

**LIBRARY**  
**Michigan State**  
**University**

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

DATE DUE	DATE DUE	DATE DUE
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____

**MSU is An Affirmative Action/Equal Opportunity Institution**

c:\circ\datedue.pm3-p.1

STATISTICAL ANALYTICAL PROCEDURES USING  
INDUSTRY SPECIFIC INFORMATION:  
AN EMPIRICAL STUDY

by

Robert D. Allen

A DISSERTATION

Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

Department of Accounting

1992

1918-887

## ABSTRACT

### STATISTICAL ANALYTICAL PROCEDURES USING INDUSTRY SPECIFIC INFORMATION: AN EMPIRICAL STUDY

by

Robert D. Allen

The current study examines the use of statistical analytical procedures (SAPs). SAP models are developed using information from a sample of nine electric utilities. The study incorporates both financial and nonfinancial information in the prediction models. The primary objectives of the study were 1) to compare the performance of alternative statistical prediction methods, 2) to test the consistency of the models across companies, 3) to evaluate the usefulness of pooled prediction models, and 4) to evaluate the relative performance of quarterly and monthly prediction models.

Prediction models were developed for three accounts: revenue, fuel expense and production expense. These accounts were identified by practitioners as areas in which significant audit effort is expended. Therefore, SAPs have are believed to have the potential to decrease audit effort for these accounts.

The performance of eight alternative prediction methods was compared. Five of the prediction methods were regression methods, a sixth statistical method used was the Census X-11 time series method, and the remaining two methods were nonstatistical prediction methods.

The results indicated that the statistical prediction methods performed better than the nonstatistical methods. In particular, First-differences regression achieved predictions that were more accurate and more consistent than any of the other prediction methods.

The results of the pooled models indicated the potential benefits of combining information from multiple companies in an industry to generate account balance predictions. The pooled models were more accurate than the predictions obtained using individual company data in some situations.

Finally, the results indicated that monthly prediction models tended to achieve more accurate predictions than quarterly prediction models. However, this result was not true for all of the prediction models.



## ACKNOWLEDGEMENTS

First, I would like to thank the members of my dissertation committee, Dr. D. Dewey Ward (Chairman), Dr. Alvin A. Arens, and Dr. Frank Boster, for their invaluable encouragement and help. I would also like to acknowledge Professors Susan Haka, William E. McCarthy, and Edmund Outslay, who served as mentors during the entire doctoral program, for their advice, friendship and support.

I am also grateful to the Institute of Internal Auditors Research Foundation for their generous funding of this research project. Their financial support made possible the timely completion of the dissertation.

I express appreciation to the practitioners who helped me obtain the information required to conduct this project. I am particularly indebted to Jan Umbaugh of Deloitte and Touche, and David Scott and Grant Trexler of Price Waterhouse.

I express my appreciation to my parents, Terry and Carol Allen, for their examples, patience and nurturing throughout my life. My mother was also very helpful in providing editorial support as the project neared completion. The support of many other family members in the form of letters and telephone calls is gratefully acknowledged.

Finally, I am grateful to my wife, Naomi. Without her patience, love and encouragement this project would not have been possible. I am also grateful to our children, William, Andrew, Jonathan and Natalie, for their understanding during the painstaking process of completing this project.

## TABLE OF CONTENTS

LIST OF TABLES .....	ix
LIST OF FIGURES .....	x
1. <u>INTRODUCTION</u> .....	1
1.1 <u>Overview of Types of Analytical Procedures</u> .....	4
1.1.1 SAPs versus nonstatistical analytical procedures .....	5
1.1.2 Time-series and cross-sectional models .....	9
1.1.3 Types of predictor variables .....	10
1.2 <u>Overview of Prior Research in Analytical Procedures</u> .....	12
1.2.1 Effectiveness of alternative SAP methods .....	13
1.2.2 Consistency of SAP performance .....	13
1.2.3 Benefits of using data from multiple companies .....	15
1.2.4 Level of aggregation of analytical procedure models ...	15
1.3 <u>Objectives of the Current Study</u> .....	16
1.3.1 Effectiveness of alternative SAP methods .....	16
1.3.2 Consistency of SAP models across multiple companies ..	17
1.3.3 Effectiveness of using pooled data .....	17
1.3.4 Quarterly versus monthly prediction models .....	18
1.4 <u>Overview of Research Methodology</u> .....	18
1.5 <u>Summary and Organization of the Dissertation</u> .....	20
2. <u>LITERATURE REVIEW</u> .....	22
2.1 <u>Research in Analytical Procedures</u> .....	22
2.1.1 Descriptive studies .....	23
2.1.2 Nonstatistical analytical procedure studies .....	24
2.1.3 SAP studies .....	27
2.1.4 Simulation studies .....	31
2.1.4.1 Simulated accounting data and simulated errors .....	32
2.1.4.2 Real accounting data with simulated errors .....	33
2.2 <u>Objectives of the current study</u> .....	33
2.2.1 Alternative SAP methods .....	34
2.2.1.1 Regression .....	34
2.2.1.2 Census X-II: A time-series model .....	35
2.2.1.3 Other methods .....	37
2.2.2 Evaluating the consistency of SAP performance .....	37
2.2.2.1 The predictive ability of SAP methods ...	38

2.2.2.2	Identifying robust predictor variables . . . .	42
2.2.2.3	Assessing the technical validity of SAPs . .	43
2.2.3	SAP prediction models using pooled data . . . . .	44
2.2.3.1	More accurate predictions . . . . .	44
2.2.3.2	Identification of errors not identified by individual company SAPs . . . . .	46
2.2.3.3	Use of more current base-period data . . . .	47
2.2.4	Quarterly and monthly estimation models . . . . .	47
2.3	<u>Summary</u> . . . . .	48
3.	<u>METHODOLOGY</u> . . . . .	49
3.1	<u>Scope of the Current Study</u> . . . . .	49
3.1.1	Reasons for selection of the electric utilities industry . .	50
3.1.2	Accounts modeled . . . . .	53
3.1.3	Predictor Variables . . . . .	55
3.1.3.1	Identifying suitable descriptor variables . .	55
3.1.3.2	Explanation of predictor variables . . . . .	57
3.1.4	Characteristics of sample companies . . . . .	63
3.2	<u>Prediction Models</u> . . . . .	64
3.2.1	Regression models: functional form and explanation . .	65
3.2.1.1	Ordinary least squares regression . . . . .	65
3.2.1.2	Cochrane-Orcutt . . . . .	66
3.2.1.3	First-differences . . . . .	67
3.2.1.4	Unit-weighted regression (UWR) . . . . .	68
3.2.1.5	Unit-weighted regression with combined factor variables . . . . .	69
3.2.2	Census X-11 Model . . . . .	71
3.2.3	Naive models . . . . .	72
3.3	<u>Tests Performed to Meet the Objectives of the Study</u> . . . . .	74
3.3.1	Method performance . . . . .	74
3.3.1.2	Simulation analysis . . . . .	78
3.3.1.2.1	Materiality and error seedings . . . .	79
3.3.1.2.2	Investigation rules . . . . .	79
3.3.2	Model consistency . . . . .	81
3.3.2.1	Prediction performance . . . . .	81
3.3.2.2	Consistency of predictor variables . . . . .	82
3.3.2.3	Diagnostic testing . . . . .	83
3.3.3	Pooled data . . . . .	85
3.3.4	Quarterly vs. monthly data . . . . .	88
3.4	<u>Summary</u> . . . . .	88

4.	<u>RESULTS: METHOD PERFORMANCE</u> .....	90
4.1	<u>Performance of Alternative Prediction Methods</u> .....	91
4.1.1	Procedures used to compare prediction methods .....	91
4.1.2	Results .....	94
4.1.3	Implications .....	99
4.2	<u>Simulation Analysis</u> .....	101
4.2.1	Simulation procedures .....	103
4.2.2	Simulation results .....	106
4.2.3	Implications .....	114
4.3	<u>Annualized Prediction Errors as a Percentage of Materiality</u> ..	115
4.3.1	Procedures to annualize predictions .....	116
4.3.2	Results of annualized predictions .....	116
4.3.3	Implications of the results .....	119
4.4	<u>Summary</u> .....	120
5.	<u>RESULTS: SAP CONSISTENCY, ALTERNATIVE MODELS</u> .....	123
5.1	<u>Model Consistency</u> .....	123
5.1.1	Consistency of SAP model predictions .....	125
5.1.2	Consistency of predictor variables .....	131
5.1.2.1	Variables consistently improving prediction accuracy .....	132
5.1.2.1.1	Results of significant predictor variables .....	132
5.1.2.1.2	Implications .....	133
5.1.2.2	Incremental benefit of nonfinancial predictor variables .....	137
5.1.2.2.1	Results .....	139
5.1.2.2.2	Implications .....	144
5.1.3	Diagnostic testing .....	145
5.1.3.1	Autocorrelation .....	145
5.1.3.2	Continuity .....	148
5.1.3.3	Heteroscedasticity .....	149
5.1.3.4	Multicollinearity .....	149
5.1.3.6	Alternative test and summary of diagnostic testing results .....	154
5.2	<u>Pooled Models</u> .....	157
5.2.1	Pooled model procedures .....	158
5.2.2	Pooled model results .....	159
5.2.3	Implications of pooled model results .....	160

5.3	<u>Quarterly versus Monthly Models</u> .....	162
5.3.1	Procedures used to compare monthly and quarterly models .....	163
5.3.2	Results of monthly and quarterly prediction models ....	164
5.3.3	Implications .....	164
5.4	<u>Summary</u> .....	166
6.	<u>SUMMARY, IMPLICATIONS, CONTRIBUTIONS, LIMITATIONS, AND SUGGESTIONS FOR FUTURE RESEARCH</u> .....	168
6.1	<u>Summary of the Results and Implications</u> .....	169
6.1.1	Alternative SAP prediction methods .....	169
6.1.2	Consistency of SAP models .....	171
6.1.3	Pooled models .....	174
6.1.4	Quarterly versus monthly models .....	175
6.2	<u>Primary Contributions and Limitations</u> .....	175
6.2.1	Contributions of the current study .....	176
6.2.2	Limitations of the current study .....	177
6.3	<u>Suggestions for Future Research</u> .....	178
6.4	<u>Summary</u> .....	180
	REFERENCES .....	180

## LIST OF TABLES

Table 4.1	Revenue: Prediction MAPEs and Rankings . . . . .	95
Table 4.2	Fuel Expense: Prediction MAPEs and Rankings . . . . .	96
Table 4.3	Production Expense: Prediction MAPEs and Rankings . . . . .	98
Table 4.4	Simulation Results: Annual Error Seed Condition . . . . .	107
Table 4.5	Simulation Results: Quarterly Error Seed Condition . . . . .	110
Table 4.6	Simulation Results: Monthly Error Seed Condition . . . . .	112
Table 4.7	Annualized Prediction Results . . . . .	117
Table 5.1	Consistency of SAP Models: Revenue . . . . .	128
Table 5.2	Consistency of SAP Models: Fuel Expense . . . . .	130
Table 5.3	Robust Predictor Variables: Revenue . . . . .	134
Table 5.4	Robust Predictor Variables: Fuel Expense . . . . .	135
Table 5.5	Incremental Benefit of Nonfinancial Information . . . . .	142
Table 5.6	Autocorrelation Diagnostic Testing Results . . . . .	147
Table 5.7	Continuity Test Results . . . . .	150
Table 5.8	Heteroscedasticity Test Results . . . . .	151
Table 5.9	Multicollinearity Test Results . . . . .	152
Table 5.10	Normality Test Results . . . . .	153
Table 5.11	Diagnostic Test Summary: Regression Model . . . . .	156
Table 5.12	Comparison of Pooled Models and Individual Company Models . . . . .	161
Table 5.13	Comparison of Monthly Models to Quarterly Models . . . . .	165

## LIST OF FIGURES

Figure 1.1	Mix of Audit Tests Used to Obtain Sufficient Competent Evidence . . . . .	2
Figure 1.2	Nonstatistical and Statistical Analytical Procedures . . . . .	6
Figure 2.1	SAP Studies Using Actual Data . . . . .	40
Figure 2.2	Diagnostic Tests . . . . .	45
Figure 3.1	Predictor Variables . . . . .	58
Figure 3.2	Summary of Current Study . . . . .	73
Figure 3.3	Objectives, Methods, and Metrics . . . . .	75
Figure 3.4	Diagnostic Testing . . . . .	84
Figure 3.5	Individual vs. Pooled Model Predictions . . . . .	87
Figure 5.1	Predictor Variables . . . . .	140

## **Chapter I**

### **1. INTRODUCTION**

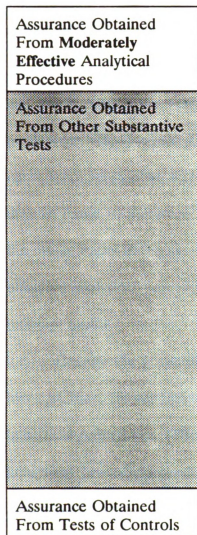
The competitive pressures that exist in the market for audits have induced members of the auditing profession to search for low-cost methods of obtaining audit assurance. In today's competitive audit environment there is a high demand for audit procedures that are both efficient and effective. Analytical procedures have been advocated by academics and practitioners as one possible means of obtaining audit assurance at low cost, compared to other audit procedures. The potential of analytical procedures to provide assurance at a relatively low cost is evidenced by their increased use on actual audits in recent years (Biggs and Wild, 1984). An additional indication of the potential of analytical procedures is the recent adoption of SAS 56 by the Auditing Standards Board which requires that analytical procedures be used on all audits (AICPA, 1988).

One reason analytical procedures are so appealing to practitioners is that they are generally performed more quickly and efficiently than other types of audit tests. Substantive tests of balances, for example, tend to be more expensive to perform than analytical procedures. Figure 1.1 illustrates that the assurance required to issue an opinion comes from a combination of evidence obtained from tests of controls, substantive tests, and analytical procedures. When more assurance can be obtained

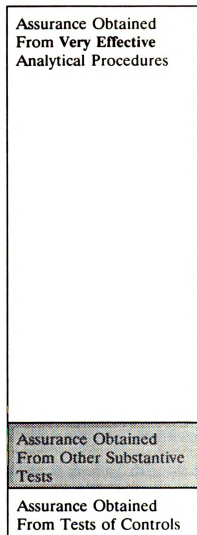


**Figure 1.1**  
**Mix of Audit Tests Used to**  
**Obtain Sufficient Competent Evidence**

**Panel A: Expensive Audit**



**Panel B: Inexpensive Audit**



Inexpensive



Expensive

from analytical procedures, less substantive testing is required, and therefore the audit is less costly. However, the nature of the assurance derived from analytical procedures depends on the effectiveness of the analytical test. A comparison of Panels A and B indicates that more assurance is obtained from very effective analytical procedures (Panel B) than from moderately effective analytical procedures (Panel A). As a result, in Panel B less assurance is required from the more expensive substantive tests. In Panel A less assurance was obtained from the moderately effective analytical procedures, and a greater amount of assurance must come from substantive tests. More substantive testing will be required in Panel A than in Panel B; therefore, the audit performed in Panel A is more expensive than the one in Panel B. Figure 1.1 underscores the potential benefits of developing more effective analytical procedures. More effective analytical procedures lead to more efficient audits.

Despite the potential of analytical procedures to provide audit assurance at low cost, there remain unanswered questions regarding the effectiveness of statistical analytical procedures. The current study examines 1) the consistency of statistical analytical procedure predictions, 2) the potential benefits of pooling data from multiple companies, 3) the level of data aggregation appropriate for analysis, and 4) the effectiveness of competing methods of conducting analytical procedures.

The remainder of this chapter is divided into five sections. The first section provides background information regarding different types of analytical procedures. This background information is necessary to understand the remaining sections of

this chapter. The second section contains a brief discussion of the contributions and limitations of previous research addressing the development of effective analytical procedures. This overview of the literature provides important perspective about the current use of analytical procedures. The overview of the literature also highlights important areas in which the current state of knowledge is inadequate. The third section is a discussion of the specific objectives of the current study, and how they address the limitations of prior studies mentioned in section two. An overview of the research approach used to meet the study's objectives is described in section four. Finally, the fifth section summarizes the chapter and presents the organization of the remainder of the dissertation.

### **1.1 Overview of Types of Analytical Procedures**

There are several different approaches used in conducting analytical procedures. The purpose of this section is to provide insight regarding the types of analytical procedures that have been researched in the existing auditing literature. An explanation of the various types of analytical procedures will assist the reader in understanding the succeeding sections that explain how the current study contributes to existing research related to analytical procedures.

As mentioned, the nature of analytical procedure applications used in auditing practice and examined in the accounting literature varies greatly. For example, some studies incorporate the use of statistical analytical procedure (SAP) methodologies, while others use simpler, nonstatistical approaches. Some statistical approaches incorporate time-series models, while others use cross-sectional models.

Furthermore, approaches vary by the nature of data that are available and incorporated into analytical procedure models. This section contains a discussion of each of these alternative approaches to conducting analytical procedures. Section 1.1.1 discusses the differences between SAPs and nonstatistical analytical procedures. Section 1.1.2 explains the differences between time-series and cross-sectional analytical procedure models, and section 1.1.3 describes the different types of predictor variables that may be incorporated into analytical procedures.

#### **1.1.1 SAPs versus nonstatistical analytical procedures**

One factor that influences the relative effectiveness of analytical procedures is the nature of the model used to generate predictions. Figure 1.2 demonstrates how analytical procedures vary in their level of complexity by providing examples of simple, nonstatistical approaches and more complex statistical approaches. Auditors may use nonstatistical approaches, or they may incorporate sophisticated statistical models to make predictions of account balances. Traditionally, auditors have favored use of nonstatistical approaches of conducting analytical procedures (Biggs and Wild, 1984). As indicated in Figure 1.2, these nonstatistical procedures include comparison of current year account balances to prior year. Statistical approaches include Ordinary Least Squares (OLS) regression, autoregressive-integrated-moving-average (ARIMA), simultaneous equations, and Census X-11.<sup>1</sup>

---

<sup>1</sup>Census X-11 is a time-series prediction model developed by the United States Department of the Census. Dugan, Gentry, and Shriver (1985) first introduced X-11 as a technique useful to auditors in conducting analytical procedures. The X-11 model is less time consuming to employ than other time-series models such as ARIMA (Dugan, Gentry, and Shriver, 1985).

**Figure 1.2**  
**Nonstatistical and Statistical Analytical Procedures**

**Nonstatistical  
Analytical Procedures**

<b>Single Account Analysis</b>	<b>Ratio Analysis</b>
Comparison of individual accounts to assess overall reasonableness of account balances.	Comparisons of relationships between accounts to assess overall reasonableness of account balances.
<b>Example:</b>  Compare current year Operating Expenses with prior year.	<b>Example:</b>  Compare current year Gross Margin Percentage to prior year.

**Statistical  
Analytical Procedures**

<b>Multiple Regression Analysis</b> <b>ARIMA</b> <b>Census X-11</b> <b>Structured Equations</b>
Statistical models that allow the modelling of relationships between variables.
<b>Example:</b>  <b>Sales</b> $\text{Rev.} = f(W, X, Y, Z)$  where,  $W = \text{Prior year Revenue}$ $X = \text{CPI}$ $Y = \text{Industry Growth}$ $Z = \text{Number of Stores}$

There are a number of advantages of employing statistical approaches as opposed to nonstatistical approaches. A discussion of these advantages is important to understanding how analytical procedure effectiveness may be enhanced. One of the primary advantages to using a statistical approach to conducting analytical procedures is that multiple relationships between variables may be examined simultaneously. It is more difficult for the auditor to assess the impact of many changes that take place simultaneously. For example, assume the auditor wants to assess the reasonableness of a company's gross margin (sales less cost of goods sold) for each month of the audit period. Some of the factors that affect gross margin include sales price, sales quantity, and inventory purchase price. Sales price and sales quantity are directly related to gross margin, while there is an inverse relationship between purchase price and gross margin. In a given month, sales prices might increase, sales quantity decrease and purchase prices increase. The effect these changes have on gross margin is difficult to determine due to the confounding effects of the sales price increase and quantity decrease. On the other hand, use of a statistical approach such as multiple regression analysis will allow the auditor to examine the effects of simultaneous changes in sales price, quantity and purchase price. The change in gross margin that would be expected from changes in each of the associated variables is easily determinable. In short, SAPs facilitate the simultaneous examination of the relationships of multiple variables.

Further, SAP procedures tend to be more objective than nonstatistical approaches. A commonly used nonstatistical approach is to compare the current year

account balance to the balance reported in the prior year. Differences that exceed a pre-specified percentage are then investigated and explained. Typically, the process of investigation and explanation entails asking the client why the account balance changed. Assume for example that the auditor plans to investigate differences between the current and prior year that exceed 10 percent. The current year accounts receivable balance is 15 percent higher than the prior year. The auditor would then seek an explanation why the accounts receivable balance increased by 15 percent. Such reasons might include an increase in sales volume or less stringent credit policies.

The process of seeking explanations for the observed changes may not always be objective. Wallace (1983) contains an amusing example of the lack of objectivity that may accrue from using soft, subjective evidence:

A CPA mistakenly asked why an expense item was down, when in fact it was up, relative to prior years. The client provided a list of explanations as to why that account might be low. Upon discovery of the error, the CPA returned to the client, explaining the prior mistake, and asked why that same expense item was, in fact, up. With little effort, the client developed a list of explanations of why the account that had been previously "explained" to be low, could similarly be "explained" to be high! (Wallace, 1983, p. 26).

Obtaining the explanations of the reasons for changes in account balances and documenting these reasons in the working papers can be a time-consuming process. The relationships that cause account balance changes can often be quantified and incorporated into SAP models, which often alleviates the need for explaining changes in the working papers. There is no need for written explanations of changes that are already explained by data incorporated into the SAP model.

On the other hand, SAPs are more costly than nonstatistical approaches to conducting analytical procedures. Use of SAPs requires more information and more time than most nonstatistical approaches. The assurance obtained from employing SAPs must be greater than the assurance obtained from using nonstatistical approaches to justify the use of SAPs.

To summarize, SAPs tend to be more precise and less biased than nonstatistical approaches to conducting analytical procedures. However, SAPs tend to be more costly to employ than nonstatistical approaches. In the current study, the performance of various SAPs is assessed. Nonstatistical methods are not examined because SAPs appear to have much greater potential than nonstatistical methods in providing audit assurance, as argued in the preceding discussion.

The preceding section highlighted differences between SAPS and nonstatistical approaches to conducting analytical procedures. The next section describes another fundamental way in which analytical procedures vary; namely, the section discusses the difference between time-series and cross-sectional approaches to conducting analytical procedures.

#### **1.1.2 Time-series and cross-sectional models**

Time-series applications use data from multiple points in time to generate predictions. An example of a time-series approach would be to incorporate 36 months of historical data to generate predictions for monthly balances for the current audit period. Such an approach facilitates the identification of relationships and trends in the data that occur over time.



Cross-sectional applications use data from a single point in time to generate predictions. An example of a cross-sectional approach would be for the auditor to examine annual operating results of a national retail chain by location. Examination of same period data by location may help the auditor identify stores with unusual characteristics that might be selected for more extensive audit testing.

The time-series approach is the predominant method used in the current study because the approach lends itself to meeting the objectives of the current study. The time-series approach captures the relationships and trends present in the data that were collected. The relationships and trends estimated from the data set are used to predict selected financial statement account balances. It also should be noted, however, that a combined time-series and cross-sectional approach is used to examine the effectiveness of analytical procedures that incorporate data from multiple companies.

### **1.1.3 Types of predictor variables**

To date, most research studies in analytical procedures have used only data that is included in the financial statements to generate predictions of the correct account balances (Akresh, et.al., 1988). Only a few studies have used both financial and nonfinancial information sources to conduct analytical procedures research (Albrecht and McKeown, 1977; Akresh and Wallace, 1981; Neter, 1981; Wild, 1987). These studies have, in general, achieved better predictions (and lower prediction errors) than studies that only use financial data.

**This class of studies has been questioned due to their use of internal-to-the-company independent variables. Kinney explains, "this need to use internal variables raises questions of audit logic as well as data problems for audit researchers" (Kinney, 1983, p. 198). By audit logic problems, Kinney means relying on predictor variables that are provided by the client organization. The assertion is that the use of internally generated predictor variables is problematic.**

**However, the use of internal predictor variables does not necessarily create a problem for the auditor. SAS 56 mentions factors that should influence the auditor's beliefs regarding the reliability of data. The factors mentioned are:**

- Whether the data was obtained from independent sources outside the entity or from within the entity.**
- Whether sources within the entity were independent of those who are responsible for the amount being audited.**
- Whether the data was developed under a reliable system with adequate controls.**
- Whether the data was subjected to audit testing in the current or prior year.**
- Whether the expectations were developed using data from a variety of sources (SAS 56, par. 16).**

**Auditors must consider the reliability of all data on which they rely. This does not mean, however, that data should not be used because they are collected internal to the company. Internally generated accounting data may be easily substantiated in many instances. The current study examines the relative prediction performance of utilizing data from a variety of different sources including the financial statements, operating and production data, environmental data and macroeconomic data.**

The preceding section is an overview of different approaches used to conduct analytical procedures. The perspective provided by this section is important to understanding how the current study contributes to the existing body of analytical procedures research. The next section is an overview of prior research related to the effectiveness of analytical procedures. The section also identifies the deficiencies of prior research and the areas in which further research is needed.

## **1.2 Overview of Prior Research in Analytical Procedures**

A review of the existing analytical procedures research suggests that further research is warranted to examine the application and use of analytical procedures. This section highlights four of the most significant areas in which the current body of literature is incomplete or inconclusive. The four areas identified are:

1. What statistical methods generate the most accurate predictions of account balances to conduct analytical procedures?
2. How consistent is the prediction performance of analytical procedures when applied to many companies in the same industry?
3. Is it possible to improve the effectiveness of analytical procedures by pooling data from multiple companies?
4. What level of aggregation of data is most appropriate when conducting analytical procedures?

Research that addresses these questions in a conclusive manner will be extremely helpful to practicing auditors whose goal is to provide efficient and effective audit services. The research related to each of these four questions is discussed next.

