GRAHAM, Curtis Carl, 1935-
APPLICATIONS OF MULTIPLE REGRESSION AND CORRELATION ANALYSIS FOR FORECASTING AND CONTROLLING MANUFACTURING OVERHEAD AND OTHER INDIRECT COSTS.

The University of Oklahoma, Ph.D., 1968
Economics, commerce-business

University Microfilms, Inc., Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

APPLICATIONS OF MULTIPLE REGRESSION AND CORRELATION ANALYSIS FOR FORECASTING AND CONTROLLING MANUFACTURING OVERHEAD AND OTHER INDIRECT COSTS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

CURTIS C. GRAHAM

Norman, Oklahoma

1968

APPLICATIONS OF MULTIPLE REGRESSION AND CORRELATION ANALYSIS

FOR FORECASTING AND CONTROLLING MANUFACTURING

OVERHEAD AND OTHER INDIRECT COSTS

APPROVED BY

[illegible]

[illegible]

[illegible]

T. P. Herrick

James E. Hibdon

[illegible]

DISSERTATION COMMITTEE

ACKNOWLEDGMENT

The writer wishes to extend his sincere appreciation to the dissertation committee members, Professor Vernon H. Upchurch, Chairman, Professors Ed F. Crim, Jr., Theodore P. Herrick, Jr., James E. Hibdon, James M. Murphy, and Walter F. Scheffer for the many hours each spent reviewing and developing suggestions that are a part of this study. Special thanks are due Professor Vernon H. Upchurch for both his comments and encouragement during the writing of this dissertation and for the financial and moral support he provided throughout the writers period of academic work. The writer is also deeply indebted to Professor Ed F. Crim, Jr., for his intensive efforts and guidance that have contributed immeasurably to the validity of this dissertation. To Professor Theodore P. Herrick, Jr., the writer extends a special thanks both for his obtaining an H. B. Earhart Fellowship for the writer and for his personal encouragement.

In addition, the writer wishes to thank his student assistants, Miss Sylvia A. Lawton, Miss Linda G. Wyrick, and Mr. Bill H. Ozborn, for their incessant efforts over the past year. Finally, the writer is deeply grateful to his wife Anne for her understanding, urging, and personal sacrifices, without which this dissertation would not have been possible.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF SCHEDULES

LIST OF CHARTS

Chart | Page

Chart		Page

Chart Page

APPLICATIONS OF MULTIPLE REGRESSION AND CORRELATION ANALYSIS FOR FORECASTING AND CONTROLLING MANUFACTURING OVERHEAD AND OTHER INDIRECT COSTS

CHAPTER I

INTRODUCTION

Management bases its plans and operating decisions on the information provided by the accounting system and other internal and external information systems. If the firm is to be managed effectively, the accounting system must generate appropriate, accurate, and timely reports that reflect the results of management's plans and operating decisions. It follows that if improvements are made in the capabilities of the accounting system to provide additional and better information, management is in an improved position to manage. Increased efficiency in the selection, employment, and control of the firm's economic resources should result.

Traditionally, accounting systems have been designed to provide reports for both external and internal users of financial information. But the emphasis has been placed on the information needs of external users. The accounting profession "has been largely concerned with historical costs, the accountability for assets, the measurement of liabilities, and determination of the resulting net income."[1] But internal management has demanded and is demanding additional information in the form of better analysis and better data on which to base analysis.

Contemporary managerial accountants are seeking to develop integrated information systems capable of generating appropriate data to assist management in solving increasingly complex business problems. The advent of computers, now widely employed, has provided the necessary computational capabilities needed by managerial accountants to fulfill the internal accounting function. But the managerial accountant is confronted with an institutional framework that, like law, is bound by precedent and steeped in tradition. Professor Hector R. Anton aptly summarizes the current situation as follows:

> Increasing use of quantitative analyses in business for managerial decision-making has posed a significant dilemma for accountants. On the one hand, accountants are faced with a huge institutional structure, deeply rooted in legal and conventional theory and practice, and oriented mainly towards the preparation of multipurpose, general financial statements. This accounting structure has not been particularly effective in furnishing information that can be used directly, without substantial adjustments by management for decision-making, or indeed for many kinds of control. On the other hand, accountants have been and are faced with the task of incorporating these methods and techniques in their tool kits or surrendering them to other disciplines. Given the institutional framework, the task of incorporation is extremely difficult.[2]

Although managerial accountants are embroiled in a professional dilemma, an evolutionary transition is occurring wherein multi-dimensional

---

[1]Robert M. Trueblood, "Operations Research—A Challenge to Accounting," The Journal of Accountancy, Vol. CIX, No. 5 (May, 1960), p. 49.

[2]Hector R. Anton, "Activity Analysis of the Firm: A Theoretical Approach to Accounting (Systems Development)." Contemporary Issues in Cost Accounting, ed. Hector R. Anton and Peter R. Firmin (Boston: Houghton Mifflin Company, 1966), p. 517.

accounting systems are being developed. These systems will incorporate certain operations research techniques and concurrently will seek to satisfy the multifarious information needs of management. Professor Charles T. Horngren envisions the business information system of the future as:

> ...a single, multi-purpose system with a highly selective reporting scheme. It will be tightly integrated and will serve three main purposes: (1) routine reporting to management, primarily for planning and controlling current operations...; (2) special reporting to management, primarily for long-range planning and nonrecurring decisions...; and (3) routine reporting on financial results primarily for external parties....[3]

The evolutionary transition toward integrated and multi-dimensional accounting information systems has only commenced. But this transition is manifested by the "separation of budgetary procedures from the accounting system, and their use, especially in variable budgeting for manufacturing costs, can even be termed an alternative cost system."[4] In addition, "more and more extensive use of statistical cost studies, engineering economy, and special operations research programs may also be cited."[5] The recent publication of Richard Mattessich's book[6] relating to simulation of the firm with computerized budget models is another indication of the trend in managerial accounting toward multi-dimensional accounting information systems.

Because of this trend in internal accounting information systems, accountants must examine existing quantitative and statistical techniques to

---

[3]Charles T. Horngren, Accounting for Management Control: An Introduction, (New Jersey: Prentice-Hall Inc., 1965), p.5-6.

[4]Anton, op. cit, p. 520

[5]Ibid.

[6]Richard Mattessich, Accounting and Analytical Methods, (Homewood, Illinois: Richard D. Irwin, Inc., 1964).

establish their applicability for generating different types of information for various purposes. One statistical technique which this writer believes may be appropriate is multiple regression. Although this technique may have broader implications, interest here will be centered on its applicability in generating budget information.

Preparation of budgets involves managerial accountants in virtually all larger firms. The accountant may be required to prepare, or assist in the preparation of, sales and cost forecasts, and to assist in establishing acceptable performance ranges. One of the more difficult budgeting problems for accountants in manufacturing firms is that of forecasting and suggesting control parameters for overhead costs. It is within this context, and in view of the need of additional information, that the applicability of multiple regression analysis will be examined.

Objectives

The specific purpose of this dissertation is to explore possible practical applications of multiple linear and multiple curvilinear regression and correlation techniques in forecasting and controlling overhead costs in an actual manufacturing firm having multiple production and service departments. If it is concluded that valid estimating equations can be determined, then it should be possible to develop mathematical models simulating respective overhead cost functions. If these models are determined valid, they may then be integrated into a computerized accounting system as an overhead cost model and employed to generate cost forecasts for planning and cost control.

Conversely, it may be found that multiple regression and correlation is not a practical technique to forecast and assist in the control of factory overhead costs. Under these circumstances it would be hoped that the reasons "why" such a situation existed could be determined. It is quite likely that valid reasons for such a condition could be isolated. Thus it would be the intent of this work to stipulate the amount, type, and form of data which would be necessary for valid regression analysis.

The regression and correlation tools to be employed have been in existence for a long time and articles have been written setting forth conceptual applications relating to accounting information.[7] Robert A. Knapp, one of the earlier advocates of regression and correlation as a practical method for generating accounting information, says of this technique and its practicability:

> Correlation analysis has been described as one of the most powerful tools available for business analysis, planning, and forecasting. It is in general use by econometricians, operations researchers, and management consultants; however, there is too little evidence of its day-to-day use as a tool for forecasting, or for budgeting and measuring the operating elements, the assets and the liabilities of business.[8]

Without question there is conceptual precedent supporting the proposed applications for multiple regression analysis. Now needed is an experiment in reality. It is the writer's view that an application of multiple regression analysis, using the data of an actual manufacturing firm and

---

[7]See for example: George J. Benston, "Multiple Regression Analysis of Cost Behavior," Accounting Review, XLI, No. 4 (October, 1966), p. 657-672.

[8]Robert A. Knapp, "Forecasting and Measuring with Correlation Analysis," Financial Executive, XXXL, No. 5 (May, 1963), p. 13.

coupled with a detailed report of the results obtained, would comprise a significant contribution to the field of managerial accounting.

The Contemporary Situation

Contemporary manufacturing firms, in general, have attained a reasonable level of sophistication in accounting for the direct materials and direct labor costs associated with product production. One of the most sophisticated techniques employed in accounting for materials and direct labor costs is standard costs. If the standard unit costs for these resources have been properly developed, it is likely that their costs can be accurately accounted for, forecasted, and controlled. But this is not the case for overhead costs.

Larger manufacturing organizations employing either a job-order or process-cost system are usually segregated into a series of production departments and a series of service departments. Both the production departments and the service departments incur direct overhead charges for such items as indirect labor, supplies, and depreciation. In addition, when a step method of reapportionment is utilized, all the production departments and many of the service departments must absorb indirect overhead costs. Eventually, of course, these overhead charges are attached to the units produced in the production departments.

Because overhead charges contain such a vast array of different costs, from different sources, standard cost techniques are not appropriate, except under extremely stable operating conditions. Such a condition of stability is indeed unusual in the present day economy. Thus, difficult and in many cases relatively insurmountable problems are encountered in forecasting, controlling, and allocating overhead costs. Unfortunately the overhead costs associated with producing a given product or a group of products are increasing faster than those for direct materials and labor. For many firms, it is not unusual for 40 to 60 per cent of the total production cost to be composed of overhead costs.

Management and accountants fully agree that an information vacuum now exists in planning and control for overhead costs. But this same vacuum also exists in the other non-production segments of manufacturing firms. These problems exist primarily because of the complex relationships and interrelationships involved in the activities of the producing and nonproducing segments of the firm. Therefore, if management and accountants better understood the relationships existing between the various levels of output and the required services of the indirect cost elements, they could better plan and control accurately the costs attaching to these elements.

Perspective

It is because the fundamental characteristic of regression and correlation analysis is one of defining and measuring relationships between variables that its potential applicability in a business environment is being explored. As noted earlier, conceptual applications of the tools have been previously advocated. However, these applications have been oriented toward determining a total product cost or toward developing a financial budget as opposed to an operating budget. This writer is not cognizant of any previous studies, of this type, relating to either conceptual or practical applications of regression and correlation in planning and controlling specific overhead cost elements.

The primary orientation of this dissertation is toward the development of techniques which may be used to obtain practical solutions, in a practical way, to ordinary business information problems. To sustain this purpose, it was necessary to develop the procedures in a realistic environment and such an environment could not properly be simulated. Therefore, an intermediate sized manufacturing firm was selected. Because of the confidential nature of the data employed in subsequent chapters this real and very cooperative firm requested that a hypothetical name be contrived. Therefore, the fictitious name of Grace Company will be employed. A second request made by the Grace Company was that no actual or simulated operating data be presented. Thus, no schedule or exhibits of data appear in any of the succeeding chapters. But numerous charts representing actual operating costs are presented. These charts, developed for the purposes of better understanding the relationships between the variables employed, are an integral part of the analysis. But again, in compliance with the company's request, no dollar or unit scales have been shown on the charts. The source of the data presented on the charts is in every case the general ledger, factory ledger, or some integral record source within the Grace Company.

The effect on the structure of this dissertation, resulting from the fact that an actual case firm is providing all empirical data, is profound. The source, in fact, literally dictated the organizational structure of the work and influenced the timing of many procedures. Although the use of an actual company compounded this writer's difficulties considerably, employment of simulated data would not have been consistent with the objectives in this dissertation.

Cost Model of the Grace Company

The Grace Company produces two products and several models of each. The company is organized into four broad cost sections: (a) shipping and warehousing, (b) manufacturing, (c) sales, and (d) administration. Standard costs are employed in accounting for the direct labor and materials costs. Overhead is attached to production on the basis of direct labor dollars. A predetermined overhead rate is employed; however, the company has found it necessary to adjust this rate frequently. No standard costs or any similarly related procedures are employed in planning and controlling shipping and warehousing or selling costs. Budgets for these cost segments are based on experience ratios tempered by the company's expectations of future activities and costs.

In view of the objectives of this dissertation, a three-segment cost model of the Grace Company will be developed, having the following general form for planning and control.

Manufacturing:

- Direct materials - standard costs
- Direct labor - standard costs
- Factory overhead costs:
  - Fixed costs - conventional estimating methods
  - Variable and semi-variable costs - estimating equations to be determined.

Shipping and Warehousing:

- Fixed costs - conventional estimating methods
- Variable and semi-variable costs - estimating equations to be determined.

Selling:

Fixed costs - conventional estimating methods
Variable and semi-variable costs - estimating equations to be determined.

If appropriate and valid estimating equations can be developed for the variable and semi-variable cost elements contained above, then in effect, a sophisticated model for forecasting and controlling costs of the Grace Company will have been generated. Further, it will have been shown that, in this instance, regression and correlation are significant and practical tools for the accounting analyst.

Structure of the Dissertation

In Chapter II the assumptions underlying regression and correlation will be presented. The intent here is to give the reader an understanding of the tools that will be employed subsequently and a basis on which an evaluation of their practical applications may be made.

The second section of this chapter is devoted to pointing out and explaining some of the many difficulties that may be encountered when attempting to employ correlation analysis in a business environment. Some of the problems discussed in this section include multicollinearity, autocorrelation, those encountered in time-series analysis, and those resulting from errors of measurement. The purposes here are to determine, examine, and understand the problems that may be encountered when time-series data are employed in the analysis, and to illustrate techniques for eliminating certain types of errors.

Finally, the last section of Chapter II contains a brief discussion of contemporary standard cost accounting systems, how they are developed and their purpose. The intended contributions of this section were general in character. Most contemporary business firms consider a standard cost system to be of a relatively high degree of sophistication. Therefore, this discussion is made to indicate the arbitrary and perhaps subjective manner in which overhead cost rates are often derived for inclusion in standard costs. In addition, since many firms utilize standard costs systems to differing degrees, the reader could better determine the types of information that might be generated by such a system and that could be used in correlation analyses. Finally, it is hoped the discussion will indicate the potential contribution that correlation analysis may make toward improving the validity of overhead cost rates in a standard cost system.

Chapter III will be the first of five chapters required for the analysis of the Grace Company's shipping and warehousing costs, sales department costs, and the manufacturing overhead costs. The shipping and warehousing cost will be subjected to analysis in Chapter III. The general approach will be to review the function of the department in an effort to ascertain why costs are incurred and to provide insight into the most likely types of explanatory variables. The accounts employed for cost collection will be presented and reviewed as to their nature and accuracy. Based on these detailed account reviews, a group of possible explanatory variable series will be selected. After this phase is completed, actual computations will be commenced with a view toward developing appropriate estimating equation(s) for the department's costs. Each estimating equation will be evaluated. If it is judged both valid and efficient, it will then be employed to restate the historical series used in its computation. Provided the estimating equation is capable of restating past experience in a prescribed manner, it will be accepted as a cost model for the dependent variable cost series.

The sales department costs will be reviewed and if appropriate subjected to regression and correlation analysis in Chapter IV. The objective in this chapter is to develop a cost model of the sales department costs. The approach will be similar to that of Chapter III in that the initial phase will involve determining the functions performed in the department and reviewing the accounts employed to collect the associated costs. Based on the reviews of the functions and accounts, several functional cost classes will then be examined in depth and where appropriate, estimating equations will be developed and tested.

The manufacturing overhead costs will be analyzed in Chapters V, VI, and VII. Because of the numerous and varied types of costs included in this section, it will be necessary to organize and develop a systematic pattern of analysis. This will be the objective in Chapter V. Here the classification and reporting structure employed by the company will be examined to determine its relevancy for purposes of regression and correlation analysis. Each of the accounts comprising the overhead cost sector will be superficially examined as to the services provided and its cost behavior over time.

This examination will be used as a basis for classifying each cost as fixed, immaterial, stochastic, non-operating, or variable. All costs, except those classified as variable, will be examined subsequently in depth and, if the initial classification remains valid, will be excluded from further analysis. In the final section of Chapter V, those costs initially classified as variable will be further sub-classified into one of three categories. These sub-classifications will be made on the basis of the respective costs' apparent sensitivity to variations in production volume. By doing this it should be possible to select several explanatory variables that would be closely related to each of the accounts in a particular category. Further, it is expected that the analysis problems encountered should be similar for all the accounts in that respective category.

In Chapter VI those overhead cost accounts demonstrating the greatest sensitivity to direct variation in production volume will be subjected to regression and correlation analysis. Several possible explanatory variable series will be presented and reviewed. The overhead cost accounts will then be examined in depth and subjected to regression analysis. As efficient estimating equations are developed, they will be tested for their ability to restate past cost experience. Then if deemed valid, the estimating equation will be accepted as a cost model of its respective dependent cost variable.

Those costs considered of the second order in their degree of sensitivity of variation in production volume will be analyzed in the first section of this chapter. In the main, the objective in this chapter is the same as before, to develop valid and efficient estimating equations for each overhead cost for which it is deemed appropriate.

Although the primary objective in Chapters III through VII is to develop efficient estimating equations to serve as cost models, there is a secondary and perhaps as important an objective. That is, development of a particular cost model for the Grace Company is not vitally important; the lessons taken in its development are important. Each cost account represents a new challenge with different problems for which solutions must be determined and their relationships to the dependent variable measured. By fulfilling the primary objective, it may be possible to determine the applicability and practicability of regression and correlation analysis in the business environment.

Chapter VIII consists of three sections, the first of which contains an examination of the relative merits of regression and correlation in forecasting various costs. But the relative merits of a particular tool can be determined only in light of the characteristics of other tools used for the same purposes. Therefore, several conventional methods of forecasting costs will be reviewed. Comparisons will be drawn between the forecasting capabilities of these conventional methods and those of regression and correlation. In addition, a complete cost forecasting model for the Grace Company will be presented.

The characteristics of estimating equations developed from multiple linear and multiple curvilinear regression and correlation, that should be of assistance in cost control, will be discussed in section two. Here the possible impacts of a realistic forecasting technique on the attitudes of supervisory personnel and how this relates to cost control will be examined. Finally, the standard error of the estimate will be placed in a control context and several of the implications will be discussed. In the third section of Chapter VIII the limitations of regression and correlation techniques for forecasting and controlling costs will be reviewed.

Chapter IX will be comprised of the summary and conclusions. The hypothesis is that multiple linear and multiple curvilinear regression and correlation are valid techniques for improving the accuracy of overhead and other indirect costs forecasts and, in addition, represent significant improvements in cost control procedures. It is hoped that by the conclusion of Chapter IX considerable light will have been shed on the validity of this hypothesis.

CHAPTER II

RESEARCH METHODOLOGY AND THE CONTEMPORARY INFORMATION ENVIRONMENT

Introduction

To achieve the objectives previously set forth in Chapter I, research was conducted on two fronts. First, a thorough review of library materials was conducted. Emphasis was placed on materials providing a strong basis in cost accounting theory, on the techniques of employing standard cost systems, and on existing methods of analyzing cost information for management purposes. Cost accounting texts, periodical articles, and research studies were perused for related materials. Concurrently, library research pertaining to multiple regression and correlation and curvilinear regression and correlation techniques was conducted.

An actual case study was then carried out in conjunction with the Grace Company personnel. Actual operating data from one of the Company's plants was utilized in the models. The models were developed and tested on computers made available by the University of Texas at Arlington.

Quantitative Tools Employed

In order to develop valid cost models, capable of describing and perhaps predicting the cost performance of a service department and also to provide a measure of the probable accuracy of the cost models, it was necessary to utilize multiple regression and correlation analysis, and in some cases, multiple-curvilinear regression and correlation methods. Actual techniques and procedures applied in using these tools will be developed in subsequent chapters of this dissertation. However, because of the possible relative unfamiliarity of some readers with the concepts of regression and correlation, the lack of prior applications of the tools in this area, and since they are an integral part of this dissertation, a discussion of the tools is necessary.

The Assumptions Underlying Regression and Correlation

The regression technique employed in this paper is often termed "the least squares method." Several assumptions which underlie the method will be briefly discussed. The purpose is to provide a theoretical basis for the evaluation of techniques employed in subsequent chapters. Further, it is vital that the implications for the validity of parameter estimates be known when the basic assumptions are not appropriate for a specific data population.

To provide perspective for the assumptions, a hypothesis will be stated to the effect that a linear relationship exists between the variables (Y) and (X). The population equation form is stated as

$$Y_i = a + bX_i + u_i \qquad i = 1, 2, \ldots, n.$$

The constant term a and the regression coefficient b are the unknown population parameters. The disturbance term u_i is a random variable. The equation used to estimate the hypothesized relationship is written as

$$Y_c = \hat{a} + \hat{b}X$$

where $\hat{a}$ and $\hat{b}$ are estimates of the true population parameters. The Y_c term is the computed value of Y_i, given X_i and the parameter estimates $\hat{a}$ and $\hat{b}$.

The necessary assumptions are:

$$E(u_i) = 0 \quad \text{for all } i$$

$$E(u_i u_j) = \begin{cases} 0 & \text{for } i \neq j;\ i, j = 1, 2, \ldots, n \\ Q^2 & \text{for } i = j;\ i, j = 1, 2, \ldots, n \end{cases}$$

X is a set of fixed numbers

X has rank $K < n$.

The first assumption states that the u_i are random variables with an expected value of zero. The second assumption is a double assumption and states that the u_i are uncorrelated or have zero covariance for $i \neq j$, and that the u_i have constant variance Q^2 for $i = j$. The third assumption is that X is a set of nonrandom fixed numbers. The fourth assumption states that the number of X variables employed must be less than the number of sample observations and that no exact linear relations exist between the X variables.[1]

The foregoing assumptions provide the least-squares estimating equation with several interesting properties. The first property of interest is the character of the $\hat{a}$ and $\hat{b}$ terms. Note that Y_i is treated as a random variable and the independent variable X_i is a constant. Thus repeated samples might result in an infinite number of Y_i values for a constant X_i value.

---

[1] J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company, Inc., 1963), p. 106-107.

For each Y_i value an $\hat{a}$ and $\hat{b}$ could be computed so there would be a sampling distribution for each of these parameters which are themselves means. Because of the first assumption of $E(u_i) = 0$ the regression line passes through the means of the sampling distributions of the Y_i values. Hence $E(\hat{a} + \hat{b}X_i) = a + bX_i$, since $\hat{a}$ and $\hat{b}$ are unbiased estimates of a and b respectively.[2]

Because of the second assumption of zero covariance and constant variance with a zero mean, the $\hat{a}$ and $\hat{b}$ parameters are the minimum variance, linear, unbiased estimators of a and b.[3] Observe that these properties do not require that the distribution of the u_i values for any given X_i be normal. However, it is necessary that some form of probability distribution be assumed when significance tests are being made during regression analysis.

Problems Inherent in Regression and Correlation Analysis

The theory of regression and correlation analysis was developed within a laboratory environment where conditions could be altered in accordance with the wishes of the researcher or mathematician. Although correlation techniques have been widely accepted and applied in the natural sciences, they have not been so readily accepted in the social sciences and even less by the business community. Perhaps some of the reluctance, on the part of business to accept correlation as a practical tool, may be attributed to its complexity, but this is probably not a major factor. Instead, most arguments against the applications of correlation in business

---

[2]Ibid., p. 9-17.

[3]Ibid., p. 9-17.

stem from (a) the problems characteristic to and inherent in correlation, and (b) the fact that business data are quite unstable when considered in relationship to the phenomena in nature often termed "laws." After all, business data exist because of decisions and actions made by man himself. Consequently, he may completely alter any of these conditions he originally established literally over-night. To the extent that management continues to follow the same paths over an extended period of time, correlation is applicable; but if management alters its direction significantly, correlation techniques may become worthless, unless new stability is established.

The assumptions which underlie regression and correlation were mentioned previously. In view of these assumptions it is now appropriate to examine some of the complexities involved in applying these same tools in the analysis of business data. Perhaps it will become apparent why these tools have not been widely applied for business data analysis.

Cause and Effect

For purposes of illustration, assume a hypothetical linear relationship employing two independent variables that may be expressed as $Y = a + bX_1 + cX_2$. Now, the assumption was made that the value of Y is a function of the values of X_1 and X_2. It may also be noted that the values of X_1 and X_2 are determined independently of Y. However, these considerations may not hold true in a particular set of business data, and yet a high coefficient of multiple determination may be derived. A high coefficient of determination does not necessarily prove that a cause and effect relationship exists. The relationship may exist because Y, X_1, and X_2 all fluctuate with a fourth variable that may not even be contained in the equation![4] The relationship may exist because the three variables may in fact be interdependent rather than X_1 and X_2 being independent.[5] Finally, the apparent relationship may be attributable just to chance.[6] Consequently, it is necessary to examine the relationships between the variables to determine that they are logical relationships.

Multicollinearity

A second problem often encountered pertains to the relationship between the two independent variables X_1 and X_2. Many times an interrelationship exists between the independent variables which is neither apparent nor logically obvious to the researcher prior to computation of the regression equation. In fact, the relationship may be so indiscernible that the researcher may not even be cognizant of its existence. Confusion on the part of the researcher may then occur when he attempts to evaluate the significance of each partial coefficient of determination in light of the value obtained for the multiple coefficient of determination. For example, a test of the partial coefficient of determination for X_2 may indicate that its contribution was significant. Yet, if X_2 were deleted from the estimating equation, it might be found that the decrease in the multiple coefficient of determination was definitely insignificant. However, interdependence among the independent variables may not always be reflected in such a manner.

---

[4] Frederick C. Croxton and Dudley J. Cowden, Applied General Statistics (2nd Ed., New Jersey: Prentice Hall Inc., 1960), p. 470.

[5] Ibid.

[6] Ibid.

"When the explanatory variables are highly correlated with one another, it is very difficult, and often impossible, to estimate the separate relationships of each of the dependent variables. This condition is called multicollinearity."[7] However, high correlation between two or more independent variables does not necessarily prove interdependence, but it may make any interpretation of $R^2_{1.23}$ more difficult. For this reason, efforts must be made to determine if correlation does exist among the independent variables.

The existence of multicollinearity may be determined in some cases by examining the various explanatory variable series. If two or more of the series appear to move together, they are probably correlated. A second method is to develop an estimating equation by actually correlating two explanatory variables with one another. A third approach is to examine the standard errors of the regression coefficients. In the case of high multicollinearity between X_1 and X_2 the standard errors may actually dominate the respective regression coefficients even though $R^2_{1.23}$ may be relatively high. In general, it appears that as multicollinearity increases so does the magnitude of the standard errors, as related to the regression coefficients. Finally, when the X_1 and X_2 series are perfectly correlated the mathematical requirements are such that the regression coefficients may not be computed.

What are the implications of multicollinearity for business data analysis? The state of the art is such that no clear answer is available.

---

[7] George J. Benston, "Multiple Regression Analysis of Cost Behavior," Accounting Review, XLI, No. 4, (October, 1966), p. 666.

However, it appears that the existence of multicollinearity will influence the size of the standard errors of the regression coefficients. If this occurs, the validity of the regression coefficients as estimates of marginal cost would seem questionable. On the other hand, if one desires to employ the estimating equation as a whole for forecasting, "then intercorrelation of explanatory variables may not be too serious, provided it may reasonably be expected to continue in the future."[8]

Autocorrelation

A third problem often encountered when conducting regression analyses of business data is autocorrelation. If the successive observed values of any variable X are correlated, it may result in an autocorrelated disturbance term. Autocorrelation of the disturbance term is often incurred when annual observations are employed but it would seem to be an almost inescapable condition when monthly data are being used. For example, the direct labor cost for any single month is probably a function of the immediately preceding month's total direct labor cost rather than just the result of the volume of output for the current month. Complete independence between the monthly observations of such business data should not be expected to exist under normal circumstances. Therefore, an autocorrelated disturbance term may be a highly probable condition.

"The main point is that autocorrelated series give us less information per observation than do completely random ones."[9] This point leads one to

---

[8] Johnston, op. cit., p. 207.

[9] Mordecai Exekiel and Karl A. Fox, Methods of Correlation and Regression Analysis, II (New York: John Wiley & Sons, 1959), p. 334.

believe that it might generally be necessary to obtain a larger number of observations for business data analysis than would normally be required, but this may not be possible as typically there are only a limited number of appropriate observations. The consequences of this reduced information state for each observation in the use of ordinary least-squares may be: (a) an underestimate of the sampling variances of the regression coefficients, if the usual least-squares formulas are employed and (b) the usual t and F tests of significance may not be valid since the true number of degrees of freedom may not be known. The extent to which each of these effects may actually be manifested in a given estimating equation is not easily determinable.

It is obvious that autocorrelation may be a serious problem, but its severity is related to the purposes for which the estimating equation is being developed. If one intends to use the estimating equation for forecasting only, then limited autocorrelation may not be a major problem. But if one desires to employ the regression coefficients as marginal costs, then autocorrelation may be a major problem. The usual least-squares method will provide unbiased estimates of a and b, but their associated sampling variances may be underestimated if the disturbance terms are autocorrelated.[10] This condition may place in question the validity of the regression coefficents as marginal costs.

There are tests which may be made to determine if autocorrelation does exist. Two of these tests are the (a) determination of a coefficient of

---

[10]Johnston, op. cit., p. 187-188.

autocorrelation and (b) the von Newman ratio technique.[11] Once it is determined, by one of these methods, that ..."a given time series is characterized by autocorrelation, can it be removed? Several methods have been suggested in the literature on this problem, but often their validity in practical situations is doubtful."[12] One method of eliminating autocorrelation in the residuals, the one that will be used in subsequent chapters, is to employ a variable in the regression equation representing time.[13] Hence the multiple regression equation would be written as

$$Y = a + bX_1 + cX_2 + dX_3 + u$$

where X_3 denotes time and the disturbance term u is assumed random. However, this method may or may not eliminate the serial correlation in the X variables. Thus, "a regular pattern may remain in the residuals so that autocorrelation persists."[14]

Autocorrelation will almost certainly exist in the data to be analyzed in the following chapters. It is hoped that much of its effects will be eliminated by the foregoing method of including a time variable in the estimating equation. To undertake other steps of coping with the problem would require considerable time, time which a company's budgeting executive may not have available. However, steps will be carried out in subsequent

---

[11]Ezekiel and Fox, op. cit., p. 135.

[12]Michael J. Brennan, Preface to Econometrics, (2nd Ed., Cincinnati: South-Western Publishing Company, 1965), p. 311.

[13]Ibid., p. 359-360.

[14]Ibid., p. 361.

chapters wherein the computed estimating equations will be used to restate past experience. This technique should be helpful in determining if the proposed remedial procedure has been effective in eliminating autocorrelation.

In the preceding pages of this section several points that may create confusion or lead to incorrect conclusions in the process of evaluating an estimating equation have been discussed and, it is hoped, clarified. It was first stated that even though a high coefficient of determination may be computed for a given estimating equation, it does not necessarily prove a cause and effect relationship exists between Y and $X_{1\ldots,n}$. Such a relationship can only be established on the basis of logic.

The second potential problem discussed was multicollinearity. This problem was said to exist when two or more of the explanatory variable series were correlated. Methods of determining when multicollinearity existed were discussed and its associated implications for the validity of the estimating equation. Finally, the most difficult problem of autocorrelation was discussed. It was pointed out that autocorrelation almost certainly exists in time series data and that its existence in the data violated one of the basic assumptions underlying regression analysis. Although measures of the degree of autocorrelation are available and though conceptual techniques have been developed for eliminating autocorrelation, these procedures are not always valid.[15] Indeed autocorrelation may remain as a sticky problem. On the other hand, it was stated that the consequences of autocorrelated residuals might not be as severe if the estimating

[15] Ibid., p. 359.

equation was to be used only for forecasting purposes. Further, these consequences might be mitigated by the inclusion of a time variable in the estimating equation.

A second group of problems exists in correlation analysis of business data, because of the basic character of business information. That is, decisions are made and actions are taken, the results of which are expressed over succeeding time periods in dollar terms. Consequently, the researcher is faced with the difficulties of time series analysis and perhaps, the problem of stabilizing the factor in which his data is often stated--the dollar.

Price Level Changes

Information expressed in current dollar terms is quite hazardous to work with if the time period covered is more than a few years. The difficulty here is that the base of the unit of measure changes because of price changes. The effects of this on the standard error of the estimate, the coefficients of determination and partial determination, are often difficult to perceive. Unless the data is restated for price level changes, so that it is in "common dollar" terms, it is not possible to determine how the coefficient of determination has been affected by changes in the price level.

In order to understand this issue clearly, assume a hypothetical case wherein a need exists to correlate sales salaries to net sales over a five-year period. Sales salaries are computed by multiplying the number of salesmen times their respective fixed salaries. It is management's desire that sales salaries should not exceed some fixed percentage of dollar sales. Therefore, total sales salaries is the dependent variable Y, and total net sales is the independent variable X. Assume that on the average for the five-year period, sales salaries have increased 4 per cent a year. Assume further that the number of salesmen has not changed. If sales have also increased an average of 4 per cent a year over the same period, then the coefficient of determination may be quite high. Now change the situation and assume that units sold are used instead of sales dollars. If the average increase of 4 per cent per year in total net sales is due entirely to price increases, reflecting inflation, and if no increase in the volume of units sold has occurred, then the coefficient of determination will be lower. It may then appear that salaries had gone up unjustifiably. In this instance one variable, unit volume, which has a stable measuring base, is being correlated with a second variable, total sales salaries, which does not have a stable measuring base. However, if the price level had increased an average of 4 per cent a year over the five-year period and the sales salaries were adjusted for this change, then the correlation between the two variables might be improved considerably. Therefore, when correlating two or more time series, if one or more of the variables are expressed in other than dollar terms, it may be necessary to adjust the variables expressed in dollar terms for price level changes. Conversely, such adjustments may, on occasion, result in poorer correlation. A second conclusion that might be drawn is that if all the variables are expressed in dollar terms, it may not be necessary to adjust for price level changes since the variables would be expressed in similar dollars.

There is, of course, always the question of whether or not adjustments should even be made for price level changes. If one is interested in the number of dollars relationship as opposed to the dollars of purchasing power, then adjustments may not be necessary. This type of situation is probably the case in most budgeting problems. However, if price level changes have occurred over the period of analysis, it is likely that restating the data in the form of constant dollars will improve the fit. This is particularily true if one or more of the variables are not expressed in dollar terms. If the dollars are adjusted for price level changes, then the most current price level should be used. Such a procedure should yield a more accurate forecast of the unadjusted dollars involved in future activities.

Time Series Factors

Time series data contain such factors as trends, cycles, and seasonal movements, and of course when correlation is involved, autocorrelation may be a factor. The considerations applicable to autocorrelation were discussed previously; therefore, in this section some of the implications for regression analysis of trend, cyclical, and seasonal movements in the data will be considered.

In conducting regression and correlation analyses the researcher must be careful to include a sufficient number of observations for each variable, but it is also possible to include too many observations. For example, if the monthly maintenance costs, for a certain firm, were to be correlated with the machine hours, it might be necessary to take only 24 to 36 observations (2 to 3 years) in order to get an approximation of what the maintenance costs might be during the next 6 months. To collect data originating farther back in the past might be to include data in the analysis that were not relevant. This would be particularily true if a large degree of technological progress had occurred within the firm or perhaps even if significant change in the product mix had taken place. Note that the period of time covered in the maintenance cost analysis was only two to three years. This may be contrasted to the normal period of time employed in many time series analyses of 15 to 30 years. Over such an extended period of time the elements of trend, cycles, and seasonal fluctuations may be reasonably well isolated. But this may not be the case over a two or three-year period of time. For example, the two or three-year time period covered in the correlation analysis will probably not be sufficiently long to encompass the completion of one or more cycles. Of course, the time required for several cyclical movements to occur will depend on the industry itself and the economy in general. Unless one or more complete cycles are present in the time series analyzed, it is probably not necessary to explicitly consider the implications of cyclical movements on the correlation results. But the researcher must be cognizant that the phase of the cycle contained in the computation may inappropriately influence the estimating equation.

"However, when correlating time series, we must take cognizance of the fact that trend is usually present in annual data and that both trend and seasonal variation, as well as irregular fluctuations, may be found in monthly data."[16] If time series data are correlated but are left unadjusted for trend and seasonal variations, the resulting coefficients reflect the composite relationships between the fluctuations in the variables themselves and their seasonal and trend variations. This results in a "masking" of the correlation coefficients' true values. The effects on the values of the coefficients may be quite pronounced and in differing directions, depending upon the trends in the data. The four possible effects follow:

> 1. The fluctuations of two time series may be positively correlated, but the trends may be in opposite directions. Correlating the data without adjusting for trend...will result in lowering the positive correlation coefficient or may even change it to a negative coefficient, if the trends are marked in relation to the fluctuations.
>
> 2. The fluctuations of two time series may be positively correlated, and the trends may be in the same direction. Correlating the data without adjusting for trend...will result in increasing the positive correlation coefficient.
>
> 3. The fluctuations of two time series may be negatively correlated, but the trends may be in the same direction. Correlating the data without adjusting for trend...will result in lowering the negative correlation coefficient or may even change it to a positive coefficient if the trends are pronounced in relation to the fluctuations.
>
> 4. The fluctuations of two time series may be negatively correlated and the trends may be in opposite directions. Correlating the data without adjusting for trend...will result in increasing the negative correlation coefficient.[17]

Since the effects of not eliminating trend on the correlation coefficients may be quite significant, provisions must be made to eliminate

---

[16]Croxton and Cowden, _op. cit._, p. 562.

[17]_Ibid._, p. 567-569.

the possible bias. "The effect of trend may be eliminated by correlating percentages of trend instead of correlating the raw data."[18] A second method of eliminating the effect of trend is to add a third independent variable to the hypothetical equation, making it

$$Y = a + bX_1 + cX_2 + dX_3 + u,$$

where the variable X_3 represents time. By employing a variable for time in the estimating equation, one is able to derive a partial coefficient of determination that reflects the correlation between the variables Y, X_1, and X_2, when time (X_3) is, at least conceptually, held constant. The second approach will be employed in this dissertation.

Seasonal movements remain a consideration in that if their effect is not eliminated the resulting correlation coefficients may be biased. Seasonal effects may be removed by incorporating variables in the equation representing the respective seasons. By removing seasonal effects as well as those of trend, the correlation coefficients should properly reflect the relationships between the various independent variables and the dependent variables.

The technique used to eliminate the effects of seasonal movement is analogous to the method employed to remove the effects of trend. The difference is that several seasonal variables may be needed rather than just one, as in the case of trend. For example, four dummy seasonal variables might be used where:

S_1 = 1 for November, December, January, and February, and it will have a value of zero for all other months.

---

[18] <u>Ibid</u>., p. 563.

S_2 = 1 for March and April, and it will have a value of zero for all other months.

S_3 = 1 for May, June, and July, and zero for all other months.

S_4 = 1 for August, September, and October; for all other months it will have a value of zero.

Note that the dummy seasonal variables employed do not coincide with the conventional seasons of spring, summer, fall, and winter. The seasonal variables employed may be determined by reviewing the seasonal fluctuations contained in the data being analyzed. Needless to say, it may be difficult to determine the proper seasonal variables for some firms. Therefore, it may be necessary to use an iterative process in order to determine the proper time periods to be associated with each of the seasonal dummy variables.

The primary purpose of determining the timing and magnitude of seasonal variation is to eliminate its effects from the data. Thus the sampling variance of Y may be reduced, a process which in turn provides a smaller standard error of the estimate and a larger coefficient of determination. Conceptually, seasonal variation may be accounted for by the use of dummy variables, but in actual practice, this is an extremely difficult procedure. One reason it is difficult is that no absolute separation appears in most data between any two seasons. Second, within the same firm different seasonal patterns may be associated with the various elements or activities. Further, variations that appear to be seasonal variations may occur in the data when in reality some other factor may be causing the variation. For example, many firms employ a 13-week quarter for accounting purposes. This means there are two four-week months and one five-week month in each quarter.

Further, these accounting months do not correspond to the related calendar months. Given the case of a fluctuating variable and all other things being held constant, the accounting month having five weeks will probably contain the largest number of transactions, dollars, units, or hours. Other biasing elements such as holidays, vacations, and retooling periods may be introduced. In such cases the inclusion of a number-of-working-days variable or a working-day index may be advisable, depending on the type of variables being used in the analysis.

There may be occasions when it is not as necessary to adjust for seasonal variation, as when all the variables in the analysis appear to contain similar seasonal variation. In fact, not to adjust for seasonal variation in such cases may be desirable. In some cases the seasons may vary in timing and duration from one year to the next or they may not be clearly definable. On balance, unless seasonal variation is extreme and reasonably constant in magnitude and timing over the period of analysis, it may not be necessary or practical to eliminate the effects of the seasonal variation from the data.

Time series data usually contain irregular data that are the results of unusual occurrences either within or outside the firm. Irregular items are of a non-recurring nature—at least in the short-run. An example might be the case of a firm that furnishes a car for its sales manager. If the car requires major repairs in any one month and if the repair cost is placed in the normal automotive expense account, then for that month an irregular item has been included in the selling expense data. In a second case, assume that several warehouse and shipping employees are required to work one full week each year taking a physical inventory count, but that their wages are not placed in a special account for the cost of inventory counting. Thus for that particular week the wages account for the warehouse and shipping employees would contain irregular data.

Obviously, the researcher must be careful to exclude irregular or irrelevant information from the accounts before undertaking regression analysis. Of course, it is much easier to say that all irrelevant items should be eliminated from the data than it is to remove them effectively. A complete analysis of the respective accounts, coupled with interviews of the proper personnel, should be helpful in this regard. The objective is to restate the original data so that they will contain only information which is of a continually recurring character.

Measurement Errors

A problem often encountered by the analyst preparing data for regression analysis, and a difficulty also quite often met by most budgeting personnel, is that of measurement error. It is likely that measurement errors of one type or another will appear in almost all business data. The possibility of clerical errors in restating data in preparation for analysis always exists, but aside from this type of error, there is always the possibility of error in the data itself. Obviously "the books must balance," but this fact alone does not preclude classification errors in recording expenses. Nor does it compensate for short-comings in the chart of accounts. For example, assume a firm employs four stockroom clerks on a monthly salary; two of the clerks forego their annual two-weeks' vacation and instead are paid double for the two-week period. Certainly such costs should be separately classified for purposes of correlation analysis. If the chart of accounts contains an account for vacation expense, then this account should be charged for the wages paid the two clerks. If it is not charged to the vacation expense account, then a clerical error has occurred. If the chart of accounts does not contain a vacation expense account, then the chart of accounts is deficient and the expense should be cast out. Aside from the system's deficiencies, great care must also be taken to discover and correct simple clerical errors.

Errors in measurement of the data for the dependent variable are critical and bias the degree of correlation. However, it is even more critical that errors of measurement in the data of the independent variables be eliminated. The reason is that if significant errors exist, the resulting coefficients for each respective variable are of little value in predicting costs. "If we are attempting to develop a model for budgeting purposes, it is vital that the marginal costs attached to the independent variables be accurate."[19]

The Number and Length of Time Periods

Two final considerations relevant to correlation analysis of business data remain to be discussed. The first of these deals with the "number of time periods" to be included in the analysis, and the second deals with the "length of the time periods."

---

[19]Benston, *op. cit.*, p. 665.

The number of time periods is important for two reasons. First, consideration must be given to the mathematical and statistical requirements of correlation analysis as reflected by the following.

> As a minimum, there must be one more observation than there are independent variables to make regression analysis possible.... Of course, many more observations must be available before one could have any confidence that the relationship estimated from the sample reflects the 'true' underlying relationships. The standard errors, from which one may determine the range within which the true coefficients lie (given some probability of error), are reduced by the square root of the number of observations.[20]

But how many more observations should be included in order to have any confidence in the resulting independent variable coefficients, the correlation coefficients themselves, and for conducting significance tests? There is no "absolute number;" rather, it is many times a matter of the number of observations available or of the practical advisability of selecting observations too far in the past. In other words, this is one of those points in correlation analysis where the judgment of the researcher is of primary importance. For example, this writer has encountered published research reports pertaining to business time series analysis that employed monthly data and that used as few as six observations with one independent variable and others using up to sixty observations. Assuming there is no real limit to the number of observations, the researcher would probably utilize at least eight or ten more observations than there are independent variables. The purpose, of course, is to have more degrees of freedom, which in turn may give the researcher more confidence in the results

[20] Ibid., p. 663.

he obtained. However, a word of caution is appropriate; confidence and sample size can be intelligently determined only in light of the variation in the population from which the observations are taken.

There is usually a practical limitation on the number of observations available when one is dealing with time series business data. As mentioned previously, it is important that antiquated data be excluded. Business data taken from too far in the past stand a good chance of not being relevant, because of changes in the technology of production, changes in product lines, and other changes in the cost structure, or because of the way in which functions are accomplished. Further, the estimating coefficients are in fact averages; so if they are being used to forecast future performance, it is likely that the more recent history would provide a better basis for prediction. In summary, if monthly data are being used, it is probably best to include only the most recent two or three years' data in the analysis or approximately 24 to 36 monthly observations. Again the cautionary "it depends" is relevant.

The <u>length</u> of the time periods employed is primarily a function of the firm's cost accounting system. In some instances it may be difficult to determine that a certain variable exists until it is viewed in the proper time perspective. For example, depreciation expense for manufacturing machinery, computed on an annual straight line basis, would appear to be a fixed cost with little variability, but this same expense, computed monthly, using machine hours, might literally "jump out" at the researcher as a variable.

Ideally the length of the time periods would be determined in the following manner.

> (a) The time periods (1, 2, 3, ...m) chosen should be long enough to allow the bookkeeping procedures to pair output produced in a period with the cost incurred because of that production....
>
> (b) The time periods chosen should be short enough to avoid variations in production within the period. Otherwise, the variations that occur during the period will be averaged out, possibly obscuring the true relationship between cost and output.[21]

In the case of most manufacturing companies, a month will probably be the shortest time period that may be used because of requirement (a) above. But, many firms do accumulate direct labor hours and units of production on a weekly basis, so it might be possible to utilize weekly observations in some cases. Again, the choice would depend upon the relationships that were being tested. A one-month period would probably be the maximum length of time that should be used in most analysis. Otherwise criteria (b) above would more than likely not be satisfied. It is possible that in some cases, as in the construction industry, quarterly data might be satisfactory, but in few situations would it be possible to use annual observations.

To conclude this discussion pertaining to many of the problems inherent in correlation analysis of business data, it should be noted that the researcher must be constantly aware of the numerous pitfalls and be ready to exercise his value judgments on many occasions. Admittedly, there are many difficulties encountered in correlation analysis, some of which require in-depth perceptiveness on the part of the researcher to recognize that they actually exist, but no insurmountable difficulties prevail. During the

---

[21] *Ibid.*, p. 663.

course of examining the remaining chapters of this dissertation, the reader should note how many of the problems discussed are encountered and subsequently resolved.

Accounting For Costs

Since in subsequent chapters of this dissertation suggestions will be made with the intent of improving the validity and value of cost accounting reports, it is probably appropriate at this point to discuss the contemporary state of the art. However, the discussion will not pertain to companies which produce primarily under defense contracts, companies having only one primary product, or companies involved in the extraction of natural resources.

Purposes of Cost Accounting

Cost accounting systems are maintained for a number of reasons. In this section, four general reasons will be presented, but they are not intended to be all-inclusive. The first major purpose is to assist in determining product costs. In a plant producing several different products, this determination has proved to be a practical impossibility if product cost is considered to be comprised of materials, labor, and overhead. Accountants and engineers together have generally been successful in establishing, with considerable accuracy, the cost of direct materials and direct labor embodied in a particular product. But the task of allocating that group of cost termed "overhead" has proved insurmountable. Obviously, estimating has been done and is being done by most firms, but probably no firm would argue that its method is absolutely valid. This aspect of cost accounting will be discussed at some length in a subsequent section.

The second purpose of cost accounting might be control. In its isolation from production, management must develop techniques for judging efficiency relative to output and direct costs. Control is exercised over materials, direct and indirect labor, and other resources through cost data. Control of these and other non-manufacturing costs is difficult, and it is to the topic of control that Chapter VIII will be devoted.

A third purpose of cost accounting is planning. Through past knowledge gained from cost accounting data and industrial engineering studies, management can estimate future costs. Budgets based on past experience and knowledge may be developed. These budgets reflect management's goals, serve as a control tool, as a method of communicating responsibilities, and as a basis for evaluating efficiency.

Fourth, cost accounting may provide data for special decisions that management is often called upon to make. In this group might fall make or buy decisions, capital investment decisions, and decisions either to continue or discontinue producing a certain product.

Of the reasons for maintaining cost accounting records, probably the second and third stand out as being the most successful in their present applications. On the other hand, cost control usually has been truly effective only in the production departments. Control over administrative and other overhead costs is usually much less effective, even where flexible budgeting, considered to be of a fairly high degree of sophistication, is employed. Cost data for planning suffers from many of the ills inherent in absorption product costing. Planning through budgets, even though such sophisticated devices as learning curves and simple correlation are employed, often leaves management in an embarrassing position. For often the future development does not result in the figures reflected in the budget because of the invalidity of the data on which the budget was based originally, or perhaps, because it was not updated to reflect subsequent management decisions, or because of management's inability to control effectively various costs. The latter reason is quite common to service department costs.

