analyst's or the manager's loss function requires knowledge of forecasting models, of investment models, and of levels of loss associated with different levels of forecast error. At this time, we have no pertenient research findings which provide insights into these functions. Thus, further analysis of these functions is not possible.

However, the concepts of linear versus nonlinear and symmetric versus asymmetric functions may be applied to help analyze the loss function for the third primary group of users: creditors and investors.

Creditors may employ analysts' forecasts of earnings in loan decisions. This situation is similar to that of managers who employ forecasts of earnings in estimates of cash flows. For the creditor, overestimates of earnings produce losses if these estimates are employed in estimates of cash flows, and cash flows are not adequate to service the loan. Additionally, overestimates may result in an underpriced loan. Underestimates produce losses, in the sense of an opportunity loss, if the forecasts of earnings limit the loan amount or preclude its initiation.

Investors (in addtion to analysts and managers) utilize analysts' forecasts of earnings as inputs to the process of predicting share prices. For example, two models which employ estimates of earnings to derive share prices are the Whitbeck-Kisor Model [Whitbeck and Kisor, 1963], and the Wells Fargo Model [Fouse, 1976]. In both of these valuation models, estimates of earnings are employed with other variables to determine share price.

Niederhoffer and Regan [1972] provided evidence that share prices are dependent on both earnings changes and analysts' forecasts of earnings. In a study of the 50 best-performing, and the 50 worst-performing stocks for the year 1970, the authors provided evidence that firms which registered the highest increases in share price (best-performing) were characterized by substantial analyst underestimates. Conversely, the worst-performers were those firms which had been substantially overestimated.

At the individual security level, forecast inaccuracy results in a loss. When earnings forecasts are employed in stock valuation models, forecast error yields over- or underestimates of share value. In the event share prices were overestimated, two types of loss may result. For the investor who purchased the security based on the analysts' forecast of earnings, a real loss occurs around the date of the actual earnings announcement, when markets adjust (reduce) the price of shares to reflect earnings. An opportunity loss results for the investor who, given perfect information, would have sold the securities short.

In the event share prices were underestimated, the reverse situation occurs. Loss is in the form of an opportunity loss for the investor who would have purchased more of the securities given perfect information on earnings. A real loss is incurred by the investor who sold the securities short.

These relationships were noted in previous research efforts. Ball and Brown [1968, p.175] suggested that:

> If an investor knew the sign of the change in earnings per share twelve months in advance of its public release, he could earn an abnormal return of 8.3% by investing long in positive earnings change firms and selling short in negative earnings change firms.

At the individual security level, forecast inaccuracy is associated with loss, and the related risk estimate is total risk. However, the characteristics of individual securities are more important in terms of their effect on the distribution of a portfolio return. In this context, the related risk estimate of forecast error is systematic risk.

Portfolio theory, in the context of an efficient market [Fama, 1970], suggests that investors are price protected. They are rewarded for bearing risk which cannot be eliminated through formation of a diversified portfolio. Security risk estimation is crucial for portfolio formation, in the sense that investors are assumed to be risk averse, and to form portfolios which correspond to intended levels of risk.

Forecast error may be viewed as a surrogate for risk. As suggested by Elton and Gruber [1972a, p. 316], the inability to predict earnings is more consistent with the postulates of subjective risk assessment than earnings instability (risk is not created by a predictable pattern of unstable earnings).

Thus, at the individual security level, forecast inaccuracy affects loss, and may serve as an estimate of total risk. At the portfolio level, forecast error may be viewed as a surrogate for systematic risk. Further, the relationship between forecast error and systematic risk may be used to infer the properties of the investor loss function.

At this time, there is little pertenient empirical evidence concerning this loss function. As was previously noted, the function expresses the relationship between forecast error and losses from decisions which utilized the forecast. This relationship cannot be observed. However, the relationship between forecast error and systematic risk may be employed to infer the general properties of this function.

One major issue associated with the concept of the investor loss function is the shape of the function. If the loss function is assumed to be linear, then error metrics which are defined in linear terms are appropriate for use. If the loss function is assumed to be nonlinear, then the error metric should reflect that relationship.

One purpose of this study is to empirically compare error metrics which correspond to linear versus nonlinear investor loss functions. If, initially, the assumption is made that the function expresses a symmetric relationship, then previously defined error metrics provide a starting point for analysis.

Assumption of a symmetric loss function does not impose an unrealistic constraint. For example, if an investor employs analysts' forecasts in estimating returns, and makes decisions to buy or to sell short based on this estimate, losses will result, at the individual security level, and at any level of forecast inaccuracy. The presumption that the cost of error is the same for both long and short positions provides a basis for metric estimation and analysis.

This study compares the more commonly defined linear and nonlinear error metrics. Of equal importance is the definitional form of the internal variables. Specifically, the forecast statistic employed in the error metric must also correspond to the general linear and nonlinear cases. The next subsection discusses the choice of this forecast statistic.

Definition of the Internal Forecast Statistic

Previous efforts have identified the forecast statistic as the mean forecast for a group of analysts. The concept of user loss function suggests the cost of forecast error should be a major determinant in metric definition. This same cost should be considered when defining the forecast statistic incorporated into the error metric.

Hamburg [1983] argues that certain consensus-forecast measures are more appropriate when loss functions are linear than when these functions are nonlinear. If consensus earnings forecasts are predictions of an observation picked at random from the distribution of all analysts' forecasts for a firm, and the cost of error is both symmetric and linear, then the forecast median should define the internal forecast parameter.

Hamburg's analysis rests on the minimization of cost. In the situation where investor loss functions are assumed to be linear, that is, the cost of error varies proportionally with error size regardless of the sign of the error, and the distribution of analysts' forecasts is asymmetric, the minimum cost prediction would be the median.

Two basic assumptions are required for this analysis of error metrics which correspond to investor loss. The first assumption is that the investor loss function is symmetric, thus, over- and underestimates result in the same level of loss. The second assumption is that the underlying distribution of analysts' forecasts is asymmetric, therefore, the mean observation is not equal to the median observation. (In the event that the distribution is symmetric, the mean and median observations are equal, thus, differences in error metrics defined using the mean versus the median forecast do not exist.)

Implicitly, this analysis suggests that: 1) investors who face linear loss functions should employ the median forecast in applications which utilize the forecast amount of earnings; further, 2) error metrics which correspond to linear loss functions should incorporate forecast median as the internal forecast parameter.

Hamburg suggests that under the assumptions of a linear, symmetric loss function, and an asymmetric distribution of forecast observations, the least cost prediction would be the prediction which minimizes absolute error. In this case, the median forecast would minimize average absolute deviations, and the mean deviation about the median would be the measure of the minimum cost of error. This point is shown in the following analysis.

Let X_1, $X_2,\ldots,X_N$ define N observations of an asymmetrical distribution of analysts' forecasts such that:

1) $X_1 < X_2 < \ldots < X_N$

2) The median of the distribution may be defined as $M_d = X_{(N/2)} + X_{(N/2)} + 1 / 2$ if N is an even number, and,

$M_d = X_{(N+1/2)}$ if N is an odd number.

3) Assume two predictions are made at A and A' such that: $A < A'$, and neither is equal to M_d; $A' = X_{j+1}$; and, $X_j < A < X_{j+1} < M_d$ for the odd number case. (Note that the analysis as provided for the odd number case can easily be extended to the even number case.)

The cost function corresponding to a prediction equal to A is shown by equation 3.1.

$$\text{Cost}(A) = \sum_{i=1}^{j} (A-X_i) + \sum_{i=j+1}^{N} (X_i-A) \tag{3.1}$$

Now assume a second prediction at A', where, by definition, $A' > A$, but is less than the median value. The new cost function is expressed by equation 3.2.

$$\text{Cost}(A') = \sum_{i=1}^{j+1}(A'-X_i) + \sum_{i=j+2}^{N} (X_i-A') \tag{3.2}$$

This cost function may be transformed to equation 3.3 by adding the quantity (A - A) to both terms on the right hand side and rearranging terms.

$$\text{Cost}(A') = \sum_{i=1}^{j} (A-X_i+A'-A) + \sum_{i=j+1}^{N} (X_i-A-A'+A) \tag{3.3}$$

Removing the term (A-A') from the summation results in equation 3.4.

$$\text{Cost}(A') = \sum_{i=1}^{j} (A-X_i) + (j+1)(A'-A) + \sum_{i=j+1}^{N} (X_i-A) - (N-j-1)(A'-A) \tag{3.4}$$

Rearranging and collecting terms yields equation 3.5, which is expressed in terms of prediction A and a term which reduces cost as A' moves toward the median.

$$\text{Cost} = C(A') - \underset{>0}{[N-2(j+1)]}\underset{>0}{[A'-A]} \qquad (3.5)$$

By definition, A' > A, thus, A'-A > 0; also, j + 1 < N/2, thus, N-2(j+1) > 0; therefore, [N-2(j+1)][A'-A] is a positive number. In every case, as the prediction approaches the median, total cost is reduced. At A' equal to the median, cost is at the minimum value.

This same analysis may be made for values greater than the median, or for the case where N is an even number. Once again, the results would indicate that, assuming a symmetric underlying loss function and an asymmetric distribution of observations, use of the median results in a least cost prediction.

In the same manner, a special case of the Gauss-Markov theorem [Johnston, 1972] may be employed to establish the mean as the least cost predictor when the underlying loss function is quadratic. A quadratic function is representative of the class of nonlinear functions in which larger errors exact greater penalities.

Hamburg suggests if the cost of the error varies according to the square of the error, use of the mean results in a lower average of squared deviations about it than any other predictor. In this situation, the variance may be interpreted as the average cost of error per observation, and the average amount of error would be zero.

The cost of error may be minimized by minimizing the variance of the error term. The following analysis provides evidence that in the special case of the Gauss-Markov theorem in which $X_i = 0$, variance is minimized by using the mean value as the prediction.

Given a cost function, C, defined in the quadratic case (as shown in equation 3.6), the objective would be to select the value of Y_i which minimizes cost.

$$\text{Cost} = \sum_{i=1}^{N} (Y_i - A)^2 \tag{3.6}$$

Taking first (C') and second (C'') derivatives of 3.6 yields equations 3.7 and 3.8.

$$C' = -2 \sum_{i=1}^{N} (Y_i - A) \tag{3.7}$$

$$C'' = 2 \tag{3.8}$$

As described by equation 3.8, the function is at its minimum point when C''>0. The value of Y_i which minimizes the cost function is found by setting equation 3.7 equal to zero and solving for A. Equations 3.9 to 3.12 provide the solution.

$$C' = -2 \sum_{i=1}^{N} (Y_i - A) = 0 \tag{3.9}$$

$$\sum_{i=1}^{N} (Y_i - A) = 0 \tag{3.10}$$

$$Y_i = NA \tag{3.11}$$

$$A = Y_{mean} \tag{3.12}$$

Thus, the minimum cost prediction is defined by the mean of the distribution of observations, when the underlying loss function is defined as the special case of nonlinear loss, that is, the quadratic loss function.

Theoretical support is thus provided for two error metrics, depending upon assumptions of user loss. In the linear case, the Median Error forms were shown to provide an error form which best corresponds with notions of linear user loss. In the same manner, if user loss is considered to be nonlinear, the Quadratic Mean Error was shown to provide a measure of error best associated with notions of nonlinear user loss.

Other nonlinear expressions of error may also be considered. Fractional power expressions, such as the square root function, and logarithmic expressions of error metrics also provide forecast errors which are nonlinear. However, for the purposes of this study, the quadratic expression was considered to adequately represent the class of nonlinear error metrics.