### **1.2.1 Effectiveness of alternative SAP methods**

Some studies have been devoted to identifying prediction methods that provide the most accurate predictions of account balances (Albrecht and McKeown, 1976; Kaplan, 1978; Kinney, 1978; Wild, 1987; Wheeler and Pany, 1990). These studies generally conclude that regression is the preferred statistical method for use in practice. However, the results of Wheeler and Pany (1990) suggest that the Census X-11 time-series method is preferred to regression in certain circumstances. However, Wheeler and Pany (1990) incorporate only a limited set of financial data in arriving at their conclusions. Because of the limited data set used, further testing is warranted to conclusively evaluate the relative performance of Census X-11 and regression. The current study compares the performance of these alternative SAP prediction methods.

### **1.2.2 Consistency of SAP performance**

In order to address the consistency of SAP performance two elements must be present. First, data must be collected from multiple companies. Second, the models should incorporate the information available to auditors, including both financial and nonfinancial information. These elements are discussed in the next two paragraphs.

One common difficulty of conducting analytical procedures research is obtaining the information required to perform the analysis. Thus, research related to the effectiveness of SAPs for multiple companies is difficult to accomplish due to limited availability of data. Because of this difficulty, most studies use only data that

is readily available in the financial statements (Akresh et.al., 1988). Other nonfinancial data are often excluded from the analysis. It is not surprising that many such studies conclude that analytical procedures are not very effective in developing account balance expectations (Loebbecke and Steinbart, 1987; Kinney, 1987). Before concluding that analytical procedures are not effective, it may be important to use data sets that include both financial and nonfinancial information to generate predictions.

There are a few analytical procedures studies that incorporate both financial and nonfinancial data. These studies indicate that analytical procedures may be very effective in accurately predicting account balances (Akresh and Wallace, 1981; Neter, 1981; Albrecht and McKeown, 1976; Wild, 1987). However, each study relies on data from a single company to generate individual account balance predictions. Therefore, the results of these case studies may not be generalizable to other firms or industries. It is unclear whether the positive results of these studies are isolated success stories, or whether they represent the types of predictions that are possible for most or all companies.

The current study incorporates both of the elements described above. The study incorporates information from multiple firms. The current study also uses a diverse set of information for predictions. The information collected includes both financial and nonfinancial predictor variables. Hence, a more complete evaluation of the consistency of SAP predictions is possible in the current study compared to prior studies.

### **1.2.3 Benefits of using data from multiple companies**

An additional limitation of the case studies mentioned above is that they have not examined the potential benefits of using information from multiple companies to generate account balance predictions. At present, there are no studies that have successfully combined data from multiple companies to generate useful account balance predictions for analytical procedures. The combining of data from multiple companies may provide the following advantages:

- reduction of model building costs
- identification of certain errors not detected by individual company models
- use of more current base-period data

The current study evaluates the effectiveness of simultaneously incorporating data from multiple companies into an analytical procedures framework. This approach may improve the effectiveness of analytical procedures.

### **1.2.4 Level of aggregation of analytical procedure models**

The auditor must decide the level of data aggregation that is appropriate for each analytical procedure application. Prior research studies have addressed the aggregation issue; however, the results of these studies are somewhat inconclusive. Wild (1987) concludes that monthly models generate more accurate account balance predictions than quarterly models. This suggests that disaggregate data is superior to aggregate data for conducting analytical procedures. However, another study suggests that monthly data may be inferior to quarterly data because quarterly data are reviewed by independent auditors and monthly data are not (Wheeler and Pany, 1990). These studies suggest that further research is needed to determine the

appropriate level of data aggregation for conducting analytical procedures. The current study addresses this issue by comparing the performance of prediction models constructed from monthly and quarterly data, respectively.

Further research must address previously unanswered questions related to the application of analytical procedures. The next section of this chapter describes how the objectives of the current study address some of the unresolved issues identified above.

### **1.3 Objectives of the Current Study**

The current study addresses the four areas highlighted in Section 1.2 where the analytical procedures literature is either incomplete or inconclusive. This section relates these areas to the specific objectives of the current study. The study has four primary objectives as follows: 1) to compare the effectiveness of alternative SAP methods, 2) to assess the consistency of SAP models for multiple companies in a single industry, 3) to assess the effectiveness of using pooled data, and 4) to compare the relative accuracy of quarterly and monthly prediction models. Each of these objectives is discussed in turn.

#### **1.3.1 Effectiveness of alternative SAP methods**

The first objective is to compare the effectiveness of alternative SAP methods. In a competitive market, auditors need to use procedures that are efficient and effective. Therefore, SAP methods must have the potential to predict account balances accurately. A comparison of various predictions methods is performed in

the current study. This comparison should identify the prediction methods with the most potential benefit to practitioners.

### **1.3.2 Consistency of SAP models across multiple companies**

The second objective is to assess the consistency of SAP models for multiple companies in a single industry. For statistical analytical procedures to be cost effective for use by auditors, they must apply to multiple companies. It is not likely that auditors will apply analytical procedures that are only useful to a single client. Thus, the current study evaluates the prediction performance of analytical procedures for multiple companies in a single industry. Prior studies are inadequate in addressing the consistency of analytical procedures because most of these studies used inadequate data sets, consisting exclusively of financial information. The few studies that did use both financial and nonfinancial data were case studies and are, therefore, inadequate in addressing the consistency of analytical procedures. The current study improves on prior research by using data sets from multiple companies, within a selected industry, including both financial and nonfinancial variables.

### **1.3.3 Effectiveness of using pooled data**

The third objective is to assess the effectiveness of using pooled data from multiple companies to estimate parameters for a single prediction model. Such multi-company models may facilitate use of SAPs when structural changes take place within a company or industry. For example, a structural change in a company after the prediction model for that company was developed would render the model ineffective. If a model was developed using industry-wide characteristics, then the



model should be less sensitive to changes in individual companies. Pooling data also allows the use of more current data, which mitigates the possibility of imprecision due to structural change. Multi-company models also may contribute towards reducing model building costs since they may be used by practitioners on multiple companies in the industry.

#### **1.3.4 Quarterly versus monthly prediction models**

The fourth objective is to compare the relative accuracy of predictions using quarterly and monthly prediction models. Some studies suggest that predictions from disaggregate (monthly) data are more precise than predictions from aggregate (quarterly) data (Wild, 1987). However, another study asserts that predictions from quarterly data are superior to those from monthly data (Wheeler and Pany, 1990). The current study will provide further empirical evidence regarding the relative accuracy of quarterly and monthly data in generating account balance predictions.

#### **1.4 Overview of Research Methodology**

Data from nine electric utilities are examined in the current study. Companies from a single industry are examined because it is not feasible that analytical procedures could be developed for multiple industries. Such multiple industry models are not feasible because of the many differences that exist between industries. The relationships that must be identified for accurate predictions are often masked by noise created by the inter-industry differences.

Data from a sample of the electric utilities are examined in the current study. Analysis of data from multiple companies will allow an evaluation of the consistency

of the performance of SAP models. Firms were selected to obtain a sample which is representative of the industry as a whole. Therefore, the electric utilities selected in the sample are located in various geographic locations and regulatory environments. The companies also vary in size and in the types of electric generating facilities.

Monthly data are collected which include both financial and nonfinancial information. Forty-eight months of data are collected from each utility. Each data set includes financial statement, operating, production, environmental, and price information. The nature of specific data items collected is discussed in greater detail in Chapter 3.

The prediction performance of various regression models and Census X-11 are compared with two nonstatistical models that serve as baseline prediction models. The performance of the most accurate models is also addressed by "seeding" varying levels of material errors into the recorded account balances. Predetermined investigation rules determine whether the auditor investigates differences between recorded and predicted balances. If the auditor investigates an account balance when no error has been seeded, then a type I error occurs. Similarly, if the auditor fails to investigate an account balance that has been seeded with a material error, then a type II error obtains. The incidence of type I and type II errors provides further evidence regarding the effectiveness of analytical procedures in signalling financial statement errors.

### **1.5 Summary and Organization of the Dissertation**

The purpose of the current study is to develop more effective analytical procedures. Analytical procedures with increased effectiveness will lead to more efficient audits. An overview of the literature suggests four research questions that were not addressed or were addressed inconclusively in prior studies:

1. Which statistical methods generate the most accurate predictions of account balances for purposes of conducting analytical procedures?
2. Are analytical procedures consistently effective when applied to many companies in the same industry?
3. Is it possible to improve the effectiveness of analytical procedures by simultaneously incorporating data from multiple companies?
4. What level of aggregation of data is most appropriate when conducting analytical procedures?

The research objectives of the current study address these questions. The objectives are:

1. To compare the effectiveness of alternative SAP methods.
2. To assess the consistency of SAP predictions for multiple companies in a single industry.
3. To assess the effectiveness of using pooled data.
4. To compare the relative accuracy of quarterly and monthly prediction models.

The objectives are addressed by analyzing a 48 months of financial and nonfinancial information from each of a sample of nine investor-owned electric utilities. The prediction performance of analytical procedures is examined for both a model construction and a "hold-out" period. A simulation analysis is also conducted in which errors of varying magnitudes are "seeded" into recorded account balances. The

incidence of both type I and type II errors will be examined to assess the effectiveness of analytical procedures in identifying material errors.

The remainder of the dissertation is divided into five chapters. Chapter 2 is a review of the literature. Chapter 3 describes the study's methodology. Chapters 4 and 5 present the results of the study. Chapter 6 contains a summary of the dissertation, conclusions, limitations, and suggestions for further research.

## **Chapter II**

### **2. LITERATURE REVIEW**

As indicated in Chapter 1, the overall objective of this study is towards improved effectiveness of analytical procedures. This chapter contains a discussion and analysis of academic research papers dealing with analytical procedures. A review of this research literature indicates that further research is needed to improve the effectiveness of analytical procedures. The chapter is divided into three main sections. The first section discusses individual research reports related to analytical procedures. Section two relates the research articles presented in section one with the objectives of the current study. The third section contains a summary of Chapter Two.

#### **2.1 Research in Analytical Procedures**

This section describes the existing auditing research in analytical procedures that relates to the current study. The analytical procedures research studies are classified as descriptive studies, nonstatistical studies, statistical studies and simulation studies. Descriptive studies are important in the context of the current study because they describe current practice. An understanding of current practice provides a useful starting point from which to develop analytical procedures that are even more effective. Accordingly, a discussion of descriptive studies is presented in section 2.1.1. Sections 2.1.2 and 2.1.3 discuss nonstatistical and statistical studies respectively. These sections demonstrate that SAPs have much greater potential to provide

additional cost-effective audit assurance compared to nonstatistical approaches. Section 2.1.4 contains a discussion of the simulation studies which have been conducted to test the effectiveness of analytical procedures in signalling material errors. All four sections provide the basis for the development of the study's objectives, which are discussed in Section 2.2.

### **2.1.1 Descriptive studies**

One purpose for conducting descriptive studies in analytical procedures is to learn how analytical procedures are used in practice to gain insight regarding how such procedures might be improved. It is difficult to suggest improvements for current practice before a good understanding of practice is obtained. The first descriptive study dealing with SAPs was published more than 15 years ago (Stringer, 1975). Stringer commented on the experiences of the accounting firm of Haskins and Sells using regression analysis for conducting analytical reviews. Stringer reported that more than 10,000 applications were processed during 1974. Reactions about the use of regression were generally favorable. Stringer's study conveyed the impression that regression-based analytical procedures are widely used. Stringer did not indicate the frequency of use of regression compared with simpler procedures.

Biggs and Wild (1984) found that simple procedures such as scanning financial statement data and ratio analysis are used in practice with greater frequency than statistical approaches. Their results indicate that quantitative techniques such as regression analysis and time-series models were only used by a small percentage of auditors. The results presented in Daroca and Holder (1985) are similar to those of

Biggs and Wild (1984) in that "exotic procedures requiring extensive mathematical techniques or additional data generation are only rarely employed in either audits or reviews" (Daroca and Holder, 1985, p. 92). Biggs and Wild (1984) indicate that less experienced practitioners are more likely to use quantitative techniques than more experienced practitioners. Perhaps more experienced auditors are less familiar and less comfortable with using quantitative techniques than less experienced auditors who may have more training in quantitative methods. Further research is needed to provide insight regarding the relative costs and benefits of SAPs and nonstatistical analytical procedures.

Tabor and Willis (1985) indicate that the use of analytical procedures has increased between 1978 and 1982. The study also indicates an increase for the same period in the use of quantitative methods of conducting analytical procedures. The auditors participating in the study all agreed that the use of analytical procedures will increase in the future. Forty-three percent of participating auditors stated they believe that analytical procedures will be less costly with increased use of the microcomputer. These findings suggest that SAPs will be used more in the future.

### **2.1.2 Nonstatistical analytical procedure studies**

Nonstatistical analytical procedures use simple comparisons between a few items. For example, the accounts receivable balance from the current audit period can be compared with the prior year balance. The auditor can then assess whether the change in accounts receivable is warranted, given his or her knowledge of other factors such as changes in credit policy or sales volume. The focus of nonstatistical

analytical procedures is 1) to direct the auditor's attention to segments of the audit that warrant examination and 2) to reduce the level of substantive tests when results are satisfactory. These nonstatistical studies fall into two categories: 1) those in which real data are used to generate predictions of account balances and ratios. (i.e., Loebbecke and Steinbart, 1987; Kinney and Salamon, 1987), and 2) those that investigate ex-post the effectiveness of nonstatistical procedures in identifying material audit adjustments in actual audits (i.e., Hylas and Ashton 1982; Wright and Ashton, 1989).

Loebbecke and Steinbart (1987) examined the effectiveness of a set of nonstatistical procedures using real accounting data in combination with simulated errors. The study focused on five different types of errors found to occur commonly in Coakley and Loebbecke (1985). Annual data for firms selected in the study are available on the COMPUSTAT data base. Four experiments were used to test the effectiveness of these "attention directing" analytical procedures. The first experiment tested the effectiveness of a simple ten percent change rule, which simply means that the auditor investigates changes greater than ten percent and does not investigate changes less than ten percent. In Experiment Two, eleven methods of generating account predictions were developed and tested. In Experiments Three and Four, the focus was to develop better investigation rules as opposed to better predictions. The results of all four experiments indicated that simple nonstatistical procedures were not effective enough to be used as a justification for the reduction of other substantive testing. The current study extends the analysis performed by



Loebbecke and Steinbart (1987) by determining if statistical analytical procedures are effective enough to justify the reduction of other substantive testing.

Kinney (1987) also investigated the effectiveness of nonstatistical analytical procedures. The focus of Kinney's case study was on the use of accounting ratios. Kinney used 48 periods of monthly data from a single firm for the analysis. Three investigation rules were used: 1) simple percentage change rule, 2) statistical standardized change rule and 3) a pattern analysis of cross-sectional changes in several ratios. Not surprisingly, effectiveness is closely linked to the relative size of seeded errors. This result emphasizes the importance of using disaggregated data, since errors are more likely to stand out as unusual when compared to smaller subannual balances as opposed to annual balances.

Hylas and Ashton, (1982) is an empirical study that reports on the nature of 281 errors requiring financial statement adjustments on 152 actual audits. The approach analyzed ex post the reasons for the occurrence of each error. The results of the study indicate that a high percentage of the errors were signaled with nonstatistical procedures.

Wright and Ashton (1989) improved on the methodology used by Hylas and Ashton (1982) by (among other improvements) providing information of the circumstances surrounding the use of nonstatistical analytical procedures. The study also examined the extent to which the proportion of signaled errors is conditional upon internal control strength. The results indicate that when controls are strong, analytical procedures involving internal accounting data are more likely to signal

errors. With weak controls, evidence external to the accounting records signals a greater proportion of errors. Hylas and Ashton (1982) and Wright and Ashton (1989) did not examine the performance of statistical approaches in signalling errors. The current study compares the performance of various alternative SAP methods in signalling material errors.

### **2.1.3 SAP studies**

In addition to the nonstatistical analytical procedures studies just discussed, there are many studies that examine the effectiveness of various statistical approaches to conducting analytical procedures. Alternative statistical approaches include regression, ARIMA (autoregressive-integrated-moving-average), simultaneous equations, and Census X-11.<sup>1</sup> The specific details of studies utilizing these methodologies are described next.

Kinney (1978) used regression, univariate ARIMA, and bivariate ARIMA to generate revenue predictions for six railroads. The predictions were generated exclusively from monthly revenues of each of the six railroads. Regression and bivariate ARIMA lead to more accurate predictions than univariate ARIMA and naive prediction models. Bivariate ARIMA was found to generate predictions that were slightly more accurate than regression; however, the author inferred that regression is the preferred method since it is less time-consuming to use than

---

<sup>1</sup>Census X-11 is a time series prediction model developed by the United States Department of the Census. Dugan, Gentry, and Shriver (1985) first introduced X-11 as a technique useful to auditors in conducting analytical procedures. The X-11 model is less time consuming to employ than other time series prediction models such as ARIMA.

**ARIMA.** In the current study, a more complete comparison of regression and time-series models is performed due to the inclusion of both financial and nonfinancial information into the prediction models.

Albrecht and Mckeown (1977) used monthly financial statements and other data to generate predictions. They compared the performance of regression, univariate ARIMA and bivariate ARIMA on three independent data sets provided to them by the accounting firm of Haskins and Sells. The results indicated that regression and bivariate ARIMA performed better than univariate ARIMA and naive martingale and submartingale models. Neither bivariate ARIMA nor regression emerged as clearly superior. The primary concern with this study is that each model is constructed with data from a single firm. The current study incorporates data from multiple firms for a more conclusive examination of the consistency of model predictions.

Akresh and Wallace, (1981) present the results of a case study that predicts certain income statement account balances for a gas and electric utility company. The primary methodology used is regression analysis. The authors also tested the usefulness of structured simultaneous equation methods. The authors incorporated monthly financial statement data as well as operating and environmental data to generate predictions. The results indicate that regression analysis performs well in generating predictions of the selected account balances. However, the only criteria used for model performance were goodness of fit measures because no out-of-sample predictions were conducted in the study. Another limitation of the study is that it

captures data from a single firm. The study also relied on budgeted data in generating predictions. The justification for using budgeted account balances as predictor variables was that budgeted data are subject to several levels of management review. Auditors would probably be unwilling to rely extensively on budgeted data for analytical procedures designed as test-of-details substitutes. The current study is an improvement over Akresh and Wallace (1981) in two important ways. First, more rigorous testing of the models is performed in the current study. Second, the consistency of predictions is tested more conclusively by including data from multiple firms. Inclusion of data from multiple firms provides greater evidence of the generalizability of the prediction models.

Neter (1981) used regression analysis to develop both time-series and cross-sectional models. The time-series application is an accounts receivable prediction model. The cross-sectional application of sales outlets is used to identify those whose performance is unusual. The primary emphasis in this study is the development of models that might be used by auditors given their time-budget pressures. It was not Neter's intent to develop highly sophisticated models, but to concentrate on applications that could be implemented in reasonable amounts of time by practicing auditors. It is difficult to evaluate the predictive ability achieved by the time-series application since the only measures of predictive ability are goodness of fit measures. The results are not tested on a holdout sample. Nevertheless, the time-series prediction models appear to perform very well. R-squares greater than .90 were reported. Mean percentage errors range between 4 and 15 percent. Neter's cross-

sectional study of sales outlets investigated the usefulness of prediction models that test the profit and loss accounts of a company with many sales outlets. The models facilitate identification of stores with unusual characteristics that may require further follow-up. The current study improves on Neter (1981) by including data from multiple companies and by testing prediction models in a hold-out period.

Wheeler and Pany (1990) test the relative effectiveness of regression and Census X-11 in conducting analytical procedures. They provide the first empirical evidence of the usefulness of X-11 for conducting analytical procedures. Expectation models are developed for both account balances and ratios. One focus of the study is to induce "best case" conditions for predictions by including single industry firms and quarterly data. The data incorporated in the study are financial statement data from the COMPUSTAT database. The results of the study indicate that X-11 predicts better than regression for ratios, but the reverse is true for account balances. The results also indicate that neither method is reliable in signaling material quarterly errors. However, when an annual material error is introduced, both X-11 and regression are reliable in signalling such errors. The primary limitation of Wheeler and Pany (1990) is that only financial information was included in the prediction models. The current study includes both financial and nonfinancial information in the prediction models.

In summary, regression, Census X-11, simultaneous equations, and ARIMA methods have been applied to analytical procedures. Results regarding the prediction performance of these methods indicates that further research is needed

to determine their relative effectiveness. Specifically, the relative performance of regression and Census X-11 has not been examined with a data set that includes both financial and nonfinancial information. The current study will make such a comparison.

Regression analysis has emerged as the most used statistical method of conducting analytical procedures in actual practice (Biggs and Wild, 1984; Stringer, 1975). The research literature indicates that other methods may rival regression in predictive performance; however, after considering the ease of use of regression compared to other methods, regression is the current method of choice for use by practitioners who desire to use statistical means of generating account predictions for analytical procedures (Kinney, 1983). Therefore, regression is used extensively in the current study and serves as a benchmark for other prediction methods. The prediction performance of other statistical methods, such as Census X-11, are compared against regression.

#### **2.1.4 Simulation studies**

Simulation studies examining the effectiveness of analytical procedures provide mixed results. These simulations are of two types: 1) simulated accounting data with simulated errors, and 2) real accounting data with simulated errors. These two types of studies are fundamentally different with respect to the input data used to generate account balance predictions. With simulated accounting data, the analytical procedure predictions are generated from synthetic data. With real accounting data, the predictions are generated using actual historical accounting numbers. The

effectiveness of the procedures examined in both types of studies is sometimes examined by introducing simulated errors into the "recorded" account balances to determine the incidence of type I and type II errors. A type I error occurs when the model signals that an error is present when no error has been seeded into the "recorded" account balance. Similarly, a type II error occurs when the model fails to signal an error when an error has been seeded into the account.

#### **2.1.4.1 Simulated accounting data and simulated errors**

Knechel (1986) examined the effectiveness of various approaches to conducting analytical procedures. The prediction performance of nine nonstatistical approaches were compared with four models based on regression analysis. Simulated accounting numbers were used for all 13 approaches and simulated predictor variables are used to generate account balance predictions for the regression models. Effectiveness was evaluated by comparing the incidence of type I and type II errors. The results indicate that regression models perform better than nonstatistical methods of conducting analytical procedures.

Knechel (1988) links the results of SAPs to the quantity of other procedures and assesses the combined effectiveness of SAPs used in combination with dollar unit sampling. The results indicate that use of regression combined with other substantive procedures improves audit effectiveness above that achieved through dollar unit sampling alone. Both of the Knechel studies (1986, 1988) use artificially generated accounting data. They also assume a high correlation between artificially generated independent and dependent variables ( $R^2 = .95$ ). The restrictive assumptions which

were employed in these studies limit the external validity of their results. The current study uses 48 months of actual accounting data, as opposed to artificially generated accounting numbers.

#### **2.1.4.2 Real accounting data with simulated errors**

Other studies use real accounting data to generate expectations, and errors are artificially introduced to evaluate prediction effectiveness. As indicated, two of these studies conclude that nonstatistical analytical procedures are not very effective (Kinney, 1987; Loebbecke and Steinbart, 1987). A third study, (Wheeler and Pany, 1990), suggests that the lack of analytical procedure effectiveness in Kinney (1987) may be due to measurement error in monthly data used in the study. Wheeler and Pany (1990), therefore, use quarterly data and still find that analytical procedures were not very effective in identifying material quarterly errors. Wheeler and Pany used a restrictive data set. They included only variables that are readily available in the financial statements, and did not use nonfinancial data or external data. The lack of performance of analytical procedures in detecting errors reported in this study may result from limited data sets rather than measurement error.

## **2.2 Objectives of the current study**

This section describes the objectives of the current study, and their relationship to prior studies. Some of the limitations of prior studies are mentioned, and the suggested contributions and improvements of the current study as it relates to prior work is also presented. The next four subsections relate prior research efforts to the study's primary objectives which are: 2.2.1) to compare the predictive



performance of alternative SAP methods, 2.2.2) to evaluate the consistency of SAPs, 2.2.3) to examine the use of pooled data, and 2.2.4) to compare the predictive performance of quarterly models with monthly models.

### **2.2.1 Alternative SAP methods**

In the current study, the performance of various regression models will be compared with Census X-11. Such a comparison is important because model selection will impact on prediction effectiveness. Prior studies have compared the effectiveness of various SAP methods. Subsection 2.2.4.1 contains a discussion of regression; Subsection 2.2.4.2 is a discussion of Census X-11. A discussion of other methods used in prior research articles is presented in Subsection 2.2.4.3.

#### **2.2.1.1 Regression**

Regression is the most widely used method of conducting SAPs (Biggs and Wild, 1984). Some of the advantages of using regression are addressed in Albrecht and McKeown (1977). Its primary advantage over time-series models is that multiple independent variables can easily be incorporated to make predictions of the account balance of interest. It is flexible enough to incorporate time-series properties such as seasonality and trend parameters, and it allows the exploration of certain nonlinear relationships between variables (Albrecht and McKeown (1977). Unlike time-series models, regression makes use of prediction variables from the audit period in generating predictions. Regression coefficients are estimated using base-period data. Regression coefficients are combined with predictor variables from the

audit period to generate predictions of the account balance of interest in the audit period.

#### **2.2.1.2 Census X-II: A time-series model**

Prior studies have experimented with the use of ARIMA time-series models in conducting analytical procedures (Albrecht and McKeown, 1977; Kinney, 1978). These methods are somewhat inaccessible for use in practice. Due to the computation effort and model building skill required to employ ARIMA, Kinney (1978, p. 59) concluded that "ARIMA-based models used for analytical review in auditing seem to be potentially beneficial but not as a generally applicable alternative to regression." Kinney (1983, p. 199) states, "given the relatively restrictive assumptions of time-series models and the relative simplicity of training for and application of regression, regression is more likely to be the preferred alternative for widespread practical use."

More recently, another time-series model (Census X-11) has been identified as potentially useful for use in conducting analytical procedures (Dugan, et al, 1985). Census X-11 is a time-series model that captures many of the benefits of ARIMA models. "Like ARIMA, the X-11 model decomposes time-series data into its trend-cycle, seasonality, and irregular components" (Wheeler and Pany, 1990, p. 582). The X-11 model is much less time consuming to apply and requires fewer observations to obtain valid predictions than ARIMA (Dugan, et al., 1985). The X-11 procedure is widely available, as evidenced by its inclusion as a procedure on the SAS statistical package (SAS, 1984). One primary disadvantage of X-11 (and other time-series

models) is that it relies exclusively on lagged observations of the dependent variable in developing predictions. Accordingly, other explanatory variables cannot be included in the X-11 prediction models.

