THE EFFECT OF METHOD OF FORECAST REVISION
ON FORECAST VALIDITY: A LABORATORY STUDY

DISSERTATION
Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By
Russell Morris Barofield, B.B.A., M.A.

* * * * * *

The Ohio State University
1969

Approved by
Adviser
Department of Accounting
ACKNOWLEDGEMENTS

The process of writing a dissertation seems to consist of (1) the identification of interesting problems, (2) the discovery of acceptable solutions, and (3) the routine task of preparing the written results in polished form. In all three of these areas I received valuable help.

Professor T. J. Burns provided counseling throughout my graduate program, thereby helping me obtain the background necessary to deal with research problems. My committee chairman, Professor J. L. Livingstone, deserves a great deal of my thanks for helping me to identify problems and to work toward their solutions. Professor Y. K. Li, another committee member, also was valuable in this respect.

Finally, I owe a very real debt of gratitude to my wife Betty. Not only did she ease the tedious nature of the task by doing all the typing, but she also was tolerant of the large portion of my time which had to be devoted to the completion of my graduate program.
VITA

December 22, 1943 ... Born -- Boulder, Colorado

1965 ................. B.B.A., The University of Texas, Austin, Texas

November, 1965 ...... C.P.A. examination passed

1965-66 ............... Teaching Assistant, Department of Accounting, The Ohio State University, Columbus, Ohio

1966 ................... M.A., The Ohio State University, Columbus, Ohio

1966-69 ............... Teaching Associate, Department of Accounting, The Ohio State University, Columbus, Ohio

PUBLICATIONS

"Comments on a Measure of Forecasting Performance"
Accepted for publication by Journal of Accounting Research for Autumn, 1969 issue.

FIELDS OF STUDY

Major Field: Accounting

Psychology

Economics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgments</td>
</tr>
<tr>
<td>Vita</td>
</tr>
<tr>
<td>List of Tables</td>
</tr>
<tr>
<td>List of Figures</td>
</tr>
<tr>
<td>Chapter</td>
</tr>
<tr>
<td>I. Introduction</td>
</tr>
<tr>
<td>Purpose of Research</td>
</tr>
<tr>
<td>Definition of Terms</td>
</tr>
<tr>
<td>II. Status of the Literature</td>
</tr>
<tr>
<td>Research from Accounting and Marketing</td>
</tr>
<tr>
<td>Two Relevant Articles</td>
</tr>
<tr>
<td>Relevant Psychological Literature</td>
</tr>
<tr>
<td>III. A Model of Biasing Behavior</td>
</tr>
<tr>
<td>A Simple Model</td>
</tr>
<tr>
<td>A Model with a Response Continuum</td>
</tr>
<tr>
<td>Some Theorems of the Model</td>
</tr>
<tr>
<td>IV. An Experiment and Related Data Analysis</td>
</tr>
<tr>
<td>Chapter Purpose</td>
</tr>
<tr>
<td>Experiment</td>
</tr>
<tr>
<td>Data Analysis</td>
</tr>
<tr>
<td>Conclusions</td>
</tr>
<tr>
<td>V. Summary and Possible Extensions</td>
</tr>
<tr>
<td>Appendix</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>Bibliography</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hypothetical Density Function for Output</td>
<td>22</td>
</tr>
<tr>
<td>2. Expected Value of Performance for Various Forecasts</td>
<td>22</td>
</tr>
<tr>
<td>3. Probability of R's Adjustment on any Trial</td>
<td>31</td>
</tr>
<tr>
<td>4. Experimental Design with Actual Data</td>
<td>59</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Tree Diagrams for Discrete Response Case and Corresponding Transition Matrix</td>
<td>32</td>
</tr>
<tr>
<td>2.</td>
<td>Predicted vs. Actual Mean Responses for the First 35 Trials</td>
<td>73</td>
</tr>
<tr>
<td>3.</td>
<td>Predicted vs. Actual Mean Responses for the Second 35 Trials</td>
<td>74</td>
</tr>
<tr>
<td>4.</td>
<td>Two Predicted Limiting Distributions vs. the Actual Distribution of Responses for Case I</td>
<td>81</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Purpose of the Research

Quite often accountants are assigned partial responsibility for deciding what data should be gathered and processed by the accounting system. Recognizing this responsibility accountants have attempted to devise criteria for evaluating sources of data. Four criteria recently suggested by an American Accounting Association committee are relevancy, freedom from bias, quantifiability, and verifiability. ¹ This same committee, possibly reflecting a trend in accounting, placed major emphasis on the relevancy criterion. Perhaps this emphasis explains why articles in the accounting literature seem to indicate a growing interest in forecasts or projections as accounting data. For example, Cooper, Dopuch and Keller suggest that the public might benefit if firms were required to prepare pro forma financial statements for external

reporting purposes. 2

In examining a source of forecasts such as a division manager, a salesman or some other potential participant in a budgeting process, the accountant, if he is primarily concerned with relevance, is really interested in the amount of potential information available from the source being considered. In the case of forecasts, accountants are aware that environmental conditions frequently exist within the firm which cause the forecaster to desire to bias his forecast. The presence of bias in forecasts could be a problem of considerable importance because of the large number of forecasters who, while operating under various payoff structures, can be involved in preparing a statement such as an investment schedule or a pro forma statement of financial position. Secondly, consider the large number of people who in the hierarchical structure of the budgeting process act as recipients to remove, as best they can, bias from the forecasts they receive.

One question that might be asked is, "Does the

---

method used to revise forecasts in an attempt to remove bias affect the amount of information present in the forecasts?" This research will investigate a possible link between R's (the recipient's) method of forecast revision and the validity of F's (the forecaster's) forecast. Theoretically, the link is that R's method of forecast revision affects F's biasing behavior, which in turn affects the validity of F's forecast. R's method of forecast revision is chosen as the independent variable in this study because the accountant may frequently have control over the method of forecast revision used by a recipient. Occasionally the accountant may be the recipient (possibly from his position as chief budgetary officer) or he may be in an advisory or staff relationship to the recipient. In either case, if method of forecast revision does affect forecast validity, the accountant could act to improve the validity of a source of forecasts by recommending, or by using, a particular method of forecast revision.

The type of forecast to be studied is the personal forecast. Some of the forecasts processed and analyzed by accountants are arrived at through the use of statis-

---

3Measures of accuracy and validity are thought of as possible measures of information content. They will be defined shortly.
tical techniques. Generally, these techniques create forecasts by extrapolating historical data into the future. Examples of such techniques are exponential smoothing, moving averages and least square procedures. Other forecasts are created by individuals without the aid of statistical techniques. More importantly, these other forecasts may be created in a manner which is not observable. These non-observable or personal forecasts are subjective in nature and rely primarily on an individual's knowledge of the particular item being forecasted. An example of a personal forecast would be a sales forecast prepared by the "sales force composite" method. In this method, each salesman is asked to forecast the level of sales he feels will occur in his territory in some future period of time. The individual personal forecasts are then reviewed, sometimes revised, and then summed to form the company's sales forecast.

An experiment has been designed to test the null hypothesis that the method of forecast revision chosen by R will not affect the validity of F's forecast. The data from this experiment will also be used to examine a model of F's biasing behavior which does provide support for the alternative hypothesis that R's choice of method of forecast revision will affect the validity
of F’s forecast. This model is a modified version of a model frequently used in psychology to describe learning behavior. The development of the model is described in Chapter III. The test of the null hypothesis and the examination of the data for correspondence with the model is contained in Chapter IV.

**Definition of Terms**

Several terms have been used in this chapter without being defined. The purpose of this section is to define and relate the terms "accuracy," "validity" and "intentional bias."

**Accuracy**

Accuracy concerns the magnitude of error or incorrectness of the forecast. For example, if $F_1$ represents a forecast of the value of a variable at time $i$ and $A_1$ represents the actual value of a variable at time $i$, then $(F_1 - A_1)^2$ would be a measure of the accuracy (or inaccuracy) of this particular forecast. The accuracy of a particular source of forecasts (such as $F$) might be the average squared difference between $F_1$ and

---

4The larger the measure of accuracy the less the accuracy of the forecast. Possibly, it is less confusing to call the statistic a measure of inaccuracy.
Thus, the accuracy of any particular source of forecasts would equal 
\[ \frac{1}{N} \sum_{i=T-N+1}^{T} (A_i - F_i)^2 \] where \( N \) is the given number of most recent forecasts and \( T \) represents the number of the most recent forecast.

**Validity**

Validity is defined in terms of how well the forecast can be used to predict the actual. This definition allows the forecast to be included in some form of prediction equation. One measure of validity might be the correlation between past forecasts and actuals. Such a measure of validity implies a linear prediction equation where past data is used to estimate parameter values.

To illustrate the difference between accuracy and validity, consider the following sequence of forecasts and actuals:

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>5</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>27</td>
</tr>
<tr>
<td>8</td>
<td>26</td>
</tr>
</tbody>
</table>
The accuracy of these forecasts is very poor. However, a brief examination should convince the reader that the forecast has a high degree of validity since the two variables have a negative correlation of 1.0. If this relationship holds for all forecasts, then a linear regression equation of Actual = \( 34 - \text{Forecast} \) would allow a user of the forecast to predict the actual exactly, even though the forecast is inaccurate.

**Bias**

To define bias assume the existence of an "ideal observer" who (1) has adopted the stated goal of the firm as his own, and (2) chooses that figure to forecast which maximizes the expected value of the firm's performance. The actual forecaster develops an expectation and the forecast of the ideal observer represents an unintentional form of bias. Differences between the forecaster's reported forecast and his expectation will be called intentional bias. Presumably, such differences would result from a consideration of the effect various forecasts might have on the achievement of personal goals.

This research is concerned with intentional bias. The model described in Chapter III is a model of inten-
tional biasing behavior, and the experiment described in Chapter IV has been designed in an attempt to minimize the presence of unintentional bias. Hereafter, unless stated otherwise, the presence of unintentional bias is assumed away and the term bias can be taken to mean intentional bias.

Given certain assumptions, it can be shown that bias will affect both the accuracy and validity of a forecast. The assumptions are

1) $q$ is a random variable representing the forecast of the ideal observer.

2) $s$ is a random variable representing the actual event being forecasted.

3) $x$ is a random variable which represents the amount of bias added to $q$ to obtain F's forecast. $x$ is assumed to be independent of both $q$ and $s$.

Assume the set of $\{ (F_i - A_i) \mid i = T-N+1, T-N+2, \ldots, T \}$ constitutes a random sample from a population of forecast errors. Then, the statistic $\frac{1}{N} \sum_{i=T-N+1}^{T} (F_i - A_i)^2$ can be thought of as an estimator of $E [(q + x - s)^2]$. This expectation is commonly known as the second moment about zero of the random variable $(q + x - s)$. The accuracy of the ideal

\[ \text{The assumption of independence will be tested in Chapter IV.} \]
observer's forecast is \( E[(q - s)^2] \). The accuracy of the forecaster's forecast is \( E[(q + x - s)^2] \). The two measures of accuracy are related in the following manner:

\[
E (q + x - s)^2 = E (q - s)^2 + E(x^2) + 2 E(x) \{ E(q) - E(s) \}
\]

From this relationship it is apparent that the accuracy of F's forecasts is a function of the mean and second moment about zero of the bias random variable \( x \). Furthermore, with the reasonable assumption that \( E(q) \) equals \( E(s) \), one can conclude that the accuracy of F's forecasts varies directly with the second moment about zero of the bias random variable \( x \). Finally, the second moment about zero of any random variable, including \( x \), is related to its variance in the following manner:

\[
E(x^2) = \text{Variance}(x) + \left[ E(x) \right]^2
\]

Thus, the accuracy of F's forecasts can be affected by a change in either the mean or the variance of \( x \).

The inclusion of the bias random variable in...