Metric Definitions

This chapter has provided the basis for alternative definitions of the error metric. The discussion relating to user loss function concluded with two major computational forms, linear Median Error and nonlinear Quadratic Mean Error.

Previous studies have employed a variety of metric forms beyond the two theoretically based forms identified above. This study will compare the theoretically based metric forms with the more common measures incorporated into previous research efforts. A complete list of metric abbreviations and definitions follows.

(1) Median Error (F1)

This form of the metric corresponds to the general linear case. The definitional form is shown in equation 3.13.

$$F1 = |\text{ Median Forecast} - \text{PEPS }| \qquad (3.13)$$

where:

Median Forecast = Median forecast from the IBES data base.

PEPS = Primary EPS corrected for splits and dividends from the IBES data base.

This form and all subsequent forms were defined on a yearly basis, then averaged across five years.

(2) Mean Error (F2)

The second form of this metric employs the mean of the forecast in order to compare error metrics defined using forecast medians with those defined using forecast means. Equation 3.14 expresses this metric.

$$F2 = | \text{Mean Forecast} - \text{PEPS} | \qquad (3.14)$$

(3) Relative Measures of Error (F3, F4, F5, F6)

Previous efforts have also expressed error relative to either the forecast statistic or the actual earnings achieved. Relative metrics provide one means by which error may be compared across firms with earnings of different magnitudes. Expressing error relative to actual or forecast earnings should not affect cost of error minimization. Equations 3.15 through 3.18 provide the definitions of relative metric forms.

Median Error Relative to Actual EPS (F3)

$$F3 = \frac{|\text{Median Forecast} - \text{PEPS}|}{|\text{PEPS}|} \qquad (3.15)$$

Mean Error Relative to Actual EPS (F4)

$$F4 = \frac{|\text{Mean Forecast} - \text{PEPS}|}{|\text{PEPS}|} \qquad (3.16)$$

Median Error Relative to Forecast EPS (F5)

$$F5 = \frac{|\text{Median Forecast} - \text{PEPS}|}{|\text{Median Forecast}|} \qquad (3.17)$$

Mean Error Relative to Forecast EPS (F6)

$$F6 = \frac{|\text{Mean Forecast} - \text{PEPS}|}{|\text{Mean Forecast}|} \qquad (3.18)$$

(4) Nonlinear Measures of Error

The general nonlinear case will assume an underlying investor loss function which corresponds to a quadratic loss function. Again, error metrics may be expressed in absolute terms or in terms relative to either the forecast statistic or the actual reported earnings. Since cost of error analysis requires use of the mean forecast (for the nonlinear case), and no previous efforts have employed the median in assessments of error for the general nonlinear case, only the mean forecast will be employed in definitions of nonlinear metrics. Equations 3.19 through 3.21 provide definitions of these metrics.

Quadratic Mean Error (F7)

$$F7 = (\text{Mean Forecast} - \text{PEPS})^2 \qquad (3.19)$$

Quadratic Mean Error Relative to Actual EPS (F8)

$$F8 = \frac{(\text{Mean Forecast} - \text{PEPS})^2}{|\ \text{PEPS}\ |} \qquad (3.20)$$

Quadratic Mean Error Relative to Forecast EPS (F9)

$$F9 = \frac{(\text{Mean Forecast} - \text{PEPS})^2}{|\ \text{Mean Forecast}\ |} \qquad (3.21)$$

Two other forms of quadratic error, in which F4 and F6 are squared, are included in an analysis of the investor loss function. These forms are not included in nonparametric tests of rank, since squaring F4 and F6 only scale the results, thus, the results of nonparametric tests of rank are the same as F4 and F6. Chapter V employs these error metrics in a regression analysis, and provides both definitions of the metrics and an explanation of their use.

Chapter Summary

Chapter III defined user loss functions and identified the relationship between forecast error (over- or underestimates) and user loss. Forecast error was shown to affect loss, and to provide a proxy for security risk.

This chapter also provided theoretical support for use of two error metrics. In the event a linear function is assumed to represent user loss, the median forecast was shown to minimize cost of the error. In the event a nonlinear function was assumed to represent user loss, forecast mean minimized the cost of error. Seven other metrics were introduced due to their use in previous studies.

Chapter IV presents the framework for error metric analysis in terms of research objectives and hypotheses designed to test the objectives. Chapter V reports the results of the hypotheses tests, and Chapter VI provides interpretation of and limitations that pertain to this study.

CHAPTER IV

RESEARCH METHODOLOGY

The purpose of this study is to provide an analysis of error metric selection by: 1) testing the consistency of error metrics in assessing the relative accuracy of analysts compared with a simple mechanical model; 2) testing the effects of mean versus median forecast definition on linear error metrics; 3) determining if alternative error metrics change rank ordering of firms; and, 4) isolating which error metric, if any, is most closely associated with systematic risk. This chapter provides a detailed discussion of these objectives, presents the hypotheses, and outlines the means by which the hypotheses are tested.

Research Objective One

Research Objective One addresses the effects of error metric selection on the results of previous research efforts. The objective may be stated as:

1) To analyze the effects of error metric selection on conclusions drawn from previous studies of comparative forecast accuracy of analysts with a mechanical model.

With respect to Research Objective One, the following question is raised:

1) Do alternative error metric definitions provide

consistent results in comparisons of relative accuracy of analysts with a mechanical model?

Certain previous studies, notably Brown and Rozeff [1978] and Imhoff and Pare [1982], included comparisons of accuracy for analysts and mechanical models. In these studies, the Friedman Test was employed to provide an assessment of significant differences between forecast agents. Subsequent analysis of mean ranks was used to identify the superior forecaster; the agent with the lower mean rank was favored.

A Friedman Test performed in this manner has three possible outcomes: 1) there is no significant difference in forecast agents, thus, neither agent is favored; 2) there is a significant difference in forecast agents, and analysts are favored; and, 3) there is a significant difference in forecast agents, and the mechanical model is favored.

The current study tests the consistency of error metrics applied to this general setting. Consistency occurs when all error metrics yield the same results. For example, a consistent pattern results when, under all error metrics: 1) analysts are favored; 2) the mechanical model is favored; or, 3) there is no significant difference in the forecast agents.

<u>Hypothesis Statement Relating to Research Objective One</u>

Hypothesis One provides evidence of metric consistency (defined as the same agent being favored for all error metrics). The hypothesis may be stated as:

H_{o1}: Choice of error metric does not alter consistency of findings of comparative analyst/model forecast performance.

H_{a1}: Choice of metric form alters consistency.

Agents are defined, in this test, as analysts, and as a naive, no-change model. The naive, no-change model was selected due to its simplicity, and its use in previous studies. The no-change model represents a very limited class of models to which analysts have been compared. Since it is invariant to any change in earnings, use of this model may provide a setting whereby significant differences will be most easily observed.

The Friedman Test was used to test Hypothesis One. The purpose of this test is to determine whether there is any consistent relational pattern between the forecast agents. In applying this test, analyst error metrics and no-change error metrics were ranked across each firm. Over all firms, the mean rank of each error metric was computed. From these mean ranks, test statistics with Chi-square distributions were compared. The resultant Friedman Significance Levels indicated the probability that the two error metrics (analyst or no-change) came from essentially the same population. Table 4.1 provides an application of the Friedman Test to this study.

Table 4.1
Ranks of K Metric Forms

Firms (N)	F(1) Analyst	F(2) No-Change
Firm 1	(These columns contain the rank of each metric across K definitions of metric. For example:)	
.		
.		
.		
.		
Firm N-1	2	1
Firm N	1	2
Mean Ranks	$\overline{\text{MRF1}}$	$\overline{\text{MRF2}}$

where:

N = the total number of firms in the sample

K = related measures of forecast error (analyst or no-change)

MRF = the mean rank of forecast error

For example, F1 and no-change form 1 (NC1) directly correspond in metric definition with only the agent altered. F1 was defined as |Median Forecast - EPS|, and NC1 is defined as $|EPS_{t-1} - EPS_t|$ where t indicates the year. Similar definitions are documented in Table 4.2 for all metric forms across both forecast agents.

Table 4.2
Corresponding Analyst and No-Change Metric Forms

<u>Analyst Form</u>	<u>No-Change Form</u>
F1=|Median - EPS|	$NC1=\|EPS_{t-1} - EPS_t\|$
F2=|Mean - EPS|	
F3=|Median - EPS|/|EPS|	$NC2=\|EPS_{t-1} - EPS_t\|/\|EPS_t\|$
F4=|Mean - EPS|/|EPS|	
F5=|Median - EPS|/|Median|	$NC3=\|EPS_{t-1} - EPS_t\|/\|EPS_{t-1}\|$
F6=|Mean - EPS|/|Mean|	
$F7=(Mean - EPS)^2$	$NC4=(EPS_{t-1} - EPS_t)^2$
$F8=(Mean - EPS)^2/\|EPS\|$	$NC5=(EPS_{t-1} - EPS_t)^2/\|EPS_t\|$
$F9=(Mean - EPS)^2/\|Mean\|$	$NC6=(EPS_{t-1} - EPS_t)^2/\|EPS_{t-1}\|$

where all forms are computed yearly, then averaged across the time period.

A two-tailed test is appropriate since, *a priori*, there is no theoretical reason to expect one forecast agent to be favored. Failure to reject the two-tailed null would suggest that choice of error metric does not alter consistency. Rejection of the null would provide evidence that the pattern of accuracy was not consistent. Such a finding would indicate that the results of analogous comparative studies can be affected by choice of error metric. Any inconsistency provides evidence which rejects the null.

Research Objective One tests error metrics in a general setting which was employed in previous research studies. If Hypothesis One is rejected, error metric selection will be examined in the more specific setting where error metrics are related to risk assessment. Research Objectives Two through Four provide such an examination.

Research Objective Two

Research Objective Two addresses the effects of employing forecast mean versus forecast median in linear error metrics. The objective may be stated as:

2) To determine if forecast errors produced from error metrics which employ forecast median differ significantly

from those which employ forecast mean.

With respect to Research Objective Two, the following question is raised:

2) Do significant differences exist in error metrics defined using mean versus median forecasts?

Chapter III provided theoretical justification for use of forecast median in the event that a linear function was assumed to represent user loss. The assumptions of this analysis were: 1) the underlying loss function was symmetric and linear; and, 2) the underlying distribution of analysts forecast observations was asymmetric.

Cost of error minimization, in the linear case, implied that in the event that outliers forced asymmetry (whereby, the mean was not equal to the median), use of forecast median resulted in the lowest cost of error. A comparison of corresponding error metrics, in which only the internal forecast parameter differs, provides evidence of the distributional properties of analysts' forecasts. (The IBES data base does not report sufficient detail to directly determine this distribution.)

Comparison of corresponding error metrics provides one method of inferring the distributional properties of analysts' forecasts across all firms in the sample. Significant differences in corresponding error metrics implies an asymmetric distribution, while no difference in corresponding error metrics implies a symmetric distribution. If significant differences are noted, subsequent tests will determine if alternative error metrics produce forecast errors which are more highly associated with security risk.

Hypothesis Test Relating to Research Objective Two

Hypothesis Two provides one method by which significant differences in error metrics may be isolated. The hyopthesis is stated as:

H_{o2}: There is no significant difference in central tendency of forecast errors produced by error metrics defined using forecast median versus forecast mean.

H_{a2}: A significant difference exists.

The Friedman Test was used to address Hypothesis Two. This test is appropriate for data which is at least ordinal in scale, taken from N related groups, measured under K treatments. The Friedman Test is a nonparametric test of central tendency. A two-tailed test is appropriate since, *a priori*, there is no theoretical reason to expect one of the corresponding error metrics to produce forecast errors which are consistently larger or smaller than the other.