Desirable Characteristics of a Cost System

In the previous section it was stated that the general purpose for maintaining a cost accounting system is to provide product cost determination, planning, control, and evaluation of operating efficiency. For a cost accounting system to provide valid data for all these purposes, it must have several specific characteristics. Several of these characteristics have been aptly stated by Howard M. Carlisle as follows:

> Cost identification and control should be based on sound organization concepts and must be consistent with the authority relationships manifested by the formal organizational structure. Cost generated by each department or organizational unit must be made the responsibility of that department or organizational unit through proper coding and identification. ...The system should be flexible enough to permit cost to be collected for any special purpose or activity.... A system should be avoided that is rigid and not adaptable to special circumstances.... The cost accounting system, above all, must be consistent with the manner in which management estimates, directs, and controls programs.[22]

Therefore, it would seem a cost system should be capable of readily identifying and classifying all costs. These costs should be determined

---

[22] Howard M. Carlisle, "Cost Accounting for Advanced Technology Programs," *Accounting Review*, XLI, No. 1 (January, 1966), p. 119-120.

and assigned to the responsibility centers, which themselves must be consistent with the organizational intent of management. The system must be so flexible as to incorporate easily interim changes in the types of costs or in the organization of production. Finally, it would seem to this writer that the system should accumulate data by its behavioral character in order to provide information for control and evaluation, based on performance reports, and for special information needs, as in the case of pricing decisions.

Standard Costs--in General

Probably the most sophisticated cost accounting technique employed by most firms is a standard cost system. Standards are established for direct labor and direct materials for each product. Most firms are able to establish reasonably accurate standards for these two elements by employing time and motion studies for direct labor time and engineering-materials-use studies for direct materials.

The final standard for each product is expressed by production departments in dollar terms and reflects both the standard quantity and standard cost of each factor. Once these standards are developed, it is then possible to compute variances from standard. Variances will always exist but they nevertheless constitute a valuable cost control procedure, and the system itself provides a valid product cost for the quantities of direct labor and materials embodied in the final product.

Changes in the standards must be made as the elements relating to the variables comprising the standards change. For example, if the price of a basic raw material changes significantly, the material cost standard should be adjusted. Failure to adjust standards periodically will soon result in established standard costs becoming almost useless.[23]

The establishment of standard costs, by products, for overhead costs is a much more complex task. Even after overhead standards are established, their value may be quite questionable. Generally it is considered desirable to establish standard costs by production departments. Costs allocated to production departments may then be distributed to the various products passing through the department on the basis of direct labor hours or dollars, of machine hours, or of material costs.

Development of a Standard Overhead Rate

The first step in establishing a standard overhead rate is to determine the level or volume of production on which the standards must be based. It appears that most firms will have four basic levels of productive capacity on which the standard overhead rate may be based.

1. Rated Capacity is the rate of production that would result if equipment were operated continuously at the maximum output for which it was designed. This level of activity can be sustained only for short periods.[24]

2. Practical attainable capacity represents the full practical utilization of equipment that can be sustained under attainable operating conditions. In determining practical attainable capacity, rated hourly capacity is first reduced

---

[23]Adolph Matz, Othel J. Curry, and George W. Frank, Cost Accounting, (South-Western Publishing Company, Cincinnati, 1962), p. 588.

[24]"Research Report 39," Accounting for Costs of Capacity, National Association of Accountants, (New York: May, 1963), p. 21.

by allowances for predictable and unavoidable operating interruptions such as maintenance, repairs, set-ups, and operating personal time.[25]

3. Normal capacity for most firms is the rate of activity needed to meet average sales demand over a period long enough to encompass seasonal and cyclical fluctuations (often, three to five years). It is usually somewhere between 75 and 90 per cent of practical capacity.[26]

4. Scheduled production is the output currently planned or scheduled for a given unit of equipment or production center. It may be expressed as a percentage of practical attainable capacity as well as in units in which production is measured.[27]

Rated capacity is not a realistic foundation on which to base a standard. Nor is scheduled production a satisfactory basis since a standard should reflect expectations over a three to five-year period. However, if sales are expected to be reasonably constant over several years, then scheduled production may afford the most realistic basis. It may be that practical attainable capacity, normal capacity, and scheduled production are the same. However, it is generally considered best to base standards for allocating cost on either practical attainable capacity or normal capacity. Responsibility for determining capacity levels is normally an engineering function.

Admittedly, it is often difficult to determine capacity levels, but an estimate is absolutely essential if management is to develop valid plans. Further, this writer is of the opinion that management is in a better

---

[25] Ibid., p. 21.

[26] Charles T. Horngren, Cost Accounting - A Managerial Emphasis, (New Jersey: Prentice Hall Inc., 1962), p. 310.

[27] Ibid., p. 22.

position to develop plans to evaluate total revenue earned if practical attainable capacity is adopted as the overhead base. The reasoning underlying this writer's preference is that management should be fully cognizant of its organization's complete capacity to produce over a longer period of time. Estimated sales and production quotas for a three to five-year period, which serve as the basis for normal capacity, may be altered by making additional advertising expenditures, by introducing new products, or by any one of numerous other ways which do not require material additions to fixed assets. The financial reports should report a penalty when management fails to utilize fully its long-term capacity to produce. The use of normal capacity as a base often ensures that this possible shortcoming of management will not appear on the internal financial reports.

When the practical attainable capacity base is employed, the overhead rate will be smaller than if normal capacity is used. As a result overhead will generally be underabsorbed, this difference being the cost of capacity management failed to employ--or the loss from idle capacity. Preferably, management would hold this figure to a minimum and in an expanding economy should be expected to do so.

After an appropriate level of capacity has been established, the next step is to analyze overhead costs. The motive here, of course, is to enable the cost accountant to make an equitable and objective allocation of overhead department costs to the respective producing departments. The analysis first requires segregating accounts by their variability and by departments. That is, each cost account in each overhead department must be closely examined to determine if it is fixed, semi-fixed, or variable cost. This is an onerous task, usually requiring considerable time and effort. To complicate the analysis further, there may be a number of different overhead departments in the firm, and since each department has a different function, the nature of each department's cost accounts may vary.

In establishing standards for overhead rates through overhead cost analysis, it is not normally sufficient to look only at past cost experience. Historical costs may be used only as a guide to future overhead costs.[28] Given estimated future production and the capacity required, it is generally a fairly simple matter to determine what accounts will be fixed. For example, depreciation, fire insurance, and property taxes usually fall into this category. But, many types of costs fall into a semi-variable classification. The prime example here is indirect labor of various types. In the past one of the most sophisticated approaches to analysis of this type of cost has been made by using bivariate straight-line correlation.[29] The same method may be used for variable costs, as in the case of some types of supplies and power. Nevertheless, it should be noted that few strictly variable cost items are found in the overhead department's costs because of the very nature of overhead functions.

After the foregoing classifications are made, it may be possible to determine, within a reasonable range, the total cost of each overhead department when production is at the level of either practical attainable capacity or normal capacity. The next phase is to determine whether these

---

[28] Matz, Curry, and Frank, loc. cit., p. 542.

[29] Ibid.

costs are to be reapportioned by the step method or by the direct method. Conceptually and for reporting purposes, the step method of reapportionment is the more desirable of the two, but the net effect on total production cost will be the same under either reapportionment method.

Once the reapportionment method is decided, the next step is to determine the base on which the overhead is to be applied to each production department. This process in itself presents difficulties when there are several products and several producing departments. Several of the more common bases for reapportioning costs are direct labor hours, direct labor dollars, primary materials cost, machine hours, number of departmental employees, electrical power usage, and the square feet of floor space occupied. In any case, "...reapportionment should be based on the relative benefits received by the producing departments from their utilization of the various other departments' services."[30]

The overhead costs apportioned to the various production departments can then be attached to the units as they are produced or pass through the individual departments. Bases used to make this last allocation are standard direct labor hours or standard direct labor dollars. However, these are not the only bases that may be employed; rather, it depends on the functions performed on the product in the particular department. No single good basis may appear in a particular department; the basis which seems to be the lesser evil may of necessity be selected. Some flexibility is permitted in that the same basis does not have to be used in each production department.

[30]Horngren, loc. cit., p. 417.

For example, in one production department the basis might be standard machine hours, and in a second production department the basis might be standard direct labor hours.

A total standard cost for each product, by production departments, may be developed through the foregoing procedures. Obviously there are numerous variations to the procedures previously set forth. One such variation is to charge certain overhead costs, such as maintenance, directly to the production departments, based on maintenance employees' time tickets. Such costs might be termed direct overhead costs and thus would not be contained in the lump sum reapportionment of other non-direct overhead costs. Another variation occurs when a single standard overhead rate is employed to assign overhead costs to each product, regardless of which production department it passes through. This last method is widely used, even though it is less valid. The argument given to justify its application is practicability.

In the immediately preceding pages an attempt has been made to describe briefly the standard cost technique employed by contemporary cost accountants. Standard product costing, coupled with flexible budgeting, represents, more or less, the maximum effort that is presently being made by most firms. Indeed, these tools represent such an expense that few medium sized firms feel they can afford them. But assuming all larger firms employ such techniques to some extent, this writer is of the opinion that even these firms generally do not know the real cost of producing a specific product, nor do they have sufficient data to control properly either manufacturing or non-manufacturing overhead costs or to make many special decisions regarding such problems as new product development, product pricing, and whether to buy or make.

It should be observed that standard costs are not computed or developed for non-manufacturing costs such as those incurred in the administrative and sales sectors of the firm. Accountants have made significant contributions in developing procedures for management to employ in evaluating these types of costs; however, there is much room for improvement. Work has been done in this area but no techniques comparable to the standard costs used in production cost control have been developed. Consequently, management has few measures of efficiency which can be applied to control or reduce these costs. A vacuum of control procedures exists.

Conversely, the department supervisors in the administrative and sales sectors often find their departments are called upon to perform services of which they are not capable because of personnel or other resource shortages. Surely both top management and department supervisors would materially benefit from the development of quantitative techniques that would assist not only in cost control, but also in cost justification.

Summary

The objectives of this chapter were: (a) To acquaint the reader with the assumptions underlying correlation and regression analysis, (b) to present several of the problems involved in regression and correlation analysis of business data, and (c) to acquaint the reader with the concepts of standard costs, and in particular, with the concept of a standard overhead rate.

In pursuing the three objectives of this chapter the assumptions underlying the least-squares method of correlation and regression analysis were presented. It is first hypothesized that Y_i is a function of X_i and their average relationship may be expressed as $Y_i = a + bX_i + u_i$ for $i = 1, 2, \ldots, n$, in the linear case.

The first assumption is that the expected value of the residuals is zero, that is, $E(u_i) = 0$ for all i. Second, it is assumed that the residuals have a zero mean, zero covariance and constant variance or,

$$E(u_i u_j) = \begin{cases} 0 & \text{for } i \neq j;\ i, j = 1, 2, \ldots n \\ Q^2 & \text{for } i = j;\ i, j = 1, 2, \ldots n. \end{cases}$$

The fourth assumption is that X is a set of fixed numbers, and the final assumption is that X has rank $K < n$.

A few possible problems and misconceptions relating to regression and correlation were next discussed. It was stated that a high coefficient of determination, obtained when Y is regressed on X, did not prove a cause and effect relationship existed. The implications of multicollinearity, autocorrelation, and the use of time series data in regression analysis were set forth. It was stated that each of these items might generate major problems in specific cases. But in general, the problems involved could be resolved or at the worst circumvented, so that it should be possible in most cases to develop a valid estimating equation for cost functions.

In order to accomplish the third objective of this chapter, the general procedures and techniques for developing a standard cost system were presented. A standard cost system was recognized as being the highest state of sophistication in general use by cost accountants. Its virtues and a few of its short-comings were discussed.

CHAPTER III

DETERMINATION OF A SHIPPING AND WAREHOUSING COST MODEL

Introduction

A mathematical model of the Grace Company's overhead cost structure will be developed in this and the four succeeding chapters. The general approach will be to isolate the organizational structure of the company into three broad functional groups. These group classifications are functions of : (a) the formal organizational and responsibility structure, and (b) the organization of the accounting information flows and reporting structures. The three cost-group classifications are:

1. Shipping and warehousing costs
2. Sales department costs
3. Manufacturing service department costs

In this chapter the shipping and warehousing cost group will be subjected to analysis. Data for a 28-month period will be utilized for the analysis. Each cost item within the shipping and warehousing cost group will be critically examined as to: (a) its nature and purpose incurred, (b) its variation over the 28-month period, (c) its relationship to the total cost structure of the group, and (d) its relationship to activities or other costs outside its own cost group.

Once each of the determinations in (a) through (d) is made for each cost item or group of costs in the shipping and warehousing cost group, it should be possible to develop a single mathematical estimating equation for the department as a whole. If the model so developed is to be considered valid, then it must be capable of restating the group's past cost experience. Tests will be conducted to determine if the developed model does indeed have this capability within acceptable limits. If it does, then it will be considered a valid model.

After the estimating equation for the first cost group has been shown to be sufficiently descriptive of that group's cost behavior, similar procedures will be followed in constructing models for the three remaining cost groups. It is to be expected that each cost group will have different inherent complexities, but the approach to a solution will be similar in each case. For the shipping and warehousing costs model, considerable detail will be given regarding the individual accounts comprising the cost group and the process of determining the explanatory variables, which will not be given in the discussion of the remaining two cost groups contained in subsequent chapters.

Shipping and Warehousing Functions

Costs are incurred in this department in order to provide the following services:

1. Warehousing of all finished products.
2. The unloading or receiving of finished products shipped in from associated plants.

3. Loading and shipping of finished products to dealers and other associated plants.

4. Stacking and moving finished products as received from either outside sources or the production line.

Other functions occasionally performed by the department's personnel are limited janitorial work, inventory counting, maintenance of certain equipment, minor repairs to the building, and the handling of certain office supplies. When these other functions are performed to any extent, the workers note the time spent on their time cards, and the appropriate accounts are charged. Nevertheless, some of the costs of performing these services are probably not classified correctly and thus are incorrectly reported in the accounts. Interviews with the department supervisor indicate that this type of distortion has been held to a minimum.

Accounts Employed

Twenty account titles are employed by the company to accumulate the shipping and warehousing costs. Most of these accounts contain small annual totals and involve only a few transactions. Consequently, several of the accounts were consolidated, but it is felt that consolidation will have no bearing on the validity of the estimating equation. A second minor change was made relating to the terminology employed by the Grace Company. That is, for reporting purposes, the company grouped the accounts under two headings, controllable and uncontrollable costs. Because it was felt that such a grouping was inappropriate for use in correlation analysis, the accounts have been regrouped by their variability. After making the consolidations and regroupings, the account titles used in the analysis appear as follows:

Variable Costs	Fixed Costs
Wages	Supervisory Salaries
Payroll Taxes	Depreciation
Supplies and Small Tools	Insurance
Machinery Maintenance	Property Taxes
Lights, Gas, and Water	
Sundry Expenses	

Classification of the various cost accounts by their individual variability is an important first step in developing the estimating equation. The next step involves conducting an analysis of each account. In general those costs considered as fixed costs will be reviewed first. The purpose here is to determine if they are approximately constant or whether they react to internal production changes or other influencing factors.

Fixed Costs

Supervisory salaries were maintained at a relatively constant balance over the 28-month period. It includes the salary of the departmental foreman and the assistant departmental foreman. The assistant also handles all the filing, typing, and report preparation of various types done in the Shipping and Warehousing Department. Both men work on a fixed monthly salary and normally do not receive overtime. Occasionally one or both will work on a legal holiday or on Sunday and will receive additional compensation for those times; however, such cases are unusual and are a function of the timing of a shipment rather than the volume of shipping. A second reason for variation in this account results from periodic raises being granted.

Depreciation expense is recognized on both the building space occupied by the department and on the equipment which it utilizes. Building depreciation is computed on a straight-line basis and is based on the cost of the structures occupied by the department. In the case of the Grace Company, the department utilizes buildings especially constructed for shipping and warehousing purposes; therefore, it is not necessary that building depreciation be allocated on a floor space basis. But even if this were necessary, it would not invalidate the analysis. Depreication expense recognized on the equipment used within the department is also computed on a straight-line basis. Only major equipment such as fork lifts and hoists is depreciated; other items are expensed when purchased by charging their costs to the supplies and small tools account. Major maintenance charges incurred on equipment are capitalized and depreciated over the remaining life of the equipment. Minor repairs and maintenance charges are expensed when incurred.

Insurance and property tax expenses are both relatively fixed. The insurance expense includes charges for fire coverage on both the inventory and the facilities employed. Coverage on the inventory is designed to encompass inventory levels for both peak and low periods, but because the aggregate value is reasonably constant there is little fluctuation in monthly charges. Second, the premium is computed on an annual average inventory value and thus would not vary when allocated monthly, unless allocated on the basis of inventory valuation. However, the company does not make an allocation based on inventory valuation, and it is felt that such an allocation is not really necessary for the analysis. Insurance on the buildings is determined separately from insurance on other plant buildings, and no allocation problems appear present.

Property taxes are paid only on the buildings and equipment. However, tax assessments are levied on the entire plant as a whole and thus must be allocated. Property taxes on the buildings are allocated on a floor space basis, whereas on equipment they are allocated on the basis of the net book value of the equipment employed in each department.

In summary, the foregoing costs, classified as fixed costs, are not responsive to changes in volume of shipments, order size, number of workers, or other variable factors. Rather, they are a function of external influences or managerial decisions. This conclusion becomes quite apparent upon examination of Chart 3.1.

Variable Costs

The variable cost group represents the challenge in regression analysis. If the estimating equation is to be expected to provide eith meaningful marginal costs or total cost forecasts, it is vital that this group of cost accounts be correctly stated. After all, these costs will comprise the dependent variable Y in the equations to follow and the values of any $r^2_{1.23...n}$ and $R^2_{1.23...n}$ are definitely related to the propiety of the Y data. Therefore, the variable cost accounts must be examined for irregular items and other clerical errors.

It is expected that the explanatory variables will consist of the units handled by the department. Thus, restatements of the variable costs series may be desirable in order to promote homogeneity in the series as related to the possible explanatory series. Only after this is done may regression analysis be commenced.

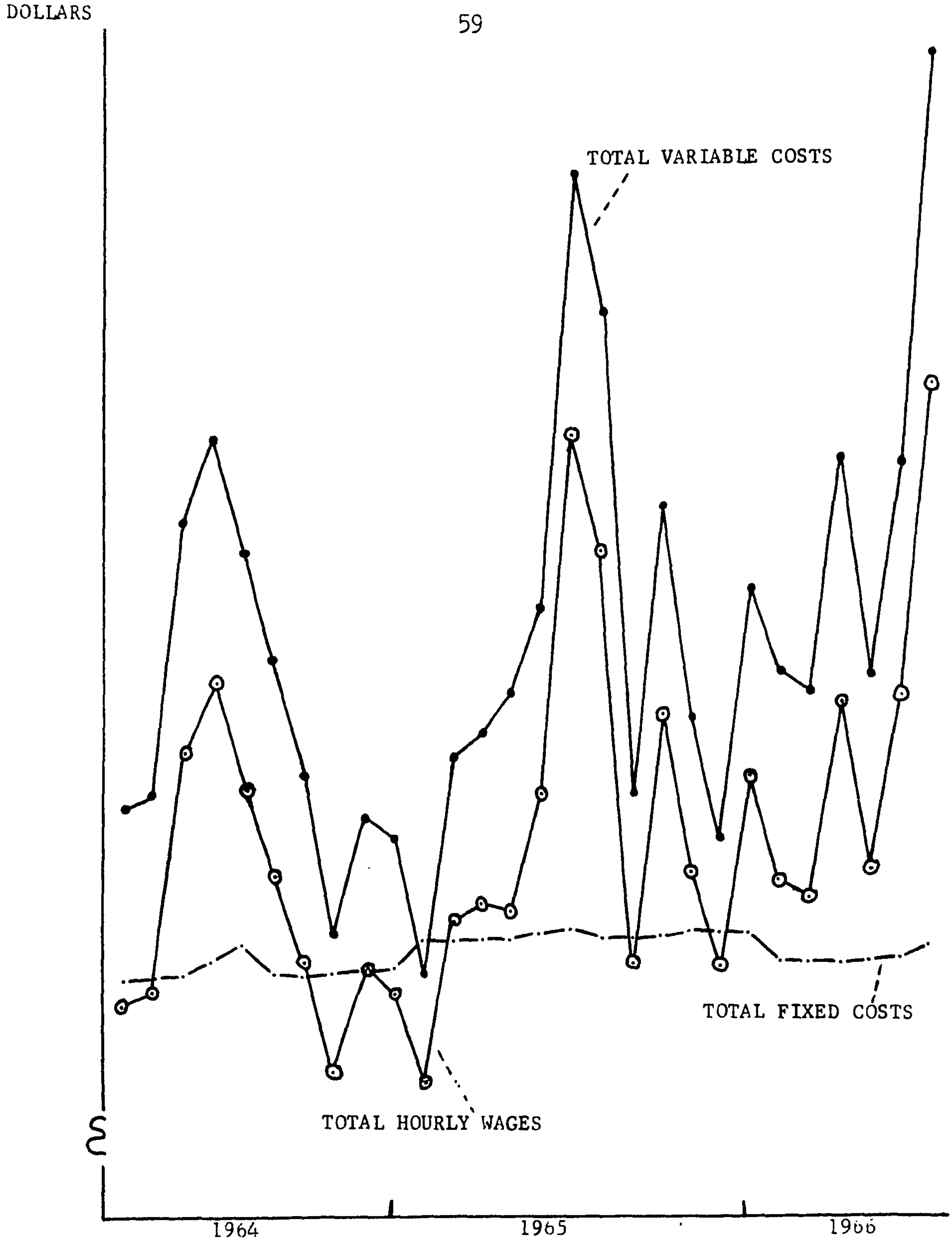

Chart 3.1. Total Variable Costs, Total Hourly Wages, and Total Fixed Costs of the Shipping and Warehousing Department, March 1964 - June 1966.

It should be noted that at this point in the analysis no hypothesis has been set forth regarding the equation form to be employed. However, it is likely that the final equation form will be both multiple and curvilinear. But this awaits determination. On the other hand, it was previously stated that the most likely explanatory series would involve the units handled. Using these and possibly other explanatory variables, it is hoped that an estimating equation capable of accurately forecasting the total variable costs of the department over the next 12 months can be developed. These considerations and objectives should be kept in mind as the analysis proceeds.

Variable Account Analyses

An analysis of the wages account revealed that both wage payments for straight-time hours and overtime hours, including the overtime premium, were included. In addition, the account contained wages paid to the night shift that was added in May of 1966. For purposes of regression analysis, the ideal situation would be for the wages account to include only straight-time hourly wages paid, and separate accounts should be maintained for such items as overtime premiums and night-shift premiums. Such a procedure would improve the data's homogeneity. Since this was not the case and since many firms follow similar procedures, it was decided that no breakdown or segregation of the wage costs would be attempted until after several correlations runs were made.

Aside from the foregoing aggregation problem, analysis revealed that over the 28-month period several workers had been paid for their vacations in lieu of taking off any time. These payments would cause irregularities in the data and bias the coefficients to be computed. Fortunately, the timing and amount of these vacation payments were readily determinable and so were cast out. Another factor which would have biased the results was a 4.0 per cent increase in the hourly wage rate for all hourly workers. This occurred in August, 1965 as a result of a new wage contract negotiated by the labor union. It was felt that in order to promote homogeneity in the data, it would be desirable to restate wages paid prior to August, 1965 in terms of current wage rates. This should enhance the estimating equation's ability to forecast total cost in the immediate future. However, it would not have been possible to make a valid restatement of the wages into present rates had it not been that information pertaining to the total straight-time and overtime hours worked each month was available.

After the foregoing adjustments were made, the balances in the wages account represented the monthly wages paid for straight-time hours, overtime premiums, and the night-shift premium. Admittedly, this is a conglomeration of sorts, yet the Grace Company uses this procedure in its accounting system, and if forecasts are to be prepared using an estimating equation, the forecasts must be compatible with their environment. Otherwise, the forecasted figures would not be comparable with future cost information to be generated by the accounting system.

The next account analyzed was payroll taxes, which includes the employer's contribution under the Federal Insurance Contributions Act and the Federal and State Unemployment Compensation Acts. No other items are included. Of course, the payroll taxes are directly related to the wages paid, but unfortunately the relationship is not constant. Ideally, for the purposes of correlation analysis, for every dollar of wages paid there would be some constant sum of payroll taxes expense. But that is not the case since FICA was paid on only the first $6,600 of each employee's wages in 1966. The rates, at which the FICA was paid, also varied over the three-year period. In addition the unemployment insurance was paid on only the first $3,000 of each worker's earnings. Thus during the third and fourth quarters of each year, reductions occur in payroll taxes that are in no way related to the work done in the department. As a result, if exactly the same volume of shipments was made in January as in December, and if exactly the same hours were worked by the same laborers, the total expense of wages and payroll taxes for each month would be different. This obviously would be reflected in the estimating equation and would reduce both the coefficients of partial and multiple determination. Now the Company recognizes payroll taxes on a monthly basis and it was determined that this, in general, was done accurately. Given this procedure and the fact that some sort of adjustment must be made in the timing of the recognition of payroll taxes expense, if its biasing effect was to be eliminated or diminished, it was determined that the following alternatives existed:

1. The tax incurred for each employee could be determined and then allocated on a proportional basis over the entire year.
2. The total payroll tax expense incurred for a year could be allocated proportionately to each month, based on the wages paid during each month.
3. The total payroll tax expense incurred each quarter could be allocated proportionately to each of the three months in the quarter, based on the wages paid during each month.

The first alternative procedure was rejected on the grounds that it would be impractical in most situations and a consistent methodology is desirable. Although alternative two was the simplest, it was not used because many of the employees did not work the entire year for one reason or another. The third alternative retained, to some extent, the same difficulty as did the second, but it appeared to be the most practical method for diminishing the bias as related to the possible explanatory variables and thus was adopted.

The supplies and small tools expense account is charged monthly for invoiced items received during the month. There are no provisions made in the accounting system to recognize charges for these items as they are exhausted. Interviews conducted with the department foreman revealed that some supplies were purchased monthly, others were purchased in sufficient quantities to last several months, and small tools were purchased or replaced as necessary. Consequently, the monthly expense reported was quite large in some months and almost non-existent in others. There was absolutely no relationship between the monthly expense recognized and the actual use of supplies and small tools. This condition could not be tolerated in the regression computations, yet the expense was significantly large and it could not be eliminated without materially affecting the estimating equation coefficients. It was, therefore, necessary to include the expense, but some method of smoothing had to be derived.

Additional interviews with the departmental foreman clarified the types of items considered as supplies or small tools and indicated appropriate use patterns. It was then decided that their cost could be spread over the months based on the proportion of the total labor hours worked each month to the total annual hours. Because of the nature of some items contained in the expense, this approach could not be considered absolutely valid, but it was most rational and practical of the several alternative methods considered.

Expenses incurred for lights, gas, and water are recorded directly from the statements received. Charges are based on meter readings for the Shipping and Warehousing Department alone. Although there were certain minimum charges by the utility companies, there was considerable variability in the account. Perhaps a certain portion of this expense could be considered fixed since it would normally exist each month, but it was felt that inclusion of the entire expense would not materially mitigate the total variation in the other strictly variable costs. For this reason, the entire expense was classified as a variable cost.

The machinery maintenance account is comprised of routine maintenance on hoists, fork lifts, and other minor equipment employed in the department, as well as major overhaul costs of this same equipment. The costs of major overhauls were significant, but they were not incurred at definite intervals. No rational basis for allocating these charges over certain months or periods was determinable. Consequently, these costs were eliminated from the account as irregular items while routine maintenance charges were retained.

The expense items classified as sundry were not significant in amount, but did contain considerable variability. After examining the various expenses comprising the account, it was concluded that it should be classified as a variable cost and that the timing of the expense recording appeared correct.

After reading the foregoing it should be obvious that considerable time and effort was spent conducting account analyses and interviews. However, this is a vital part of any effort to conduct correlation analysis. As a result of this phase of the work, several changes were made in the original data, as presented by the company, and this researcher's confidence in its accuracy was greatly improved. On the other hand, because of the smoothing of certain expense charges and the composite nature of several of the variable cost accounts this researcher is fully aware that there are inadequacies in the data. But these or similar data inadequacies are to be expected when analyzing the cost data of almost any operating firm.

Chart 3.1 appropriately summarizes the operating costs of the Shipping and Warehousing Department after the adjustments were made. An examination of the chart provides considerable information about the nature of the costs and their interrelations. For example, there is considerable variation in the total variable costs over the entire period. Costs appear to peak in June of each year, but the low points do not occur consistently in the same months. Therefore it would be difficult to establish definite seasonal patterns which could be employed as constants in the estimating equation. Yet there is obviously a significant amount of seasonal fluctuations. Second, the fact that the low cost points never fall to zero definitely indicates that either there is some constant volume of work or that, regardless of the work load, management has elected to maintain a minimum work force with a minimum productive capacity.

The relationship between total variable costs and total wages is quite apparent. Hourly wages fluctuate almost identically with total variable costs and obviously comprise a significant proportion of the total. However, a perusal of the two series will reveal that the amplitude of upward fluctuations in the total costs is greater than that of wages. This is to be expected since there are other variable costs contained in the total variable cost series. It should be noted that the amplitude of downward fluctuations in total variable costs is also greater, but not consistently so. Further, in April of 1965 the relationship between the series is negative rather than positive. This behavior indicates an irregular item may be present in the data.

Fixed costs, as a group, demonstrate very little variability over the entire period. Variability in the series stems from the supervisor's being required to work on Sundays or other legal holidays during certain times of the year, from changes in supervisory salary levels, and from changes in insurance and property tax rates.

In summary, Chart 3.1 reveals that the fixed costs are indeed fixed and that the trends of the total variable costs and wages series are both positive and quite similar. Fluctuations of the two series are almost identical in timing but differ somewhat in amplitude. For purposes of determining an estimating equation, it would seem that both series should be a function of the same independent variable or variables. Consequently, the total variable costs series will be used to determine an estimating equation by employing selected independent variables that are logically causal factors for all costs contained within the dependent cost group.

Preliminary Selection of Explanatory Variables

It has already been stated that initially the total variable cost series will be the dependent variable and that the units handled will be one of the explanatory variables. The next step is to determine the exact nature of the independent variable or variables. Three general considerations in selecting the independent variables are: (a) the purposes for which the estimating equation are being developed, (b) the complexity of the dependent variable, and (c) the types and validity of available information.

If the function of the estimating equation were only to forecast future total variable costs, one might attempt to determine independent variables that would provide a high coefficient of multiple correlation. That is, the two series would move together even though they might be completely unrelated. Such a procedure is extremely dangerous and is not recommended. But this equation will be used not only for forecasting and budgeting but also for control purposes. In view of these requirements, it is necessary that the independent variable(s) employed either logically explain or cause variation in the dependent variable. Obviously such a requirement may in many instances considerably complicate the search for independent variables.

The dependent variable, total variable cost, is also quite complex in that it represents the composite balances of a number of costs. Ordinarily it would be inadvisable to employ such a complex variable; however, in this instance it may justifiably be assumed that all its components are directly related to the same basic set of causal factors.

The third general consideration, availability and validity of data, may often be and, in this case, is a most difficult aspect of the analysis.

As stated in Chapter II, it is vitally important that the various data series used as independent variables be accurate. The errors of measurement and timing must be eliminated from the data of the independent variable(s) series if at all possible. Such a rigid accuracy requirement will obviously make it difficult, in a practical sense, to employ all the independent variables one might consider desirable. Such was the case in the analysis of the Shipping and Warehousing Department's costs as will be noted later. On the other hand, in real-life situations it is rarely possible to employ a sufficient number of independent variables so as to completely explain all variation in the dependent variable. "We choose instead to represent Y as an explicit function of just a small number of what are thought to be the more important X's independent variables and let the net effect of the excluded variables be represented by u."[1]

In light of the three foregoing general considerations, it is obvious that any series of data to be selected as a possible independent variable must possess the characteristics of causation, availability, and accuracy. Causation is of primary importance and once a data series logically qualifies in this regard, it must then be evaluated in terms of availability and accuracy. A search for possible independent variables possessing the causation characteristic resulted in the following list of potential candidates. Each series is assumed to be accumulated on a monthly basis.

A. Total units of Product A shipped to dealers.
B. Total units of Product A shipped to associated plants and warehouses.

---

[1]J. Johnston, Econometric Methods, (New York: McGraw-Hill Book Company Inc., 1963), p. 6.

C. Total units of Product A received from associated plants and warehouses.
D. Total units of Product A received from the production line.
E. Total units of Product B shipped to dealers.
F. Total units of Product B shipped to associated plants and warehouses.
G. Total units of Product B received from the associated plants and warehouses.
H. Total units of Product B received from the production line.
I. The proportion of LTL shipments to total truckload shipments.
J. The proportion of LCL shipments to total carload shipments.
K. The total monthly tonnage received and shipped via truck.
L. The total monthly tonnage received and shipped via rail.
M. The average number of items per sales order, per shipping order, or some related ratio.
N. The number of shipping orders.
O. The number of receiving reports.
P. The number of hourly laborers.
Q. The per cent of overtime hours worked to total-straight time hours worked.
R. The total hours worked each month.
S. The ratio of hourly laborers to total equipment investments.

Obviously there are numerous ratios which could be developed by combining various ones of the foregoing data series. It is also quite likely that the foregoing list does not include all acceptable independent variable candidates, but it is felt that sufficient variables are contained in the list to permit development of an estimating equation. It should be noted that most or all of the independent variable series are in some way related to one or more of the other variables in the list. Further, the series (P) through (R) are explicitly related to the total labor cost, which is a component of total variable cost, and which it is hoped is a $f(X_a \ldots X_o)$. It is admitted that independence between all the independent variables themselves does not necessarily exist. Indeed, such conditions will normally be present in business data because of the very nature of the business, but this does not preclude valid analyses.

The availability of data for each of the foregoing series was the next consideration in the selection of the independent variables. Information pertaining to the series (A) through (H) was available from inventory reports. No card or tape inventory records were maintained by the company's data-processing department. To have attempted verification by examining shipping and receiving reports would have been a practical impossibility. Verification of the accuracy of each series was still possible because two independent lists of the series are compiled within the company. Based on the inventory reports, adjustments were made to several of the series, after which they appeared to be reasonably accurate. The data series (P) through (R) were not available through any formal company report so they were compiled directly from the company's payroll records by this writer.

None of the remaining data series were available from any company reports covering the entire period used for the regression analysis. However, some data for shorter periods were available for several of the series. These data were compiled within the Shipping and Warehousing Department by its own personnel and not within the formal accounting system. Tests were made of the accuracy of the data, and it was determined that the series were not sufficiently accurate to be employed in the regression analysis. The series involved here were (I) through (L). The series (M), (N), and (O) presented other difficulties which prevented their being used in the analysis. That is, the records were either destroyed after one year or they were filed in a manner incompatible with other monthly series or costs.

In summary, of all the possible data series which were potential independent variables, only the data series (A) through (H) and (P) through

(R) remained after the criteria of causation, availability, and accuracy were applied. Though this way may seem to limit the potential applicability of any resulting estimating equation, it should be remembered that for the purposes of expediency and practicability it is best to reduce to a minimum the number of independent variables employed. For this reason, only the remaining data series were employed in the following regression analysis, the results of which indicate their considerable significance in determining total variable costs.

Developing an Estimating Equation

At this point in the analyses of the shipping and warehousing costs, it has been determined that the dependent variable (Y) is to be total variable costs and several independent variable candidates ($X_1 \ldots X_n$) have been selected. The data for the dependent and explanatory variable series have been compiled and determined to be reasonably accurate, that is, free of timing and clerical errors. Therefore, all preliminary procedures necessary for actually computing an estimating equation have been accomplished.

In developing the estimating equation, the procedures and each objective are set forth as follows:

1. Chart the dependent and independent variable series to provide visual assistance in establishing; (a) the possible existence of a relationship between the (Y) and ($X_1 \ldots X_n$) variables, and (b) the form of mathematical equation which might be the most efficient.

2. Take the dependent variable (Y), total variable cost, and develop an equation of explanatory variables of the form $Y_c = a + bX_1 + \ldots c_nX_n$ such that $\Sigma(Y - Y_c)^2$ is a minimum or that $R^2_{1 \ldots n}{}^c$ is a maximum.

3. Evaluate the coefficient of multiple determination ($R^2_{1 \ldots n}$) using an F-Test of significance at the .05 probability level.

4. Determine a standard error of the estimate (S_y) for the total estimating equation such that $\pm 1.96S_y \leq 25$ per cent of $Y_{c1...n}$ for any value of $X_{1...n}$.
5. Determine a coefficient of partial determination ($r^2_{1.n}$) for each $X_{1...n}$ and evaluate using an F-Test of significance at the .05 level.
6. Evaluate the standard error for each regression coefficient using a t-test of significance at the .05 level.

It may not be necessary to complete all six of the foregoing procedures for each trial equation. Rather, it should be sufficient initially, to compute only the coefficient of multiple determination and the standard error of the estimate for each set of explanatory variables. If both of these measures satisfy the prestated requirements, then the coefficients of partial determination and the standard errors of the regression coefficients will be evaluated.

Information pertaining to the number of units or product items handled is contained in the (A) through (H) data series. Plotted on Chart 3.2 are the series total variable cost, total units of Product A shipped, and total units of Product A shipped and received. Note that the data series (A) and (B) have been combined to form a new series which was plotted and that the data series (A), (B), and (C) were also combined. The (A) and (B) series were combined after examining the two series and determining that each series individually had little relationship to the (Y) variable. Upon personally observing the work being done within the department and after interviews with the department foreman, it was concluded that there was very little difference in the labor time required to ship a unit to a dealer as opposed to an associated plant. Further, combining the series reduced the number of variables and therefore increased the degrees of freedom.

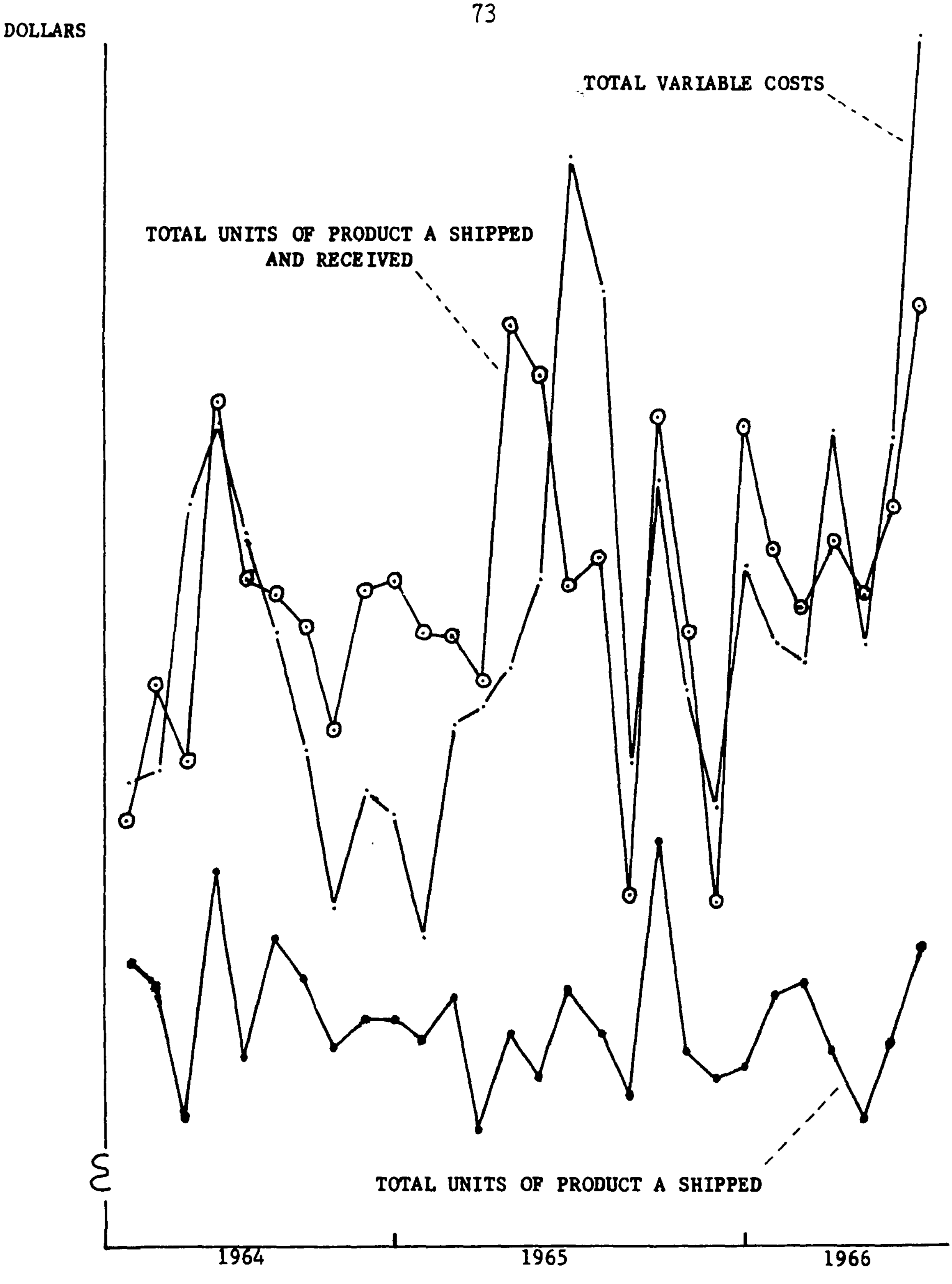

Chart 3.2. Total Variable Costs, Total Units of Product A Shipped, Total Units of Product A Shipped and Received of the Shipping and Warehousing Department, March 1964 - June 1966.

Examination of the plotted series indicates that the trend of each is similar, but that the trend of the (Y) variable appears slightly more positive. Fluctuations in the series are not particularly uniform, as to time, particularly in the early months; however, their uniformity seems to have improved significantly by August, 1965 and thereafter. These irregularities during the period March, 1964 through July, 1965 may partially be attributable to the gradual decline in production of Product A and the gradual increase in the number of units received from associated plants. After July, 1965, the proportion of units of Product (A) produced as opposed to the number received from associated plants remained relatively constant.

In Chart 3.3 the series for Total Variable Costs (Y), Total Units of Product (B) shipped, and the Total Units of Product (B) shipped and received are plotted. Again several of the original series were combined. The grounds for making the combination were similar to those for Chart 3.2. But the primary purpose in this case was to promote consistency in the types of series plotted to improve comparability. Because the proportion of units of Product (B) received from associated plants is relatively insignificant the procedure is particularily relevent. Inspection of the plotted series reveals that a relationship definitely exists. In general, the fluctuations are similar as to time; however, there are irregularities between them.

Chart 3.4 contains curves for Total Variable Cost (Y) and for the Total Units of Product (A) and (B) Shipped and Received. This chart combined into one series, the four explanatory variable series plotted on Charts 3.2 and 3.3. It now becomes obvious that a reasonably close relationship exists between the variables. This relationship is much more obvious in Chart 3.4

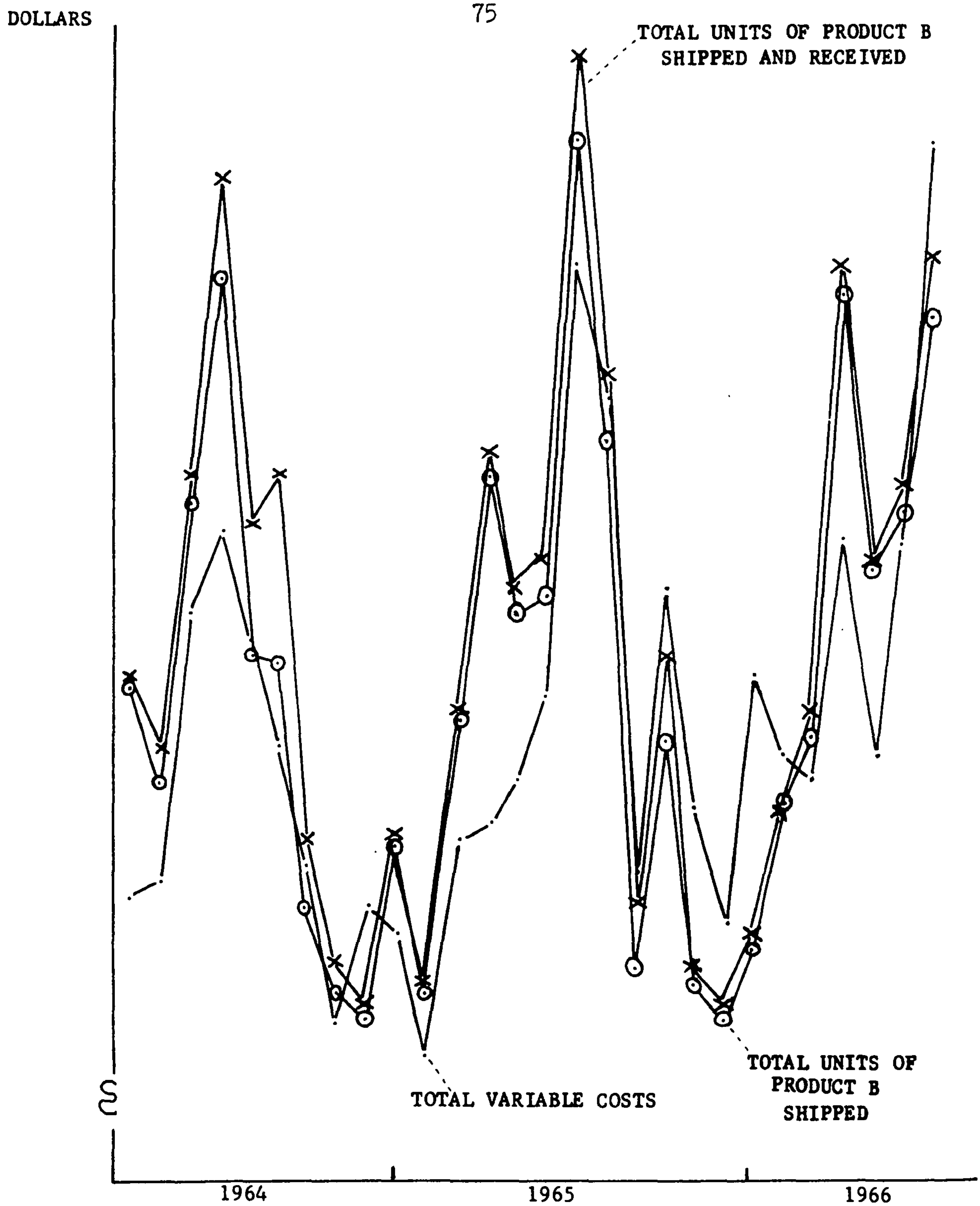

Chart 3.3. Total Variable Costs, Total Units of Product B Shipped, and Total Units of Product B Shipped and Received of the Shipping and Warehousing Department, March 1964 - June 1966.

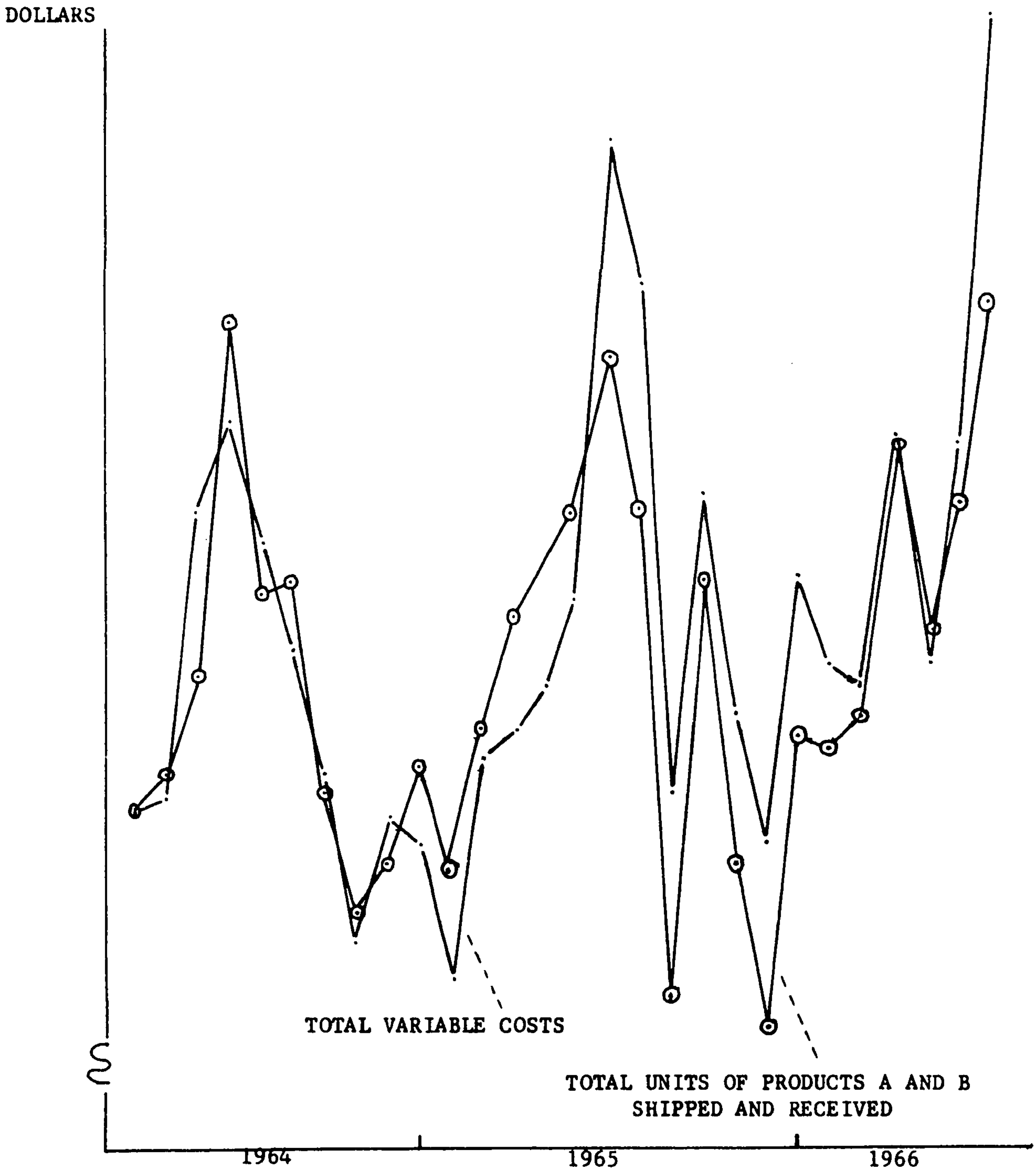

Chart 3.4. Total Variable Costs, Total Units of Products A and B Shipped and Received of the Shipping and Warehousing Department, March 1964 - June 1966.

than in either Chart 3.2 or 3.3. However, combining the explanatory variables to such a degree may result in the estimating equation's not providing as much cost information as it could have, had the series not been combined.