---

\(^6\)The assumption is reasonable if the firm's attitude toward forecast error is such that the direction of error is irrelevant. Under these conditions, one would expect the average error of the ideal observer to be zero.
the forecast will also affect the validity of the forecast. The correlation coefficient $r_{qs}$ represents the validity of the "ideal observer's" forecast. The correlation coefficient $r_{qs}$ represents the validity of the forecaster's forecast. The development of the relationship between $r_{qs}$ and $r_{qs}$ is as follows:

$$r_{A,B} = \frac{E (A - \mu_A) (B - \mu_B)}{\sigma_A \sigma_B}$$

where $\mu_A = E(A)$, $\mu_B = E(B)$

$$r_{q+x,s} = \frac{E (q + x - \mu_{q+x}) (s - \mu_s)}{\sigma_{q+x} \sigma_s}$$

where $\sigma_{q+x} = \sqrt{\sigma_q^2 + \sigma_x^2}$ since $q$ and $x$ independent

$$= \frac{E (q - \mu_q) (s - \mu_s)}{\sigma_s \sigma_q} + \frac{E (x - \mu_x) (s - \mu_s)}{\sigma_s \sigma_q \sqrt{\sigma_q^2 + \sigma_x^2}}$$

$$= E \left[ (s - \mu_s)(q - \mu_q) \right] \cdot \frac{1}{\sigma_s \sigma_q \sqrt{\sigma_q^2 + \sigma_x^2}}$$

because $s$ and $x$ independent
\[
\frac{E \{(s - \mu_s) (q - \mu_q)\}}{\sigma_s \sigma_q} \cdot \frac{\sigma_q}{\sqrt{\sigma_q^2 + \sigma_x^2}}
\]

\[
= \frac{\sigma_q}{\sqrt{\sigma_q^2 + \sigma_x^2}} \ r_{x,y}
\]

Thus,

\[
r_{q+x,s} = \frac{\sigma_q}{\sqrt{\sigma_q^2 + \sigma_x^2}} \ r_{q,s}
\]

This relationship shows that the validity of the forecaster's forecast is inversely related to the variance of the bias random variable \( x \).

Summary of Chapter

The purpose of this chapter was to state the research problem and to define terms which will be used throughout the study. The purpose of the research was stated to be to investigate a possible link between the method of forecast revision chosen by R and the validity of F's forecast. The relationship developed in this chapter between the validity of F's forecast and the variance of the bias random variable \( x \) will be shown to provide part of this possible link.
CHAPTER II
STATUS OF THE LITERATURE

Introduction

The purpose of this chapter is to discuss some of the published research which has dealt with or can be applied to the general topic of bias in forecasts. The discussion is not limited to the effect of method of forecast revision on F's forecast validity because it appears that no research has previously been done on this subject.

Research Drawn from Accounting and Marketing

Little systematic research has been done in accounting or marketing on the general subject of bias in forecasts. What little that has been said on the subject has appeared primarily in the budgeting and sales management literature. A statement typical of these sources is one by Welsch. In discussing the accuracy of salesmen's forecasts, he states

On the other hand, some salesmen are too optimistic, or conversely, some may turn in low estimates as a matter of self protection. They may not give adequate attention to the problem and thus improperly evaluate the
general market potential. These tendencies can be largely overcome through a program of budget education.¹

A more specific statement, which apparently resulted from a questionnaire sent to member firms by the National Industrial Conference Board, is

It is difficult for the sales force to bring to the sales forecasting job the impartial, objective viewpoint so necessary for good forecasting. There seems to be some evidence, for example, that if the forecast is to become a basis for setting quotas, salesmen will underestimate demand in order to make their later performance appear good by comparison. One company found that estimates averaged 30% to 40% below actual performance. On the other hand, when a product is in short supply, salesmen's forecasts tend to have an upward bias, inspired by the desire to get larger allotments of goods.²

These two statements do not indicate that accountants know a great deal about the subject of bias in forecasts. However, the statements are definitely typical of the accounting and marketing literature. The more rigorous work on the subject has been done in two articles from the literature of management and


organization theory. These two articles are the subject of the next section.

Two Relevant Articles

Cyert, March and Starbuck have published the results of some laboratory research designed to test "the effects of biased and unbiased pay-off structures on estimation within an organization." Subjects were grouped into three-man teams. The task of the group was to estimate the area of a rectangle. One member (Length Estimator) estimated one dimension of the rectangle basing his estimate on the length of a line projected on a screen which only he could see. The second member (Width Estimator) made an estimate of the other dimension of the rectangle in a similar manner. These two members reported their estimates to the third member whose responsibility it was to make the area estimate. Subjects were individually rewarded on the basis of error in area estimation. All pay-off structures gave the highest reward for an accurate forecast (absolute error within 5% of actual area), but rewards declined in an asymmetric manner about an accurate

---

forecast. Pay-off structures were assigned to individual subjects so that conflict within the group was common. Finally, the only feedback to subjects was the percentage error in area estimation.

The results of this study led the authors to conclude that the recipient of biased information is remarkably able to adjust for the presence of bias. They say

In particular, it seems clear that in an organization of individuals having about the same intelligence adaptation to the falsification of data occurs fast enough to maintain a more or less stable organizational performance.

...For example, we—as well as others—have argued that the fact of internal informational bias has to be dealt with explicitly in a theory of the firm. These results cast doubt on such an argument. They indicate that perhaps such phenomena, important as they may be to an understanding of the internal operation of the organization, may not be of particular significance to a theory of organizational choice.¹

The authors were justifiably cautious in the manner in which they stated their conclusions. Enough questions about experimental design and data analysis can be asked to place some of the results in doubt. An unrealistic aspect of the experimental design is that length and width estimators received feedback only on

¹Ibid., p. 264.
error in area estimation. The area estimator would seem to be at an advantage in the sense that he knew after each trial what change in his own forecast would have caused him to receive the highest payoff. This is not true in the situation of either line estimator. What effect a change in the quality of feedback would have on the results of this experiment is impossible to say.

Also, the data had some puzzling characteristics which the authors did not attempt to explain. For example, in their analysis the authors calculated separate regression equations of line estimates on actual line lengths for those subjects favoring (a) underestimation of area, (b) overestimation of error and (c) accurate estimation of error. The expected slope coefficient for the underestimation case would be less than one and the expected intercept term would be zero. The expected slope coefficient for the overestimation case would be greater than one. The actual slope coefficient was not only greater than one but also slightly greater than the slope coefficient for the overestimation case. The authors did not attempt to explain this result.

The questions just mentioned were raised not to criticize the research done by Cyert, March and Starbuck
but to point out that the results of this study indicate only that further research on the subject is needed.

One example of additional research is a field study by Lowe and Shaw. Their conclusions do not apply directly to the issues concerning Cyert, March and Starbuck, but their conclusions do deal with the subject of bias in forecasts. The contributions of this field study were (1) to report on some factors which appear to cause bias to be incorporated into forecasts and (2) to report that

Some examples from individual cases enable us to conclude tentatively that forecasting bias was not fully excluded from the budget because of the difficulty of recognizing the existence and extent of bias.

An alternative explanation put forth by Lowe and Shaw is that the existence and extent of bias was recognized but that the superior hesitated to reject the judgement of a subordinate because of the possible effect the rejection might have on the subordinate.

This appears to say that in deciding what forecast to


6 Ibid., p. 312.
report for budgetary purposes the superior and the subordinate are operating under payoff structures biased in the same direction. Cyert, March and Starbuck might hypothesize that the superior, in making decisions other than what figure to accept for budgetary purposes, would base his decision on an adjusted forecast. Lowe and Shaw provide no information bearing directly on this hypothesis.

**Relevant Psychological Research**

The articles previously discussed initially assumed and then supported the hypothesis that the personal goals and objectives of the forecaster can cause bias to be present in forecasts. An obvious approach to removing bias from forecasts is to specifically include in performance measures a reward for forecasting accuracy. Ijiri, Kinard and Putney have made just this suggestion in a recently published article. This section will discuss some of the possible effects performance measures of the type Ijiri, Kinard and Putney suggest might have on the behavior of a forecaster.

---

The point of importance is that the functional form of the performance measure may cause the forecaster to systematically deviate from reporting the figure the firm would consider the optimal forecast. Included in the discussion will be some references to relevant work done in psychology.

A specific algebraic formulation of one performance measure suggested by Ijiri, Kinard and Putney is

\[ P = -k_1 |A - F| + k_2 A \]

- \( P \) = the level of performance attained
- \( A \) = actual level of operations in units
- \( F \) = expected or forecasted level of operations in units
- \( k_2 \) = the contribution or profit margin per unit
- \( k_1 \) = the penalty coefficient for misestimation

\( k_1 \), in the general formulation of the performance measure, could be either a constant function or a function which varies with the magnitude of forecast error.

In discussing \( k_1 \), the authors state:

The penalty coefficients for misestimation must be determined after careful study of the forecasting displacement costs involved, as well as other factors such as profit margins. It

---

8Ijiri, Kinard and Putney originally used \( E \) to denote the expected level of operations in units. The symbol \( E \) will be reserved here to denote the expectation operator of statistics.
is difficult to make a general statement about how one should actually calculate the penalty coefficients; however, in general, we can say that (1) if the forecasting displacement costs involved are large, the penalty coefficients should also be made large relative to the operating performance measure... (2) the penalty coefficients should not be made so large as to discourage operations beyond the estimated level insofar as such operations are desirable from the standpoint of the organization as a whole.9

Two questions arise about the case where \( k_1 \) is a constant:

1. What value, in a normative sense, should the forecaster predict?

2. In practice, is it likely that the performance measure \(-k_1 |A - F|\) will encourage the forecaster to predict the value that our normative theory specifies?

Question 1

Assume a forecast represents a choice of a measure of central tendency from the subjective probability density function \( f(p) \) of unit operations \( p \).10

---

9Ijiri, Kinard and Putney, op. cit., p. 9.

10 We will assume \( f(p) \) satisfies the requirements of a probability density function. A list of these requirements may be found in R. V. Hoag and A. T. Craig, Introduction to Mathematical Statistics (New York: The Macmillan Co., 1965), p. 11. We should point out that experimental tests frequently show that subjective probability distributions violate some of these requirements. See W. Edwards, "Behavioral Decision Theory" Annual Review of Psychology, XII (1961), 473-98.
The penalty for forecast errors is measured by the portion of the formula \(-k_1 |A - F|\). If we assume that the forecaster wants to maximize his performance measure, then the forecaster should forecast \(p = b\) such that the expected value of \(k_1 |p - b|\) is a minimum. It can be shown that \(E (k_1 |p - b|) = k_1 E (|p - b|)\) is a minimum when \(b\) is the median value of the probability density function \(f(p)\). \(^{11}\)

An example of a performance measure having the mode as an optimal prediction would be: \(^{12}\)

\[
\text{Performance} = \begin{cases} 
-100 & \text{if } A \neq F \\
0 & \text{if } A = F 
\end{cases} + k_2 A
\]

To illustrate that with this measure the mode is an optimal strategy, assume that the forecaster feels that operations will consist of 0 to 3 units of product being sold and that the probability of each level of sales occurring can be described by the discrete probability distribution.

\(^{11}\)The proof that \(E (|p - b|)\) is a minimum for \(b\) equal to the median is a problem assigned in introductory texts in mathematical statistics. See E. Parzen, Modern Probability Theory and Its Applications (New York: John Wiley and Sons, 1960), p. 213.

\(^{12}\)This performance measure is the one measure discussed in this paper which does not satisfy the conditions specified by Ijiri, Kinard and Putney, op. cit., p. 3. Their conditions require that the degree of error affect the performance measure.
density function $f(p)$ shown in Table 1.

<table>
<thead>
<tr>
<th>(p)</th>
<th>(f(p))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8/20</td>
</tr>
<tr>
<td>1</td>
<td>4/20</td>
</tr>
<tr>
<td>2</td>
<td>4/20</td>
</tr>
<tr>
<td>3</td>
<td>4/20</td>
</tr>
</tbody>
</table>

To maximize performance, the forecaster should predict \(b\) such that \(\$0 f(b) - \$100 (1 - f(b))\) is a maximum.