Failure to reject the null would imply that there is no significant difference in forecast errors produced by error metrics defined using forecast median versus forecast mean. This result would be attributed to an approximately symmetrical distribution of forecast observations, in which for each firm, the forecast mean was approximately equal to the forecast median.

Rejection of the null would provide evidence that one of the corresponding error metrics produced forecast errors which differed significantly from the other. This finding would suggest the need for further analysis to determine the definition which is most highly associated with security risk.

Research Objective Three

Research Objective Three addresses the effects of alternative error metrics on firm ranking by forecast error. The objective is stated as:

3) To determine if forecast errors resulting from alternative error metrics provide significantly different estimates of risk.

With respect to Research Objective Three, the following question is raised:

3) Do alternative error metrics change ranking of firms by forecast error across all firms?

As discussed in Chapter II, forecast error may be viewed as a surrogate for security risk. Tests of rank association may be employed to determine if alternative error metrics provide different estimates of risk. If alternative error metrics provide significantly different predictions of risk, choice of error metric may lead to incorrect investment decisions.

Investors, in an efficient market [Fama, 1970], are price protected; they are rewarded for bearing risk which cannot be eliminated through formation of a diversified portfolio. Security risk estimation is essential for portfolio formation. If risk is incorrectly estimated, or if error metrics yield forecast errors which differ in predictions of risk, the investor may hold a portfolio which is either more or less risky than intended.

One method by which error metrics may be tested for consistent predictions of risk is provided by H_{O3}. Relative rankings of firms, by forecast error, and resultant correlation coefficients of bivariate rankings indicate the degree to which alternative error metrics agree in this estimate of risk.

Hypothesis Statement Relating To Research Objective Three

In response to research question three, the following hypothesis is stated:

H_{O3}: There is no significant difference in the rank order of firms across all firms when alternative forecast error metrics are employed.

H_{a3}: A significant difference exists.

Tests of rank order are considered to be appropriate since a difference in ranking, due solely to error metric definition, could significantly affect predictions of risk which draw on measures of forecast error. Spearman's Rank Order Correlation Coefficients (Rhos), and Kendall's Correlation Coefficients (Taus) were used to address this hypothesis. These tests are appropriate when K pairs of values may be ranked from smallest to largest. Tables 4.3 and 4.4 outline the application of Spearman's Rho and Kendall's Tau to Hypothesis Three.

Table 4.3
Rhos Between Bivariate Observations

Firms (N)	Metric F_i	Metric F_j
.	(These columns contain the rank of each firm in ascending order down all firms for each metric form. For example:)	
.		
.		
.		
Firm N-1	50	20
Firm N	100	300

Rank Order Correlation Coefficients, called Rhos, are then computed. Significance levels indicate the probability that the treatments (alternative error metrics) have significantly affected rank order.

Rhos estimate the degree of agreement in ranking between two variables. Under the null hypothesis related to this test, Rhos are tested for significant differences from zero. Thus, ranking may be different between corresponding pairs of variables, even though correlation significance levels are at .001. For example, the correlation between F1 and F2 may be .9990, indicating that the two variables rank firms in a similar manner, while the correlation between F1 and F6 may be .5000. The level of association between F1 and F6 may still be significantly different from zero, however, relative ranking has been changed.

Evidence of differences in ranking is also provided by Kendall's Tau in which the degree of difference in ranking is computed for each bivariate observation. In computing Tau, the observations are designated as concordant if, in all cases, the rank of treatment K is larger than the rank of treatment K+1. Discordance is exhibited when, in some cases, the ranking is reversed. Significance levels are interpreted in the same manner as the significance levels of Rho.

For example, in the current study, Taus were computed in the following manner for each pair of error metrics. All firms were first ranked in descending order on error metric F(i), for example, F1. Corresponding metric F(j), for example, F2, was paired, by firm, with F1, resulting in a ranking of firms on F1 with corresponding values for F2.

Values of F2 for each firm were then compared with all observations of F2 falling below in the ranking. If F2 was greater than the next value down in the ranking, the pair was considered to be concordant. If F2 was less than the next value below, the pair was considered to be discordant. This procedure was repeated for each subsequent F2, and all observations below. Table 4.4 provides an application of Tau to the current study.

Table 4.4
Taus Between Bivariate Observations

Firms (N)	F1	F2	Concordance or Discordance
Firm A	.75	.77	
Firm B	.74	.75	Concordant
Firm C	.69	.80	Discordant
.	.	.	.
.	.	.	.
Firm N-1	.	.	.
Firm N	.	.	.

From the concordant or discordant designation, a correlation coefficient, Tau, was computed. This coefficient indicated the number of times F1 and F2 differed in ranking bivariate metrics across all firms.

For both Rho and Tau, a statistically significant, positive relationship may be exhibited, but differences in rank order may still exist.

Failure to reject the null that there is no difference in rank ordering among metric forms may suggest that metric form definition is not an important consideration in risk assessment. Rejection of the null implies that alternative error metrics could provide different predictions of risk. If significant differences are noted, Research Objective Four identifies the error metrics which is most closely associated with market beta.

Research Objective Four

Previous studies have identified a statistically significant, positive relationship between forecast error and systematic risk. This relationship supports the view that the risk of a firm is related to an earnings surprise, thus, risk may be partially composed of those factors which contribute to forecast inaccuracy (or vice versa).

Research Objective Four utilizes this empirical relationship in a further investigation of error metrics. The objective may be stated as:

4) To determine if a particular error metric produces forecast errors which are more highly correlated with systematic risk.

With respect to objective four, the following research question is raised:

4) Which error metric, if any, yields a forecast error measure which is most closely associated with systematic risk?

If forecast error is employed as a surrogate for security risk, then, the relationship between alternative error metrics and market beta may provide an opportunity to identify a preferred error metric.

<u>Hypothesis Statement Relating to Objective Four</u>

In response to research question four, the following hypothesis is stated:

H_{04}: There is no significant difference in the degree of association between alternate measures of error and measures of systematic risk.

H_{a4}: At least one definition of error exhibits a greater degree of association.

Spearman's Rho and d-statistics were used to address H_{o4}. Measures of systematic risk were computed for each firm using the market model, as shown in equation 4.1.

$$R_i = a + B_i(R_m) + e_i \qquad (4.1)$$

where:

R_i = The monthly return on the i^{th} security with dividends

B_i = Market beta for firm i

R_m = The equally weighted monthly returns for the market with dividends

Forecast errors and market betas were computed for the five year period beginning in 1979 and ending in 1983. As with all of the other hypotheses tests, the forecast error metric was an equally weighted five year average. Market beta regressed 60 monthly security returns upon 60 monthly market returns.

Forecast errors were tested for the degree of association with market beta using two statistical tests. In the first, Spearman's Rank Order Correlation Coefficients were computed for each forecast error and market beta pair. In the second, a nonparametric test of differences of regression coefficients of determination, R^2s, was performed.

Using a nonparametric test described by Rao and Miller [1971, p. 109], R^2s for two related regressions may be tested for significant differences. In the simple regression case, R^2 corresponds to the square of Pearson's Correlation Coefficient. In the event there are no ties, this parametric definition of correlation reduces to Rho. Thus, this nonparametric test indicates significant differences in correlation coefficients. The test statistic, denoted as d, is defined in equation 4.2.

$$d = N/2 \ln |RSS1/RSS2| \qquad (4.2)$$

where:

N = the sample size

RSS = the residual sums of squares for each regression

ln = the natural logarithm

The residual sums of squares, utilized in the d-statistic, were generated by the following regression, as shown in equation 4.3.

$$FE_i = a + C_i(B_{m,i}) + e_i \qquad (4.3)$$

where:

FE_i = The i^{th} definitional form of forecast error

C_i = Regression parameter

$B_{m,i}$ = Market beta for the i^{th} firm

e_i = The error term

a = Regression intercept

Failure to reject the null implies that, in the captial markets setting, the definition of the error metric is not an important factor in risk prediction. This finding would suggest that the concepts of linear versus nonlinear investor loss functions are either not adequately characterized by the more common error metrics, or the functions are not important considerations.

Rejection of the null would suggest that, in this empirical setting, one error metric may best represent security risk. This metric would be viewed as the form which is most consistent with market perceptions of risk.

Sample Selection

The sample consists of all firms which simultaneously meet the following restrictions.

(1) All firms must have a fiscal year end of December 31.

(2) Complete data must be available on the Institutional Brokers Estimate Service (IBES, Lynch, Jones and Ryan) tape, and the Center for Research on Security Prices (CRSP, University of Chicago) tape.

The sample was initially selected from the IBES tape, then matched to CRSP. The December 31 fiscal year end requirement reduced computer search time, and this arbitrary designation was not expected to systematically bias the results of this study, although most retailers would be excluded.

The IBES Summary History Tape was developed by Lynch, Jones and Ryan, and currently provides forecast information for about 2800 companies. Special care is taken by Lynch, Jones and Ryan to ensure that consistency is observed between the forecast of earnings and the actual earnings reported. For example, all individual forecasters for each firm are asked to indicate if primary or fully diluted EPS is being reported. If the majority of individual forecasters are providing forecasts of primary EPS, all analysts are asked to submit forecasts on this basis.

In addition, analysts at Lynch, Jones and Ryan update the history tape on a yearly basis to correct actual reported EPS, and analysts forecasts for stock dividends and stock splits. Thus, this data tape provides summary information which allows consistent comparisons of information.

In this study, all firms with complete data for the years 1979 to 1983 were selected from the IBES tape. Summary statistics for analysts forecasting primary EPS were then compiled, and included annual forecasts of median earnings, mean earnings, and the number of forecasters. Industry category was based on the classification scheme provided by the monthly hardcopy reports included as part of the summary tape service. A total of 766 firms met the first set of criteria, of which 91 were classified as regulated utilities. (Concentration by industry provides one method of examining error metrics for industry effects.)

These firms were then matched against the CRSP tape. A total of 529 firms met the dual selection criteria of having complete data on both tapes.

In addition to the full data set, and in response to the criticisms of Brown, Foster and Noreen [1985] who suggested that outliers may drive the results of some studies, a truncated data sample was also tested. The truncated sample consisted of all firms with linear error less than or equal to $1.00 for F1 and F2, linear error less than or equal to 100% for F3 through F6, quadratic error less than or equal to $2.00 for F7 and quadratic error less than or equal to 200% for F8 and F9. These parameters were selected because they reduced skewness, but did not reduce the sample below 500 firms.

Chapter Summary

Chapter IV provided hypotheses tests which are employed in an analysis of error metrics. The hypotheses tests were designed to discover the effects of error metric selection in a variety of settings.

Hypothesis One compares the error metrics in a previously tested empirical setting. This hypothesis tests the properties of all metric forms to produce consistent results. Failure to reject the null indicates that all forms are consistent. Rejection of the null implies that different results are obtained when alternative error metrics are selected; thus, the results of a comparative analysis using the Friedman Test and a no-change model depends on the error metric selected.

Hypothesis Two provides evidence of the differences in internal parameter selection. Failure to reject this hypothesis suggests that differences in mean versus median forecast errors may not exist due to the approximately symmetrical distribution of analysts' forecasts. Rejection of the null implies that significant differences exist, and further investigation of these error metrics may provide evidence which supports use of one form over the other.

Hypothesis Three provides evidence concerning the effects of alternative error metrics on firm ranking. This test indicates the degree of agreement in firm ranking between corresponding pairs of error metrics, and failure to reject the null implies that the error metrics rank order firms in a similar manner. Rejection of the null suggests that error forms alter rank order of firms, and that error metric selection can affect risk assessment.