Wheeler and Pany (1990) compare the performance of the X-11 model with regression and four naive (martingale and submartingale) models. They use actual, quarterly data from five single-industry companies, seeded with artificial errors. Their results indicate that X-11 performs better than regression in minimizing type I and type II error rates.

The performance of regression may be significantly improved by incorporating a richer set of predictor variables. Wheeler and Pany (1990) constructed the models exclusively with data from the COMPUSTAT database. In each regression model, the dependent variable is estimated from lagged observations of that variable, an industry statistic, and one other financial statement variable. Due to the data constraints imposed by use of COMPUSTAT, little attention could be given to identifying other variables that account for changes in the variable of interest. One objective of the current study is to identify a richer set of variables that provide predictions of the account balance of interest. This difference in focus is likely to improve the performance of regression compared to X-11.

Additionally, the five companies in Wheeler and Pany (1990) appear to have very low irregular (unexplained) components in their revenue streams. The favorable performance of X-11 may not be generalizable, assuming the irregular components of their sample companies are small compared to other companies.

### **2.2.1.3 Other methods**

The current study also incorporates two nonstatistical prediction methods. The two nonstatistical methods are referred to as the martingale and submartingale models. These two methods use prior period account balances as predictions of the current period. The performance of the statistical prediction models is compared with the nonstatistical prediction methods. The nonstatistical methods serve as a baseline prediction for comparative purposes.

Researchers have compared the predictive performance of other statistical methods that are not examined in the current study. These methods include structured (simultaneous equation) models and ARIMA techniques (discussed previously). Simultaneous equation models link systems of related regression equations. ARIMA techniques rely on lagged values of the account under audit in forming current expectations. Neither ARIMA nor structured models have emerged as clearly superior to regression in their predictive performance (Wild, 1987; Albrecht and McKeown, 1977; Kinney, 1978). These methods also require more time and effort to implement than regression (Kinney, 1983). The current study examines only those methodologies thought to be realistically practical for use by auditors. Accordingly, ARIMA and simultaneous equation techniques are not included in the analysis.

### **2.2.2 Evaluating the consistency of SAP performance**

The consistency of SAP performance is an important factor that effects their usefulness to practitioners. Practitioners are only likely to use methods and

procedures which are useful to multiple clients. Statistical analytical procedures will only be useful to practitioners if these procedures are consistently effective for most or all firms in a specific industry.

In the current study, the consistency of SAPs is examined in the current study in three ways in each the following subsections. First, consistency is examining by evaluating the predictive ability of SAPs applied to multiple companies in a single industry and identifying the characteristics that lead to good (and poor) predictions. Second, consistency is examined by identifying specific variables that are robust predictors of account balances for multiple companies within the selected industry. Third, robustness is examined by testing the technical validity of SAPs for each application.

#### **2.2.2.1 The predictive ability of SAP methods**

The results of research in SAPs provide mixed signals of the effectiveness of SAPs. The level of achieved precision varies greatly from one study to another. According to Kinney (1983), studies utilizing both financial and nonfinancial data as predictor variables achieve greater precision than SAP studies using only financial data as predictor variables. The paragraph that follows contains a discussion of the SAP studies which do incorporate both financial and nonfinancial information as predictor variables.

There are four primary SAP studies that use both financial and nonfinancial data as predictor variables (Albrecht and McKeown, 1977; Akresh and Wallace, 1981; Neter, 1981; and Wild, 1987). Figure 2.1 includes details of these four studies

(Studies One through Four). Each of these studies uses actual (as opposed to simulated) accounting data combined with other financial and nonfinancial data to make predictions. These studies achieve high  $R^2$  values, low prediction errors, and accurate predictive ability. These studies also indicate that nonfinancial data improve the predictive performance of SAPs. Three of the studies only evaluate model predictions from the model estimation period, and do not test the models in a "holdout" period (Akresh and Wallace, 1981; Neter, 1981; Albrecht and McKeown, 1977). The fourth (Wild, 1987) assesses the predictive ability of models in both a base period and a prediction (audit) period; however, the study does not attempt to investigate the effectiveness of SAPs in detecting errors. In the current study, model performance is assessed in the prediction (audit) period. The industry characteristics associated with accurate and inaccurate predictions are identified. In addition, the ability of the best prediction models to detect errors is evaluated by the artificial seeding of errors.

One noteworthy attribute of all four of the aforementioned studies is that the SAP models developed in these studies rely on data from a single company.<sup>2</sup> These authors acknowledge the difficulty of generalizing their conclusions to other companies and industries. The current study examines whether it is possible to achieve similar precision levels for all or most companies in a given industry. The

---

<sup>2</sup>One minor exception is Albrecht and McKeown (1977). In this study, independent models were developed from three different firms. Each prediction model was constructed with data from a single company.

**Figure 2.1**  
**SAP Studies Using Actual Data**

Study Num	Num Accts	Num Cos	Data Aggre-gation	Nonfinancial Financial	Method	Sim-ulation
1.	3	1	Monthly	Fin and Non	Regression	No
2.	8	1	Monthly	Fin and Non	Regression & Structured	No
3.	2	1	Month.** & Quarterly	Fin and Non	Regression	No
4.	14	1	Monthly & Quarterly	Fin and Non	Regression & Structured	No
5.	1	6	Monthly	Financial	Regression & ARIMA	No
6.	15***	5	Quarterly	Financial	Regression & X-11	Yes
7.	3	9	Monthly & Quarterly	Fin and Non	Regression & X-11	Yes

1. Albrecht and McKeown (1977)
2. Akresh and Wallace (1981)
3. Neter (1981)
4. Wild (1987)
5. Kinney (1978)
6. Wheeler and Pany (1990)

---

7. **Current Study.**

---

\* 3 account predictions, each with data from a single company.

\*\* Some "monthly" amounts are quarterly variables repeated 3 times.

\*\*\* 7 accounts, 8 ratio predictions.

current study also attempts to identify factors that affect the relative precision of predictions across companies.

One way of assessing the generalizability of the predictive performance of SAPs is to assess model performance for companies in a single industry. A single industry is selected as the unit of study for two primary reasons. The first is audit efficiency. Accounting firms may be able to leverage modeling techniques across multiple audit clients. Second, the examination of SAP performance of multiple companies within a single industry will allow a more comprehensive evaluation of the robustness of SAP models than has been performed in the aforementioned case studies. Furthermore, Loebbecke (1987) points out another important reason for conducting industry studies in analytical procedures:

Most extant research relating to analytical procedures has been done in the context of commercial and manufacturing companies. It probably is not appropriate to generalize the results of those studies to special industry groups. Moreover it may be that some techniques that aren't particularly effective for commercial and manufacturing companies are very effective within other industries.

As mentioned, prior studies indicate that it may be possible to make accurate predictions of account balances (Akresh and Wallace, 1981; Neter, 1981; Wild, 1987; Albrecht and McKeown, 1977). Nevertheless, do the results of these studies represent independent "success stories" with the use of SAPs, or are they indicative of the predictive performance that may be obtained by applying SAPs on most or all audits?



### **2.2.2.2 Identifying robust predictor variables**

Akresh, et. al., (1988) emphasize the importance of conducting analytical procedures research using data other than internal financial data as predictor variables. They suggest that predictive performance may be improved by including other relevant data into the procedure. "Little is known about what the other relevant data might be, how to obtain them, and how to incorporate such data in the most effective manner" (Akresh, et. al., 1988, p. 31). Other than Studies One through Four in Figure 2.1, there are no other studies which incorporate both financial and nonfinancial predictor variables. Studies One through Four only incorporate data from a single firm. Studies Five and Six use data from multiple companies; however, these studies do not include both financial and nonfinancial predictor variables. In the current study, an extensive set of data is collected. These data include both financial and nonfinancial predictor variables and are collected from multiple companies in the selected industry.

Application of regression as an analytical procedure with financial and nonfinancial data requires that the following tasks be performed before any analysis can be performed: 1) identification of suitable predictor variables, 2) collection of predictor variables, and 3) data input into a format appropriate for analysis. A portion of the costs of performing these tasks may be eliminated if robust predictor variables for specific industries are known in advance. Identification of variables that are useful predictors also provides the auditor with greater assurance that predictions are not based on spurious correlations between variables. Such assurance is

particularly important for cases in which many descriptor variables are used in regression applications. SAS 56 states that "it is important for the auditor to understand the reasons that make relationships plausible because data sometimes appear to be related when they are not, which could lead the auditor to erroneous conclusions" (AICPA, SAS 56). In the current study, robust predictor variables are identified for each account modeled in the selected industry. Identification of variables that are robust predictors of a given account balance for all companies (or a subset of companies with certain characteristics) in an industry reduces the possibility of relying on SAP predictions based on data that appear to be related when, in fact, they are not related.

#### **2.2.2.3 Assessing the technical validity of SAPs**

The technical validity of SAP models is an important factor affecting the usefulness of these procedures to auditors. Technical validity refers to the "robustness of the technique in the face of such problems as nonlinear relationships, multicollinearity, autocorrelation, heteroscedasticity, and nonnormality" (Elliot, 1977, p. 68). For example, in regression applications, adjacent observations are assumed to be independent. One study found that this assumption may be violated often (Albrecht and McKeown, 1977). Violations of the assumptions of regression may lead to inaccurate predictions. In the current study, diagnostic testing is conducted to assess the frequency and magnitude of these potential problems. Where applicable, corrective action is taken. Figure 2.2 lists each of the diagnostic tests that

are performed. Also listed are possible corrective procedures that will be employed to deal with these problems if and when they occur.

### **2.2.3 SAP prediction models using pooled data**

Prior studies have not examined the potential usefulness of simultaneously incorporating data from multiple firms into analytical procedure models. Such an approach requires a comprehensive data set including both financial and non-financial information. The primary reason prior research has not addressed multiple-company analytical procedure models is a lack of available data. Figure 2.1 lists SAP papers utilizing actual accounting data. This figure demonstrates that all studies incorporating both financial and nonfinancial information are single-firm case studies. In the current study, both financial and nonfinancial information are collected from multiple companies, which makes possible an evaluation of multiple-company analytical procedure models (hereafter called pooled models).

Effective pooled models are likely to lead to: 1) more accurate predictions, 2) identification of certain errors that may not be identified by individual firm SAP models, and 3) use of more current base-period data. Each of these potential advantages of pooled data are discussed next in Subsections 2.2.3.1 through 2.2.3.3, respectively.

#### **2.2.3.1 More accurate predictions**

Use of pooled models may lead to more accurate predictions. Pooling data yields more observations for analysis which, in turn, leads to greater statistical power.

**Figure 2.2**  
**Diagnostic Tests**

<u>Statistical Problem and Suggested Diagnostic Tests</u>	<u>Possible Corrective Action</u>
<b>Autocorrelation of Residuals:</b>	
<b>Durbin-Watson Test</b>	<b>First-Differences or Cochrane-Orcutt model regression (See Kinney, 1978)</b>
<b>Lack of Continuity:</b>	
<b>Chow Test</b>	<b>Identify sources of structural change, and eliminate observations from the model, if appropriate.</b>
<b>Heteroscedasticity:</b>	
<b>Goldfeld-Quandt Test</b>	<b>Exclude the descriptor variable(s) causing the problem.</b>
<b>Multicollinearity:</b>	
<b>Haitovsky Test</b>	<b>Use Unit-Weight regression or Unit- Weight regression combined with confirmatory factor analysis to identify the appropriate descriptor variable(s).</b>
<b>Normality:</b>	
<b>Kolmogorov-Smirnov Test</b>	<b>Identify omitted variables.</b>

A structural change in a company after the prediction model for that company was developed would render the model ineffective. If a model was developed using pooled data, then the models should be less sensitive to changes in an individual company. Changes taking place in one firm will not have as drastic an impact on the predictions estimated by a pooled model as they would have on the predictions of an individual firm model. Thus, pooling data may lead to more accurate predictions than individual company prediction models.

#### **2.2.3.2 Identification of errors not identified by individual company SAPs**

Use of pooled data may facilitate identification of certain errors that would not be discovered using an individual company SAP model. If, for example, material errors occur and persist from one period to another in a given company, individual company SAP models may fail to detect the error. A SAP model that makes extensive comparisons of the relationships between many variables from that company may not signal the errors if the errors are somewhat consistent over time. An error that is persistent in nature may be more difficult to detect because the relationships between variables may not stand out as unusual. Moreover, when data from multiple companies are pooled together, comparison of these same relationships increases the likelihood that the error will be detected. This is the case because the relationships would be different for companies whose financial statements do not contain the same type of persistent material errors, and the industry model would reflect this.

### **2.2.3.3 Use of more current base-period data**

Another potential advantage of using models developed with pooled data is that predictions with more current base-period data is facilitated. The application of most SAPs requires approximately 36 observations to estimate model parameters (Stringer, 1975). With a single company application, three years of monthly base-period data is required for estimation. Yet, if data from three homogeneous companies are pooled together, 36 observations are obtained with data from a single year. A reduction in the noise entering the model because of structural change may be accomplished through pooling. Use of pooled data from multiple companies with similar characteristics may increase the usefulness of SAPs for a higher percentage of companies by reducing the time span of the required base-period data.

### **2.2.4 Quarterly and monthly estimation models**

Existing research is inconclusive about the prediction performance of quarterly and monthly estimation models. One study argues that quarterly data may be superior to monthly data for predictive purposes since the former is subject to review by external auditors and the latter is not. The advantage is that quarterly review reduces the possibility of measurement error in the data (Wheeler and Pany, 1990). Additionally, quarterly data may contain less measurement error than monthly data due to the temporal aggregation of monthly cut-off errors. Kinney and Salamon (1979) suggest that measurement error increases the incidence of type I and type II error.

Another competing belief is that temporally disaggregate data allow more detailed analysis than aggregate data. Expectations derived from detailed analysis have a greater chance of detecting errors than do broad comparisons (SAS 56, par. 6). Results in the literature support the notion that disaggregate data lead to more precise predictions of account balances than aggregate data (Knechel, 1988; Wild, 1987). Wild (1987), for example, compares the performance of monthly and quarterly models. His findings suggest that monthly data provide more accurate predictions than quarterly data. These results are based on data from a single company and should be interpreted with caution. The assertion of Wheeler and Pany (1990) contrasted with the results of other studies (Wild, 1987; Kinney and Salamon, 1979) suggest that further investigation be undertaken to determine the relative accuracy of monthly or quarterly predictions.

Accordingly, the current study examines the relative performance of monthly and quarterly predictions using pooled data. The comparison of the relative performance of quarterly and monthly models will provide more conclusive evidence regarding the prediction accuracy of these models.

### **2.3 Summary**

This chapter evaluates the extant research related to analytical procedures. The literature review demonstrates the need for further research relative to analytical procedure effectiveness. The chapter also indicates how the objectives of the current study address some of the limitations of existing research. The next chapter contains a discussion of the research methodology that will be used in the current study.

## **Chapter III**

### **3. METHODOLOGY**

In the current study, statistical analytical procedure (SAP) models were developed for a sample of electric utilities using both financial and nonfinancial data. The inclusion of a variety of financial and nonfinancial information is believed to significantly improve prediction accuracy and increase the usefulness of SAPs to practitioners. The four primary objectives of the current study are:

1. To examine the performance of several alternative prediction methods.
2. To examine the consistency of SAP predictions across multiple accounts and multiple companies.
3. To compare the prediction performance of pooled prediction models with individual company prediction models.
4. To compare the performance of monthly and quarterly prediction models.

This chapter describes the methodological procedures followed in the current study. The chapter is divided into four sections. The first section describes the scope of the current study. The second section describes the prediction methods that were incorporated into the current study. The third section describes the specific tests performed to meet the four primary objectives of the current study. The fourth section contains a summary of the chapter.

#### **3.1 Scope of the Current Study**

This section describes the selection of industry, accounts modeled, information collected, and sample firms that were incorporated into the current study. Subsection



3.1.1 indicates the reasons why the electric utilities industry was selected for examination in the current study. Subsection 3.1.2 describes the accounts that were modeled. Subsection 3.1.3 describes the data collection procedures. Subsection 3.1.4 describes the characteristics of sample companies.

### **3.1.1 Reasons for selection of the electric utilities industry**

Data were collected from a sample of investor-owned electric utilities. Examination of a single industry was considered necessary to obtain precise estimates. A review of prior analytical procedures studies suggests that individual industries are the best units of analysis for study, based on the results of prior research (Akresh, et. al., 1988). The electric utility industry was selected for five reasons which are enumerated in the following paragraphs.

First, the utility industry has recently been identified by one Big-6 accounting firm as one of the five industries in which increased use of statistical analytical procedures (SAPs) has been recommended.<sup>1</sup> These industries were singled out because of the likelihood of reducing total audit effort through employing such an approach. For these industries, there are predictor variables, independent of the accounting function, that are expected to be useful in predicting certain account balances. Therefore, assurance may be obtained at a lower cost from performing SAPs than from other detailed substantive tests.

Second, the electric utility industry merited study due to its size and importance in financial markets. "The process of generation, transmission and

---

<sup>1</sup>The other industries are banking, retail, insurance and mining.

distribution of electricity is the nation's largest industry" (Phillips, 1984, p. 571). Furthermore, the industry requires tremendous amounts of investor capital to finance the construction of generating plant capacity. During the period 1982 through 1986 the industry required nearly \$12 billion each year in new capital (Hyman, 1988).

Third, the electric utility industry was selected due to the high degree of internal control strength found in the industry. Large, investor-owned utility companies are generally regarded as possessing strong systems of internal control (AICPA, 1990, par. 20). It was important to select an industry with strong internal controls to reduce the possibility of measurement error in the data. Kinney and Salamon (1979) found that measurement error in independent variables leads to a greater incidence of type I and type II errors. Errors in the dependent variable in the base period leads to increased type II errors. Unaudited data are more suspect to measurement error than audited data. The current study utilized sub-annual (monthly) account balances. Sub-annual account balances are unaudited. It is, therefore, important that the industry selected be composed of companies with strong systems of internal control in order to reduce the possibility of measurement error in both independent and dependent variables.

Fourth, electric utilities were selected due to data availability. The data requirements for the study were substantial. Kinney (1987) alluded to the continuing difficulty of obtaining actual, disaggregate accounting data for purposes of conducting analytical procedures research. However, because electric utilities are regulated, the collection of some of the data was facilitated. Federal and state regulations require

that the activities of utilities be measured separately from the activities of other lines of business. This facilitated the collection of unconsolidated financial statements even for utilities that are involved in other lines of business.

The abundance of publicly available annual data may have impacted utility company officials' willingness to provide monthly data for purposes of this research. Investor-owned public utilities are required to report a wide variety of financial and production data in SEC Form 10K and FERC Form 1 (an annual report required by the Federal Energy Regulation Commission). These sources were not adequate for purposes of this study since most of the data contained therein are annual. These data may also be less sensitive due to the fact that utilities are natural monopolies. Company officials in other industries may not have been as willing to supply similar production and operating data due to competitive pressures.

Fifth, the electric utility industry is composed of a relatively homogenous set of companies. Electric utilities vary in size, type of generating facilities, geographical region, and regulatory environment. However, they are homogenous in that their primary purpose is to generate, transmit, and distribute electricity to their customers. The probability of constructing accurate prediction models using pooled data is greater when the data come from a homogenous set of companies than a nonhomogenous set of companies. The current study is unique in its in-depth examination of a single industry. It was important that a homogenous industry be selected. If the current study is successful in identifying accurate prediction models,

then future research should pursue the development of statistical analytical procedures in less homogenous industries.

### **3.1.2 Accounts modeled**

Utility company experts provided information to identify the accounts that were modeled in the current study. These experts all had specific experience with public utilities. In total, 17 utility company experts provided information, including eight public accounting partners, four public accounting managers, four utility company employees and one academician. Experts were interviewed to identify accounts in which considerable audit effort is normally expended. The paragraphs which follow explain the logic for selecting the accounts that were modeled in the current study. Subsection 3.1.3 explains the selection of predictor variables for the current study.

The utility industry experts were interviewed to identify the accounts to be predicted. Additionally, one Big-6 firm provided a report which listed total audit effort expended on various accounts and other audit activities for their utility clients.

This report indicated two primary areas in which audit effort is expended. The two areas are 1) property and 2) operating revenues and expenses. Operating revenues and expenses took an average of 13.5 percent of total audit effort. Property accounts took an average of 10.4 percent of total audit effort. The next highest accounts were accounts receivable (4.4 percent of audit effort) and inventories (3.5 percent of audit effort). Due to the tremendous data requirements of modeling each account, it was considered beyond the scope of the study to make predictions for all of the accounts

mentioned in the previous paragraph. The next two paragraphs give the specific reasons for the accounts that were predicted.

Based on the report of audit effort and other discussions with auditors, revenues, operating revenue and expense accounts were considered to be the areas in which analytical procedure predictions were expected to have the most potential to reduce total audit effort. Revenue was selected as one of the account balances of interest due to its size and material importance. The most material operating expense account is fuel expense. Another significant operating expense account is production expense. These two accounts make up approximately 65% of the total dollar value of operating expenses. Therefore, total electric revenues, fuel expense and production expense are the three accounts for which predictions were made in the current study.

Property accounts were not selected for prediction. Changes in the property accounts generally result from new construction or capital improvements. Most of the audit effort expended for property accounts is related to these construction and improvement projects. Thus, property accounts are, in general, not suited for analytical procedure predictions. Therefore, property was not modeled in the current study.

Other material income statement accounts that were candidates for modeling were interest expenses and depreciation expenses. It is likely that accurate prediction models could be developed for these accounts (Akresh and Wallace, 1981). However, the accuracy of these account balances can be easily verified by

recomputation. It was, therefore, considered unlikely that modeling these accounts would reduce audit effort. Accordingly, these accounts were not modeled in the current study.

Prior research indicated that other accounts, such as accounts receivable, could also have been modeled (Neter, 1981). However, additional information, beyond that required for income statement account predictions would have been required from participating utility companies. Inclusion of other accounts would have greatly enlarged the scope of the study and may have had an adverse effect on the willingness of companies to participate in the study. Therefore, the current study included prediction models for only three accounts 1) revenues, 2) fuel expense, and 3) production expense. The development of accurate prediction models for these account balances has the potential to significantly reduce audit effort.

### **3.1.3 Predictor Variables**

This subsection contains an explanation of the predictor variables used in the current study. Subsection 3.1.3.1 describes the process used to identify suitable predictor variables. Subsection 3.1.3.2 is a description of each of the variables included in the study.

#### **3.1.3.1 Identifying suitable descriptor variables**

Interviews with utility company experts and a prior study (Akresh and Wallace) provided the basis for selecting the predictor variables for revenues, fuel expense, and production expense. Interviews were conducted with a total of 17 utility company experts, including Big-6 managers and partners, utility company

management, and academicians. There was considerable consensus among these experts regarding the predictor variables that should be incorporated. However, any time a predictor variable was suggested by a single practitioner, it was included in the study if it was possible to obtain the data. All of the experts suggested that some measure of volume and rates be included. Other variables were suggested by only a single expert, including load factor, capacity factor, and unemployment.

All of the variables used to predict revenues and production expense in the Akresh and Wallace (1981) case study were also incorporated in the current study. Akresh and Wallace (1981) identified several variables that were found to be useful in making revenue and expense predictions for a single utility. These include kilowatt-hour production, degree days, budgeted revenues and expenses, CPI, and fuel-CPI.

Based on the suggestions of the utility company experts, other variables were incorporated into the current study including residential, commercial and industrial rates, load-factor, capacity factor, lagged revenues and expenses, geographic region, type of generating facilities, and unemployment.

The information ultimately collected was monthly data for the years 1986 through 1989 (48 months). Specific variables included in the data analyses are listed in Figure 3.1. The variables included in the figure are discussed in greater detail in the next subsection.

There was one variable suggested by two experts that was not included in the current study. Two experts, whose clients had some hydroelectric facilities, suggested

the use of rainfall data as a predictor of fuel expense, the idea being that rainfall would be negatively correlated to fuel expense. Greater levels of rainfall would lead to more electricity from hydroelectric plants, thereby reducing the need for other fuels to fire nonhydroelectric plants. This variable was not incorporated for two reasons. First, hydroelectric power represents a small percentage of the total electricity produced by the firms in the sample and by the entire electric utility industry. Second, rainfall is not always a good indicator of hydroelectric production. Water levels on federal waterways are carefully managed. If the water level is low, then high levels of rainfall may not lead to higher hydroelectric output. Conversely, if the water level is high, hydroelectric production may be high notwithstanding low levels of rainfall.

### **3.1.3.2 Explanation of predictor variables**

The variables incorporated into the analytical procedure models are listed in Figure 3.1 and are of two types, financial and nonfinancial. Prior research has indicated that auditors tend to utilize readily available financial statement data in conducting analytical procedures. The current study examined the improvements in prediction performance resulting from inclusion of nonfinancial information which is not normally incorporated by auditors for analytical procedures (Biggs and Wild, 1984). A description of each of these variables is contained in this subsection.

Revenues, fuel expense and production expense were the three account balances of interest in the current study. Lagged observations of the account balance of interest from the preceding year were included as descriptor variables for each



**Figure 3.1**  
**Predictor Variables**

**Revenue Predictors:**

**Financial Statement  
Predictor Variables:**

Lagged Revenues  
Fuel Expense  
Production Expense

**Nonfinancial Statement  
Predictor Variables:**

Residential rate factor  
Commercial rate factor  
Industrial rate factor  
Heating degree days  
Cooling degree days  
KWH generated  
KWH sold  
Number of customers  
Unemployment  
CPI  
Fuel CPI  
Budgeted revenue

**Fuel Expense Predictors:**

**Financial Statement  
Predictor Variables:**

Revenues  
Lagged Fuel Expenses  
Production Expenses

**Nonfinancial Statement  
Predictor Variables:**

Heating degree days  
Cooling degree days  
KWH generated  
KWH sold  
Number of customers  
Unemployment rate  
CPI  
Fuel-CPI  
Capacity factor  
Load factor  
Budgeted production expense  
Geographic region\*  
Type of generating facilities\*

\* only incorporated in pooled models

**Figure 3.1 (cont'd)**  
**Predictor Variables**

**Production Expense Predictors:**

**Financial Statement**  
**Predictor Variables:**

Revenues  
Fuel Expenses  
Lagged Production Expenses

**Nonfinancial Statement**  
**Predictor Variables:**

Heating degree days  
Cooling degree days  
KWH generated  
KWH sold  
Number of customers  
Unemployment rate  
CPI  
Fuel-CPI  
Capacity factor  
Load factor  
Budgeted production expense  
Geographic region\*  
Type of generating facilities\*

\* only incorporated in pooled models

month, consistent with Wheeler and Pany (1990). Incorporating lagged observations of the dependent variable as an independent variable allowed the regression models to take advantage of the trend and cyclical time-series properties of prior-year account balances.