Table 2 shows that the modal value \(b = 0\) maximizes the portion of the performance measure which relates to forecasting accuracy.\(^{13}\)

<table>
<thead>
<tr>
<th>(b)</th>
<th>($0 f(b) - $100 (1 - f(b)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-$60.00</td>
</tr>
<tr>
<td>1</td>
<td>-$80.00</td>
</tr>
<tr>
<td>2</td>
<td>-$80.00</td>
</tr>
<tr>
<td>3</td>
<td>-$80.00</td>
</tr>
</tbody>
</table>

\(^{13}\) Implicit in this discussion is the assumption that the forecaster's choice of a value of \(p\) to predict has no influence on the value of \(p\) which actually occurs.
Alternative measures of performance can be devised which make the optimal choice in prediction the mean. As a general statement about measures of central tendency, to be right as often as possible, predict the mode; to minimize absolute error, predict the median; and to minimize squared error, predict the mean.\textsuperscript{14}

The value that a forecaster should predict will not always be one of the three common measures of central tendency mentioned above. For example, when the value of \( k \) depends upon the direction or sign of forecast error, the optimal prediction will rarely be one of the measures of central tendency. An example of such a value of \( k \) might be:

\[
\text{Performance} = -k_1 \left| A - F \right| + k_2 A
\]

where \( k_1 = \begin{cases} \$2 & \text{when } A \leq F \\ \$1 & \text{when } A > F \end{cases} \)

In this case, specific knowledge of the probability density function of operations, \( f(p) \), would be required before the optimal prediction strategy could be specified. To illustrate, consider the probability density function \( f(p) = (1/50)p, \ 0 \leq p \leq 10 \). The mean, median and mode

of this distribution are 6.67, 7.07 and 10.0, respectively. To find the optimal strategy, we must choose b such that \( E (k_1 \mid p - b) \) is a minimum.

\[
E (k_1 \mid p - b) = \int_{0}^{10} k_1 \mid p - b \mid f(p) \, dp
\]

\[
= \int_{0}^{10} k_1 \mid p - b \mid (1/50)p \, dp
\]

\[
= b \int_{0}^{10} \frac{1}{2} (p - b) (1/50)p \, dp + \int_{0}^{10} \frac{1}{2} (b - p) (1/50)p \, dp
\]

which simplifies to \((b^3/100) - b + (20/3)\).

Setting the first derivative with respect to b equal to zero, we find a maximum or a minimum value \(b^*\) which satisfies \((3b^*^2/100) - 1 = 0\), or \(b^* = 5.77\). Notice that the optimal prediction in this case is lower than any of the three measures of central tendency discussed above.

---

15 The mean, median and mode are derived as follows:

Mean = \( E(p) = \int_{0}^{10} p f(p) \, dp = \int_{0}^{10} (1/50)p^2 \, dp = 6.67\)

Median = \( \int_{m}^{10} (1/50)p \, dp = (m^2/100) = \frac{1}{2} \) or \( m = 7.07\)

Mode = 10.0 since \( f(p) \) is a strictly increasing function of \( x\).

16 The second derivative is \((6b/100)\) which is positive for values of \( b \) in the domain of \( f(p) \), so \( b^* = 5.77 \) is a minimum.
Question 2

The second question deals with whether or not a performance measure like \( \text{Performance} = -k_1 |A - F| + k_2A \) will actually encourage the forecaster to predict the value which is specified by our normative theory. There is some evidence that it may not. Peterson and Miller, in a sequence of two laboratory experiments using students as subjects, tested the subject's ability to adopt prediction strategies. Each of three groups of subjects had either the mean, median or mode as an optimal prediction strategy. They summarized their results as follows:

The clear conclusion... is that S's are surprisingly accurate in minimizing expected cost under mode and median cost conditions, but not under mean cost conditions.\(^{17}\)

One possible explanation offered for the failure of subjects to adopt the mean prediction strategy was that the "apparatus (used in the experiment) emphasized error too much and cost not enough."\(^{18}\) A second experiment provided only a small amount of support for this explanation. In both of these experiments, subjects

\(^{17}\)C. Peterson and A. Miller, "Mode, Median and Mean as Optimal Strategies," *Journal of Experimental Psychology*, LXI (October, 1964), 366.

\(^{18}\)Ibid.
whose optimal strategy was to predict the mean consistently erred in the direction of the median. Apparently, more complex measures such as \( \text{Performance} = -k_1 (A - F)^2 + k_2 A \) hinder a subject's ability to develop optimal strategies.

**Summary**

Actually, there has been little systematic research into the problem of bias in forecasts, and no research into the effect of the recipient's method of forecast revision on the presence of bias in forecasts. The research concerning bias in forecasts has been described in this chapter. The research which is the subject of this thesis does not build on any one article or research project but, in a sense, is viewed as an initial step to provide an understanding of one particular aspect of the bias problem. This effort is contained in the following chapters.
CHAPTER III
A MODEL OF BIASING BEHAVIOR

Chapter Purpose

The purpose of this chapter is to present a model of intentional biasing behavior. The approach will be to assume that the reported forecast is the sum of (1) that figure \( V \) which \( F \) thinks is an unbiased estimate of sales and (2) a bias component \( X \). Thus, the reported forecast for time period \( n \) is \( (V_n + X_n) \). The model presented here tries to provide a prediction of \( F \)'s choice of \( X_n \). This model provides support for the view that \( R \)'s method of forecast revision will affect the validity of \( F \)'s reported forecast.

The basic model is drawn from the psychological literature of learning models. More specifically, the model might be labeled a one-element stimulus-sampling learning model with a response continuum. The major portion of the development and testing of this specific model can be found in Suppes (1960), Suppes and Atkinson (1960), Suppes and Frankmann (1961), Suppes and Rouanet
Several well-known learning models could have been used to describe F's biasing behavior. Three such models are (1) the linear model, (2) the one-element stimulus-sampling model and (3) the n-element stimulus-sampling model. Characteristics which favor the stimulus-sampling models over the linear model are:

1. The stimulus-sampling models more explicitly provide a theoretical description of the process underlying learning,
2. Stimulus-sampling models are generally easier to work with mathematically.

The two stimulus-sampling models differ in their assumption of the number of elements which make up a set of stimuli. The assumption of a single element seems to correspond most with the experimental environment set up (and described in Chapter IV) to test F's biasing behavior. The item in the experiment which corresponds with the single stimulus element is R's "An introductory discussion of stimulus sampling theory can be found in Atkinson, Bower and Crothers, An Introduction to Mathematical Learning Theory (New York: John Wiley & Sons, 1965), Chapter 8."

1 An introductory discussion of stimulus sampling theory can be found in Atkinson, Bower and Crothers, An Introduction to Mathematical Learning Theory (New York: John Wiley & Sons, 1965), Chapter 8.

request for a forecast. R, on each trial, asks "What is your forecast?" The terminology and the manner in which the request is conveyed to F never varies. Thus, the experimental environment seems to support the one-element model. For this reason, the one-element model is used in this research.

Finally, all three models have several common predictions. In fact, all three predict the same mean response curve and the same asymptotic distribution of responses. The models do differ in their sequential predictions. However, because of the time required to receive and record a single response, it was not feasible to collect in this experiment the amount of data necessary to test sequential predictions.

A Simple Model

A stimulus sampling model with a discrete number of bias responses X will be presented first for purposes of illustration. The assumptions of the simple model are

1. Each subject (S) has a set of stimuli. Each element of this set of stimuli is conditioned to a particular response. For the models in this research I assume a stimulus set containing a single stimulus element.

2. When called upon to make a response, S samples a single stimulus. The response
to which this stimulus is conditioned is made.

3. If the correct response is made, the sampled stimulus remains conditioned to that correct response with probability one. If an incorrect response is made, with probability $\theta$ ($\theta$ constant over trials) the sampled stimulus becomes conditioned to the response which would have been correct if made. With probability $(1 - \theta)$ the stimulus remains conditioned to the response made.

An example of a forecast bias problem stated in the framework of stimulus sampling theory will now be given. Assume that $F$ makes an accurate personal estimate of the mean of the distribution of possible sales. $F$ may report this figure, bias it upward 1 unit or bias it upward 2 units.

Furthermore, $R$ may accept a forecast or may adjust it downward 1 unit. $F$ feels his biasing behavior is successful if the adjusted forecast is exactly one unit greater than $F$'s personal estimate of the mean of the distribution of possible sales.

I will assume that $R$'s behavior on any single trial can be described by the discrete probability distribution shown in Table 3.
TABLE 3
PROBABILITY OF R'S ADJUSTMENT
ON ANY TRIAL

<table>
<thead>
<tr>
<th>Adjustment</th>
<th>Probability of Adjustment*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>b</td>
</tr>
<tr>
<td>-1</td>
<td>a</td>
</tr>
</tbody>
</table>

* a + b = 1

The assumptions along with the density function shown in Table 3 allow the construction of a transition matrix for a Markov chain. The derivation of transition probabilities and the resulting transition matrix are shown in Figure 1.

The properties of Markov chains can be used to state asymptotic response probabilities and to develop other predictions of the model. For example, the matrix shown in Figure 1 is that of an absorbing Markov chain. Or, in other words, state 0 is a transient state. Once F leaves state 0, he will never return. If F starts in state 0, his mean time to absorption is

\[ \frac{1}{1 - (a + b(1 - 0))} = \frac{1}{b0}. \]

The probability of

---


4 Ibid., p. 46.
Adjustment of $F$ at time $n+1$

Adjustment of $F$ at time $n$

\[
\begin{array}{ccc}
0 & +1 & +2 \\
0 & a + b(1-c) & b(e) & 0 \\
+1 & 0 & a(1-e) + b & a(e) \\
+2 & 0 & b(e) & a + b(1-e) \\
\end{array}
\]

**Figure 1**: Tree diagrams for discrete response case and corresponding transition matrix
absorption is 1. Thus, the model predicts that upon reaching an asymptotic level of behavior, response 0 will never be made. If \( u_1 \) represents the limiting probability of \( F \) making response 1, then the asymptotic behavior of \( F \) may be calculated by solving the following system of two equations:

\[
\begin{align*}
\frac{\partial u_1}{\partial u} &= (a(1 - \theta) + b) u_1 + b(\theta) u_2 \\
\frac{\partial u_2}{\partial u} &= a(\theta) u_1 + (a + b(1 - \theta)) u_2
\end{align*}
\]

with the restriction that \( u_1 + u_2 = 1 \).

The solution is \( u_1 = b \) and \( u_2 = a \). Notice how the behavior of \( R \) effects the asymptotic predicted behavior of \( F \). \( R \) makes response (-1) with probability \( a \) and response (0) with probability \( b \). The model predicts that \( F \) will make response (+2) with probability \( a \) and response (+1) with probability \( b \). So, the adjustment behavior of \( R \) determines the bias behavior of \( F \). Secondly, the particular values of \( a \) and \( b \) determine the variance of \( F \)'s biasing behavior which in turn is the specific item which (under assumptions previously stated) has been shown to affect the validity of \( F \)'s forecast.

\[5\text{Ibid., p. 72.}\]
A Model with a Response Continuum

An undesirable characteristic of the model just discussed is the arbitrary limitation placed on the number of bias and adjustment responses available to F and R. To avoid this characteristic, the model will be modified to include a response continuum. The major change is an assumption that each stimulus is conditioned to more than one response. This conditioning is described by what Suppes has labeled a "smearing distribution." This smearing distribution is actually represented by a probability distribution which indicates the probability of a response given the particular stimulus sampled.

For purposes of this research, the smearing distribution is assumed to be a normal probability distribution with constant variance. Reinforcement changes conditioning by changing the mean of the smearing distribution. For convenience sake, the stimulus will be said to be conditioned to the mean of the smearing distribution. The assumptions of the simple model may now be modified. The new assumptions are

---

Each S has a set of stimuli. Each element of this set of stimuli is conditioned to a particular range of responses. This conditioning is represented by the smearing distribution which is a normal probability distribution with constant variance. For this model I assume a stimulus set containing a single stimulus element.