The form of equation which would be most efficient in describing the relationships illustrated in the charts is not readily determinable. For this reason, the initial equation form will be multiple and linear, and then perhaps curvilinear. If satisfactory results cannot be obtained using these equation forms, it may be necessary to make transformations to logarithmic, reciprocal, or $\sqrt{Y}$ forms. Such transformations do not present major computational problems since they may be made by amending the header cards of the computer program.

Before commencing computation of the estimating equation, it will be necessary to change the equation notation from the general form $Y_c = a + bX_1 + \ldots cX_n^n$, which has been used for convenience, to:

$$X_{c1.2\ldots n} = a_{1.2\ldots n} + b_{12.3\ldots n}X_2 + \ldots b_{1n.2\ldots n}X_n$$

for linear equations. For non-linear equations the notation will be the same, except that the subscripts will also include the exponent for each independent variable, if it has a value of other than one.

Selection of Explanatory Variables

The first set of computations will attempt to determine the significance of shipments as a cause of costs. Based strictly on logic, it would seem the total number of units handled by the department should largely determine the total cost incurred. If this is valid, then how do the shipments out relate to costs? Let:

X_1 = Total variable cost

X_2 = Total units of Product (A) shipped

X_3 = Total units of Product (B) shipped

and the equation form employed is

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3. \qquad \text{(Eq. 1)}$$

Using 28 observations for each variable of the equation, the computed coefficient of multiple determination $R^2_{1.23} = .6782$ and $S_{y1.23} = \$829.19$. Using the F-Test where $n_1 = 2$ and $n_2 = 25$, it is clear that $R^2_{1.23} = .6782$ is significantly greater than zero. Although the $R^2_{1.23}$ value is satisfactory, the $S_{y1.23}$ is somewhat larger than desired. Therefore, different equation forms will be tried to see if the standard error of the estimate can be reduced by improving the goodness of fit.

Inclusion of X_2^2 and X_3^2 terms in the equation resulted in an $R^2_{1.2^2 23^2 3}$ = .7020 and an $S_{y1.2^2 23^2 3}$ of \$831.95. Adding X_2^3 and X_3^3 terms to the equation resulted in an $R^2_{1.2^3 2^2 23^3 3^2 3} = .7297$ and an $S_{y1.2^3 2^2 23^3 3^2 3} = \$829.15$. It appears that this route is unfruitful as a method of reducing S_y.

The foregoing values reflect not only the seasonal and cycle variations in the series, but also those of trend. The effects of time may be partially eliminated by adding a fourth term X_4 to the equation for time. Employing a linear equation of the form

$$X_{c1.234} = a_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 \qquad \text{(Eq. 2)}$$

resulted in a $R^2_{1.234} = .7694$ and a $S_{y1.234} = \$716.31$. The net effect of adding a time variable was an increase of .0912 in the coefficient of multiple determination and a decrease of \$112.88 in the S_y. The net changes in both the measures are definitely significant. If no further regression analyses were conducted, a reasonable degree of accuracy in cost forecasting could be expected, using the above linear equation. However, significant additional information for planning may be revealed through other equations.

In the next series of computations the number of independent variables will be expanded and initially no adjustments will be made for trend. The variables are:

X_1 = Total variable cost

X_2 = Total units of Product (A) shipped

X_3 = Total units of Product (A) received

X_4 = Total units of Product (B) shipped

X_5 = Total units of Product (B) received

Equation form:

$$X_{c1.2345} = a_{1.2345} + b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5 \quad \text{(Eq. 3)}$$

Using a linear equation resulted in an $R^2_{1.2345} = .7231$ which is only .0449 higher than the $R^2_{1.23}$ for equation (1). The $S_{y1.2345} = \$801.95$ which again is but slightly smaller than that of equation (1). Still using a linear equation but adding X_6 for time resulted in an $R^2_{1.23456} = .8024$ as compared to an $R^2_{1.234} = .7694$ for equation (2) or an increase of .0330. It appears that even though the inclusion of variables representing the units received did increase the multiple coefficient of determination, these variables alone will not explain all the variation in total variable cost. Further, it should be noted that the additional amount of variation which was explained by their inclusion was relatively small. This type of behavioral pattern might be due to multicollinearity. A review of Charts 3.2 and 3.3 indicates that such could be the case. If the units shipped of each product are highly correlated with the units received series, then the coefficient of multiple determination should not be decreased materially by combining the related series.

A second possible explanation of the foregoing behavioral pattern might be attributable to the indirect effects of the units received variables acting in conjunction with units shipped variables. This is explained by Robert Ferber as follows:

> ...The influence each independent variable exerts on the multiple correlation coefficient may be direct or indirect. Direct influence is exerted...when an independent variable affects the multiple relationship solely through its own variation. Indirect, or joint, effects arise when some of the variation in the dependent variable is explained by the co-ordinated, or interacting, influence of several independent variables. The net effect of an independent variable on the multiple relationship is the sum of its direct effect and of its various indirect effects.[2]

It appears that the best approach now open is to combine the units shipped and received series for each product. After making the combination, the variables to be used in the next series of iterations are:

X_1 = Total variable costs

X_2 = Total units of Product (A) shipped and received

X_3 = Total units of Product (B) shipped and received

X_4 = Time

---

[2]Robert Ferber, Statistical Techniques in Market Research, (New York: McGraw-Hill Book Company, Inc., 1949), p. 363.

The following equation form will be used:

$$X_{c1.234} = a_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4 \qquad \text{(Eq. 4)}$$

The first computational effort did not include the time variable X_4. The $R^2_{1.23} = .7229$ for this equation is almost identical to the $R^2_{1.2345} = .7231$ for equation (3). It appears that combining the units-shipped and the units-received series of each product into two single independent variable series does not decrease their composite ability to explain variation in the dependent variable series. This supports the previously stated rationale for making the combination.

The $S_{y1.23} = \$769.33$ for equation (4) as opposed to an $S_{y1.2345} = \$801.95$ for equation (3) represents a \$32.65 decrease. This reduction, though not large, may be attributed to reducing the number of independent series from four to two.[3] An added benefit is that the degrees of freedom were increased by the reduction in the number of variables.

Adding the time variable X_4 to the equation provided an $R^2_{1.234} = .7901$ and an $S_{y1.234} = \$683.51$. To this point, only linear regression forms have been employed. The next logical step is to include in equation (4) additional terms making it both multiple and curvilinear. The purpose in doing this is to determine to the best possible extent, the mathematical relationship between the dependent and independent series. Presumably, "the expeditious selection of appropriate curve types is contingent upon logic and a knowledge of the business element interactions."[4] However, it

---

[3]Croxton and Cowden, p. 548.

[4]Robert A. Knapp, "Forecasting and Measuring with Correlation Analysis," <u>Financial Executive</u>, Vol. XXXI, No. 5 (May, 1963), p. 17.

appears impossible to determine the proper relationship by inspection when working with the type of time series data shown on Charts 3.2, 3.3, and 3.4. Therefore, when working with such complex curves as the ones plotted, iteration is required.

A number of tests were subsequently made by adding or deleting terms having different exponents although the same variables were employed. A summary of the runs follows. The general form of equation used was again:

$$X_{c1.234} = a_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4$$

Iteration	Terms Added	$R^2_{1.2...,n}$	$S_{y1.2...,n}$
a.		.7901	$683.51
b.	X_3^2	.8133	658.39
c.	X_3^3, X_3^2	.8124	674.87
d.	X_2^2	.7984	684.09
e.	X_2^3, X_2^2	.7986	699.28
f.	X_4^2	.7929	693.54
g.	X_4^3, X_4^2	.7993	697.90
h.	X_3^2, X_4^2	.8225	656.38
i.	X_2^2, X_3^2, X_4^2	.8277	662.00
j.	X_3^3, X_3^2, X_4^3, X_4^2	.8326	668.56
k.	X_3^3, X_3^2, X_2^2, X_4^3, X_4^2	.8351	680.76

Of the foregoing equations it appears as though equation (h) offers the best fit when both the coefficient of multiple determination and the standard error of the estimate are considered. But an evaluation of the individual coefficients of partial determination and the standard errors of the regression coefficients indicated that the equation form did not provide a sufficiently valid fit. A perusal of the runs (b) through (k) makes several points clear. First, the variables employed cannot be made to explain over approximately 80 per cent of the total variation. Second, the standard error of the estimate seems stable for these variables. An examination of the standard errors reveals considerable variation exists in the regression coefficients except in the strictly linear case. In view of the behavior of the standard errors it appears that multicollinearity may exist between X_2 and X_3.

In the final analysis it appears that of the iterations the linear equation form provides the best fit. This equation appears as follows:

$$X_{c1.234} = \$911.38 + \underset{(16.58)}{\$46.47X_2} + \underset{(.163)}{\$1.1638X_3} + \underset{(.297)}{\$.4058X_4} \qquad \text{(Eq. 5)}$$

For equation (5) the $R^2_{1.234} = .7901$ and the $S_{y1.234} = \$683.51$.

In an attempt to decrease the standard error of the estimate and concurrently eliminate some of the possible multicollinearity existing between X_2 and X_3, these variables will be combined to form a single explanatory variable. Therefore, the next series of iterations will employ the following variables:

X_1 = Total variable cost

X_2 = Time

X_3 = Total units of Products A and B shipped and received.

These variables and a linear equation form provided a $R^2_{1.23} = .7581$ which is .0320 less than that of equation (5). The $S_{y1.23} = \$718.92$ as compared to the $S_{y1.234} = \$683.51$ for equation (5). It appears that by combining the X_2 and X_3 variables of equation (5) some information may have been lost.

Additional iterations, using the above variables, were carried out for the purpose of determining a more efficient equation form. Of these iterations the most efficient equation form appeared curvilinear, employing X_2^3, X_2^2, X_2 terms for time and a single linear X_3 term. This equation provided a $R^2_{1.2^3 2^2 23} = .7803$ and a $S_{y1.2^3 2^2 23} = \$714.21$. However, another iteration provided a $R^2_{1.2^3 2^2 23^2 3} = .8125$ and a $S_{y1.2^3 2^2 23^2 3} = \$674.71$. But for the curvilinear equation forms a considerable amount of variation existed in the regression coefficients of the X_2 and the X_3 terms. It now appears that a variable representing units handled will not be sufficient in the sense of decreasing both the standard error of the estimate and the standard error(s) of the regression coefficients.

Since total hourly wages are a major proportion of the total variable costs, it appears desirable that some variable to assist in explaining this element should be included in the computations. Obviously total wages are closely related to the number of units handled, but there are also other considerations. For example, hourly wages are determined on a weekly basis, whereas the unit data is on a monthly basis. Thus, overtime premium may be paid during one week when unit volume is heavy and in other weeks unit volume may be less than could be handled on straight-time alone. Overtime premium paid is also a function of the number of employees during any one week, yet other things being equal, the larger the number of employees, the less likely overtime premium will be incurred.

In view of the foregoing, a variable representing the per cent of overtime hours worked to total straight-time hours worked each month will be employed in the next series of iterations. This is series (Q) in the list of explanatory variables. It is hoped that inclusion of this variable will assist in reducing the standard error of the estimate and, of course, increase the coefficient of multiple determination. Thus, the variables to be employed are:

X_1 = Total variable cost

X_2 = Time

X_3 = Units of Products A and B shipped and received

X_4 = Per cent of overtime hours to straight-time hours

The first iteration employed a linear equation form which provided a $R^2_{1.234}$ = .7884 and a $S_{y1.234}$ = \$686.22. Neither of these statistics represents a profound improvement over those previously obtained. However, the $r^2_{13.24}$ and the $r^2_{14.23}$ were both significant at the .05 level.

Several different iterations were then carried out to determine an estimating equation offering a better fit. Of the equations computed the following appeared the most efficient.

$$X_{c1.234^24} = \$1,259.85 + \underset{(16.39)}{\$77.887}X_2 + \underset{(.1195)}{\$.5775}X_3 - \underset{(3.509)}{\$13.372}X_4^2 + \underset{(1,083.90)}{\$4,126.70}X_4 \qquad \text{(Eq. 6)}$$

For equation (6) the $R^2_{1.234^24}$ = .8735, which is definitely satisfactory. The $S_{y1.234^24}$ = \$541.94 and is over \$130 less than any of the other standard errors of the estimates computed to this point. The individual coefficients of partial determination were impressive. The computed F value for each is shown below. For each, n_1 = 1 and n_2 = 23.

	Computed F Value	.05 F Value
$r^2_{12.34^24}$	22.568	4.279
$r^2_{13.24^24}$	23.342	4.279
$r^2_{14^2.234}$	14.520	4.279
$r^2_{14.234^2}$	14.554	4.279

In view of the foregoing, each of the coefficients of partial determination may be regarded statistically significant at the .05 level. The standard errors associated with the respective regression coefficients appear quite satisfactory. A t-test was carried out for each regression coefficient and each was significantly different from zero at the .05 level.

An essential characteristic of any heuristic model is its ability to duplicate past experience. Chart 3.5 indicates the extent to which equation (6) duplicates the historical variable costs of the Shipping and Warehousing Department. Close examination of the computed cost curve reveals that the costs attached to shipping and receiving Product (B) have a stronger influence on the computed curve than do the costs associated with Product (A). This was to be expected, as far more units of Product (B) are handled, they are much larger, and weigh considerably more.

Of the 28 ($X_1 - X_{c1.234^24}$) differences, 15 are above the X_1 curve and 13 are below. Assuming normality, the dispersion of the differences appears satisfactory. Of the 28 differences, 18 occur within a range of ± 1.00 $S_{y1.234^24}$ and the others occur within the ± $1.96S_{y1.234^24}$ range. In view of the criteria for acceptance, it appears estimating equation (6) may be justifiably accepted as a model of the shipping and warehousing total variable cost group.

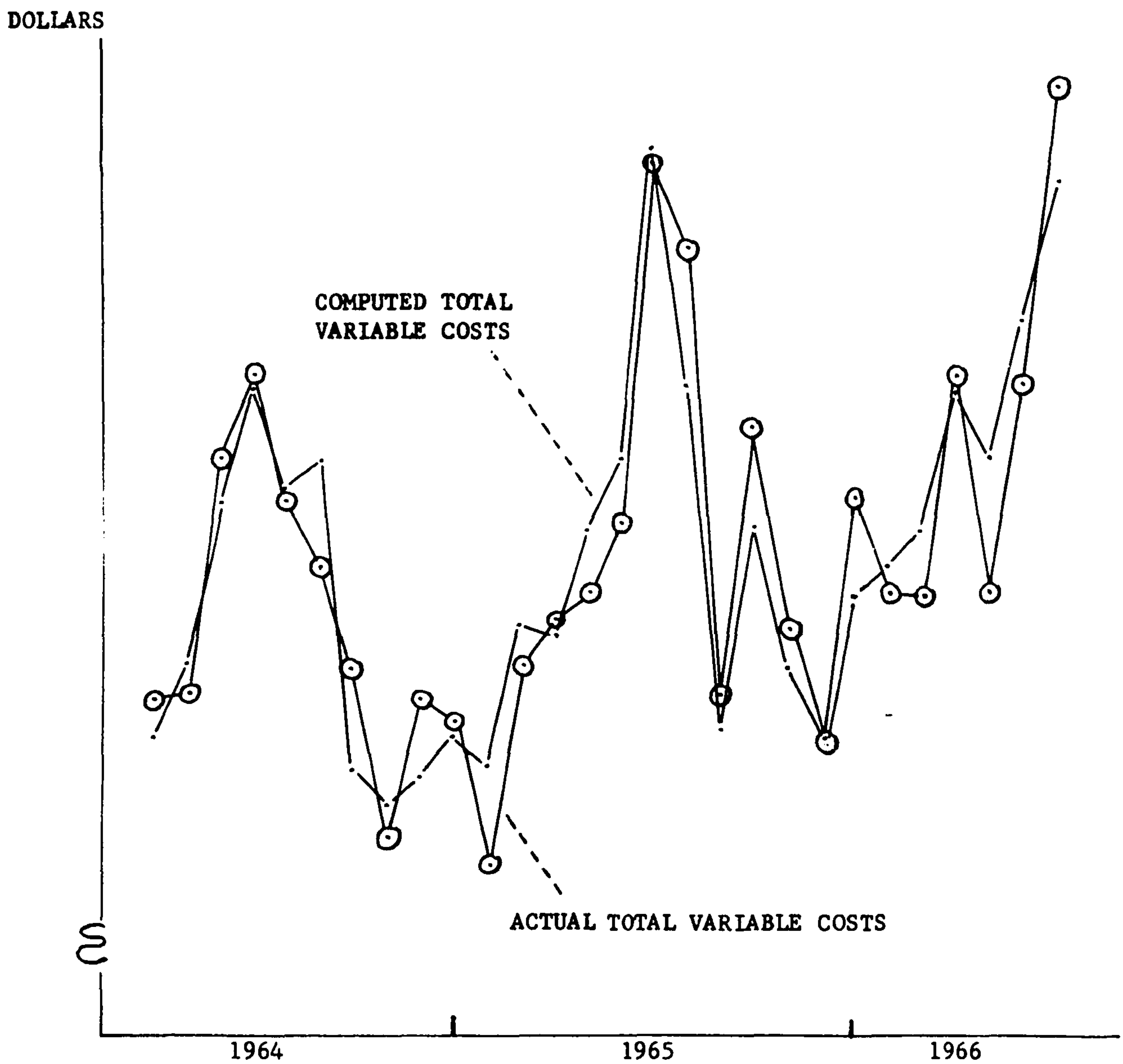

Chart 3.5. Actual Total Variable Costs and Computed Total Variable Costs of the Shipping and Warehousing Department, March 1964 - June 1966.

Although estimating equation (6) has been accepted as offering the best fit, it is appropriate that several of its characteristics and related procedural steps be further discussed. First it should be noted that approximately 87 per cent of the variation in X_1 is explained by the variables included in estimating equation (6). This explained variation represents the combined explanatory ability of all the independent variables, including both their direct and indirect effects. It is likely that some multicollinearity exists between the X_3 and X_4 variables, but its effects are minimal. This statement is based on the relative size of the individual standard errors. Further, unusually high variance in the regression coefficients may occur for reasons other than multicollinearity. For example, variance in the X_3 regression coefficient may stem from the existence of a reasonably stable work force which must be paid for forty hours' work each week, regardless of the units handled during a given week. Thus in one week (w) number of units might be handled, and in the immediately succeeding week (v) number of units might be handled, but the total labor cost might be the same in each week. Fluctuations of this type will cause variation in the X_3 regression coefficient. Presumably these elements should balance out over the course of a month but this may or may not be the case. A similar rationale could be applied to the X_4 variable as an explanation of the associated variance.

The approximately 13 per cent of unexplained variation in X_1 is presumably attributable to the exclusion of relevant explanatory variables from the estimating equation (6). But other factors may be partially responsible for this unexplained variation. For example, measurement errors that were not corrected in the earlier steps of the analysis may remain in the dependent and independent series. Such errors would also have a bearing on the size of the standard errors and the standard error of the estimate. It is also possible that the most efficient equation form has not been employed. It should be obvious at this point that a large population of possible estimating equation forms exists. Equation (6) was merely selected as the most efficient of all the equation forms hypothesized and computed.

A second consideration is the form of the variable terms. The variable of interest, in terms of form, is X_4, which represents the per cent of overtime hours worked to the straight-time hours worked. The reasons underlying inclusion of this variable were presented earlier so this aspect needs no further discussion. But the fact that it is employed in the X_4^2 and X_4 form is of interest and the negative sign of the second degree term is of particular interest. It should be noted again that the various terms for a single variable all work together to explain the relationship between the dependent variable and the specific explanatory variable. Therefore, the signs attaching to a particular term of an explanatory variable may not be particularly relevant. However, when only single linear terms are employed for each variable, a logical explanation of the sign taken on by the regression coefficient is necessary.

When two or more terms of the same variable are employed, the signs associated with the terms might be different, depending on the form of the terms. This may also be related to the type of input data being used in the series. For example, in estimating equation (6) the X_4^2 and X_4 terms represent series of percentages. Assume one observation provided a value of .20, then the X_4^2 term would in effect be .04 and the X_4 term .20. As the percentage increases, note the manner in which the relative weight of the two terms changes. Finally, the values of the regression coefficients themselves should be brought into the discussion. The regression coefficient for X_4^2 is \$13.372 and for X_4 is \$4,126.70. The second degree term has little weight in determining the total cost attributable to overtime premium. Further, even though the X_4 term's regression coefficient is large, the value of X_4 never exceeds .40, which means the highest possible value for both X_4 terms combined is \$1,648.54. Therefore, it appears that some fractional exponent would be more appropriate for X_4 and that only one term could have been employed. On the other hand, it may have taken a large number of additional iterations to determine the exact fractional exponent and little if any increased accuracy in forecasting capabilities may have accrued. But a negative sign may not have been present in the estimating equation.

Summary

The primary objective in this chapter was to develop an estimating equation(s) for costs incurred in the Shipping and Warehousing Department of the Grace Company. To this end, the accounts employed in this department by the Grace Company were analyzed for clerical and other measurement errors. Each account was reviewed for its behavioral characteristics and accordingly classified as a fixed or as variable costs.

The fixed costs were then individually discussed and subsequently excluded from further analysis. The fixed costs were exempted from regression analysis on three grounds. First, they may be accurately forecasted by conventional budgeting methods. Second, fixed costs do not vary with internal operational influences. Third, since fixed costs do not demonstrate variation, it is neither relevant nor practical to submit them to regression analysis.

It was stated that initially the explanatory variable series to be employed would relate to the number of units handled by the department each month. Since the primary function of the department is to ship, store, and receive the two products of the firm, it was felt that all the variable costs incurred by the department could be aggregated. Thus, one dependent variable series representing total variable costs would be used. Each of the variable costs was examined for irrelevant data in an attempt to maximize the homogeneity of the aggregated dependent variable series and several adjustments were made.

A list of possible explanatory variable series was then compiled. Each of these series was then examined in view of the three criteria of causation, availability, and accuracy. The next phase of the analysis involved computing a series of estimating equations employing different equation forms and various explanatory variables. In the final analysis it was concluded that estimating equation (6) provided the best fit or estimate of the total variable costs series. Estimating equation (6) is shown below in its entirety along with its associated coefficient of multiple determination and standard error of the estimate.

$$X_{c1.234^24} = +\$1,259.85 + \$77.887X_2 + \$.5775X_3 - \$13.372X_4^2 + \$4,126.70X_4$$

$$R^2_{1.234^24} = .8735 \qquad S_{y1.234^24} = \$541.94$$

CHAPTER IV

DETERMINATION OF A SALES DEPARTMENT COST MODEL

The objective in this chapter is to develop estimating equations which may be used as cost models of the various elements within the sales department. The intended use of these cost models will be forecasting total annual and monthly costs for the respective cost elements. The nature of the cost elements within the department and the nature of the explanatory variables to be employed are such that interpretation of the regression coefficients as marginal costs might be a questionable procedure. But this latter point cannot be verified until the estimating equations are determined.

Each cost group within the Grace Company will present different types of problems. But the methodology employed will be similar to that of Chapter III, and of course, regression and correlation analyses will be conducted. It is hoped that the following analyses will clearly reflect the appropriateness of the regression and correlation analyses as a budgeting tool for costs of this nature.

Sales Department Functions

Costs are incurred by the Sales Department in order to perform the following broad functions:

1. To promote sales of the Grace Company's products to dealers. No sales are made to wholesale organizations or through company owned retail outlets. In general, the sales territory encompasses a five-state geographical area; however, the parent company makes sales throughout the entire United States and abroad.

2. To assist dealers in handling especially difficult service and installation problems involving the company's products.

3. To assist dealers in planning and preparation of blueprints when they are submitting contract bids for especially complex projects.

4. To assist, inform, and educate the company's dealers regarding new sales techniques, financing their operations, and proper accounting procedures.

5. Costs are also incurred by this department for preparation of the sales orders, back-order notices, shipping reports, sales invoices, replacement of defective in-warranty products, etc.

Accounts Employed

Twenty-eight account titles are employed by the company in which the Sales Department costs are accumulated. A few of the accounts contain reasonably small balances relative to the department's total operating cost, but because of the varying nature of the functions performed, none of the accounts were consolidated. This procedure is, of course, different from that employed in analyzing the Shipping and Warehousing Department costs.

Although the company does not classify the cost accounts in the Sales Department according to functions or nature, it was felt that such a procedure would be beneficial in conducting the initial analyses of the department's costs and would subsequently assist in developing cost estimating equations. A list of the accounts employed, grouped by functions and nature, follows:

Class A

Salesmen's Salaries
Salesmen's Commissions
Salesmen's Supplies
Salesmen's Travel Expenses
Salesmen's Entertainment Expenses

Class B

Advertising Expenses
Advertising Charged to Dealers

Class C

Plan Service Expense
Engineering Salaries
Service Salaries
Service Travel Expense
Dealer Service School Expense

Class D

Clerical Salaries
Stationery, Printing, and Office Supplies
Postage
Telephone and Telegraph
Repairs--Office Equipment

Class E

Executive Salaries
Executive Travel Expense
Repairs--Executives' Cars

Class F

Exhibit Expense
Salesmen's Sales Meetings
Dealers' Sales Meetings

Class G

Depreciation--Furniture and Fixtures
Depreciation--Automobiles
Depreciation--Buildings
Equipment Rental
Liability Insurance

Obviously all costs are incurred with one broad objective in mind--to secure sales, but the foregoing account classifications were developed under the assumption that all costs are not directly related to fluctuations in the sales volume. Further, the accounts were grouped by nature as well as by functions to assist in analysis. It may well be that many of the accounts will necessarily be reclassified after a detailed account analysis is conducted, as their individual true nature will only then be determinable.

Fixed Costs

Of the seven cost classes comprising the Sales Department's total costs, three appear to be almost entirely of a fixed nature. Cost Class E is comprised of salaries and travel expenses for the department's executives. Salaries in this group have remained relatively constant over the entire three-year period. Fluctuations in its total is attributable to variation in the monthly travel expense incurred. However, monthly travel expense is extremely uniform, with practically all variation being attributable to the number of working days in each month. On an annual basis, the total variation in the Class E costs has been less than 5 per cent over the three-year period. Therefore, Class E costs have been treated as programmed costs of a fixed nature.

Cost Class F consists entirely of programmed costs. The company budgets funds for these costs at the beginning of each year. Monthly charges that will accumulate to the 12-month budgeted total are made. But since the actual expense is incurred during one or two months of the year, it was necessary to adjust the monthly budgeted expense for the difference between the actual and budgeted expense incurred.

The Class G costs are strictly fixed. The company recognizes a portion of the Class G costs monthly on the basis of budget estimates. At the end of the calendar year, the December expenses are adjusted for any previous estimating errors. Therefore, for purposes of this analysis, it was necessary to restate the monthly estimated expenses to actual.

Variable Accounts

The remaining cost classes contain programmed, semi-variable, and variable costs in each class. The reason for not segregating the various costs by variability is that it would be desirable to associate the costs as a group with their cause. Functional classification should also decrease the number of estimating equations that it will be necessary to develop.

Cost Class A contains five accounts, all relating directly to the sales function. Each account was reviewed for recording and timing errors. Except for the Salesmen's Supplies expense, the monthly data reported on the books were used in the analysis without adjustment. As for supplies, the company recognizes this expense when the supplies are purchased rather than used. After discussing the matter with members of the Sales Department, it was decided that the supplies expense should be allocated to each month, based on sales. The effects of errors in allocating the supplies expense averages less than one-half of one per cent of the total Class A costs.

Advertising expenses comprise Class B. All advertising is conducted through the dealers. Expenses are incurred for co-op newspaper, radio, and television ads. Brochures of several types are prepared and mailed by a contracting printer who invoices the company for the entire amount. The company in turn charges its dealers for a portion of this cost. In general, the company charges the dealers for their proportion of the brochure cost at the time the printer's invoice is received. Therefore, the timing of debits and credits to advertising expense should be accurate. No adjustments were made to advertising expense, as it appeared on the books for purposes of correlation analysis.

The Class C costs consist almost entirely of programmed costs. Exceptions are the Plan Service Expenses and the Service Travel Expense; however, no adjustments were made to either of these accounts for possible errors in the timing of their recognition. The Dealer Service School Expense is incurred during one month of each year. Provisions are made in the annual budget for this expense, but not on a monthly basis. For purposes of this analysis, the actual expenses for this activity were allocated over the entire year.

The Class D costs required extensive review, and considerable time was spent before this researcher was satisfied as to the accuracy and timing of the reported expenses. Major adjustments were made in the timing of the recognition of postage expenses, and minor changes were made in the timing of the recognition of telephone expenses. Postage is charged to expense at the time purchased. Fortunately the mailing department maintained a monthly record of the postage used by the Sales Department as well as by other departments. This information served as the basis for adjusting the postage expense. The telephone expense was adjusted on the basis of records maintained by the Company's PBX operator.

The clerical salaries and equipment repair expense accounts were not adjusted. Clerical salaries expense is incurred for secretarial work, order clerks and other sundry functions. In examining this account it was found that there were occasions when two clerks were being paid for doing the same job. This happened only when a new clerk or secretary had been employed to replace one who was leaving. However, no adjustments were made for this, on the assumption that it was a normal and frequent occurrence.

On the other hand, such a procedure may be expected to decrease the value of the coefficient of determination to be computed.

The account which proved most difficult to restate accurately for analysis purposes was the Stationery, Printing, and Office Supplies Expense. The Department purchases these items in sufficient quantities to last for several months to a year. The supplies are placed in a storeroom and drawn on as needed. No inventory or other similar records are maintained. However, a large proportion of the total supplies consists of data processing forms and order and invoice forms. Many of the forms were prenumbered and accurate records had been kept on the number used each month. Therefore, based on this information and interviews with relevant personnel regarding use patterns for other supplies, adjustments were made to this expense account so that it appears reasonably correct.

Considerable time and effort was spent in conducting the foregoing account analyses and interviews. As a result of this phase of the analyses, several changes were made in the original data, as reported on the company's books, but this researcher's confidence in the accuracy of the data was greatly improved. On the other hand, because some of the adjustments made were based on incomplete information and because of the composite character of the several Cost Classes, the data may still be somewhat inaccurate and not so amenable to correlation analysis as one might desire. Therefore, it is highly probably that the values obtained for the various coefficients of determination will be somewhat lower than those obtained in the Shipping and Warehousing Cost Analysis.

At this point in the analyses of the Sales Department Costs, several classes of costs have been established; the accounts have been reviewed, and adjusted when necessary, for timing and clerical errors. The next step is to chart each cost class and analyze its behavior patterns. By doing this, it should be possible to determine and select logically the independent variables. A preliminary estimating equation will then be developed for the cost class under analysis. If a satisfactory estimating equation cannot be developed for the cost class then under analysis, another class will be subjected to analysis. If by following this procedure, satisfactory estimating equations are developed for one or more of the cost classes, these classes will be set aside and the remaining classes will be subjected to further analysis until a satisfactory disposition is made.

A Possible Explanatory Variable

Since it is the costs of the Sales Department that are being analyzed, it would seem that several of the cost classes should be related to the net-sales series. Therefore, net sales for the 36-month period of January 1, 1964, through December 31, 1966, are plotted on Chart 4.1. None of the Sales Department's costs are shown on Chart 4.1 because of the differing magnitude of the various series. The function of Chart 4.1 is merely to reveal the timing and amplitude of fluctuations in the net-sales series. Examination of this chart will provide a point of reference for analyzing Charts 4.2, 4.4, and 4.6, which contain the various sales expense series. It should be noted that the most probable independent variable series is being introduced in conjunction with the dependent variable series, which of course, differs from the analysis pattern for the Shipping and Warehousing costs.

Examination of Chart 4.1 readily reveals that the company's net sales contain considerable seasonal variation. October, November, December, and January are the months during which sales are lowest. However, no single one of the four months is consistently the lowest sales month. June is consistently the peak sales month. Other months, after which the direction of the sales curve changes, are March and September, except that in 1967 sales did not increase again in September as in previous years. The sales trend for the peak month of June, and in annual terms, is positive. But the minimum points on the sales curve appear approximately horizontal. In the high dollar sales months, fluctuations, in relative terms, are of considerable magnitude.

For purposes of regression and correlation analysis, this type of curve, representing time series data, presents several difficulties. For example, the fluctuations are extreme, but their timing is inconsistent over the three-year period. Further, seasonal variation apparently exists, but it also varies in timing and magnitude. Therefore, if the net-sales series is to be employed as an explanatory variable and if one desires to remove seasonal variation, numerous difficulties may be encountered. On the other hand, it may not be necessary to smooth the data for seasonal variation, provided the same variation is contained in the dependent variables. At this point, the researcher should be aware of the possible problems and keep these in mind when conducting analyses of the respective dependent variable series.

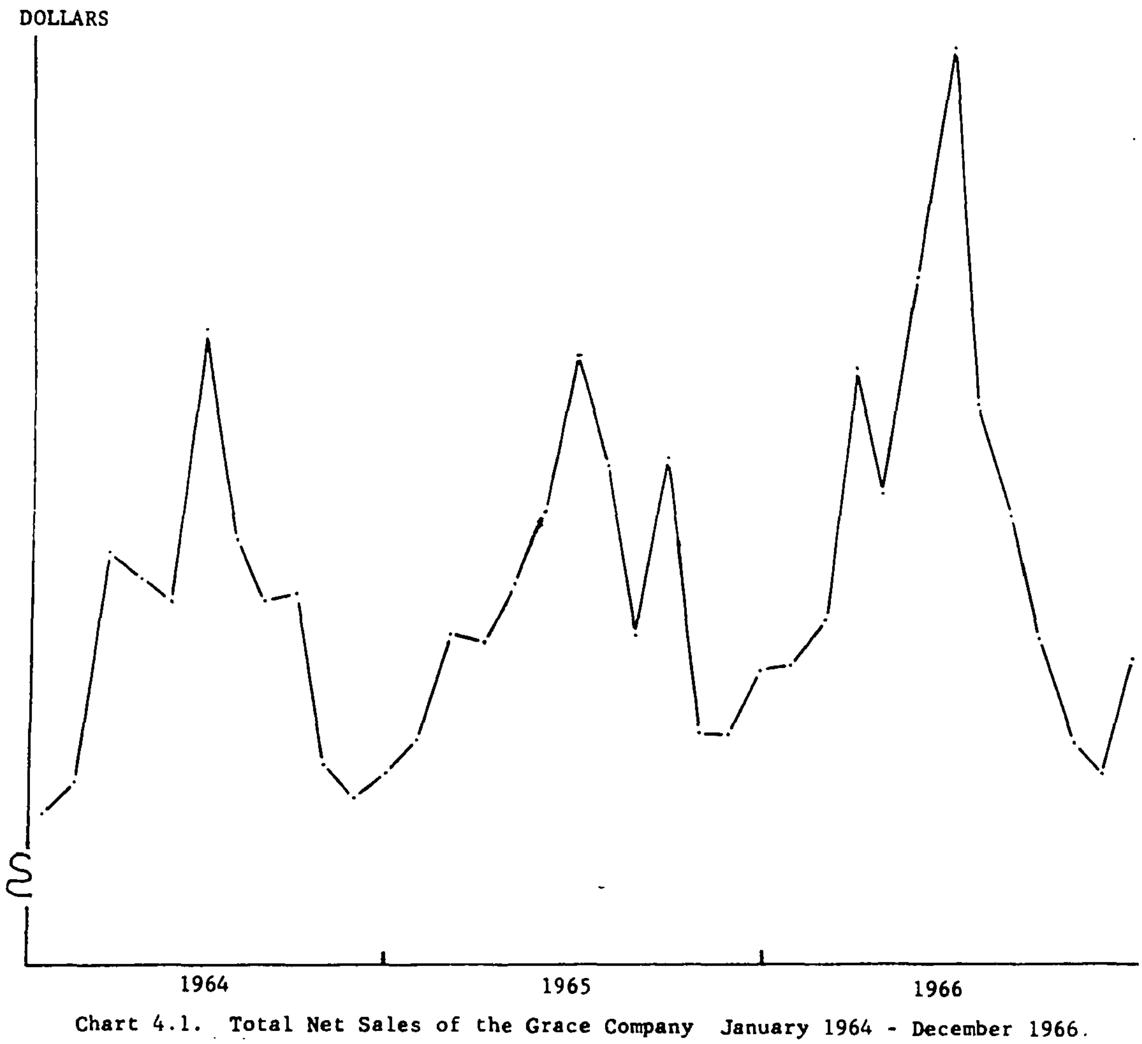

Chart 4.1. Total Net Sales of the Grace Company January 1964 - December 1966.

<u>The Class A Cost Equation</u>

The Class A cost series is presented on Chart 4.2. This cost class contains the salesmen's salaries account that is a programmed cost. Each salesmen is paid a base salary plus a commission on his sales. The base salary differs slightly between salesmen, but since the turnover among the salesmen has been quite low over the three-year period, this difference is not significant for computational purposes. The sales commission rate also differs between salesmen and is computed as a straight percentage of sales. However, because of the different commission rates paid, the varying proportion of total sales sold by each salesman, and the turnover of sales personnel, an r^2 of 1.00 would not result from correlating the commissions and net-sales series.

The accounts for travel expenses and entertainment of dealers appear to be programmed costs of a semi-variable nature. Limitations are placed on the salesmen regarding entertainment expenditures. Travel expenses are paid based on a fixed mileage rate and a per diem. The expenditure patterns for both these expenses logically should be related to sales, but in addition, they should also be related to the number of working days in each month.

Examination of Chart 4.2 reveals a definite positive trend in costs. Because the average sales commission rate has increased periodically, and because travel and entertainment expense has also increased, the minimum points on the series are not horizontal as in the case of the net-sales series. There are pronounced peaks in the series curve for March, June, September, and December. This behavior adds validity to the previously

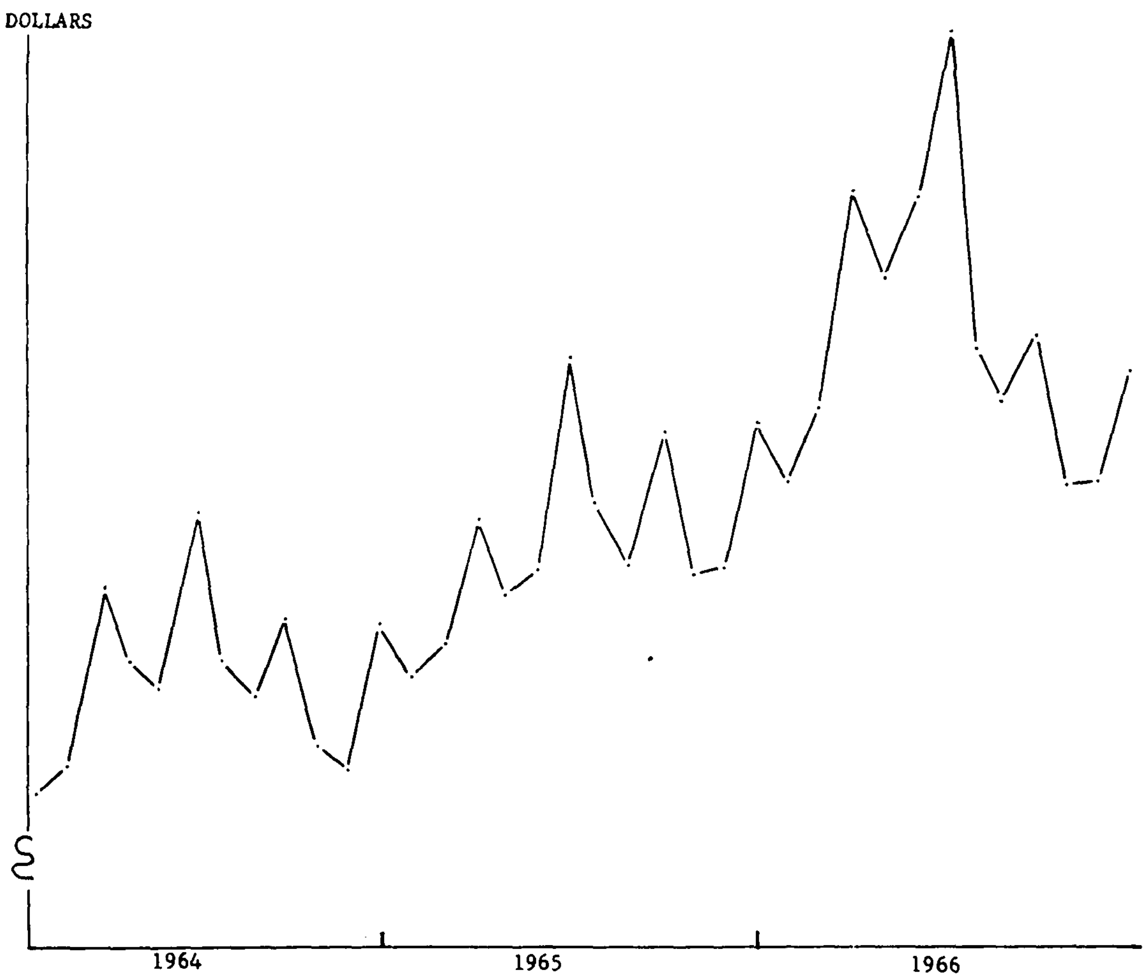

Chart 4.2. Class A Costs of the Sales Department Costs January 1964 - December 1966.

mentioned hypothesis that working days per period might be a possible independent variable.

Based on the foregoing discussion, it would seem that the Class A costs are a function of at least three quantifiable series, i.e., net sales, time, and number of working days per month. The first trial estimating equation will be of the form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

where:

X_1 = Total monthly Class A costs

X_2 = Time

X_3 = Monthly net sales

By using this type of equation initially, there will be less likelihood of introducing distortion from the possible existence of multicollinearity between net sales and the number of working days per month.

Using a 36-month period, the following estimating equation was computed:

$$X_{c1.23} = \$27,413 + \$355.88X_2 + \$.0123X_3 \qquad \text{(Eq. 10)}$$

The associated $R^2_{1.23} = .9167$, which is extremely high, and the $S_{y1.23} =$ \$1,739.79. The .95 confidence limits of $\pm\ 1.96S_{y1.23} = \$3,410$ is less than 10 per cent of any of the $X_{1...36}$ values. Tests of significance were made at the .05 level for $R^2_{1.23}$, $r^2_{12.3}$, and $r^2_{13.2}$. All three statistical measures were significant by wide margins.

To check the possibility of a curvilinear relationship between X_1 and X_3, two more estimating equations were computed. The same variables were used, but additional terms of a higher order were included for X_3. In the first test an X_3^2 term was added. The resulting $R^2_{1.23^23}$ was only .35 per cent larger than that provided by equation (10), which is not statistically significant. In turn, there was a slight decrease in $S_{y1.23}$. In the second test, both X_3^3 and an X_3^2 term were added to equation (10). Inclusion of these terms did not improve the fit of the estimating equation nor did it reduce the standard error of the estimate. Therefore, on both logical and mathematical grounds, it may be concluded that X_1 is related to X_2 and X_3, and further, that the relationship between X_1 and X_3 is quite probably linear.

Even though equation (10) does provide a "good fit," additional tests were run using a working-day index in an effort to increase the value of R^2 even more. The equation form used initially was:

$$X_{c1.234} = a_{1.234} + b_{12.34}X_2 + b_{13.24}X_3 + b_{14.23}X_4$$

where:

X_1 = Total monthly Class A costs

X_2 = Time

X_3 = Monthly net sales

X_4 = Working-day index

These variables provided the following estimating equation:

$$X_{c1.234} = \$18{,}345 + \underset{(22.957)_2}{\$362.19X_2} + \underset{(.00097)_3}{\$.0105X_3} + \underset{(289.156)_4}{\$1{,}300.13X_4} \qquad \text{(Eq. 11)}$$

For equation (11) the $R^2_{1.234}$ = .9490 per cent, or 3.23 per cent more than $R^2_{1.23}$. The value of $r^2_{14.23}$, which is the per cent of previously unexplained variation in X_1 that was explained by the addition of X_4, may be computed as follows:

$$r^2_{14.23} = \frac{R^2_{1.234} - R^2_{1.23}}{1 - R^2_{1.23}} = \frac{3.23}{8.33} = 38.77\%$$

To test the statistical significance of $r^2_{14.23} = 38.77\%$ for equation (11), the F-Test should be used.

$$F = \frac{R^2_{1.234} - R^2_{1.23} \div m - (m - 1)}{1 - R^2_{1.234} \div (N - m)}$$

$$F = \frac{3.23 \div 1}{8.33 \div 32} = \frac{3.23}{.26} = 12.42$$

Using $n_1 = 1$ and $n_2 = 32$ the F-Table, at the .05 probability levels, provides a value of 4.15 and at the .01 probability level, a value of 7.50. Since for $r^2_{14.23} = 38.77\%$, F= 12.42, it appears that X_4 does explain a statistically significant percentage of the previously unexplained variation.

For equation (11) the $S_{y1.234}$ = \$1,383.08, which is \$356.71 less than $S_{y1.23}$. The confidence limits set at $1.96S_{y1.234}$ = \$2,711 is less than 7.2% of the Class A Costs for the lowest month and for the highest month it is less than 4.4%. Estimating equation (11) provides a very impressive mathematical fit and definitely appears statistically valid. However, before it is accepted as a model for the Class A costs, the possibility of a curvilinear relationship existing between X_1 and X_4 must be tested. In making these computations, the same form of equation was used. In the first test an X_4^2 term was included, in the second X_4^3 and X_4^2 terms were included, and in the third test X_3^2 and X_4^2 terms were included. None of these equation forms provided statistically significant increases in $R^2_{1.234}$ or meaningful decreases in $S_{y1.234}$. Therefore, it appears that equation (11) provides the best fit attainable in view of existing information constraints.

The final step in testing the validity of equation (11) is to use it to restate the Class A historical costs. The purpose for doing this is to determine the actual distribution of the $X_1 - X_{c1.234}$ values. Assuming a normal distribution, the unexplained variations between X_1 and $X_{c1.234}$ should be evenly distributed above and below the $X_{c1.234}$ curve. In addition, very few, if any, of the variations should be greater than 1.96 $S_{y1.234}$. If either of these two conceptual criteria is violated by equation (11), then satisfactory explanations must be developed or perhaps some type of adjustments will be required.

Chart 4.3 compares the actual cost experience with that computed using estimating equation (11). Of the 36 $X_1 - X_{c1.234}$ differences, 19 are above the X_1 curve and 17 are below. During the most current year in the series there are 5 differences above and 7 below the X_1 curve. The dispersion of the differences is quite good. The magnitude of the variances between the $X_{c1.234}$ and X_1 curve are well within acceptable limits. At the .95 confidence level, $1.96S_{y1.234}$ = \$2,711. Only one of the 36 variances fell outside this range, and 72% of the variances fell within a range of ± $1.0S_{y1.234}$. Therefore, it appears that equation (11) may be accepted as a valid model of the Class A Costs.