Fuel expense and production expense were used as predictor variables for revenue. Similarly, revenues and production expense were incorporated as predictor variables for fuel expense; revenues and fuel expense were used as predictors for production expense.

Rate factors were determined by computing the estimated monthly bill for the average residential, commercial and industrial customer. The estimated average bills were computed for each month and customer class by incorporating information from each company's published rate schedules. The estimated average bill was then divided by the number of customers to arrive at the rate factor for the month.

Degree day information is published monthly by the national weather service. Most utility companies routinely collect this information for large population centers in their service areas. In the current study, degree day information was obtained directly from participating utility companies. Degree days were computed as follows:

Monthly Heating degree days:

$$\sum_{i=1}^n [65 \text{ degrees} - ((HT_i + LT_i) / 2)], \text{ if } (HT_i + LT_i) / 2 < 65, 0 \text{ otherwise.}$$

Monthly Cooling degree days:

$$\sum_{i=1}^n [((HT_i + LT_i) / 2) - 65 \text{ degrees}], \text{ if } (HT_i + LT_i) / 2 > 65, 0 \text{ otherwise.}$$

where,

- $n$  is the number of days in the month.
- $HT$  is the high temperature for day  $i$ .
- $LT$  is the low temperature for day  $i$ .

Net kilowatt hours (KWH) generated by each company was also collected. Net KWH generated includes all KWH generated by company plant facilities plus net purchases and interchange. Net purchases constitutes the difference between sales to and from other utility companies. Interchange (or wheeling) refers to the transfer of electricity from a second utility to a third utility across company transmission lines. KWH sold refers to the number of KWH sold to residential, commercial, industrial, and other customers. The primary difference between the two is that KWH sales do not include transmission line losses and company uses of electricity.

Utility companies routinely keep track of the number of customers served. The average number of customers for each month was collected from each participating company.

Unemployment figures were collected by state for each company in the sample. One of the utility companies in the sample operates in two states, so the weighted-average unemployment figure was computed, based on the number of customers served by that company in each of the two states. The number of customers served in each state was collected from the 1989 Directory of Electric Utilities. Unemployment was expected to be negatively associated with revenues and

expenses. High unemployment rates are likely to lead to much lower industrial demand for electricity.

Consumer Price Index (CPI), and Fuel-CPI are published by the U.S. government. CPI was suggested to be positively correlated with the demand for electricity. Fuel-CPI was also expected to be positively correlated with both revenues and expenses. The relationship on the expense side is obvious; increases in the price of fuel will lead to higher fuel expense. On the revenue side, fuel adjustment clauses permit some electric utilities to automatically raise rates based on increases in the price of fuel.

Capacity factor and load factor are measures of plant utilization. They are defined as follows:

$$\text{Capacity Factor} = \text{AKWH} / \text{PC}$$

$$\text{Load Factor} = \text{AKHW} / \text{PL}$$

where,

AKWH = Average KWH per hour = KWH production / hours in period.

PC = Plant Capacity in megawatts.

PL = Peak load for the period in megawatts.

The higher the capacity and load factors, the higher the estimated use of more expensive peaking plant facilities. Therefore, the expected cost per kilowatt hour is expected to be higher than for low capacity and load factors.

Budgeted income statement information was also requested from each company. The incorporation of budgeted data as predictor variables may be questioned based on the lack of objectivity of budgeted data. However, a competing

point of view is that budgeted data are subject to review by multiple lines of management, and are therefore considered useful predictor variables (Akresh and Wallace, 1981). In the current study, budgeted data were incorporated into the prediction models. Budgeted revenue was used as a predictor of revenues, budgeted fuel expense as a predictor of fuel expense, and budgeted production expense as a predictor of production expense.

Two additional variables (geographic region and type of generating facilities) were included as expense predictor variables for the pooled models, the expectation being that firms with different characteristics may exhibit different cost function behavior. Two of the more significant differences which exist between utilities are the geographic regions in which they operate, and the type of generating facilities in operation. Accordingly, dummy variables were introduced in the pooled prediction models to capture these characteristics.

#### **3.1.4 Characteristics of sample companies**

Information was collected from nine investor-owned electric utilities. The objective of company selection was to obtain a sample representative of the population of investor-owned electric utilities. The companies included in the sample are located in various parts of the United States. Two firms are located in the West, three are located in the Midwest, and four are located in the South. The sample firms also vary in size. Assets ranged from \$918 million to \$20.5 billion. Average assets for the sample were \$6.6 billion. Annual net income of the sample firms

ranged from a low of \$42 million to a high of \$694 million. Average annual net income for the sample was \$282 million.

To obtain the nine data sets, controllers or chief financial officers from 14 electric utilities were contacted by telephone and asked to provide information for the study. Four companies declined to participate in the study. One company that agreed to participate did not provide complete information. The data from that company were not analyzed. Of the nine participating companies, three were unwilling to provide budgeted data. Additionally, three companies did not provide 1985 income statement information required for lagged observations of the account balances of interest. Therefore, a trend-dummy variable was incorporated as a surrogate for lagged observations of the dependent variable for these three companies.

The amount of information requested from each company was substantial. Each participating firm copied approximately 400 pages of printed material to comply with the data request for the current study. Companies participating in the study provided information on condition of anonymity. Accordingly, the names of participating firms are withheld.

### **3.2 Prediction Models**

This section describes the prediction models incorporated into the current study. There were a total of eight models tested. Six of the models are statistical models, including five regression models and the Census X-11 model. Two naive models were also estimated. The predictions for the statistical models were

generated using 36 months of base-period data from the period January, 1986 through December, 1988. The performance of these models was tested on a "hold-out" period from January through December of 1989. The functional form and an explanation of each of the models is presented in Subsections 3.2.1 through 3.2.3. Subsection 3.2.4 contains a summary of the current study.

### **3.2.1 Regression models: functional form and explanation**

Five different regression models were implemented for prediction purposes. The five regression models include 1) Ordinary Least Squares, 2) Cochrane-Orcutt, 3) First differences, 4) Unit-weighted regression, and 5) Unit-weighted regression with combined factor variables. The functional form and an explanation of each model is provided in Subsections 3.2.1.1 through 3.2.1.5 respectively. Subsection 3.2.1.6 contains a summary of the regression models.

#### **3.2.1.1 Ordinary least squares regression**

Ordinary least squares (OLS) regression is the most widely used methodology for conducting statistical analytical procedures (Wild, 1987). Most auditors are at least remotely familiar with the methodology as a result of statistical training from their university curriculum. The functional form of the OLS model is as follows:

$$y_t = \alpha + \beta_n x_{nt} + e_t$$

where,

- $y_t$  are the observed values of the account balance of interest (revenues, fuel expense, and production expense) in month  $t$ .
- $x_{nt}$  are observed values of  $n$  independent variables, in month  $t$ .
- $\beta_n$  are the  $n$  regression coefficients.



- $\alpha$  is the regression constant.
- $e_t$  are the residual terms, distributed  $(0, \sigma_e)$ .

The base period was January, 1986 through December, 1988 (36 months). The prediction (audit) period was January, 1989 through December, 1989. The variables included in each regression are presented in Figure 3.1. Those variables found to improve predictions were retained in the individual company models. The resulting sets of regression coefficients from the estimation period were used to make predictions for 1986 through 1989.

### 3.2.1.2 Cochrane-Orcutt

The functional form of the Cochrane-Orcutt model is as follows:

$$y_t - \delta y_{t-1} = \alpha(1 - \delta) + \beta_n(x_{nt} - \delta x_{nt-1}) + e_t$$

where,

- $y_t$  are the observed values of the account balance of interest (revenues, fuel expense, and production expense) in month  $t$ .
- $x_{nt}$  are observed values of  $n$  independent variables, in month  $t$ .
- $\beta_n$  are the  $n$  regression coefficients.
- $\alpha$  is the regression constant.
- $e_t$  are the residual terms, distributed  $(0, \sigma_e)$ .
- $\delta$  is the autoregressive parameter satisfying  $\text{abs}(\delta) < 1$ .

The Cochrane-Orcutt model was utilized to improve model performance when autocorrelation was present in a traditional OLS model. Usually the presence of autocorrelation is an indication that significant predictor variables are omitted. One

important contribution of the current study was to collect a comprehensive set of information useful for prediction. Accordingly, autocorrelation is not expected to be a significant problem. However, in the event that significant autocorrelation is present, the Cochrane-Orcutt model introduces an autoregressive parameter which compensates for significant autocorrelation.

### **3.2.1.3 First-differences**

The functional form of the First-differences model is as follows:

$$(1 - B)y_t = \beta_n(1 - B)x_{nt} + e_t$$

where,

- $y_t$  are the observed values of the account balance of interest (revenues, fuel expense, and production expense) in month  $t$ .
- $x_{nt}$  are observed values of  $n$  independent variables, in month  $t$ .
- $\beta_n$  are the  $n$  regression coefficients.
- $e_t$  are the residual terms, distributed  $(0, \sigma_e)$ .
- $B$  is a backshift operator such that  $B^k y_t = y_{t-k}$

The First-differences model provides another potential correction for autocorrelation.

The model estimates differences between adjacent observations of the data set. As mentioned above, autocorrelation is not expected to be a problem. However, the first differences model may provide better predictions than OLS when significant autocorrelation is present.

Both Cochrane-Orcutt and First-differences were incorporated as potential corrections for autocorrelation. The only way to determine which of the two is

superior in controlling for autocorrelation in the context of the current study was to test both methods using the data collected. The incidence of autocorrelation was measured using the Durbin-Watson test statistic for OLS, Cochrane-Orcutt and First-differences. The effects of significant autocorrelation on prediction accuracy was also measured.

#### **3.2.1.4 Unit-weighted regression (UWR)**

The functional form of the UWR model is as follows:

$$y_t = \beta Z(x_{nt}) + e_t$$

where,

- $y_t$  are the observed values of the account balance of interest (revenues, fuel expense, and production expense) in month  $t$ .
- $x_{nt}$  are observed values of  $n$  independent variables, in month  $t$ .
- $\beta$  is regression coefficient for the combined variables.
- $e_t$  are the residual terms, distributed  $(0, \sigma_e)$ .
- $Z(x)$  is the standardized value of  $x$  where  $Z_t = (x_t - \mu_x)/\sigma_x$  and  $\mu_x$  is the mean of  $x$ .

Schmidt (1971,72) indicated that Unit-weighted regression is preferable to OLS regression when the number of predictor variables is large relative to the number of observations in the model. Prior studies have established the importance of limiting the number of observations in the base period when estimating account balances for analytical procedures (Stringer, 1975; Akresh and Wallace, 1981; Kinney, 1978). These studies have demonstrated that the use of a relatively short (36-month) base period is better than using longer base periods for conducting analytical procedures.

In the current study, a 36-month base period was utilized. The current study also incorporated a large set of predictor variables, compared to many other studies (Wheeler and Pany, 1991; Kinney, 1978, Albrecht and McKeown, 1976). Thus, the number of predictor variables was large, relative to the number of observations. Therefore, if the results of Schmidt (1972) generalize to the current study, then Unit-weighted regression should provide more accurate predictions than OLS regression.

### **3.2.1.5 Unit-weighted regression with combined factor variables**

The functional form of the UWR with combined factor variables model is as follows:

$$y_t = \beta[F(x_{nt}) + x_{mt}] + e_t$$

where,

- $y_t$  are the observed values of the account balance of interest (revenues, fuel expense, and production expense) in month  $t$ .
- $x_{nt}$  and  $x_{mt}$  are standardized, observed values of the independent variables, in month  $t$ .
- $\beta$  is the regression coefficient for the combined, standardized variables.
- $e_t$  are the residual terms, distributed  $(0, \sigma_e)$ .
- $F$  are factor variables composed of two or more indicators.

One potential difficulty of applying simple OLS regression is that high standard errors occur when multicollinearity is present. The employment of confirmatory factor analysis is one means of reducing the potentially harmful effects of multicollinearity. Confirmatory factor analysis was incorporated in the current study to combine highly correlated variables into a single factor. The factor was then

used as a predictor variable in the regression model. Combining highly correlated individual variables into single factors reduces the problems associated with multicollinearity.

The factors were identified by first standardizing and then combining variables found to be highly correlated in the base period. Each factor was tested for internal consistency and parallelism. The factors identified were as follows:

Factor One =  $f(\text{KWH generated, KWH sold, Budgeted Revenue, Budgeted Fuel, and Number of Customers})$

Factor Two =  $f(\text{Residential Rates, Commercial Rates})$

#### **3.2.1.6 Summary of regression**

Each of the five regression models were estimated for all nine firms in the sample, for each of the three accounts. The regression models were compared to the three time-series models (i.e. Census X-11, martingale, and submartingale models). The statistical models (five regression models and Census X-11) were estimated using 36 months of base-period data from the period January, 1986 through December, 1988. These models were tested on a "hold-out" period from January through December, 1989.

The performance of the regression models described in Subsections 3.2.1.1 through 3.2.1.5 were compared against the performance of three time-series models. The Census X-11 time-series model is described next in Subsection 3.2.2. The other two time-series models, the martingale and sub-martingale models, are both described in Subsection 3.2.3.

### 3.2.2 Census X-11 Model

The Census X-11 model is a multiplicative decomposition method of forecasting. Decomposition methods attempt to break down data series patterns into subpatterns. With the X-11 procedure, each data series is decomposed into its trend-cycle, seasonal, and irregular components. The trend-cycle component is composed of the long-term trend in the time-series and business cycle. The seasonal component is composed of the intra-year variation, which is constant from year to year. The irregular component is the remaining variation not explained by the trend-cycle or seasonal components.

The trend-cycle, seasonal, and irregular components were estimated using time-series data from the 36-month base period (1986-1988). Estimates for the audit period (1989) were then computed based on the values of the three components computed from the base-period data.

The specification of the model is:

$$y_t = TC \times S \times I$$

where,

- $y_t$  is the account balance of interest at time  $t$ .
- TC is the trend-cycle (long-term variation in the series).
- S is the seasonal (intra-year variation in the series).
- I is the irregular (unexplained) component.

The X-11 method for making the audit period estimates was modified slightly in the current study, consistent with Wheeler and Pany (1990). To estimate the year

ahead trend-cycle, the normal X-11 procedure utilizes a univariate regression with time as the independent variable and the historical X-11 trend-cycle component as the dependant variable. This procedure extrapolates the linear trend, but does not identify any cycle which exists in the data series. The residual component of the univariate regression is assumed to be the cycle. Makridakis, Wheelwright, and McGee (1983) suggest fitting sine waves to estimate the cycle in the data. This procedure was performed by regressing the cycle component (i.e., the residual component from the univariate regression) on a series of sine waves. Fitting the sine waves extracts any cycle in the data.

### 3.2.3 Naive models

In order to provide a basis for comparison, monthly martingale and submartingale models were included in the analysis. The specification of the models is:

$$\text{Martingale:} \quad y_t = y_{t-12}$$

$$\text{Submartingale:} \quad y_t = y_{t-12} + [y_{t-12} - y_{t-24}]$$

Unlike the regression models, the martingale, submartingale and Census X-11 models only utilize prior observations of the dependant variable for predictions. The regression models incorporate a rich set of financial and nonfinancial data as predictor variables to generate predictions.

### 3.2.4 Summary

A summary of the current study is presented in Figure 3.2. This figure reviews the accounts modeled, prediction methods, number of companies and time period of

**Figure 3.2**  
**Summary of Current Study**

3 Accounts Predicted	8 Methods	9 Companies	48 Months
Revenues	OLS Regression	All electric utility companies	36 month base period, beginning January, 1986, ending January, 1988. Base-period data were used to estimate model parameters
Fuel Expense	Cochrane-Orcutt Regression	From 3 different geographic regions and various regulatory environments	12 month prediction period from January through December 1989. Model parameters from base period combined with data from prediction period to estimate account balances in audit period
Production Expense	First-Differences Regression	Varying sizes	
	Unit-Weighted Regression	Varying types of generating facilities	
	Unit-Weighted CFA Regression		
	Census X-11		
	Martingale		
	Sub-Martingale		



the current study. Eight prediction methods were used in the current study to develop predictions for revenues, fuel expense, and production expense accounts. Data were collected from nine investor-owned electric utilities for a 48-month period beginning January, 1986 and ending December, 1989.

### **3.3 Tests Performed to Meet the Objectives of the Study**

This section contains a discussion of the procedures that were performed to meet each of the objectives of the current study. A summary of these procedures is presented in Figure 3.3. The figure specifies methods and metrics used to meet each of the four objectives of the study.

The tests used to assess the relative performance of alternative SAP methods are discussed in Subsection 3.3.1. Subsection 3.3.2 presents the tests performed to assess model consistency. Subsection 3.3.3 explains the tests for assessing the performance of pooled data models. Subsection 3.3.4 presents the tests performed to assess the performance of quarterly and monthly prediction models.

#### **3.3.1 Method performance**

The relative performance of the five regression methods, Census X-11 and the two naive methods was also examined. The best method was identified by comparing the mean absolute percentage errors (MAPEs) of all prediction methods. The methods used to determine the "best" prediction model are presented in Subsection 3.3.1.1. Once the best model was identified, a simulation analysis was performed to determine the effectiveness of statistical analytical procedures in signalling material errors. The simulation analysis procedures are presented in Subsection 3.3.1.2.

**Figure 3.3**  
**Objectives, Methods, and Metrics**

Objective	Method	Metric
<b>1. Performance of alternative methods</b>		
<b>1A. Compare prediction performance of alternative methods.</b>	Ordinary least squares, Cochrane-Orcutt, First-differences, Unit-Weighted, Unit-Weighted CFA, Census X-11, martingale and submartingale	<p>Mean Absolute Percentage Error (MAPE) = <math>(y_t - \hat{y}_t)/y_t</math>  where,  <math>y_t</math> = recorded value  <math>\hat{y}_t</math> = predicted value</p> <p>Lowest average MAPE yields the best prediction method.</p>
<b>1B. Assess the ability of best method in detecting material errors.</b>	<p>Two Phase Simulation Analysis.</p> <p>Phase I  No errors are seeded into account balances. If an "investigation rule" signals an investigation, then a type I error has occurred.</p> <p>Phase II  Material errors of three magnitudes are artificially seeded into recorded account balances. If an "investigation rule" fails to signal an investigation, then a type II error has occurred.</p>	<p>Simple Investigation Rule:</p> <p>If <math>(y_t - \hat{y}_t)/y_t &gt; 10\%</math>, then investigate.</p> <p>Statistical Investigation Rule:</p> <p>If <math>(y_t - \hat{y}_t)/s_y &gt; Z_{1-\alpha}</math>, then investigate.</p> <p>where,  <math>s_y</math> is the base period standard deviation of the series, <math>y_t</math>.</p> <p>The incidence of type I, type II and combined errors is measured.</p>

**Figure 3.3 (cont'd)**

Objective	Method	Metric
<b>2. Consistency of SAP models.</b>		
<b>2A. Consistency of predictions</b>	Examine the MAPEs by company and account. Comment on the factors which appear to be related to good and poor predictions.	Mean absolute percentage error (MAPE)  $\text{MAPE} = (y_t - \hat{y}_t)/y_t$ where, $y_t$ = recorded value $\hat{y}_t$ = predicted value
<b>2B1. Consistency of predictor variables.</b>	Examine the strength of the association between the predictor variables and the account balance of interest.	t-values associated with individual variable predictions
<b>2B2. Consistency of financial and nonfinancial predictor variables.</b>	Examine the incremental benefit of introducing nonfinancial predictor variables into the prediction models.	Compare MAPEs of models predicted from financial variables only with the MAPEs of models predicted from both financial and nonfinancial predictor variables.
<b>2C. Incidence and effect of violations of the assumptions of the statistical models.</b>	Test for violations of the assumptions of statistical models, including: Autocorrelation Continuity Heteroscedasticity Multicollinearity  Note: Figure 3.4 contains a more detailed description of diagnostic procedures.	Durbin-Watson statistic Chow test statistic Goldfeld-Quandt stat. Haitovsky statistic

**Figure 3.3 (cont'd)**

Objective	Method	Metric
<p>3. Compare the performance of pooled models with the performance of individual firm models.</p>	<p>Compare individual firm MAPEs with pooled MAPEs.</p> <p>Note: Figure 3.5 contains a detailed description of the differences between individual firm prediction models and pooled prediction models.</p>	<p>Mean absolute percentage error (MAPE)</p> $\text{MAPE} = (y_t - \hat{y}_t)/y_t$ <p>where,</p> <p><math>y_t</math> = recorded value  <math>\hat{y}_t</math> = predicted value</p> <p>Lower MAPEs indicate better predictions.</p>
<p>4. Compare the prediction performance of monthly and quarterly SAP models</p>	<p>Estimate monthly prediction models and quarterly prediction models. Compare the relative prediction performance of each.</p>	<p>Mean absolute percentage error (MAPE)</p> $\text{MAPE} = (y_t - \hat{y}_t)/y_t$ <p>where,</p> <p><math>y_t</math> = recorded value  <math>\hat{y}_t</math> = predicted value</p>

### **3.3.1.1 Identification of best prediction method using MAPEs**

The best prediction method was identified by comparing mean absolute percentage errors (MAPEs). MAPEs were computed as the difference between the predicted account balance and the recorded account balance, divided by the recorded account balance. A ranking of MAPEs from lowest (most accurate) to highest (least accurate) was made for each firm and account. These rankings were used to determine the most accurate prediction method.

The next subsection describes the simulation procedures which were performed to assess the performance of the estimation models described in Subsection 3.3.1.2. The simulation procedures provide insight regarding the ability of the prediction models to signal material errors in the account balances of interest.

### **3.3.1.2 Simulation analysis**

The artificial seeding of errors is a method of evaluating the ability of the models to identify type I and type II errors. A type I error occurs when the model signals that an error is present when no error is present. Likewise, a type II error occurs when the model fails to signal an error when one is present. Seeding of errors was accomplished by specifying materiality levels and investigation rules in advance. The current study incorporated three materiality levels and two investigations rules in estimating type I and type II error rates (Wheeler and Pany, 1990; Loebbecke and Steinbart, 1987; Kinney, 1987). Subsection 3.3.1 describes the methods of computing materiality and the error seedings for the simulation. Subsection 3.3.2 describes the

investigation rules used to signal the presence or absence of an investigation for the simulation.

#### **3.3.1.2.1 Materiality and error seedings**

Consistent with Wheeler and Pany (1990), the largest of the following materiality definitions was utilized to create best-case conditions for analytical procedure performance:

- "audit gauge" [ $1.6 \times (\text{greater of total assets or revenues})^{2/3}$ ] (Elliot, 1983)<sup>2</sup>.
- 10 percent of net income (Holstrum and Messier, 1982).
- 10 percent of average earnings over a three-year period (Kinney, 1979).
- .5 percent of revenues (Wheeler and Pany, 1990).

Error seedings of three magnitudes were each added to monthly predictions. The three magnitudes correspond to annual, quarterly, and monthly material errors. Annual materiality was determined by computing  $M^*$ , which was the largest amount obtained from the above four materiality methods.  $M^*/4$  was considered to be a quarterly material error, and  $M^*/12$  was considered to be a monthly material error. Thus, annual, quarterly and monthly material errors were seeded into each monthly account balance.

#### **3.3.1.2.2 Investigation rules**

The auditor must establish a decision rule for subsequent investigation of unusual differences between predicted and recorded amounts. The current study

---

<sup>2</sup>This procedure for computing materiality was developed by KPMG Peat Marwick.

compared model performance using two different decision rules that were employed in prior studies: the percentage change rule and the statistical rule (Kinney, 1987; Wheeler and Pany, 1990). The percentage change rule signals an investigation when the predicted account balance,  $\hat{y}_t$ , differs from the recorded account,  $y_t$ , by more than a critical percentage set by the auditor (10 percent in the current study). Therefore, an investigation would take place if

$$(y_t - \hat{y}_t)/y_t > 10 \text{ percent.}$$

The statistical rule signals an investigation when the standardized difference between the recorded and predicted account balance exceeds the critical Z value, which is based on the auditor's specified risk level,  $\alpha$ . In the current study,  $\alpha = .05$ . An investigation would take place if  $(y_t - \hat{y}_t)/s_y > Z_{1-\alpha}$ , where  $s_y$  is the base period standard deviation of the series,  $y_t$ .

The percentage of type I and type II errors for each of three error magnitudes and two investigation rules was computed for the best regression model, the Census X-11 model and the martingale model. The submartingale model was not examined in the simulation because the martingale model exhibited consistently lower MAPEs than the submartingale model.

In the current study, a type I error could only occur when the account balance of interest was **not** seeded with error. If the investigation rule signalled an error when no error was seeded into the account balance of interest, then a type I error occurred. The type I error percentage was computed by dividing the sum of all

months in which an investigation was signalled (when no error was seeded) by the total number of months not seeded with material errors.

In the current study, a type II error could only occur when the account balance of interest was seeded with a material error. If an error was seeded into the account balance of interest and the investigation rule did **not** signal an investigation, then there was a type II error. The type II error percentage was computed by dividing the sum of all months in which no investigation was signalled (when a material error was seeded) by the total number of months seeded with material errors.

The type I and type II error rates were computed for each error magnitude and each investigation rule. Consistent with prior studies, type I and type II error rates were combined as a final assessment of each model's ability to appropriately signal the presence or absence of material errors (Loebbecke and Steinbart, 1987; Wheeler and Pany, 1990).

### **3.3.2 Model consistency**

The consistency of SAP models was assessed by 1) examining the prediction performance of individual company SAP models, 2) by evaluating the consistency of predictor variables, and 3) by performing diagnostic tests of the assumptions of the models. These are discussed in Subsections 3.3.2.1 through 3.3.2.3, respectively.

#### **3.3.2.1 Prediction performance**

The prediction performance of the SAP models was assessed by examining mean absolute percentage errors (MAPEs) for each company's best regression model



and each Census X-11 model. Companies were ranked based on their MAPEs in the model building period from lowest to highest. The adjusted multiple correlation coefficient ( $R^2$ ) was reported for each model. Furthermore, the characteristics of companies that were associated with good and poor predictions were also identified.