Hypothesis Four provides evidence which suggests that one form of error may better correspond with market risk. Failure to reject the null indicates that all forms are interchangeable in this degree of association. Rejection of the null provides evidence that one form may best represent security risk.

Chapter V presents the results of hypotheses tests. Chapter VI provides a summary of the conclusions, and indicates limitations which may affect this study.

CHAPTER V

EMPIRICAL RESULTS

The following null hypotheses were designed to address research questions raised with respect to the research objectives:

H_{o1}: Choice of error metric does not alter consistency of findings of comparative analyst/model forecast performance.

H_{o2}: There is no significant difference in central tendency of error metrics defined using forecast median versus forecast mean.

H_{o3}: There is no significant difference in ranking of firms by forecast error when alternative error metrics are employed.

H_{o4}: There is no significant difference in degree of association between alternative metric forms and systematic risk.

Nonparametric tests were utilized for each hypothesis, since a normal distribution was rejected for each metric form. (In each case, the results of two-tailed Kolmogorov-Smirnov One-Sample test of the distribution indicated that the probability of a normal distribution was equal to zero.)

Descriptive Statistics

Descriptive statistics for each error metric in the complete sample, the truncated sample, and the subset of utilities are presented in Table 5.1. In addition to the statistics, metric definitions are provided.

Table 5.1
Error Metric Descriptive Statistics

Metric	Metric Mean	Maximum Value	CV%	Skewness
All Firms (n=766)				
F1=|Med-A|	.86$	10.2	115	3.4
F2=|Mean-A|	.86$	9.8	115	3.3
F3=|Med-A|/|A|	87%	24.5	253	6.2
F4=|Mean-A|/|A|	88%	24.3	255	6.2
F5=|Med-A|/|Med|	49%	38.9	355	17.4
F6=|Mean-A|/|Mean|	59%	130.1	802	26.7
F7=(Mean-A)^2	3.07$	301.1	445	15.5
F8=(Mean-A)^2/|A|	219%	145.0	381	10.7
F9=(Mean-A)^2/|Mean|	238%	886.4	1360	26.9
Truncated Sample (n=510)				
F1	.39$	.98	59	.5
F2	.39$	.97	59	.5
F3	22%	.97	82	1.7
F4	22%	.95	82	1.7
F5	17%	.84	68	2.0
F6	17%	.84	69	2.0
F7	.33$	1.67	109	1.5
F8	18%	1.78	134	2.8
F9	12%	1.40	115	3.5
Utilities Subset (n=91)				
F1	.37$	1.42	63	2.4
F2	.36$	1.43	64	2.5
F3	30%	12.24	428	9.3
F4	30%	12.42	432	9.4
F5	14%	.52	65	2.3
F6	14%	.52	66	2.3
F7	.29$	4.51	198	5.7
F8	69%	52.41	790	9.6
F9	10%	1.04	161	4.3

where:

Med = Median Forecast

Mean = Mean Forecast

A = Actual EPS

CV% = Coefficient of Variation

Skew = Skewness

Metric Mean = the mean metric value where all forms were computed on a yearly basis, then averaged across the time period from 1979 to 1983. The truncated sample and the subset of utilities are expressed in the same units as the entire sample.

An analysis of metric mean values, ranges, coeffiecients of variation, and values for skewness provides insights into the underlying distribution of analysts' forecasts. The IBES data base does not report sufficient detail to determine this distribution for each firm, thus, its properties must be inferred.

A comparison of corresponding linear error metrics (defined as F1 with F2, F3 with F4, and F5 with F6) provides evidence that the underlying distribution of analysts' forecasts is approximately symmetrical. A symmetric distribution is characterized by equal mean and median values. As is indicated in Table 5.1, corresponding error metrics which differ only in the definition of the internal forecast parameter (mean versus median), exhibit approximately equal values.

For example, since F1 and F2 differ only by the definition of this internal parameter, and the metric means for these forms are approximately equal, the underlying distribution of analysts' forecasts must be approximately symmetrical. This result is confirmed by the truncated sample, the subset of utilities, and an analogous comparison of F3 and F4 for all data sets. F5 and F6 exhibit larger differences because two changes are incorporated into these error metrics (both the numerator and the denominator change).

The implications of this finding suggest that cost minimization in the linear case may be a theoretical issue only. As discussed in Chapter III, the underlying user loss function is required to be symmetric, and the underlying distribution of forecast observations must be asymmetric in order to test differences in error metrics defined by median versus mean forecasts. No difference in the metrics will exist if the mean forecast is equal to the median forecast. Thus, the symmetric character of analysts' forecasts results in an empirically empty issue in terms of cost of error minimization and loss function analysis, because error metric defined using median versus mean are not appreciably different.

Properties of the distributions of the error metrics confirm this finding. Each error metric is constrained to yield positive values. Thus, the minimum value for all error metrics is equal to zero. An analysis of ranges of corresponding linear forms suggests that F1 and F2, in addition to F3 and F4, and to a lesser extent, F5 and F6 all exhibit approximately equal ranges.

The values for the coefficients of variation (defined as the standard deviation divided by the mean) provide comparative measures of the variablilty of corresponding error metrics. Again, corresponding error metrics exhibit approximately equal values for this statistic, in addition to exhibiting approximately equal values for skewness.

A comparison of skewness values between the complete sample and the truncated sample provides evidence that outliers may affect an analysis performed on the entire sample. In a symmetrical distribution, skewness is equal to zero. Positive values of skewness indicate that the median observation is less that the mean observation. Thus, larger positive skewness values suggest that outliers are shifting the mean toward the right (positive) tail of the distribution.

Individual firms with extreme values of error may not be representative of the population of all firms. Factors specific to these firms may have caused the larger error. Truncating the sample reduces skewness values, thus, is one method by which the effects of outliers may be removed.

Descriptive statistics for the truncated sample and the subset of the most stable industry are also presented in Table 5.1. An analysis of the subset of utilities suggests that the degree of error introduced by less stable industries is considerable. Metric mean values for this subsample are similar to the truncated sample, and other distributional parameters suggest greater variabililty than that noted for the truncated sample.

In summary, the descriptive statistics provided in this section suggest that:

1) The underlying distribution of analysts' forecasts is approximately symmetrical, thus, differences in mean and median forecast errors are insignificant.

2) Error metric distributions for the entire sample are skewed, thus, the results from the truncated sample data may provide results which are more representative of the population of all firms.

While the truncated data set may produce results with greater external validity, certain insights may be gained from performing hypotheses tests on the entire sample. The results of hypotheses tests performed on the entire sample, the truncated sample, and the subset of utilities are presented in the sections which follow.

Research Objective One

Research Objective One addresses the issue of consistency of error metrics in comparing forecast agents. In H_{01}, analysts' forecast errors are compared with forecast errors generated from a naive, no-change model. This analysis assumes that the no-change model is representative of the class of mechanical models to which analysts have been compared. While other mechanical models have provided better estimates of earnings (e.g., Foster [1978] provides an excellent summary of the various mechanical models which have been tested in previous studies), the no-change model is used to be represent a limited class of mechanical models, and has performed well in previous studies.

A four-year subset of analysts' error metrics and corresponding no-change error metrics was computed for the complete data set, the truncated sample, and the subset of utilities across the years 1980 to 1983. The four year subset was necessary because the no-change model employs the actual EPS from 1979 as the forecast for 1980. Analysts' errors were computed as before, and corresponding no-change metrics were computed where EPS_{t-1} was employed as the forecast, and EPS_t was considered to be the actual value. For example, F1 was still equal to |Median-EPS| averaged across four years, while NC1 was equal to $|EPS_{t-1}-EPS_t|$, also averaged across four years.

Error metrics were considered to be consistent if, for all error metrics: 1) there was no significant difference in forecast agents, thus, neither agent was favored; 2) there was a significant difference in forecast agents, and analysts were favored; or, 3) there was a significant difference in forecast agents, and the no-change model was favored. Any instance in which the favored agent changed across forecast errors provided evidence which rejected the null.

Table 5.2 reports the results of Friedman Tests for comparisons of forecast agents under alternative error metrics. The null is rejected for the entire sample, and the truncated sample. Only the subset of utilities provided consistent results that analysts are favored under all error metrics.

Table 5.2
Analysts Compared with the No-Change Model
Friedman Tests

Metric Form F(i)	Metric Form NC(j)	Mean Rank F(i)	Mean Rank NC(j)	Significance	Favored Agent
All Firms					
1	1	1.54	1.46	.051	N/S
2	1	1.53	1.47	.129	N/S
3	2	1.60	1.40	.000	NC
4	2	1.58	1.42	.000	NC
5	3	1.35	1.65	.000	A
6	3	1.36	1.64	.000	A
7	4	1.56	1.44	.001	NC
8	5	1.60	1.40	.000	NC
9	6	1.45	1.55	.009	A
Truncated Sample					
1	1	1.48	1.52	.320	N/S
2	1	1.47	1.53	.110	N/S
3	2	1.56	1.44	.045	NC
4	2	1.54	1.46	.060	N/S
5	3	1.34	1.66	.000	A
6	3	1.35	1.65	.000	A
7	4	1.51	1.49	.740	N/S
8	5	1.55	1.45	.061	N/S
9	6	1.45	1.55	.008	A
Utlilties Subset					
1	1	1.32	1.68	.001	A
2	1	1.27	1.73	.000	A
3	2	1.34	1.66	.002	A
4	2	1.33	1.67	.001	A
5	3	1.15	1.85	.000	A
6	3	1.15	1.85	.000	A
7	4	1.25	1.75	.000	A
8	5	1.26	1.74	.000	A
9	6	1.22	1.78	.000	A

where:

N/S = not significant at a .05 level

A = Analysts

NC = No-Change

(Significance levels of .05 or more are considered to indicate no significant difference in forecast agents.)

For the entire sample, inconsistency is noted among error metrics. No significant difference in forecast agents is documented for F1 with NC1, or F2 with NC2. The no-change model is favored for error metrics F3, F4, F7 and F8, while analysts are favored for error metrics F5, F6 and F9. This pattern of results rejects the null that error metrics do not alter consistency for the entire sample.

Similar results are noted for the truncated sample. No significant difference between forecast agents is noted for error metrics F1, F2, F4, F7 and F8, while analysts are favored for F5, F6 and F9, and the no-change model is favored for F3.

These results indicate that error metric selection affects the relative ranking of forecast agents when analysts are compared to a no-change model, and the Friedman Test is used to determine significant differences. Interestingly, five of the nine error metrics in the truncated sample indicate that analysts and the no-change model forecast earnings with approximately equal accuracy, thus, neither agent is favored.

Of the comparisons which yield significant differences, the no-change model is favored by F3. This result is surprising since F4 yields no significant difference, and F3 differs from F4 only by the internal forecast parameter. Closer examination of the significance levels indicates that F3 is only marginally significant.

As with the entire sample, the metric forms in the truncated sample which are expressed relative to the forecast (F5, F6 and F9) favor the analyst. This result indicates that analysts may be favored when metrics express error as a percentage of forecasts, as opposed to a percentage of the actual earnings achieved.

Both the entire sample and the truncated sample provide evidence that the results of a comparative study, which employs a no-change model and the Friedman Test, are dependent on error metric selection. For example, a comparative analysis which determined the superior forecast agent using F3, and possibly F4, would conclude that the no-change model provides superior forecasts. Conversely, if F5, F6 or F9 were employed, the conclusions would favor the analyst.

Only in the subset of utilities is consistency noted. For this subsample, the null cannot be overturned. Under every definition of forecast error, analysts are favored. This result is somewhat surprising, since previous studies have noted that the stability of earnings in this industry might favor a mechanical model. One possible explanation for analysts' superiority is the time period selected. The period from 1979 to 1983 was one of instability for oil prices, and mechanical models could fail to incorporate the dynamic nature of this industry. The increased variablilty noted in the descriptive statistics reflects this instability.