The Class B Cost Equation

The Class B expenses (advertising) are shown on Chart 4.4. Again, annual variation is present, with the low expenditure months being August, October, November, and December. This is a somewhat different pattern from that of either the Class A costs or the net-sales series. However, the

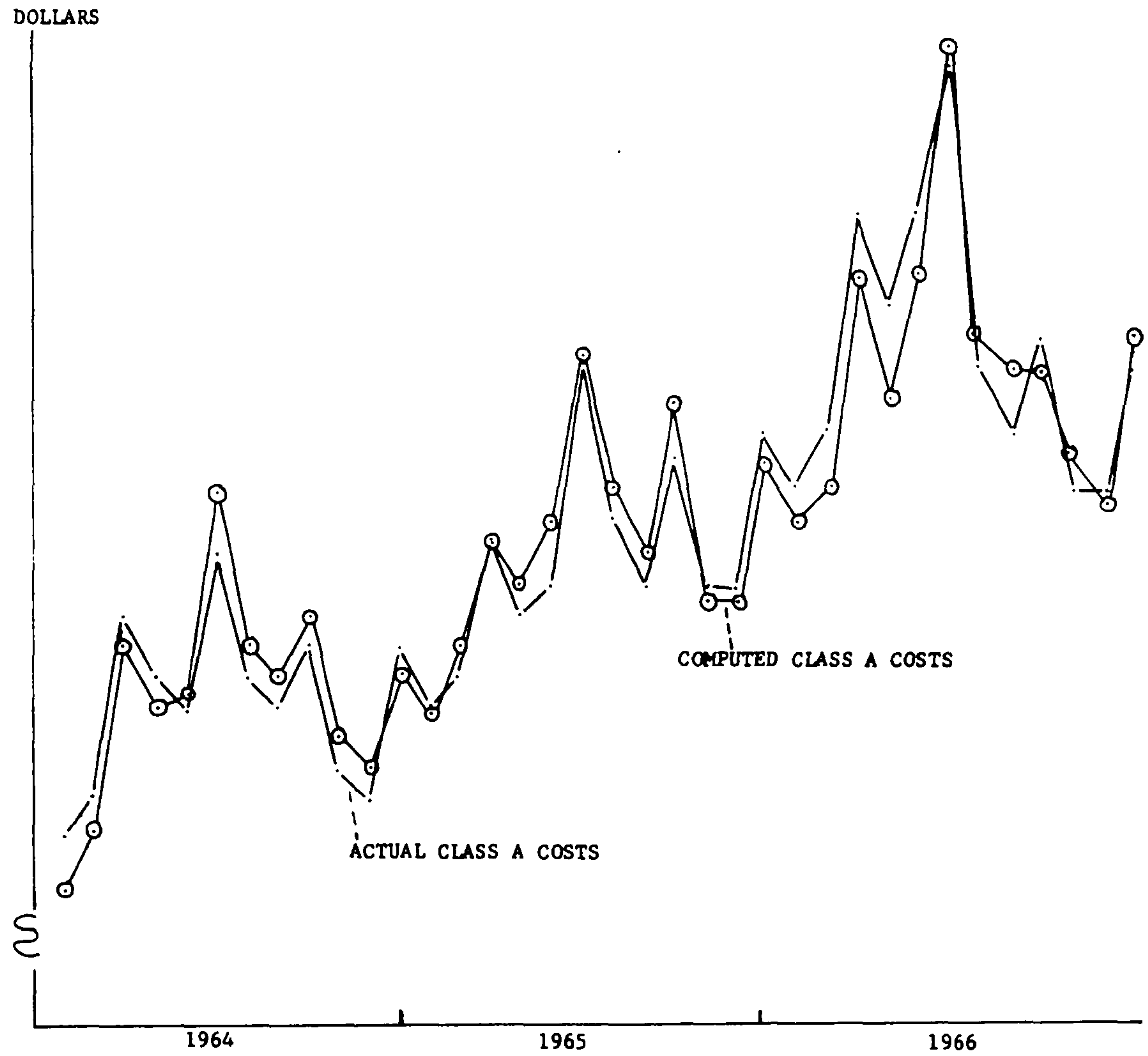

Chart 4.3. Actual Class A Costs and Computed Class A Costs of the Sales Department Costs, January 1964 - December 1966.

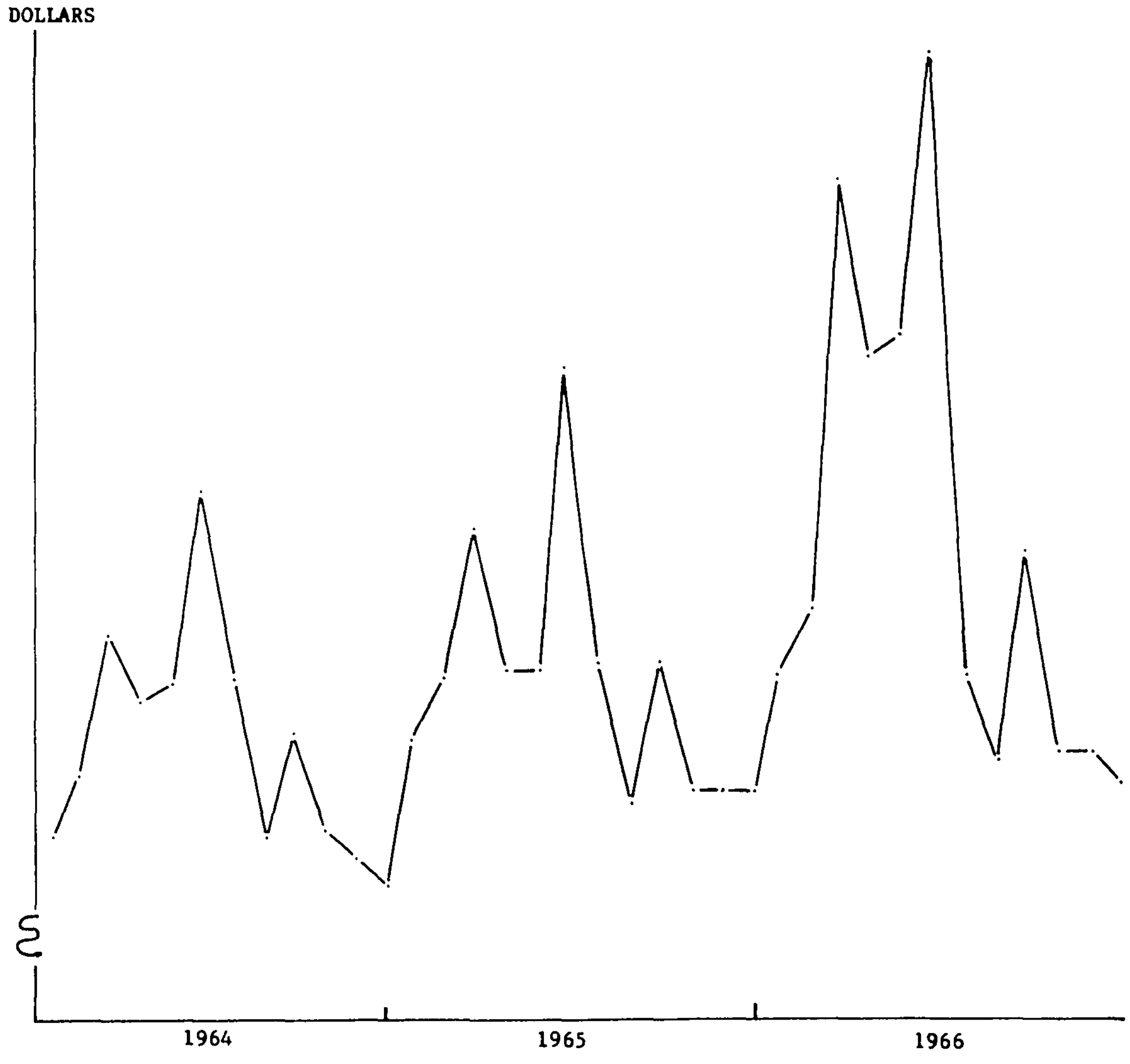

Chart 4.4. Class B Costs of the Sales Department Costs. January 1964 - December 1966.

peak months are the same in all three. The trend in the peak months is profoundly positive, while for the minimum months the trend, though positive, is not nearly so pronounced.

The magnitude of the fluctuations in the advertising expense is extremely large. In fact, the expenses incurred during June of each year are over four times as large as the expenses incurred during the low months of the same year. When large fluctuations of this type exist, there is a high probability that the standard error of the estimate, computed from the estimating equation, will be quite large. This may hold true even though the R^2 value is quite high. In other words, examination of Chart 4.4 indicates that difficulty may be encountered in narrowing the confidence limits to a meaningful value.

Logically, advertising expense should be closely related to sales and this appears to be partially true. However, one would also expect a lead-lag relationship of some type to exist, but for the most part, this does not seem to be the case. Thus, it appears the net-sales series may be a significant explanatory variable, but it will not be necessary first to determine a lead-lag relationship.

It also appears that the number of working days in each month may be a relevant explanatory variable. There are peaks during those months having the most working days, except for December, which is a minimum-cost month. The general annual pattern of advertising expenditures appears to be one wherein the expenditures commence increasing in January until a peak is reached in March. In April and May high expenditures are made and then in June the annual peaks reached. Thereafter, monthly expenditures fall rapidly and remain low except for September. Very real implications for the estimating equation result from this behavioral pattern. That some type of index of seasonal variation must be developed is obvious if a valid estimating equation is to be computed.

The first computational effort will employ the following equation form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

where:

X_1 = Monthly advertising expenses

X_2 = Time

X_3 = Monthly net sales.

The estimating equation provided was:

$$X_{c1.23} = \$-16,694 + \$97.86X_2 + \$.0367X_3 \qquad \text{(Eq. 12)}$$

For equation (12), $R^2_{1.23}$ = .7066, which is significant at the .05 probability level. The $S_{y1.23}$ = $6,797.30, which is quite large when related to expenditures in the low months. In an effort to reduce the standard error of the estimate and increase R^2, but still employing the same variables, two additional tests were run. In the first iteration an X_3^2 term was added and in the second, X_3^3 and X_3^2 terms were included. The first iteration increased the coefficients of multiple determination by .0202, which is marginally statistically significant. The second iteration resulted in no increase whatsoever in R^2. Neither of the tests reduced $S_{y1.23}$ by a meaningful amount. Therefore, based on these results, one may conclude that a relationship between X_1 and X_3 does exist and that it is linear.

The next equation test used the following variables:

X_1 = Monthly advertising expense

X_2 = Time

X_3 = Monthly net sales

X_4 = Working-day index

The linear estimating equation computed from these variables provided an $R^2_{1.234} = .7241$ or an increase of .0175. The .0175 is again marginally statistically significant, but is of little real value. In addition, the $S_{y1.234} = \$6,694$ which is only slightly lower than $S_{y1.23}$. On the basis that the relationship between X_1 and X_4 might be curvilinear, a second computation was made which employed X_4^2 and X_4 terms. The estimating equation computed was:

$$X_{c1.234^24} = \$405,721 + \$90.97X_2 + \$.0345X_3 + \$5,934.69X_4^2 - \$100,389X_4 \qquad \text{(Eq. 13)}$$

Equation (13) is considerably different from equation (12) because of the (a) term and the X_4^2 and X_4 terms. However, the $R^2_{1.234^24} = .7548$, which is a .0482 increase over the $R^2_{1.23}$ of equation (12). Individually, $r^2_{14^2.234}$ and $r^2_{14.234^2}$ explain a statistically significant percentage of the previously unexplained variations. After testing the coefficient of multiple partial determination $r^2_{14^24.23}$, it appears that the explanatory variable X_4 should be retained in the estimating equation.

Although the estimating equation explains .7548 of the variation in X_1, it does not provide a standard error of the estimate that is reasonable. The $S_{y1.234^24} = \$6,412$, is too large if any type of budget control is to be exercised on a monthly basis and the standard errors are excessive. In order to reduce the range, in dollar terms, of the standard error of the estimate and the standard errors, it appears that some type of seasonal or cyclical index must be employed. Of course, there are other alternatives, such as converting each variable series to pure numbers, or each variable series could be converted into deviations from its respective mean and then correlated. But these methods are best suited for application by the professional researcher and involve numerous subleties, conversions, and reconversions of the data, all of which require vast amounts of time, which the practicing accountant may not have available. Therefore, it is desirable to develop solutions that retain the data series in their original raw forms.

Since the advertising expense appears to have what might be called an annual cycle, it was decided that a simple explanatory variable using time might be employed. The form of the variable was merely to list the months as 1 through 12 each year, starting again the second year with 1 through 12 and so forth. The resulting coefficient for this variable would consist, in general terms, of an average for each month.

Using the variables of:

X_1 = Monthly advertising expense

X_2 = Time

X_3 = Monthly net sales

X_4 = Working-day index

X_5 = Annualized cycle

and the linear equation form:

$$X_{c1.2345} = a_{1.2345} + b_{12.345}X_2 + b_{13.245}X_3 + b_{14.235}X_4 + b_{15.234}X_5,$$

provided a $R^2_{1.2345} = .8298$. The net increase over $R^2_{1.234}2_4$ of equation (13) was .0750, which is definitely significant. Each $r^2_{1n.n}$ for each explanatory variable was significant. In addition, the $S_{y1.2345}$ = \$5,342, representing a reduction of \$1,070. However, the standard error of the estimate is still large, and further, the coefficients for X_2, X_3, and X_4 were all significantly changed. This latter point suggests the variable X_5 introduced into the computations data having some degree of multicollinearity with each of the other three independent variables. For these reasons, this type of annualized cycle was eliminated as a possible independent variable.

By reviewing Chart 4.4 and further reflecting on the nature of advertising expense, it becomes undeniably obvious that one must consider the timing of advertising expenditures as being largely a function of management decisions. If that is true it should be possible to design an index which would, in effect, quantify the effects of past management decisions. Therefore, an index was developed by employing the following steps:

1. Sum the monthly advertising expenses to arrive at an annual total.
2. Determine the percentage of each month's expense to the annual total.
3. Compute a three-year average percentage for each month, to arrive at an annual index.

In computing this index, it was noted: (a) that a remarkably similar percentage was computed for the same months in each of the three years, and

(b) that over 65% of the total annual expenditures were consistently made during the first half of the year.

An estimating equation was then computed, using the following variables:

X_1 = Monthly advertising costs

X_2 = Time

X_3 = Monthly net sales

X_4 = Working-day index

X_5 = Seasonalized index

A linear equation form was used in this first test. The $R^2_{1.2345}$ = .8540, which is a .0992 increase over that provided by equation (13). The $S_{y1.2345}$ = \$4,946 representing a reduction of \$1,466. In total, these results are quite favorable. However, tests of the various coefficients of partial determination revealed that $r^2_{14.235}$ was no longer statistically significant. If one examines the rationale supporting inclusion of the X_5 variable, then it becomes quite logical that the inclusion of X_5 should eliminate the need for the X_4 variable. In order to verify this conclusion, calculations were made using only:

X_1 = Monthly advertising costs

X_2 = Time

X_3 = Monthly net sales

X_4 = Seasonalized index

and the same linear equation form as before. For this case, $R^2_{1.234}$ = .8532 or only .0008 less than $R^2_{1.2345}$. In addition, the $S_{y1.234}$ = \$4,884, which is a \$62 decrease. Therefore, the conclusion that the need for a working-day index was eliminated by the addition of a seasonalized index appears correct.

Now that the most likely explanatory variables have been determined, the next step is to determine the best possible fit. That is, the relationship between the dependent variable X_1 and each independent variable must be determined.

To this end, two tests were run to determine the relationship between X_1 and X_4. In the first test an X_4^2 term was included and in the second test both an X_4^3 and an X_4^2 term were employed. Neither of these tests indicated that the relationship existing between the two was not linear.

In previous tests, there appeared to be a possible nonlinear relationship between X_1 and X_3, so an additional test was made, using an X_3^2 term. The resulting estimating equation is as follows:

$$X_{c1.23^23 4} = \$5{,}743.76 + \underset{(91.05)}{\$375.75X_2} + \underset{(.000000741)}{\$.000001521X_3^2} - \underset{(.01114)}{\$.0261X_3}$$

$$+ \underset{(423.69)}{\$2{,}488.205X_4} \qquad \text{(Eq. 14)}$$

The $R^2_{1.23^234} = .8707$, which is statistically significant at the .05 probability level. The $S_{y1.23^234} = \$4{,}665.67$ and is higher than desired for the low expenditure months but quite satisfactory for the higher expenditure months. In view of the practical constraints existing in regard to explanatory variable information and the amplitude of the fluctuations in the dependent variable series, it appears that the $4,665.67 value for the standard error of the estimate is the minimum that can be derived.

Tests of significance were made on each of the coefficients of partial determination at the .05 level and all were determined statistically significant. In addition, each regression coefficient was tested, relative to their respective standard errors, using the t-test and each was significant at the .05 level. Therefore, in statistical and analytical terms, it appears that equation (14) demonstrates reasonably good fit. As a final step it is necessary to calculate all the $X_{c1.23^2 34}$ values and plot them against the X_1 values. This information appears on Chart 4.5. Examination of this chart reveals that of the 36 $X_1 - X_{c1.23^2 34}$ variances, 14 are above the computed curve and 22 are below. Though the distribution of the variances above and below the computed curve are skewed, it is not cause for deep concern; in fact, it was expected. First of all, the amplitude of the upward fluctuations in the X_1 series is extreme, relative to the values in the low months. Since the estimating equation is computed around an arithmetic mean, the extreme values are weighted more heavily--; thus an upward bias may exist. Second, the dispersion of the variances during the last twelve-month period was even. That is, 6 were above the $X_{c1.23^2 34}$ curve and 6 were below.

The range of the variations, in terms of dollars, was small, relative to the value of $S_{y1.23^2 34}$. Assuming a normal distribution, two variations fell beyond the $\pm\, 1.96S_{y1.23^2 34}$ range and 30 of the 36 monthly variations, or 83.3%, fell within $\pm\, 1.00S_{y1.23^2 34}$. For the entire three-year period, the total error $\longrightarrow 0$. In view of these results, equation (14) will be accepted as a model for the advertising expenses (Class B).

The Class C Cost Equation

The Class C costs will be analyzed next. This class, shown on Chart 4.6, is composed of expenses related to customer services. The services

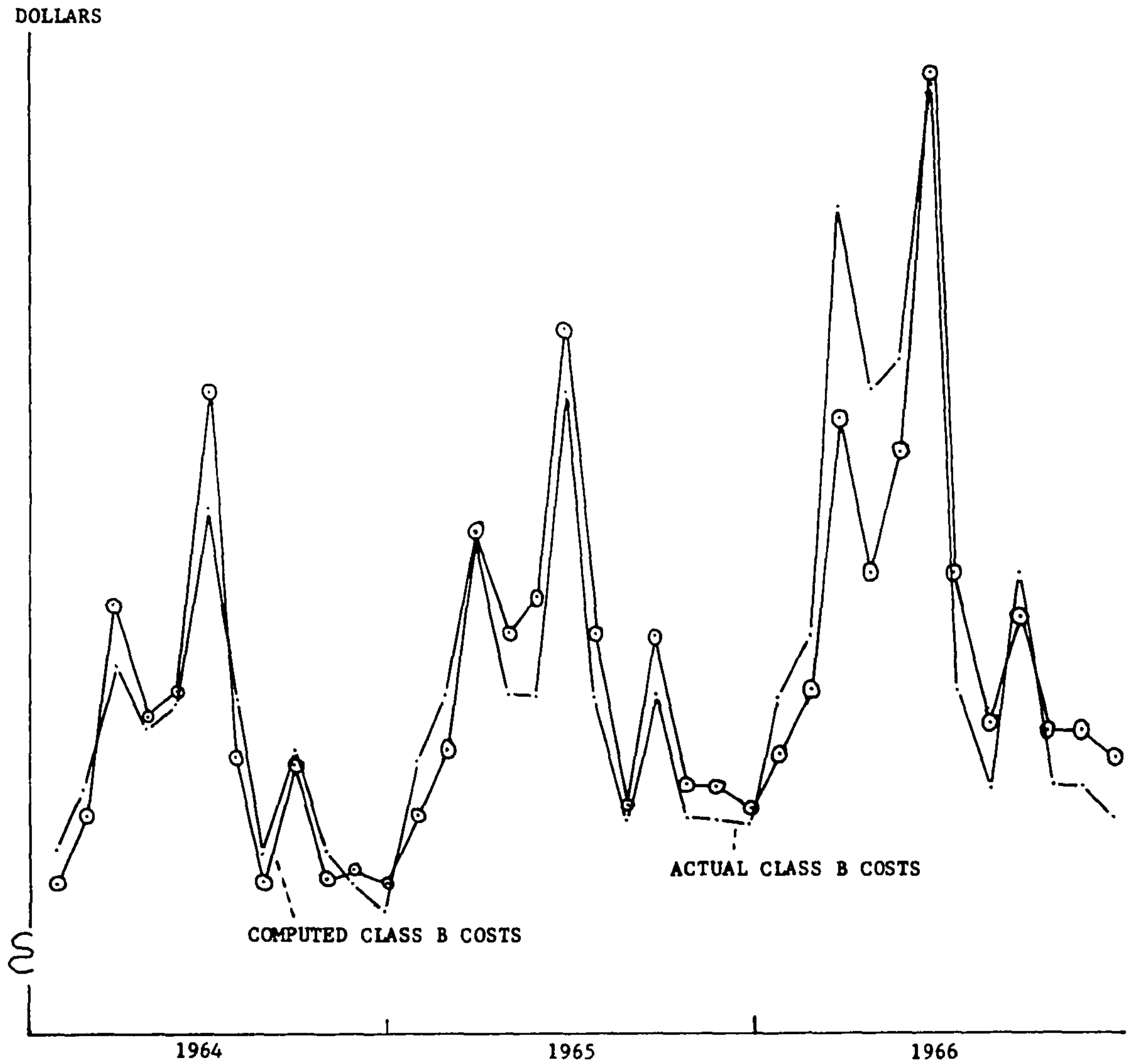

Chart 4.5. Actual Class B Costs and Computed Class B Costs of the Sales Department Costs, January 1964 - December 1966.

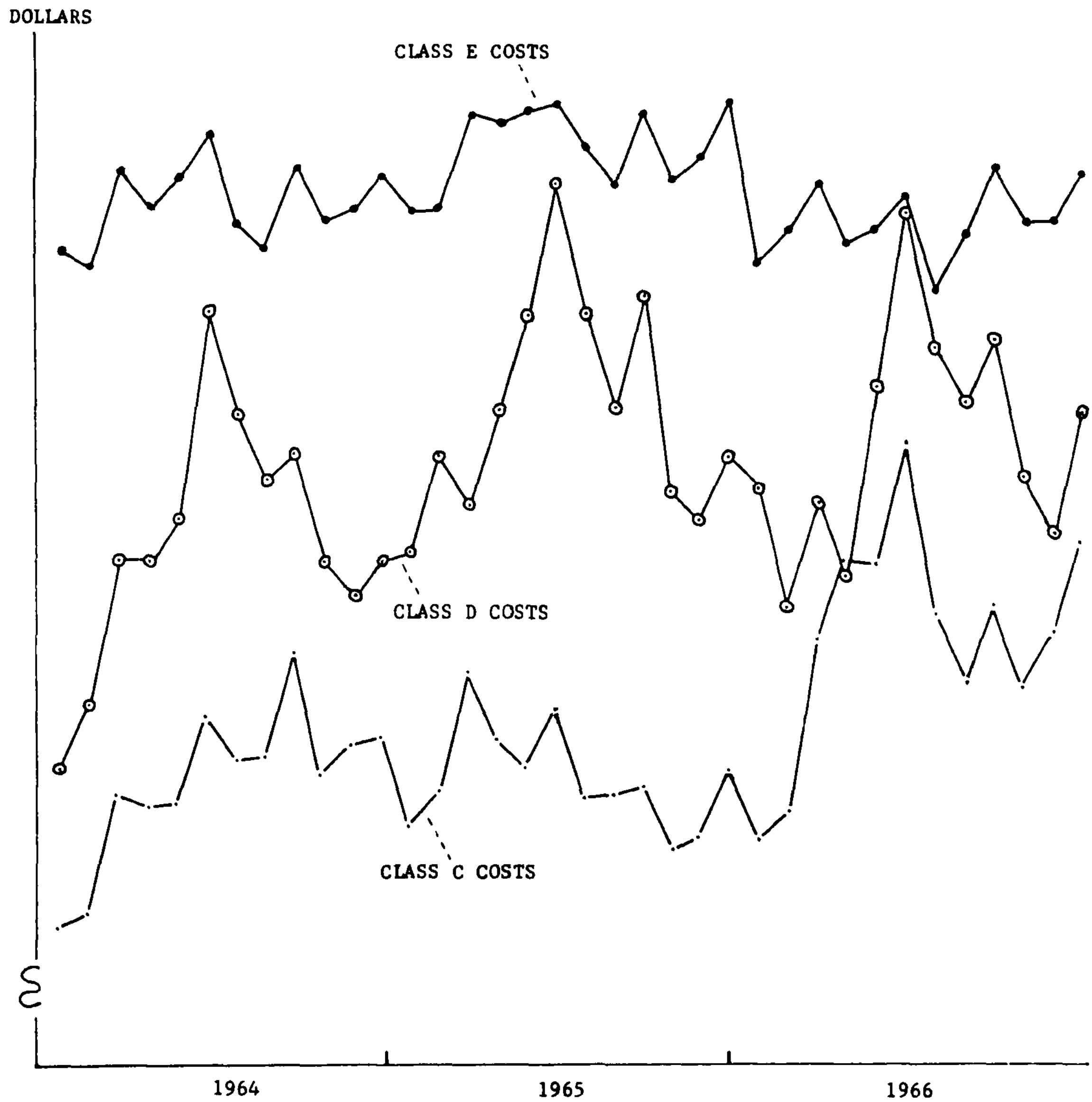

Chart 4.6. Class C Costs, Class D Costs, and Class E Costs of the Sales Department Costs, January 1964 - December 1966.

performed may be segregated into two groups: (a) Assistance in preparing plans and blueprints for dealers that are bidding on large contracts, and (b) Assisting the dealers in servicing the company's products both during and after installation and concurrently training dealer service personnel. Except for the Dealer School Expense, all the accounts demonstrate significant variability. The engineering and service salaries are no exception. A basic nucleus of engineering and service personnel is maintained, but as the work load increases, engineering and service personnel are drawn from various departments in the factory. Appropriate proportions of the salaries of those people on loan are charged to the Class C salaries expense accounts.

Upon referring to Chart 4.6, it may be seen that fluctuations in the Class C series have been consistent. During 1964 and 1965, the amplitude of the fluctuations was similar, but in 1966, any similarity vanished. The 1966 step-up in costs resulted, at least partially, from an increase in the number of personnel employed. It appears the trend of the series is almost horizontal for the first two years and then the trend curve, conceptually, should be discontinuous, becoming slightly positive in 1966.

Possible explanatory variables candidates are the net-sales series, the working-day index, a series representing the number of service calls requested and plans prepared, and perhaps an employment index. Of the foregoing series, the most logical candidate would be the number of service calls requested and plans prepared. However, interviews with several of the personnel in the department revealed that no records of this type had been maintained and further that it was quite likely that records of this type would be of little assistance anyway because of the differing lengths of time spent on each call and the travel time required to reach the problem sites. Therefore, this series was discarded as a possible explanatory variable.

The net-sales series would seem to be an appropriate variable, since it reflects both the quantity and type of units being shipped each month. Since one of the department's major functions is to assist in installation problems, there should be a relationship between the two series.

The working-day index is included on the grounds that travel expenses were larger during the months having the greatest number of working days. An employment index seems logical on the grounds that salaries are a function of the number of employees and that travel expenses would be greater as the number of personnel traveling increases.

Admittedly the foregoing series, which will perhaps be used as explanatory variables, are not causal variables to the extent one might desire. However, since data for the actual causal variables are not available, they must of necessity be employed. Of course this will not destroy the usefulness of the estimating equation as one may justifiably assume that the relationship(s) existing between the explanatory variables employed and the true causal variables will continue to exist in the future.

With these considerations in mind, the first calculations were made, using the following variables:

X_1 = Monthly Class C costs

X_2 = Time

X_3 = Monthly net sales

The equation form employed was:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

For this equation the $R^2_{1.23} = .5952$, which was somewhat better than expected. The $S_{y1.23} = \$478.43$, which was reasonably small in view of the $R^2_{1.23}$ value. Including an X_3^2 term in the equation resulted in an $R^2_{1.23^23}$ = .6122 and an $S_{y1.23^23} = \$475.48$. The 1.70 per cent increase in R^2 should probably not be considered significant at this R^2 level. Inclusion of higher order X_3 terms in additional calculations failed to increase significantly R^2. Therefore, it appears that the net-sales series has a linear relationship to X_1 and that it also explains a statistically significant percentage of the total variation in X_1.

The next step will be to test the effect of adding to the equation an X_4 term representing a working-day index. A linear equation form was again employed. The estimating equation developed follows:

$$X_{c1.234} = \$1,679.59 + \$36.54X_2 + \$.00096X_3 + \$207.32X_4 \qquad \text{(Eq. 15)}$$

This equation provided an $R^2_{1.234} = .6478$ and a $S_{y1.234} = \$453.13$. The coefficients of partial determination for each explanatory variable were statistically significant. However, it was felt that a larger percentage of the variation in X_1 should be explainable. Therefore, several additional tests were run, using the same independent variables but adding various combinations of higher order X_2, X_3, and X_4 terms. Several of these tests did provide larger coefficients of multiple determination, but the increases were small. Therefore, it appears that it will be necessary to include in the computations a fifth variable representing an employment index. The employment index represents the number of full time employees each month, based on a forty-hour work week.

Thus, the following variables were used in a linear equation form.

X_1 = Monthly Class C costs

X_2 = Time

X_3 = Monthly net sales

X_4 = Working-day index

X_5 = Employment index

The estimating equation computed provided an $R^2_{1.2345} = .9588$, which is a .3110 increase over that of the previous equation. In addition, the new $S_{y1.2345}$ = \$157.38, which is indeed excellent. Mathematically and statistically this equation provides an excellent fit. It now appears that sufficient explanatory variables are being employed, but tests should be made to determine what form of equation will be the most efficient. Before introduction of the employment index series into the equation, tests were made to determine the relationships between X_1, X_3, and X_4, but they should be retested for the purposes of determining if significant multicollinearity now exists.

Five additional estimating equations were computed, using the same variables, but higher order terms were used. In the first test an X_5^2 term was included and the results indicated that the relationship between X_1 and X_5 was linear. In the second test an X_3^2 term was included and in the third test an X_4^2 term was added. Both of these tests indicated linear relationships. In the fourth test an X_2^2 term was included. This was done after reexamining the X_1 series and concluding that it might contain a curvilinear trend, relative to the X_3 series. This proved to be correct so a fifth test was made using X_2^2 and X_3^2 to see if any material change in R^2 would occur. The results of the fifth test were negative in that the X_1 and X_3 relationship still appeared linear, after the curvilinear trend was considered. Based on the foregoing tests, the following estimating equation provided the best fit.

$$X_{c1.2^2 2345} = -\$1,904.96 + \underset{(.262)}{\$.5363} X_2^2 - \underset{(5.596)}{\$11.731} X_2 + \underset{(.0000691)}{\$.00019765} X_3$$

$$+ \underset{(31.75)}{\$257.0258} X_4 + \underset{(38.61)}{\$577.169} X_5 \qquad \text{(Eq. 16)}$$

Equation (16) provided an $R^2_{1.2^2 2345} = .9636$, which is definitely statistically significant. The $S_{y1.2^2 2345} = \$150.53$ and $1.96\ (150.53) = \$295.04$, which is less than .0700 of X_1 for the lowest month. Tests of significance were conducted for each of the coefficients of partial determination and all were significantly greater than zero at the .05 level. But it must be observed that X_4 and X_5 were the most significant. A t-test was used to test the regression coefficients relative to their respective standard errors and each was significant at the .05 level.

There appears to be multicollinearity between X_3 and X_5; however, this was anticipated and should not be viewed with alarm in this particular case. The reason is that equation (16) is to be used only for the purposes of describing the monthly Class C costs and not for marginal measurements of the various independent variable coefficients. Therefore, even if multicollinearity does exist between X_3 and X_5, it does not destroy the predictive capabilities of the estimating equation, so lomg as the relationship between X_3 and X_5 remains relatively constant. In view of the fact that 36 monthly observations were used in the calculations, it would seem that such an assumption should be valid for future months as well.

The actual historical experience and the computed curve are plotted on Chart 4.7. Of the 36 $X_1 - X_{c1.2^2 2345}$ differences, 20 are above the computed curve and 16 are below. Given that a dispersion of 18 differences above or below the computed curve would be exactly even, it would seem that the dispersion actually obtained is quite satisfactory. For the most current twelve-month period the dispersion of the differences was 7 above and 5 below the computed curve.

For the three-year period, the total error in dollar terms $\longrightarrow 0$. This, of course, is to be expected in using correlation analysis, but its validity is enhanced by the smallness of the $S_{y1.2^2 2345}$. Only one difference occurred outside the $\pm\ 1.96S_{y1.2^2 2345}$ confidence interval and 75.0 per cent of the differences were within $\pm\ 1.00S_{y1.2^2 2345}$. In view of these results, it would appear that equation (16) may be justifiably accepted as a model for Cost Class C.

The Class D Cost Equation

The next Sales Department cost segment subjected to analyses was the Class D costs. The general accuracy of the component accounts, in terms of measurement and timing, was discussed in a previous section. What must now be reviewed is the nature of the various component accounts.

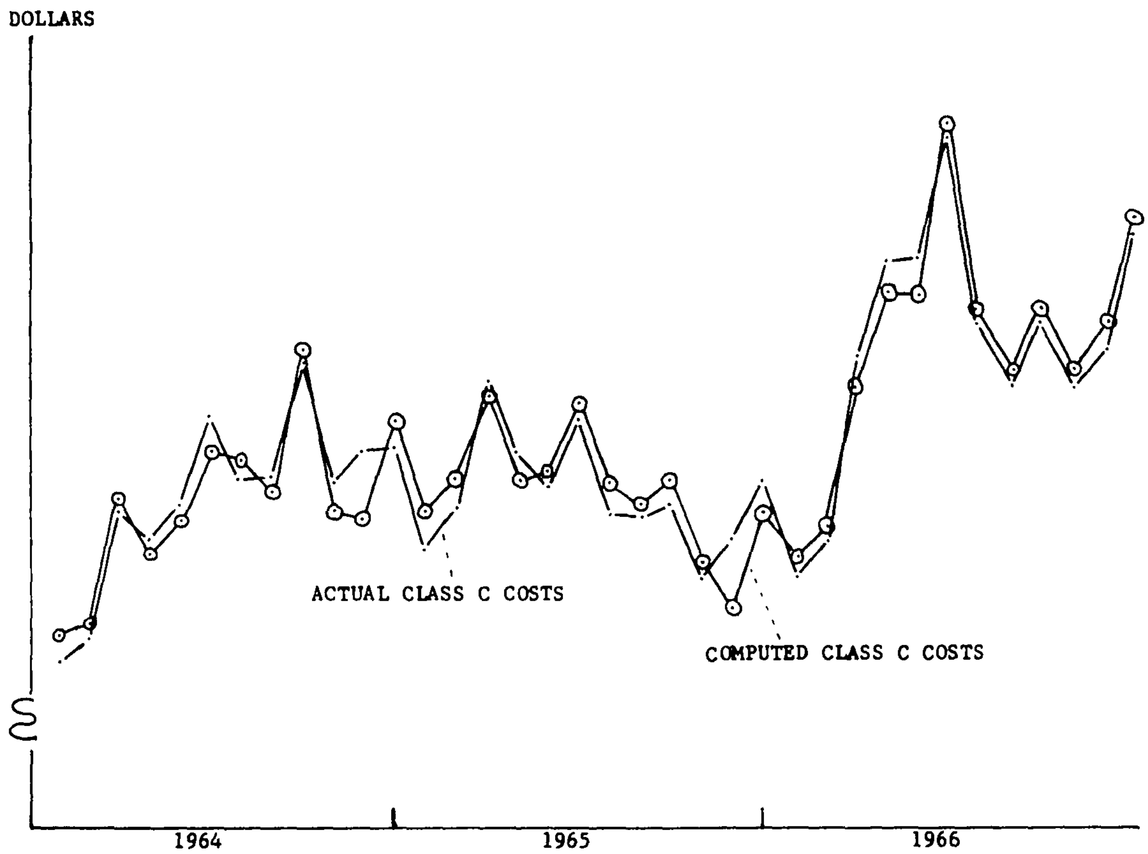

Chart 4.7. Actual Class C Costs and Computed Class C Costs of the Sales Department Costs, January 1964 - December 1966.

The Class D Cost series appearing on Chart 4.6 demonstrates considerable variability over the three-year period. This variability belies the generally accepted opinion that these types of cost may be safely regarded as programmed fixed costs. All the component costs demonstrate variability and, at first glance, the total costs may appear to fluctuate almost in accord with the net-sales series. However, logic insists that these costs should not move directly with net-dollar series; rather each series probably moves with a third set of variables. For example, it would seem that clerical salaries should be primarily a function of the work load, that letters, sales orders, and invoices prepared should be a function of the number of orders received and so forth. All the costs of this class are interrelated and may or may not be related to net-dollar sales.

It was decided that this cost class could perhaps be fairly accurately reproduced by utilizing two explanatory variables. The first series, and the one which should be by far the most significant, is the total number of orders and invoices handled. This series also includes the number of replacement orders and credit memos. The data for this series were taken from records maintained by the company's data processing section. As was mentioned previously, tests establishing that the tabulations for the series were relatively accurate were made.

The second series to be used will be the working-day index. This series was used in the Class A Costs analysis and because of the different number of working days in the various months, appears to be an appropriate explanatory variable for this case as well.

The first calculations were made employing the following variables:

X_1 = Monthly Class D costs

X_2 = Time

X_3 = Number of forms processed

Equation form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

This linear, three variable equation provided an $R^2_{1.23} = .5396$. The $r^2_{12.3}$, the $r^2_{13.2}$, and the $R^2_{1.23}$ were all significant at the .05 level. The $S_{y1.23} = \$980.55$, which is not as small as was desired, that is $1.96S_{y1.23}$ is equal to approximately $.40X_1$ in the lowest month.

The inclusion of an X_3^2 term in the equation provided an $R^2_{1.23^23} = .6112$. This is a net increase of .0716, which is definitely significant. However, the $S_{y1.23^23} = \$915.06$, which is not materially less than $S_{y1.23}$. Based on further analyses, it appears that neither the R^2 would be materially increased nor the S_y would be materially decreased by including additional X_3 terms in the equation. Therefore, the next step is to include additional explanatory variables in the computations. To this end, the following variables were employed.

X_1 = Monthly Class D expenses

X_2 = Time

X_3 = Number of forms processed

X_4 = Working-day index

Using a linear equation form, the $R^2_{1.234} = .7953$, which is a statistically significant increase of .1841. In addition the $S_{y1.234} = \$407.61$, which is quite satisfactory. Next an X_4^2 term was included in the equation, which increased $R^2_{1.234^24}$ to .8254 and decreased the standard error of the estimate to \$382.45. Finally, both an X_3^2 and an X_4^2 term were included in the estimating equation. Addition of these two terms resulted in an $R^2_{1.23^234^24} = .8285$, but the $S_{y1.23^234^24}$ changed slightly. These results, of course, indicated the inclusion of an X_3^2 was not desirable. Several additional calculations were made using increasingly higher order X_3 and X_4 terms. The results of these computations indicated that the inclusion of higher order X_3 terms would not improve the equation's efficiency. Therefore, it was decided that the following equation could be safely accepted as being descriptive of the X_1 series.

$$X_{c1.234^24} = \$-28{,}437 + \underset{(6.337)}{\$18.23X_2} + \underset{(.0572)}{\$.526X_3} - \underset{(182.475)}{\$422.23X_4^2} + \underset{(3{,}152.10)}{\$7{,}283.81X_4} \qquad \text{(Eq. 17)}$$

The $R^2_{1.234^24} = .8254$ and the $S_{y1.234^24} = \$382.45$. The $r^2_{12.34^24}$, $r^2_{13.24^24}$, $r^2_{14.^2234}$, and the $r^2_{14.234^2}$ were each statistically significant at the .05 level. The X_{c1} values from equation (17) were plotted against the X_1 values on Chart 4.8.

Examination of this chart reveals a reasonably good fit exists between the two series. Of the 36 $X_1 - X_{c1.234^24}$ differences, 22 are above the computed curve and 14 are below. For the most current twelve months the dispersion is even, with 6 differences above and 6 below the computed curve. The range of the 36 differences appears satisfactory. Only two differences fell outside the $\pm$ $1.96S_{y1.234^24}$ range and .6667 of the differences occur within $\pm$ $1.00S_{y1.234^24}$. Therefore, it appears that equation (17) may be employed as a satisfactory model of the Class D Costs.

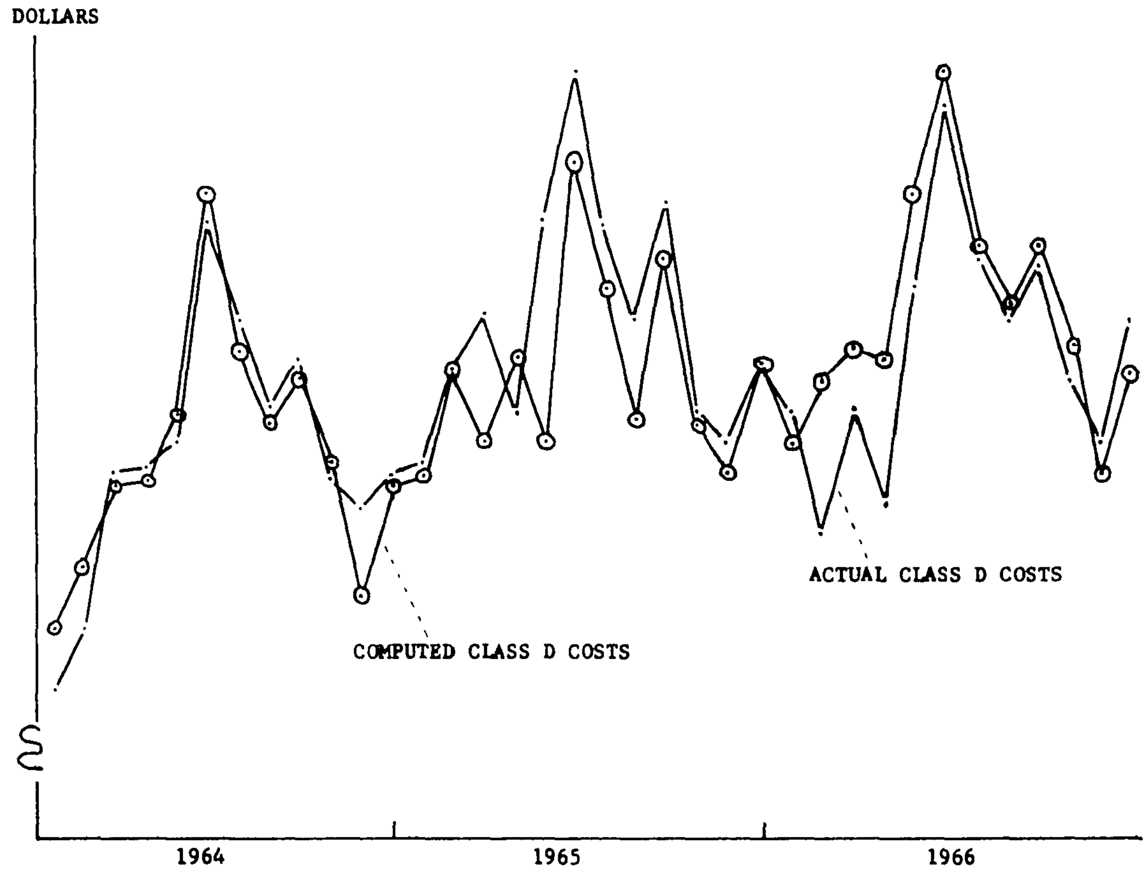

Chart 4.8. Actual Class D Costs and Computed Class D Costs of the Sales Department Costs, January 1964 - December 1966.

The Class E, F, and G Costs

Cost Class E is comprised of the executives' salaries and their travel expenses. In general, the salaries account is quite fixed in nature. In this account some variation may be attributable to transfers in and out of different positions. Fluctuations in this Cost Class are primarily attributable to variations in the travel expense items. On an annual basis, hardly any variation in the three twelve-month totals appears. In addition, no relationship seems to exist between these expenses and sales or other fluctuating variables; instead, the only really visible relationship is with the number of working days per month, and perhaps there is a slight seasonal variation. Because of the nature of this cost class and the relative immateriality of the travel expenses, it seems appropriate to treat these costs as programmed costs for which no estimating equation need be developed.

The two remaining Cost Classes, F and G, are programmed fixed costs in that only insignificant monthly variances occur. In annual terms, there are significant fluctuations, but they result from management decisions and not from some quantifiable external or internal variable factor. Therefore these costs may be reliably predicted in advance without estimating equations of the types developed for the first four cost classes.

Summary of the Sales Department Cost Analyses

In the preceding sections four of the functional areas within the Sales Department have been analyzed and estimating equations developed for each. It is appropriate at this point to discuss the differing nature and purposes of the four estimating equations derived. In this regard, it should be noted again that correlation may be used to determine the relationships between two or more series, and at the same time, it may assist in explaining the cause for the dependent series. With this in mind, observe that the estimating equation for the Class A Costs (Salesmen's salaries and related expenses) may logically explain the cause of these costs. On the other hand, perhaps the reverse is true. That is, sales would not exist if it were not for the expenditures for Class A costs. Therefore, one must admit that the relationship between Class A costs and sales is not merely one of cause and effect; rather there is a two-way interchange of influences between the variables. The estimating equation for Class A Costs cannot be considered a means of completely explaining that cost, but it may partially explain the causes and it certainly may be used for establishing parameters for cost control.

The same considerations are relevant to the Class D costs (Clerical and related expenses), but perhaps not to the same degree. The estimating equation for these costs, in this writer's opinion, does explain much of the cause for this cost. Yet, the two-way street also exists in this segment. Further it is quite probable that several relevant explanatory variables were left out of the estimating equation because appropriate information was not available.

In the cases of the Class B costs (Advertising expense) and Class C costs (Planning and Service expenses), the function of the estimating equations was to provide cost control parameters. Examining the nature of the independent variables employed should convince one of the validity of this statement. However, these estimating equations do reflect the cognizance of the two-way flow of influence; they also reflect the admission that such costs are primarily a result of management policies. Therefore, it is appropriate to conclude that the estimating equations developed for the Sales Department are primarily designed for cost control purposes and not for explanatory purposes.

The estimating equations which are to be used as the Sales Department cost model are summarized below.

<u>Class A Costs</u>:

$$X_{c1.234} = \$18,345 + \$362.19X_2 + \$.0105X_3 + \$1,300.13X_4 \qquad \text{(Eq. 11)}$$

$$S_{y1.234} = \$1,383.08$$

<u>Class B Costs</u>:

$$X_{c1.23^2 34} = \$5,743.76 + \$375.75X_2 + \$.000001521X_3^2 - \$.0260X_3 + \$2,488.20X_4 \qquad \text{(Eq. 14)}$$

$$S_{y1.23^2 34} = \$4,665.67$$

<u>Class C Costs</u>:

$$X_{c1.2^2 2345} = -\$1,904.96 + \$.5363X_2^2 - \$11.731X_2 + \$.00019765X_3 + \$257.0258X_4 + \$577.169X_5 \qquad \text{(Eq. 16)}$$

$$S_{y1.2^2 2345} = \$150.53$$

<u>Class D Costs</u>:

$$X_{c1.234^2 4} = -\$28,437 + \$18.23X_2 + \$.526X_3 - \$422.23X_4^2 + \$7,283.81X_4 \qquad \text{(Eq. 17)}$$

$$S_{y1.234^2 4} = \$382.45$$

CHAPTER V

DEVELOPMENT OF AN OVERHEAD COST MODEL FOR THE MANUFACTURING SERVICE DEPARTMENT - I

Standard materials and direct labor costs are widely used by manufacturing firms. Many of these same firms employ a standard overhead rate, but in name only is it a "standard." In reality, it may be found that a predetermined overhead rate is being used but its computation in no way involved the quasi-scientific and complex techniques utilized to develop the materials and direct labor standards. The multiplicity of functions and thus costs, necessitated by the normal production activities of contemporary manufacturing firms, makes the development of a true standard overhead rate a difficult, if not impossible task. This is especially true in the case of multiproduct manufacturing firms.

Budgeting personnel encounter numerous difficulties in forecasting overhead costs, and management is constantly facing dilemmas while attempting overhead cost control. To make matters worse, overhead costs are becoming an even larger proportion of the total costs of producing a given product. It is hoped that the analyses to be conducted in this and the two succeeding chapters will provide an insight into the possible applications of regression and correlation techniques for budget and cost control in the overhead cost sector.

This chapter will offer a general discussion of the accounts and procedures employed by the Grace Company in accounting for overhead costs. This will include a brief synopsis of the accounts and the nature of the charges made to each. The accounts will then be classified as representing fixed or variable costs. Those accounts classified as fixed costs will be reviewed and discussed, and if appropriate, subsequently excluded from further analysis. Steps will then be taken to sub-classify further those accounts regarded as variable costs in preparation for additional analysis and development of a specific estimating equation.

Chapters VI and VII will contain the in-depth analysis of the variable overhead cost accounts, including the development and testing of the respective estimating equations.

Reporting Procedures

The classification and reporting structure within an individual company's accounting system often dictates which functions are manufacturing services; not the nature of the functions themselves. If one were to conduct a regression analyses on another company, there might well be more service departments than are contained within this analysis of the Grace Company's factory accounts, or there could be fewer. Functions considered as services, for purposes of this dissertation, should presumably be prescribed by the Grace Company accounting system. However, as will be explained later, such is not the case.

Within the factory-costs sector of the Grace Company, as shown by the chart of accounts, there are seven production departments and four service departments. The four service departments employed by the Grace Company are:

1. Tools and Maintenance Department
2. Returned Goods Costs
3. Stockroom Costs
4. General Factory Costs

The company's procedure for accumulating costs relevant to each of these four sectors, though not unique, is both interesting and pertinent to this analysis. For example, the tool and maintenance costs consist of charges for wages associated with trucking parts and equipment about the factory, various supplies and small tools not chargeable to a particular work order, and supervisory salaries. No charges are entered to the account for machinery depreciation, property taxes, insurance, payroll taxes, or other related costs. The total costs accumulated in the account are allocated each month to the production departments on the basis of the total direct labor dollars.

An analysis of the Stockroom Costs account reveals that it contains only the wages of that department's employees. The Returned Goods Costs represents the wages and materials costs incurred to rework defective units rejected by the inspection personnel. These costs are in turn allocated directly to the responsible production department in the month incurred. The General Factory Costs account contains a large number of different and unrelated costs. The total of this account is also allocated to the various production departments.

In essence, it appears that for purposes of correlation analysis, it will not be practical to utilize the four-service-department framework prescribed by the chart of accounts. Rather, it will be necessary to analyze the Factory Overhead subsidiary ledger accounts with a view toward grouping related accounts into meaningful cost sections. This approach will result in the various overhead accounts' being treated in conformity with the manner in which the costs are actually accumulated prior to allocation.