### **3.3.2.2 Consistency of predictor variables**

A variable was considered to be a consistency good predictor if it significantly improved model predictions for a high percentage of firms in the sample. The significance of each variable was evaluated by examining the variable's t-statistic. T-statistics greater than 1.00 were found to improve predictions and were therefore considered to be significant. The consistency of individual predictor variables was assessed by computing the percentage of times each predictor variable was significant.

The current study also examined separately the consistency of nonfinancial predictor variables. This analysis provided insight into the benefits of incorporating nonfinancial data into analytical procedure models. The consistency of nonfinancial predictor variables was examined by comparing the performance of models estimated from both financial and nonfinancial information with the performance of models estimated from financial information only. These models were evaluated by comparing the MAPEs from the financial-variables-only models with the financial-and-nonfinancial-variables models.

### **3.3.2.3 Diagnostic testing**

Additionally, the diagnostic tests presented in Figure 3.4 were performed to test the assumptions of the regression models employed in the study. Diagnostic tests were performed to examine the incidence of 1) autocorrelation, 2) continuity, 3) heteroscedasticity, and 4) multicollinearity. Each of the statistical tests employed are described in the next four paragraphs, respectively.

Autocorrelation of the regression residuals refers to the tendency of the residuals to move in a systematic pattern. Autocorrelation typically results from omitted descriptor variables. The Durbin-Watson (DW) test statistic signals the presence or absence of autocorrelation. Accordingly, DW test statistics were computed and are reported for the best fitting regression models.

Tests for continuity investigate whether changes have occurred in the model over time. In the current study, the Chow test statistic was computed to assess continuity by comparing the first 24 observations in each data set to the last 24 observations. If the Chow test indicated a lack of continuity, then the first twelve observations from the original model building period were deleted, and the regression model was estimated again. This shortens the base period, and thereby reduces the possibility of structural changes adversely affecting the continuity of the model. Adjusted R-square statistics of the original model were then compared to the newly estimated model to determine which of the two models should be retained.

**Figure 3.4**  
**Diagnostic Testing**

<u>Statistical Problem and Suggested Diagnostic Tests</u>	<u>Possible Corrective Action</u>
<b>Autocorrelation of Residuals:</b>	
Durbin-Watson Statistic	First-Differences or Cochrane-Orcutt model regression (See Kinney, 1978)
<b>Lack of Continuity:</b>	
Chow Test Statistic	Eliminate from the model observations giving rise to structural change.
<b>Heteroscedasticity:</b>	
Goldfeld-Quandt Statistic	Exclude the descriptor variable(s) causing the problem.
<b>Multicollinearity:</b>	
Haitovsky Statistic	Use confirmatory factor analysis to group highly correlated variables into a single factor.

Heteroscedasticity refers to the tendency of one or more descriptor variables to move systematically with the error term. The Goldfeld-Quandt statistic tests each predictor variable to determine if it moves systematically with the error term. If the Goldfeld-Quandt statistic was significant for a given predictor variable, then the variable was dropped and a revised model was recomputed. Adjusted R-squares of the original and revised models were then compared to decide which model to retain.

Multicollinearity is present when there is a high degree of correlation between independent variables. The presence of multicollinearity may not be harmful to model predictions. The Haitovsky statistic measures the incidence of multicollinearity and was reported in the current study. This statistic only signals the presence of multicollinearity. It does not indicate whether multicollinearity is harmful to the model. However, to deal with the potentially harmful effects of multicollinearity, variables found to be highly correlated were combined into single factors using confirmatory factor analysis. This procedure was described in Subsection 3.2.1.5. Unit-weighted regression with combined factors was incorporated to reduce the effects of harmful multicollinearity.

The diagnostic procedures described above were designed to examine violations of the assumptions of regression. As circumstances warranted, the above mentioned corrective actions were also taken in an attempt to improve predictions.

### **3.3.3 Pooled data**

This subsection explains the procedures used to assess the effectiveness of pooling data. In prior studies, statistical analytical procedure predictions have been

exclusively for individual firms. In the current study, the performance of individual firm predictions is compared with the performance of pooled prediction models.

Figure 3.5 contains an explanation of the difference between individual company prediction models and pooled prediction models. A pooled data model simultaneously incorporates observations from multiple companies. The additional explanatory power gained from additional observations may improve predictions and enhance the generalizability of the models, thereby increasing their usefulness to practitioners. Accordingly, models were estimated using pooled data from multiple companies in the sample<sup>3</sup>. In order to assess the effectiveness of predictions using pooled data, MAPEs from individual firm models were compared to MAPEs from pooled models.

Pooled models were estimated using ordinary-least-squares regression. The other regression methods were not applicable because pooling the data changes the time-series properties of the data. Therefore, the individual firm OLS results are used as a basis for comparison.

Pooled data models may be estimated with more current base-period data than individual models. Using more current base-period data reduced the possibility of structural changes harming model predictions. Accordingly, the pooled models were estimated again using only 1988 base-period data. Once again, the effectiveness

---

<sup>3</sup>Pooling was performed with data from the five firms that provided all data including budgeted data and lagged (1985) income statement data.

**Figure 3.5**  
**Individual vs. Pooled Model Predictions**

Individual Company Regression Models	Pooled Regression Models
<p>For each of the three accounts, regression coefficients, <math>\alpha</math> and <math>\beta_n^*</math>, were estimated individually for each utility company in the sample (9 times). The regression coefficients estimated from company 1 data were used to make predictions for company 1, etc.</p>	<p>Regression coefficients, <math>\alpha</math> and <math>\beta_n^*</math>, were estimated one time using the data from all nine companies in the sample. The single set of regression parameters is used to make predictions for all nine firms.</p>
<p>* n denotes that a parameter estimate is made for each of the predictor variables listed in Figure 3.1.</p>	

of these predictions was examined by comparing MAPEs from individual firm models with MAPEs from 1988-pooled models.

#### **3.3.4 Quarterly vs. monthly data**

Regression models were estimated with quarterly data to compare the prediction performance of monthly and quarterly models. The models were estimated from pooled data only. Quarterly estimates were not possible for individual company models due to insufficient observations in the model building period ( $n=12$ ; three year model building period, four quarters each year). The performance of the quarterly and monthly models was evaluated by comparing MAPEs of the quarterly and monthly models in both the model building and prediction periods.

#### **3.4 Summary**

The methodology utilized in the current study was described in this chapter. The first section contained a description of the selected industry, the accounts modeled, the information collected and the characteristics of sample companies. Prediction models for revenues, fuel expense and production expense account balances from nine investor-owned electric utilities were estimated. The information collected to make predictions included budgeted and actual financial statement data, operating and production information, environmental and economic variables. The second section described the prediction models utilized in the current study. The prediction performance of five regression models, Census X-11 and two naive models

were compared in the current study. The third section described the specific tests performed to meet the objectives of the current study.

The analysis of the results follow in the next two chapters. Chapter four presents the results of the best prediction model as well as the results of the simulation analysis. Chapter five contains the results related to 1) the consistency of SAP models, 2) pooled models, and 3) the relative performance of quarterly and monthly prediction models.



## **Chapter IV**

### **4. RESULTS: METHOD PERFORMANCE**

In the current study statistical analytical procedure (SAP) models were developed for a sample of electric utilities. SAP models may provide auditors with a means of obtaining audit assurance at a lower cost than through performing other substantive tests. The specific objectives of the current study were to 1) to compare the performance of a number of alternative SAP prediction methods, 2) assess the consistency of SAP models, 3) to assess the performance of pooled SAP models, and 4) to compare the performance of quarterly and monthly prediction models.

This chapter contains the results related to the first objective of the current study. SAP method performance was analyzed in three ways. Three different measurements were used because multiple measurements of prediction accuracy provide more conclusive evidence regarding the relative performance of the methods tested. First, performance was examined by comparing the prediction accuracy of each of the eight methods using monthly predictions. The results of this analysis are presented Section 4.1. Second, performance was analyzed by assessing the ability of the prediction methods in detecting material errors artificially seeded into account balances. The results of the simulation analysis are presented in Section 4.2. Section 4.3 compares the accuracy of the prediction methods using annualized predictions. Section 4.4 contains a summary of the chapter.

#### **4.1 Performance of Alternative Prediction Methods**

In the current study, the performance of eight prediction methods was compared. This section contains the results of that comparison. Subsection 4.1.1 contains a discussion of the methods used to compare the performance of the prediction methods. Subsection 4.1.2 contains a discussion of the results. Subsection 4.1.3 is a discussion of the implications of these findings.

##### **4.1.1 Procedures used to compare prediction methods**

The eight methods that were compared include: 1) OLS regression, 2) Cochrane-Orcutt regression, 3) First-differences regression, 4) Unit-weighted regression, 5) Unit-weighted regression with combined factor variables, 6) Census X-11, 7) a martingale model, and 8) a submartingale model. The procedures used to test the performance of these methods are presented in this subsection.

Methods one through five are regression methods. Ordinary-least-squares regression is the classical regression model. Cochrane-Orcutt and First-differences are alternative regression methods that were developed to deal with the potentially harmful effects of autocorrelation. These methods may improve prediction accuracy in the current study due to the manner in which they treat the time-series properties of the data. Unit-weighted regression and Unit-weighted regression with combined factor variables were developed to deal with multicollinearity. These methods may significantly improve predictions due to the manner in which they control the effects of multicollinearity.

Census X-11 is a time series method developed by the U.S. Department of the Census. This method captures the time series behavior of prior observations of the account balance of interest to generate predictions of current account balances. The method has been applied extensively in the statistics literature and has been recently advocated for use as a prediction method for analytical procedures (Dugan, et. al., 1985; Wheeler and Pany, 1990).

The five regression methods and Census X-11 are classified as statistical prediction methods. The remaining two, the martingale and submartingale methods, are naive methods included for comparative purposes. The statistical methods require more data a considerably more effort than the two naive methods. Therefore, the statistical methods must perform better than the naive methods in order to be considered useful. The model specifications for each of these methods was presented in the preceding chapter in Section 3.2.

Prediction estimates were computed for each of the nine companies in the sample, for 48 months, for each of the eight methods, and for each of the three accounts (revenues, fuel expense, and production expense) for a total of over 10,000 monthly predictions. The prediction models were developed individually for each company using 36 months of base period data covering the period from January, 1986 through December, 1988. The model parameters were estimated using the base period data only. The models were tested on a hold-out period (referred to as the prediction period) beginning January, 1989 and ending December, 1989.

Prediction accuracy was evaluated using the prediction period. Using goodness of fit measures from the model-building period may be misleading. It is possible that models exhibiting "good fit" perform poorly on out of sample data. Goodness of fit is necessary, but not sufficient for good predictions. Thus, the current study tests the models developed on a hold-out sample to provide a more stringent test of model performance.

Mean absolute percentage errors (MAPEs) from the prediction period were used to evaluate the performance of each prediction method. MAPEs were computed by taking the absolute value of the difference between the prediction and the recorded account balance, divided by the recorded account balance. This gives the absolute value of the error. The absolute value of the percentage error was computed for conservatism. Thus, overstatement and understatement errors were not allowed to counterbalance. Using the percentage error allows the comparison of account balances of varying sizes, which was necessary given the varying sizes of companies in the sample.

Two criteria were used to evaluate the eight methods using MAPEs. First, the average MAPEs across all nine companies in the sample were examined for each of the three accounts. The lowest average MAPE denotes the most accurate prediction method. In addition to examining the average MAPEs, a ranking of methods was performed for each company. The purpose for the ranking was to measure the consistency of individual method performance. A method may perform well on average and still be unacceptable to auditors due to occasionally inaccurate

predictions. Accordingly, the ranking of methods for each company provided a means to examine the consistency of the performance of each method.

The comparison of alternative prediction methods is an important step towards the determining which methods will be most useful to auditors. Auditors will favor the use of methods which provide consistently accurate predictions. Furthermore, to be considered useful to auditors, the statistical methods must outperform less complicated, easier to employ methods such as the martingale and submartingale methods.

#### **4.1.2 Results**

The performance of the eight methods is presented in Tables 4.1 through 4.3. These tables present the results for revenues, fuel expense, and production expense, respectively. Panel A of each of these tables includes the MAPEs for each company and for each method. Panel A also includes the ranking of each method, based on the average MAPEs. Panel B includes the ranking, by company, of all eight methods. Panel B also includes the worst ranking for each method. The worst ranking provides a measure of the consistency of the performance of each method. The results for revenue and fuel expense are discussed next, followed by a discussion of production expense.

Panel A of Tables 4.1 and 4.2 demonstrate that First-differences is the best prediction method for revenues and fuel expense, respectively. The average MAPE using First-differences is 3.8 percent for revenue and 8.7 percent for fuel expense.

### Table 4.1

### Panel A: Prediction MAPEs

## Company

Prediction Method	1	2	3	4	5	6	7	8	9	A V G	R A N K
OLS	5.4	2.9	2.3	3.5	7.9	4.2	3.7	3.7	3.1	4.1	2
CO	4.4	1.9	1.4	3.4	8.0	3.5	3.4	12.5	3.1	4.6	3
FD	5.0	2.0	4.5	3.6	6.4	3.7	5.2	4.4	2.6	3.8	1
UWR	5.5	5.4	2.3	10.0	9.8	3.5	8.7	10.7	10.6	7.4	6
UW CFA	4.2	2.7	4.0	8.1	7.9	4.6	14.5	14.7	8.3	7.7	7
X-11	4.9	4.4	3.2	4.6	5.4	4.1	6.7	8.7	4.6	5.2	4
MART	8.1	5.7	5.4	6.1	7.8	3.3	7.7	10.2	8.4	6.9	5
SUBM	12.7	7.1	5.6	10.1	13.2	4.9	14.9	12.0	6.3	9.7	8

### Panel B: Ranking of Prediction MAPEs

## Company

[illegible]

**Table 4.2**  
**Fuel Expense: Prediction MAPEs and Rankings**

**Panel A: Prediction MAPEs**

Company

Pred. Method	1	2	3	4	5	6	7	8	9	A V G	R A N K
OLS	29.4	12.7	11.8	6.9	16.4	5.8	45.9	11.3	7.2	16.4	6
CO	32.9	11.9	16	7.5	17.5	5.7	12.1	9.3	3.7	13.0	2
FD	14.2	8.1	4.2	4.9	14.1	5.2	12.9	10.3	4.2	8.7	1
UWR	20.1	11.8	5.6	17.6	22.8	5.3	42.3	24.7	7.4	17.5	7
UW CFA	13.4	7.4	5.4	4.5	26.5	8.2	46.9	23.9	6.4	15.8	5
X-11	14.3	24.5	7.3	7.0	17.6	10.1	28.8	14.7	16.4	15.6	4
Mart	23.8	9.4	5.6	8.7	21.1	6.5	18.3	22.5	6.3	13.6	3
Sub Mar	42.7	14.2	20.7	13.1	27.3	11.5	51.0	26.2	15.5	24.7	8

**Panel B: Ranking of Prediction MAPEs**

Company

Pre- diction Method	1	2	3	4	5	6	7	8	9	Worst Rank
OLS	6	6	6	3	2	4	6	3	5	6
CO	7	5	7	5	3	3	1	1	1	7
FD	2	2	1	2	1	1	2	2	2	2
UWR	4	4	4	8	6	2	5	7	6	8
UWCFA	1	1	2	1	7	6	7	6	4	7
X-11	3	8	5	4	4	7	4	4	8	8
MART	5	3	3	6	5	5	3	5	3	6
SUBM	8	7	8	7	8	8	8	8	7	8

First-differences achieved the lowest average MAPEs, followed by OLS regression Cochrane-Orcutt, and Census X-11, respectively.

First-differences also achieved the most consistent predictions. Panel B of Tables 4.1 and 4.2 indicate that the ranking, considering all nine companies in the sample, is never worse than fourth for revenues and second for fuel expense. In both cases, First-differences achieved the best "worst" ranking for both revenue and fuel expense.

The submartingale method provided the least accurate predictions of the eight methods tested. The average MAPE using the submartingale method is 9.7 percent for revenue and 24.7 percent for fuel expense. All of the statistical prediction methods performed better than the submartingale prediction method. However, the martingale method performed surprisingly well. The martingale method was more accurate, on average, than two of the statistical methods (Unit-weighted and Unit-weighted CFA) for revenues. The martingale method was more accurate, on average, than four statistical methods (OLS, Unit-Weighted, Unit-weighted CFA, and Census X-11) for fuel expense. Although the average MAPEs obtained by the martingale method were surprisingly low, the method exhibits inconsistent performance. This lack of consistency is evidenced by the worst rankings of the martingale method. The "worst" ranking was seventh (out of eight) for revenue and sixth for fuel expense.

The performance of the statistical methods were not adequate for production expense predictions. Panel A of Table 4.3 indicates that none of the statistical



**Table 4.3****Production Expense: Prediction MAPEs and Rankings****Panel A: Prediction MAPEs**

Company

Pred. Method	1	2	3	4	5	6	7	8	9	A V G	R A N K
OLS	14.0	16.2	20.0	10.6	26.9	10.2	17.5	7.7	28.0	16.8	6
CO	16.1	19.9	24.7	11.3	45.1	1.5	17.0	8.6	37.4	21.3	8
FD	11.9	21.4	13.5	11.5	22.8	10.3	16.1	7.7	21.6	15.2	5
UWR	18.7	11.7	14.4	13.5	15.6	9.0	17.6	10.5	8.2	13.2	4
UW CFA	11.5	11.0	15.5	11.2	19.4	7.6	16.1	9.7	8.8	12.3	2
X-11	15.0	9.1	11.2	7.6	19.1	9.2	19.8	5.0	16.6	12.5	3
Mart	6.7	11.2	11.1	8.6	27.5	9.4	10.0	7.6	8.0	11.1	1
Sub Mart	12.8	18.0	43.3	10.9	38.4	19.2	24.4	8.1	15.8	21.2	7

**Panel B: Ranking of Prediction MAPEs**

Company

Pre-diction Method	1	2	3	4	5	6	7	8	9	Worst Rank
OLS	5	5	6	3	5	5	5	4	7	7
CO	7	7	7	6	8	7	4	6	8	8
FD	3	8	3	7	4	6	2	3	6	7
UWR	8	4	4	8	1	2	6	8	2	8
UWCFA	2	2	5	5	3	1	3	7	3	7
X-11	6	1	2	1	2	3	7	1	5	7
Mart	1	3	1	2	6	4	1	2	1	6
Sub Mart	4	6	8	4	7	8	8	5	4	8

methods performed as well as the martingale method. The average MAPE for the martingale method was 11.1 percent. The next best method was Unit-weighted CFA, which achieved an average MAPE of 12.3 percent, followed by Census X-11 at 12.5 percent. None of the statistical methods performed very well. For example, First-differences, which was the best prediction method for revenues and fuel expense, did not perform as well as the martingale method. A t-test of the mean difference between the martingale and First-differences predictions indicates that the martingale method predicted significantly better than First-differences ( $t = 6.16$ ). The lack of performance of the statistical methods was further evidenced by the lack of consistency of the performance of all of the statistical methods indicated in Panel B of Table 4.3. The best "worst" rank for all five statistical methods was never better than seventh.

#### **4.1.3 Implications**

There are two primary implications of the results just presented. The first is that some accounts are not suited for SAP predictions. The second implication is that First-differences was found to be the most accurate prediction method and should therefore be strongly considered as the method of choice for auditors using SAPs. These implications are discussed in greater detail in the next two paragraphs.

The results indicate that statistical prediction methods are not appropriate for all account balances. The account balances that were selected for prediction in the current study were selected based on the potential of reducing total audit effort. However, the statistical predictions for production expense were not accurate enough

to justify their implementation. None of the statistical methods performed as well as the "naive" martingale model. The inclusion of a wide array of both financial and nonfinancial information into the prediction models actually performed worse than the martingale method. The martingale method is a naive prediction method which simply predicts the current month's account balance using the account balance from last year's monthly balance.

The First-differences method was found to be the most accurate prediction method. In addition, First-differences exhibited the most consistent performance compared to other methods. The use of this method is, therefore, recommended as the method of choice for auditors employing statistical prediction methods.

Another interesting implication was the performance of the regression methods, compared to the Census X-11 method. The results of the current study partially contradict the findings of a prior study (Wheeler and Pany, 1990). Wheeler and Pany (1990) asserted that the X-11 method performed better than regression. However, their study failed to incorporate information beyond that readily available in the financial statements. Therefore, it was unclear whether regression was inferior to X-11 in their study because of 1) incomplete information or 2) the fact that X-11 is, in fact, a superior prediction method. The current study indicates that the inclusion of other, nonfinancial information in the current study lead to more accurate predictions for regression. The performance of the regression methods was significantly better than X-11. First-differences and Cochrane-Orcutt provided more accurate predictions than Census X-11. The superiority of regression methods over

Census X-11 has more intuitive appeal than the findings of Wheeler and Pany (1990). The regression methods incorporate more information to generate predictions than the X-11 method. Intuition suggests that methods which incorporate more information in the predictions should predict better than methods which incorporate less information. The results of the current study demonstrate that the methods that incorporate the most information into the prediction models yielded the most accurate predictions.

The surprisingly low average MAPE achieved by the martingale method is, perhaps, due to the consistency of utilities exhibited over time. The martingale method predicts that the current month's account balance will be equal to the account balance from the same month in the preceding year. However, the performance of the martingale method is not as accurate for adjacent periods with very different weather conditions. For example, an exceptionally cold winter followed by a mild winter will lead to inaccurate predictions using the martingale method. This was evidenced by the inconsistency of the predictions for some firms. Even though the average MAPEs were relatively low, the method is not as useful as the average MAPE indicates due to its inconsistent performance.

#### **4.2 Simulation Analysis**

The preceding section compared the performance of the eight prediction methods by comparing their average MAPEs using individual monthly predictions. Prediction accuracy is only one way of measuring the performance of the eight methods. Another means of measuring the performance of the prediction methods

is to evaluate their ability to properly signal errors. This section contains the results of a simulation analysis in which errors were artificially seeded into the account balances of interest. The performance of each method in properly signalling the presence or absence of errors was evaluated.

The preceding section indicated that revenue and fuel expense were suitable accounts for SAP predictions. None of the statistical prediction methods were adequate in predicting production expenses. Therefore, the simulation analysis does not include results for production expenses.

The preceding section indicates that First-differences provides the most accurate predictions of revenue and fuel expense. The other four regression methods were not as accurate as First-differences. Therefore, the simulation analysis only compares the performance of First-differences, Census X-11, and the martingale method.

Similar to a prior study (Wheeler and Pany, 1990) "best-case" conditions were imposed. The best-case conditions imposed in Wheeler and Pany (1990) were that only single-industry companies were included in the sample. Furthermore, their study used four methods of computing materiality, and incorporated the largest of the four as the materiality measure. Three additional best-case conditions were used in the current study that were not imposed in the Wheeler and Pany (1990) study. First, monthly data are used instead of quarterly data. Second, both financial and nonfinancial data were used in the current study. Wheeler and Pany (1990) include only information readily available in the financial statements. Third, the best

prediction method was selected from a field of six statistical methods, instead of two. Inclusion of best case conditions provided a setting in which signalling capabilities were maximized.

To summarize, the simulation analysis compares the performance of First-differences, Census X-11 and the martingale method in accurately signalling material errors. The simulation analysis incorporates monthly predictions for revenues and fuel expense, but does not include production expense predictions. The simulation analysis provides additional insight into the performance of the prediction methods because it tests the ability of the prediction methods to properly identify the presence or absence of material errors that have been artificially seeded into the account balances of interest. Best-case conditions were imposed to maximize the signalling performance of the prediction methods.

This section is divided into three subsections. Subsection 4.2.1 summarizes the procedures performed in the simulation analysis. Subsection 4.2.2 presents the results of the simulation analysis, and Subsection 4.3.3 discusses the implications and importance of the results.

#### **4.2.1 Simulation procedures**

This subsection briefly summarizes the procedures used in the simulation analysis. The following paragraphs describe the error seeding procedures, the investigation rules and the methods for computing materiality that were used in the current study. The artificial seeding of errors, the investigation rules, and the materiality definitions used in the current study parallel those used in prior studies

(Wheeler and Pany, 1990; Kinney, 1987; Loebbecke and Steinbart, 1987). Using procedures parallel to those of prior studies provides a basis for comparison.

The simulation analysis was conducted in two phases. In the first phase, no errors were seeded into the account balances of interest. If an investigation rule signalled an investigation when no error was seeded, then a type I error has occurred. In the second phase, material errors of three magnitudes were seeded into the account balances of interest. If an investigation rule indicates that no error is present, then a type II error has occurred. The incidence of type I and type II errors was considered independently for each of the 12 monthly predictions from the prediction period.

In the current study, two investigation rules were incorporated: the percentage change rule and the statistical rule. The percentage change rule signals an investigation when the predicted account balance,  $\hat{y}_t$ , differs from the recorded account,  $y_t$ , by more than a critical percentage (CP) set by the auditor (5, 10, and 15 percent in the current study). Therefore, an investigation would take place if  $(y_t - \hat{y}_t)/y_t > CP$ .

The statistical rule signals an investigation when the standardized difference between the recorded and predicted account balance exceeds the critical Z value, which is based on the auditor's specified risk level,  $\alpha$ . In the current study,  $\alpha = .10$ ,  $.33$  and  $.5$ . An investigation would take place if  $(y_t - \hat{y}_t)/s_y > Z_{1-\alpha}$ , where  $s_y$  is the base period standard deviation of the series,  $y_t$ .

The selection of simulation rule introduces an inevitable trade-off between type I and type II errors. Some investigation rules lead to numerous investigations which lead to a high number of type I errors; however, the same simulation rule would lead to relatively few type II errors. As the investigation rule is relaxed, there will inevitably be less type I errors, but more type II errors. The parameters of the investigation rules (5, 10 and 15 percent for the percentage change rule, and .10, .33 and .5 for the statistical rule) were selected to allow the trade-off between type I and type II errors to be evident. This allowed an examination of error signalling across a wide range of investigation rule possibilities.

The current study incorporated four methods of computing materiality. The methods, all used in prior studies, are as follows:

- "audit gauge" [ $1.6 \times (\text{greater of total assets or revenues})^{2/3}$ ] (Elliot, 1983)<sup>1</sup>.
- 10 percent of net income (Holstrum and Messier, 1982).
- 10 percent of average earnings over a three-year period (Kinney, 1979).
- .5 percent of revenues (Wheeler and Pany, 1990).

Each of these methods were computed for all nine companies in the sample. The largest amount computed for each company was used as the definition of materiality to provide best case conditions for signaling errors.