If the error metrics are defined in a different manner, where the linear forms are not constrained to be positive, (the absolute value operator is removed from the numerator) then the same analysis, for the utilities, between forecast agents yields inconsistent results. Table 5.3 presents the results of a comparative analysis of analysts and the no-change model in this setting.

The results presented in Table 5.3 are not considered in the hypothesis test. Rather, they are presented as supporting evidence that alternative error metrics, in addition to those selected for this study, may also affect the results of a comparative analysis. In addition, these results imply that use of the absolute value operator is an important consideration.

Table 5.3
Analysts Compared with No-Change Model
Unconstrained Linear Error Metrics
Friedman Tests (Subset Utilities Only)

Metric Form F(i)	NC(j)	Mean Ranks F(i)	NC(j)	Significance	Favored Agent
1	1	1.92	1.08	.000	NC
2	1	1.91	1.09	.000	NC
3	2	1.91	1.09	.000	NC
4	2	1.90	1.10	.000	NC
5	3	1.93	1.07	.000	NC
6	3	1.92	1.08	.000	NC
7	4	1.29	1.71	.000	A
8	5	1.37	1.63	.016	A
9	6	1.23	1.77	.000	A

The decision to use the absolute value operator in defining error metrics was based on problems associated with negative error metrics. Negative error metrics reduce mean error, even though the absolute error is greater than zero. For example, errors of -100% and 100% equal 0% error, on average. Use of the absolute value operator results in average error, in this case, of 100%.

However, as documented by Brandon and Jarrett [1977], some previous studies have employed error metrics which do not incorporate the absolute value operator. Table 5.3 suggests the need to define error with the absolute value operator, since the absence of this operator results in different forecast agents being favored.

In summary, the results of H_{01} imply that determining a superior forecast agent, in a diversified sample, is error metric dependent. This finding suggests the need to determine one metric form, or group of forms, which should be employed in comparative studies. The results of the hypotheses related to Research Objectives Two and Three confirm this finding, and Research Objective Four provides evidence that one group of forms might better represent security risk.

Research Objective Two

Error metrics consistent with linear user loss are defined using forecast median (F1, F3, and F5) versus the forecast mean (F2, F4, and F6). Table 5.4 presents the results of Friedman Tests performed on corresponding error metrics (F1 with F2, F3 with F4, and F5 with F6).

Table 5.4
Comparisons of Median to Mean Forecasts
Friedman Tests

Error Metrics F(i)	F(j)	Mean Ranks F(i)	F(j)	Significance
All Firms				
1	2	1.49	1.51	.745
3	4	1.49	1.51	.588
5	6	1.48	1.52	.295
Truncated Sample				
1	2	1.49	1.51	.565
3	4	1.49	1.51	.790
5	6	1.48	1.52	.330
Utilities Subsample				
1	2	1.54	1.46	.402
3	4	1.53	1.47	.600
5	6	1.54	1.46	.463

As was expected, given the implications of the descriptive statistics, the null that there is no difference in error metrics defined using forecast median versus forecast mean cannot be rejected. The distribution of analysts' forecast observations is approximately symmetrical, thus, no significant differences are noted between the corresponding error metrics.

In response to these results, a divergence technique was employed to select only those cases in which the distribution was somewhat asymmetrical (see Comiskey, Walkling and Weeks, 1986). Table 5.5 presents the results of Friedman Tests of the largest mean/median differences for the entire sample. (Insufficient cases for the truncated sample and the subset of utilities eliminated these data sets from the analysis.)

Table 5.5
Largest Median to Mean Differences
Friedman Tests

Difference Variable	Metric Forms F(i)	Metric Forms F(j)	Mean Ranks F(i)	Mean Ranks F(j)	Cases	Sig
>.249	1	2	1.80	1.20	5	.180
	3	4	1.27	1.73	11	.132
	5	6	1.40	1.60	10	.527
>.099	1	2	1.57	1.43	21	.513
	3	4	1.39	1.61	28	.257
	5	6	1.50	1.50	18	1.000
>.049	1	2	1.54	1.46	61	.522
	3	4	1.44	1.56	54	.414
	5	6	1.49	1.51	37	.869

The results in Table 5.5 were derived by first computing a difference variable, denoted as DIFF. In all cases, DIFF was defined as the absolute value of the difference in error results. Thus, DIFF for F1 and F2 was equal to |F1-F2|, while other differences were defined in a similar manner.

DIFF was preset to three levels to select approximately the ten largest differences (DIFF greater than .249), the 20 largest differences (DIFF greater than .099), and the 50 largest differences (DIFF greater than .049). In every case, the null could not be overturned.

The same results occured when DIFF was defined in a different manner. When DIFF was used to partition the sample on the absolute value of median forecasts less the mean forecasts (DIFF = |Median-Mean|), before computation of the forecast error, similar results were noted, and the null could not be overturned in any case.

Failure to reject the null at any level of difference or at any reasonable level of significance implies that the distribution of consensus forecasts is approximately symmetrical. Additional evidence of this assertion is included in the number of cases for each level of DIFF. In only five instances did the difference in forecast error forms between F1 and F2 exceed .249. Further, of the 766 total cases, 705 exhibited differences of less than .050.

Thus, the analysis of the least cost predictor for this data cannot be conclusively tested. While the least cost predictor for the linear case may be the consensus forecast median instead of the consensus forecast mean, sufficient cases with significantly large differences are not available from the current data. However, certain inferences may still be drawn from subsequent tests of metrics in which the internal forecast parameter is defined as the median. For this reason, these metrics will be included in all subsequent tests.

In summary, the results of tests of Research Objective Two confirm the conclusions drawn from the descriptive statistics. The underlying distribution of analysts' forecasts is approximately symmetrical, thus, mean forecasts equal the median forecasts. Accordingly, forecast mean versus forecast median is empirically an empty issue in terms of loss function analysis.

Research Objective Three

Research Objective Three provides evidence that error metrics alter firm ranking. As discussed in Chapter II, forecast error may be viewed as a surrogate for security risk. Thus, differences in firm ranking, due solely to error metric definition, may lead to incorrect risk assessments.

One method by which error metrics may be tested for consistent risk assessment for all firms is provided by H_{03}. Relative rankings of firms by forecast error, and resultant correlation coefficients of bivariate rankings suggests the degree to which alternative error metrics agree in assessment of this risk measure. Table 5.6 provides matrices of Spearmans Rho coefficients, and Table 5.7 presents Kendall's Tau results.

Table 5.6
Rhos Between Alternative Error Metrics

All Firms

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9987	.7974	.7951	.8100	.8082	.9908	.9142	.9453
F2		.7977	.7971	.8100	.8100	.9919	.9158	.9467
F3			.9991	.9441	.9435	.7996	.9566	.9113
F4				.9429	.9439	.7988	.9565	.9109
F5					.9988	.8011	.9064	.9354
F6						.8009	.9069	.9357
F7							.9286	.9534
F8								.9692

Truncated Sample

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9968	.7086	.7052	.6904	.6862	.9808	.8808	.9095
F2		.7092	.7103	.6903	.6910	.9835	.8847	.9130
F3			.9977	.9587	.9567	.7092	.9260	.8986
F4				.9568	.9593	.7099	.9273	.9001
F5					.9971	.6683	.8620	.8852
F6						.6687	.8635	.8867
F7							.9064	.9230
F8								.9779

Utilities Subset

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9936	.7608	.7422	.8168	.8060	.9671	.8906	.9073
F2		.7617	.7549	.8146	.8136	.9675	.8965	.9116
F3			.9888	.9540	.9511	.7733	.9265	.8895
F4				.9463	.9579	.7621	.9235	.8875
F5					.9928	.8138	.9179	.9286
F6						.8089	.9214	.9301
F7							.9344	.9489
F8								.9814

(All coefficients are significant at the .001 level)

Table 5.7
Taus Between Alternative Error Metrics

All Firms

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9948	.5997	.5967	.6215	.6193	.9195	.7496	.8015
F2		.6009	.5999	.6224	.6227	.9251	.7535	.8058
F3			.9776	.8057	.8029	.6018	.8195	.7376
F4				.8017	.8062	.6008	.8191	.7375
F5					.9736	.6124	.7339	.7851
F6						.6125	.7353	.7861
F7							.7715	.8186
F8								.8629

Truncated Sample

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9572	.5175	.5133	.5061	.5017	.8827	.7005	.7387
F2		.5193	.5194	.5068	.5077	.8929	.7082	.7467
F3			.9624	.8326	.8262	.5184	.7666	.7220
F4				.8258	.8351	.5182	.7678	.7240
F5					.9575	.4845	.6779	.7093
F6						.4850	.6806	.7119
F7							.7397	.7622
F8								.8825

Utilities Subset

	F2	F3	F4	F5	F6	F7	F8	F9
F1	.9437	.5801	.5615	.6330	.6171	.8616	.7323	.7568
F2		.5784	.5686	.6279	.6257	.8665	.7369	.7638
F3			.9253	.8217	.8187	.5844	.7768	.7231
F4				.8134	.8358	.5731	.7714	.7206
F5					.9398	.6161	.7578	.7695
F6						.6116	.7571	.7708
F7							.7812	.8154
F8								.8955

(All coefficients are significant at the .001 level)

Significant positive relationships exist between each metric form and all others, but the large range of values implies that differences in rank order do exist. For example, for the entire sample, the value of Rho F1,F2 is equal to .9987, indicating a very high level of agreement in ranking, while the value of Rho F1,F3 is equal to .7974, implying that while the correlation is still significantly positive, the rank order has been altered to a greater degree. Similar results are noted for the truncated sample and the subset of utilities.

Rho and Tau both measure the degree of association of bivariate rankings. Tau indicates the degree of discordance in the sample. As explained in Chapter IV, in the current study, Taus were computed in the following manner for each pair of error metrics. All firms were first ranked in descending order on error metric F(i), for example, F1. Corresponding metric F(j), for example F2, was paired, by firm, with F1, resulting in a ranking of firms on F1 with corresponding values for F2.

Values of F2 for each firm were then compared with all observations of F2 falling below in the ranking. If F2 was greater than the next value down in the ranking, the pair was considered to be concordant. If F2 was less than the the next value below, the pair was considered to be discordant. This prodecure was repeated for each subsequent F2 and all observations below. The resulting statistic, Tau, measures the degree of discordance, which indicates the number of times F1 and F2 differed in ranking bivariate metrics across all firms.

The values for Tau confirm the differences in ranking noted in the analysis of Rho. Although the significance levels indicate a positive, statistically significant relationship in each bivariate observation, differences in ranking exist. For example, in the entire sample, Tau values range from .5967 for Tau F1,F4 to .9776 for Tau F3,F4, indicating that discordance was greater for Tau F1,F4.

The differences in ranking are more pronounced in the truncated sample. Tau F3,F4 exhibited the greatest agreement in ranking with a value of .9624. The lowest agreement in ranking occurs with F6,F7 which exhibits a Tau value of .4850.

The results of H_{O3} indicate that selection of error metrics can affect risk assessment. If forecast error is employed as a surrogate for security risk, and choice of error metric affects assessments of risk, then, securities with inappropriate risk may be selected in portfolio formation. Thus, measures of forecast error should be tested to determine if one form, or group of forms, exhibits a higher degree of association with market beta. Research Objective Four tests for the degree of association of each error metric market beta.

Research Objective Four

Previous efforts have identified a positive, statistically significant relationship between forecast error and systematic risk. Under the assumption that this relationship provides an opportunity to identify a preferred error metric, objective four seeks to determine the metric form which is most highly associated with systematic risk.