The factory overhead costs control is segregated into two sections for reporting purposes: controllable costs, and non-controllable costs. Each of these sections contain variable, semi-variable, and fixed costs of one type or another. In addition, portions of the various costs contained in the controllable cost section are allocated either directly to the production departments or to the General Factroy Costs account, which is subsequently reapportioned to the respective production departments. These two cost classifications, used by the company for reporting purposes, are neither realistic nor relevant for purposes of correlation analysis. With this in mind, the following schedule of accounts was prepared in order to set out the accounts actually employed and to indicate the cost-flow patterns. Each account was assigned a number in order to expedite subsequent discussions of the individual accounts. Actual dollar figures are not shown because of the confidential nature of the information involved.

Accounts Employed

As reflected by Schedule A, there are 36 different accounts contained in the factory overhead ledger. It may not be necessary to develop an estimating equation for each of the 36 accounts. Many of the accounts are of a fixed nature and, therefore, are not amenable to correlation analysis. A number of the individual variable cost accounts are probably functions of

SCHEDULE A

FACTORY OVERHEAD COST ACCOUNTS

Account Title	Account Number	Account Total	Production Departments	Tool and Maintenance	General Factory
Direct Supervision	50	$ XXX	$ XX	$ X	$ X
Machinery Repair	51	XXX	XX	X	X
Supplies	52	XXX	XX	X	X
Trucking Expenses	53	XXX	XX	X	X
Inspection Expenses	54	XXX	XX		X
Rework Costs	55	XXX	XX		X
Time Study Allowances	56	XXX	XX		
Time Lost - Injuries	57	XXX	XX	X	X
Overtime Premium Paid	58	XXX	XX	X	X
Night-Shift Premium	59	XXX	XX	X	X
Machinery Rental	60	XXX	XX		X
Lights, Water, Gas	61	XXX			X
Indirect Supervision	62	XXX			X
Payroll Taxes	63	XXX			X
Time Study Salaries	64	XXX			X
Factory Engineering	65	XXX		X	X
Planning Salaries	66	XXX			X
Purchasing Salaries	67	XXX			X
Personnel Expenses	68	XXX			X
Stockroom Salaries	69	XXX			X
Safety and First-Aid	70	XXX			X
Employee Pensions	71	XXX			X
Group Insurance	72	XXX			X
Telephone and Telegraph	73	XXX			X
General Factory Travel	74	XXX			X
Vacation Payments	75	XXX			X
Depreciation - Cars	76	XXX			X
Repairs-Company Cars	77	XXX			X
Depreciation - Machinery	78	XXX			X
Building Repairs	79	XXX			X
Depreciation - Building	80	XXX			X
Insurance - Fire	81	XXX			X
Insurance - Liability	82	XXX			X
Plant Security Salaries	83	XXX			X
Property Taxes	84	XXX			X
Scrap Recovery (Deduct)	85	(XXX)			(XX)
Totals		$XXXXX	$XXX	$XX	$XXX
Departmental Allocations			X	(XX)	(XXX)
Totals			$XXX	-0-	-0-

the same or similar explanatory variables. Finally, several of the accounts are likely to be interrelated, because of mutual causation variables or because they manifest similar activities. In such instances it may be best to combine the related cost accounts.

With these considerations in mind, it is appropriate to identify the type of costs recorded in each account. Once this is determined, it may be possible to group the accounts logically and thereby reduce the number of estimating equations that it will be necessary to compute.

Direct Supervision--50: Contains charges for the salaries of the production department foremen and assistant foremen.

Machinery Repair--51: Comprised of charges for labor and parts used for repairs. Labor charges include time spent by the production workers in repairing their own machines and also time spent by personnel from the tool and maintenance department. Parts charges also include expenditures for dies which cost $200 or less.

Supplies--52: Contains charges for all types of indirect supplies utilized in the plant production departments, general departments, and the tool and maintenance department. Cost for such items are charged to this account. In addition, some very minor indirect materials are included in this account along with charges for administrative supplies of one sort or another.

Trucking--53: Contains labor charges for general cleaning, unloading rail cars, handling stores, etc. Charges to this account represent indirect labor throughout the factory and are taken from the labor summary.

In addition, this account contains labor charges for time spent by production workers, personnel of the tool and maintenance department, personnel of the stockroom, and other indirect labor personnel who are paid on an hourly wage basis and who on occasion perform any of the foregoing tasks.

Inspection—54: Contains charges for the salary of the department supervisor and the wages paid the factory inspectors. The supervisor's salary is charged into the general cost section while all wages paid to the factory inspectors are allocated to the individual production department in which they work.

Rework—55: Contains charges for direct labor required to correct previous production errors. When determinable, charges for this labor are allocated directly to the production department responsible for the faulty product; however, when this practice is not determinable or appropriate, the charges are placed in the general cost section.

Time Study Allowances—56: Contains charges representing allowances granted to incentive pay workers to bring their earnings up to the standard amount per hour where it has been determined that the standard time per unit was set too low. These allowances also include charges for delays caused by failure to have materials on hand, and occasionally, machine down time.

Time Lost Injuries—57: Contains charges for all indirect and direct labor time lost because of injuries. This account total is reallocated to the respective production departments or overhead departments, based on the time lost from this cause by the employees of the separate departments.

Overtime Premium Paid--58: Contains charges for all overtime wages paid to direct and indirect factory personnel. Charges for overtime paid to supervisory personnel are included in the regular supervision accounts. Although this account reflects the total overtime premium paid, each production department and the various overhead departments are allocated their appropriate share, based on the actual overtime incurred in the respective departments.

Night-Shift Premium--59: Contains charges for the night-shift premium paid to both direct and indirect laborers working on the night shift. This account is the only one of the 36 factory accounts that does not contain data for the period March, 1964, through December, 1966. Although a night shift had been operating for several years, the night-shift premium had not been reported separately prior to August of 1965.

Machinery Rental--60: Contains charges for machinery used in production and in the service segments of the factory on a rented basis. These charges are shown as direct charges to the various departments receiving benefits of the rented machinery and equipment.

Utilities Expense--61: Contains invoiced charges for electrical power, lights, gas, and water. Instead of departmentalizing this cost, the company charges it to the general factory overhead and subsequently reapportions it to the various production departments.

Indirect Supervision--62: Contains charges for one-half of the general managers's salary, one-half of the assistant general manager's salary, all of the factory manager's salary, and all of the production manager's salary. No other salaries or expenses are charged to this account.

Payroll Taxes--63: Contains charges for the employer's share of the F.I.C.A. Taxes, State Unemployment Insurance, and the Federal Unemployment Insurance. These taxes are the total for all direct, indirect, and supervisory personnel within the factory.

Time Study Salaries--64: Contains charges for the salaries of all time-study personnel and assorted clerical help within the department.

Factory Engineering Expenses--65: Contains all salaries paid to the factory engineering department personnel, including salaries of necessary clerical help, and all associated administrative supplies expense. Because one of the engineers is assigned to the tool and maintenance department, his salary is reapportioned as a part of that department's cost. The balance of the department's cost is included in the general factory costs.

Planning Department Expense--66: Contains charges for salaries of engineers and other planning department personnel, including secretarial and clerical assistants. Supplies used by the departments are also contained in the account.

Purchasing Department Expense--67: Contains charges for salaries of all purchasing agents, stenographic and clerical personnel, and all supplies.

Personnel Department Expense--68: Includes the salary of the personnel manager, the assistant personnel manager, stenographers, and clerical personnel. Also included are the supplies used within the department and the cost of advertisements for employee recruitment.

Stockroom Salaries--69: Contains charges for salaries and wages paid to all stockroom personnel, including supervisory and administrative clerks. No charges are made to this account for supplies of any type.

Safety and First Aid--70: Contains charges for first aid supplies located within the plant, ambulance, fees, etc.

Employee Pension Cost--71: Contains charges for annual pension costs incurred by the company in accordance with the employee union contract. Non-unionized personnel also have a retirement program requiring that deposits be made to a pension trust. These deposits are included in this expense account.

Group Insurance--72: Contains the costs of group insurance premiums paid by the company on all factory employees, including those of the shipping department.

Telephone and Telegraph--73: Contains charges for this service as invoiced by the company. The Grace Company maintains a record of the number of extensions both within the factory and in other offices that are utilized by factory personnel. In addition, a record is maintained of all long distance calls placed by factory personnel. A summary of all long distance calls is prepared each month and the associated cost is allocated between the various cost sectors.

General Factory Travel--74: Contains charges for all travel and away-from-home living expenses incurred by all factory personnel.

Vacation Payments--75: Contains charges for payments made to all direct, indirect, and supervisory personnel for vacation time. This item is recognized as an expense on an estimated basis for eleven of the twelve months. In December of each year, an adjustment is made to the account to correct the estimated expense deductions to actual.

Depreciation - Company Cars--76: Contains charges for depreciation taken annually on all company cars utilized by factory personnel. The sum-of-the-years digits method of depreciation is used for tax purposes, but the straight-line method is used for book purposes.

Repairs - Company Cars--77: Contains charges for all maintenance and repair expense related to the company cars. Operating expenses for these cars are reported in the general factory travel account.

Plant Security Salaries--84: Contains charges for salaries to all security and plant watchmen.

The remaining account titles appear adequately descriptive of the cost contained therein. However, it should be noted that depreciation on both the machinery and on the buildings is charged to general factory costs and subsequently allocated back to the production departments on the basis of standard direct labor dollars. The Grace Company viewed this procedure as a significant deficiency in its cost system. Therefore, on January 1, 1967, a procedure was adopted wherein depreciation of machinery was allocated to the department in which the machine was used, but building depreciation continues to be allocated as before.

Insurance premiums paid for fire and liability protection are carried into the general factory overhead account and subsequently reallocated on the basis of standard direct labor dollars. This same procedure is also used for property taxes and is consistent with the procedure employed to recognize building depreciation.

Fixed Costs

The fixed costs may be segregated into a committed cost and a programmed cost class. The committed cost class contains those costs of a sunk nature, such as machinery and building depreciation, and long-term costs associated with top level factory personnel. The programmed fixed cost class includes items of a shorter duration, in terms of time, but which are not directly related to volume of production.

Chart 5.1 contains three series. The bottom series is comprised of Machine Rental--60, and depreciation of, Cars--76, Machinery--78, and Buildings--80. Technically, the depreciation should and does fluctuate from one month to the next, but since it is treated a function of time by the company, and because accurate monthly data was not available, its presentation on an annualized basis seems justifiable. The machinery rental account was analyzed and found to be a reasonably fixed charge. Rentals are not paid on the basis of volume, but on the length of time used. The company rents only a few pieces of equipment, and these were rented for the entire period of analysis, except in 1965, when several additional pieces of machinery were rented and subsequently purchased on January 1, 1966. This partially explains the fluctuation in the charted series. The balance of the fluctuation is attributable to new machinery acquisitions which are being depreciated, using the straight-line method.

The middle series on Chart 5.1 is comprised of charges for: Insurance --81 and 82, Property Taxes--85, and Employee Pension Costs--71. Of these

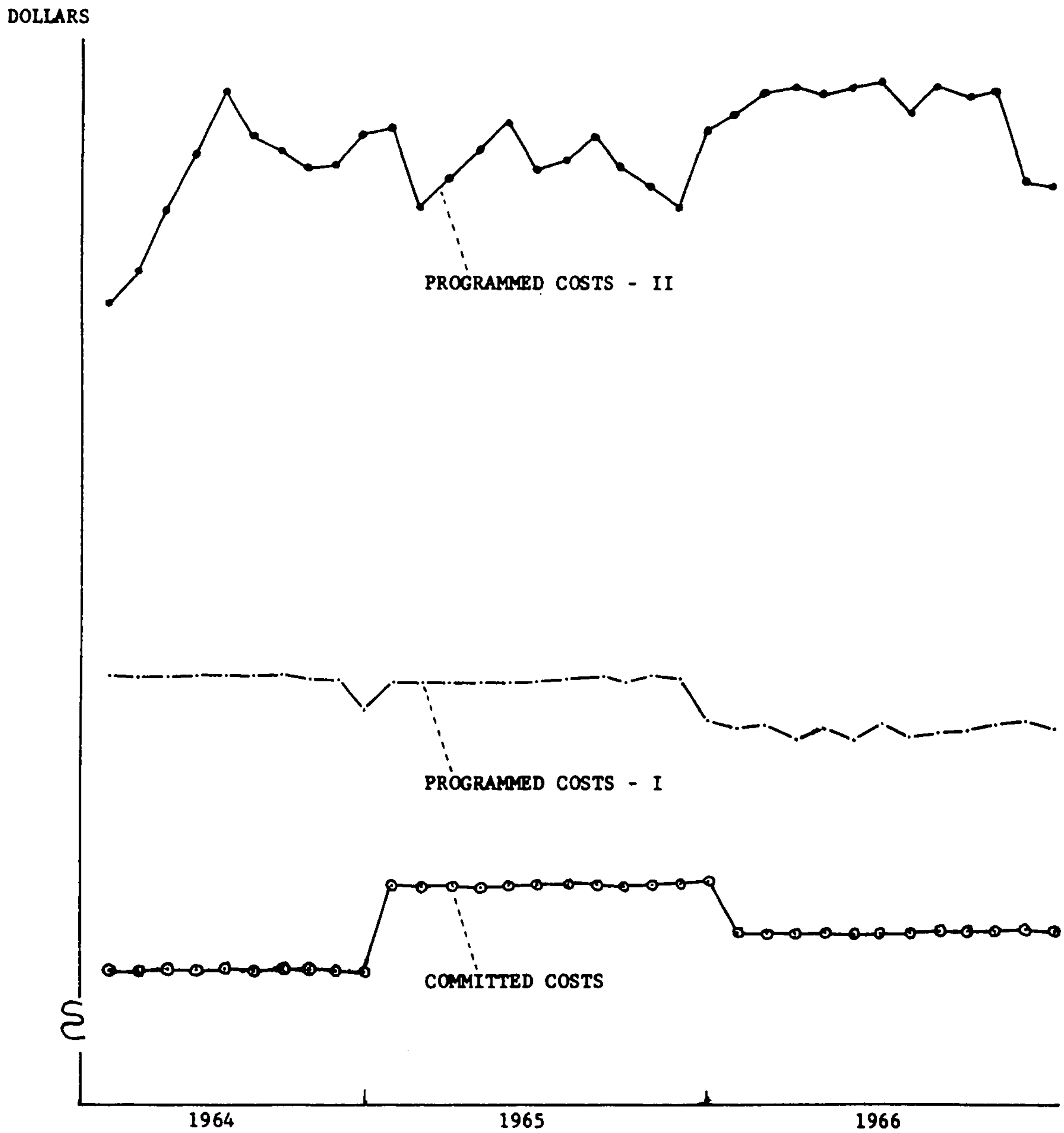

Chart 5.1. Programmed Costs - I, Programmed Costs - II, and Committed Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

three costs, management has little or no control over the property taxes, no direct control over the employee pension costs, and only limited control over the cost of insurance. It seems unique that the employee pension cost would not fluctuate to any great extent, but that is the case. The reason is that the company pays monthly premiums for all employees based on the average number of employees in each class. The fluctuations appearing in this series are attributable to changes in the average number of employees per class, a reduction in property taxes in 1966, and slight variations, on an annual basis, in the insurance costs. However, a perusal of this series readily reveals its stable nature and lack of sensitivity to internal influences.

The highest series is comprised of salaries, some supplies and insignificant sundry departmental expenses. Specifically, the accounts contained in the series are Indirect Supervision--62, Purchase Department Expense--67, Planning Department Expenses--66, Engineering Department Expenses--65, Time-Study Salaries--64, Personnel Department Expenses--68, and Plant Security Costs--84. An examination of the plotted series reveals a slightly upward trend and some variation, for which there is no clear seasonal pattern. Each of these accounts was subjected to analyses prior to classification as programmed costs. The analysis of the indirect supervision account revealed it to be absolutely stable over a one-year period. For the three-year period the only variation occurring resulted from annual increments in salaries. The expenses of the purchasing department demonstrated more variability, but analysis revealed that the variation was attributable to the timing of supplies charges and increases in employee salaries. The annual cost of supplies in this department changed less than one-tenth of one per cent over the three-year period. Annual stability, in the view of this writer, is quite remarkable.

The Engineering Department expenses varied considerably. Analysis and interviews revealed that supplies for both engineering and the time-study department were contained in this account. Further, personnel of the time-study department were responsible to the engineering department supervisor and worked concurrently in both departments. Since the time-study salaries also demonstrated variability, it was obvious that the payroll department had encountered difficulty in properly classifying salaries between these two departments. Therefore, for purposes of this analysis, the two departments were combined. This procedure eliminated much of the two departments' variability. The remaining variation may be attributed to salary increase, the difficulty of replacing trained personnel, charges of salaries to the sales department on occasion, and the timing of supplies charges. However, here again, the supplies expenses varied only slightly between years.

Expenses associated with the Personnel Department consist of salaries, supplies, and a very small amount of classified advertising and other sundry expense.

After a thorough analysis of the foregoing accounts, it was the view of this writer that each one unquestionably qualified as either a committed or programmed cost; that each could be quite accurately predicted without developing an estimating equation; and that each was highly impervious to changes in production volume within the range of the past three years of experience.

Immaterial, Stochastic, and Non-Operating Costs

Aside from the foregoing accounts which have been justifiably excluded from further analysis, there are additional accounts that should be excluded either on a stochastic basis or as trivia, or even on both grounds, and others that may be excluded because they contain non-operating costs. For example, it would seem that costs incurred for Time Lost from Injuries--57, Safety and First Aid--70, and Repairs to Company Cars--77 should be excluded on both grounds. The annual sum of these items is absolutely insignificant, relative to total overhead costs, and the timing of their occurrence is not predictable. Therefore, they may safely be excluded from the analysis, even though they are certain to occur over longer periods of time.

Triviality, or lack of significance, provides valid grounds for excluding other overhead accounts. In the case of the Grace Company, it appears only two other accounts fall within this range. Specifically, these accounts are the Telephone and Telegraph Expense--73, and the General Factory Travel Expense--74. Both accounts demonstrate some variation between months, and each appears to have a positive but stable trend. However, the combined monthly and annual expenditures are absolutely insignificant relative to the total overhead cost. In addition, it is doubtful that their respective variations could be correlated with any plant-wide activity. Therefore, it is felt that these two accounts should not be subjected to correlation analysis.

Non-operating costs are the third type of expense which should be exempted from regression and correlation analysis. Various costs incurred by the firm cannot be related to production nor can they be considered certain to occur in significant amounts during any one- or two-year period. However, these same costs are certain to occur over longer periods and in significant amounts. The account to be eliminated on this basis is Building Repairs--79. Costs of this type may be the result of time and normal wear, but the timing of incurrence is largely a function of management decision. Costs may also be incurred because of fire or other catastrophes over which there is no human control. Therefore, there is logical justification for not submitting the Building Repair Expense--79 to correlation analysis.

As has been previously stated, correlation is not an effective tool for analyzing non-repetitive processes or occurrences that are due to chance. Most firms will have a few accounts of this nature and to attempt correlation with some exogenous variable would probably be futile and impractical. Correlation is not a particularly useful tool for analyzing small, but stable expense items. However, the degree of "smallness" relies in each case on a value judgment and must be related to the magnitude of the independent variables being employed. It is not the intent of this writer to imply that small expense items should not be analyzed in any manner or considered in preparing a budget. Indeed, small expense items, expenses of a stochastic nature, and non-operating expenses must be considered in the budget. But it is better merely to compute a trend for such items or establish a budget limit based on value judgments within the company's policy framework.

Variable Cost Accounts

Accounts included in this section demonstrate considerable variation over each one-year period. Reasons for variation in the different factory overhead cost accounts are numerous; many of the relationships are quite complex, and in some cases, practically indeterminable. It does not appear plausible that any single explanatory variable could rationally be expected to explain a large proportion of the variation in any one cost or even of the composite sum of all the variable costs. On the other hand, it would seem that the selection of a large number of explanatory variables for the different costs would be impractical when related to efficiency in preparing the budget. Therefore, the first step in preparation for conducting the actual correlation analyses should be to segregate those overhead accounts containing significant variation into categories. Classification in the various categories would be a function of the relative sensitivity of a specific cost to variations in the production activities.

Category I—In the first category would be those costs directly associated with production of the product(s) in the sense that the related services are absolutely essential and consist primarily of indirect labor. Overhead accounts that might be placed in this category are: Trucking—53, Inspection—54, Rework—55, Time-Study Allowances—56, and Stockroom Expense—69. It would seem that these costs should be closely related to such explanatory series as direct labor hours and dollars of materials usage.

Category II—The second category would consist of those costs which logically should be sensitive to other internal influences as well as to either the direct labor hours or materials usage series. Overhead accounts that might be classified in this category are: Overtime Premium—58, Night-Shift Premium—59, Machinery Repair—51, Indirect Supplies—52, Utilities Expense—61, Scrap Recovery—86, and Direct Supervision—50. All of these accounts will probably require that at least two or more explanatory variables be employed before acceptable estimating equations can be developed.

Category III—The third category of overhead costs will exist because other costs were first incurred and only indirectly because production took place. This category would include such costs as: Employee Group Insurance—72, Vacation Pay—75, and Payroll Taxes—63. Logically, these fringe benefits should be related to production efforts, but because of their very nature more accurate estimates may often be developed by employing methods other than regression analysis.

Summary

The Grace Company's formal reporting structure for overhead cost accounts was not satisfactory for purposes of regression and correlation analyses. The four-account structure employed by the company was broken into 36 different overhead cost accounts. Several of these 36 accounts contain charges from a single source such as insurance or property taxes, and others, such as machinery repairs, contain charges from numerous sources.

Each of the 36 accounts was then reviewed as to the source and behavior of its charges. Based on this review, each account was placed in one of three broad classes. The first classification was fixed costs, which was further sub-divided into committed or programmed costs. A second classification was established for those accounts containing immaterial balances and costs of a stochastic, or non-operating nature. Finally, a number of the accounts were placed in a third broad classification as variable costs. The variable cost class was subsequently sub-divided into three categories and several accounts were placed in each.

Those accounts placed in the fixed cost and the second classification will be excluded from further analysis. Of those accounts classified as variable costs, the Category I accounts will be subjected to regression and correlation analysis in Chapter VI. The Category II and III accounts will be similarily analyzed in Chapter VII.

When the analysis of the manufacturing service departments of the Grace Company was first commenced, it seemed as though a complex maze of costs was being confronted. It was immediately obvious that some sort of organization would have to be introduced into the account classifications if this task was ever to have an end. Therefore, the foregoing classification scheme that reflects both the natural characteristics of the various costs and the capabilities of the statistical tools to be employed was devised. If the classification system developed is valid, then the objective of this chapter has been achieved.

CHAPTER VI

DEVELOPMENT OF AN OVERHEAD COST MODEL FOR THE MANUFACTURING SERVICE DEPARTMENT - II

Determination of the Category I Estimating Equations

Those accounts contained in Category I will be analyzed in this chapter. Starting with these accounts should develop information that may be useful in the analyses of the Category II and III accounts. The general methodological approach will be first to determine and review several data series as potential explanatory variable candidates for the Category I accounts. The individual cost accounts will then be reviewed for measurement errors and for their cost sources and behavioral characteristics. Each account will be treated in full before starting analysis on a succeeding account. Finally, a summary of the analysis will be presented, but this will be deferred until the end of Chapter VII when all three variable cost categories may be viewed as a whole.

Explanatory Variable Candidates

Logically, there should be some direct relationship between such costs as Trucking--53 or Inspection--54, and at least one of the explanatory series of direct labor hours, standard direct labor dollars, direct materials usage, or units produced. These explanatory series are shown on

Charts 6.1, 6.2, 6.3, 6.4, and 6.5. Each of these explanatory series must be analyzed and evaluated as likely candidates to be employed with the dependent variables.

Perusal of the total direct labor hours series reveals considerable seasonal and monthly variation. On the whole, the trend appears slightly positive, although June of 1965 contains the peak number of hours for the entire 34-month period. The base month in each year is August, a state which stems from the company's policy of stopping production for a two-week vacation each year. Much of the apparent seasonal variation in the series is attributable to the division of the year into four thirteen-week quarters, consisting of two four-week months and a third five-week month. The remaining variation is attributable to seasonal influences and, of course, to other elements that may or may not be determinable and that are likely to be insignificant. Experience gained from analyses of the two previous cost groups indicates that it probably will not be necessary to de-seasonalize the series, but that in some instances a working-day index may be appropriate.

Upon perusal of the standard direct labor dollars series and the dollars of material usage series, it becomes obvious that the same statements may be made regarding these two series as were made about the direct labor hour series. However, it is interesting to note that the trend is slightly more positive in the latter two series. Note that the standard direct labor dollars series peaks in 1966, even though direct labor hours actually decline. At first glance this appears to be a major discrepancy in the data but further inquiry provided a reasonable explanation. That

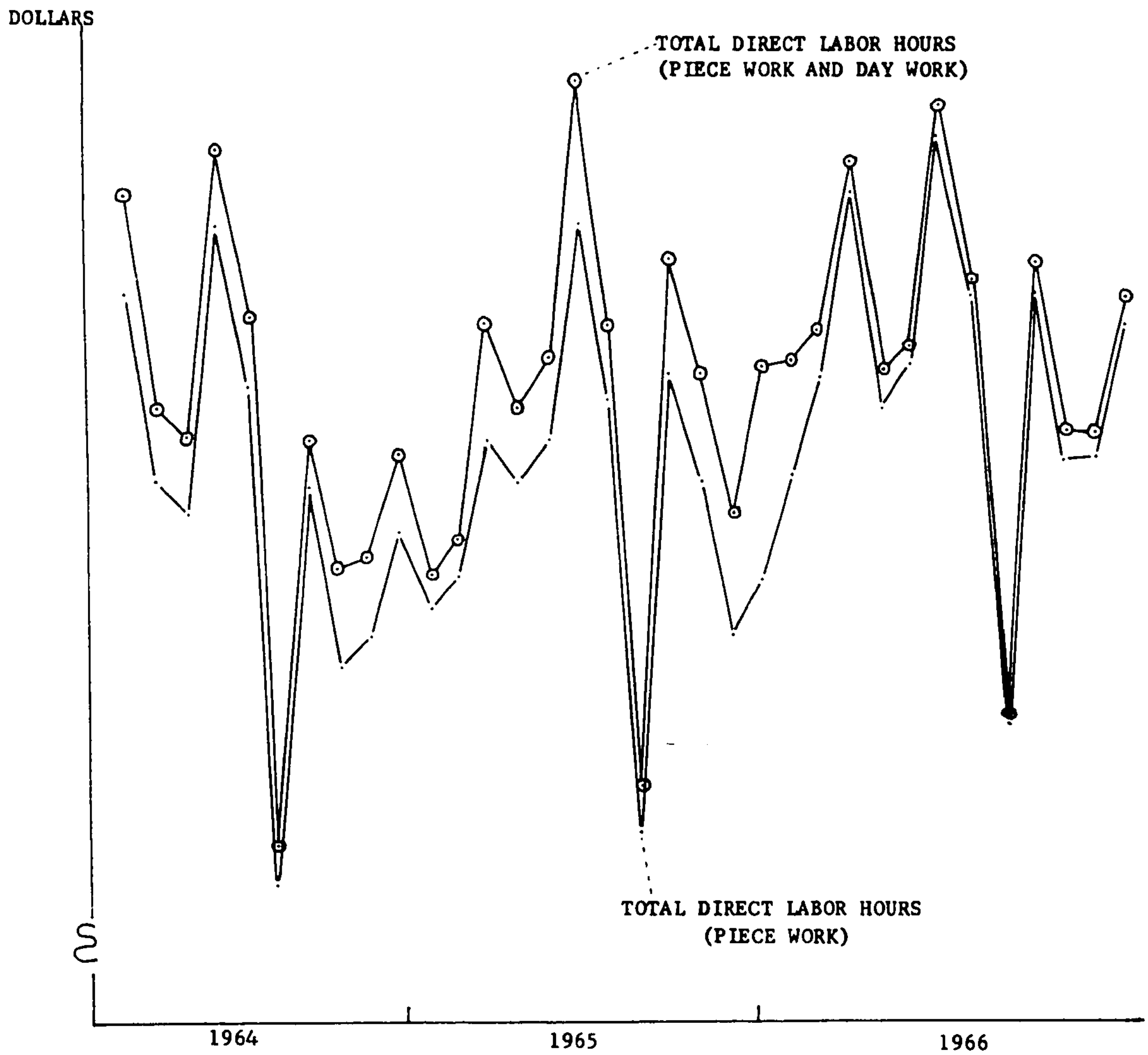

Chart 6.1. Total Direct Labor Hours (Piece Work) and Total Direct Labor Hours (Piece Work and Day Work) of the Manufacturing Department, March 1964 - December 1966.

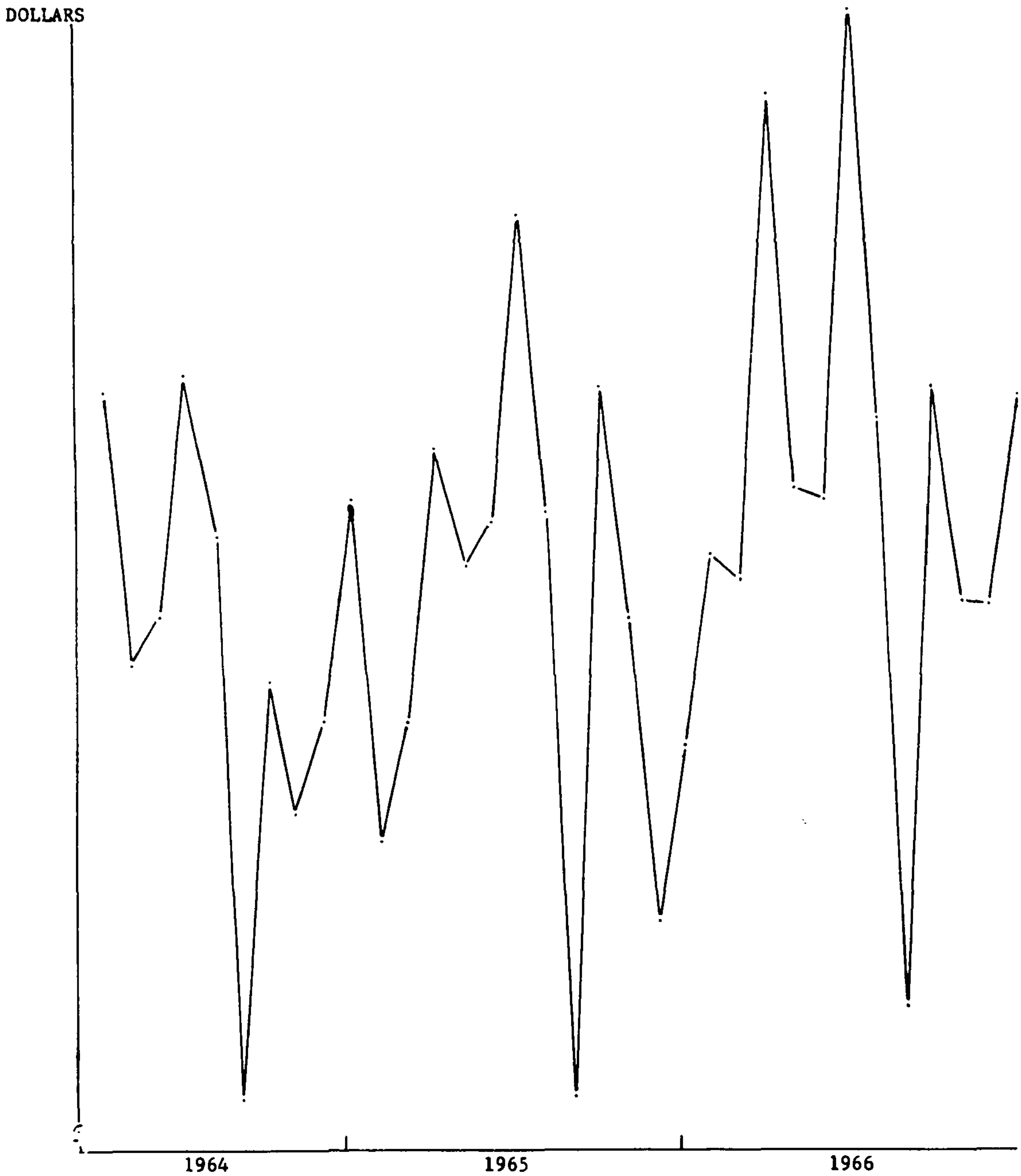

Chart 6.2. Standard Direct Labor Dollars of the Manufacturing Department, March 1964 - December 1966.

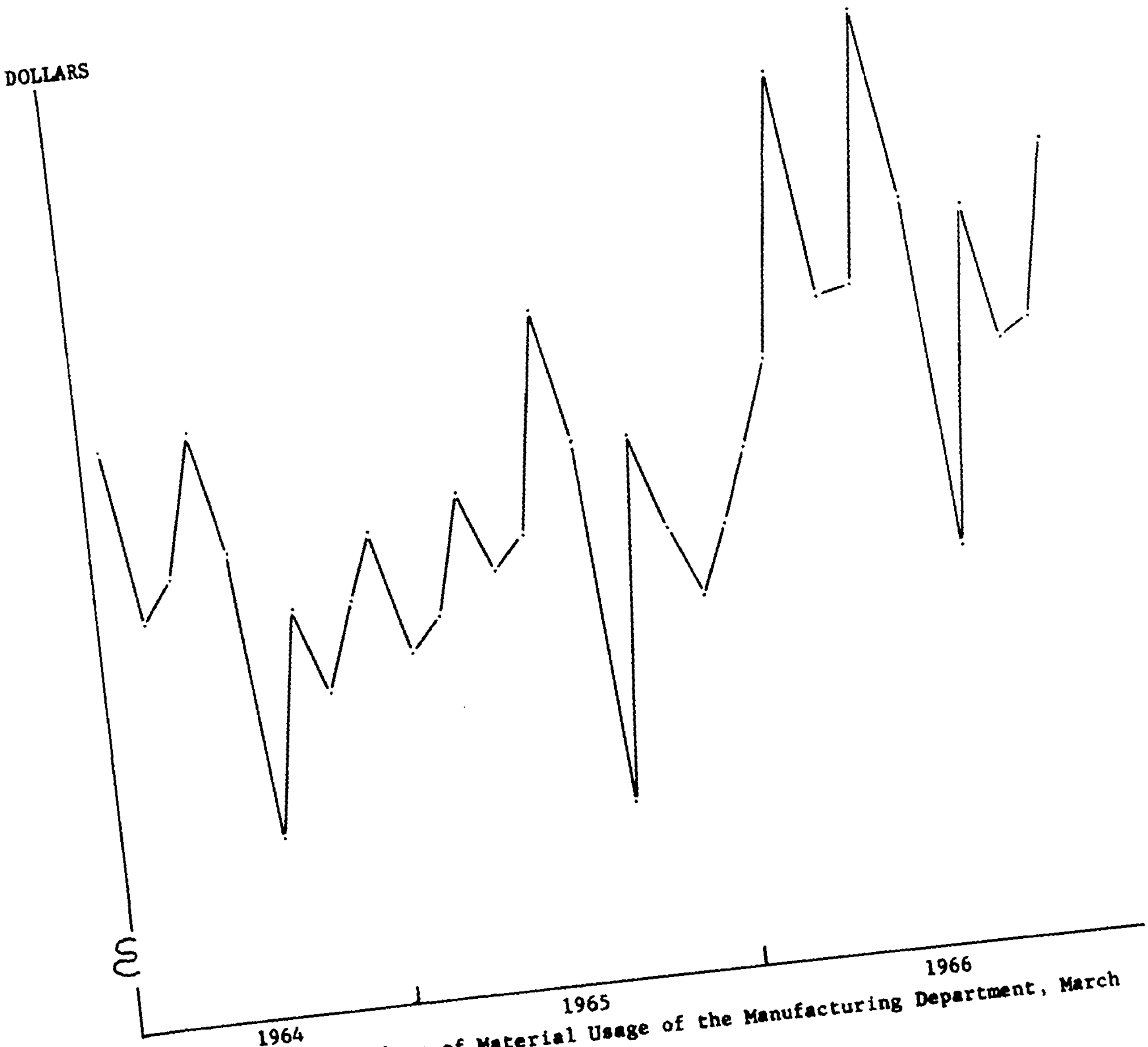

Chart 6.3. Dollars of Material Usage of the Manufacturing Department, March 1964 - December 1966.

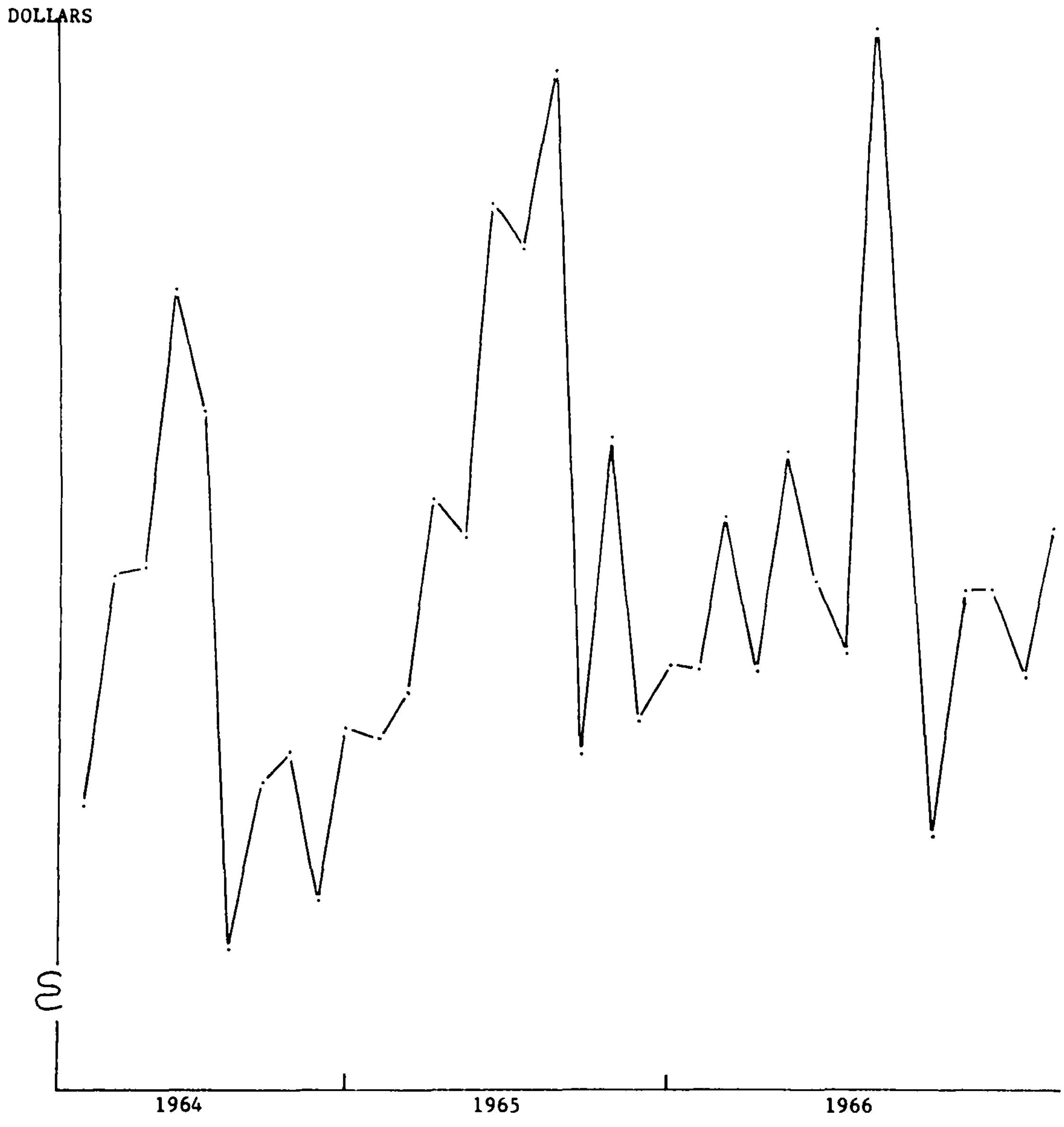

Chart 6.4. Total Units of Product A of the Manufacturing Department March 1964 - December 1966.

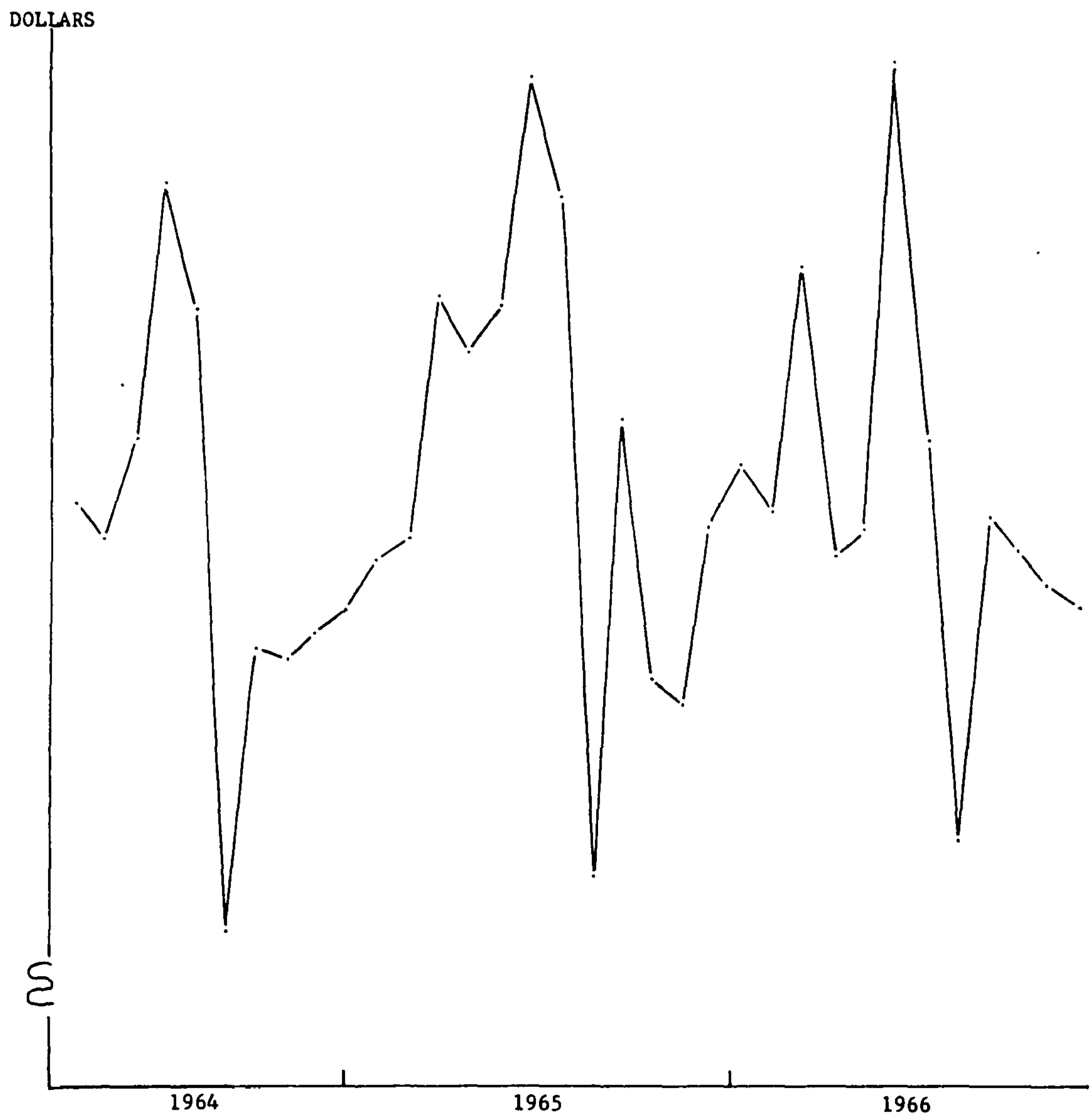

Chart 6.5. Total Units Produced of the Manufacturing Department, March 1964 - December 1966.

is, early in 1966, additional and more efficient equipment that reduced the labor requirements per unit was installed. This resulted in reducing the total direct labor required while increasing the quantity of materials usage. In addition, commencing in September, 1965, an across-the-board increase in the hourly rate of pay was incorporated into the labor standards. A third element also distorting the relationships between the three series is that the product design changes instituted in 1966 called for a larger and more complex unit. A fourth distorting factor was that the sales mix of Product B models shifted during 1966 so that a greater proportion of the sales consisted of the larger models than in previous years. In summary, distortion in the relationships between the three series is primarily attributable to increased automation, increased hourly wage rates, increased size and complexity of the products, and a shift in the product sales mix.

What are the implications of these changes for correlation analysis of the overhead costs? It is hoped they will be minimal. If the proper explanatory variable(s) are selected for each overhead cost, and if the overhead costs were sensitive to the factors of production, the overhead cost relationships should have remained stable—despite the alterations in the combinations of the two primary factors of production.

Estimating Equation—Trucking Expenses—53

Material handling is the primary function of the personnel whose wages are charged to this account. However, these employees perform other functions, some of which were previously discussed. In addition to the changes for regular workers employed in the department, there are numerous occasions when the direct labor workers will do clean-up work or material handling. The cost of this time is also charged to trucking. Chart 6.6 contains the trucking series. In general the fluctuations in this series follow the same pattern as both the labor hours and the materials-usage series. Logically, the trucking expenses should be most closely related to the materials-usage series, which appears to be valid when the series are compared. Because of the likelihood of multicollinearity between the labor-hours series and the materials-usage series, it is necessary that only one of the two be used in this particular instance. Therefore, because of both the logical and graphical relationships demonstrated, the first computational effort will be made using the following:

X_1 = Trucking Expenses--53

X_2 = Time

X_3 = Dollars of materials usage

and a linear equation of the form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3 \qquad \text{(Eq. 18)}$$

The variable X_2 for time is used as before, to eliminate the effects of time on the relationship between the X_1 and X_3. In this case, the data covers the 34-month period from March, 1964, through December, 1966. The data time period for each observation is one month.

For equation (18), $R^2_{1.23}$ = .7056 and the $S_{y1.23}$ = \$1,561.21. $R^2_{1.23}$ is statistically significant at the .05 level. The value of $1.96S_{y1.23}$ = \$3,059.97 is quite satisfactory when related to the individual monthly observations. Although these results are good, it must be kept in mind that

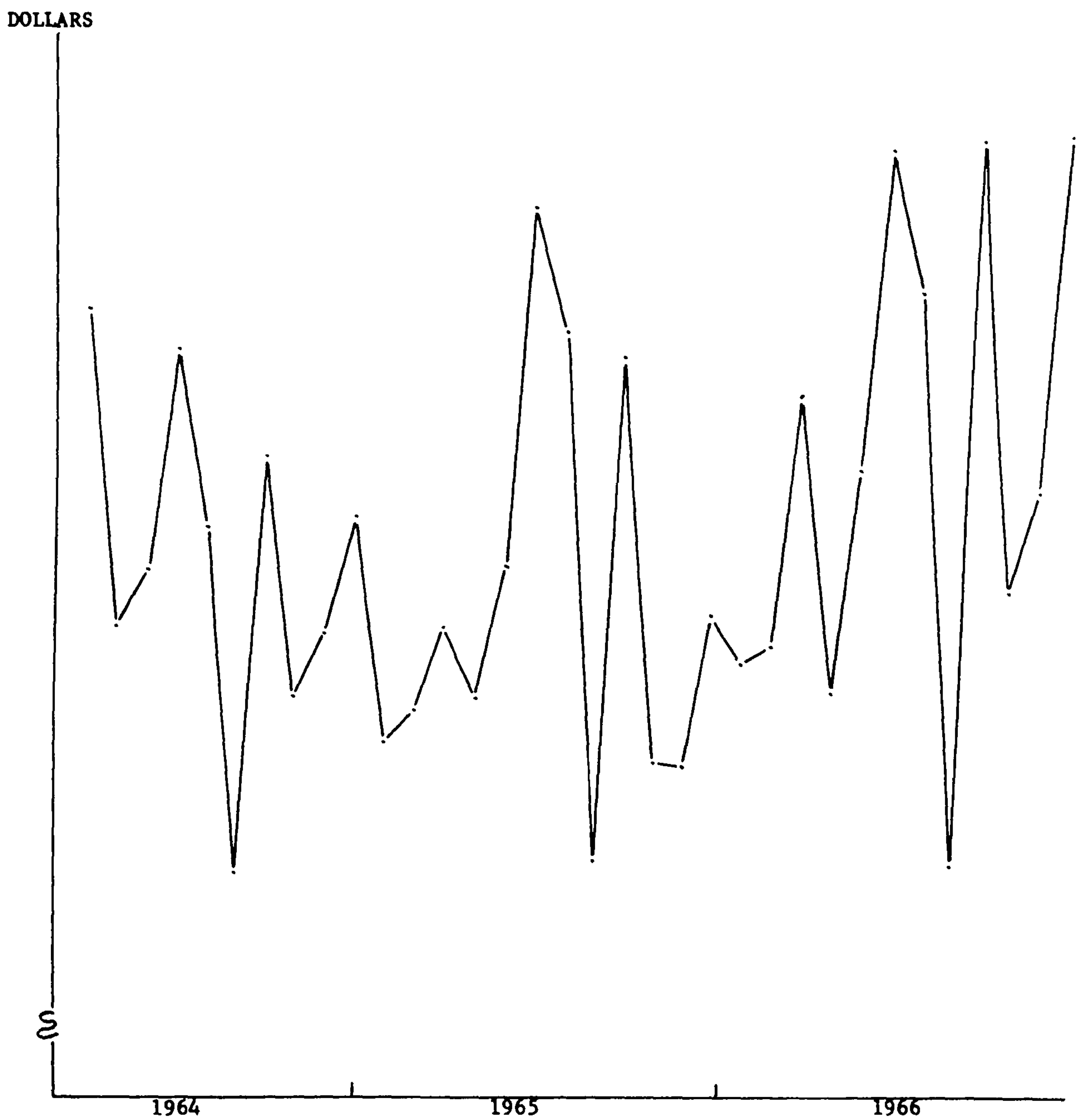

Chart 6.6. Trucking Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

a simple linear equation form was employed. The next step is to refine the foregoing results by determining the most efficient equation form, but continuing to employ the X_2 and X_3 variables.