When called upon to make a response, S samples a single stimulus. The probability of a particular range of responses is determined by the smearing distribution.

If the correct response is made, the conditioning of the sampled stimulus remains the same. If an incorrect response is made, with probability $\theta$ ($\theta$ constant over trials) the mean of the smearing distribution becomes the response which would have been correct if made. With probability $(1 - \theta)$ conditioning remains the same.\(^7\)

Two reasons exist for assuming the smearing distribution has a normal shape. First, the normal distribution is mathematically tractable. This reason can be allowed to play a major role in model formation when the modeling effort is an initial one. Second, it is reasonable to assume that one cause of the smearing phenomenon is variations in F's perception of the same reinforcement. Here I will rely on a statement by Green and Swets (1966). They say

Since we often think that sensory events are composed of a multitude of similar, smaller events, which are by and large independent,

\(^7\)Ibid., pp. 271-72.
the central limit theorem might be invoked to justify the assumption of a Gaussian distribution of net effects.

The important result of the central limit theorem referred to by Green and Swets is that the sum of \( n \) independent and identically distributed random variables, each having mean \( m \) and finite variance, has a distribution which approaches normality as \( n \) approaches infinity.\(^8\)

---

Some Theorems of the Model

The following random variables are used in the model. Let:

\( x_n \) - The bias included in a forecast on trial \( n \). The density function of the response \( x_n \) is \( r_n(x) \). The mean value of \( r_n(x) \) is denoted by \( \mu_n \).

\( z_n \) - The mean of the smearing distribution on trial \( n \). The density of \( z_n \) is \( g_n(z) \). \( k(x;z) \) is the smearing distribution.

\( y_n \) - The reinforcement on trial \( n \). This random variable indicates to the subject the correct response (degree of bias) on trial \( n \). The density function of \( y_n \) is \( f_n(y) \).

---


Two theorems have been proved concerning the axioms (assumptions) and these random variables. They are

Theorem 1 \[ r_{n+1}(x) = -\infty^{+\infty} k(x;z) g_{n+1}(z) \, dz \] (3.1)

Theorem 2 \[ g_{n+1}(z) = (1 - \theta) g_n(z) + \theta f_n(z) \] (3.2)

**Environmental Assumptions**

Certain assumptions are necessary to relate the model just presented to the environment of the biasing situation. These assumptions are

1. F may not know the actual level of sales which will occur but he does feel he knows an unbiased estimate of sales. F forms his reported forecast by adding a bias factor to the unbiased estimate of sales. This bias factor is independent of the unbiased estimate.

2. F is personally rewarded or reinforced if the adjusted forecast is c to d units less than the actual (c < d). Here, I will argue that reinforcement is an "on-off" situation with reinforcement occurring over a limited range of bias.

3. After each forecast, F sees the adjustment made by the recipient R, and the actual level of sales. Knowledge of (a) a good estimate of sales, (b) the adjustment just made by R, and (c) the actual level of sales allows F to perceive what amounts of bias would have caused him to be

---

10 P. Suppes and R. C. Atkinson, op. cit., pp. 273-74 has the proofs of Theorem 1 and Theorem 2.
successful at forecasting on this trial. This perception is represented by the random variable $y_n$ in the model.

The reinforcement random variable $y_n$ needs a little additional discussion. The value of $y_n$ on trial $n$ simply shows $F$ what amount of bias on trial $n$ would have resulted in a successful forecast. Consider the following:

$u_n$ - a random variable representing the amount by which $R$ revises $F$'s forecast on trial $n$.

$v_n$ - a random variable representing the unbiased estimated of sales on trial $n$.

$w_n$ - a random variable representing the difference between unbiased estimate and actual level of sales on trial $n$. In other words, the random variable $(v_n + w_n)$ represents the actual level of sales on trial $n$. $w_n$ is assumed to be distributed $n(0, \lambda^2)$.

$m - m = (c + d)/2$. $F$ is successful if the revised forecast is $c$ to $d$ units below actual level of sales. The midpoint $m$ of this interval is chosen as the single point which is representative of the interval.

If $F$ incorporated bias equal to $y_n$ into his trial $n$ forecast, then
Actual - Revised Forecast = m

or, \((v_n + w_n) - (v_n + y_n + u_n) = m\)

\(w_n - y_n - u_n = m\)

or, \(y_n = w_n - u_n - m\) \(3.3\)

**Specification of** \(y_n\) **for Experimental Purposes**

Model predictions will be developed for two forms of \(y_n\). These will be labeled Case I and Case II.

**Case I**

\(y_n\) will be considered as being independent of \(x\) and \(n\) and as having a distribution which is normal. Whenever \(y_n\) is independent of \(n\) and \(x\), reinforcement is said to be noncontingent.

Conceptually, Case I corresponds to an attitude on the part of \(R\) that \(F\) is biased on the average some number of units but that the average bias does not change over time.

**Case II**

In the second case, \(u_n\) will be assumed to be \(F\)'s forecast error on trial \(n - 1\). In other words,
\[ u_n = \text{Actual}_{n-1} - F's \text{ Forecast}_{n-1} \]
\[ = (v_{n-1} + w_{n-1}) - (v_{n-1} + x_{n-1}) \]
\[ = w_{n-1} - x_{n-1} \]

Therefore,
\[ y_n = (w_n - u_n - m) = (w_n - w_{n-1} + x_{n-1} - m) \quad (3.4) \]

This form of \( y_n \) reflects the view that \( R \) feels \( F \) changes his biasing behavior frequently and that \( R \) must constantly try to adjust for these changes.

**Model Predictions**

Theorem 3 contains the model's predictions for Case I and Theorem 4 contains the model's predictions for Case II.

**Theorem 3**

If:
(a) \( k(x; z) \) is distributed \( n(z, \sigma^2) \), and
(b) \( f_{n}(y) \) is distributed \( n(\mu, \delta^2) \) not dependent on \( n \) and independent of \( x_n \) for all \( n \),

Then: \( r(x) \), the limiting form of \( r_n(x) \) is distributed \( n(\mu, \sigma^2 + \delta^2) \).
Proof

By Theorem 1 \( r_n(x) = \int_{-\infty}^{\infty} k(x;z) \varphi_n(z) \, dz \)

which by Theorem 2

\[ r_n(x) = \int_{-\infty}^{\infty} k(x;z) \left( (1 - \theta) \varphi_{n-1}(z) + \theta f(z) \right) \]

or, \( r_n(x) \)

\[ = (1 - \theta) \int_{-\infty}^{\infty} k(x;z) \varphi_{n-1}(z) \, dz + \theta \int_{-\infty}^{\infty} k(x;z) f(z) \, dz \]

\[ = (1 - \theta) r_{n-1}(x) + \theta \int_{-\infty}^{\infty} k(x;z) f(z) \, dz \quad (3.5) \]

Using the method of difference equations,

\[ r_n(x) \]

\[ = (1 - \theta)^n r_0(x) + \theta \left( \int_{-\infty}^{\infty} k(x;z) f(z) \, dz \right) \frac{(1 - (1-\theta)^n)}{\theta} \]

\[ = (1 - \theta)^n r_0(x) + \int_{-\infty}^{\infty} k(x;z) f(z) \, dz \left( 1 - (1-\theta)^n \right) \quad (3.6) \]

and, \( \lim_{n \to \infty} r_n(x) = \int_{-\infty}^{\infty} k(x;z) f(z) \, dz \quad (3.7) \)

Now \( k(x;z) f(z) \) can be shown to form a bivariate normal distribution with common mean \( \mu, \) having variance \( \sigma^2 + \delta^2 \) and \( z \) having variance \( \delta^2. \) The marginal of a bivariate normal is also normal.\(^{11}\) The marginal of \( x \) would therefore be distributed \( n(\mu, \sigma^2 + \delta^2) \) where

\( \mu = E(y) \). The proof that \( k(x;z) f(z) \) forms a bivariate normal is in Appendix A. This completes the proof of Theorem 3.

**Theorem 4**

If: \( f_n(y) = f_n(w_n - w_{n-1} + x_{n-1} - m) \)

Then: \( r(x) \), the limiting form of \( r_n(x) \), does not exist. \(^{12}\)

**Proof** (By Contradiction)

Assume \( r(x) \) exists, then \( \lim_{n \to \infty} E(x_n) = E(x) = C \) (C a constant)

By Theorem 1

\[ r_{n+1}(x) = -\infty \int_{-\infty}^{\infty} k(x;z) g_{n+1}(z) \, dz \]

and by Theorem 2

\[ r_{n+1}(x) = (1 - \theta) \int_{-\infty}^{\infty} k(x;z) g_n(z) \, dz + \theta \int_{-\infty}^{\infty} k(x;z) f_n(z) \, dz \]

and \( E(x_{n+1}) \)

\[ = (1 - \theta) E(x_n) + \theta E(w_n - w_{n-1} + x_{n-1} - m) \]

\(^{12}\)Ibid., p. 187.
\[ E(x_{n+1}) = (1 - \theta)\mu_n + \theta (0 - 0 + \mu_{n-1} - m) \]

or, \[ \mu_{n+1} = (1 - \theta)\mu_n + \theta \mu_{n-1} - \theta m \]

This nonhomogeneous second order difference equation has the following general solution:

\[ \mu_n = c_1 + c_2 (-\theta)^n + (-\theta m / \theta + \theta) n \]

(3.8)

and \[ \lim_{n \to \infty} \mu_n \] clearly does not exist; a fact which contradicts the original assumption.

**Discussion of Model Predictions**

**Theorem 3 (Case I)**

Theorem 3 shows that the model predicts an asymptotic form of behavior for \( F \) in the form of a probability distribution with mean \( \mu \) and variance equal to the sum of the variances of the smearing distribution and the reinforcement random variable \( y \). This prediction is appealing for several reasons. First, the prediction is independent of \( \theta \) which is one of the two model parameters which must be estimated using experimental data. Second, the prediction is such that several tests of the model are quickly apparent. \( F \)'s responses may be compared with the predicted \( r(x) \). Also, Suppes and
Frankmann (1961) have pointed out that that normative
decision theory would encourage F to make response \( \mu \)
on each trial.\(^{13}\)

Certain assumptions are necessary to assume that
normative decision theory would encourage F to make bias
response \( \mu \). Recall \( y_n \) is a random variable whose
value on any trial informs F what amount of bias on
that trial would have resulted in a successful fore­
cast. In effect, F is successful if he errs in predicting
the value of \( y_n \) by no more than \((d - c)/2\) units. If F
is fully aware of the characteristics of \( y_n \) (independence,
normal shape with known mean) then he maximizes the
probability of success by predicting \( \mu \) on each trial.
The probability is a maximum because, for a unimodal,
symmetrical distribution like \( y_n \), maximum probability
for any interval occurs when the interval is centered
at the modal (also mean) value.

The act "predict \( \mu \)" will also have the greatest
expected value if F assigns more utility to success
than to failure. F's choice of acts affects only the
probability of success. Thus, the expected value of

---

\(^{13}\) P. Suppes and R. W. Frankmann, "Test of Stimulus
Sampling Theory for a Continuum of Responses with
Unimodal Noncontingent Determinate Reinforcement," Journal
act "predict \( \mu \)" is the maximum expected value because that act has the highest probability of success. In brief, normative decision theory would encourage F to make bias response \( \mu \) assuming F knows the characteristics of \( y_n \) and assuming F assigns more utility to success than to failure. F's responses may be examined for correspondence with choice of \( \mu \) units bias on each trial.

Finally, note that the model predicts that the behavior of R, as represented by \( f_n(y) \), affects the biasing behavior of F. \( E(f_n(y)) \) determines the mean of \( r(x) \), and the variance of \( r(x) \) varies directly with the variance of \( f_n(y) \). This variance, as stated several times before, affects the validity of F's forecast.