Table 5.8 presents the results of H_{04}, and provides Spearman Rank Order Correlation Coefficients between each form of error and market beta. The results of related d-statistics reject the null that there is no significant difference in the degree of association between alternative error metrics and market beta.

Table 5.8
Relationship of Error Metrics
with Systematic Risk (Market Beta)

	All Firms (n=529)	Truncated Sample (n=345)	Utilities Subset (n=91)
F1	.4008(.001)	.2156(.001)	.2256(.016)
F2	.4060(.001)	.2264(.001)	.2444(.010)
F3	.5145(.001)	.4088(.001)	.2113(.023)
F4	.5174(.001)	.4144(.001)	.2152(.021)
F5	.5412(.001)	.4121(.001)	.2655(.040)
F6	.5434(.001)	.4201(.001)	.2549(.033)
F7	.4090(.001)	.2318(.001)	.2413(.011)
F8	.4829(.001)	.3546(.001)	.2320(.014)
F9	.4885(.001)	.3418(.001)	.2315(.014)

(Note: sample sizes have been reduced because each firm was required to have complete data on both the IBES and the CRSP tapes. Significance levels are in parentheses.)

The results of d-statistics provides evidence which rejects the null. In each case, the highest value for Rho was tested against all other correlation coefficients, using the d-statistic, to determine if one form was significantly greater than all other forms. Table 5.9 provides the results of d-statistics computed for F6 with all other measures of error.

Table 5.9
Results of d-statistics Between
F6 and All Other Error Metrics

	All Firms	Truncated Sample	Utilities
F1=|Med-A|	919.5	268.6	81.1
F2=|M-A|	924.1	267.6	79.9
F3=|Med-A|/|A|	484.7	153.6	238.4
F4=|M-A|/|A|	478.2	151.7	239.4
F5=|Med-A|/|Med|	2.7*	.4*	.6*
F6=|M-A|/|M|	---	---	---
F7=$(M-A)^2$	521.4	424.5	159.6
F8=$(M-A)^2/\|A\|$	278.5	292.8	372.1
F9=$(M-A)^2/\|M\|$	901.2	77.2	48.2

(Note that Med=Median, M=Mean, A=EPS, and * The Chi-Square distribution with one degree of freedom at an alpha level of .05 is equal to 3.841)

In the entire sample, F6 exhibits the greatest degree of association with market beta. This correlation coefficient is significantly different from all others except F5. This result indicates that error metrics expressed relative to analysts' forecasts provide risk surrogates which are most closely associated with market beta.

In the truncated sample, as with the utilities subsample, the same result is noted. Results of d-statistics calculated for F6 with all other coefficients indicates that this coefficient is significantly different from all others, except for F5.

However, the results of correlation coefficients between forecast error and market beta are reduced for the truncated sample, when compared with the entire sample. This result is counter-intuitive (because eliminating those data points which reflect extreme error should remove the effects of firms which may not be representative of the population of all firms, and should increase the correlation of forecast error with market beta). These results suggest the need for further analysis to determine the approximate shape of the user loss function.

As described in Chapter III, the related investor loss function may be linear or nonlinear, symmetric or asymmetric. Additionally, the concept of a threshold level suggested that the function could be discontinous.

Cost of error minimization required the assumption of a symmetric loss function for either the linear or the nonlinear case. This assumption implied that the loss of overestimates was equal to the loss of underestimates in any instance of investor loss which was related to use of a forecast of earnings.

The relationship between forecast error and market beta is assumed to impound market perceptions concerning this loss function. Since, in the aggregate, investor loss functions are unobservable, the relationship between forecast error and market beta is used as one possible technique to infer the shape of the aggregate investor loss function.

The symmetry of the investor loss function is inferred through a partition, for a single year, of the sample into three groups of: 1) the 50 largest positive forecast errors; 2) the 50 largest negative forecast errors; and, 3) the 50 smallest forecast errors. Each group is then included in separate simple regressions with market beta, and the slope coefficient is used to determine the symmetry of the distribution of forecast errors for that year.

Only one year of the sample is employed, since multiple years would require the use of an average forecast error which would confound the analysis for firms with overestimates in some years and underestimates in other years. This study employed the year 1981 to represent the years 1979 to 1983. (The analysis was also performed for the year 1983 with no substantial differences in results).

In this analysis, the absolute value operator is removed from the numerator of each metric form, allowing negative error metrics to result. In addition, two supplemental metric forms are included. Table 5.10 provides the definitions of the adjunct error metrics which are employed in this analysis.

Table 5.10
Adjunct Error Metrics

AF1	=	Median - EPS
AF2	=	Mean - EPS
AF3	=	Median - EPS / |EPS|
AF4	=	Mean - EPS / |EPS|
AF5	=	Median - EPS / |Median|
AF6	=	Mean - EPS / |Mean|
AF7	=	(Mean - EPS)^2
AF8	=	(Mean - EPS)^2 / |EPS|
AF9	=	(Mean - EPS)^2 / |Mean|
AF10	=	(Mean - EPS / EPS)^2
AF11	=	(Mean - EPS / Mean)^2

The supplemental error metrics are included, since this analysis does not entirely utilize nonparametric tests of rank; thus, the numerical values for each metric are employed, instead of being replaced by their ranks. As was indicated in Chapter III, AF10 is the square of F4, while AF11 is the square of F6. Nonparametric tests of rank provide the same results for the adjunct forms as for the original forms.

The sample was first partitioned on the sign of a difference variable (ADIFF) defined as the difference in the mean forecast value less the actual value (ADIFF = Mean - EPS). From this partitioning scheme, the sample was divided into firms which were overestimated (ADIFF > 0), firms which were underestimated (ADIFF < 0), and firms which were forecast with relative accuracy (ADIFF about equal to 0). ADIFF was defined using the forecast mean, since no significant differences were noted between forecast mean and forecast median error metrics.

Subsamples were then formed by ranking the firms on ADIFF to form the 50 highest overestimates (POS 50), the 50 highest underestimates (NEG 50), and the 50 estimates closest to the actual EPS (MID 50). Error metrics AF1 to AF11 were then computed on each group.

An alternate partitioning scheme was also employed in which forecast errors, instead of ADIFF, were used to select POS 50, MID 50, and NEG 50. In this scheme, the sample was partitioned on ADIFF, then forecast errors were computed, and firms were ranked on the size of the forecast error. This second scheme resulted in different firms being selected for each error metric, and is presented as confirmatory evidence that correlation coefficients differ between overestimates and underestimates.

Table 5.11 presents the results of Spearman's Rank Order Correlation Coefficients for firms ranked on ADIFF, and Table 5.12 presents Spearman's Correlation Coefficients for firms ranked on the size of the forecast error.

Table 5.11
Spearman's Rhos Between Error Metrics and
Market Beta Ranked on ADIFF

	POS 50	NEG 50	MID 50
AF1=Med-A	.3113(.002)	.2058(.076)	.0586(.344)
AF2=M-A	.3099(.002)	.1715(.117)	-.0158(.457)
AF3=Med-A/|A|	.2456(.043)	.0151(.459)	.2294(.079)
AF4=M-A/|A|	.2555(.037)	.0291(.421)	.2261(.069)
AF5=Med-A/|Med|	.3978(.003)	.0141(.460)	.2294(.055)
AF6=M-A/|M|	.4021(.002)	.0290(.421)	.2261(.058)
AF7=$(M-A)^2$	.3099(.002)	.1715(.117)	-.0158(.457)
AF8=$(M-A)^2/\|A\|$	.4006(.002)	.1247(.195)	.1572(.138)
AF9=$(M-A)^2/\|M\|$	.4577(.001)	.0858(.277)	.1591(.135)

(Note that significance levels are in parentheses, Med=Median, M=Mean, and A=Actual EPS.)

The results provided by Table 5.11 indicate that, for the year 1981, the relationship between forecast error and beta is different between overestimates and underestimates. Further, correlation coefficients for this year indicate that overestimates exhibit a higher correlation between nonlinear error and market beta (as expressed by Rho AF9,Beta of .4577).

The coefficients between NEG 50 and MID 50 forecast errors with market beta are not significant. These results imply that the observed relationship between forecast error and market beta is not as apparent for underestimates or for forecast accuracy. This result may be due to factors specific to the time period, or may indicate that the more common forms of forecast error do not adequately capture this relationship.

Interestingly, the results noted in Table 5.11 are relatively consistent when the sample is ranked on the size of the forecast error. Table 5.12 provides the results for the second ranking scheme.

Table 5.12
Spearman's Rhos Between Error Metrics and Market Beta Ranked on Forecast Error

	POS 50	NEG 50	MID 50
AF1=Med-A	.3113(.002)	.1658(.096)	.0165(.455)
AF2=M-A	.3099(.002)	.1715(.117)	.0552(.352)
AF3=Med-A/|A|	.2456(.043)	.0151(.459)	.1938(.089)
AF4=M-A/|A|	.2555(.037)	.0291(.421)	.1876(.097)
AF5=Med-A/|Med|	.3987(.003)	.0149(.460)	.1938(.089)
AF6=M-A/|M|	.3034(.002)	.0290(.421)	.1876(.097)
AF7=$(M-A)^2$	.3099(.002)	.1715(.117)	.0552(.352)
AF8=$(M-A)^2/\|A\|$	.4600(.002)	.2647(.195)	.1174(.209)
AF9=$(M-A)^2/\|M\|$	.4577(.001)	.1858(.099)	.1193(.205)

(Note that significance levels are in parentheses, Med=Median, M=Mean, and A=Actual EPS.)

The results of Table 5.12 are similar to those noted in Table 5.11. In fact, very few of the coefficients are different. Those which have been changed (due to the different firms included when each forecast error is used to rank the firms) provide evidence that the nonlinear forms again exhibit higher correlation with market beta for overestimates.

Since the second ranking scheme does not include the same firms in each grouping, the ranking scheme in which ranks are assigned according to the rank of ADIFF is used to infer the shape of the investor loss function. In this evaluation, the assumption of a symmetric loss function is addressed. In addition, the approximate shape of the function is considered.

Each group of partitioned forecast errors was entered into a regression with market beta. Table 5.13 provides the results of slope coefficients and coefficients of determination by error metric.

Table 5.13
Slope Coeffiecients and R^2 Values
Error Metrics and Market Beta

	POS 50		MID 50		NEG 50	
	Slope	R^2	Slope	R^2	Slope	R^2
AF1=Med-A	1.62	.106	.009	.014	-.53	.059
AF2=M-A	1.59	.108	.009	.004	-.55	.062
AF3=Med-A/|A|	1.43	.023	.032	.088	-.06	.004
AF4=M-A/|A|	1.40	.022	.032	.072	-.06	.005
AF5=Med-A/|Med|	.85	.104	.027	.081	-.07	.005
AF6=M-A/|M|	.85	.117	.027	.063	-.08	.007
AF7=$(M-A)^2$	13.5	.136	.002	.001	2.98	.064
AF8=$(M-A)^2$ /|A|	9.7	.069	.004	.077	.32	.003
AF9=$(M-A)^2$ /|M|	5.81	.111	.004	.069	.55	.041
AF10=$(M-A/A)^2$	28.7	.015	.005	.112	.19	.005
AF11=$(M-A/M)^2$	3.2	.058	.004	.108	.23	.008

(Where Med=Median, M=Mean, and A=Actual EPS.)

The results in Table 5.13 indicate that the slope coefficients are different between overestimates and underestimates. This result implies that the corresponding investor loss function is not symmetric.