Charts 6.3 and 6.6, containing the two variables, should now be reexamined. It appears that the independent variable (X_3) has a reasonably linear trend, whereas the trend of the dependent variable (X_1) appears curvilinear. Further, it appears there may be a slightly curvilinear relationship between the two variables themselves, after the effects of time have been eliminated. Therefore, the original equation form of $X_{1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$ will be altered in the following computations by adding or deleting various terms. In the first iteration an X_2^2 term was added. This provided a $R^2_{1.2^2 23} = .7133$ representing a .77 per cent increase that is not statistically significant at the .05 probability level. The next iteration employed an X_2^3 and an X_2^2 term. Use of both these terms changed the shape of the curve considerably, but it provided an $R^2_{1.2^3 2^2 23} = .7498$. This represented a 4.42 per cent increase over the $R^2_{1.23}$ value originally computed, which is not a statistically significant increase at the .05 probability level. Other iterations were made using the X_2 term with larger exponents, but none of these provided a better fit. Thus it appears that the use of the three terms X_2^3, X_2^2, and X_2, cannot provide a satisfactorily efficient equation. But this behavioral pattern does indicate that some form of curvilinear X_2 term may be appropriate.

To verify the relationship between X_1 and X_3 several additional iterations were performed employing different X_3 and X_2 terms. None of the iterations provided a better fitting estimating equation. Therefore, it must be concluded that the true relationship between X_1 and X_3 is approximately linear.

Although the coefficient of multiple determination is reasonably good for the type of data being analyzed, it is not so high as might be desired. Because of the composite nature of the expense items contained in the trucking account, it is probable that the inclusion of a working-day index would significantly improve the efficiency of equation (18). An X_4 term, representing a working-day index, was added to equation (18) as the next step in the analysis. This new equation form provided a $R^2_{1.234} = .7424$ or only a .0368 increase. The $S_{y1.234} = \$1,484.55$, which is a \$76.66 decrease from that of equation (18). Although the changes occurring in the statistical measures as the result of adding an X_4 term to equation (18) were not impressive, they were in the right direction.

Additional iterations were made, using the same X_2, X_3, and X_4 variables, but changing the equation form, to test the possibility of multicollinearity between X_3 and X_4. But, the behavior of the variables in the completed iterations indicated that little multicollinearity exists. Of these iterations, the best fitting estimating equation is:

$$X_{c1.2^{3}2^{2}34} = +\$552.67 + \underset{(.1247)}{\$.32542}X_2^3 - \underset{(4.254)}{\$11.6965}X_2^2 + \underset{(.001205)}{\$.0049641}X_3 + \underset{(280.84)}{\$585.289}X_4 \qquad \text{(Eq. 19)}$$

Associated with equation (19) are a $R^2_{1.2^{3}2^{2}34} = .7807$ and a $S_{y1.2^{3}2^{2}34} = \$1,393.14$. The coefficient of multiple determination of .7807 is significantly greater than zero at the .05 level. The standard error of the estimate of \$1,393.14 is quite satisfactory in view of the criterion established in Chapter III. Each of the coefficients of partial determination, $r^2_{12^3.2^234}$, $r^2_{12^2.2^334}$, $r^2_{13.2^32^24}$, and $r^2_{14.2^323}$ were subjected to an F test with $n_1 = 1$ and $n_2 = 29$. Each was individually significant at the .05 level. The regression coefficients and their respective standard errors were subjected to a t-test and all were significantly greater than zero at the .05 level.

Estimating equation (19) appears to be statistically valid and both the $R^2_{1.2^32^234}$ and $S_{y1.2^32^234}$ have quite acceptable values. The form of equation (19) is of interest because the time terms are not represented in a linear form. Inclusion of X_2^3 and X_2^2 without including an X_2 term is a departure from prior procedures, but it is not a major departure. The reason for not including the X_2 term was that it simply was not significant. If X_2 had represented an explanatory variable other than time, there may have been grounds for retaining the linear term in the estimating equation. But in this case there seemed to be no logical need for the X_2 term; thus it was deleted.

If practicable, it is still desirable to obtain a higher R^2 value for the dependent variable Trucking Expenses—53. To this end a new explanatory series consisting of a summation of the two-unit output series was developed. In addition, it was decided that the direct-labor-hours series could also be employed. The problem involved in using either of these two series was multicollinearity between the independent variables. To minimize bias from this source, the following equation form was employed:

$$X_{c1.2^32^223(45)} = a_{1.2^32^223(45)} + \ldots b_{1(45).2^32^223}X_4X_5.$$

The X_5 term was used to designate the total output series in one group of iterations, and the direct-labor-hours series in a second group of iterations. These procedures and the resulting runs did not provide any significantly larger R^2 values than those already obtained. Further, there was no material change in the X_2 coefficient values, and the standard error of the estimate was not decreased. Therefore, the conclusion must be accepted that within the present information constraints, equation (19) affords the best fit on both statistical and logical grounds.

Before equation (19) is accepted as a model of Trucking Expenses, it is necessary to plot a computed curve against the actual curve and review the variances. The X_{c1} and X_1 curves are plotted on Chart 6.7. Of the 34 $X_1 - X_{c1}$ differences, 18 are above the X_1 curve and 16 are below. During the most recent twelve-month period, 7 of the differences are above the X_1 curve and 5 are below. Assuming normality, this variation of the $X_1 - X_{c1}$ differences above and below the curve is satisfactory. The magnitude of the variations demonstrated a better than normal distribution as 25 of the differences, or 73.5 per cent, were within ± $1.00S_{y1.23^22^34}$ and only one difference occurred outside of the ± 1.96 standard error of the estimate range.

The individual months which appear to have the larger differences are March, July, August, and September. It is, of course, quite difficult to offer a rational explanation of why differences exist during these certain months. If the explanation were quantifiable, then it could be incorporated into the estimating equation to eliminate the differences. Unfortunately such is not the case in most instances. For example, the differences

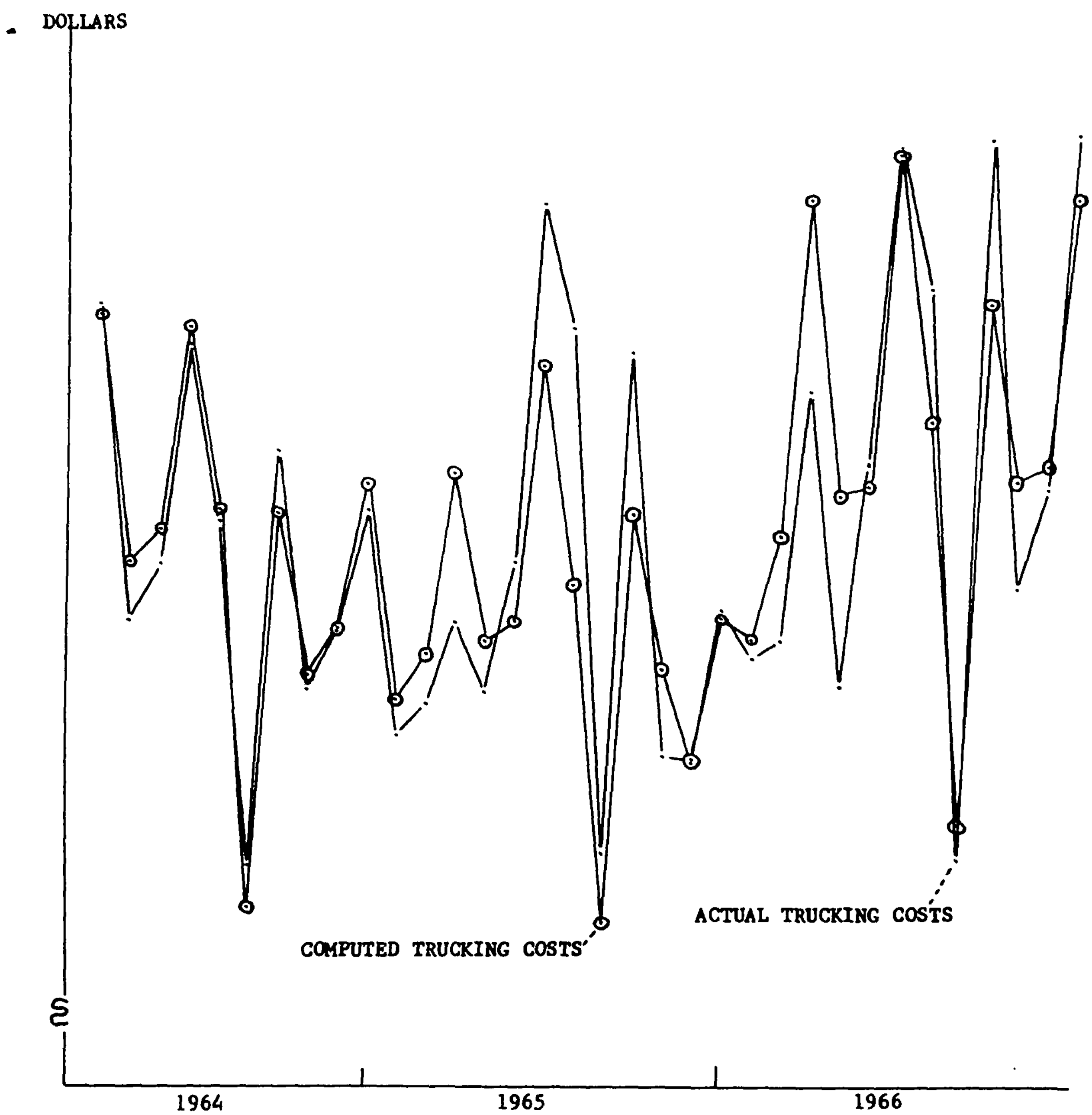

Chart 6.7. Actual Trucking Costs and Computed Trucking Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

associated with July, August, and September may be partially attributed to the lag between the time when materials are brought to the production line and the time when the materials are actually used in production. In addition, the mix of the raw materials issued may be an element. Finally, there may be a group seasonal psychological factor that affects efficiency. No doubt, present are many other factors that have a bearing on the relationships contained in the estimating equation. But in view of the existing information constraints, it appears equation (19) does provide a reasonably good fit. Therefore, it must be accepted as the model for Trucking Expenses —53.

Estimating Equation – Inspection Expenses--54

Inspection Expenses are incurred for the purpose of maintaining prescribed product quality levels. Wages are the only charges contained in the account. The series is plotted on Chart 6.8. Charges to this account are incurred in each production department, not just at the point of final output. Because of increased automation, employee education, more intense direct supervision, and improved inspection techniques, the company has been successful in gradually reducing this expense, even though product output has increased. This negative trend is apparent upon examination of Chart 6.8. It was felt that the direct-labor-hours series should be the most closely related explanatory variable series.

The initial computations employed the following variables:

X_1 = Inspection Expenses

X_2 = Time

X_3 = Total direct labor hours

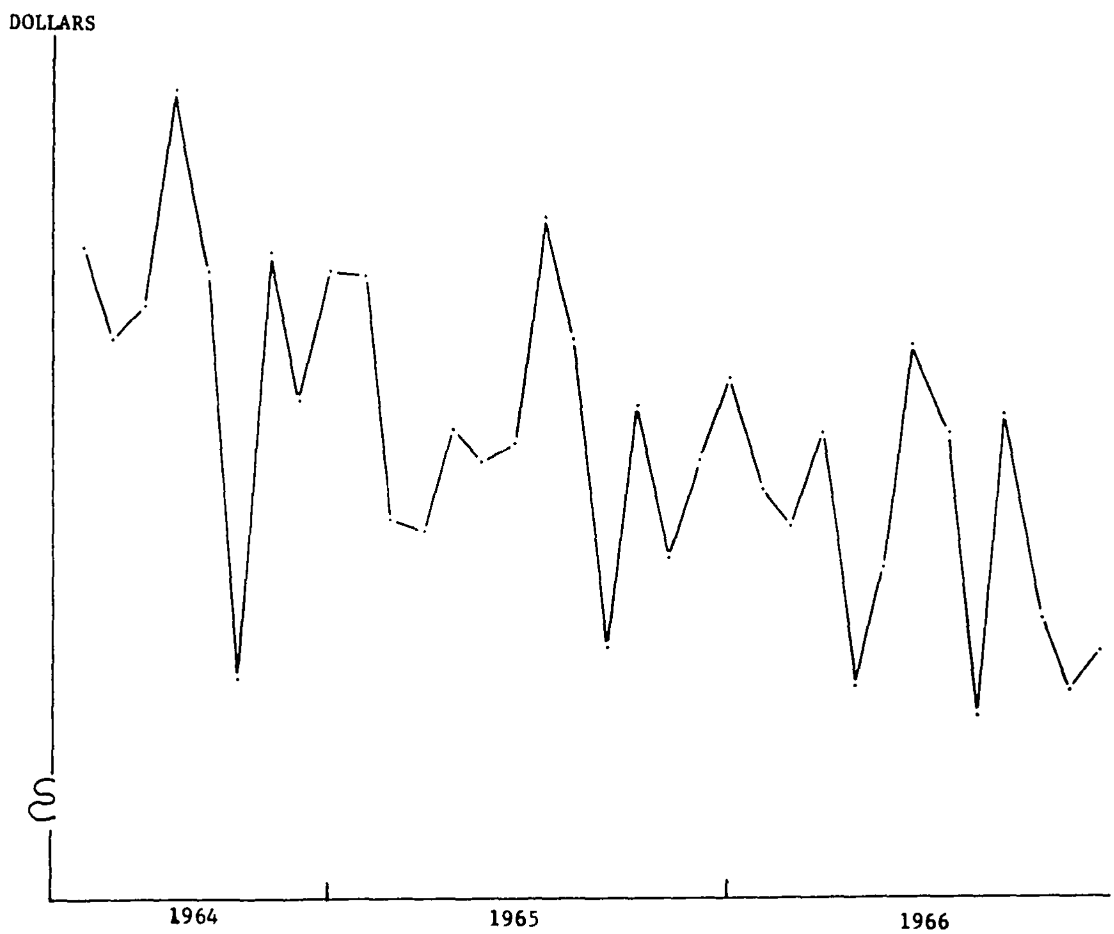

Chart 6.8. Inspection Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

in a linear equation of the form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3.$$

The first iteration provided a $R^2_{1.23} = .7705$, which is definitely significant at the .05 level, and a $S_{y1.23} = \$736.37$. Statistically, these results are satisfactory; however, it is still necessary to determine the most efficient equation form employing only the foregoing variables. A number of additional iterations were made, using additional X_2 and X_3 terms with various combinations of exponents. The procedures used in testing for significance and efficiency were identical with those used in previous account analyses. Of the equation forms used, the following equation appeared the most efficient:

$$X_{c1.23^3 3^2 3} = -\$4{,}179.58 - \underset{(12.56)^2}{\$103.49X_2} + \underset{(.0000000000533)^3}{\$.0000000001133X_3^3} - \underset{(.00000729)^3}{\$.00001544X_3^2}$$

$$+ \underset{(.314)^3}{\$.7363X_3}. \qquad \text{(Eq. 20)}$$

For equation (20) the $R^2_{1.23^3 3^2 3} = .8013$, which provides an F value of 29.24, which is definitely significant. The individual $r^2_{13^3.23^2 3}$, $r^2_{13^2.23^3 3}$, and $r^2_{13.23^3 3^2}$ were significant at the .05 probability level. The $S_{y1.23^3 3^2 3} =$ \$708.27 and is satisfactorily low, relative to the magnitude of the individual monthly observations. The standard errors for each X_2, X_3^3, X_3^2, and X_3 coefficient were acceptable as the related regression coefficients were significant at the .05 level. Therefore, based on the foregoing statistical measures, equation (20) appears to be an efficient estimating equation for inspection expenses.

To preclude the possibility of a relationship existing due to time alone, several iterations were carried out using an X_4 term representing a working-day index. This was done only for the purposes of maintaining consistency in conducting the various account analyses, not on the grounds of logic. Indeed, the nature and functions of the inspection personnel are such that a working-day index should not logically be appropriate. Examination of the results of the several iterations using the working-day index revealed that the X_4 term did not explain a statistically significant amount of variation in X_1. In addition, the inclusion of the X_4 term did not alter the previous relationships established.

In view of the foregoing, it appears that equation (20) should be accepted as the model of the inspection expenses. The next step is to compute the $X_{c1.23^3 3^2 3}$ values and plot these against the X_1 values. This information is shown on Chart 6.9. Of the 34 $X_1 - X_{c1}$ differences, 18 are above the X_1 curve and 16 were below. For the most recent twelve-month period, 6 of the $X_1 - X_{c1}$ differences were above the curve and 6 were below. Assuming normality for the three-year period and for the most current 12 months, the dispersion of the $X_1 - X_{c1}$ differences conform to a normal distribution. However, there is some apparent covariance since the variations seem to have a series tendency. On the other hand, there is no definite pattern in the occurrence of the variations. A re-examination of the data provided no tangible clues for the type of variation demonstrated. In terms of magnitude, again assuming normality, the variations are quite acceptable. That is, 67.6 per cent of the differences occur within a range of $\pm 1.00 S_{y1.23^3 3^2 3}$ and none of the differences is larger than ± 1.96 $S_{y1.23^3 3^2 3}$. Therefore, equation (20) will be accepted as a model for the Inspection Expenses—54.

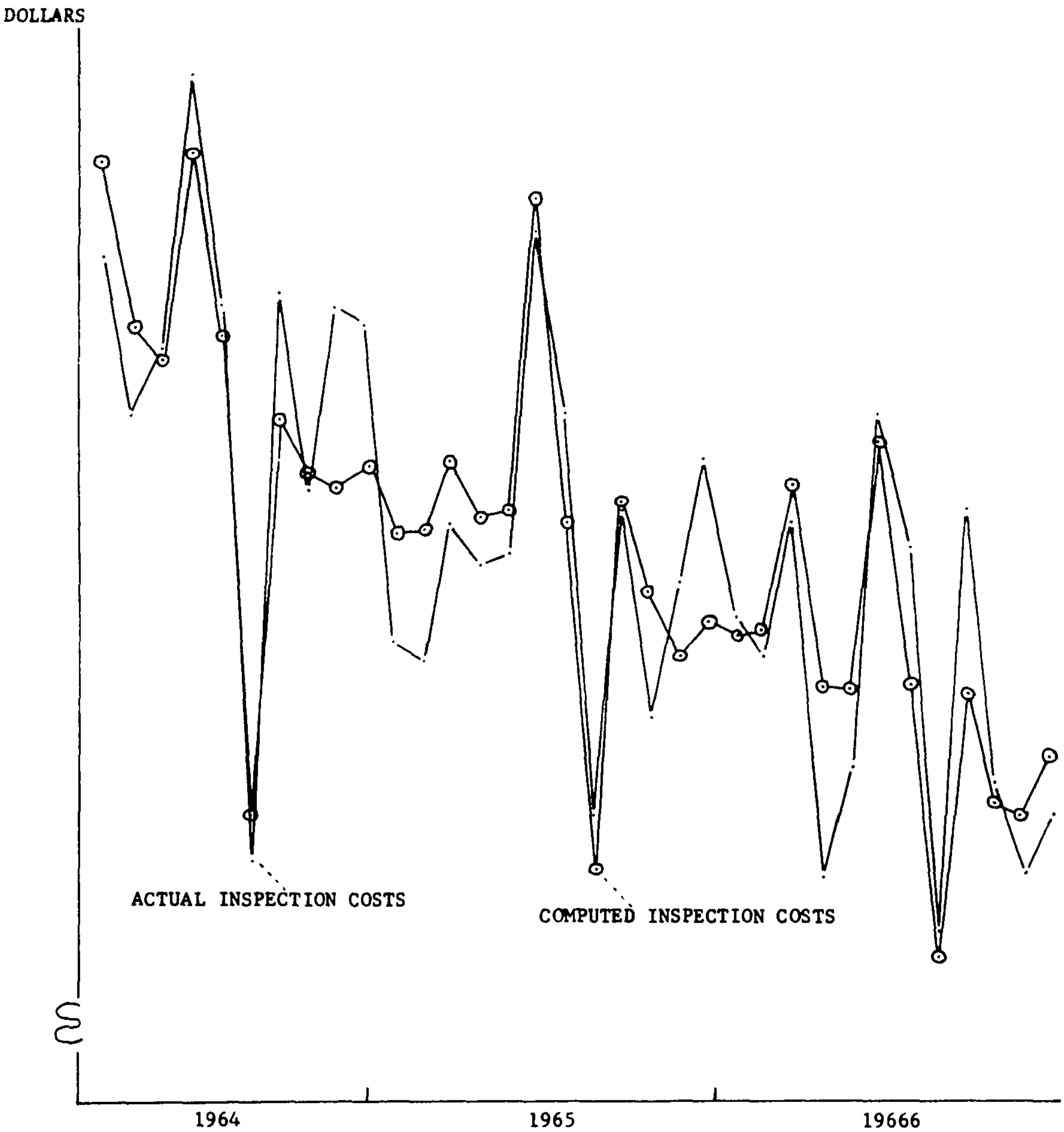

Chart 6.9. Actual Inspection Costs and Computed Inspection Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

Estimating Equation - Time-Study Allowances--56

Charges to the time-study allowances account represent the costs of direct labor associated with work performed where it has been determined the piece-work standard was set too high. Charges are also made to this account for various types of delays encountered by the direct labor workers. Monthly charges to the account appear on Chart 6.10. Perusal of the series indicates a close relationship exists with the direct-labor-hours piece-work series on Chart 6.1. Based on the nature of the cost being analyzed, this apparent relationship is definitely logical.

The first computational effort employed the following variables:

X_1 = Time-Study Allowances--56

X_2 = Time

X_3 = Direct labor hours - piece work

The following linear equation form was employed:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3.$$

This equation form provided an $R^2_{1.23}$ = .8142, which is significantly greater than zero. The standard error of the estimate ($S_{y1.23}$) was \$1,025.53, which is also quite satisfactory. Given these excellent initial results, the next step is to determine the most efficient equation form. A number of iterative computations were made employing the X_2 term with different exponents. The equation which, on balance, appeared the most reliable was:

$$X_{c1.2^223} = -\$1,477.81 - \underset{(1.7305)^2}{\$6.694X_2^2} + \underset{(61.803)^2}{\$210.128X_2} + \underset{(.00434)^3}{\$.01978X_3}. \quad \text{(Eq. 21)}$$

The $R^2_{1.2^223}$ = .8761, which is definitely significant and the $S_{y1.2^223}$ = \$861.53, which is also quite satisfactory.

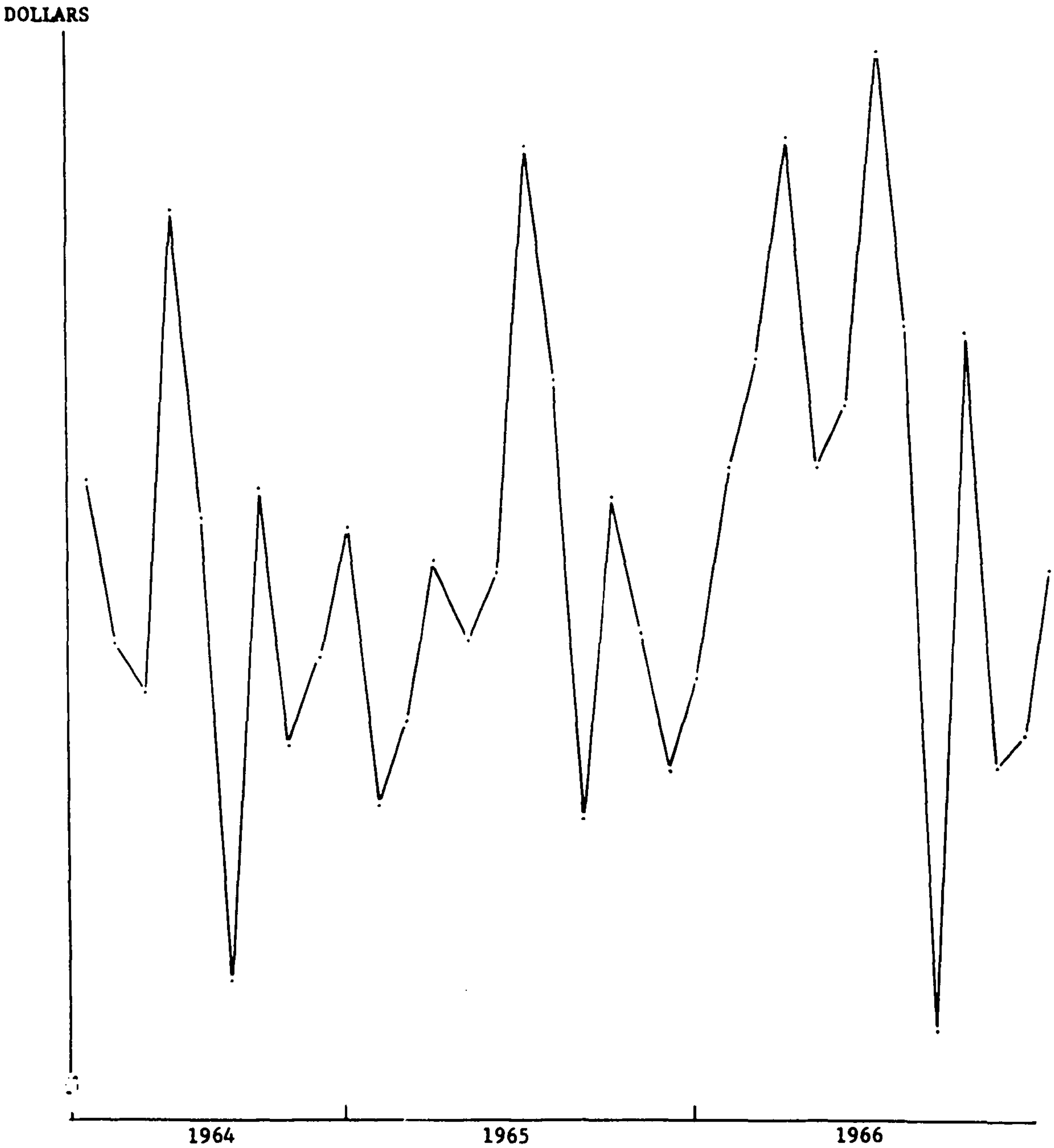

Chart 6.10. Time Study Allowances Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

The existence of a curvilinear time relationship between the two series is not surprising after re-examining the two plotted series: The explanation for this type of relationship is much less readily discernible. It may be partially attributed to the changing proportions of labor and materials required, during the three-year period, to produce a unit. These changing proportions may have occurred because of model changes and increased automation.

Although equation (21) possesses excellent statistical characteristics, it should not be immediately accepted as the model for time-study allowances. That is, it appears estimating equation (21) is relatively efficient in its capability to eliminate the effects of time on the X_1, X_3 relationship; however, X_3 is employed only in its linear form and it seems likely that a curvilinear form might be more appropriate. Therefore, a number of iterations were carried out using the X_2^2, X_2, and various combinations of the X_3 term. The most efficient of the equations computed was the following:

$$X_{c1.2^2 2 3^2 3} = +\$2{,}774.70 - \underset{(1.478)^2}{\$6.9725X_2^2} + \underset{(52.793)^2}{\$218.173X_2} + \underset{(.0000000729)^3}{\$.0000002547X_3^2}$$

$$- \underset{(.006318)^3}{\$.0018963X_3}. \qquad \text{(Eq. 22.1)}$$

For equation (22.1) the $R^2_{1.2^2 2 3^2 3} = .9128$, which is a .0367 increase over that of equation (21). The $S_{y1.2^2 2 3^2 3} = \$726.69$, represents a \$124.84 decrease from that of equation (21). That the X_2 coefficients did not materially change lends validity to equation (22.1). However, in regard to the X_3 variable, it must be noted that in equation (22.1) the X_3^2 term explains vastly more of the variation in X_1 than does the X_3 term. In fact, the X_3 term does not explain a significant amount of the variation in X_1

and X_3 is completely dominated by its standard error. Therefore, it appears a curvilinear relationship does exist between X_1 and X_3, but that the relationship may not be described accurately by the use of X_3^2 and X_3. Since the X_3 linear term is not even marginally significant, it was excluded from subsequent iterations. These iterations employed the X_2^2 and X_2 terms and a single X_3 term with an exponent ranging from 1.1 through 2.8. Of this series of iterations the following equation was the most efficient.

$$X_{c1.2^2 2 3^2} = +\$2,385.77 - \underset{(1.456)}{\$6.9647X_2^2} + \underset{(51.93)}{\$217.873X_2} + \underset{(.000000013327)}{\$.0000002331616X_3^2}$$

(Eq. 22.2)

For equation (22.2) the $R^2_{1.2^2 2 3^2} = .9125$, which is significantly greater than zero at the .05 level. The $S_{y1.2^2 2 3^2} = \$715.58$, which is extremely small when related to the monthly data observations. All of the coefficients of partial determination are significant at the .05 level. The regression coefficients and their related standard errors were subjected to t-tests. Each regression coefficient was significantly different than zero.

The form of equation (22.2) is of interest. The curvilinear time relationship has already been discussed, so there is no further need to analyze it. The curvilinear relationship between X_1 and X_3 may be attributable to several factors. First, the direct-labor-hours - piece-work series is not a prime explanatory variable. That is, the true explanatory variable is the number of units produced for which allowances are given, but it appears this series moves with the direct-labor-hours - piece-work series. Second, time-study allowances expense includes charges for time lost, in some instances, because of material shortages, etc. As the number of direct labor hours increases, it is to be expected that some types of delays would occur more frequently, the rate of delay occurrence being higher than the rate of increase in the direct labor hours. Therefore, the data indicate that efficiency in the service departments, machine operations, etc., apparently declines when the direct labor hours exceed a certain level.

In view of the foregoing analyses, it appears that equation (22.2) should be accepted as a model on both statistical and analytical grounds. The next step is to plot both the X_1 and X_{c1} series on the same chart and examine the differences. These series appear on Chart 6.11. Of the 34 $X_1 - X_{c1}$ differences, 18 are above the X_1 curve and 16 are below. For the most current twelve-month period, 6 of the $X_i - X_{c1}$ differences are above the X_1 curve and 6 are below. Assuming normality, the magnitude of the variances is quite acceptable in that 20 of the observed differences occur within $\pm\ 1.00S_{y1.2^{2}23^{2}}$ and the remainder are within $\pm\ 1.96S_{y1.2^{2}23^{2}}$. In view of the criterion established for equation acceptance, it appears that estimating equation (22.2) may be readily accepted as a model of the Time-Study Allowances—56.

Estimating Equation - Stockroom Expenses—69

Costs are incurred in this account for the purposes of receiving, maintaining, and disbursing raw materials and some factory supplies. The lack of a dollars-of-raw-materials-received series resulted in this expense being analyzed in the Category I group. However, this is an exception to the category classification criteria and it may not be an appropriate classification in another firm.

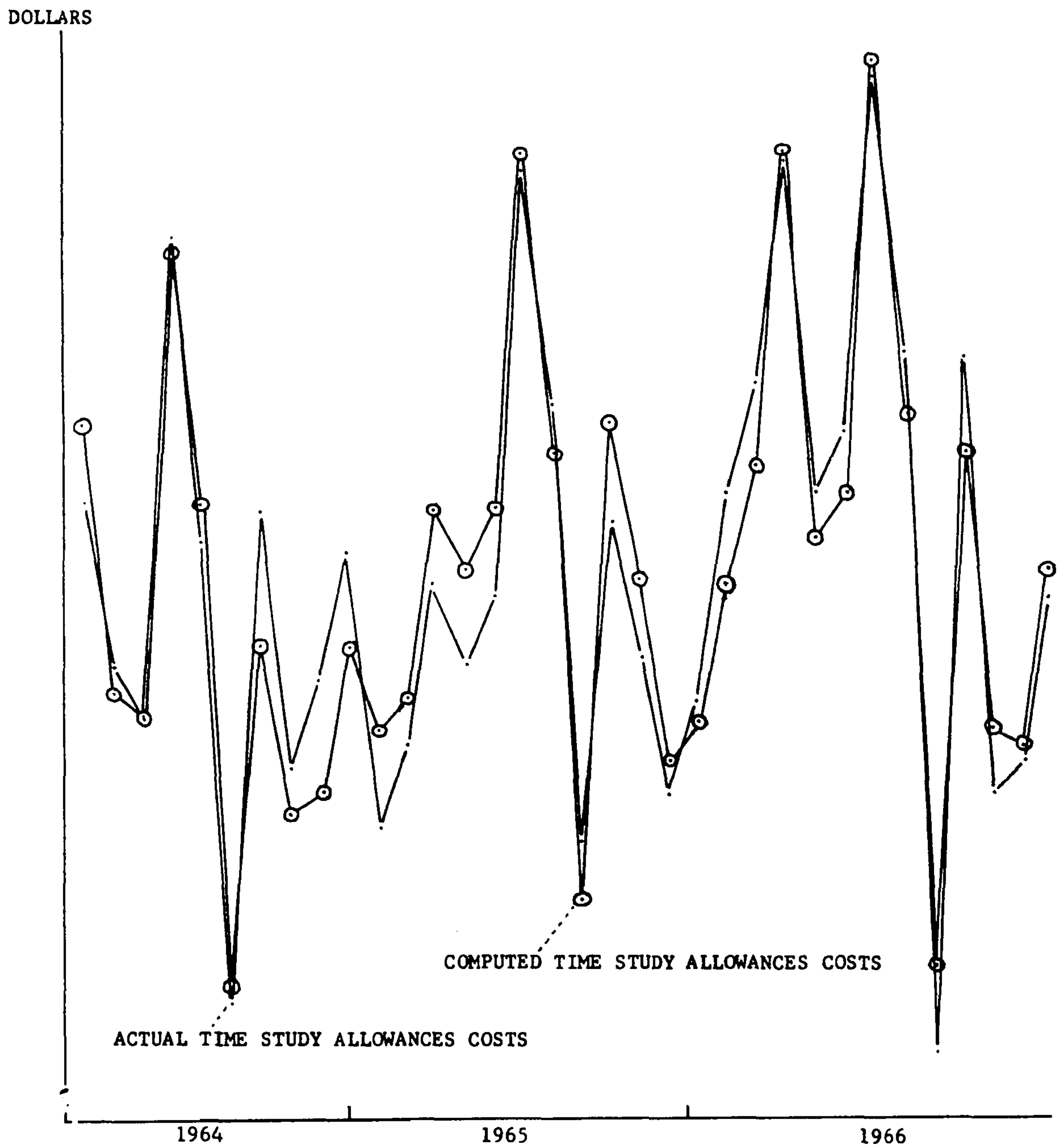

Chart 6.11. Actual Time Study Allowances Costs and Computed Time Study Allowances Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

The stockroom expense series is plotted on Chart 6.12. Perusal of the plotted series revealed a positive and somewhat curvilinear trend. Fluctuations in the series are of considerable amplitude; however, the amplitude is probably exagerated because of the scale against which the series is plotted. In any case, the fluctuations are quite consistent in terms of both timing and amplitude. Because of the nature of the functions performed by the department's personnel, it was felt the dollars-of-material usage would be the most likely explanatory variable series. However, because certain types of raw materials are received by the stockroom personnel, the material usage series alone will probably not explain a particularly high proportion of the department's total cost. A search of the company's records revealed there was no permanent file of receiving reports for even a six-month period. The procedures for establishing the monthly purchases in dollar amounts was such that a series of that type would have been inaccurate. The inaccuracy was attributable to failures in making proper cut-offs at the end of each accounting month. It appeared impractical to attempt reconstruction of a dollar or quantity series of raw materials received by the stockroom, so it is hoped an adequate estimating equation can be developed without the use of such a series. But one mitigating factor should be of significant benefit. That is, interviews with the department foreman indicated that the volume of raw materials received each month was relatively stable. Therefore, based on this information and the nature of the department personnel's other functions, it would seem that a working-day index might be an appropriate explanatory variable.

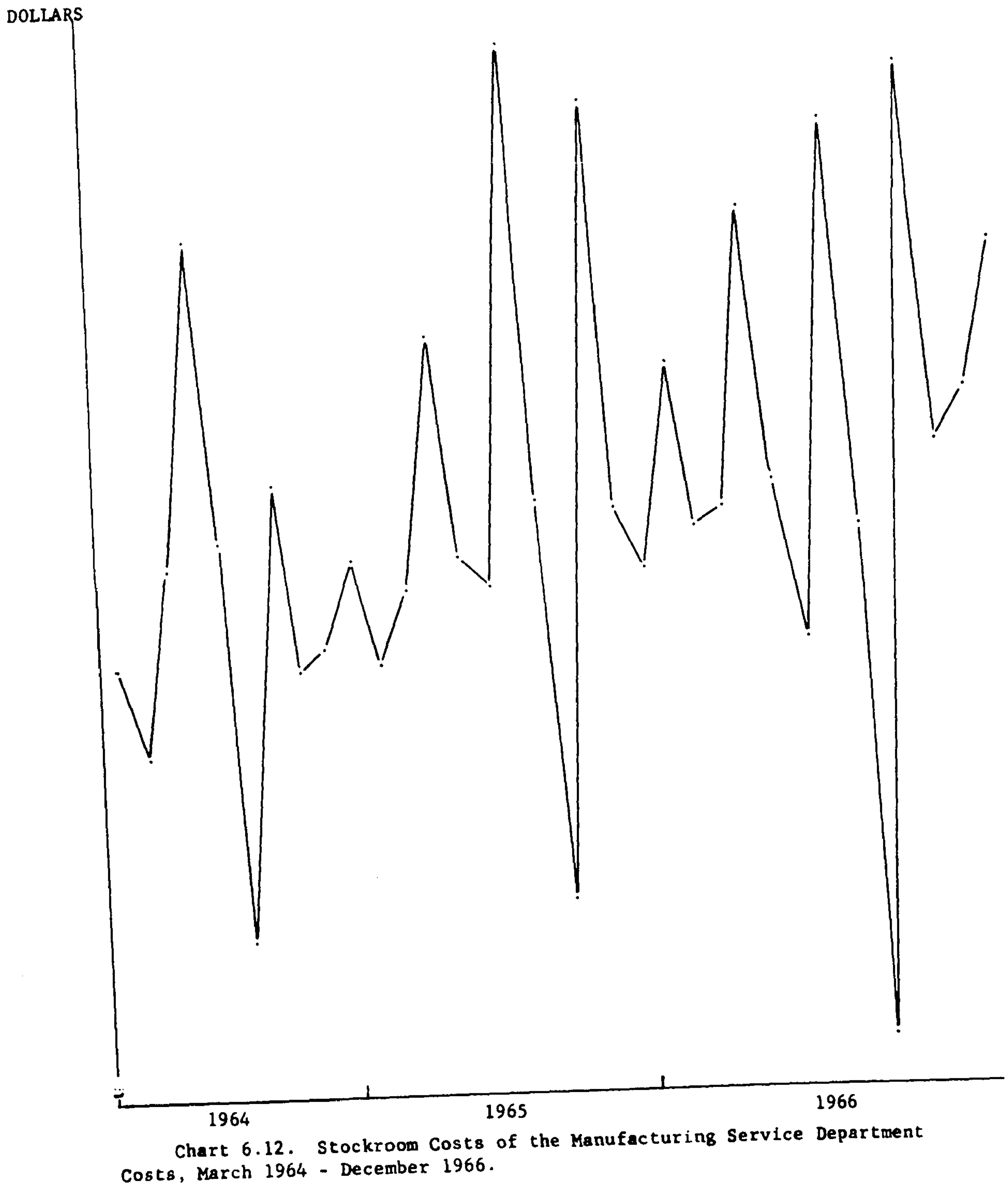

Chart 6.12. Stockroom Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

Based on the foregoing discussion, the first computational efforts will employ a linear equation form and the following variables:

X_1 = Stockroom Expenses--69

X_2 = Time

X_3 = Dollars of material usage.

The first set of computations provided an $R^2_{1.23} = .5814$, which is, of course, significantly greater than zero at the .05 level. The $S_{y1.23}$ = \$571.66, which is quite satisfactory relative to the absolute magnitude of the individual monthly observations.

To improve the efficiency of the estimating equation, a number of additional iterations were carried out, employing various combinations of the X_2 and X_3 terms with different exponents. The most efficient equation determined was of the form:

$$X_{c1.2^3 2^2 23} = a_{1.2^3 2^2 23} + b_{12^3.2^2 23} X_2^3 - b_{12^2.2^3 23} X_2^2 + b_{12.2^3 2^2 3} X_2 + b_{13.2^3 2^2 2} X_3, \qquad \text{(Eq. 23)}$$

for which the $R^2_{1.2^3 2^2 23} = .6996$ and the $S_{y1.2^3 2^2 23}$ = \$500.41. Although the X_3 variable was employed only in a linear case, analysis of the variance associated with the X_3 term indicated that a higher order term should probably be included. However, when computations were made using a higher order X_3^2 term, the increase in the per cent of variation explained in X_1 was not significant. It is felt this lack of significance may possibly be blamed on the composite character of the dependent variable. Inclusion of an X_4 term representing a working-day index should assist in clarifying the relationship between X_1 and X_3.

The next step, therefore, is to amend equation (23) by adding an X_4 term for the working-day index. The variables to be employed are:

X_1 = Stockroom Expenses--69

X_2 = Time

X_3 = Dollars of material usage

X_4 = Working-day index.

The following estimating equation was then computed.

$$X_{c1.2^3 2^2 234} = -\$582.85 + \underset{(.1135)}{\$.2510797X_2^3} - \underset{(6.139)}{\$15.6343X_2^2} + \underset{(92.19)}{\$261.725X_2}$$

$$+ \underset{(.003235)}{\$.0163458X_3} + \underset{(115.58)}{\$273.577X_4} \qquad \text{(Eq. 24.1)}$$

For equation (24.1) the $R^2_{1.2^3 2^2 234}$ = .750, representing a .05 increase over that of equation (23). In addition, the $S_{y1.2^3 2^2 234}$ = \$464.88, which was a \$35.53 decrease. The value of the coefficient of multiple determination was significant at the .05 level and the standard error of the estimate value of \$464.88 is quite satisfactory. The individual regression coefficients and their respective standard errors were subjected to t-tests and all were significant at the .05 level. In addition, all five of the coefficients of partial determination were tested for significance and all were significant at the .05 level.

The inclusion of an X_4 term did not increase the coefficient of multiple determination to the degree desired, but its explanatory contribution was significant. Tests were subsequently made to determine the most efficient form of X_4 which was found to be linear.

The next step was to re-test the validity of employing both an X_3^2 and an X_3 term now that the influence emanating from the working-day variable had been eliminated. The computations employed the following variables:

X_1 = Stockroom Salaries

X_2 = Time

X_3 = Dollars of material usage

X_4 = Working-day index.

Using these variables and the previously discussed equation form, the following equation was developed:

$$X_{c1.2^3 2^2 2 3^2 3 4} = -\$2,086.57 + \underset{(.1091)}{\$.2449}X_2^3 - \underset{(5.899)}{\$15.3021}X_2^2 + \underset{(88.563)}{\$260.399}X_2$$

$$- \underset{(.0000003393)}{\$.0000006207}X_3^2 + \underset{(.01006)}{\$.03385}X_3 + \underset{(113.081)}{\$312.794}X_4. \quad \text{(Eq. 24.2)}$$

For equation (24.2) the $R^2_{1.2^3 2^2 2 3^2 3 4} = .7775$ and the $S_{y1.2^3 2^2 2 3^2 3 4} = \$446.56$. Although the 2.75 per cent increase in variation explained by the addition of X_3^2 is quite marginal, in terms of statistical significance, it provided the best fit among the equations in this series employing a curvilinear X_3 term. A logical basis for the retention of the X_3^2 term may exist. That is, as more materials are issued, during any given period of time, it should be expected that employee efficiency will drop, thereby causing costs to increase geometrically. The exact rate of increase is, of course, subject to considerable variation; the X_3^2 term may be regarded as only an approximation of the actual average rate. However, since the X_3^2 term is not significant, equation (24.2) will not be utilized.

It is apparent that the independent variable (X_1) still contains significant unexplained variation, but it seems that sufficient variation has been explained to permit the valid use of estimating equation (24.1) for establishing annual budgets and for limited cost-control purposes. However, because of the lack of appropriate explanatory variable series, interpretation of variances from the computed stockroom expenses curve may be difficult.

The X_1 and X_{c1} curves are plotted on Chart 6.13. Of the 34 $X_1 - X_{c1}$ differences, 19 are above the X_1 curve and 15 are below. During the most current twelve-month period, 7 of the $X_1 - X_{c1}$ differences are above the X_1 curve and 5 are below. The dispersion of the differences, in terms of dollar magnitude, and assuming normality, is quite satisfactory. Of the 34 differences, 23 or 67.6 per cent, occur within $\pm$ $1.00S_{y1.2^3 2^2 234}$ and none of the differences occur outside the $\pm$ 1.96 standard error of the estimate range. This relatively tight pattern of dispersion imputes additional validity, particularly in view of the smallness of the standard error of the estimate. This same feature also mitigates the implications of the series-like, but unpatterned behavior of the $X_1 - X_{c1}$ differences.

In light of the criteria for acceptance and in view of the information constraints mentioned, estimating equation (24.1) will be accepted as a tentative model of the Stockroom Expenses—69.

Rework Costs—55

This is the last account within the Category I group that will be subjected to correlation analysis. Costs are incurred for the purpose of correcting defects in the units produced and the account includes charges for both materials and direct labor. In terms of absolute dollar amounts the charges to this account are relatively small, but they are significant. Most importantly, management is vitally concerned with controlling costs of this type. However, this is one of those accounts which might not be

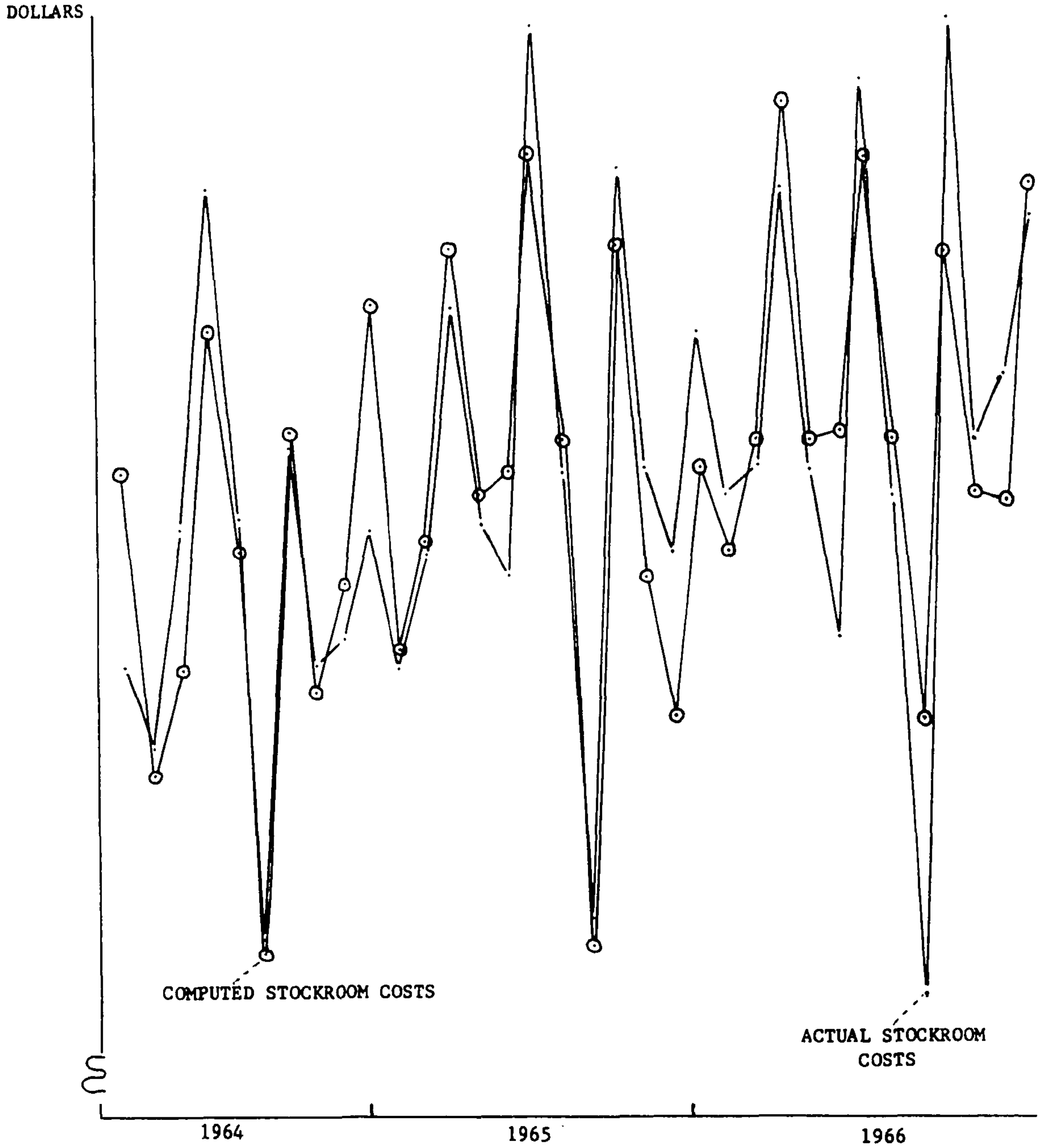

Chart 6.13. Actual Stockroom Costs and Computed Stockroom Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

significant for some other firm or for which there was no pattern of cost incurrence. In such cases it probably would not be practical to attempt correlation analysis of the account.