**Theorem 4** (Case II)

Theorem 4 shows a drastic change from Theorem 3 in the predictions of the model. Theorem 4 shows that under a Case II method of forecast revision, bias responses do not have a limiting probability distribution. For one thing, this means that the average or mean response does not approach any single response as a limit. In fact, as equation 3.7 shows, the mean response drifts constantly away from 0 towards minus infinity.
Secondly, the lack of a limiting form of $r_n(x)$ prevents the making of an exact prediction of the effect of Case II method of forecast revision on forecast validity.

For testing purposes, Theorem 4 provides no density function at asymptote against which to compare subject responses. Tests of the model will primarily be limited to testing correspondence with predicted mean responses. The tests of both this model and the Case I model are in the next chapter.

**Chapter Summary**

In this chapter, a model of F's biasing behavior has been presented. This model can be linked to the validity of F's forecast by making the additional assumption that F determines his bias figure independently of his unbiased estimate and then sums the two to form his reported forecast. In Chapter I, a relationship between the validity of F's forecast and the variance of a bias random variable was developed. For a Case I type of method of forecast revision, forecast validity and F's biasing behavior can be linked very nicely. The model for Case I predicts a bias random variable which does have a limiting form and therefore a
predicted variance. However, for Case II, no limiting form is predicted for the bias random variable. Under this condition, it seems reasonable to argue that, considered alone, the rate of drift in the mean of the bias random variable will vary inversely with the validity of F's forecast.

The next chapter describes an effort to test whether or not method of forecast revision does affect forecast validity. Part of this effort is devoted to testing the correspondence of the data with the predictions of models described in this chapter.
CHAPTER IV
AN EXPERIMENT AND RELATED DATA ANALYSIS

Chapter Purpose

An experiment was designed to test several hypotheses concerning the effect of R's method of forecast revision on the validity of F's forecast. Also, the data from this experiment will be used to examine several of the assumptions and predictions of the two models.

The hypotheses are as follows:

1. The method of forecast revision chosen by R will not affect the validity of F's forecast.

2. Given F begins each forecasting task having an unbiased estimate of the actual, the validity of F's forecast will not change as F gains experience with the task.

Assumptions and predictions of the models that will be examined are listed below. For convenience, they will be labeled hypotheses 3, 4, 5 and 6.

3. F's bias response on trial n is independent of the unbiased or "good" estimate on trial n.

4. \( \theta \), the learning parameter, remains constant between the first half and the second half of the data.

5. The mathematical expectation of equation 3.6 yields a predicted mean response curve for Case I. Assuming the reinforce-
ment random variable has mean 20, the following difference equation describes the mean response curve for Case I

\[ \mu_n = (1 - \theta)^n \mu_0 - 20(1 - (1 - \theta)^n) \]

where \( \mu_0 \) is the mean response at time zero. Also, the Case I model predicts a limiting response density function \( r(x) \) with mean 20 and variance equal to the sum of the variance of the reinforcement random variable and the variance of the smearing distribution.

6. The mathematical expectation of equation 3.8 yields a predicted mean response curve for Case II. The following difference equation describes the mean response curve for Case II

\[ \mu_n = C_1 + C_2 (-\theta)^n + (-\theta m/1-\theta)^n \]

The tests of these hypotheses, assumptions and predictions will proceed in two steps. First, an analysis of variance will be performed on validity measures derived from experimental data. This will specifically test hypotheses 1 and 2.

A second series of tests will consist of examining the correspondence of the experimental data with the two models of biasing behavior. Correspondence will be examined by testing hypotheses 3 through 6.
Experiment

Task

The nature of the task will be briefly stated here. Specific procedural matters will be discussed in following sections.

In the experiment, F, a subordinate of R, is required to submit a forecast of unit sales to R each period. Because the forecast is used to evaluate F's selling ability, F is asked to consider a successful forecast to be one somewhat lower than the actual sales figure. More specifically, F considers a forecast to be a success if actual sales exceed the revised forecast by six to ten units.

F will be operating under the two methods of forecast revision discussed in general as Case I (Theorem 3) and Case II (Theorem 4). Under Case I the figure added to F's forecast by R to arrive at the revised forecast is a random variable which is normally distributed with mean 12 and variance 36. Under Case II the figure added to F's forecast by R to arrive at the revised forecast is F's forecast error on the previous trial. Forecast error is defined to be the excess of actual over F's original forecast.
Subjects

Eighteen undergraduate accounting majors served as subjects for this experiment. Fifteen of the subjects were members of the accounting honorary Beta Alpha Psi. The other three subjects were volunteers from upper level accounting courses. All subjects agreed to serve without pay.

Three subjects were omitted from the analysis either because of apparatus failure (2) or sizeable subject error. The sizeable error made by the one subject resulted from having based almost all forecasting and biasing decisions on the wrong information. In short, the subject misread information supplied by the experimenter.

Attempts were made to reschedule the two subjects omitted because of apparatus failure. The experiment was run the week before the research was to be discussed in a Beta Alpha Psi meeting. The subjects, who were Beta Alpha Psi members, could not be rescheduled before this meeting.

Experimental Procedure

The experiment took approximately one hour. The sequence of events making up the experiment will be described in chronological order.
First, F met with the Experimenter (E) for instructions. F was given an instruction sheet to read. This instruction sheet is part of Appendix B. The purposes of the instruction sheet were

1. to explain the sequence of events which would occur once the experiment began.

2. to ask the subject to assume the role of the subordinate F who is required to send a forecast to his superior H under conditions where F desires a downwardly biased result.

3. to ask F to respond within a certain time period (10 seconds). The 10 second limit was not strictly enforced by E. The purpose was to insure that F would not consistently take 40 or 50 seconds per trial, thus failing to complete the experiment within one hour.

Before making his response on each trial, F was given what was labeled a "good" estimate of sales. The "good" estimate was the mean of a normal distribution with small variance from which the actual sales figure was randomly sampled. During the instruction period, E attempted to demonstrate the meaning of the "good" estimate by showing F a list of "good" estimates for 15 time periods and then one at a time, turning over cards which showed the actual for each of the time periods. The purpose here was to show S that the actuals
fell symmetrically about the "good" estimate and that the variance about the "good" estimate was small.

**Pretest**

Before actually running the experiment, a pretest of the experiment was performed. The pretest resulted in the instructions being rewritten and the task being made somewhat easier for F to perform.

**Actual Experiment**

Each trial (or period) began by F's being asked "What is your forecast?" To aid him in giving a forecast, F was given the "good" estimate of sales for that period. In the experiment any deviation of the forecast from the "good" estimate was considered intentional bias on the part of F. Subjects were asked in a questionnaire administered after the experiment if they felt the actuals were generally less than, greater than or neither greater than nor less than the good estimates. Eight responded "neither greater than nor less than," three responded "greater than" and four responded "less than."

When the responses are broken down by case, one difference is apparent. The Case I group's responses
were almost equally distributed among the three possible responses while all but one of the Case II group's responses were in the "neither greater than nor less than" category. Rankings of average bias responses (forecast minus "good" estimate) over the last twenty trials of subjects in Case I support their stated attitude toward the "good" estimate. In other words, most subjects who responded "actuals were generally less than" ranked high on average bias responses and most subjects who responded "actuals were generally greater than" ranked low on average responses.

This breakdown by case illustrates one possible difference between groups operating under different methods of forecast revision and also weakens the argument for the stated measure of bias. However, since subjects were not asked their impression of the average difference between actual and "good" estimate, the stated measure of bias must still be used for purposes of analysis.

After giving his forecast for the period, F received the following information:

1. The revised forecast (F's forecast revised by R)
2. The actual sales figure for the period

3. The figure "actual sales minus revised forecast of sales."

Then, F was asked to record whether or not he was successful at forecasting. This was done by checking the appropriate column on the same sheet of paper that informed F of the period's "good" estimate. F was asked to check the appropriate column on each trial to insure that the criteria for success would be remembered.

After completion of the 70 trials, F was asked to fill out a short questionnaire. Finally, after all subjects were run, the proposal for this research was presented to an audience containing most of the individuals who had been subjects in the experiment.

Generation of Data

The data needed for the experiment were (1) the "good" estimate for each of the 70 trials, (2) the actual sales for each of the 70 trials, and (3) the figure added to F's forecast under Case I to arrive at R's revision of F's forecast.

The "good" estimate and the actual sales were

---

Appendix B contains the instruction sheet, the sheet informing F of each period's "good" estimate and the short questionnaire administered to F.
initially conceived as two random variables having a bivariate normal distribution with a high degree of correlation and a regression equation with intercept zero and slope coefficient plus one. The choice of a bivariate normal distribution was not selected because it necessarily corresponds with actual situations in business settings. It was chosen because data generation was made more simple and because the bivariate normal distribution provided data which satisfied the assumptions of regression analysis.

One way of generating the data would have been to sample pairs of values from a bivariate normal distribution having the desired characteristics. Actually, the data for "good" estimates and actual sales were generated by first sampling "good" estimates from a normal distribution having mean 450 and variance 100. Then, the actual sales figure was arrived at by adding to each sampled "good" estimate a number randomly sampled from a normal distribution having mean 0 and variance 9. This procedure did result in regression equation coefficients whose 95% confidence intervals contained the predicted values.
Finally, as mentioned before, the figure added to F's forecast to obtain R's revised forecast was a random sample from a normal distribution having mean 12 and variance 36.

**Apparatus**

The apparatus used in the experiment was a typewriter-console which is a part of the General Electric time sharing system. This console is in the data center of the College of Administrative Sciences at The Ohio State University. The computer was programmed to play the role of R. No attempt was made to fool F into thinking that R was an actual person. F was told only that he was to send his forecast to his superior by typing his forecast on the typewriter.

The computer did the following:

1. Had the console ask F, "What is your forecast?"

2. Received the forecast and revised it in the predetermined manner.

3. Had the console type out
   a) the revised forecast
   b) the actual sales
   c) actual sales minus revised forecast of sales.
Data Analysis - ANOVA Design

The first test of the data will consist of a two-way analysis of variance on each subject's validity measure. Each subject's 70 responses are broken down into five blocks of 14 trials. These five blocks for each subject form a time factor while the method of forecast revision is the second factor in the two-way ANOVA. The experimental design along with the raw data is shown in Table 4.

Errors occasionally occurred in the subject's data due partly to the subject's lack of familiarity with the electric typewriter. When the experimenter felt that an error had been made and that the nature of the error was obvious, the error was corrected. Corrections were made several weeks after the experiment, so subjects were not consulted as to whether or not an error of a particular type had been made. The average number of corrected errors per subject was .47, with the range being 0 to 2.

After obvious errors in F's responses were corrected, the correlation between F's original forecast and actual sales for each block of 14 trials was computed. The correlation coefficients were transformed
<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
<th>$B_4$</th>
<th>$B_5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>0.9990</td>
<td>0.9835</td>
<td>0.9936</td>
<td>0.9796</td>
<td>0.8326</td>
</tr>
<tr>
<td>$S_2$</td>
<td>0.9982</td>
<td>0.9846</td>
<td>0.9952</td>
<td>0.9987</td>
<td>0.9510</td>
</tr>
<tr>
<td>$S_3$</td>
<td>0.9983</td>
<td>0.9840</td>
<td>0.9980</td>
<td>0.9980</td>
<td>0.9985</td>
</tr>
<tr>
<td>$S_4$</td>
<td>0.9646</td>
<td>0.9850</td>
<td>0.9954</td>
<td>0.9985</td>
<td>0.9988</td>
</tr>
<tr>
<td>$S_5$</td>
<td>0.9986</td>
<td>0.9827</td>
<td>0.9963</td>
<td>0.9674</td>
<td>0.9471</td>
</tr>
<tr>
<td>$S_6$</td>
<td>0.9849</td>
<td>0.9826</td>
<td>0.9990</td>
<td>0.9982</td>
<td>0.9987</td>
</tr>
<tr>
<td>$S_7$</td>
<td>0.9995</td>
<td>0.9831</td>
<td>0.9977</td>
<td>0.9997</td>
<td>0.9993</td>
</tr>
<tr>
<td>$S_8$</td>
<td>0.9994</td>
<td>0.9849</td>
<td>0.9989</td>
<td>0.9992</td>
<td>0.9952</td>
</tr>
<tr>
<td>$S_9$</td>
<td>0.9974</td>
<td>0.9824</td>
<td>0.9897</td>
<td>0.9982</td>
<td>0.9953</td>
</tr>
<tr>
<td>$S_{10}$</td>
<td>0.9992</td>
<td>0.8608</td>
<td>0.9940</td>
<td>0.9972</td>
<td>0.9927</td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>0.9990</td>
<td>0.9828</td>
<td>0.8894</td>
<td>0.9952</td>
<td>0.9997</td>
</tr>
<tr>
<td>$S_{12}$</td>
<td>0.9895</td>
<td>0.9855</td>
<td>0.9340</td>
<td>0.8934</td>
<td>0.6491</td>
</tr>
<tr>
<td>$S_{13}$</td>
<td>0.9980</td>
<td>0.9820</td>
<td>0.9882</td>
<td>0.9984</td>
<td>0.9978</td>
</tr>
<tr>
<td>$S_{14}$</td>
<td>0.9601</td>
<td>0.9810</td>
<td>0.7229</td>
<td>0.9347</td>
<td>0.6521</td>
</tr>
<tr>
<td>$S_{15}$</td>
<td>0.9992</td>
<td>0.9839</td>
<td>0.9980</td>
<td>0.9551</td>
<td>0.8995</td>
</tr>
</tbody>
</table>
to Fisher z scores. The parametric analysis of variance was conducted using these z scores.