A second group of tests may be applied to the error metrics to determine if the investor loss function is linear or nonlinear. Each of the error metrics, AF1 through AF11, were replaced with their natural logarithms, and the regressions repeated. The results of R^2 values for POS 50 and NEG 50 are presented in Table 5.14. (The MID 50 firms were excluded from this analysis because replacing these error forms with the natural logarithm resulted in the majority of the values being set equal to zero.)

Table 5.14
Comparison of Error Metrics and Logs of Error Metrics

POS 50

	R^2 AF(i)	R^2 LN(AF(i))
AF1=Med-A	.106	.127
AF2=M-A	.108	.130
AF3=Med-A/|A|	.023	.071
AF4=M-A/|A|	.022	.069
AF5=Med-A/|Med|	.104	.147
AF6=M-A/|M|	.117	.151
AF7=$(M-A)^2$	.136	.197
AF8=$(M-A)^2$/|A|	.069	.163
AF9=$(M-A)^2$/|M|	.111	.229
AF10=$(M-A/A)^2$	.015	.069
AF11=$(M-A/M)^2$	.058	.181

NEG 50

	R^2 AF(i)	R^2 LN(AF(i))
AF1=Med-A	.059	.049
AF2=M-A	.062	.049
AF3=Med-A/|A|	.004	.0002
AF4=M-A/|A|	.005	.0003
AF5=Med-A/|Med|	.005	.0001
AF6=M-A/|M|	.007	.0003
AF7=$(M-A)^2$	.064	.049
AF8=$(M-A)^2$/|A|	.033	.017
AF9=$(M-A)^2$/|M|	.041	.014
AF10=$(M-A/A)^2$	.005	.003
AF11=$(M-A/M)^2$	.008	.0003

For the overestimated firms, in every instance, replacing the error term with its natural logarithm improves the value of R^2. This result indicates that for overestimates, the underlying investor loss function may be viewed as nonlinear, suggesting that the penalities associated with overestimates are more than proportional to the forecast error.

Notice that forms AF1 through AF6 provide logarithmic forms of error, which, in addition to the quadratic forms of error, increases the nonlinear forms evaluated. Both nonlinear error metrics exhibit higher values for R^2, which suggests that for overestimates, the functional relationship is nonlinear.

Interestingly, underestimates do not seem to be represented by the same relationship. The values for R^2 do not improve, and in some cases, these values are smaller. Again, the function is viewed as asymmetric, and may be described as nonlinear for overestimates, and linear (or no significant relationship) for underestimates. In every case, for the overestimates, logarithmic nonlinear R^2 values exceed R^2 values for linear measures of error, while no differences are noted for underestimates.

These results provide the first empirical estimation of the aggregate investor loss function in relation to errors in forecasts of earnings. The results imply that one form of forecast error may not adequately describe the relationship between user loss and forecast error for both overestimates and underestimates.

Additionally, the results explain the reduction in correlation for the truncated sample when compared with the entire sample. As is indicated by Table 5.13, the MID 50 firms do not exhibit statistically significant correlation coefficients between forecast error and market beta. Truncation removed the extreme forecast errors, resulting in a group of firms which was similar to the MID 50 firms.

This same analysis was performed on the data provided by Neiderhoff and Regan [1972]. The authors analyzed the relationship between analysts' errors and percentage share price increase, and provided data for the 50 best-performing firms and the 50 worst-performing firms. The current analysis resulted in mixed results for the year 1970. This finding was due to the different partitioning schemes employed by the two studies.

Further analysis of these relationships is provided by using two regression equations to estimate the shape of the loss function for overestimates. These equations estimate the degree to which the functional relationships between forecast error (X) and systematic risk (Beta) may be expressed as either quadratic or cubic functions. Equations 5.1 and 5.2 provide slope coefficients with related t-statistics (in parentheses). To reduce problems of multicollinearity, forecast errors were expressed as deviations around the mean.

$$\text{Beta} = 1.39 + \underset{(3.34)}{.167X} - \underset{(-2.32)}{.024X^2} \tag{5.1}$$

$$\text{Beta} = 1.33 + \underset{(3.46)}{.179X} + \underset{(1.89)}{.012X^2} - \underset{(-.29)}{.700X^3} \tag{5.2}$$

Equation 5.1 provides evidence that for 1981, and for the specific data set evaluated, the quadratic term provides a significant coefficient in the evaluation of the nonlinear response. Slope coefficients for both the linear and the quadratic terms suggest that these coefficients are significantly different from zero.

Equation 5.2 presents the results of the cubic form of a polynomial regression with one independent variable. Addition of the cubic term does not provide a significant slope coefficient.

The results of equations 5.1 and 5.2 imply that the quadratic function may better express the relationship between forecast error and systematic risk for overestimates. However, the effects of multicollinearity cannot be completely eliminated in this analysis. Therefore, these results should be viewed as tentative.

Thus, analysis of the adjunct error metrics provides evidence that one of the basic assumptions of cost of error minimization has been violated. The loss function is not symmetric, which suggests that the common forms of error cannot be utilized for overestimates and underestimates.

The conclusions of this analysis include the following:

1) Investor loss functions are not symmetric.

2) Nonlinear measures of error best describe the relationship between forecast error and beta (the proxy for investor loss) for overestimates.

3) Linear measures of error may provide the best measures of error for underestimates.

Chapter Summary

The empirical results presented in this chapter provide evidence that error metric selection affects comparative analysis of analysts with a mechanical model, and associations of forecast error and systematic risk. The empirical findings are summarized as:

1) The underlying distrubution of analysts' forecasts in the IBES data base is approximately symmetrical. Thus, there is no difference, in the grouped data, between mean forecasts, and median forecasts.

2) Error metric distributions were positively skewed, indicating that outliers could affect subsequent analyses. Truncated data provided one method by which the measure of asymmetry could be reduced, thus, enhancing external validity.

3) Comparison of analysts with a naive, no-change model using the Friedman Test provides inconsistent results when alternative error metrics are employed. Thus, the conclusions of comparative studies which utilized this paradigm were error-metric dependent.

4) Analysis of the largest mean/median differences failed to yield significant results. Thus, even the most extreme differences in analysts' forecasts do not provide error

metrics which are significantly different.

5) Alternative error metrics change the rank ordering of firms. Thus, if error metrics are employed as surrogates for security risk, inconsistent risk assessments will result.

6) In a comparison of error metrics with market beta, the underlying investor loss function was considered to be asymmetric. Nonlinear measures of error described the relationship between loss and forecast error for overestimates, while linear measures of error were viewed as more consistent with underestimates.

Chapter VI provides an interpretation of these results. In addition, the implications and limitations of this study are presented.

CHAPTER VI

CONCLUSIONS

The purpose of this study was to empirically assess error metrics applied to analysts' forecasts of earnings. Nine metric forms were defined and employed in hypotheses tests. These tests were proposed in response to the following research objectives:

1) To analyze the effects of error metric selection on conclusions drawn from previous studies of the comparative accuracy of analysts with a mechanical model.

2) To determine if forecast errors produced from error metrics which employ forecast median differ significantly from those which employ forecast mean.

3) To determine if forecast errors resulting from alternative error metrics provide significantly different estimates of risk.

4) In the event that alternative error metrics were shown to provide significantly different estimates of risk, to determine if a particular error metric produces forecast errors which are more highly correlated with systematic risk. (This objective also served as an indirect test of investor loss functions.)

A series of four research hypotheses were proposed in response to these research objectives. Nonparametric tests were utilized to provide statistical inferences concerning the population of firms from which the sample was drawn. In total, 766 firms were included in the entire sample, 510 firms were included in the truncated sample, and 91 firms were included in the utilities subsample. All samples contained data for the years 1979 to 1983.

The results, interpretations, and conclusions of this study are presented in the paragraphs which follow.

1) The underlying distribution of analysts' forecasts in the IBES data base is approximately symmetrical. Thus, there is no difference, in the grouped data, between mean forecasts, and median forecasts.

Chapter III provided theoretical support for use of forecast median in the event that a linear function was assumed to represent user loss. The results of Chapter V indicated that, for the time period tested, analysts' forecasts of earnings were essentially identical. Most analysts provided individual forecasts of earnings for a firm which were similar to all other analysts' forecasts for the same firm.

The analysis of the minimum cost predictor in the linear case required the assumption of a symmetrical loss function, and the assumption of an asymmetrical set of observations of analysts' forecasts. The latter requirement was necessary since under a symmetric distribution, the mean and the median forecasts are equal.

Failure to support the median forecast as the least cost predictor rests with the distribution of analysts' forecasts. Of the 766 firms in the entire sample, 705 exhibited differences in F1 with F2 of less than $.05. The largest differences were only at $.250, and only five firms exhibited this degree of difference. Thus, sufficient observations which were asymmetric could not be located in this sample, therefore, the least cost prediction analysis could not be conclusively tested.

2) Comparison of analysts with a naive, no-change model using the Freidman Test provided inconsistent results under alternative error metrics. In this general test, the accuracy of analysts was compared with the accuracy of a no-change model. The form of the test was designed to reflect previous efforts in this area.

For example, Brown and Rozeff [1978], and Imhoff and Pare [1982] both utilized Friedman Tests to determine a superior forecast agent. In both of these studies, analysts were viewed as the superior forecast agent when compared with a mechanical model. However, the conclusions of these efforts must be viewed as tentative and conditional upon metric form.

In the current study, forecast agents (analysts and a naive, no-change model) were compared using the Friedman Test. The results indicated that selection of an error metric provided inconsistent results, that is to say, the favored analyst depended on the metric employed in the analysis. Thus, in this very general setting, the results indicate that the results of comparative analyses are dependent on error-metric selection.

3) Alternative error metrics significantly changed the ranking of firms. This result provided evidence that when error metrics are viewed as estimates of risk, choice of error metric can affect risk assessment (or in any other comparable use.)

In this test, Spearmans' Rho and Kendall's Tau correlation coefficients provided evidence that firm rank ordering was altered across metric forms. The results of such tests indicated that error metric form affects ranking, thus affecting risk prediction.

This result indicates that choice of error metric could affect, for example, estimates of security risk. Estimates of risk are required in portfolio formation. Therefore, if error metrics were employed in estimates of security risk, the form which provides the best estimate of security risk should be identified.

4) Since choice of metric form could affect risk prediction, additional analyses were performed to determine which group of error metrics produced forecast errors which exhibited the highest correlation with market beta. Evidence from this group of tests supported the view that the investor loss function was asymmetric, and overestimates

are best represented by nonlinear error metrics, while underestimates are best represented by linear error metrics. These results suggest that one form of error may not provide the best surrogate for risk for both overestimates and underestimates. If forecast error is viewed as a predictor of risk, then the trends of analysts to over- or underestimate earnings may provide an indication of which metric form should be employed in risk estimation.

An extension of this study should consider the trends in analysts' forecast errors. For example, if forecast error is viewed as a surrogate for risk, and analysts have consistently overestimated earnings, then nonlinear error measures may provide better estimates of risk. In a similar manner, consistent underestimates of earnings might indicate that linear measures of error provide better estimates of risk.

A second extension of this study would be one in which forecast errors are analyzed in relation to corresponding cumulative abnormal residuals. This analysis would provide supplemental information concerning the investor loss function, and would evaluate user loss in the context of abnormal returns on share price.

Certain limitations affect the conclusions of this study. These limitations are summarized as:

1) The results of this study are specific for the time period of 1979 to 1983. In addition, the analysis of

investor loss functions is both indirect, and specific for the year 1981 (although the year 1983 was also evaluated, and the results were similar).

An extension of this study would be a study in which additional time periods were selected for all analyses. Further, extending the analysis of the relationship of forecast error with market beta over a longer horizon would provide evidence of the stability of the user loss relationships noted above.