Errors of three magnitudes were seeded into the account balances of interest.

The three magnitudes of material errors were as follows:

---

<sup>1</sup>This procedure for computing materiality was developed by KPMG Peat Marwick.



$M^*$  = Annual error seed condition

$M^*/4$  = Quarterly error seed condition

$M^*/12$  = Monthly error seed condition,

where  $M^*$  is the largest of the four materiality definitions computed for each company.

#### **4.2.2 Simulation results**

This subsection contains the results of the simulation analysis. The ability of the best prediction methods to properly signal the presence or absence of varying magnitudes of material errors was evaluated.

Table 4.4 presents the results for the annual material error seed condition. If the investigation rule signalled that an error was present when no error was seeded into the account balance, then a type I error occurred. If the investigation rule failed to signal that a material error was present when an account balance was seeded with error, then type II error occurred. Table 4.4 presents the results for the annual material error seed condition. The table presents the incidence of both type I and type II errors. The adjusted sum of type I and type II errors is also presented for the best regression method (First-differences), Census X-11, and the martingale method.<sup>2</sup>

---

<sup>2</sup>In some cases, it was possible for a type I and type II error to occur for the same observation. Since this would not be possible in an actual audit, double counting was not allowed in the simulation. The adjusted sum counts only one of the two errors in the event that both a type I and type II error occurred for a given observation.

**Table 4.4**  
**Simulation Results: Annual Error Seed Condition**

**Panel A: Regression Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	41.67%	1.39%	41.67%
10% Change	18.52%	3.24%	20.37%
15% Change	10.65%	20.83%	30.09%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	1.85%	28.7%	30.56%
Alpha = .33	7.87%	9.73%	17.13%
Alpha = .50	22.21%	6.95%	27.32%
<b>Average</b>	<b>17.13%</b>	<b>11.81%</b>	<b>27.86%</b>

**Panel B: Census X-11 Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	65.28%	.93%	65.28%
10% Change	31.02%	4.17%	33.34%
15% Change	22.22%	14.13%	38.89%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	12.5%	31.95%	43.52%
Alpha = .33	27.32%	10.65%	36.11%
Alpha = .50	35.65%	6.02%	40.28%
<b>Average</b>	<b>32.33%</b>	<b>11.81%</b>	<b>42.90%</b>

Table 4.4 (cont'd)

**Panel C: Martingale Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
Simple Invest. Rules:			
5% Change	63.43%	3.71%	63.43%
10% Change	38.89%	7.41%	40.28%
15% Change	17.60%	14.36%	30.10%
Statistical Invest. Rules:			
Alpha = .10	8.80%	26.39%	33.36%
Alpha = .33	25.93%	12.50%	33.80%
Alpha = .50	43.52%	6.95%	46.76%
Average	33.03%	11.88%	41.28%

Table 4.4 indicates the superiority of regression over the time series methods. The average adjusted sum for regression, over all six investigation rules, is 27.85 percent. The adjusted sum for X-11 and the martingale method are 42.90 percent and 41.28 percent, respectively. Regression does a better job of properly signalling the presence or absence of material errors than either X-11 or the martingale method.

Regression performs especially well when the 10 percent change rule and the statistical rule ( $\alpha = .33$ ) are used. The adjusted sum for regression, using the 10 percent change rule is 20.37 percent. The adjusted sum for regression, using the statistical rule ( $\alpha = .33$ ) is 17.13 percent.

The 10 percent change rule and the statistical rule ( $\alpha = .33$ ) adequately control the incidence of type II error, while maintaining a modest type I error rate. The type II error rate is 3.24 percent for the 10 percent change rule and 9.73 percent for the statistical rule ( $\alpha = .33$ ). In the current study, type I and type II errors are weighted equally. However, auditors are much more concerned about type II errors than type I errors. The potential loss to the auditor is much greater for failing to identify a material error than for investigating an account balance further when no errors are present.

The simulation analysis also examined the signalling capabilities of the prediction methods when material errors of smaller magnitudes were seeded into the account balances of interest. Tables 4.5 and 4.6 contain the results for the quarterly ( $M^*/4$ ) and monthly ( $M^*/12$ ) error seed conditions. As expected, the signalling

**Table 4.5**  
**Simulation Results: Quarterly Error Seed Condition**

**Panel A: Regression Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
Simple Invest. Rules:			
5% Change	41.67%	27.78%	62.96%
10% Change	18.52%	52.32%	67.13%
15% Change	10.65%	55.09%	61.57%
Statistical Invest. Rules:			
Alpha = .10	1.85%	81.95%	82.87%
Alpha = .33	7.87%	61.57%	66.67%
Alpha = .50	22.21%	48.15%	64.35%
Average	17.13%	54.48%	67.59%

**Panel B: Census X-11 Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
Simple Invest. Rules:			
5% Change	65.28%	27.32%	77.32%
10% Change	31.02%	47.23%	69.91%
15% Change	22.22%	52.78%	67.13%
Statistical Invest. Rules:			
Alpha = .10	12.50%	79.17%	87.5%
Alpha = .33	27.32%	60.19%	77.31%
Alpha = .50	35.65%	46.76%	71.30%
Average	32.33%	52.24%	75.08%

Table 4.5 (cont'd)

**Panel C: Martingale Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	63.43%	17.60%	72.22%
10% Change	38.89%	38.89%	70.38%
15% Change	17.60%	46.3%	59.73%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	8.80%	79.17%	84.26%
Alpha = .33	25.93%	49.54%	67.13%
Alpha = .50	43.52%	32.87%	67.13%
<b>Average</b>	<b>33.025%</b>	<b>44.06%</b>	<b>70.14%</b>

**Table 4.6**  
**Simulation Results: Monthly Error Seed Condition**

**Panel A: Regression Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	41.67%	52.78%	83.33%
10% Change	18.06%	76.85%	91.67%
15% Change	10.65%	86.58%	94.91%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	1.85%	97.22%	98.15%
Alpha = .33	7.87%	50.47%	57.87%
Alpha = .50	22.21%	70.37%	86.58%
<b>Average</b>	<b>17.05%</b>	<b>72.38%</b>	<b>85.42%</b>

**Panel B: Census X-11 Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	65.28%	39.82%	92.59%
10% Change	31.02%	65.28%	90.74%
15% Change	22.22%	79.63%	96.76%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	12.50%	85.65%	95.37%
Alpha = .33	27.32%	74.54%	95.84%
Alpha = .50	35.65%	59.72%	91.21%
<b>Average</b>	<b>32.33%</b>	<b>67.44%</b>	<b>93.75%</b>

Table 4.6 (cont'd)

**Panel C: Martingale Results**

Investigation Rule	Percentage Type I Errors	Percentage Type II Errors	Adjusted Sum
<b>Simple Invest. Rules:</b>			
5% Change	63.43%	31.48%	87.04%
10% Change	38.89%	55.10%	88.89%
15% Change	17.60%	73.15%	88.43%
<b>Statistical Invest. Rules:</b>			
Alpha = .10	8.80%	88.89%	97.22%
Alpha = .33	25.93%	70.37%	90.28%
Alpha = .50	43.52%	52.32%	90.74%
<b>Average</b>	<b>33.03%</b>	<b>61.88%</b>	<b>90.43%</b>



capabilities are not as good when the magnitude of material error is reduced. As with the annual error seed condition, regression performs better than X-11 or the martingale method. However, the signalling accuracy is greatly reduced for the smaller error seed conditions. The average adjusted sum for regression was 27.86 percent in the annual error seed condition. However, the average adjusted sum is 67.59 percent for the quarterly error seed condition and 85.42 percent for the monthly error seed condition.

#### **4.2.3 Implications**

The prediction methods employed in the current study do a reasonably good job of detecting material errors equal to annual materiality. However, the methods are much less reliable at detecting errors which are material to an individual quarter or month.

At first glance, the results of the simulation analysis may appear disappointing. However, when considered in the context of the audit, the results are more promising. First of all, analytical procedures are not performed in isolation. They are performed along with other audit procedures. Further research should examine the combined levels of assurance obtained by combining SAPs with other audit tests.

Furthermore, SAPs could be used in the planning phases of an audit to identify specific monthly account balances which warrant additional testing. While the procedures in and of themselves may not be reliable enough to justify the elimination of other substantive tests, use of monthly data for may improve audit

efficiency by signalling the specific time periods which are most likely to contain material errors.

Statistical analytical procedures signalled large material errors with a relatively high degree of precision. SAPs performed poorly when material errors of a smaller magnitude were seeded into the account balances of interest. The poor performance of SAPs in signalling material errors of smaller magnitude may not be as serious as the percentages indicate because in the current study, the incidence of type I and type II errors was considered independently for each month. The results are more promising when the monthly predictions are annualized, instead of examining each monthly prediction independently. The results of the annualized predictions are reported in the next section.

#### **4.3 Annualized Prediction Errors as a Percentage of Materiality**

One of the limitations of Sections 4.1 and 4.2 is that monthly predictions were examined independently. The MAPEs in Section 4.1 were measured for individual monthly predictions. Likewise, the simulation analysis in Section 4.2 measured the ability of the prediction methods to properly signal errors seeded into monthly account balances. Examining each monthly prediction independently is conservative and may understate the accuracy of the prediction methods.

It is also useful to evaluate the prediction methods based on annualized predictions. The annualized approach is, more consistent with the approach that would probably be taken by the auditor. Rather than examine each month independently, the auditor would probably combine the monthly predictions into an

annual balance by summing the 12 individual monthly predictions. Accordingly, in the current study an additional analysis was performed to examine the accuracy of the annualized account balance predictions.

The remainder of this section is divided into three subsections. Subsection 4.3.1 describes the procedures used to annualize and evaluate the predictions. Subsection 4.3.2 contains the results. Subsection 4.3.2 contains the implications of the results.

#### **4.3.1 Procedures to annualize predictions**

The twelve monthly predictions were summed for each method. The difference between the annual prediction and the annual recorded balance is the prediction error for the year. As a benchmark, prediction error for each company was divided by materiality. Percentages smaller than 100 percent indicate that the annual predictions were within materiality. Percentages greater than 100 percent indicate that the annual predictions error is greater than materiality.

#### **4.3.2 Results of annualized predictions**

Table 4.7 presents the prediction errors as a percentage of materiality for First-differences, Census X-11, and the martingale method. The results are presented for each of the nine companies in the sample. Panel A contains the results for revenues, and Panel B contains the results for fuel expense. The results once again indicate the superiority of regression over X-11 and the martingale method. Prediction error as a percentage of materiality for regression is 33.59 percent for revenue and 23.05 percent for fuel expense. The worst (highest) percentage for all

**Table 4.7**  
**Annualized Prediction Results**

**Panel A: Annualized Revenue Predictions as Percentage of Materiality**

<b>Company</b>	<b>First-differences Error/Materiality</b>	<b>Census X-11 Error/Materiality</b>	<b>Martingale Error/Materiality</b>
1	4.35%	128.38%	344.72%
2	22.79%	149.03%	349.93%
3	4.57%	62.29%	298.48%
4	48.10%	40.78%	108.30%
5	66.72%	60.41%	177.98%
6	29.98%	223.83%	66.92%
7	7.17%	439.95%	194.98%
8	67.89%	198.24%	301.46%
9	50.79%	127.7%	398.5%
<b>Avg.</b>	<b>33.59%</b>	<b>158.96%</b>	<b>249.03%</b>

**Table 4.7 (cont'd)****Panel B: Annualized Fuel Expense Predictions as Percentage of Materiality**

<b>Company</b>	<b>First-differences Error/Materiality</b>	<b>Census X-11 Error/Materiality</b>	<b>Martingale Error/Materiality</b>
1	9.52%	107.91%	216.25%
2	49.59%	594.32%	235.15%
3	11.12%	78.38%	69.99%
4	32.12%	16.00%	24.58%
5	12.94%	65.17%	168.76%
6	2.69%	52.23%	57.75%
7	53.44%	378.92%	45.85%
8	46.17%	78.82%	323.78%
9	19.82%	150.73%	64.57%
<b>Avg.</b>	<b>23.05%</b>	<b>169.16%</b>	<b>134.07%</b>

nine firms is 67.89 percent. Thus, the annualized prediction error was never worse than 68 percent of the annual materiality threshold. For regression, the predictions were within materiality without exception.

Prediction errors as a percentage of materiality using Census X-11 were significantly higher than regression. Average prediction error as a percentage of materiality using X-11 is 158.96 percent for revenue and 169.16 percent for fuel expense. Ten of the 18 percentages were greater than 100 percent. Thus, 56 percent of the X-11 predictions were materially different than the recorded account balance.

Prediction errors as a percentage of materiality using the martingale method were also significantly higher than regression. Average prediction error as a percentage of materiality using the martingale method is 249.03 percent for revenues and 134.07 percent for fuel expense. Twelve of the 18 percentages were greater than 100 percent. Thus, 67 percent of the martingale predictions were materially different than the recorded account balance.

#### **4.3.3 Implications of the results**

The results of presented in this section were consistent with the results contained in Sections 4.1 and 4.2. The results of the annualized predictions, once again, demonstrate the superiority of First-differences over the other prediction methods. First-differences was significantly more accurate on average than Census X-11 and the martingale method. First-differences was also more consistent than X-11 and the martingale method in generating predictions which were within materiality limits.

#### **4.4 Summary**

The purpose of this chapter was to compare the performance of eight alternative prediction methods. These methods are as follows:

1. OLS Regression
2. Cochrane-Orcutt Regression
3. First-differences Regression
4. Unit-Weighted Regression
5. Unit-Weighted CFA Regression
6. Census X-11
7. Martingale
8. Submartingale

The performance of these eight methods was evaluated in three ways. First, the average mean absolute percentage errors (MAPEs) of each method were computed based on individual monthly predictions from a hold-out period. Second, the performance of the methods was evaluated using a simulation analysis. Errors were artificially seeded into the account balances of interest. Performance was evaluated based in the ability of each prediction method to properly identify the presence or absence of material errors. Third, the performance of the methods was evaluated by evaluating the accuracy of annualized predictions. The results of each of the three analyses were presented in Sections 4.1 through 4.3, respectively.

The results in Sections 4.1 through 4.3 were consistent in identifying First-differences as the most accurate prediction method. First-differences achieved the lowest average MAPEs using monthly predictions. First-differences was more accurate than other methods in properly signaling the presence and absence of material errors which were artificially seeded into the account balances of interest. Finally, First-differences was more accurate than other methods using annualized

predictions of the account balances of interest to evaluate the alternative prediction methods. The superior performance of regression partially contradicts the findings of a prior study (Wheeler and Pany, 1990) which indicates that Census X-11 performed better than regression. However, this finding is not surprising since the current study incorporated a variety of financial and nonfinancial predictor variables. Inclusion of such variables is expected to improve the relative performance of regression over Census X-11.

There were two other noteworthy findings in the current chapter. The first is that some accounts are not suited for statistical prediction methods. The second is that prediction methods do a better job of signalling large material errors than small material errors. Each of these findings is discussed in the next two paragraphs, respectively.

Section 4.1 demonstrated that some accounts are not suited for statistical prediction methods. The results of the current study indicate that the production expense account is not suited for statistical methods. The basis for this conclusion is that a naive model (the martingale method) achieved more accurate predictions than any of the statistical prediction methods for this account. The martingale method is much less costly to employ than statistical methods because it does not require that predictor variables be collected, nor does it require statistical analysis. Additional research is needed to determine which accounts are suited to statistical prediction methods.



The simulation analysis indicated that all of the prediction methods do a better job of signalling larger errors. The prediction methods properly signalled errors with much greater accuracy when the errors were large. However, the results were much less promising when smaller errors were introduced into the account balances of interest. Future research is needed to evaluate the performance of statistical analytical procedures with a number of varying distributions of errors.

This chapter contained the results of the first objective of the current study. The first objective was to evaluate the performance of various alternative prediction methods. The second objective of the current study is to evaluate the consistency of statistical analytical procedures. The third objective is to evaluate the performance of pooled prediction models. The fourth objective is to test the relative performance of monthly and quarterly prediction models. The next chapter contains the results of objectives two through four.

## **Chapter V**

### **5. RESULTS: SAP CONSISTENCY, ALTERNATIVE MODELS**

In the current study statistical analytical procedure (SAP) models were developed for a sample of nine electric utilities. The models were constructed using financial and nonfinancial information. The use of SAP models may provide auditors with a lower cost means of obtaining audit assurance than other types of substantive tests. The specific objectives of the current study were to 1) to compare the performance of a number of alternative SAP prediction methods, 2) assess the consistency of SAP models, 3) to assess the performance of pooled SAP models, and 4) to compare the performance of quarterly and monthly prediction models.

This chapter contains the results of objectives two, three, and four. The results of the first objective were presented in Chapter Four. The discussion in this chapter is organized according to the three objectives examined in this chapter. Section 5.1 discusses the consistency of the SAP models. Section 5.2 discusses the performance of the pooled models. Section 5.3 compares the performance of monthly and quarterly prediction models. Section 5.4 contains a summary of the current chapter.

#### **5.1 Model Consistency**

The first objective of the current study was to evaluate the consistency of SAP models. Consistency is an extremely important attribute of audit tests. SAP prediction models must perform consistently for multiple clients if they are to be

useful to auditors. Therefore, an important objective of the current study was to demonstrate the consistency (or lack thereof) of SAP models.

The consistency of SAPs is examined along three important dimensions: 1) the consistency of predictions, 2) the consistency of specific financial and nonfinancial predictor variables, 3) the consistency of violations of the assumptions of SAPs. To adequately address the consistency of SAP models along these dimensions, two elements needed to be present. First, multiple firms must have been included in the sample. Second, both financial and nonfinancial data had to be included in the SAP models. The following paragraphs explain the importance of these two attributes of the current study.

Most prior research studies have not evaluated the performance of SAPs across multiple companies. Many prior SAP studies are case studies (Wild, 1987; Neter, 1981; Akresh and Wallace, 1981; Albrecht and McKeown, 1976). Therefore, little is known about the consistency of the performance of SAPs across multiple companies. In order to assess the robustness of SAPs, data must be collected from multiple companies, as was done in the current study.

Studies which do collect data from multiple companies suffer the limitation of inadequate data sets (Wheeler and Pany, 1990; Kinney, 1978). These studies only include predictor variables which are readily available in the financial statements. Other nonfinancial predictors are ignored. Therefore, little is known regarding the consistency of the performance of financial and nonfinancial predictor variables in generating accurate predictions.

To summarize, prior studies could not address the consistency of SAP models adequately either because they were case studies, or they did not incorporate both financial and nonfinancial data into the prediction models. The current study overcomes the limitations of many prior studies by including data from multiple firms, and by incorporating both financial and nonfinancial predictor variables in the analysis.

In the current study, the consistency of SAP models was evaluated in three ways: 1) by evaluating the consistency of SAP predictions, 2) by examining the consistency of individual variables in predicting the account balances of interest, and 3) by performing diagnostic tests of the assumptions of the prediction models. The results of each are presented in Subsections 5.1.1 through 5.1.3.

#### **5.1.1 Consistency of SAP model predictions**

The primary reason for examining consistency of SAP predictions is a practical one. Practitioners are unlikely to adopt new methods unless such methods can be leveraged on multiple clients. Adopting new audit procedures is a costly undertaking. Therefore, it is important that the performance of SAP prediction models be examined for multiple firms. Prior research studies (Wild, 1987; Akresh and Wallace, 1981; Neter, 1981; Albrecht and McKeown, 1976) have examined the performance of SAP methods on individual company applications. It is unclear whether their studies are indicative of isolated successes with SAPs or whether the procedures are potentially useful on most or all firms in the respective industries examined.

In this subsection, the prediction performance of SAPs is examined across nine electric utility companies. Due to resource limitations, only a small number of companies could be included in the sample. Therefore, companies were selected to capture the differences between companies in the industry. Companies from three geographic regions (the West, Mid-West, and South) were included. These companies had varying mixes of generating facilities (i.e. nuclear, coal, gas, oil and hydroelectric facilities). The companies also varied in size. Some of the largest utilities in the United States were included in the sample along with medium and small sized utilities. Section 3.1.4 contains detailed demographic information regarding the characteristics of sample firms.

The consistency of predictions was evaluated by examining the best prediction method. The other methods were not analyzed because an auditor employing SAPS would only use one method. The assumption is that the auditor would use the best method available. The results presented in Chapter Four indicated that First-differences is the most accurate prediction method. Accordingly, the consistency of the First-differences models were examined.

The consistency of predictions was evaluated using four different measurements: 1) MAPEs from the model-building period, 2) F-statistics, 3) MAPEs from the prediction period, and 4) Annualized prediction error as a percentage of materiality. Items one and two are goodness of fit measures achieved in the model-building period. Items three and four are measures of prediction accuracy in a hold-out period.

Table 5.1 presents the regressions results of the revenue predictions for each of the nine companies in the sample. The two goodness of fit measures (the model building MAPE and the F-statistic) are presented. The MAPEs attained in the prediction period are also presented. Predictions from the prediction (hold-out) period provide an additional indication of the consistency of model predictions. Many prior studies do not examine model performance in a hold-out period.

The revenue results obtained in the current study were consistent with the results obtained in the Akresh and Wallace (1981) case study. This study indicated a high degree of prediction accuracy for revenue predictions. Akresh and Wallace reported an F-statistic of 154.48 for their revenue prediction model. The prediction accuracy of their model is somewhat inconclusive, however, because this case study did not present predictions using a hold-out period.

For revenue predictions, accuracy was consistent both in the model building period and in the prediction period. The model building MAPEs for all nine companies are all below seven percent. F-statistics, which are an indication of the overall fit of the model were all significant ( $p < .001$ ). The average F-statistic for the nine revenue models was 141.13. The average MAPE in the prediction period was 3.8 percent. All nine of the MAPEs in the prediction period were also below 7 percent. Another indication of the accuracy of the revenue models is the annualized prediction error divided by materiality. Prediction error never exceeded materiality for any of the nine firms. The average prediction error was 34 percent

**Table 5.1**  
**Consistency of SAP Models: Revenue**

Company	Model Building MAPE	F-Statistic	Prediction Period MAPE	Annualized Error/ Materiality
1	1.8%	100.91	5.0%	4.39%
2	1.7%	126.31	2.0%	22.79%
3	1.3%	207.28	1.5%	4.57%
4	1.6%	215.21	3.6%	48.10%
5	5.3%	79.58	6.4%	66.72%
6	1.1%	258.14	3.7%	29.98%
7	6.2%	50.16	5.2%	7.17%
8	1.7%	143.24	4.4%	67.89%
9	1.4%	89.32	2.6%	50.79%
Average	2.4%	141.13	3.8%	33.59%

of materiality. The revenue predictions were consistently very accurate for all nine firms.

Table 5.2 presents the regression results for fuel expense. The predictions for fuel expense were also promising, though somewhat less consistent. The average MAPE from the model-building period was 7.6 percent. The predictions for companies one and five indicate the inconsistency of some of the models. The F-statistics were smaller, though all were significant ( $p < .03$ ). The results from the prediction period were also somewhat inconsistent. Three of the prediction MAPEs were greater than 12 percent. However, all nine of the annualized prediction errors were within materiality.

Auditors are more likely to adopt SAP predictions for revenue models than for fuel expense. The revenue models exhibit a higher degree of consistency than the fuel expense models. Even though the fuel expense predictions were very accurate for some companies, they are not consistently accurate for all companies. Therefore, it is likely that auditors would be more hesitant to implement SAP models for fuel expense than SAP models for revenues.

The reason for the difference in prediction accuracy between revenue and fuel expense accounts is probably due the complexity of the cost function of electric utilities. The cost of producing electricity varies depending on a number of factors which would be difficult to measure and incorporate into a SAP model. For example, utility companies frequently interchange power at different times of day to minimize costs. The decision of whether Company A buys or sells to a Company B



**Table 5.2**  
**Consistency of SAP Models: Fuel Expense**

Company	Model Building MAPE	F-Statistic	Prediction Period MAPE	Annualized Error/ Materiality
1	14.3%	3.44	14.2%	9.52%
2	5.0%	8.63	4.9%	49.59%
3	2.7%	16.09	3.8%	11.12%
4	4.2%	88.62	4.9%	32.12%
5	14.2%	20.14	14.1%	12.94%
6	5.6%	10.74	5.2%	2.69%
7	8.5%	9.62	12.0%	23.44%
8	8.5%	11.88	8.8%	46.17%
9	5.4%	12.49	3.2%	19.82%
Average	7.6%	20.18	7.9%	23.05%

at a given time of day depends on a number of factors such as the demand for electricity, peak loads, and generating capacities. These factors vary from moment to moment, which makes it very difficult to capture the complexity in monthly SAP models. This may explain, why the fuel expense predictions were not as consistent as the revenue predictions.

### **5.1.2 Consistency of predictor variables**

One of the time consuming aspects of employing SAPs is data collection. Auditors must collect information and transform it to machine readable format. Furthermore, the identification of variables that are the best predictors for particular account balances may also be a costly activity. The auditor may have to collect, input and analyze many variables in order to identify the few variables which are consistently good predictors of a particular account balance.

Auditors would benefit by knowing, in advance, the variables which are consistently good predictors of a particular account balance. This information will allow the auditor to collect the information they need and not collect information which does not tn consistently good explanatory value.

Another important aspect of predictor variable consistency is the type of information used as predictor variables. A prior study indicates that auditors tend to use only information that is readily available in the financial statements as predictor variables (Biggs and Wild, 1984). In the current study, other nonfinancial information were also included in the prediction models. The predictions obtained when incorporating financial variables only was compared the predictions obtained

when both financial and nonfinancial predictor variables were included. This comparison provides an indication of the benefits of including nonfinancial predictor variables into SAP models.

Accordingly, in the current study, model consistency was evaluated by examining the consistency of individual predictor variables in improving the fit of the models. The consistency of predictor variables was evaluated in two ways: 1) by identifying variables which consistently improve prediction performance, and 2) by examining the incremental benefit of incorporating nonfinancial variables into the prediction models. The results of each are presented in Subsections 5.1.2.1 and 5.1.2.2 respectively.

#### **5.1.2.1 Variables consistently improving prediction accuracy**

This subsection provides evidence regarding the variables that were found to be consistently good predictors of the account balances of interest. As mentioned previously, one of the objectives of the current study was to identify robust predictor variables which could be implemented into SAP models for all companies in the selected industry. Utility company experts were interviewed to determine the set of variables that were collected for the current study. The analysis began with the entire set of variables. However, only those variables that were found to improve the fit of each of the models were retained.