The correlation coefficient was used as the measure of validity because of its close ties to the prediction equations frequently developed from regression analysis. The correlation coefficient can be thought of as a measure which tells R how much variability in actual sales can be removed for prediction purposes by using F's forecast. However, if tests indicated violations of regression analysis assumptions such as autocorrelation, R might get an additional reduction of variability by going through standard procedures such as searching for "missing variables." In these situations, the correlation coefficient would not be a sufficient measure of the amount of information present in F's forecast. To see how common violations might be, a 20% random sample of each Case was taken, and the regression equation for each sampled item was determined with actual sales being the dependent variables. The

2The Fisher r to z transformation is common when dealing with correlation coefficients. The reasoning behind this transformation can be found in W. L. Hays, Statistics for Psychologists (New York: Holt, Rhinehart and Winston, 1963), pp. 530-33.

one sample runs test was performed on the residuals as a test of autocorrelation. None of the 15 results of use of the runs test was significant at the .05 level. Lack of any significant result provides additional confidence in the validity measure.

The decision to break the 70 trials into blocks was motivated by a desire to examine the effects of experience in early trials. Also, it did not seem likely that a validity measure based on 70 trials would be practical in a forecasting situation. Assuming monthly forecasts, 70 periods would be an experience of almost six years.

The choice of five blocks was arbitrary. The 70 trials could just have easily been divided into seven blocks of ten trials. The point is important because the choice of blocks could affect the results of the tests of significance.

The correlation coefficients (validity scores) which make up the data indicate how much variability in the sample of the dependent variable can be explained by a least squares line on the sample of the independent variable. This least squares line has two free

---

parameters which are estimated from the data: the intercept term and the slope coefficient. If the 70 trials were divided into 35 blocks of two trials each, all correlation coefficients would probably be +1.0 (in this case) and no significant effect actually present in the data would be found. There is a direct relationship between the number of blocks of trials and the average correlation coefficient. Increasing the number of blocks increases the number of free parameters which must be estimated from the data (an intercept term and a slope coefficient for each block.) Generally, the more free parameters contained in the model (the model implicit with using the correlation coefficient,) the better the model fits the data.

A second factor of importance is that Case II of the method of forecast revision factor in the ANOVA corresponds with the prediction of a bias response distribution \( r(x) \) whose mean drifts over time. This drift will affect the validity measure. Increasing the number of trials under consideration increases the amount of predicted drift. This increase of drift considered alone would indicate that a lower degree of forecast validity is likely. Decreasing the number of
trials considered in measuring validity (and developing a prediction equation) might actually increase R's ability to predict the next actual, given F's next response. Also, an important implication for the ANOVA test is that decreasing the number of trials makes the least squares line more sensitive to changes in $E(r_n(x))$ and decreases the influence of drift on the validity measure.

In summary, changing the number of blocks can, in extreme cases, affect the results of the test solely because of the change in the number of free parameters available to fit to the data. An extreme example is the case of 35 blocks where all validity measures would probably by ±1.0, thus making the discovery of a significant effect impossible.

Secondly, changing the number of blocks changes the sensitivity to drift of $E(r_n(x))$ of the least squares line underlying the validity measures. On an a priori basis, decreasing the number of blocks should increase the likelihood of finding a significant effect due to method of forecast revision.
Assumptions

Three assumptions can be tested for this repeated measures ANOVA design. Two of the assumptions are homogeneity of variance assumptions. The F_max test with a significance level of .20 will be used to test these two assumptions. The two homogeneity of variance tests are for the subjects within group and blocks by subjects within groups poolings of variance estimates.

A significance level of .20 is larger than the conventional levels usually seen in research using statistical tools such as the analysis of variance. For this reason some justification of the choice of .20 is needed.

One argument for a small significance level such as .05 is that there is a large penalty associated with Type I error and a relatively small penalty associated with Type II error.\(^5\) For purposes of this research, in testing the assumptions of the model, just the opposite is the case. The penalty associated with Type II error is judged by the experimenter to be considerably larger than the penalty associated with Type I error.

\(^5\)Type I error is rejecting the null hypothesis when, in fact, the null hypothesis is true. Type II error is failing to reject the null hypothesis when, in fact, the alternative hypothesis is true.
The penalty associated with Type I error is the loss in power resulting from switching to a nonparametric test. The loss in power does not appear to be large, especially considering the small number of subjects used in this experiment.  

The penalty associated with Type II error is the use of a model when the assumptions of the model are violated. Robustness is usually put forth as an argument that the penalty associated with using the model is small for moderate violations of the assumptions. This argument cannot be well supported. In general, the case for robustness appears to have been overstated. And, in this particular case, it is generally agreed that the presence of unequal n weakens any argument that can be made for robustness. The result is, in the case of Type II error, the user of the model cannot be confident that the actual distribution that should be used to test his hypotheses differs only slightly from the F distribution that he is using in an analysis of variance.

---


7Ibid., pp. 24-25.
For these reasons, the significance level was set at the relatively large value of .20.

The third assumption is the homogeneity of covariance assumption. Actually, this test will be performed only if a conservative F test of main effects indicates no significant results and the usual F test indicates significance. The rationale behind this procedure is described in Winer. 8

**Results of the ANOVA Test**

The subjects within groups homogeneity test was significant at the .20 level. Because of this failure of one of the model assumptions, a nonparametric substitute suggested by Bradley was performed to test the group, block and group by block interaction effects. Bradley suggests using several tests to test the different main effects and the interaction effect. 9 The Wilcoxin Rank--Sum Test was used to test the group effect, and the Friedman Multi-Sample Test was used to test the block effect and the group by block interaction effect.

---


The group effect was significant at the .05 level ($W=40; n=7, m=8$). This is interpreted to mean that the method used by R (a computer) to revise F's forecast did affect the validity of F's forecast. Examination of the data shows that the higher validity scores, on the average, were in the group operating under the Case I method of forecast revision.

The time or block effect was also significant at the .05 level ($S=278; R=8, C=5$). Examination of the data indicates that the first and fifth periods made by far the biggest contributions to the statistic used to test for significance. Generally, the first time period had the highest validity score and the fifth period, the lowest validity scores. The middle three periods could not be ranked in any particular order. In other words, the biggest changes occurred between the first and second and the fourth and fifth periods. As will be seen shortly, the models predict a significant amount of change in mean response behavior for both groups only in the first period. After the first period, the Case I group is predicted to be approximately at asymptote. This provides some understanding of the change between the first and second periods.
However, the models provide no explanation for the change after the fourth period.

Finally, the interaction effect was not significant at the .05 level (S=90; R=7, C=5). This supports the view that the size of the group effect does not depend upon the particular time period considered, or that the size of the time effect does not depend upon the specific group considered.

**Data Analysis - Correspondence with Models**

**Hypothesis 3**

One assumption used in relating the model to the biasing situation is that the amount of bias added to the "good" estimate to form F's forecast is determined by F independently of the "good" estimate. The correlation between bias responses and "good" estimates for subjects operating under Case I was -.1872. A t test was used to test the null hypothesis that an r of -.1872 was not significantly different from zero. The null hypothesis was rejected at the .01 level indicating that the "good" estimate did have some effect on F's choice of bias in Case I. This effect, though significant,

\[
\text{The formula used to calculate the t statistic is: } t = \frac{(r \sqrt{n-2})}{\sqrt{1-r^2}}, \quad n \text{ for Case I was 556.}
\]
is relatively small. Knowledge of the "good" estimate reduces the variability of bias responses $100r^2\%$ or less than $4\%$.

The correlation between bias responses and the "good" estimate for Case II was .008, which is not significant at even the .20 level ($n=484$). A partial explanation for the difference between cases might be that the predicted trend factor (see the next section) would tend to disguise the effect of the "good" estimate on the amount of bias incorporated into a forecast. In any case, the significant effect for Case I is sufficient to cast some degree of doubt on one assumption of the models.

Hypotheses 4, 5 and 6

Both models yield a prediction of how mean responses vary over time. These predictions are dependent upon the value chosen for $m$, where $m$ is the single value that is representative of the criteria for success interval $+6$ to $+10$. In Chapter III it was stated that $m$ would be assumed equal to 8. However, in testing asymptotic predictions, it was found that an $m$ of +6 provides a better fit to the data. For this reason, an $m$ of +6 is also used to examine mean responses.
For Case I,

\[ r_n(x) = (1 - \theta) r_{n-1}(x) + \theta - \int_{-\infty}^{\infty} k(x; z) f(z) \, dz \]

where \(-\int_{-\infty}^{\infty} k(x; z) f(z) \, dz\) is a normal distribution with mean \(-18\). Taking expectations results in the following

\[ \mu_n = (1 - \theta) \mu_{n-1} + \theta (-18) \]

which can be represented by a difference equation of the form

\[ \mu_n = (1 - \theta)^n \mu_0 - 18 (1 - (1-\theta)^n) \]

or \[ \mu_n = (18 + \mu_0)^n - 18 \]

The method of least squares was selected for purposes of estimating \( \mu_0 \) and \( \theta \). For the one-element stimulus sampling model with a discrete number of responses and noncontingent reinforcement, there is a maximum likelihood estimate of \( \theta \). This method might be used to estimate \( \theta \) for the one-element model with a response continuum by arbitrarily dividing the continuum into a discrete number of intervals. However, the maximum likelihood method cannot be used for the Case II model. To aid comparison of models, it is desirable to

---

11The method is described in Suppes and Atkinson, op. cit., Chapter 2.
use the same method for both models. For this reason, the method of maximum likelihood was rejected. Alternative methods are to minimize least squares, to minimize a weighted least squares (minimum chi-square method), and to minimize absolute deviations. One criterion for choosing among methods would be to examine the experimenter's attitudes toward prediction errors. Actually this criterion was not used because no strong opinion was held by the experimenter concerning the weight to assign errors of various magnitude. The method of least squares was chosen primarily for its close ties with regression analysis and the associated tests of the assumptions of the regression model. These ties provided help in the analysis of mean responses, especially for Case II.