In the case of the Grace Company, the rework costs demonstrated a consistent pattern. During the high production months of March, June, and September, rework costs were lower than in the other months, particularly when related to either the direct-labor-hours series or dollars-of-materials-usage series. On a month by month basis there was no apparent relationship between rework costs and either of the two possible explanatory series mentioned. The next possible approach considered was a lead-lag relationship. For example, March rework costs were related to the February dollars of material usage. However, this approach also proved unfruitful. It was obvious that additional information was needed, although the ledger account had been previously examined and various personnel of the cost department had been interviewed. Therefore, it was decided that the plant manager should be interviewed regarding several facets of the rework costs. This interview revealed that during high production months it was not possible to rework or scrap all the defective units encountered during those same months. Therefore, it was necessary to carry over defective units to a less intensive production month, during which they could be reprocessed. During the less busy production month(s) the defective units carried over were corrected and, in addition, every effort was made to correct all defective units generated in those same months. Finally, the interview revealed that a considerable effort was made to have no defective units on hand at December 31 of each year.

Based on the information obtained in the interview and in view of the production pattern, the monthly rework costs were grouped as follows:

A. January and February

B. March, April, and May

C. June, July, and August

D. September, October, November, and December.

The Rework Costs—55 series is plotted on Chart 6.14 according to the foregoing groups. As will be noted, this grouping procedure resulted in only 11 observations being plotted as opposed to the normal 34 observations.

The next problem is to determine the appropriate explanatory variable(s) to employ. The two prime candidates are the direct-labor-hours series and the dollars-of-materials-usage series. The first series was considered and subsequently eliminated on the basis that the proportion of labor time contained in each unit had declined over the 34-month period. Further, a previous interview had revealed that a large proportion of the defects occurred in the early stages of production, before much direct labor had been added. On the other hand, a large proportion of the total materials cost had been added before the defects occurred. Therefore, a test was made wherein the annual rate of increase in the dollars of material usage was compared to the annual rate of increase in the rework costs series. Surprisingly, the rates of increase were almost identical for each series, for each year. In view of these considerations, the dollars-of-material-usage series was employed as the explanatory variable.

The first computational efforts utilized the following variables:

X_1 = Rework Costs—55

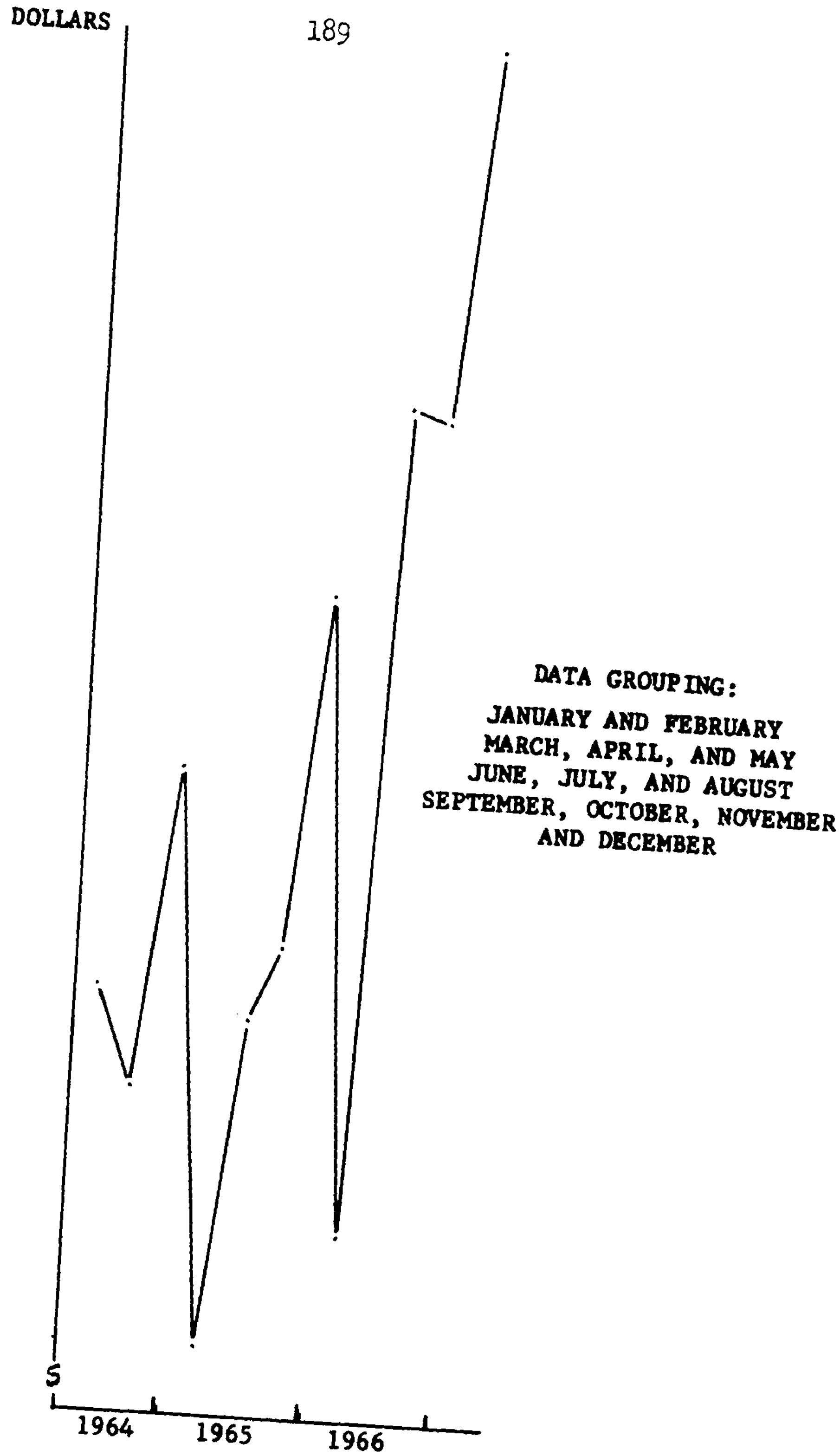

Chart 6.14. Rework Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

X_2 = Time

X_3 = Dollars of material usage

and the linear equation form:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3$$

This first iteration provided an $R^2_{1.23}$ = .9802 and a $S_{y1.23}$ = \$975.04. Both of these statistics are impressive; however, the standard error of the estimate is somewhat large relative to the value of R^2. In any case, the data was re-examined for clerical misstatements and the calculations were repeated. The results were precisely the same in the second iteration. Therefore, the conclusion must be accepted that extremely high correlation exists between the two series.

Despite the high value of R^2, it is necessary to test other equation forms before the strictly linear form may be accepted as the most efficient. In view of the relative value of the standard error of the estimate, there is a high probability that a more efficient equation form does exist. The likelihood of the time relationship's not being linear is small because of: (a) the earlier computations made regarding the annual rates of change for the two series and (b) the high R^2 value. However, several iterations which employed X_2^2 and higher power terms were subsequently carried out. These iteration were inconclusive regarding the linear time relationship.

The next step was to test the possibility of a curvilinear relationship existing between X_1 and the X_3 term. Reduction of the standard error of the estimate was the objective; increasing the R^2 was not a relevant consideration. Subsequent iterations employing a linear X_2 term and different combinations of the X_3 term provided a more efficient equation.

$$X_{c1.23^2 3} = -\$7,828.525 + \underset{(51.31)^2}{\$90.865X_2} - \underset{(.000000005277)^3}{\$.00000003621X_3^2} + \underset{(.0004743)^3}{\$.00753175X_3}. \quad \text{(Eq. 25.1)}$$

For equation (25.1) the $R^2_{1.23^2 3} = .9974$, which is phenomenal, and the $S_{y1.23^2 3} = \$375.04$. It is the reduction of \$600.00 in the standard error of the estimate that is so very significant.

The standard errors for the regression coefficients of X_3^2 and X_3 are small and consequently provide high t-values. But the standard error of the X_2 time term is such that the X_2 regression coefficient is not significantly greater than zero at the .05 level. In order to maintain consistency in methodology and in view of the purposes for inclusion of a time variable, retention of the time term X_2, was desirable. Therefore, a series of iterations were carried out using the X_3^2 and X_3 terms and various forms of the X_2 variable. These iterations indicated it was not feasible to represent time with more than one term. For that reason a series of iterations were then made using a single X_2 term with exponents ranging from .8 through 3.4. Of these iterations the following equation offered the best fit.

$$X_{c1.2^{3.4}3^2 3} = -\$7,981.66 + \underset{(.19819)^2}{\$.45661X_2^{3.4}} - \underset{(.000000005724)^3}{\$.000000041093X_3^2}$$

$$+ \underset{(.00046701)^3}{\$.00782479X_3} \quad \text{(Eq. 25.2)}$$

For equation (25.2) the $R^2_{1.2^{3.4}3^2 3} = .9979$. The $S_{y1.2^{3.4}3^2 3} = \$340.34$, represents a decrease of approximately \$25 from that of equation (25.1). Tests of the $r^2_{13^2.2^{3.4}3}$ and $r^2_{13.2^{3.4}3^2}$ indicated both were more significant than in equation (25.1). In addition, $X_2^{3.4}$ is marginally significant at the .05 level. In statistical terms, estimating equation (25.2) appears extremely valid.

The X_1 and X_{c1} curves are plotted on Chart 6.15. Of the 11 $X_1 - X_{c1}$ differences, 6 are above the X_1 curve and 5 are below. This even distribution and the lack of a pattern provide additional validity. Assuming normality, the dispersion of the differences in terms of dollar magnitudes is excellent. Of the 11 $X_1 - X_{c1}$ differences, 8 or 72.7 per cent, are within a range of $\pm 1.00S_{y1.2^3.4_3^2{}_3}$ and the remaining differences are well within the $\pm 1.96S_{y1.2^3.4_3^2{}_3}$ range. Estimating equation (25.2) was accepted as a valid model for Rework Costs--55.

From the foregoing discussion it should be obvious to the reader that this writer was quite surprised to obtain such a large coefficient of multiple determination. It was to be expected that a fairly high R^2 would be obtained because of the nature of the dependent variable being examined and its apparent relationship to the independent variable. But such results as those actually obtained are unusual. In the first place, the variation in both series was of considerable magnitude. Second, only eleven observations were employed in the analysis. This second factor is of primary importance, for as the number of observations decreases, the degree of confidence must also decrease. Generally, this reduction in confidence is reflected in the increased size of the standard error of the estimate. But in this case, the standard error of the estimate was quite small, relative to the actual X_1 values. In spite of the excellent fit demonstrated by estimating equation (25.2), it is quite likely that future periods may provide instances where the $X_1 - X_{c1}$ differences are considerably outside a range of $\pm 1.96S_{y1.2^3.4_3^2{}_3}$. These statements are not intended to repudiate the validity of estimating equation (25.2); their purpose is only to

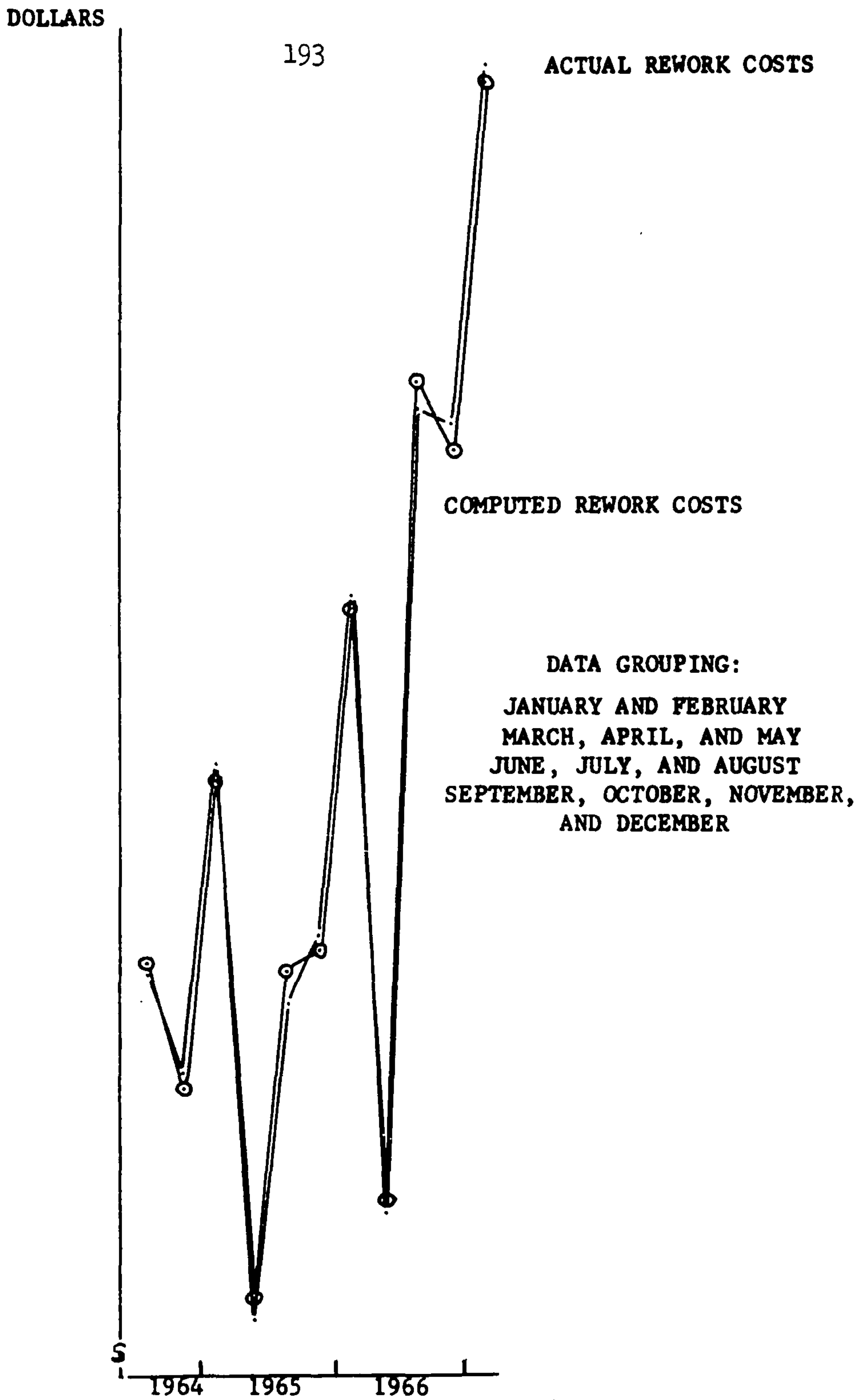

Chart 6.15. Actual Rework Costs and Computed Rework Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

caution the reader that even the highest of R^2 values does not prove causation or future applicability.

With the completion of the analysis of the Rework Costs—55, all of the Category I accounts have been examined. The next chapter contains the Category II and III account analyses, at the end of which a summary of all the variable costs account analyses will be offered.

CHAPTER VII

DEVELOPMENT OF AN OVERHEAD COST MODEL FOR THE MANUFACTURING SERVICE DEPARTMENT - III

In this chapter the variable cost Categories II and III will be analyzed. The general methodology will be similar to that of Chapter VI. Several of the explanatory variable series utilized in Chapter VI will probably be appropriate for some of the accounts to be analyzed. Thus occasional references may be made to the charts in Chapter VI.

Determination of the Category II Estimating Equations

In contrast to the analysis of the Category I accounts, in which only one primary independent variable was employed, it may be necessary to utilize several explanatory variables or derivations of the primary variables in order to develop estimating equations for the Category II accounts. The accounts to be analyzed in this section are: Direct Supervision--50, Machinery Repairs--51, Utilities Expense--61, Overtime Premium--58, Night-Shift Premium--59, Indirect Supplies--52, and Scrap Recovery--86.

The broad underlying purpose for the company incurring these costs was, of course, to produce products. It would be convenient if these costs could be directly related to either of the two primary resource inputs of direct labor or raw materials. But unfortunately, such direct relationships might not exist, at least on a monthly basis. Therefore, as was previously mentioned, it may be necessary to employ secondary explanatory variables: some may be derivatives of a primary resource series, some may be the result of past management decisions, and others may be synthetic explanatory variables. This latter type of explanatory variable may be manifested in the form of a seasonal index, an index of investment in machinery and equipment, etc. The type of explanatory variable employed is important, but it does not necessarily need to be a causation variable. What the explanatory variables must do is provide a logical basis for estimating future costs, given that production will continue.

Estimating Equation - Direct Supervision—50

Salaries paid to departmental foremen and assistant foremen are charged to this account. Although overtime wages are not paid to supervisory personnel, they do receive additional compensation when their entire department is required to work extensive overtime. In prior periods department-wide overtime has occurred at reasonably patterned intervals. Additional compensation is also paid during August of each year, as some of the foremen are required to work considerably longer hours than in other months because of the maintenance and repair work carried out.

The direct supervision series is plotted on Chart 7.1. Perusal of the series reveals a definite pattern during 1965 and 1966. For both of these years the months of January, February, March, April, and May were quite stable. During June a considerable fluctuation occurred, but in July the pattern appears restored. In August a second pattern disturbance results in a new pattern of stability in September, but at a slightly stepped up

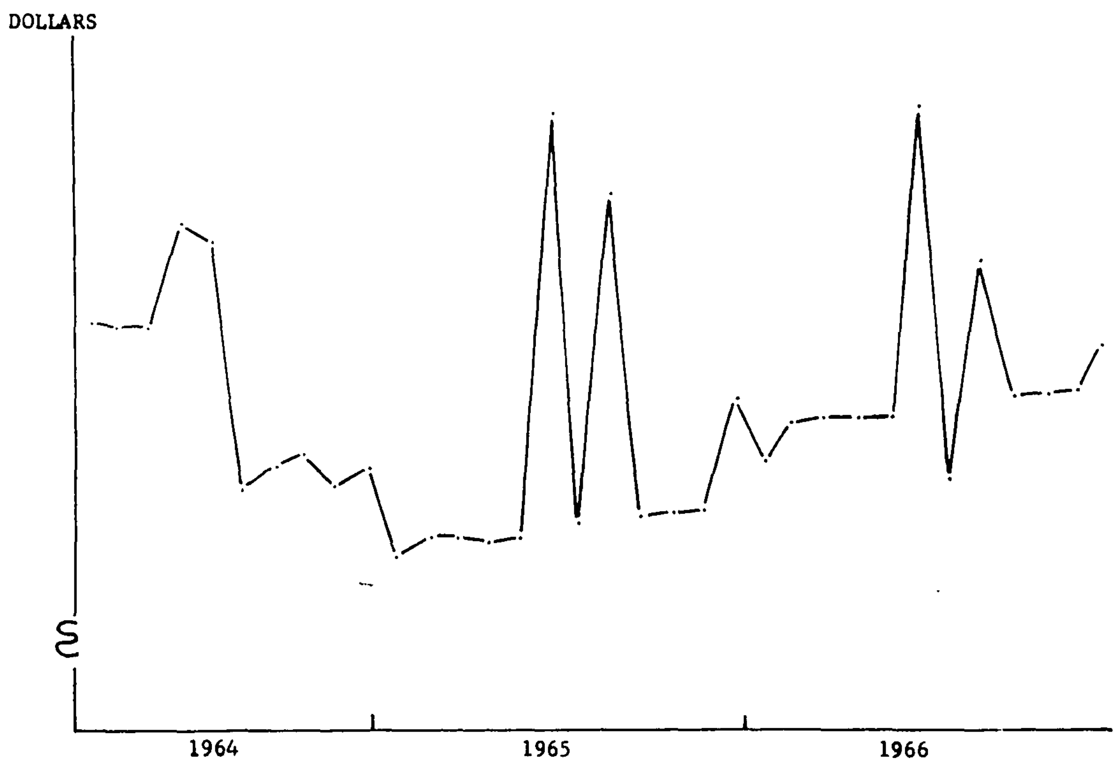

Chart 7.1. Direct Supervision of the Manufacturing Service Department Costs. March 1964 - December 1966.

level. The December fluctuation is present in both years, but it is not nearly so extreme as the June and August fluctuations. The year 1964 presents almost a complete contrast in behavior.

The next step was to consider several data series as possible explanatory variables. The series considered were the total units produced, direct labor hours, dollars of material usage, monthly average number of production workers, and overtime premium paid. None of these appeared to offer a reasonable basis for developing an estimating equation. Therefore, these same series were examined with a view toward combining two or more series or for the possibility of developing some type of ratio or index. All of these efforts proved futile. It was then obvious that additional interviews should be made.

In subsequent interviews it was learned that the unusual pattern in 1964 was the result of management's inexperience in coping with illusional prosperity. That is, during the year 1964, sales increased dramatically in the first few months; and management, in its desire to increase production, arbitrarily gave temporary percentage increases to all production-line supervisors. During June and July the percentage of increase was again raised. The drastic reduction in August occurred because of management's awareness of the inadequate profit picture and a lower production level. But for the remainder of 1964 the cost line could not be held completely because of previous commitments. Therefore, at the end of December, 1964, a new policy was instituted by management, and several supervisors left. As a result of this new policy, percentage increases are given only in June and August, except for December, when one is given primarily as a goodwill measure. These incentive increases are determined arbitrarily and are not an absolute function of production volume, etc.

In view of the foregoing, it appears that the Direct Supervision—50 account cannot be subjected to correlation analysis. A logical and persistent activity must be present before correlation analysis is beneficial; such is not the case. However, there is no reason why an equation could not be employed to represent Direct Supervision in a budget cost model. A possible equation form would be:

$$X_1 = bX_2 + cX_3,$$

where:

X_1 = Monthly direct supervision

X_2 = Number of supervisors

b = Average monthly salary per supervisor

X_3 = Percentage of salary to be paid as an incentive

$c = bX_2$.

The b term was designated an average, because a limited turnover does persist among the supervisory personnel and there are numerous salary levels. For all months except June, August, and December the cX_3 term will be zero, as it is only in those months that X_3 will have a value greater than zero. The annual budget could be determined by merely taking $X_{1...12}$.

Estimating Equation - Machinery Repair Expense—51

Charges are made to this account for the cost of supervision and wages in the tool and maintenance department, wages paid to other direct and indirect workers for time spent in repairing machinery and equipment, parts, small tools, and some of the less expensive dies. The machinery repair account obviously does not represent a single department; indeed, it is a complex overhead account. Adding to the complexity is the fact that during intensive production periods personnel of the tool and maintenance department are often used as production workers.

The company's accounting system does not accumulate costs related to the tools and maintenance department and then subsequently reapportion this cost to the appropriate production departments. Instead, labor and parts are charged directly to the production department in which the work was done. Of course, there is some expense incurred that cannot be directly related to a specific production department. When this occurs, the expense is charged to the tool and maintenance department and then subsequently reapportioned to the production departments based on standard direct labor dollars.

The machinery repair series is shown on Chart 7.2. There are 36 observations covering the period January 1, 1964, through December 31, 1966. Peak months are consistently March, August or September, and November. The overall pattern of the series is quite consistent both in terms of the timing and amplitude of the fluctuations. The series' trend is positive and slightly curvilinear. The behavioral pattern demonstrated by the series is different from that of any of the previously examined series. Therefore, the work done in developing previous estimating equations is of questionable value in developing an equation for this account. Perusal of the pattern indicated a generally constant minimum monthly cost, but the amplitude of the fluctuations is considerable. It appears that difficulty in developing an

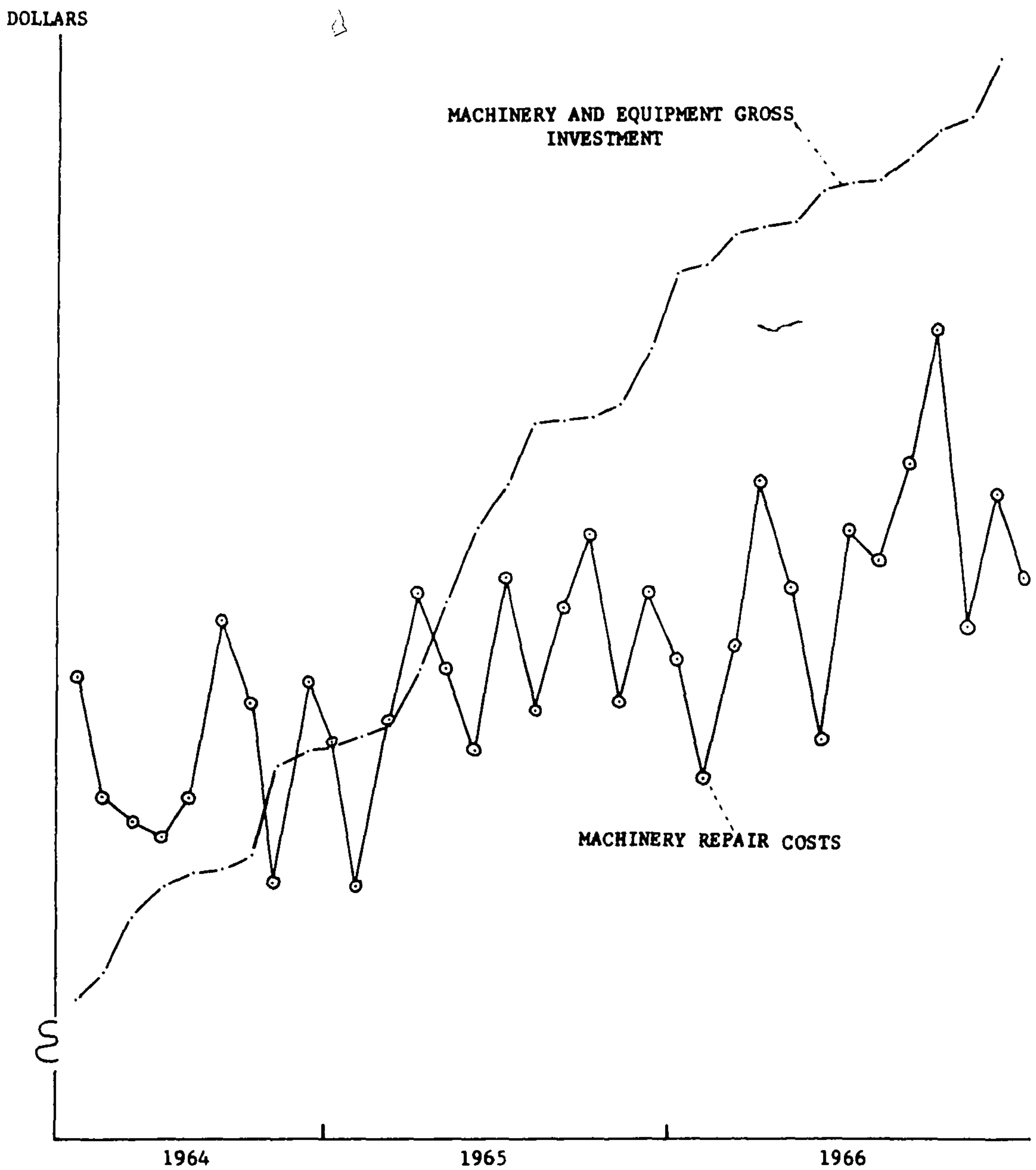

Chart 7.2. Machinery Repair Costs and Machinery and Equipment Gross Investment of the Manufacturing Service Department Costs, March 1964 - December 1966.

estimating equation for the series may be encountered because of the timing of the fluctuations and not because of their amplitude.

In selecting explanatory variable(s) one should attempt to use a primary series such as the direct-labor-hours series. But perusal of the primary series reveals that no month-by-month relationship exists, either in terms of positive or negative correlation. In addition, no immediate lead-lag type of relationship appears appropriate. A final possibility might be to determine some average level of capacity, in terms of direct labor hours, and then correlate the differences with the machinery repair account. However, such a procedure should be utilized only as a last resort.

Aside from the previously mentioned primary explanatory series, it would seem that the amount of machinery utilized in production should be a relevant factor. Presumably, the more machinery employed in production, the larger the maintenance expense, *ceteris paribus*. It would be useless to develop a series expressing only the number of machines employed throughout the plant. A machine hours series would be appropriate but the company does not develop that type of information. Therefore, a series reporting the gross investment in machinery and equipment at the end of each month will be used. In order to develop this series, it was necessary to analyze the fixed asset ledger. The analysis covered the three-year period from January 1, 1964, through December 31, 1966. During this period a small amount of machinery was retired, and no adjustments were made to the series for these items. It was also found that fully depreciated machinery, provided it was still being used, was retained on the books. Thus, no adjustments were necessary for these items. After the analysis was completed it was decided that an investment index would be more appropriate than an actual dollars series. The index was computed using an average-value-of-investment for the eighteenth and nineteenth months as the 100 per cent base. The resulting index appears on Chart 7.2 along with the machinery repair expense series.

By utilizing a machinery and equipment investment index it should be possible to establish an average rate of growth relationship that may be valid for future years. But the problem of the timing of machinery repair expense is still unsettled. The primary explanatory series have been tentatively set aside, one secondary explanatory series has already been suggested, and it now appears a synthetic explanatory variable must be developed. In view of the consistent pattern of fluctuations in the dependent series and the similarity in their amplitude, some type of cost incurrence index is appropriate. To this end a machinery repairs index was constructed in the following manner.

1. Sum the monthly expenses for each year.
2. Determine the per cent of each month's expense to its respective annual sum.
3. Sum the three percentages for the same months in each of the three years and divide by three.

This index in effect provides a percentage value that reflects the average amount of machinery repair expense incurred during any given month in the past three-year period. Note that this series does not attempt to explain either the cause or the amount of machinery repair expense—it merely represents quantification of the prior year patterns of expense incurrence. Its only purpose is to assist in budgeting. Only when definite and consistent behavioral patterns have existed in the past and when they are expected to continue in the future, is this type of index appropriate.

The first computational efforts employed the following variables:

X_1 = Machinery Repair Expense--51

X_2 = Investment Index

X_3 = Machinery Repair Index

and a linear equation form of:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3.$$

The first iteration provided a $R^2_{1.23}$ = .8370, which is quite good, and a $S_{y1.23}$ = \$1,919.73, which is also quite satisfactory. Individually the $r^2_{12.3}$ and the $r^2_{13.2}$ were of approximately equal significance. The next step was to determine the most efficient equation form. Logically, the X_3 term should be linear because of its computational nature. Several iterations were made, all of which indicated the relationship between X_1 and X_3 was definitely linear. The relationship between X_1 and X_2 is not readily determinable, either by logic or from inspection. Iterations were carried out wherein the X_3 term was retained in a linear form, and the X_2 term was varied in form. Of these iterations, the following estimating equation appeared the most efficient.

$$X_{c1.2^32^223} = +\$679{,}545.91 - \underset{(.18510)^2}{\$.662785X_2^3} + \underset{(57.799)^2}{\$204.6059X_2^2} - \underset{(5{,}961.829)^2}{\$20{,}619.497X_2}$$

$$+ \underset{(143.92)^3}{\$1{,}491.263X_3}. \qquad \text{(Eq. 26)}$$

For equation (26) the $R^2_{1.2^32^223}$ = .8855, representing a .0485 increase over that of $R^2_{1.23}$. By employing X_2^3, X_2^2, and X_2 terms approximately 29.75 per cent of the previously unexplained variation was explained. There was a similar reduction in the standard error of the estimate. For equation (26) the $S_{y1.2^3 2^2 23}$ = \$1,660.14, which is a \$259.59 reduction from the previous $S_{y1.23}$.

In statistical terms, estimating equation (26) provides an impressive fit; the coefficient of multiple determination is quite high, and the value of the standard error of the estimate is well within a satisfactory range. The individual coefficients of partial determination: $r^2_{12^3.2^2 23}$, $r^2_{12^2.2^3 23}$, $r^2_{12.2^3 2^2 3}$, and $r^2_{13.2^3 2^2 2}$ are all highly significant. In addition, the variance in each of the (b) coefficients is acceptable. Variance in the (b) coefficient values is a significant consideration if one intends to employ the (b) coefficients as marginal costs, but it has been previously declared that such use is not practical in most cases where non-linear terms have been employed. Second, variance in the (b) coefficients does not inhibit the ability of the estimating equation, as a whole, to produce past actual experience if the assumption of normality is valid. Third, when explanatory variable series, that do not contain variation in proportion to the variation contained in the dependent variable are employed, this writer's experience indicates that considerable variance will be present in the regression coefficients. Finally, when synthetic variables similar to the X_3 term are employed, one knows in advance that a high degree of variance may exist because of the methodology used in creating the explanatory variable. Ergo, when the foregoing conditions are present, one must be prepared to accept the reality of larger variances in the (b) coefficients. But such variance does not destroy the validity of the estimating equation, providing the R^2 and S_y values are within predetermined ranges.

Before equation (26) is accepted as a model for machinery repair expense, it is necessary to establish its capability of duplicating actual experience. A curve representing the actual experience or X_1 values and a second curve representing the X_{c1} values are plotted on Chart 7.3. Of the 36 $X_1 - X_{c1}$ differences, 19 are above the X_1 curve and 17 are below. For the most current twelve-month period there are 7 $X_1 - X_{c1}$ differences above the X_1 curve and 5 below. The behavioral pattern of the differences is not consistent and is not particularly important because of the even dispersion, particularly over the last twelve-month period. The magnitude of the $X_1 - X_{c1}$ differences, and assuming normality, in terms of the $S_{y1.2^32^22_3}$ range, is quite satisfactory. That is 25, or approximately 69.44 per cent, of the $X_1 - X_{c1}$ differences occurs within a range of $\pm$ $1.00S_{y1.2^32^22_3}$; and only one falls outside $\pm$ $1.96S_{y1.2^32^22_3}$. In view of the foregoing analysis, estimating equation (26) will be accepted as a model for Machinery Repair Expense—51.

Before moving on to the analysis of another account, it is appropriate to discuss the departure from the normal methodology that was made in the case of estimating equation (26). This departure consisted of the deletion of a term representing time or trend. There were several reasons for such a departure. First, one must bear in mind the purposes for conducting the correlation analysis. That is, the resulting estimating equation is to be used for budgeting and cost control purposes. Therefore, the explanatory variables employed should be so related to the dependent variable that they

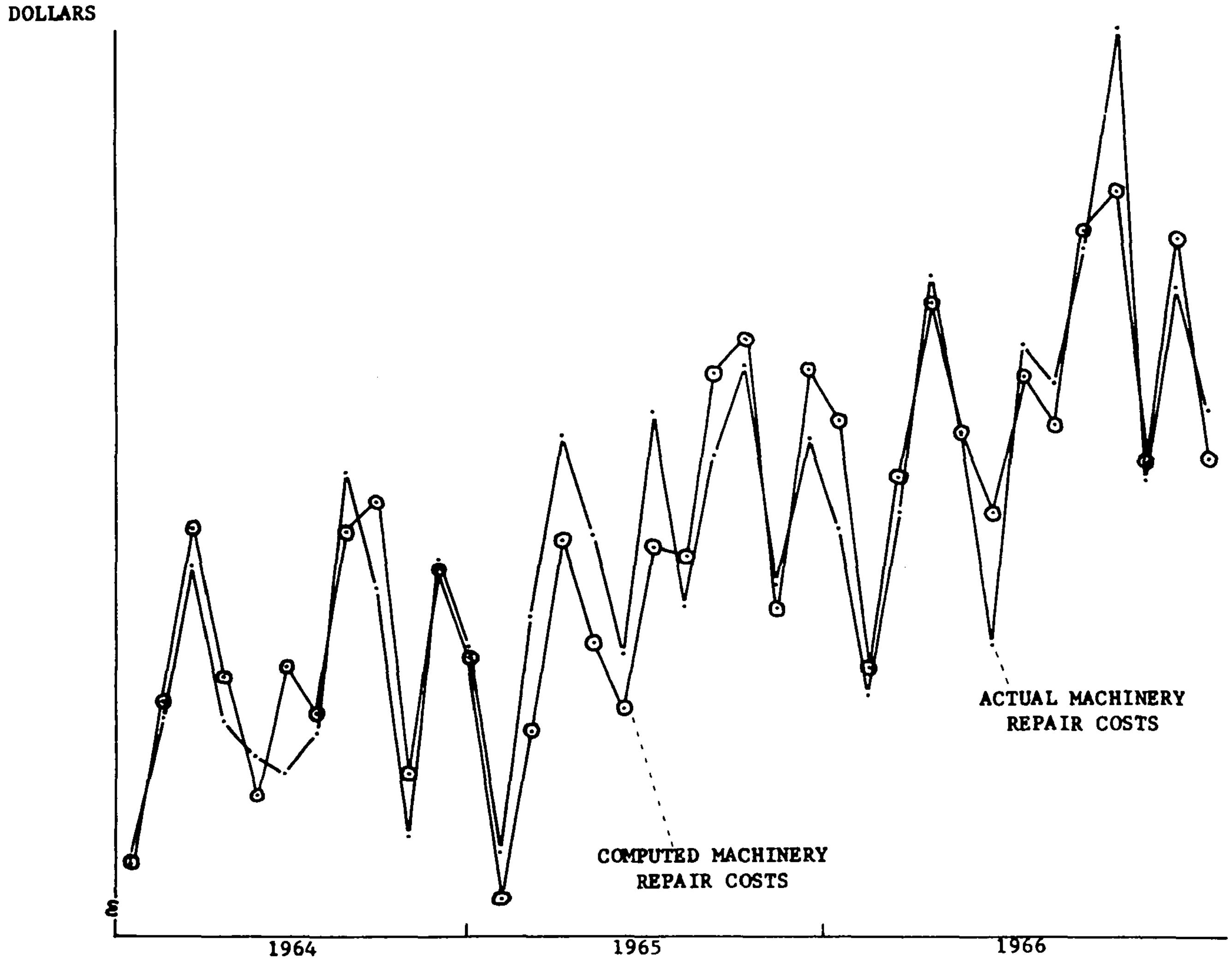

Chart 7.3. Actual Machinery Repair Costs and Computed Machinery Repair Costs of the Manufacturing Service Department Costs. January 1964 - December 1966.

will be of continuing explanatory significance in future periods. The first explanatory variable (investment index) was employed in an effort to establish a basis for estimating future machinery repair expenses, given future capital expenditures. The machinery repair index was employed to assist in estimating the timing and, to some extent, the magnitude of future expenditures. These two explanatory series both supplement and complement one another. However, assume the investment index had been deleted and in its place a series representing time had been employed. In fact, this was done. The results of such a procedure were statistically impressive, as a $R^2_{1.2^3 2^2 23} = .8887$ was derived; and the associated standard error of the estimate was slightly lower than that provided by estimating equation (26). Therefore, a statistically valid estimating equation was in hand. But assume management desired to know the probable implication for machinery repair expense of larger increments in capital investment over the next three years. That type of information would not be obtainable from an estimating equation employing only a trend and a timing index. Obviously such an equation would be quite inflexible and thus undesirable.

Second, assume that an estimating equation was derived utilizing a variable for trend, one for investment index, and one for the timing index. This was done and again the results appeared statistically valid. Of the iterations carried out, the most efficient equation provided a $R^2_{1.23^3 3^2 34}$ $= .8861$ and an $S_{y1.23^3 3^2 34}$ that was slightly higher than for equation (26). However, this approach did not alter significantly the b coefficient value for the timing index, change the value of the (a) term, or reduce the variance of the b coefficients. Further, because of the nature of the investment index it would seem logical that considerable multicollinearity would exist between it and the trend series. On the other hand, this did not appear to be the case, based on this writer's analysis of the various iterations. In any case, it was concluded that the techniques of correlation are such that the true time and investment index relationships were probably "clouded" and, therefore, neither of the associated b coefficients would be indicative of the true relationship. In view of this, it was concluded that the time term should not be included in the final estimating equation. Although this has not been the procedure previously followed, it was believed that in this case the validity, flexibility, and predictive integrity of the estimating equation, as a whole, was enhanced.

Estimating Equation - Indirect Supplies--52

Charges are made to the indirect supplies account for a multitude of sources. Interviews with production personnel revealed that approximately 80% of the total supplies expense was associated with a few items. These few items could be classified into two groups, according to the methods used for collecting cost information. In the first group only one item, actually of the nature of an indirect material, accounted for a major portion of the total indirect supplies cost. This item, for identification purposes, will be designated Supply A. In the second group are several supply items; all of these are carried in the stockroom until such time as they are requisitioned for use in the factory. The remainder of the indirect supplies expense is comprised of numerous small cost items that are purchased for specific production departments or for other factory uses, all of which are taken directly into expense at the time the invoice is received.

The timing or incurrence of indirect supplies expense appears to be fairly closely related to supply usage, when viewed over a period of one month. Supply A is purchased in liquid form with deliveries being made the latter part of each week by the vendor and placed in the company's bulk storage tanks. Invoices are rendered after each delivery, and expense is charged; so it appears this supply cost is reported in a timely manner. The other group of large dollar supply expense items is entered to expense at the end of each month based on a summary of the supply requisitions. For the balance of the indirect supplies expense there is no clerical means of relating use patterns to the expense recognition patterns. But a mitigating consideration is that most of these items are of a reoccurring nature, purchased for periods of 15 to 45 days, and numerous in type. The effect then is to generate a more or less constant but positive sloped expense curve for this portion of the indirect supplies.

The indirect supplies account has been adequately reviewed and may not be subjected to correlation analysis. Because of the nature of this expense, it is logical that the direct-labor-hours series should be closely related. A comparison of the indirect supplies series, shown on Chart 7.4, and the direct-labor-hours series indicates the two series are quite similar. Therefore, the first computational effort will employ a linear equation form and the following variables:

X_1 = Indirect Supplies

X_2 = Time

X_3 = Direct labor hours.

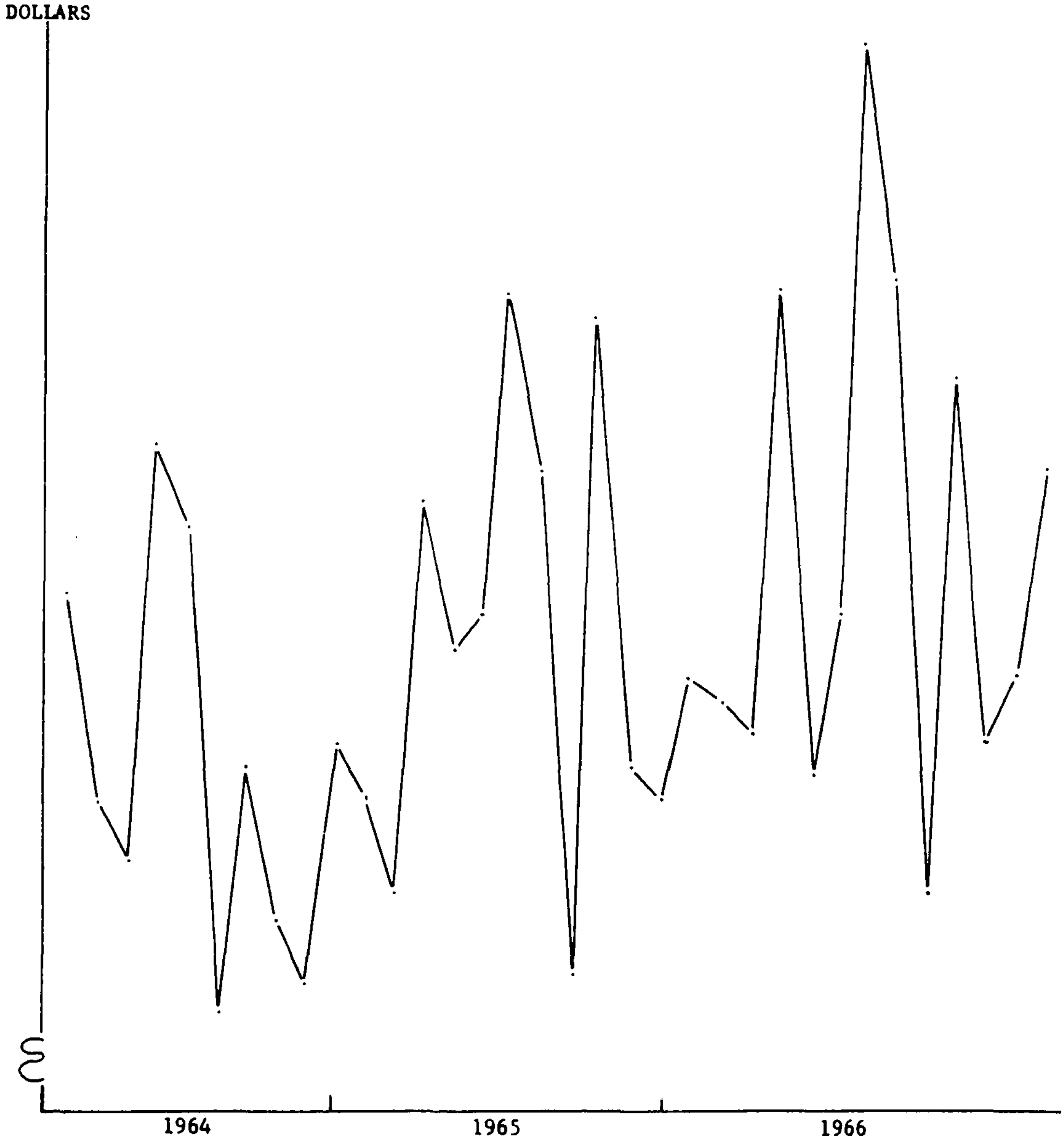

Chart 7.4. Indirect Supplies Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

By employing a linear form, the following equation was computed.

$$X_{c1.23} = -\$1,044.62 + \underset{(33.87)}{\$84.29247X_2} + \underset{(.02112)}{\$.282599X_3}$$

This first calculation provided a $R^2_{1.23} = .7763$, which is definitely significant at the .05 level, and a $S_{y1.23} = \$1,917.58$. Both of these statistical measures of fit are satisfactory and the standard errors for the regression coefficients are low. Possibly the values of both may be improved by determining the most efficient equation form. Tests were first made to determine the true character of the X_2 variable. These tests involved conducting several iterations in which the X_2 term was employed with various exponents. Based on these tests it may be concluded that the time relationship between X_1 and X_3 is approximately linear.

The next step involved making a series of iterations wherein the X_3 term was employed with various exponents. Of these iterations, the best non-linear equation determined, in statistical terms, follows:

$$X_{c1.23^23} = +\$9,569.74 + \underset{(28.733)}{\$89.628X_2} + \underset{(.000001508)}{\$.00000547498X_3^2} - \underset{(.1404)}{\$.218616X_3} \qquad \text{(Eq. 27.1)}$$

For equation (27.1) the $R^2_{1.23^23} = .8445$, representing a 6.82 per cent increase over that of the linear case. The $S_{y1.23^23} = \$1,625.32$, which is quite satisfactory when related to the absolute values of the monthly data. The standards errors of the regression coefficients of X_2 and X_3^2 respectively were satisfactory as both the X_2 and X_3^2 regression coefficients were significantly greater than zero at the .05 level. But the X_3 regression coefficient was not statistically significant at the .05 level and further, the X_3 regression coefficient had a negative sign.

The numerous iterations previously run indicated the X_2 term was approximately linear and that the X_3 variable should consist of two terms with one being curvilinear. It was concluded that the curvilinear X_3 term should probably have an exponent greater than one but less than two. Iterations were subsequently run employing linear X_2 and X_3 terms and an $X_3^{1.n}$ term. Of the equations computed, the following appears the most efficient.

$$X_{c1.23^{1.4}3} = \$13{,}699.24 + \underset{(28.58)}{\$89.3544}X_2 + \underset{(.003255)}{\$.012017}X_3^{1.4} - \underset{(.3318)}{\$.93773}X_3$$

(Eq. 27.2)

For equation (27.2) the $R^2_{1.23^{1.4}3} = .8461$, representing a .0016 increase over the $R^2_{1.23^23}$ of equation (27.1). The $S_{y1.23^{1.4}3} = \$1{,}616.86$, representing a \$9.46 decrease from that of equation (27.1). The X_2 regression coefficient changed only slightly and remained significantly greater than zero at the .05 level. A major change did occur in both the X_3 regression coefficients. But the X_3 and $X_3^{1.4}$ regression coefficients are both significantly greater than zero at the .05 level. Yet the coefficients of the linear X_3 term remained negative.

In the course of conducting this last series of iterations, it was observed that as the $X_3^{1.n}$ exponent was increased the regression coefficient of the linear X_3 term and the pure constant both decreased. This behavioral pattern indicates that the sign of the X_3 linear term relates to the magnitude of the exponent of the curvilinear X_3 term. In an earlier iteration an X_3^3 and an X_3 term were used, and in that case the regression coefficient of X_3 was positive. The purpose of this discussion has been to illustrate the difficulty of rationalizing the signs taken on by particular terms when multiple curvilinear regression is employed. In view of the foregoing, it would seem that equation (27.2) could be justifiably accepted as a model of the Indirect Supplies--52.

Although equation (27.2) is sufficient, additional effort should be expended to increase its efficiency. To this end an X_4 term, representing a working day index, was added as an explanatory variable. The reason for adding the X_4 term is to determine if a type of fixed monthly cost exists which has not already been explained by X_3, and thereby increase the R^2 value. Because of the possibility of high multicollinearity existing between X_3 and X_4, a linear equation form was employed in the initial iteration. In successive iterations different combinations of the X_2, X_3, and X_4 terms employing various exponents were included. As these iterations progressed it became apparent that high multicollinearity did exist between the independent data series. In an attempt to eliminate the effects of the interrelationships between X_3 and X_4 on the estimating equation, the two terms were used as a product. That is, an equation form of the following type was used:

$$X_{c1.2(34)} = a_{1.2(34)} + b_{12.(34)}X_2 + b_{1.(34).2}X_3X_4.$$

But this procedure did not prove fruitful in this case. It was concluded that high multicollinearity exists, because the direct-labor-hours series is quite dependent for its sum on the number of working days in each month. Not only is this a rational assumption but it is also quite probable that the average minimum points on the direct-labor-hours series are closely related, in proportional terms, to the same points on the X_1 series. What these two considerations imply is that a working day index in not a relevant explanatory variable in this case.