2) All results are specific for firms which met the dual selection criteria of having complete data on the IBES data base, and on the CRSP tape. These firms are generally larger, more established companies, thus, the results of this study may not extend to the population of smaller, or newer firms. Further, the December 31 reporting date requirement eliminated many retailers from the sample.

An extension of this study would be to replicate the analyses on other data bases. For example, both the Earnings Forecaster, and the Value Line data set include consensus mean forecasts. (However, only the IBES data tape provides median forecasts.)

3) The comparison of forecast consistency of analysts with a mechanical model included only one mechanical model. While the no-change model was viewed as representative of a limited class of mechanical models, an extension of this study would be one in which multiple mechanical models were

compared with analysts to determine if the results of previous efforts were consistent.

4) Only one form of nonlinear error was evaluated (although logarithmic forms were employed in an evaluation of the investor loss function.) An extension of this study would be one in which forecast errors produced by other nonlinear error metrics, such as a fractional power error metric, were compared with linear and quadratic metric forms.

5) One final limitation of this study is that forecast error expresses a total risk measure, while market beta expresses a systematic component. An extension of this study would be to further test error metrics using only the systematic component of forecast error, as defined by Comiskey, Mulford and Porter [1986].

In summary, this study provides evidence which supports nonlinear error forms in risk prediction for firms which have been overestimated, and linear forms for firms which have been underestimated. The results of this study provide evidence which suggests that error metric selection is an important consideration in security risk estimation. These results emphasize the need to employ only the error metrics which are most representative of security risk.

BIBLIOGRAPHY

Accounting Principles Board, Opinion Number 30, "Reporting the Results of Operations", 1973.

Albrecht, W., Johnson, O., Lookabill, L. and Watson, D., "A Comparison of the Accuracy of Corporate and Security Analysts' Forecasts of Earnings: A Comment", THE ACCOUNTING REVIEW, July, 1977, pp. 736-740.

Armstrong, J. and Beuchert, G., "Judgment Extrapolation and Amalgamated Forecasts of Corporate Earnings", Working Paper, The University of Pennsylvania, 1979.

Ball, R. and Brown, P., "An Empirical Evaluation of Accounting Income Numbers", JOURNAL OF ACCOUNTING RESEARCH, Autumn, 1968, pp. 159-178.

Barefield, R. and Comiskey, E., "The SEC Segment Disclosure Requirement and Its Impact on Forecasting Earnings", ARIZONA REVIEW, November, 1974, pp. 7-12.

Barefield, R. and Comiskey, E., "The Accuracy of Analysts' Forecasts of Earnings Per Share", JOURNAL OF BUSINESS RESEARCH, July, 1975, pp. 241-252.

Barefield, R. and Comiskey, E., "The Association of Forecast Error with Other Risk Measures", JOURNAL OF BUSINESS, FINANCE AND ACCOUNTING, Autumn, 1975a, pp. 315-325.

Barefield, R. and Comiskey, E., "The Differential Association of Forecast Error and Earnings Variability with Systematic Risk", JOURNAL OF BUSINESS, FINANCE AND ACCOUNTING, Spring, 1979, pp. 1-8.

Barefield, R., Comiskey, E. and McDonald, C., "Accuracy of Management and Security Analysts' Forecasts: Additional Evidence", JOURNAL OF BUSINESS RESEARCH, 1979, pp. 109-115.

Basi, B., Carey, K. and Twark, R., "A Comparison of the Accuracy of Corporate and Security Analysts' Forecasts of Earnings", THE ACCOUNTING REVIEW, April, 1976, pp. 244-254.

Beaver, W., Clarke, R. and Wright, W., "The Association Between Unsystematic Security Returns and the Magnitude of Earnings Forecast Error", JOURNAL OF ACCOUNTING RESEARCH, Autumn, 1979, pp. 316-340.

Beaver, W. Kettler, P. and Scholes, M., "The Association Between Market-Determined and Accounting-Determined Risk Measures", THE ACCOUNTING REVIEW, October, 1970, pp. 654-682.

Bhaskar, K. and Morris, R., "The Accuracy of Brokers' Profit Forecasts in the UK", ACCOUNTING AND BUSINESS RESEARCH, Spring, 1984, pp. 113-124.

Branch, B. and Berkowitz, B., "The Predictive Accuracy of the BUSINESS WEEK Earnings Forecasts", JOURNAL OF ACCOUNTING, AUDITING, AND FINANCE, Spring, 1981, pp. 215-219.

Brandon, C. and Jarrett, J., "Accuracy of Externally Prepared Forecasts", REVIEW OF BUSINESS AND ECONOMIC RESEARCH, Fall, 1977, pp. 35-47.

Brown, L. and Rozeff, M., "The Superiority of Analysts' Forecasts as Measures of Expectations: Evidence from Earnings", THE JOURNAL OF FINANCE, March, 1978, pp. 1-16.

Brown, L. and Rozeff, M., "Analysts Can Forecast Accurately!", JOURNAL OF PORTFOLIO MANAGEMENT, Spring, 1980, pp. 31-34.

Brown, P., Foster, G. and Noreen, E., SECURITY ANALYST MULTI-YEAR EARNINGS FORECASTS AND THE CAPITAL MARKET, American Accounting Association, 1985.

Collins, W. and Hopwood, W., "A Multivariate Analysis of Annual Earnings Forecasts Generated from Quarterly Forecasts of Financial Analysts and Univariate Time Series Models", JOURNAL OF ACCOUNTING RESEARCH, Autumn, 1980, pp. 390-406.

Comiskey, E., "Assessing Financial Quality: An Organizing Theme for Credit Analysts", JOURNAL OF COMMERCIAL BANK LENDING, December, 1982, pp. 32-47.

Comiskey, E., Mulford, C. and Porter, T., "Forecast Error, Earnings Variability and Systematic Risk", JOURNAL OF BUSINESS FINANCE & ACCOUNTING, Summer, 1986, pp. 257-265.

Comiskey, E., Walkling, R., and Weeks, M., "Dispersion of Expectations and Trading Volume", Forthcoming, JOURNAL OF BUSINESS FINANCE & ACCOUNTING.

Copeland, R. and Marioni, R., "Executives Forecasts of Earnings Per Share Versus Forecasts of Naive Models", JOURNAL OF BUSINESS, October, 1972, pp. 497-572.

Cragg, J. and Malkiel, B., "The Consensus and Accuracy of Some Predictions of the Growth of Corporate Earnings", THE JOURNAL OF FINANCE, March, 1968, pp. 67-84.

Crichfield, T., Dyckman, T. and Lakonishok, J., "An Evaluation of Security Analysts' Forecasts", THE ACCOUNTING REVIEW, July, 1978, pp. 651-668.

Elton, E. and Gruber, M., "Earnings Estimates and the Accuracy of Expectational Data", MANAGEMENT SCIENCE, April, 1972, pp. 409-424.

Elton, E. and Gruber, M., Editors, SECURITY EVALUATION AND PORTFOLIO ANALYSIS, Prentice-Hall, Englewood Cliffs, New Jersey, 1972a.

Elton, E., Gruber, M. and Gultekin, M., "Professional Expectations: Accuracy and Diagnosis of Errors", JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS, December, 1984, pp. 1-13.

Fama, E., "Efficient Capital Markets: A Review of Theory and Empirical Work", JOURNAL OF FINANCE, May, 1970.

Foster, G., FINANCIAL STATEMENT ANALYSIS, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.

Fouse, W., "Risk and Liquidity: The Keys to Stock Price Behavior", FINANCIAL ANALYSTS' JOURNAL, May, 1976, pp. 35-45.

Givoly, D. and Lakonishok, J., "Properties of Analysts' Forecasts of Earnings: A Review and Analysis of the Research", JOURNAL OF ACCOUNTING LITERATURE, Spring, 1984, pp. 117-152.

Hamburg, M., STATISTICAL ANALYSIS FOR DECISION MAKING, Third Edition, Harcourt Brace Jovanovich, New York, New York, 1983.

Imhoff, E. and Pare, P., "Analysis and Comparison of Earnings Forecast Agents", JOURNAL OF ACCOUNTING RESEARCH, Autumn, 1982, pp. 429-439.

Jaggi, B., "Comparative Accuracy of Management's Annual Earnings Forecast", FINANCIAL MANAGEMENT, Winter, 1978, pp. 24-32.

Jaggi, B., "Further Evidence on the Accuracy of Management Forecasts Vis-a-Vis Analysts' Forecasts", THE ACCOUNTING REVIEW, January, 1980, pp. 76-101.

Johnston, J., ECONOMETRIC METHODS, McGraw-Hill Book Company, New York, New York, 1972.

Lorek, K., McDonald, C. and Patz, D., "A Comparative Examination of Management Forecasts and Box-Jenkins Forecasts of Earnings", THE ACCOUNTING REVIEW, April, 1976, pp. 321-330.

McDonald, C., "An Empirical Examination of the Reliability of Published Predictions of Future Earnings", THE ACCOUNTING REVIEW, July, 1973, pp. 502-510.

Neiderhoffer, V. and Regan, P., "Earnings Changes, Analysts' Forecasts and Stock Prices", FINANCIAL ANALYSTS JOURNAL, May, 1972, pp. 65-71.

Porter, G., "Determinants of the Accuracy of Management Forecasts of Earnings", REVIEW OF BUSINESS AND ECONOMIC RESEARCH, Spring, 1982, pp. 1-13.

Richards, R., Benjamin, J., and Strawser, R., "An Examination of the Accuracy of Earnings Forecasts", FINANCIAL MANAGEMENT, Fall, 1977, pp. 78-86.

Rao, P. and Miller, R., APPLIED ECONOMETRICS, Wadsworth Publishing Company, Inc., Belmont, California, 1971.

Ruland, W., "The Accuracy of Forecasts by Management and by Financial Analysts", THE ACCOUNTING REVIEW, April, 1978, pp. 439-447.

Schrueder, H. and Klassen, J., "Confidential Revenue and Profit Forecasts by Management and Financial Analysts: Evidence from the Netherlands", THE ACCOUNTING REVIEW, January, 1984, pp. 64-77.

Theil, H. APPLIED ECONOMIC FORECASTS, Amsterdam: North-Holland Publishing Company, 1966.

Whitbeck, V., and Kisor, M., "A New Tool in Investment Decision-Making", FINANCIAL ANALYSTS JOURNAL, May, 1963, pp. 55-62.

RUTH ANN MC EWEN

EDUCATION

1986	Ph. D. Management Georgia Institute of Technology Major: Accounting Minor: Finance
1983	M.S.I.M. (Management) Georgia Institute of Technology Concentrations: Accounting and Finance
1976	B.S. Medical Technology Alabama A & M University Concentrations: Biology and Chemistry

PROFESSIONAL QUALIFICATIONS

1984	Passed CPA Exam, Georgia

EXPERIENCE

1986	Assistant Professor, Temple University
1983-1986	Graduate Teaching Assistant Georgia Institute of Technology
1982	Summer Internship C & S National Bank Research and Development Government Banking Department
1978-1982	Assistant Technical Director, Pathology West Paces Ferry Hospital, Atlanta
1976-1978	Staff Technologist, Huntsville, Alabama

HONORS, AWARDS AND ACTIVITIES

1982-1985	Regents Scholarship, Georgia Tech
1983-Present	Honorary Accounting Organization Vice President, 1985
1984	American Accounting Association Doctoral Consortium Fellow
1983-Present	Beta Gamma Sigma National Honorary Fraternity
1983	Beta Gamma Sigma National Seminar
1981-1983	Graduate Students in Management Chair, Placement Committee, 1983 Inceptor Committee, 1983 Student Representative to Faculty Graduate Committee, 1983