A computer program was written to estimate $\theta$ by successively trying various values of $\theta$ and $\mu_o$ until those values of $\theta$ and $\mu_o$ were found which provided a minimum sum of squares. All the data for Case I were used to determine mean responses per trial. The estimate of $\theta$ was .037 and the estimate of $\mu_o$ was -5. Use of the model with $\theta = .037$ and $\mu_o = -5$ reduced the sum of squares figure from 1061.05 to 448.95, a reduction of 57.68%.
Similarly, the Case II model predicts that the following difference equation (Equation 3.8, Chapter III) describes how mean responses will vary over time.

\[ \mu_n = C_1 + C_2 (-\theta)^n + \left( \frac{-0.6}{1+\theta} \right) n \]

For reasonable values of \( \theta \) (\( \theta < 0.25 \)), the term \( C_2 (-\theta)^n \) rapidly approaches zero. In fact, \((-0.25)^n\) rounded to three digits is zero for \( n \) greater than four. For this reason, the \( C_2(-\theta)^n \) term was dropped for purposes of estimating \( \theta \). The form of the equation now predicts a linear relationship between mean responses and time. The equation is

\[ \mu_n = C_1 + \left( \frac{-0.6}{1+\theta} \right) n \]

A least squares line was fit to the mean responses for Case II. The value estimated for \( \theta \) was 0.105. \( C_1 \) was estimated to be -2.06. Use of the model for Case II reduced the sum of squares from 10,943.3 to 1,681.6, a reduction of 84.65%.

Figure 2 shows the data plotted along with predicted values for both models for the first 35 trials. Figure 3, a continuation of Figure 2, shows the data plotted along with predicted values for the second 35 trials.
FIGURE 2: PREDICTED VS. ACTUAL MEAN RESPONSES FOR THE FIRST 35 TRIALS
FIGURE 3: PREDICTED VS. ACTUAL MEAN RESPONSES FOR THE SECOND 35 TRIALS
Both models provided substantial reductions in the sum of squared deviation. However, observation of Figures 2 and 3 seem to indicate a pattern in the residual differences, especially for Case II. Such patterns are frequently associated with the presence of autocorrelation in regression analysis. However, in this case it might be partly due to nonconstant \( \theta \). This observation led to a test of Hypothesis 4 and a test for other unexplained patterns in the data.

The stimulus sampling model assumes that \( \theta \) is constant. This parameter represents the probability that the subject's single stimulus element at time \( n \) will become conditioned to the correct response at time \( n \). The parameter is often thought of as a measure of rate of learning.

To test Hypothesis 4, \( \theta \) was estimated for the first half and the second half of each subject's 70 responses. Individual responses were considered estimates of mean responses. The predicted mean response difference equations and the least squares criteria were used to provide the 16 estimates of \( \theta \) for Case I and the 14 estimates for Case II. The sign test for
related samples was used to test the null hypothesis that \( \theta \) was constant.\(^{12}\)

For Case I, the exact probability of the result of the sign test, given the null hypothesis, was \( .637 \). For Case II the exact probability, given the null hypothesis, was \( .062 \), with the larger average \( \theta \) occurring for the second half of the data. As a result of this test, the assumption of constant \( \theta \) for Case I is not rejected while the assumption is placed in serious doubt for Case II. Examining Figures 2 and 3, it is plain that a smaller \( \theta \) would provide a better fit for the first half of the data in Case II, and a larger \( \theta \) would provide a better fit for the second half of the data. The value of \( .105 \) estimated for \( \theta \) for purposes of specifying the learning curve is now best considered an average value.

Part of the pattern (or apparent nonrandomness) in the residuals of Case II can be explained by assuming a nonconstant \( \theta \). However, since the manner in which \( \theta \) changes is not known, it is impossible to remove the variability due to variations in \( \theta \) and then check for additional nonrandomness in the residuals.

On the other hand, the test for Case I does not allow a rejection of the null hypothesis of constant \( \theta \); therefore, a test for nonrandomness not due to \( \theta \) can be performed. The one-sample runs test was performed on the residuals of Case I. The exact probability of the observed number of runs was .3669. This probability is not small enough to indicate a strong unexplained pattern in the data.

**Asymptotic Predictions**

The model for Case I also predicts a limiting form of behavior. This prediction is in the form of a normal probability distribution with its mean dependent on \( m \) and variance equal to the sum of the variance of the reinforcement random variable \( y \) and the variance of the smearing distribution \( k(x;z) \). The mean of \( y \) can be found by taking the expectation of equation 3.3. Thus, \( E(y) = E(w_n - u_n - m) = (0 - l2 - m) \). Originally, \( m \) was assumed to be 8, the midpoint of the range which defines F's criterion for success. However, the data in this section will provide some support for assuming \( m \) equals 6.
The variance of the reinforcement distribution is \( \text{Var}(y) = \text{Var}(w_n - u_n - m) = 9 + 36 = 45 \). The variance of the smearing distribution must be estimated from the data.

The data used must be those responses made at asymptote. The procedure used to determine what portion of the 70 trials to use is as follows. First, the predicted responses for Case I were examined. Figure 3 shows that predicted responses approach -18 very slowly. In fact, predicted responses do not appear to reach asymptote until late in the sequence of 70 trials. However, a plot of both individual and average responses over the last 40 trials does not indicate the presence of any trend. A nonparametric test of trend supports this indication. A null hypothesis of no trend could not be rejected at even the .2 level of significance.  

\[\text{Therefore, the last 40 responses for each subject were used to examine the correspondence of the data with asymptotic predictions.}\]

An obvious approach to testing correspondence with \( r(x) \) is to divide the domain of \( r(x) \) into some

\[\text{13The test used was Daniel's Test of Trend. The procedure is described in J. V. Bradley, op. cit., p. 96.}\]
number of intervals, each having an expected frequency of responses greater than 5, and then to use a chi-square statistic to test for goodness-of-fit. The degrees of freedom would be adjusted for the fact that the variance of $r(x)$ would have to be estimated from the data.

However, one assumption of the chi-square test is that the data must be from a random sample. The model used predicts a correspondence between adjacent responses by $F$. Therefore, every response in the set of 40 cannot be used. The serial correlation between every other and every third response was calculated. The serial correlation for every other trial was .266 ($N=312$) and for every third response the serial correlation was .256 ($N=312$). Both are easily significant at the .05 level.

It is apparent that we do not have independence even if every third trial is used to test goodness-of-fit. Secondly, the measure of serial correlation is not diminishing at a rapid rate. It appears, therefore, that a test of goodness-of-fit using an inferential statistical technique is not justifiable. The approach to be used here will be to take every other response
of each subject and to plot this data as a bar chart superimposed on a graph of the predicted distribution. The chi-square statistic will be calculated and used to compare the predicted r(x) with an alternative form of r(x). The chi-square statistic is a better model with no reference being made to significant fit or lack of significant fit.

The variance of the smearing distribution was estimated by calculating the sample variance of subject responses, making the necessary adjustment to obtain an unbiased parameter estimate, and subtracting the variance of the reinforcement random variable y. The sample consisted of every tenth response of the eight subjects making up Group I. The variance of k(x;z) was estimated to be 51.98 - 45 = 6.98. Figure 4 shows the bar chart representing actual responses superimposed on the predicted r(x), given r(x) has variance 51.98 and mean -18. The chi-square statistic for goodness-of-fit with 13 degrees of freedom is 82.71.

An examination of Figure 4 supports the view that an m of 6 provides a better fit than an m of 8. Recall, the criterion for success at forecasting was actual sales six to ten units greater than the revised forecast.
FIGURE 4: TWO PREDICTED LIMITING DISTRIBUTIONS VS. THE ACTUAL DISTRIBUTION OF RESPONSES FOR CASE I
The value 8 was originally chosen as being the one figure representative of this range. However, the alternative value of m equal 6 was used to test asymptotic predictions. This choice of m could be considered a reasonable choice since it is the smallest amount of bias which still leaves F successful at forecasting. The chi-square statistic for goodness-of-fit for \( r(x) \) with a mean of \((-12 - 8) \) or \(-20 \) is \( 87.49 \). Thus, a better fit to the data is provided by a choice of 6 for the value of m.

Finally, in Chapter III the assumptions necessary for normative decision theory to recommend a bias component of 20 (or 18 if m=6) were discussed. Examination of Figure 4 shows that the group of subjects did not come close to always responding -20, -18 or any other single value. Examination of the responses of individual subjects supports the same conclusion. In the experiment, subjects were not told what assumptions they could make about the reinforcement random variable. For this reason, an examination of the data cannot conclusively show whether subjects did or did not reject the "maximize expected utility" decision rule of normative decision theory.
Conclusions

The question originally asked in the first chapter concerned whether or not the method of forecast revision chosen by a recipient would affect the validity of a forecaster’s forecast. The question was supported by an a priori argument in the form of a model that presented an hypothesized link between method of forecast revision and forecast validity. This model is presented in the latter part of Chapter I and in Chapter III. Several conclusions about the question can now be made based on the experimental work described in this chapter.

First, it does appear that the validity of forecasts made by students in a laboratory setting is affected by the method of forecast revision used by a computer to prepare revised forecasts. Furthermore, forecast validity did vary over time but the effect of method of forecast revision on forecast validity did not depend on the time period. In other words, there does not seem to be an interaction between the method of forecast revision effect and the time effect. The tests of Hypotheses 1 and 2 are the bases for these conclusions.
Secondly, the tests of Hypotheses 5 and 6 indicate that a large portion of the variability in mean biasing behavior can be explained by using the mean response curves derived from equations 3.6 and 3.8 and the values of $\theta$ estimated from the data. A large portion of the variability was explained even though the data failed to support at least one assumption of each model. The test of Hypothesis 3 indicates that subjects in Case I did not treat the bias factor and the "good" estimate as being independent. The violation in this instance seems to be minor in importance and does not cast serious doubt on the model for Case I.

Tests of Hypothesis 4 allowed a rejection of the assumption of a constant $\theta$ for the Case II model. The assumption was not rejected for the Case I model. The violation of Hypothesis 4 for the Case II model is a more serious violation than the violation of Hypothesis 3 just discussed. In fact, it may indicate that the model used to describe biasing behavior may have to be substantially changed when reinforcement is dependent on responses made on previous trials. The questionnaire administered to subjects after the experiment tried to identify possible ways of modifying the model. Each
subject was asked if he had a particular strategy for determining his bias response, and if so, he was asked to identify it. Several subjects operating under Case I said they frequently responded in the -15 to -20 range. This response supports the model for Case I. However, subjects operating under Case II provided no clues as to how the model might be modified. This was a little disappointing because the results from testing Hypothesis 4 seem to indicate that the Case II model needs modification.

A third set of conclusions concerns the asymptotic predictions discussed in Hypothesis 5. The chi-square statistics calculated in testing the asymptotic predictions are probably a little large. The major cause of their size is the number of responses in the left tail of the distribution. Actually, part of the distribution's skewness to the left may have resulted from not collecting quite enough data to insure subjects' responses were at asymptote. In any case, the test of asymptotic predictions does support using $m$ equal to 6 rather than $m$ equal to 8.

Finally, no attempt has been made in this concluding section to extend the results of this study to
the situation of a forecaster operating in an actual business environment. Being a laboratory study substantially limits the conclusions one can make about actual situations. Several of the limitations of this study which do act to restrict conclusions are discussed in the next chapter.
CHAPTER V
SUMMARY AND POSSIBLE EXTENSIONS

Introduction

Conclusions that can be drawn from the experimental and modeling work have been stated in the previous chapter. These conclusions must be limited in their generality due to the laboratory nature of the experiment. Any laboratory study has limitations which restrict the generality of conclusions. Some of the more important limitations are discussed in the next section of this chapter. That section will then be followed by a section on possible extensions of the research.

Limitations

Specific characteristics of this laboratory study which hinder the extension of conclusions to real situations include (1) the use of students as subjects, (2) the decision not to use paid subjects and (3) the use of a computer to simulate recipient behavior.

Students as Subjects

Several authors, including Alpert and Birnberg and Nath, have expressed concern with the use of students
as surrogates for businessmen in laboratory experiments. Alpert performed an experiment designed to test for differences between the two subject populations. Alpert did appear to find some differences between businessmen and students and between two different categories of students. However, his results are limited in value because subjects performed only a single task. Birnberg and Nath point out that some experimenters, primarily in the area of psychology, can rely on the theories being tested to point out relevant individual differences characteristics. Then if the two groups appear to differ in terms of these characteristics, it would be beneficial to use that sub-population about which the experimenter wants to make conclusions. However, the stimulus sampling theory associated with the learning models used in this research focus on no individual differences characteristics. A review of the axioms referred to in Chapter III will support this statement.