##### **5.1.2.1.1 Results of significant predictor variables**

Table 5.3 presents significant predictor variables for revenue models by company. Three variables were found to be significant for nearly all nine companies'

revenue models. Kilowatt-hours (KWH) sold was highly significant for all nine firm models, as evidenced by the t-values presented in Table 5.3. At least one of the three Rate Factors was significant for eight firm models and Degree Days were significant for seven firm models. The remaining variables were significant for at least some firm models, though the degree of relationship between the other variables and revenue was not strong enough to warrant retaining the variables in the model for many of the firms.

Table 5.4 presents the significant predictor variables for fuel expense. Consistency strong predictor variables did not emerge as readily for fuel expense as for revenue. KWH sold was a relatively good predictor of fuel expense. It was significant for six of the nine companies. No other predictor proved significant for more than four of the nine companies.

#### **5.1.2.1.2 Implications**

The superior performance of the revenue predictions can be partially attributed to the strong degree of relationship between three independent variables (rates, kilowatt-hour production and degree days) and revenues. Likewise, the less consistent performance of the fuel expense and production expense models can be attributed to the fact that robust predictor variables did not emerge.

**Table 5.3**  
**Robust Predictor Variables: Revenue**

**T-statistics Associated With Revenue Predictor Variables**

Variable:	Company								
	1	2	3	4	5	6	7	8	9
Lagged Revenue	2.1		2.8		1.0				2.5
Fuel Expense	1.5		2.8	4.2	7.6		8.6		
Production Expense					-1.3	-3.2			-2.4
Residential Rate			4.1	2.0	-2.2		4.4		7.3
Commercial Rate	5.0	3.2			-1.5				
Industrial Rate	-1.3				1.9			5.6	-5.4
Heating Degree Days		3.6		-4.5	-3.6		-4.4	6.9	2.6
Cooling Degree Days		4.3		-2.7	-2.1	3.1	-1.0	4.7	2.8
KWH Generation								-7.6	-2.6
KWH Sold	8.4	10.1	7.5	7.3	3.9	9.2	9.4	24.1	19.1
Number Customers					2.4				-3.4
Unemployment					1.4	1.5	-2.4		
CPI					-3.0				3.4
Fuel CPI									
Budgeted Revenue	3.4				2.9	6.8			

**Table 5.4**  
**Robust Predictor Variables: Fuel Expense**

**T-statistics Associated with Fuel Expense Predictor Variables**

Variable:	Company								
	1	2	3	4	5	6	7	8	9
Revenue	-2.6		8.1		8.8		4.5		
Lagged Fuel Expense					-3.0		2.3	3.7	
Production Expense			-1.3		2.5		-1.4		
Heating Degree Days					2.2	-2.0	3.8		
Cooling Degree Days			1.7	5.9		-1.2	3.4		
KWH Generation		2.2	-6.1					18.3	1.2
KWH Sold	4.7	-1.5	3.7		-3.8	2.1			3.1
Number Customers			3.9				1.6	1.9	
Unemployment	-3.3	1.5	2.4						1.7
CPI	-4.2			-2.4			2.2	-8.5	
Fuel CPI		1.7						1.9	
Capacity Factor			-8.5	4.0			-4.7		
Load Factor			1.2						
Budgeted Fuel	3.7			6.2		4.1			

These results warrant some discussion regarding the production data which were included in the models. Not surprisingly, KWH sales was found to be the most robust predictor variable for revenues and fuel expense. Inclusion of production data significantly improved the accuracy of predictions for all three accounts. The inclusion of production data in other industry models may be even greater in other industries as described in the next two paragraphs. The next two paragraphs explain how inclusion of production data may improve the accuracy of revenue and expense models for other industries to an even greater degree than in the electric utilities industry.

For revenues, the expected benefit of including production data in the models was offset by the fact that the revenue per KWH varies a great deal. Large industrial customers pay substantially different rates than small residential and commercial customers. Rate information was included to compensate for these differences; however, inclusion of rate data did not completely control for this problem due to the complexity of the rate structures. The per KWH effect on revenue may also vary depending on the time of day, the peak demand of each customer. The use of production data for SAPS in other industries with less disparity in prices charged to customers may further improve the prediction performance achieved through including production data.

For fuel expense the benefit of including production data in the models is offset by the fact that the cost of KWH production varies depending on the types of facilities used to generate the power. Electricity generated from nuclear facilities is

typically produced at the lowest cost, followed by coal, gas and oil, respectively. Nuclear and coal fired plants are sometimes referred to as base-demand facilities, and gas and oil plants are sometimes referred to as peaking-demand facilities. Peaking facilities are more expensive to operate, but less expensive to construct than base-demand facilities. Base-demand facilities tend to be used to meet both normal and peak demands for electricity. Peaking facilities tend to be used to meet peak demands. Varying circumstances, such as unscheduled repairs or unexpected changes in demand, will cause the production mix between base and peaking facilities to change. Changes in the production mix probably caused the expense predictions to be less accurate. The prediction performance of including production data may be improved in industries in which the cost function is more constant.

Furthermore, the prediction models for electric utilities may be improved further by including more detailed information for firms with volatile revenue and cost functions. Future research should examine the potential benefits of including company specific predictor variables which would more accurately model the complexity of a specific company's revenue or cost function. Where possible, auditors should include company specific information, that more accurately reflects the complexity of the client's cost function, to improve the precision of the fuel and production expense models that were developed in the current study.

#### **5.1.2.2 Incremental benefit of nonfinancial predictor variables**

In the current study, considerable effort was expended to identify all of the variables that would be useful predictors of the account balances of interest. The



identification of variables included both financial and nonfinancial predictor variables. Prior studies have indicated that auditors tend to use only information readily available in the financial statements. The current study examines the incremental benefits of including other nonfinancial information in the predictions.

Prior research has indicated that auditors tend to use financial variables only in conducting analytical procedures (Biggs and Wild, 1984). One objective of the study is to determine the improvement in predictions that is possible by incorporating both financial and nonfinancial variables into the SAP models. This subsection compares the predictions obtained from financial variables only and the predictions obtained from both financial and nonfinancial variables.

Inclusion of nonfinancial information into SAP models may significantly improve prediction performance. Greater prediction accuracy may allow auditors to place greater reliance on SAP models and thereby allow them to reduce the amount of other, more expensive, procedures.

Some of the nonfinancial information is obtained from sources external to the firm and may therefore be more reliable than the financial information obtained from the client. For example, the degree day information collected in the current study can be obtained from the U.S. Weather Service. Use of information generated external to the firm may allow auditors to place increased reliance on SAP models than prediction models generated using information collected from the client. Increased reliance on SAP models may allow the auditor to reduce the amount of other expensive audit procedures.

The incremental benefit of including nonfinancial predictor variables was examined using a two step process. First, models were estimated using financial information only.<sup>1</sup> Second, the models were estimated again using both financial and nonfinancial predictor variables.

Figure 5.1 contains a list of financial predictor variables and nonfinancial predictor variables for each of the three account balances of interest. The performance of the two sets of models was examined by comparing the MAPEs from the prediction (hold-out) period. Comparing MAPEs from the prediction period provides a better indication of the incremental benefit of including nonfinancial predictors than using predictions from the model building period.

#### **5.1.2.2.1 Results**

Table 5.5 presents the results of a comparison of two types of models. The first type of models were constructed exclusively with financial information from the financial statements (financial information). The second type of models were constructed using both financial and nonfinancial information. The comparison provides an indication of the incremental benefit of including nonfinancial variables to the analysis.

Panel A of Table 5.5 contains the mean absolute percentage errors (MAPEs) for the revenue predictions for all nine companies. The MAPEs are lower for all nine firms when both financial and nonfinancial information were included in the

---

<sup>1</sup>In the current study, the definition of "financial information" is information that is readily available in the financial statements.

**Figure 5.1**  
**Predictor Variables**

**Revenue Predictors:**

**Financial Statement  
Predictor Variables:**

Lagged Revenues  
Fuel Expense  
Production Expense

**Nonfinancial Statement  
Predictor Variables:**

Residential rate factor  
Commercial rate factor  
Industrial rate factor  
Heating degree days  
Cooling degree days  
KWH generated  
KWH sold  
Number of customers  
Unemployment  
CPI  
Fuel CPI  
Budgeted revenue

**Fuel Expense Predictors:**

**Financial Statement  
Predictor Variables:**

Revenues  
Lagged Fuel Expenses  
Production Expenses

**Nonfinancial Statement  
Predictor Variables:**

Heating degree days  
Cooling degree days  
KWH generated  
KWH sold  
Number of customers  
Unemployment rate  
CPI  
Fuel-CPI  
Capacity factor  
Load factor  
Budgeted production expense  
Geographic region\*  
Type of generating facilities\*

\* only incorporated in pooled models

**Figure 5.1 (cont'd)****Production Expense Predictors:****Financial Statement****Predictor Variables:**

Revenues

Fuel Expenses

Lagged Production Expenses

**Nonfinancial Statement****Predictor Variables:**

Heating degree days

Cooling degree days

KWH generated

KWH sold

Number of customers

Unemployment rate

CPI

Fuel-CPI

Capacity factor

Load factor

Budgeted production expense

Geographic region\*

Type of generating facilities\*

\* only incorporated in pooled models

**Table 5.5**  
**Incremental Benefit of Nonfinancial Information**

**Panel A: MAPEs for Revenue Predictions**

Company	Financial Information Only	Financial and Nonfinancial Information	Difference
1	6.3%	5.0%	1.3%
2	5.2%	2.0%	3.2%
3	3.6%	1.5%	2.1%
4	6.2%	3.6%	2.6%
5	6.8%	6.4%	.4%
6	5.1%	3.7%	1.4%
7	9.1%	5.2%	3.9%
8	7.6%	4.4%	3.2%
9	5.2%	2.6%	2.6%
Average	6.1%	3.8%	2.3%

**Table 5.5 (cont'd)****Panel B: MAPEs for Fuel Expense Predictions**

<b>Company</b>	<b>Financial Information Only</b>	<b>Financial and Nonfinancial Information</b>	<b>Difference</b>
<b>1</b>	15.1%	14.2%	.9%
<b>2</b>	4.9%	4.9%	.0%
<b>3</b>	5.0%	3.8%	1.2%
<b>4</b>	6.9%	4.9%	2.0%
<b>5</b>	15.2%	14.1%	1.1%
<b>6</b>	6.9%	5.2%	1.7%
<b>7</b>	12.5%	12.0%	.5%
<b>8</b>	10.2%	8.8%	1.4%
<b>9</b>	4.1%	3.2%	.9%
<b>Average</b>	9.0%	7.9%	1.1%

models. The difference is significant ( $p < .01$ ) using a t-test comparing the average MAPEs. The average MAPE using financial information only is 6.7%. However, when nonfinancial information was added to the prediction models, the average MAPE decreased to 3.8%, which represents a 43% reduction of MAPE.

Panel B of Table 5.5 contains the MAPEs for the fuel expense predictions for all nine companies. The inclusion of nonfinancial information improved predictions for seven of the nine companies for the fuel expense predictions. The average MAPE obtained using financial information only was 11.5%. By including nonfinancial information to the prediction models, the average MAPE decreased to 8.7%, which represents a 24% reduction in the average MAPE.

Inclusion of both financial and nonfinancial predictor variables significantly improves the predictions obtained by including financial predictor variables only. This result held for all nine company's revenue predictions and for 7 of the 9 company's fuel expense predictions.

#### **5.1.2.2.2 Implications**

SAP predictions financial predictor variables are significantly improved by including nonfinancial predictor variables. Existing research indicates that auditors tend not to use nonfinancial predictor variables in conducting analytical procedures. The foregoing analysis indicates that inclusion of nonfinancial variables in SAP prediction models may allow auditors to place increased reliance on AP models, thereby allowing a reduction in other more expensive audit procedures. Auditors should therefore consider including nonfinancial predictor variables into SAP models.

### **5.1.3 Diagnostic testing**

This section reports the results of the diagnostic testing performed in the current study. Diagnostic tests were performed to evaluate the effect of violations of the assumptions of regression. One prior study (Elliot, 1976) speculated that statistical prediction methods would be of little value to auditors in conducting analytical procedures due to statistical problems such as autocorrelation, heteroscedasticity, multicollinearity, normality and continuity. Prior studies have not examined the effects of these statistical problems. It is, therefore, unclear whether these statistical problems occur. Furthermore, it is also unclear whether these statistical problems harm prediction accuracy in cases when the problems exist. The current study measures the incidence of these statistical problems and examines the effects of each on prediction accuracy.

Tests were conducted to assess the incidence of 1) autocorrelation, 2) continuity, 3) heteroscedasticity, 4) multicollinearity, and 5) normality. In addition, the effect of each of these potential problems on model predictions was also measured and is reported in this section. Subsections 5.1.3.1 through 5.1.3.5 will discuss the results of the diagnostic tests for each of the potential problems mentioned. Subsection 5.1.3.6 contains a summary of the diagnostic testing results.

#### **5.1.3.1 Autocorrelation**

This subsection reports on the incidence of autocorrelation and the effects of significant autocorrelation on the prediction models estimated in the current study. Autocorrelation refers to the tendency of the residuals to move in a systematic



pattern. The presence of autocorrelation is believed to significantly harm prediction accuracy. The current study examines the potentially harmful effects of autocorrelation on prediction accuracy.

The results of the autocorrelation testing are presented in Table 5.6. Panel A of Table 5.6 reports the number of cases in which autocorrelation was significant ( $n = 135$ ; 3 accounts, 5 regression models, 9 companies). In total, autocorrelation was significant in 46 of 135 cases.

Panel B of Table 5.6 indicates that the presence of autocorrelation significantly harms prediction accuracy. The average MAPE when autocorrelation is present is 13.3%. The average MAPE when autocorrelation is not present is 11.1%. The difference is significant ( $p < .10$ ;  $t = 1.29$ ). Thus, predictions are significantly less accurate when autocorrelation is present than when autocorrelation is not present. This result was true in general, but was not the case for First-differences.

The prediction accuracy of First-differences was unaffected by the incidence of significant autocorrelation. The MAPE when autocorrelation was present is 9.3%. The MAPE when autocorrelation was not present is 9.2%. First-differences provides more accurate predictions than other prediction methods whether or not significant autocorrelation is present.

There are two primary implications of these findings. First, auditors using statistical analytical procedures should be aware that the presence of autocorrelation significantly harms prediction accuracy. Accordingly, auditors should test for

**Table 5.6**  
**Autocorrelation Diagnostic Testing Results**

**Panel A: Incidence of Autocorrelation (n = 135)**

Number of Cases Autocorrelation Test Significant	46
Number of Cases Autocorrelation Test Not Significant	89
Total	135

**Panel B: Comparison of Prediction MAPEs (n = 135)**

Average Prediction MAPE Autocorrelation Test Significant	13.3%
Average Prediction MAPE Autocorrelation Test Not Significant	11.1%
Difference	2.2% *

\* Difference significant ( $p < .10$ ;  $t = 1.29$ ).

autocorrelation to determine if it is present. Second, auditors should strongly consider the use of First-differences regression. When autocorrelation was present, the First-differences was found to significantly improve predictions compared to those obtained using OLS regression. Furthermore, when autocorrelation was not present, First-differences provided more accurate predictions than either OLS or Cochrane-Orcutt.

#### **5.1.3.2 Continuity**

This subsection reports the incidence of the lack of continuity and the effects on predictions when a lack of continuity is present. Tests for continuity investigate whether changes have occurred in the model over time.

Panel A of Table 5.7 indicates that the continuity test was significant in 54 of 135 cases ( $n = 135$ ; 3 accounts, 5 regression models, 9 companies). However, the lack of continuity did not significantly harm predictions as evidenced by the MAPEs presented in Panel B of Table 5.7. The average MAPE in the prediction period when the continuity test was significant was 12.4 percent. The average MAPE when the continuity test was not significant was 11.4 percent. The two average MAPEs were not significantly different at conventional levels.

The implication is that the lack of continuity does not significantly harm predictions. However, this result should be interpreted with caution. It is possible that more significant shifts in continuity in other companies, industries or time periods, may give rise to inaccurate predictions. Auditors should still consider the possibility of changes which may occur in a prediction model over time.

Nevertheless, in the current study, continuity was not found to harm model predictions.

#### **5.1.3.3 Heteroscedasticity**

This subsection contains the results of the diagnostic tests for heteroscedasticity. Heteroscedasticity refers to the tendency for the variance of the predictions to vary at different levels of one or more independent variables.

Panel A of Table 5.8 indicates that heteroscedasticity was found to be significant in 59 of the 135 cases. Panel B of Table 5.8 indicates that the average MAPE was 13.2 percent in the prediction period when significant heteroscedasticity was present. The average MAPE when heteroscedasticity was not present was 10.8 percent.

The predictions were significantly more accurate when heteroscedasticity was present than when heteroscedasticity was not present ( $p < .08$ ). Thus, the incidence of heteroscedasticity appears to significantly harm prediction accuracy. However, an alternative analysis, which will be presented in Subsection 5.1.3.6 indicates that the incidence of heteroscedasticity does not significantly harm predictions.

#### **5.1.3.4 Multicollinearity**

This subsection contains the results of the diagnostic tests for multicollinearity. Multicollinearity refers to the tendency of predictor variables to be correlated with one another. Multicollinearity increases the standard error of the regression model.

Table 5.9 contains the results of the diagnostic tests for multicollinearity. Panel A of the table indicates that multicollinearity is significant in 46 out of 81

**Table 5.7**  
**Continuity Test Results**

**Panel A: Incidence of the Lack of Continuity (n = 135)**

Number of Cases Continuity Test Significant	54
Number of Cases Continuity Test Not Significant	81
Total	135

**Panel B: Comparison of Prediction MAPEs (n = 135)**

Average Prediction MAPE Continuity Test Significant	12.4%
Average Prediction MAPE Continuity Test Not Significant	11.4%
Difference	1.0% *

\* Difference not significant at conventional levels.

**Table 5.8**  
**Heteroscedasticity Test Results**

**Panel A: Incidence of Heteroscedasticity (n = 135)**

Number Cases Heteroscedasticity Test Significant	59
Number Cases Heteroscedasticity Test Not Significant	76
Total	135

**Panel B: Comparison of Prediction MAPEs (n = 135)**

Average Prediction MAPE Heteroscedasticity Test Significant	13.2%
Average Prediction MAPE Heteroscedasticity Test Not Significant	10.8%
Difference	2.4% *

\* Difference significant ( $p < .08$ ;  $t = 1.49$ ).

**Table 5.9**  
**Multicollinearity Test Results**

**Panel A: Incidence of Multicollinearity (n = 135)**

Number Cases Multicollinearity Test Significant	58
Number Cases Multicollinearity Test Not Significant	77
Total	135

**Panel B: Comparison of Prediction MAPEs (n = 135)**

Average Prediction MAPE Multicollinearity Test Significant	13.2%
Average Prediction MAPE Multicollinearity Test Not Significant	10.9%
Difference	2.3% *

\* Difference significant ( $p < .09$ ;  $t = 1.44$ ).

**Table 5.10**  
**Normality Test Results**

**Panel A: Incidence of the Lack of Normality (n = 135)**

Number of Cases Normality Test Significant	1
Number of Cases Normality Test Not Significant	134
Total	135



cases. Panel B indicates that the average MAPE when multicollinearity is present is 13.3 percent. the average MAPE when multicollinearity is not present is 11.1 percent. The difference of 2.2 percent is significant ( $p < .07$ ).

#### **5.1.3.5 Normality**

This subsection contains the results of the diagnostic tests for normality. Normality refers to the assumption of regression that the prediction errors be normally distributed. Table 5.10 indicates that normality was not a problem. The test for normality was only significant once in a total of 135 cases, which is less than one percent. The effects on prediction accuracy were, therefore, not reported.

#### **5.1.3.6 Alternative test and summary of diagnostic testing results**

An alternative test was also performed to test the impact of each of 1) autocorrelation, 2) continuity, 3) heteroscedasticity, 4) multicollinearity, and 5) normality on model predictions. MAPEs from the prediction period were regressed onto the presence or absence of each of these five potential problems. Dummy variables were used as the independent variables, where 0 denoted the absence of the potential problem and 1 denoted the presence of the potential problem. The values of the dummy variables were determined by employing the diagnostic tests presented in Figure 2.2. If the statistic was significant at a confidence level of 95%, then the related item was coded with a 1, indicating the presence of the potential problem. If the statistic was not significant at a confidence level of 95%, then the related item was coded with a 0, indicating the absence of the potential problem.

The results of the regression model are reported in Table 5.11. The significantly positive T-statistics indicate that high levels of autocorrelation and multicollinearity lead to higher MAPEs. Consistent with the results presented in subsections 5.1.3.1 and 5.1.3.4, the incidence of autocorrelation and multicollinearity harm model predictions. The T-statistic associated with the heteroscedasticity coefficient was not significant. Thus, after controlling for the effects of autocorrelation and multicollinearity, the harmful effects of heteroscedasticity were not present.

The insignificant T-statistics for continuity and normality are consistent with the results presented in Subsections 5.1.3.2 and 5.1.3.5. The incidence of continuity and normality was not found to be related to prediction accuracy.

In summary, the diagnostic tests for autocorrelation and multicollinearity indicated that these potential problems harm model predictions. When autocorrelation is present, the predictions obtained from First-differences were more accurate than the predictions obtained from any other prediction method. Furthermore, predictions were more accurate when multicollinearity was not present than when multicollinearity was present. After controlling for the effects of autocorrelation and multicollinearity, heteroscedasticity did not significantly harm prediction accuracy. Tests for continuity and normality indicated no significant harm to model predictions.

**Table 5.11**  
**Diagnostic Test Summary: Regression Model**

Dependent variable: MAPE from the prediction period (n = 135)

Independent variables	T-Statistic	p
Autocorrelation	2.17	.01
Continuity	-.78	NS
Heteroscedasticity	-.66	NS
Multicollinearity	2.27	.01
Normality	.63	NS

## **5.2 Pooled Models**

This section contains the results of the models generated from pooled data. Heretofore in the current study, the predictions for each company have been estimated using information from a single company. For example, the predictions for Company Three were generated using data from Company Three. The predictions for Company Four were generated using data from Company Four, etc. This individual company approach has been followed in prior studies because most of these prior studies only collect data from a single company.

In the current study, a pooled modeling approach is also possible because data were collected from multiple companies in the same industry. Prior research suggests that employment of analytical procedures at the industry may be appropriate (AICPA, 1988). Therefore, a pooled approach is used in the current study in addition to the individual company approach used in prior studies.

Pooling the data from multiple companies in the same industry provides two primary advantages over individual company models. First, data may be estimated with more current base-period data than individual models. Using more current base-period data reduces the possibility that structural changes, that occur over time, adversely affecting model predictions. Second, pooled models may signal errors that would not be signaled by individual company models. For example, recurring errors in a given company's financial statements may not stand out as unusual when examined in isolation. However, when combined with information from other similar companies, the errors are more likely to stand out as unusual. Therefore, this section

of the current study assesses the performance of pooled models. The performance of pooled models was compared to the performance of individual company models.

This section is divided into three parts. Subsection 5.2.1 describes the procedures used to develop the pooled models. Subsection 5.2.2 contains the results of the results of the pooled models. Subsection 5.2.3 presents the implications of the pooled model findings.

#### **5.2.1 Pooled model procedures**

In order to obtain pooled predictions, data were grouped together from multiple companies to estimate a single prediction model for each of the three account balances of interest. A single revenue was used to generate revenue predictions for each of the companies. A single fuel expense model was used to generate fuel expense predictions for each of the companies, etc. The performance of each of the pooled models was compared to the performance of individual company models for each of the companies in the pooled group.

Five companies were included in the pooled model group. Four of the companies were not included in the pooled model group because some data were not available. Budgeted data were not available for three of the companies, and lagged observations of the account balances of interest were not available for one of the companies. Therefore, these companies could not be included in the pooled models.

The performance of the pooled models was evaluated by computing average mean absolute percentage errors (MAPEs) for the companies in the pooled group sample. The MAPEs were generated using predictions from the hold-out period.

MAPEs were computed by taking the absolute value of the difference between the predicted monthly account balance and the recorded monthly account balance, divided by the recorded monthly account balance. The MAPEs achieved from the pooled models were compared to the MAPEs achieved from the individual company models of the companies in the pooled group.

The pooled models were estimated using Ordinary Least Squares regression. The First-differences and Cochrane-Orcutt were not appropriate method selections because the pooled models possess both time-series and cross-sectional components. OLS regression was also used to estimate the individual company models in order to provide a fair comparison of pooled and individual company prediction models.

### **5.2.2 Pooled model results**

Table 5.12 presents the comparison of the pooled models with the individual company models. Panel A presents the results for the prediction period, and Panel B presents the results for the model-building period.

The results in the prediction period were mixed. Panel A indicates that in the prediction period, the pooled average MAPE was 9.7%, which is significantly lower than individual company average MAPE of 14.3%. This result was true for fuel and production expenses, but did not hold for revenue. The pooled model results were better for fuel expense ( $p < .11$ ) and production expense ( $p < .07$ ). The opposite was true for revenue. The individual firm revenue models were better than the pooled revenue models as evidenced by the individual company MAPE of 3.8% compared to the pooled MAPE of 5.7%.

The results from the model-building period are not consistent with the results from the prediction period. Panel B presents the results from the model-building period for each of the three account balances of interest. The results indicate that the individual company models were significantly better than the pooled models for all three accounts in the model building period ( $p < .01$ ). The opposite was true for fuel and production expenses in the prediction period, as indicated in Panel A.

### **5.2.3 Implications of pooled model results**

The pooled models were more accurate than individual company models for fuel expense and production expense. The predictions for both of these accounts have been less accurate than revenues throughout the study. The implication is that, the potential improvement from using pooled models appears to be greatest for accounts that cannot be predicted with a high degree of accuracy. Pooled models more accurately predict difficult to predict accounts than individual company models. Therefore, if the auditor is unable to obtain the desired level of precision using individual company models, the predictions may be improved by pooling information from one or more similar companies.

One important implication for accounting researchers is the importance of measuring model performance in a hold-out period. The results from the model building-period were different than the results from the prediction period. Results based on analyses which do not include a prediction period may lead to erroneous conclusions. In the current study, for example, the results from the

**Table 5.12**  
**Comparison of Pooled Models and Individual Company Models**

**Panel A: Prediction Period**

Account	Individual Company MAPE	Pooled MAPE
Revenue	3.8	5.7
Fuel Expense	16.1	9.5****
Production Expense	23.2	13.8***
Average	14.3	9.7**

\*\* Significantly lower than Individual Company MAPE ( $p < .05$ ).

\*\*\* Significantly lower than Individual Company MAPE ( $p < .07$ ).

\*\*\*\* Lower than Individual Company MAPE ( $p < .11$ ).

**Panel B: Model Building Period**

Account	Individual Company MAPE	Pooled MAPE
Revenue	1.5	3.8
Fuel Expense	5.6	9.5
Production Expense	7.6	10.8
Average	4.9*	8.0

\* Significantly lower than Pooled MAPE ( $p < .01$ ).



model-building period suggest that individual company models are always superior pooled prediction models. However, when the performance of the same models is measured in a hold-out period, the performance of the pooled models were found to be significantly better than individual company models for two of the three accounts.

### **5.3 Quarterly versus Monthly Models**

The level of data aggregation that is appropriate for analytical procedures is not always clear. Auditors must choose the level of data aggregation that is most appropriate. Some types of analytical procedures are performed using only annual account balances. Other analytical procedures rely on quarterly or monthly comparisons. In general, SAPs require monthly or quarterly data; annual data are not appropriate for most SAP applications.

Prior research comparing the performance of quarterly and monthly prediction models is inconclusive. The results of Wild (1987) indicate that monthly models are more accurate than quarterly prediction models. On the other hand, Wheeler and Pany (1990) assert that quarterly predictions are more accurate than monthly predictions. This assertion is based on the idea that measurement error is more likely to be contained in monthly data since monthly account balances are unaudited. Quarterly balances are subject to review by independent auditors and are, therefore, less likely to contain measurement error than monthly account balances.

This section contains the results of the performance of monthly and quarterly prediction models. The section is divided into three parts. Subsection 5.3.1 presents

the procedures used to compare the performance of monthly and quarterly prediction models. Subsection 5.3.2 presents the results. Subsection 5.3.3 presents the implications of these findings.

#### **5.3.1 Procedures used to compare monthly and quarterly models**

This subsection contains the procedures used to compare the performance of monthly and quarterly prediction models. In the current study, the performance of monthly prediction models was compared to quarterly prediction models. Monthly models were estimated using monthly data points. Quarterly models were estimated using quarterly data points. Performance was measured and evaluated using MAPEs from both the prediction period and the model- building.

The first step in constructing the quarterly models was to appropriately aggregate the monthly account balances and predictor variables. For the account balances of interest and some of the predictor variables, this required summing three monthly data points. For other variables, it was appropriate to average the three data points instead of summing them. Variables which were summed included the three account balances of interest, heating and cooling degree days, KWH generated, KWH sold, budgeted revenues and expenses, and lagged revenues and expenses. Variables which were averaged included the rate factors, number of customers, unemployment, CPI, Fuel-CPI, capacity factor, and load factor.

The comparison of quarterly and monthly models could only be performed for pooled models. It was not possible to estimate individual company quarterly models due to the lack of sufficient observations ( $n = 12$ ; 4 quarters X 3 years of data in the

model building period). Prior research indicates that 24 to 36 data points in the model building period are required to develop adequate prediction models (Stringer, 1975; Albrecht and McKeown, 1976; Akresh and Wallace, 1981).

### **5.3.2 Results of monthly and quarterly prediction models**

This subsection contains the results of the quarterly and monthly prediction models. Table 5.13 presents the MAPEs achieved from quarterly models and monthly models. The results were inconclusive. The monthly pooled MAPEs for revenue and fuel expense were lower than the quarterly pooled MAPEs in both the model building period and the prediction period. This would indicate that monthly prediction models are superior to quarterly prediction models. However, the opposite was true for production expense. The quarterly MAPEs were lower than the monthly MAPEs. The results from the prediction period presented in Table 5.15 were consistent with the results from the model-building period (not presented).

### **5.3.3 Implications**

The results for revenue and fuel expense were consistent with the findings of Wild (1987). The monthly models more accurately predicted the account balances of interest than the quarter prediction models. The results for production expense were consistent with the assertion of Wheeler and Pany (1990), that quarterly predictions are more accurate than monthly predictions.

The predictions for production expense were never as accurate as the predictions for revenue and fuel expense. Thus, the more accurate the prediction model, the more likely that monthly prediction models are superior to quarterly

**Table 5.13**  
**Comparison of Monthly Models to Quarterly Models**

**Prediction Period MAPEs**

Account	Quarterly MAPE	Monthly MAPE
Revenue	9.8	5.7*
Fuel Expense	15.4	9.5**
Production Expense	9.7#	13.8
Average	11.6	9.7

\* Significantly lower than quarterly MAPE ( $p < .10$ ).

\*\* Significantly lower than quarterly MAPE ( $p < .04$ ).

# Significantly lower than monthly MAPE ( $p < .05$ ).

prediction models. The results suggest that the inverse is also true. For less accurate predictions, quarterly prediction models are likely to be superior to monthly prediction models. Further research is required determine conclusively the relative accuracy of monthly and quarterly prediction models.

#### **5.4 Summary**

A discussion of the analyses and results for each of the three objectives of that were presented in this chapter. The objectives of the current study examined in this chapter were 1) to evaluate the consistency of SAP models, 2) to evaluate the performance of pooled models, and 3) to compare the performance of quarterly and monthly models.

The primary findings related to each of these four objectives are summarized as follows:

**Model Consistency:** The revenue prediction models were more robust than the fuel or production expense models. The production expense predictions were particularly disappointing, as evidenced by MAPEs in the prediction period which were higher than MAPEs achieved using naive prediction models.

Production information (KWH production) emerged as the most robust predictor variable for all three accounts. However, the strength of the relationship between KWH production was strongest for revenue and fuel expense. The strength of the relationship between KWH production and production expense was much lower. Consistent predictor variables did not emerge for production expenses.

The diagnostic testing indicated that predictions were improved by reducing the incidence of autocorrelation and multicollinearity. Predictions using First-differences were relatively consistent whether or not autocorrelation was present. Diagnostic tests for continuity and normality indicated that these potential problems did not significantly harm predictions.

**Pooled Models:** The results indicated that pooled models predicted more accurately than individual company prediction models for fuel and production expense. The reverse was true for revenue. The implication is that pooled models appear to work best for accounts that cannot be modelled with a high degree of accuracy. Individual company models tended to perform best for account balances that were modelled with a high degree of accuracy.

**Quarterly versus Monthly Models:** The results indicated that monthly prediction models were, in general, more accurate than quarterly prediction models. The performance of quarterly prediction models was found to improve for accounts with lower prediction accuracy.

The next chapter contains a summary of the primary conclusions, implications, contributions and limitations of the current study. The chapter also suggests areas for future research resulting from the current study.

## **Chapter VI**

### **6. SUMMARY, IMPLICATIONS, CONTRIBUTIONS, LIMITATIONS, AND SUGGESTIONS FOR FUTURE RESEARCH**

In the current study, statistical analytical procedures (SAPs) were developed and tested for a sample of nine electric utilities. Both financial and nonfinancial data were collected from each sample company for the period January, 1986 through December, 1989. The information was used to predict revenue, fuel expense and production expense account balances. The primary objectives of the study were to 1) to compare the performance of alternative SAP methods 2) evaluate the consistency of the SAP models across sample companies, 3) to evaluate the performance of pooled models, and 4) to compare the performance of quarterly and monthly prediction models.

This chapter contains a summary of the primary research findings, and the implications of these findings. The chapter also describes the contributions and limitations of the current study, as well as suggestions for future research. The chapter is divided into four sections. Section 6.1 contains a summary of the results and the implications of the results. Section 6.2 contains a discussion of the contributions and limitations of the current study. Section 6.3 contains suggestions for future research resulting from the current study. Section 6.4 contains a final summary.

## **6.1 Summary of the Results and Implications**

This section contains a summary of the primary results of the current study as well as the implications of the results. The section is organized around the four objectives of the current study. Subsection 6.1.1 summarizes the results of the robustness of SAP models. Subsection 6.1.2 summarizes the results and of the pooled models. Subsection 6.1.3 summarizes the results of the comparison of quarterly and monthly prediction models. Subsection 6.1.4 summarizes the findings regarding the accuracy of alternative SAP prediction methods.

### **6.1.1 Alternative SAP prediction methods**

The performance of eight prediction methods was evaluated. Six of the eight methods were statistical methods, including five regression methods and Census X-11. The remaining two methods were naive prediction methods (the martingale and submartingale methods), which served as benchmarks for the statistical prediction methods.

Method performance evaluated using three alternative measurements: 1) by comparing monthly mean absolute percentage errors (MAPEs), 2) by assessing ability of the methods to properly detect seeded errors, and 3) by comparing annualized predictions. The use of three alternative measures provided more conclusive evidence of method performance than performance evaluation using a single measure. The results of the three measures of performance were consistent. The primary results and implications of these findings are presented in the following paragraphs.



In general, the prediction performance of the SAP methods dominated the naive models. First-differences was the most accurate prediction method for revenue and fuel expense. In addition to achieving more accurate average predictions, First-differences also exhibited greater consistency than other prediction methods.

The results of the current study partially contradict the results of a prior study (Wheeler and Pany, 1990). The prior study indicates that Census X-11 performs better than regression. However, in the current study, First-differences and Cochrane-Orcutt regression generated more consistently accurate predictions than Census X-11. The primary reason for the improved performance of regression relative to X-11 is that both financial and nonfinancial information were included in the current study. The Wheeler and Pany (1990) study only incorporates information which was readily available in the financial statements. This finding underscores the importance of including nonfinancial predictor variables when using statistical prediction methods.

The seeding of artificial errors indicated that SAPs were found to be useful in signalling annual material errors seeded into monthly account balances. SAPs were significantly less useful in signalling the presence of quarterly and monthly errors. Combined type I and type II error rates achieved in the current study may appear to be higher than would be acceptable in practice. The results indicated that the SAPs developed in the current study should not be performed in isolation. The error rates achieved were not low enough to justify the complete exclusion of other substantive tests. Other substantive procedures would be required to reduce type

II error rates to a tolerable level. The use of SAPs may, however, justify a significant reduction of other substantive tests.

The results imply that SAP predictions are not appropriate for all accounts. One suitable benchmark for the appropriateness of using SAPs is whether the errors in the prediction period are significantly better than a naive model. Statistical methods are more costly to employ than the naive methods. Therefore, the statistical methods must perform better to justify their use.

#### **6.1.2 Consistency of SAP models**

This subsection contains a summary of the results and implications of the consistency of SAP models. The consistency of SAP models was evaluated by examining model performance across multiple companies. Robustness was evaluated in by examining 1) the consistency of the model performance, 2) the consistency of significant predictor variables, and 3) the consistency of diagnostic tests of the assumptions of the models. The results and implications of each are summarized next.

Revenue models generated the most accurate predictions, followed by fuel expense and production expense, respectively. The revenue models exhibited very consistent performance for all nine firms. For example, the average mean absolute percentage error (MAPE) for all nine firms' revenue models was 2.4 percent in the model building period. Some of the fuel and production expense predictions exhibited performance nearly as good as the revenue models. However, the predictions were not consistently as good for all companies.

The statistical prediction methods performed reasonably well for fuel expense. The statistical methods performed significantly better than the naive methods both in the model-building period and a hold-out period. However, the fuel expense predictions exhibited less consistency than revenue predictions, as evidenced by MAPEs for some of the companies being greater than ten percent.

The results for production expense were especially disappointing. None of the statistical prediction methods performed as well as one of the naive methods. Thus, statistical prediction methods are not recommended for production expenses due to the inconsistency of the predictions obtained.

The consistency of the revenue predictions is indicative of their usefulness to auditors. Models such as should allow auditors to reduce the amount of substantive testing for revenues, and the related receivables. These models were found to be generalizable to all of the firms in the sample.

The results also indicated the variables that were the most consistent predictors of each account balance. KWH production emerged as the most robust predictor variable for all three accounts. Other robust predictor variables for revenues were degree days and rate data. Other than KWH production, there were no variables that emerged as consistently strong predictors for fuel and production expenses for the nine sample firms. Degree days and budgeted fuel expense were moderately robust predictor variables for fuel expense. CPI and trend emerged as moderately robust predictors for production expense.

The implication is that the degree of measurement error associated with predictor variables was believed to significantly effect prediction accuracy. The relatively high degree of measurement error associated the best predictor variables for production expenses (i.e. CPI, trend) was believed to be the primary reason for the less accurate predictions for this account. Conversely, the relatively low degree of measurement associated with the best predictor variables for revenues (i.e. KWH production, rates, degree days) was believed to account for more accurate predictions for this account.

The inclusion of nonfinancial predictor variables was found to significantly improve prediction accuracy. Prediction accuracy was significantly better when both financial and nonfinancial information were used in prediction models than when only financial information was included.

Diagnostic were performed to identify the incidence and effects of 1) autocorrelation, 2) continuity, 3) heteroscedasticity, 4) multicollinearity, and 5) normality. The diagnostic tests performed revealed that autocorrelation and multicollinearity significantly reduced prediction accuracy. When autocorrelation was present, use of First-differences regression resulted in lower average MAPEs in the prediction period than other prediction methods. The incidence of continuity, heteroscedasticity, and normality did not significantly influence prediction accuracy. The presence of multicollinearity was found to harm prediction accuracy.

There are two primary implications of the diagnostic tests. First, auditors should use First-differences regression when significant autocorrelation is present.

The predictions for these methods were found to be significantly more accurate than the predictions obtained using other methods when significant autocorrelation was present.

Second, the incidence of multicollinearity was found to significantly harm prediction accuracy. The effects of multicollinearity may be reduced by eliminating highly correlated predictor variables from the models. Auditors should be aware of the potentially harmful effects of multicollinearity. When high levels of multicollinearity are present, the auditor should consider refining the prediction model by eliminating the one or more highly correlated variables from the analysis.

In general, violations of the assumptions did not significantly harm prediction accuracy. The incidence of continuity, heteroscedasticity and normality problems did not significantly harm prediction accuracy either in the model-building period or in the prediction period. The incidence of autocorrelation did significantly harm prediction accuracy in the prediction period. However, the harmful effects of autocorrelation were found to be avoidable by using the First-differences prediction method.

### **6.1.3 Pooled models**

The performance of models generated using individual company data were compared to the performance of models generated using data from multiple companies (pooled models). The results of the analysis were mixed. The results indicated that pooled models were less accurate than individual company models for accounts for which consistently accurate predictions were possible (revenues). Pooled

models were found to be more accurate than individual company models for accounts that could not be modeled with a high degree of consistency (fuel expense and production expense).

The implication is that if the precision achieved by using individual company data is not sufficient for a specific account, the auditor may improve prediction accuracy by pooling data from other similar companies. Pooling was found to be more useful for applications in which the prediction accuracy was low.

#### **6.1.4 Quarterly versus monthly models**

The performance of quarterly and monthly prediction models was compared. The results of the analysis were inconclusive. The monthly models performed better than the quarterly prediction models for revenue and fuel expense predictions both in the model-building period and in the prediction period. However, the quarterly models performed better than the monthly models for production expense predictions.

The implication is that monthly models perform better for accounts for which consistently accurate performance is possible. Quarterly prediction appear to perform better for accounts that are predicted with less accuracy.

## **6.2 Primary Contributions and Limitations**

This section describes the primary contributions and limitations of the current study. Subsection 6.2.1 contains a discussion of the primary contributions of the current study. Subsection 6.2.2 contains a discussion of the limitations of the current study.

### **6.2.1 Contributions of the current study**

There are four primary contributions of the current study. First, the attributes of the current study made possible a meaningful evaluation of the consistency of SAP models. Second, the current study evaluated model performance using more stringent tests than have been used in many prior studies. Third, the current study evaluated the usefulness of pooled models. Fourth, the current study made a more meaningful comparison of alternative prediction methods than has been accomplished in prior studies. Each of these contributions is discussed in greater detail in the following paragraphs.

The current study included both 1) financial and nonfinancial information, and 2) multiple firms in the sample. The inclusion of both of these elements allowed the current study to address two important objectives. First, the consistency of SAP models could only be addressed through the inclusion of multiple firms in the sample and inclusion of financial and nonfinancial information. Identifying robust predictor variables, and robust prediction models could not be accomplished without both of these elements. Likewise, the benefits of pooling data for SAPs could only be assessed through the inclusion of multiple firms in the sample.

A second important contribution of the current study was the use of strict tests of the models in a "hold-out" period. Many prior studies base their results wholly, or in part, on goodness of fit criteria from the model building period (Wheeler and Pany, 1990; Akresh and Wallace, 1981; Neter, 1981; Albrecht and McKeown, 1976). In addition to goodness of fit criteria, the current study also tested the models in a

"hold-out"period. This analysis revealed that the best fitting models in the prediction period did not always achieve the most accurate predictions in the "hold-out"period.

A third contribution of the current study was the pooling of data. Pooling allowed predictions for multiple firms to be obtained from a single model. For two of the three accounts modelled in the current study, predictions were significantly more accurate using pooled data than using individual company data. A fourth contribution of the current study was a more meaningful comparison of the relative performance of regression and Census X-11. A prior study comparing the performance of regression and Census X-11 included only financial information. The study concluded that "X-11 predicted better than any other expectation model used [including regression]." By including both financial and nonfinancial information, the results of the prior study were contradicted. Regression outperformed Census X-11 in both 1) signalling material errors, and 2) prediction accuracy using MAPEs to measure prediction accuracy.

#### **6.2.2 Limitations of the current study**

One limitation of the current study is that the inferences made may not be generalizable to other industries. Further research is needed to determine the usefulness of SAPs for other firms and industries. The model-building methodologies used in the current study should be useful in examining the usefulness of SAPs in other industries.

Another limitation is that the sample of electric utilities may not be representative of all electric utilities. Only a small number of companies could be





included in the sample. To reduce the effects of a small sample, utilities with varying production facilities, and in different geographic regions were included in the sample in an effort to make the sample as representative of the population as possible. Nevertheless, the sample may not capture all of the important characteristics of the population of electric utilities.

Another limitation of the current study is that the models were developed without the benefit of company specific information which may be available to auditors. Other company specific information, not available in the current study, may further enhance the usefulness of individual company prediction models. The results of the current study may understate the usefulness of SAPs. The performance of SAPs were evaluated in isolation in the current study. In practice, SAPs would be combined with other substantive tests. Auditors would have the benefit of the combined assurance obtained through using SAPs along with other audit tests.

### **6.3 Suggestions for Future Research**

The current study indicates the need for further research in four areas. First, additional industry studies are needed. Second, Cost-benefit studies which measure and evaluate the benefits obtained from using SAPs compared to the costs of these methods are needed. Third, research which evaluates the level of data aggregation that is appropriate for SAP models is needed. Fourth, additional research is needed to determine the usefulness of pooled models relative to individual company models. Each of these four areas in which further research is needed will be discussed in the following paragraphs.

Identification of the industries and accounts in which development of SAPs is appropriate. SAP predictions should be significantly better than naive predictions in a "hold-out" period to be considered potentially useful to auditors.

Research is needed which examines the costs and benefits of applying SAPs. This research should address the costs of employing SAPs compared to the savings obtained through reduction of other substantive audit tests. Particular attention should be focused on identifying 1) the incremental costs of using SAPs as opposed to more traditional analytical procedures and 2) the incremental benefits of SAPs over other analytical procedures. Studies should evaluate the level of assurance provided by different types of statistical and nonstatistical analytical procedures. Such studies should also attempt to address the affects that alternative analytical procedures have on the extent of other substantive tests.

More research is needed to identify the levels of data aggregation that are most appropriate when using analytical procedures. The current study examined the temporal level of data aggregation by comparing the performance of quarterly and monthly prediction models. The results of the current study were inconclusive regarding temporal data aggregation. Further research is needed to determine the level of temporal data aggregation that is most appropriate. Furthermore, research is needed to determine whether segmented information might lead to more accurate prediction models than company wide models. For example, the expense predictions in the current study may have been more accurate if production information had been available by plant. The level of data aggregation that is appropriate may be

## REFERENCES

- Akresh, A, J.K. Loebbecke, and W.R. Scott, 1988. "Audit Approaches and Techniques," *Research Opportunities In Auditing: The Second Decade*, Edited by A.R. Abdel-Khalik and I. Solomon.
- \_\_\_\_\_, and W. Wallace, 1981. "The Application of Regression Analysis for Limited Review and Audit Planning," Symposium IV, University of Illinois, pp. 68-129.
- Albrecht, W.S., and J.C. McKeown, 1977. "Toward an Extended Use of Statistical Analytical Reviews in the Audit," *Symposium on Auditing Research II, University of Illinois*, pp. 53-69.
- American Institute of Certified Public Accountants. 1988. Statement on Auditing Standards No. 56: Analytical Procedures. AICPA.
- American Institute of Certified Public Accountants. 1980. Statement on Auditing Standards No. 31: Evidential Matter. AICPA.
- \_\_\_\_\_, 1990. Exposure Draft: The Confirmation Process. AICPA.
- Arens, A., and J. Loebbecke, 1991. *Auditing: An Integrated Approach, 5th Edition*. Prentice Hall Publishers, Englewood Cliffs, New Jersey.
- Arrington, E., W. Hillison, and R. Icerman, 1983. "Research in Analytical Review: The State of the Art," *Journal of Accounting Literature*, pp. 151-185.
- Belsley, D. A., E. Kuh, and R. E. Welsch, 1980. *Regression Diagnostics, Identifying Influential Data and Sources of Colinearity*, Wiley, New York, Chapter 3.
- Biggs, S. F., and J. J. Wild, 1984. "A Note on the Practice of Analytical Review," *Auditing: A Journal of Practice and Theory* (Spring), pp. 69-79.
- Daroca, F., and W. Holder, 1985. "The Use of Analytical Procedures in Review and Audit Engagements," *Auditing: A Journal of Practice and Theory*, (Spring), pp. 80-92.
- Dugan, M., J. Gentry, and K. Shriver, 1985. "The X-11 Model: A new Analytical Review Technique for the Auditor" *Auditing: A Journal of Practice and Theory* (Spring): 23-37.

- Elliot, R., 1979. "Discussants Response of The Effect of Measurement Error on Regression Results in Analytical Review," *Symposium III. University of Illinois*, pp. 49-64.
- \_\_\_\_\_, 1983. "Unique Methods: Peat Marwick International," *Auditing: A Journal of Practice and Theory (Spring)*: pp. 1-12.
- Holstrum, G., and W. Messier, 1982. "A Review and Integration of Empirical Research on Materiality," *Auditing: A Journal of Practice and Theory, (fall)*: pp. 45-63
- Hyman, L., 1988. *America's Electric Utilities: Past, Present and Future*, Public Utilities Reports, Inc.
- Kaplan, R., 1978. "Developing a Financial Planning Model for an Analytical Review: A Feasibility Study," *Symposium on Auditing Research III, University of Illinois*, pp. 3-30.
- Kinney, W. R., 1978. "ARIMA and Regression in Analytical Review: an Empirical Test," *The Accounting Review* (January), pp. 48-60.
- \_\_\_\_\_, 1979. "The Predictive Power of Limited Information in Preliminary Analytical Review: An Empirical Study," *Journal of Accounting Research* (Supplement), pp. 148-165.
- \_\_\_\_\_, 1983. "Quantitative Applications in Auditing," *Journal of Accounting Literature*. pp. 187-204
- \_\_\_\_\_, 1987. "Attention Directing Analytical Review Using Accounting Ratios: A Case Study," *Auditing: A Journal of Practice and Theory (Spring)*, pp. 59-73.
- \_\_\_\_\_, and G.L. Salamon, 1979. "The Effect of Measurement Error on Regression Results in Analytical Review" *Symposium III. University of Illinois*: 49-64.
- Knechel, W. R., 1986. "A Simulation Study of the Relative Effectiveness of Alternative Analytical Review Procedures," *Decision Sciences* (Summer), pp. 376-394.
- \_\_\_\_\_, 1988. "The Effectiveness of Statistical Analytical Review as a Substantive Accounting Procedure: A Simulation Analysis," *The Accounting Review* (January), pp. 74-95.

- Neter, J., 1981. "Two Case Studies on Use of Regression for Analytic Review," Symposium IV, University of Illinois, pp. 292-337.
- Loebbecke, J.K., 1987. "Research Opportunities in Auditing: Analytical Procedures" *Prepared for the American Accounting Association Audit Section*, (March).
- \_\_\_\_\_, and Steinbart, 1987. "An Investigation of the Use of Preliminary Analytical Review to Provide Substantive Audit Evidence," *Auditing: A Journal of Practice and Theory*, pp. 74-89.
- Neter, John, 1981. "Two Case Studies on Use of Regression for Analytical Review," *Symposium IV*, University of Illinois, pp. 292-348.
- Phillips, C., 1988. *The Regulation of Public Utilities*, Public Utilities Reports, Inc.
- SAS Institute Inc., 1984. *SAS/ETS User's Guide, Version 5 Edition*, Cary, NC: SAS Institute Inc., pp. 551-602.
- Schmidt, F.L., 1971. "The Relative Efficiency of Regression and Simple Unit Predictor Weights In Applied Differential Psychology," *Educational and Psychological Measurement*, pp. 699-714.
- \_\_\_\_\_, 1972. "The Reliability of Differences Between Linear Regression Weights in Applied Differential Psychology," *Educational and Psychological Measurement*, pp. 879-886.
- Stringer, K. W., 1975. "A Statistical Technique for Analytical Review," *Journal of Accounting Research* (Spring), pp. 1-13.
- Tabor, R. H., and J. T. Willis, "Empirical Evidence on the Changing Role of Analytical Review Procedures," *Auditing: A Journal of Practice and Theory*, (Spring), pp. 93-109.
- Wallace, W., 1983a. "Analytical Review: Misconceptions, Applications and Experience--Part I," *CPA Journal*, January, 1983, pp. 24-37.
- Wallace, W., 1983b. "Analytical Review: Misconceptions, Applications and Experience--Part II," *CPA Journal*, February, 1983, pp. 18-27.
- White, H., 1980. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and A Direct Test for Heteroscedasticity," *Econometrics*, 48, pp. 817-838.

- Wild, J. J., 1987. "The Prediction Performance of a Structural Model of Accounting Numbers," *Journal of Accounting Research* (Spring), pp. 139-160.
- Wheeler, S., K. Pany, 1990. "Assessing the Performance of Analytical Procedures: A Best Case Scenario," *The Accounting Review* (July), pp. 557-577.