Weick, in discussing the use of students as subjects, expressed the following opinion:

Unless the experimenter is explicit about personal variables that moderate his phenomenon, and unless these variables assume quite different values in his subject population and his referent population, the use of a college student is not a serious limitation.⁴

It seems best still to view the use of students as a limitation, but one cannot easily evaluate the magnitude of the limitation because the theory used as a basis for describing forecasting and biasing behavior does not focus on any individual differences characteristics.

Subjects without Pay

In planning the experiment recognition was given to the fact that the conditions which brought the subjects to the experiment might affect subject behavior.⁵ Primarily, it was considered most desirable that the subjects approach the task with a willingness to strive for success. Using students required to serve as subjects seemed likely to reduce this willingness. Asking students to serve as volunteers seemed to be the best


⁵Ibid.
alternative. Whether the students volunteered because of interest in the experiment and as a favor to the experimenter or because of interest in the experiment and the offer of small payments related to the degree of success was considered less important than avoiding required participation.

Observation of subject behavior during the experiment and subsequent examination of the data seem to indicate that the subjects did strive for success throughout the experiment. Whether or not pay would have substantially increased this willingness or have affected subject behavior in some other manner is unknown. For this reason, the particular reward structure (no pay) used in this experiment is still considered a possible limitation of the study.

Computer as Recipient

A computer was used to simulate recipient behavior in this experiment partly for control reasons and partly for the advantages provided by the speed and accuracy of the computer.

The computer acted as a control by (1) always asking each subject for his forecast on each trial in exactly the same manner and (2) by taking the same
amount of time in both Case I and Case II to inform F of his revised forecast after F reported his forecast.

Secondly, the computer was advantageous because it allowed quick, accurate replies to F's forecasts, thus allowing a sizeable amount of data to be collected in a relatively short period of time. The printout of the computer also provided an accurate record of all responses on the part of F and the recipient as an aid to data analysis.

The disadvantage of the computer is that F's behavior may be different when he knows that he is dealing with a computer rather than with an actual person. One possibility might be that F would be more confident that there is a pattern to the behavior of the computer. He might then approach the forecasting task in a different manner.

For this experiment the advantages of the computer seemed to outweigh the disadvantages. However, identification of this limitation provides a likely area for future extensions of the research.
Summary of Limitations

The limitations just discussed, along with other limitations common to most laboratory studies, limit the generality of the results of this research. The major contribution of this research probably is that it points out an approach to a relevant and interesting accounting problem, and that it provides some justification for extending the research into related areas, including the actual forecasting situation. Possible extensions are discussed in the following section.

Extensions

The results of this study identify a sizeable number of possible research topics for future development. One topic might be a replication of the Cyert, March and Starbuck study discussed in Chapter II with significant changes being made in the quality of feedback provided line estimators. The major advantage of such a change is that it might allow the use of modeling techniques such as those used in this research to provide predictions of behavior.

A second area for research would be to continue the laboratory approach but either try to remove some of the possible limitations of the present study or to
extend the present work. For example, with the proper physical facilities, an individual working for the experimenter might be substituted for the computer as the forecast recipient. Control advantages of the computer might be retained by allowing the person acting as recipient to have access to on-line computer facilities without the forecaster knowing.

The present study could be extended by periodically changing the reinforcement random variable used in Case I. This might be done to see how subjects would respond to periodic changes in recipient behavior. Such changes can be thought to correspond with the attitude on the part of the recipient that F is biased on the average some specific number of units, but that periodically this estimate of average bias must be changed. Changes in reinforcement have been examined in more abstract setting in learning experiments. Such experiments could provide predictions of behavior in a corresponding forecasting situation.

Finally, a third topic area would be a field study examining the effect of method of forecast revision on forecast validity. Situations would have to be identified where reward structures encourage the fore-
caster to want to bias his forecast and the recipient to want to adjust the forecast in an attempt to remove bias. Such situations might be found where sales estimates are made, where cash flows are projected and where new product cost are estimated. In these situations, methods of forecast revision would have to be identified and forecasters would be grouped by the method used to revise their forecasts. Then, statistical tests could be performed to test the effect of method of forecast revision on forecast validity.

These three areas of possible future research demonstrate that the present study can be thought of as only a beginning. In each case, the emphasis can easily be put on a measure of the amount of information available from a particular source of accounting data. Such studies should be of continuing interest to accountants.
The purpose of this appendix is to show that

\( k(x; z) \) forms a bivariate normal when:

(1) \( k(x; z) \) is distributed \( n(z, \sigma^2) \)

(2) \( f(z) \) is distributed \( n(\mu, \delta^2) \)

\[
k(x; z) f(z) = \frac{1}{2\pi\sigma\delta} e^{-\frac{(x-z)^2}{2\sigma^2} - \frac{(z-\mu)^2}{2\delta^2}}
\]

write \((x-z)^2\) as \((x-\mu + \mu - z)^2 = (z-\mu)^2 - 2(z-\mu)(x-\mu) + (x-\mu)^2\)

\[
k(x; z) f(z) = \frac{1}{2\pi\sigma\delta} e^{-\frac{1}{2} \left[ \begin{array}{c} \frac{1}{\delta^2} + \frac{1}{\sigma^2} \end{array} \right] (z-\mu)^2 - \left[ \begin{array}{c} \frac{2}{\sigma^2} (z-\mu)(x-\mu) + \frac{1}{\sigma^2} (x-\mu)^2 \end{array} \right]}
\]

let \( \rho^2 = \frac{\delta^2}{\sigma^2 + \delta^2} \)

\[
k(x; z) f(z) = \frac{1}{2\pi\sigma\delta} e^{-\frac{1}{2} \left[ \begin{array}{c} \frac{\sigma^2}{\delta^2 + \sigma^2} \end{array} \right] (z-\mu)^2 - \left[ \begin{array}{c} \frac{\sigma^2}{\delta^2 + \sigma^2} \end{array} \right] \left( \frac{1}{\sigma^2} + \frac{1}{\delta^2} \right) - \frac{2}{\sigma^2} (z-\mu)(x-\mu) + \frac{(x-\mu)^2}{\sigma^2} \}
\]
\[ k(x; z) f(z) = \]
\[ = \frac{1}{2 \pi \sigma \delta} \left[ \left( \frac{z-\mu}{\delta^2} \right)^2 - \frac{2}{\delta^2 + \sigma^2} (z-\mu)(x-\mu) + \frac{1}{\sigma^2 + \delta^2} (x-\mu)^2 \right] \]
\[ = \frac{1}{2 \pi \sigma \delta} \left[ \frac{1}{1-\rho^2} \left( z-\mu \right)^2 - \frac{2 \rho}{\delta^2 + \sigma^2} \frac{(z-\mu)}{\delta} \frac{(x-\mu)}{\sqrt{\delta^2 + \sigma^2}} + \right. \]
\[ \left. \frac{1}{\sqrt{1-\rho^2}} e^{-\frac{1}{2} \left( \frac{1}{1-\rho^2} \right)} \left[ \right] \right] \]
\[ = \frac{1}{2 \pi \sigma \delta} \left[ \frac{1}{\sqrt{\delta^2 + \sigma^2}} \sqrt{1-\rho^2} e^{-\frac{1}{2} \left( \frac{1}{1-\rho^2} \right)} \right] \]
\[ = \frac{1}{2 \pi \sigma \delta} \left[ \frac{1}{\sqrt{\delta^2 + \sigma^2}} \sqrt{1-\rho^2} \right] \]
\[ = \frac{1}{2 \pi \sigma \delta} \left[ \frac{1}{\sqrt{\delta^2 + \sigma^2}} \sqrt{1-\rho^2} e^{-\frac{1}{2} \left( \frac{1}{1-\rho^2} \right)} \right] \]
which is a bivariate normal distribution. The marginals are distributed in the following manner:

\[ x \sim n(\mu, \delta^2 + \sigma^2) \]

\[ y \sim n(\tilde{\mu}, \delta^2) \]
QUESTIONNAIRE

1. Mark on the scale below your opinion of the "good" estimate.

[Scale]

The actual usually was less than the "good" estimate

The actual was greater than the "good" estimate

The actual usually was greater than the "good" estimate about as often as the "good" estimate.

it was less than the "good" estimate.

2. Did you develop a method or strategy for determining your own forecast?
   Yes ________ No ________

3. If yes on question 2, please describe your method or strategy briefly.

4. Do you think that your superior had some method or strategy for revising your forecast?
   Yes ________ No ________

5. If yes on question 4, try to briefly describe your superior's method or strategy.
This is an experiment designed to measure forecasting ability. The experiment will be conducted at a typewriter-console on the second floor of this building.

Assume you are the Columbus sales representative of a single-product manufacturing concern. Each period your superior asks you to send him a forecast of the number of units of product you expect to be sold in Columbus in the upcoming period. Your superior examines your forecast, revises your forecast and then uses the revised forecast to evaluate your selling ability.

In this situation you would like to have the revised forecast be somewhat lower than the actual level of sales. Specifically, you should assume that your forecasting in a particular period is successful if:

\[
\text{ACTUAL SALES - REVISED FORECAST OF SALES} = +6 \text{ to } +10
\]

Try to be successful on as many trials as possible.

To aid you in your forecasting, you will be shown a figure which is a "good" estimate for sales in the upcoming period. Your superior has no knowledge of this figure. The experimenter will describe to you in what sense this figure is a "good" estimate.

The sequence of events which make up a period is as follows:

1. You will be asked "What is your forecast."

2. Decide what figure you will forecast. Communicate this figure to your superior by typing the number on the typewriter keyboard. After typing the number, hit the carriage return key. The experimenter will allow you 10 seconds to complete this step.

3. The typewriter will then provide you with three (3) numbers from your superior. These are:
   a) the revised forecast
   b) the actual, and
   c) actual - revised forecast

4. Note on the form provided whether or not you were successful at forecasting.
The sequence of events will be repeated until you are no longer asked "What is your forecast." This task is a difficult one. Do not get discouraged. With practice, your forecasting ability should improve. Remember, you should assume that your forecasting in a particular period is successful if:

ACTUAL SALES - REVISED FORECAST OF SALES = +6 to +10

Try to be successful on as many trials as possible. Do you have any questions? If you do, ask the experimenter now. The experimenter will answer no questions during the time the experiment is being conducted.
<table>
<thead>
<tr>
<th>PERIOD</th>
<th>&quot;GOOD&quot; ESTIMATE</th>
<th>SUCCESSFUL?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>378</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>387</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>499</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>283</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>538</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>361</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>204</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>439</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>484</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>253</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>439</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>407</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>486</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>203</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>539</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>361</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>391</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>329</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>637</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>390</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>474</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>268</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>608</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>534</td>
<td></td>
</tr>
<tr>
<td>PERIOD</td>
<td>&quot;GOOD&quot; ESTIMATE</td>
<td>SUCCESSFUL?</td>
</tr>
<tr>
<td>--------</td>
<td>----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>26</td>
<td>532</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>475</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>463</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>530</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>450</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>553</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>478</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>515</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>428</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>366</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>578</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>560</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>427</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>547</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>533</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>550</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>447</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>491</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>361</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>363</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>384</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>501</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>533</td>
<td></td>
</tr>
<tr>
<td>PERIOD</td>
<td>&quot;GOOD&quot; ESTIMATE</td>
<td>SUCCESSFUL?</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>-------------</td>
</tr>
<tr>
<td>51</td>
<td>593</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>496</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>215</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>431</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>355</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>256</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>657</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>471</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>410</td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>342</td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>490</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>434</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>446</td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>370</td>
<td></td>
</tr>
<tr>
<td>66</td>
<td>452</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>409</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>434</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>476</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>416</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>267</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>451</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>371</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>484</td>
<td></td>
</tr>
</tbody>
</table>
SELECTED BIBLIOGRAPHY

Books


<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Publisher</th>
<th>Location</th>
<th>Year</th>
</tr>
</thead>
</table>


Journals