It now appears that the $R^2_{1.23^{1.4}3} = .8461$ of equation (27.2) is the highest value obtainable in view of existing information constraints. The unexplained variation in X_1 is not particularly large and might be partially rationalized by noting that clerical errors relating to the timing of recognition of the Supply A expense may exist. Further, there are probably occasions when some supply items, which are not carried in the materials inventory, are purchased in unusually large quantities, thereby creating distortion. But these factors do not materially impugn the validity of equation (27.2).

Chart 7.5 compares the plotted X_1 curve and the $X_{c1.23^{1.4}3}$ curve for equation (27.2). Of the 34 $X_1 - X_{c1}$ differences, 18 are above the X_1 curve and 16 are below. For the most recent twelve-month period, 7 of the $X_1 - X_{c1}$ differences are above the X_1 curve and 5 are below. For the entire three-year period the error in measurement $\rightarrow 0$. This dispersion of the differences is quite satisfactory, particularly in view of their pattern. In terms of dollar magnitude, the dispersion of the differences approximates the assumed normal distribution as 22, or 64.7 per cent, occur within the $\pm 1.00S_{y1.23^{1.4}3}$ range and 33, or 97.0 per cent, occur within the $\pm 1.96S_{y1.23^{1.4}3}$ range. Therefore, estimating equation (27.2) will be accepted as a tentative model for Indirect Supplies--52.

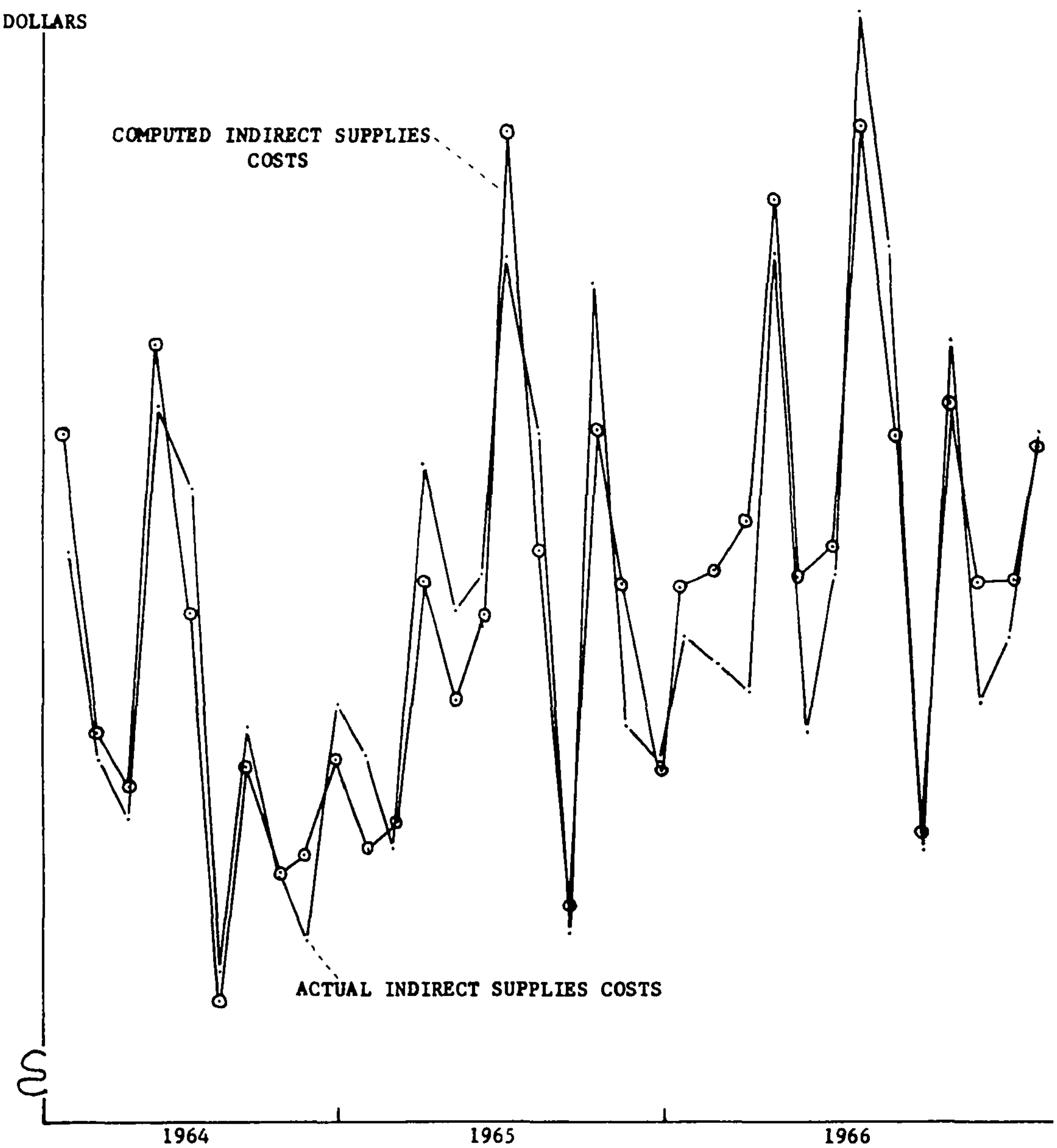

Chart 7.5. Actual Indirect Supplies Costs and Computed Indirect Supplies Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

Estimating Equation - Scrap Recovery--86

There are several different methods of accounting for scrap materials. The Grace Company utilizes a most simple method in that scrap materials are collected daily and then picked up by the buyer when a truckload has been accumulated. The company merely credits the scrap recovery account for the weekly proceeds. The scrap recovery account is a credit balance account in the factory overhead ledger.

Scrap materials are created almost entirely in two of the seven producing departments. These two producing departments manufacture component parts for the final product. An effort is made to maintain a small inventory of these intermediate component parts. For this reason the incurrence of scrap proceeds does not necessarily coincide with the material usage series previously employed as an explanatory variable. Three basic raw materials are utilized in the components, and it is from these that almost all the scrap is created. Records are maintained which lump the proceeds from two of the scrap materials and the third is set out separately.

In conducting the analysis of this account it quickly became apparent that the dollars-of-material-usage series previously used would be inappropriate because of the component parts inventory. Therefore, the three raw materials series mentioned above, and which might be designated X, Y, and Z, were each examined as explanatory candidates. It was found that during 1964 the company had not maintained a dollars-of-material-usage series for each of the three materials. Rather, during this period the three had been combined. However, during the years 1965 and 1966 they were shown separately, and over this period, the proportional use rates remained relatively constant. These considerations indicated the three series could be combined for the entire three-year period for purposes of correlation analysis. Such a procedure would also provide a more valid explanatory variable series, plus a larger number of observations.

An analysis of the scrap recovery account revealed that scrap proceeds from materials X and Y had been lumped together and that it was not always possible to determine what proceeds had been derived from which materials. Therefore, for a second reason it was necessary for correlation purposes to combine all scrap receipts. The analysis also revealed that because regular pickups of the scrap material were not made by the buyer, no clear cut-off point for the scrap proceeds appeared. In reality, it seemed that for most of the accounting periods the cut-off, or matching of material usage to scrap proceeds, was reasonably accurate, but for some it was not. Interviews established that the company tried to sell the scrap in the month created, but this was not always accomplished. However, there was a clean cut-off as of the end of December of each year.

Examination of the scrap recovery series and the dollars of material usage of X, Y, and Z indicated that it would be necessary to aggregate the months. The necessity for aggregation obviously stemmed from the data inadequacies previously mentioned. After reviewing the behavior of the two series, it appeared the following scheme of grouping would be the most appropriate:

January and February

March, April, and May

June, July, and August

September, October, November, and December.

The period covered in the analysis is March, 1964, through December, 1966. After grouping the monthly data according to the foregoing scheme, there were only eleven data values. However, this is a sufficient number of observations for purposes of correlation.

The first correlation computational efforts employed an equation form of:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3,$$

where

X_1 = Scrap Recovery

X_2 = Time

X_3 = Dollars of material usage of X, Y, and Z.

This equation form provided a $R^2_{1.23} = .9320$, which is quite significant, particularly in view of the small number of observations. The $S_{y1.23}$ = \$2,129.29, which is satisfactory.

The next step was to improve the efficiency of the estimating equation. To this end different combinations of the X_2 and X_3 terms, using various exponents, were employed in a number of iterations. Of the equation calculated the highest coefficient of multiple determination derived was $R^2_{1.2^2 2 3^2 3} = .9697$. But for this equation the regression coefficients of X_2^2, X_2, and X_3^2 were not statistically significant at the .05 level. However, the X_2^2 term was almost significant. Therefore, an additional iteration was carried out employing only an X_2^2 and an X_3 term. This equation appears as follows:

$$X_{c1.2^2 3} = -\$1,218.20 + \underset{(26.13)^2}{\$65.638X_2^2} + \underset{(.002965)^3}{\$.0126507X_3} \qquad \text{(Eq. 28)}$$

For estimating equation (28) the $R^2_{1.2^2 3}$ = .9422 and the $S_{y1.2^2 3}$ = \$1,919.57. The coefficient of multiple determination is significant at the .05 level and the $S_{y1.2^2 3}$ = \$1,919.57 is quite satisfactory. The $r^2_{12^2.3}$ and the $r^2_{13.2^2}$ are both statistically significant at the .05 level. The standard errors for the respective regression coefficients are satisfactorily low, and both regression coefficients are significantly greater than zero at the .05 level.

Curves representing the $X_1 - X_{c1}$ series are plotted on Chart 7.6. Of the eleven $X_1 - X_{c1}$ differences, 4 are above the X_1 curve and 7 are below. Although this is not an ideal dispersion of the differences, it is satisfactory. After all, the best dispersion attainable would have been a 5 to 6 pattern. In terms of magnitude, the dispersion of the differences appears satisfactory. Of the eleven differences, 8 occur within a range of $\pm$ 1.00$S_{y1.2^2 3}$ and the remaining three are well within the $\pm$ 1.96$S_{y1.2^2 3}$ range. In view of the foregoing, it seems estimating equation (28) should be accepted as a model for Scrap Recovery—86.

<u>Estimating Equation - Utilities Expense—61</u>

Charges are entered to this account for electrical power, lights, gas, and water. The company uses a considerable amount of electricity and gas in the production process but hardly any water. Unfortunately this account also includes utilities charges used by the administrative sectors. However, such charges should represent only a small proportion of the total expense, and in addition, they should be relatively fixed. Therefore, it is not a significant consideration.

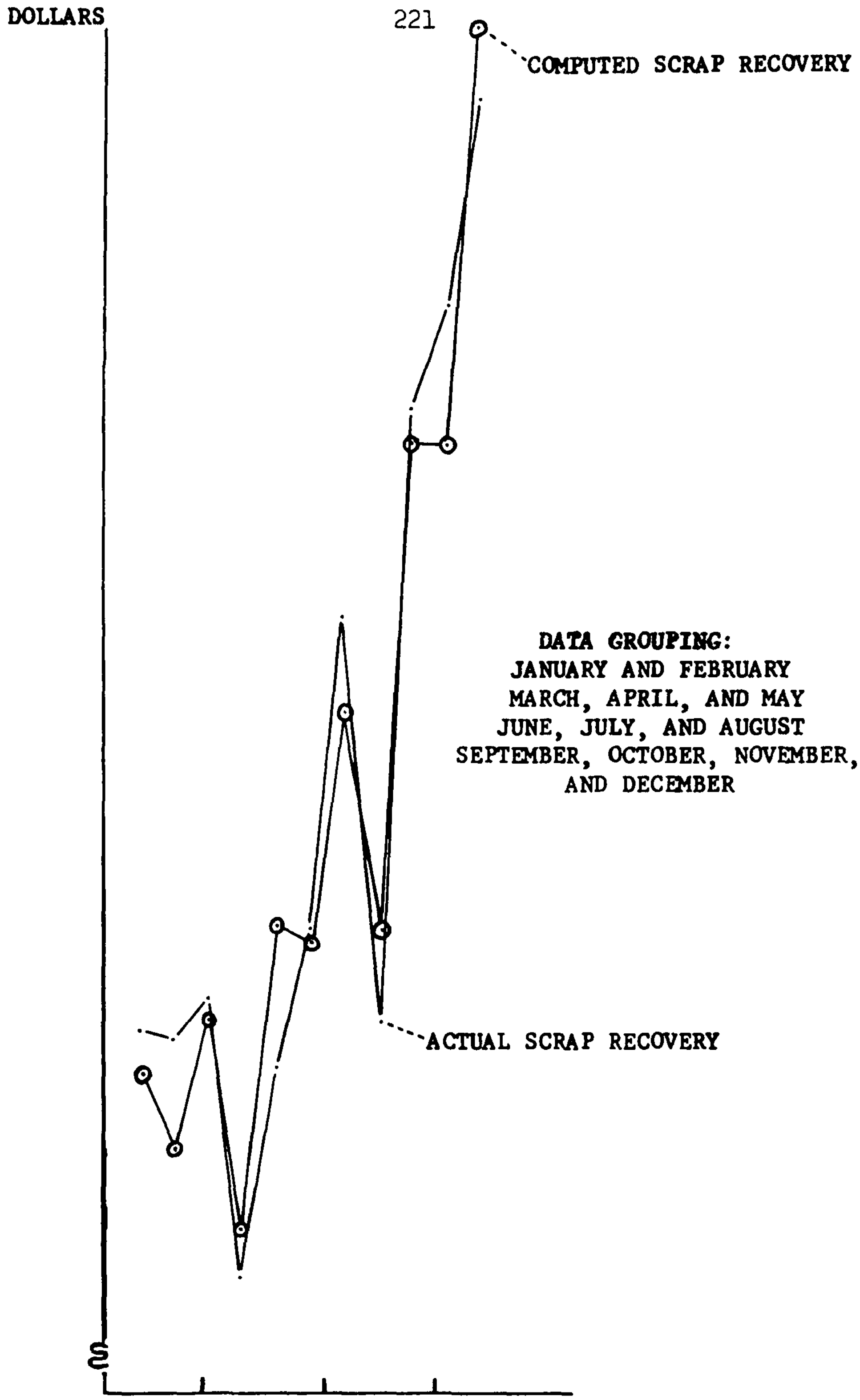

Chart 7.6. Actual Scrap Recovery and Computed Scrap Recovery of the Manufacturing Service Department Costs, March 1964 - December 1966.

An examination of the utilities account revealed that monthly charges are posted from a voucher register in total. Because of the manner in which utility companies bill, it was necessary to examine the original invoices. This examination revealed that the charges for gas had been recorded as expense in the month following actual consumption. The same procedure also occurred occasionally with the electricity charges. Because of these errors it was necessary to restate completely the monthly charges. A second difficulty encountered, and one that was not amenable to adjustment, related to the cut-off dates on the meter readings. For example, the electricity readings were made on the third or fourth of each month, the gas, on the 25th or 26th, and the water, on the 30th or 31st. These dates did not necessarily correlate with the company's accounting period. The discrepancies introduced by this factor remain in the data, as no valid method of adjustment was determinable.

The utilities expenses, after being recast, appear on Chart 7.7. Note that only a single series is plotted to represent all three expenses. This was done in order to maintain conformity between the estimating equation and the company's reporting system. The series demonstrates reasonable variability and a positive trend. Peaks appear in the series during June and July, indicating a responsiveness to production levels, and in December and February, indicating a responsiveness to climate conditions.

Each of the individual component series demonstrates a different type of variability. The electricity series reaches its peak in June of each year and does contain significant variability. In contrast, the peak month for gas consumption was February; however, June was also a peak month. The

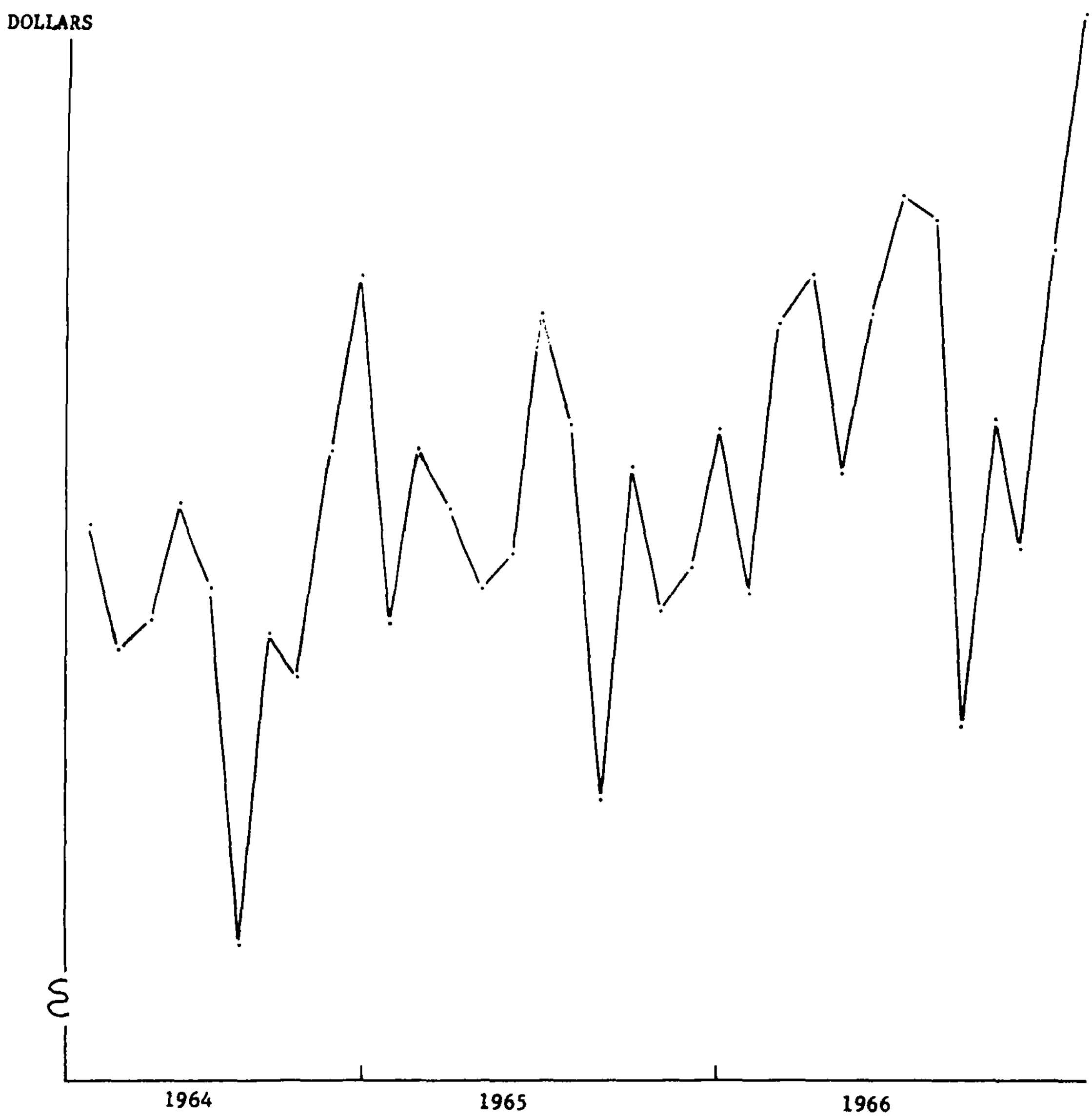

Chart 7.7. Utilities Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

series representing water consumption was approximately stable, being somewhat proportionally higher during the summer and for those months containing the greatest number of days. In general, the utilities series represents a conglomerate of expenses, generated by a number of factors.

In seeking appropriate explanatory variables it was necessary to consider the nature of the individual utility expense components. Because considerable electricity is used in production, it seems logical that the dollars-of-material-usage series would be a possible explanatory series. Although the direct-labor-hours series might be appropriate, it probably would be less satisfactory because of the changing proportions of labor and capital employed in production. Gas consumption, because of the manner in which gas is used in production, should also be related to the dollars of material usage. However, gas consumption should also be related to climate conditions. Therefore, a seasonal index assigning various weights to the individual months based on the consumption of gas for heating purposes was developed. This index was prepared by assuming no gas was used for heating during August and that all gas consumption was associated entirely with production. Although this assumption may not have provided an absolutely valid basis for such months as December, January, and February, it does provide a reasonable point of departure.

In view of the foregoing, the first computational efforts employed the following variables:

X_1 = Utilities Expenses—61

X_2 = Time

X_3 = Dollars of material usage,

and the equation form of:

$$X_{c1.23} = a_{1.23} + b_{12.3}X_2 + b_{13.2}X_3.$$

The first iteration provided a $R^2_{1.23}$ = .6908 and a $S_{y1.23}$ = \$286.83, which is quite satisfactory. It now seems that both a logical and statistically significant relationship exists between X_1 and X_3. The relatively small value obtained for the standard error of the estimate is, to a large degree, attributable to the relative variation in the independent variable series. But its size is also a function of the fit between the two series.

The usual procedure was next employed to determine the most efficient equation form. Based on the numerous iterations carried out, it appears the relationship between X_1 and X_3 is approximately linear, only after eliminating the effects of time which required an X_3^2, X_2^2, and X_2 term. There appeared to be a slight need for a second negative X_3 term with a fractional exponent greater than one but less than two. This does fit logically, in that as electricity consumption increases, the rate per K.W.H. decreases above a predetermined level. Inclusion of such a term might improve the fit, but probably not significantly.

The next step was to add an X_4 term representing the seasonal index previously discussed. The first iteration employed a linear equation form to maintain consistency in methodology. This first iteration provided a $R^2_{1.234}$ = .7561, representing a 6.53 per cent increase over the $R^2_{1.23}$ = .6908 previously computed. This 6.53 per cent increase is significant, although it is not so large as was hoped. The $S_{y1.234}$ = \$257.36 is somewhat less than the $S_{y1.23}$ = \$286.83. It will obviously be difficult to further decrease the standard error of the estimate value because of its present minimum size.

To improve the efficiency of the estimating equation and concurrently to test for multicollinearity, iterations were made employing different combinations of the four variables. In the final analysis, and as was expected, the following equation appears the most efficient.

$$X_{c1.2^3 2^2 234} = \$3{,}149.35 + \underset{(.05483)}{\$.1752695X_2^3} - \underset{(2.9582)}{\$9.655598X_2^2} + \underset{(44.659)}{\$149.0858X_2}$$

$$+ \underset{(.00012602)}{\$.001014126X_3} + \underset{(5.78919)}{\$18.03887X_4}. \qquad \text{(Eq. 29)}$$

For equation (29) the $R^2_{1.2^3 2^2 234} = .8234$, which seems to be the highest attainable value in view of the existing data constraints. The $S_{y1.2^3 2^2 234}$ = \$226.67, which is excellent relative to the absolute values of the monthly observations. Variance associated with the regression coefficients was satisfactory. Tests of significance on the individual coefficients of partial determination for each term revealed each was significant at the .05 probability level. Therefore, estimating equation (29) seems to provide a valid statistical description of the utilities expenses.

It was previously mentioned that the direct-labor-hours series might also be a suitable explanatory variable. To test this possibility, several iterations were completed using the direct-labor-hours series. Because of past experience it is known that multicollinearity exists between the dollars-of-material-usage and the direct-labor-hours series, so that they were not employed together. In any case, the results of these iterations verified that the direct-labor-hours series was not so significant an explanatory variable as is the dollars-of-material-usage series. That is, in each equation case which has been previously discussed in this section, the resulting coefficient of multiple correlation was consistently 6.0 to 8.0 per cent lower, and the associated standard error of the estimates was higher. However, the regression coefficient value for the seasonal index did not change significantly in any of the iterations. Therefore, it appears equation (29) is the most efficient of those computed. For clarification purposes the variables employed in estimating equation (29) are shown below.

X_1 = Utilities Expenses--61

X_2 = Time

X_3 = Dollars of material usage

X_4 = Seasonal index

Chart 7.8 contains the X_1 and the X_{c1} series. Between the two series are a few disparities, but in terms of dollar magnitude, they may be attributable to chance. Of the 34 $X_1 - X_{c1}$ differences, 17 are above the X_1 curve and 17 are below. For the most recent twelve-month period, 7 of the $X_1 - X_{c1}$ differences are above the X_1 curve and 5 are below. In terms of magnitude, 24, or 70.5 per cent, of the differences occur within a range of $\pm$ 1.00$S_{y1.2^3 2^2 234}$, and none of the differences are greater than $\pm$ 1.96 $S_{y1.2^3 2^2 234}$. In view of the foregoing, estimating equation (29) is acceptable as a model of Utilities Expenses--61.

<u>Overtime Premium--58</u>

Charges are made to this account for the premium portion of all payments for overtime hours worked by all hourly wage workers throughout the

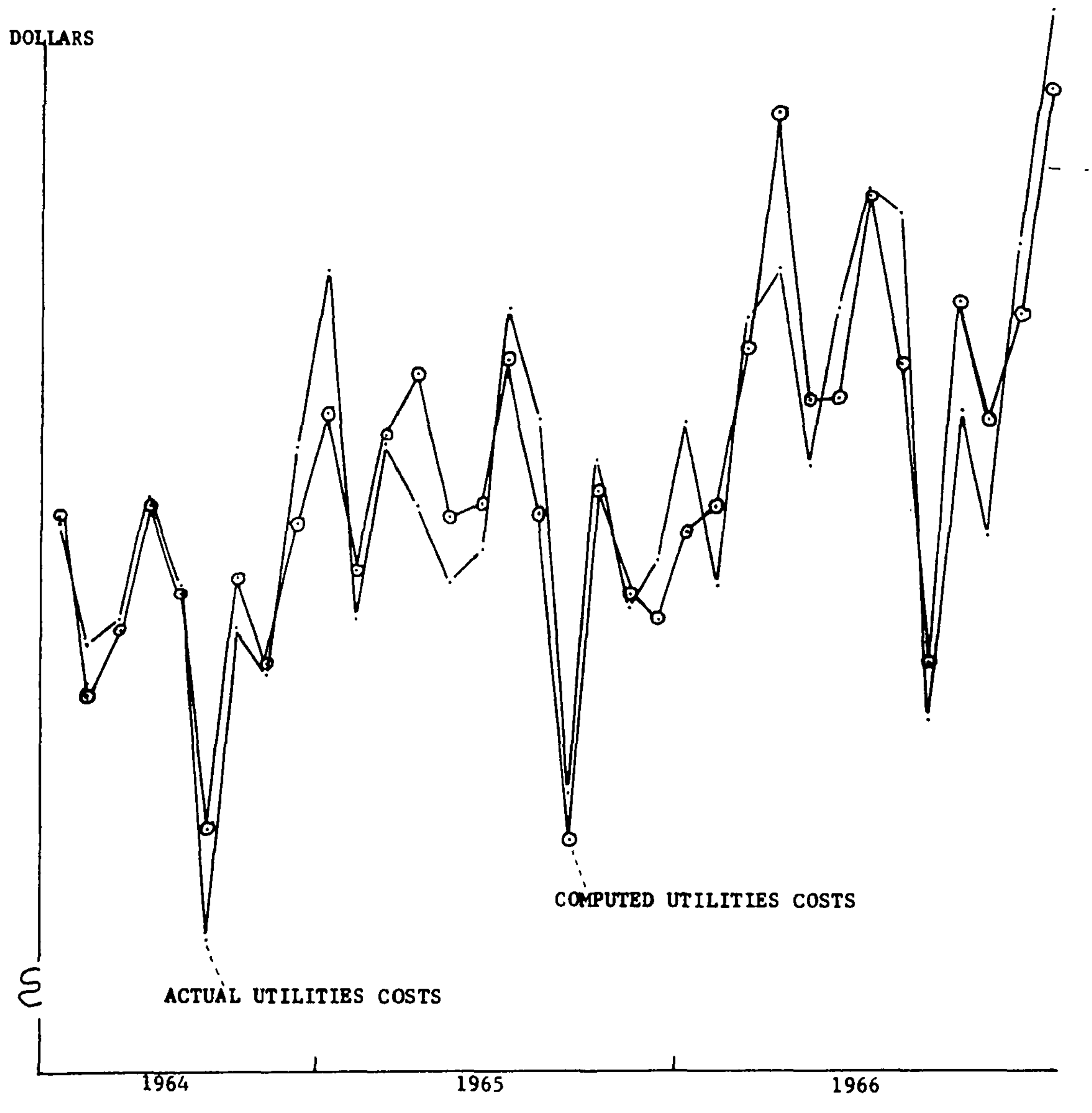

Chart 7.8. Actual Utilities Costs and Computed Utilities Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

factory. Supplemental payments made to supervisory, professional, and clerical personnel are not included in this account. Because the charges to this account are for all factory workers, the dependent series in a sense, contains heterogeneous data. After reviewing the weekly and monthly reports reflecting transactions in this account, there appears to be no practical way of determining the amount of premium applicable to direct labor as opposed to indirect labor. A second factor contributing to the lack of homogeneity was the segregation of the night-shift premium commencing in August, 1965. Because of the records maintained by the company, it was not practical to adjust the overtime premium account for this change for the months preceding August, 1965. Therefore, the two series were combined.

In developing an estimating equation for the revised overtime and night-shift premium account it seemed a basic and easily determinable explanatory variable should be employed. The obvious candidate would be a labor hours series. In terms of data, the company reports direct-labor-piece-work hours and a total labor hours series. A day-work-labor hours series developed from source records by this writer, when combined with the direct-labor-piece-work hours, provides a total direct labor hours series. At first glance, the most logical series would be the total labor hours series, since the dependent variable contains all overtime premiums. However, the total labor hours series would not be easily quantifiable for future periods, whereas the total direct-labor-hours series could be computed using forecasted production. Therefore, it was decided that the latter series should be employed as an explanatory variable. This necessitated the assumption that overtime hours would occur in the service departments at approximately the same time as they occurred in the production sectors. The validity of this assumption could not be verified by examining the raw data.

A second explanatory variable selected was merely the number of hourly workers within the factory. Computations were then carried out using three variables: total direct labor hours, number of employees, and a time variable. Monthly data were used, which is in conformity with previous account analyses and covered the period of March, 1964, through December, 1966. These calculations did not provide any significant coefficients of correlation, nor were the standard errors of the estimates within acceptable ranges.

Subsequent interviews with the concerned budgeting personnel revealed that the company had not developed any budgeted figures for overtime and night-shift premiums. Two basic reasons justified such a policy. First, management felt that to budget overtime premium would be, in a sense, implying that overtime should be worked. Management had supported this contention to the point of not permitting any overtime work in the production departments during January, February, or March of 1965. However, by April of 1965 such a back-log of orders was on hand that overtime was forced, and during subsequent months the cost of overtime premiums ballooned. The second reason that overtime was not reflected in the annual budget was that the budgeting personnel had not been able to develop an accurate method of preparing estimates. The nature of the costs itself, coupled with the departmentalization of the production departments and the character of the production process, had effectively fouled all efforts of the company's personnel to prepare accurate forecasts. Therefore, this writer was not able to obtain any assistance on the matter from the company's budgeting personnel. It was obvious that a new approach had to be taken.

Since overtime premium is incurred on a weekly basis and not on a monthly basis, it was decided that the data should be restated by weeks. Based on weekly reports generated by the company's accounting system, the overtime and night-shift premium account was restated for the period January, 1965, through December, 1966. No data were developed for 1964, as the years 1965 and 1966 provided a total of 104 weekly observations, a more than sufficient number. The month of August was also deleted from the series on the grounds that night-shift premium is minimal during this period and that overtime premium paid is not a function of production but rather a function of maintenance work. Exclusion of August reduced the number of observations to 96. By employing the two years 1965 and 1966, and by excluding the month of August in each year, it is felt the remaining data would be as homogeneous as is possible under the circumstances. Chart 7.9 contains the overtime and night-shift premium series for all three years. Examination of the plotted series, shown by monthly periods, readily reflects the similarity between the actual costs for the two years, 1965 and 1966.

The next step was to convert the direct-labor-hours series to a weekly basis. A second change made at this time involved restating the total direct-labor-hours series to an average-hours-worked-per-week-per-employee series. The purpose here was to determine at what point overtime premium costs seemed to increase.

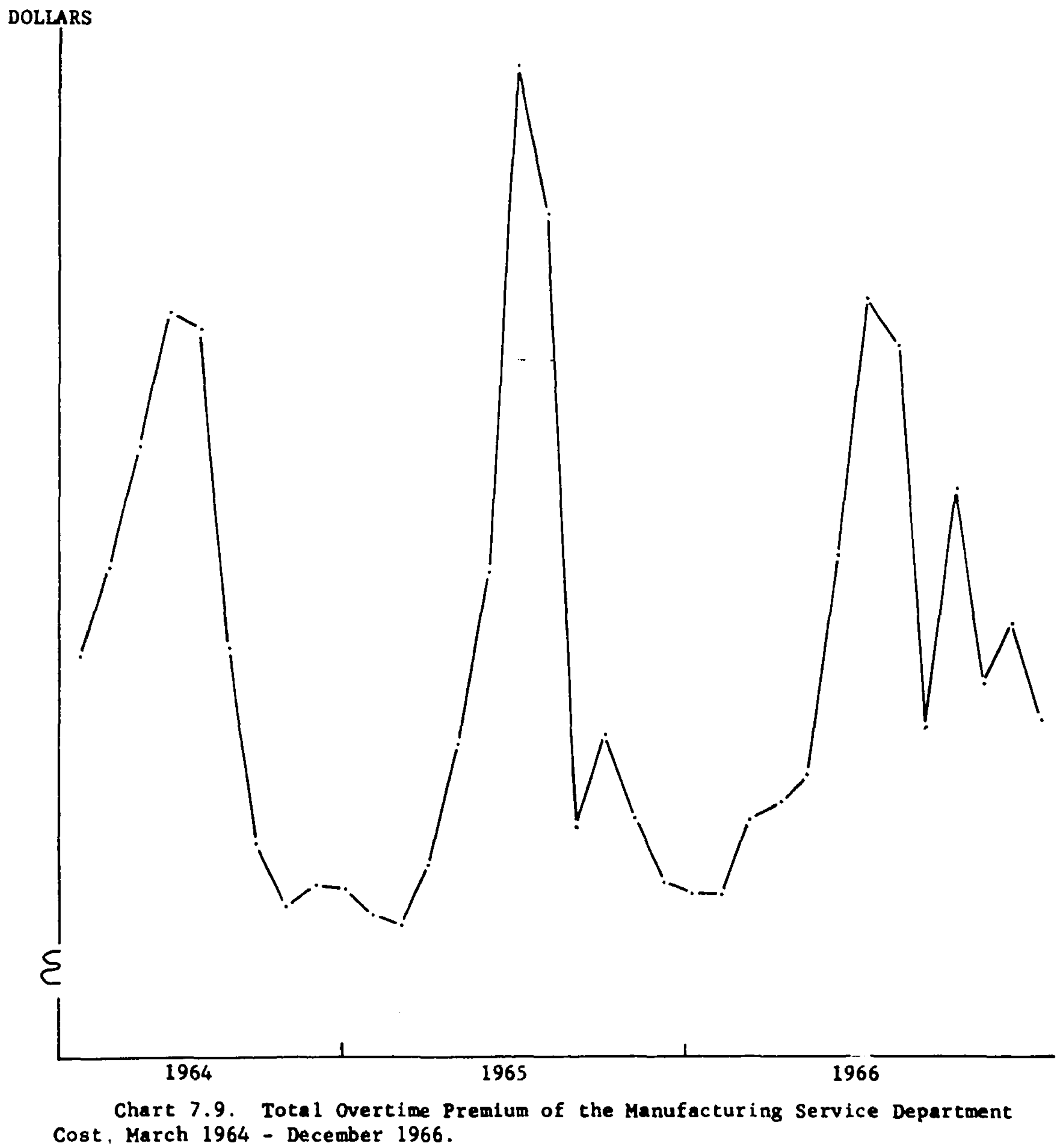

Chart 7.9. Total Overtime Premium of the Manufacturing Service Department Cost, March 1964 - December 1966.

It was decided that the number of factory employees series should be retained as an explanatory variable. This series was, therefore, restated in weekly terms. However, the exact accuracy of this series is somewhat questionable because it is prepared as of the last working day of each week. Since there are employees that are terminated and are added during the week, it is necessary to assume that the effect of terminations and additions of employees off-set one another, both in terms of numbers and hours worked. This assumption could not be verified by a check of the personnel reports because these records reflect only monthly turnover figures. However, the relatively large number of observations employed lends validity to the assumption.

The first group of iterations employed the following variables.

X_1 = Overtime and Night-Shift Premium--58

X_2 = Time

X_3 = Average hours worked per employee

X_4 = Number of employees

The first calculations, using a linear equation form, provided a $R^2_{1.234}$ = .5667, which at the .05 level is significantly greater than zero. The $S_{y1.234}$ = \$1,224.80, which is too large relative to the lower observed weekly values, but satisfactory relative to the higher observed weekly values.

A series of iterations was then carried out employing various combinations of exponents with the three explanatory variables. The purpose of these iterations was to determine the most efficient equation form and, concurrently, to test for significant multicollinearity. But these iterations indicated that significant multicollinearity did not exist between the independent variables. As for the most efficient equation form, it appeared the X_2 variable should be supplemented by a squared term. Inclusion of an X_2^3 term and higher powers did not improve the fit. The same behavior pattern applied to both the X_3 and X_4 terms. Of the estimating equations computed, the following equation appeared to be the most efficient.

$$X_{c1.2^2 2 3^2 3 4^2 4} = +\$101,051.96 + \underset{(.19768)}{\$.494824X_2^2} - \underset{(21.149)}{\$53.0315X_2} + \underset{(5.0539)}{\$25.83155X_3^2}$$

$$- \underset{(358.90)}{\$1,480.7650X_3} + \underset{(.23459)}{\$.774377X_4^2} - \underset{(160.994)}{\$494.1750X_4}. \quad \text{(Eq. 30)}$$

For equation (30) the $R^2_{1.2^2 2 3^2 3 4^2 4} = .7410$, which represents an increase of 17.34 per cent over that of the linear case. The individual coefficients of partial determination for each term were all significant at the .05 probability level. Further, the decrease in the standard error of the estimate over that previously obtained was excellent. That is, the $S_{y1.2^2 2 3^2 3 4^2 4} =$ \$963.84, but it was still much higher than was desired.

Although the individual statistical measures of accuracy, i.e., the coefficient of multiple determination and the individual coefficients of partial determination, were well within acceptable ranges, the standard error of the estimate was too high. The cause of this condition is not readily apparent; however, its existence was anticipated by this writer based on a thorough examination of the raw data prior to commencing the correlation computations.

The magnitude of the standard error of the estimate may be attributed to three fundamental factors. First, the amplitude of the fluctuations in the dependent variable series is quite extreme in relative terms. There is no comparable variation in either the X_3 or X_4 variable series. The independent variable series, by their very nature, are limited in the amplitude of their fluctuations. When such a condition exists, the standard error of the estimate may be high even though the most efficient equation form has been determined. Second, the relationship between the dependent and independent variables appeared to be changing over the period of analysis. That is, the average-number-of-hours-worked-per-week-per-employee seemed to be increasing relative to the overtime premium series. Although the X_2 variable may account for some of this variation, it should not account for a significant proportion. This changing relationship will cause an increase in the standard error of the estimate. The changing relationship between X_1 and X_3 might be attributable to the changing proportion of labor and capital employed, more efficiency in the employment of direct labor personnel, and a reduction in overtime premium paid to the indirect laborers.

A third factor could be that only 74.01 per cent of the variance in X_1 has been explained. Obviously, other explanatory variables might be included in the estimating equation if they were determinable. However, information constraints have limited the number of possible causation variables.

Although equation (30) does provide reasonably high coefficients of partial determination and a satisfactory coefficient of multiple determination, two facets of the equation require further considerations and might assist in solving the variance problem. First is the question relating to its form. The inclusion of the X_2^2 and X_2 terms seems appropriate because a curvilinear time relationship does exist between the X_1 and the X_3 and X_4 variables. The X_3^2 term may be logically rationalized in that as the average hours worked per direct labor worker increases, the overtime premium paid to all workers should increase at a faster rate. However, the X_4^2 term does not appear rational, at least to the same degree. It would seem logical that this term should have a linear relationship to X_1, after the effects of time are eliminated. On the other hand, it must be remembered that $r^2_{14^2.2^2 23^2 34}$ was definitely significant. During the iterations conducted earlier an equation form was employed which was similar to equation (30) except that no X_4^2 term was employed. That equation form provided a $R^2_{1.2^2 23^2 34} = .7083$ and a $S_{y1.2^2 23^2 34} = \$1,015.45$. Yet when an X_4^2 term was added, the R^2 increased only .0318, which is minimal; but the coefficient of partial determination for X_4^2 must be closely related to some variable that is not included in the estimating equation.

The second problem also related to the first, as will be shown subsequently, is the magnitude of the constant term in equation (30). Its value was \$101,051.96, which is absolutely illogical in view of the nature of the independent variable. Such a value for the constant term may be regarded as a forced value based on the explanatory variables used and their respective regression coefficient values. In general, it is a meaningless figure in most cases, since it is primarily a function of the equation form and the value of the coefficients. Just as a note of caution, this is the case only in multiple correlation, not simple correlation.

In view of this, it should be noted that when only an X_4 linear term was employed, and where the $R^2_{1.2^2 23^2 34} = .7083$, the constant term had a value of \$8,988.34. This is far different from the \$101,051.96 derived when an X_4^2 and an X_4 term were both employed. Therefore, in view of the foregoing, it was concluded that in this one case the estimating equation should be forced through the point of origin. This procedure merely eliminates the pure constant term in the estimating equation and, concurrently, will proportionately adjust the regression coefficient values. It should also assist in eliminating the effects of variables that have not been included in the estimating equation, and particularly, their effects on X_4. It is hoped that this procedure will also result in a material reduction in the value of the standard error of the estimate.

Two iterations were carried out wherein the estimating equation was calculated through the point of origin. The first iteration employed the same equation form as equation (30). This resulted in a $R^2_{1.2^2 2 3^2 3 4^2 4} =$ .9021. In addition, all of the regression coefficients were altered to some extent. The X_2 coefficients altered because of the shift in the curves, the X_3 coefficients decreased proportionately, but the X_4 coefficients were dramatically reduced. Further, the coefficient of partial determination for X_4^2 was no longer significant. These are the results that were expected and as such, added validity to the previous discussions.

The next iteration employed only a linear X_4 term and provided a $R^2_{1.2^2 2 3^2 3 4} = .9003$. The individual coefficients of partial determination were all quite significant. The new $S_{y1.2^2 2 3^2 3 4} = \$1,026.06$, which is still much too large. These statistics are associated with the following estimating equation:

$$X_{c1.2^2 2 3^2 3 4} = +\underset{(.1778)}{\$.866765}X_2^2 - \underset{(18.339)}{\$95.4685}X_2 + \underset{(1.615)}{\$17.6991}X_3^2 - \underset{(104.23)}{\$879.0080}X_3 + \underset{(5.5696)}{\$37.7005}X_4. \qquad \text{(Eq. 31)}$$

What has been accomplished is the development of an estimating equation that appears valid. The values of the explanatory variables appear logical, particularly in view of the assumptions made, and the multicollinearity does not appear to be significant. This equation would be quite appropriate, except that the value of the standard error of the estimate is much too large. Even if the equation did exactly duplicate past experience, this researcher could have no confidence in any forecasts which might be subsequently developed. Therefore, equation (31) will be discarded as invalid. New approaches must be considered.

It may be recalled that the initial data employed was in terms of one-month periods. Restating these data into weekly periods appears to be a valid procedure. The problem, therefore, appears to be in the data's factory-wide aspect. That is, it appears the variables employed might be appropriate if they were segregated both by weeks and by production departments. But this information is not available within the Grace Company's reporting structure. If this information were available, then a satisfactory estimating equation could probably be developed.

A second approach to development of an estimating equation for the overtime and night-shift premium account could be made by developing one or more synthetic variables. Such an approach always exists as a possibility, but because of the inconsistent attitude of management toward this expense, the validity of synthetic variables would be highly questionable.

A third approach would be to examine other series of all types that could be found within the firm. Possibly one or more unrelated series might be employed to develop an excellent estimating equation, in a mechanical sense. But the usefulness of an equation developed in that manner would also be extremely questionable.

A great deal of time and effort has been expended by this writer toward developing a valid estimating equation for the Overtime and Night-Shift Premium--58. At every turn difficulties and information constraints have been encountered. Because of these factors, it has not been possible to develop a valid estimating equation for this account. If additional and much more detailed information had been available, this writer is confident that a satisfactory estimating equation could have been developed.

Analyses of the Category III Accounts

It was previously determined that several types of factory overhead costs exist because other manufacturing costs were first incurred. In the case of the Grace Company, there were three such accounts. There is a question regarding the appropriateness of regression as a tool for predicting future costs of this nature. Its applicability is questionable because these costs are, to a large degree, a function of external factors and have only a secondary relationship to activities within the firm. Before correlation is discarded as an appropriate tool for forecasting such accounts, it seems that the individual accounts should be thoroughly analyzed.

Employee Group Insurance--72

Charges are made to this account for the company's share of the group insurance premium paid on all factory employees participating in the program. The company pays 80 per cent of the total premium regardless of the employee's classification as either an hourly wage worker or as a fixed salary employee.

Examination of the group insurance expense account revealed monthly charges to be exceptionally constant over extended periods of several months, but in other months considerable variation was present. This inconsistent behavior required clarification. Subsequent examination of source documents and withholding patterns revealed that several clerical errors had occurred which, in turn, resulted in the accounts demonstrating considerable variation. Necessary correcting adjustments were made that caused the monthly expense pattern to be much more consistent. The adjusted series is plotted on Chart 7.10. Although the dollar scale is not shown, it is permissible to state that the chart was plotted with the scale of $300 increments and was started at zero. The purpose in giving this information was merely to indicate to the reader the lack of variation in this expense series. On an annual basis, the normal operating budget period, the average monthly deviation was less than 3 per cent of the monthly average expense.

For the years 1964 and 1965, the monthly expense behavior seems to respond to the number of employees in the work force, the number of new employees added each month, and finally, some psychological factor that apparently influenced a few employees to subscribe to the insurance in January of each year. This pattern was interrupted during 1966, when the employee turnover rate increased drastically. For example, during the month of April, 1966, the employee turnover rate increased approximately four times the normal April rate. This was caused by severe competition in the labor market stemming from defense contractor's hiring much larger amounts of labor than in previous years.

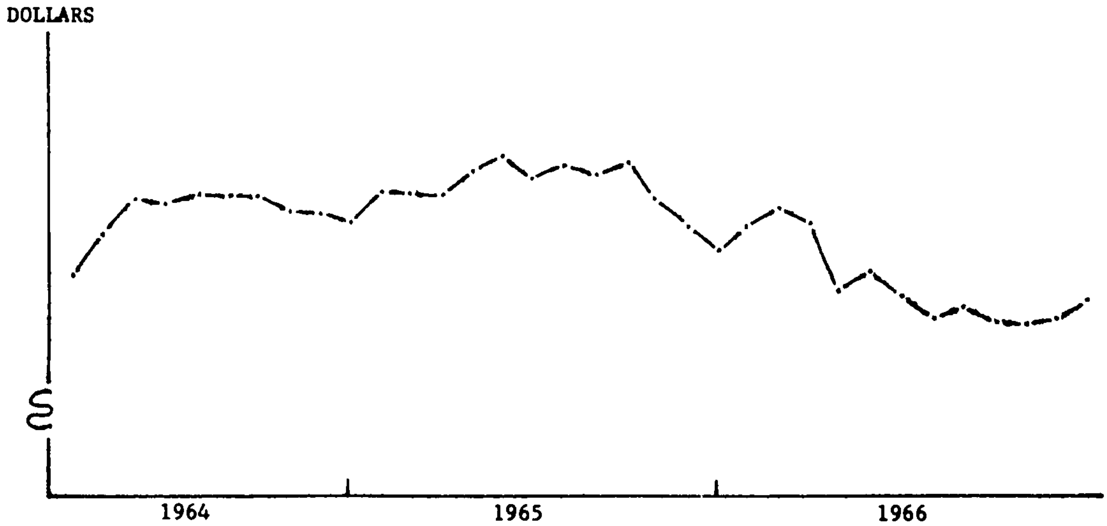

Chart 7.10. Employee Group Insurance Costs of the Manufacturing Service Department Costs, March 1964 - December 1966.

On balance, it appears that correlation analysis is not an appropriate tool for developing forecasts of the employee group insurance expense, in this particular case. This conclusion is predicated on the foregoing factors, summarized as follows:

A. The employee group insurance expense series does not exhibit significant variation.

B. The expense is not a function of quantifiable activities within the firm or of management decisions. Rather, it appears to be a function of employee opinion, the stability of the company's "hard core" labor force, and the demands for labor by other firms competing in the labor market.

The foregoing reasons do not provide a basis for automatically precluding the use of correlation either conceptually or mathematically, but they do raise the question of practicability. It has been previously declared that the primary purpose for developing estimating equations is to assist in the budgeting function. In the case of many programmed costs, correlation has not been employed because it simply was not necessary in order to derive valid budget estimates. The employee group insurance expense does appear to be of the nature of a programmed cost, other things being equal, and as such, use of correlation techniques is not necessary. However, if one elected to use correlation, then the considerations mentioned in (B) would necessarily need to be quantified. To do this would require development of several probability series, development of which is outside the scope of this paper. Therefore, it is felt that the most practical treatment of the Employee Group Insurance--72 would be as a programmed cost. Conventional methodology could no doubt be employed to develop accurate annual estimates for budget purposes.

Vacation Payments--75

The Grace Company accrues a liability each month for the estimated vacation payments expense. These monthly accruals are debited to the vacation payments account, resulting in an equal monthly charge to expense. As of December 31, each year the accrued and expense accounts are presumably adjusted for previous estimating errors, to the degree possible. Since the adjustment to the expense account is a function of the accrual adjustment, it is appropriate to examine the way the latter adjustment is determined. Interviews revealed that on December 31, the prior year's actual payments are analyzed. An estimate is made of the average number of employees for the next year and is then computed using proportions. After the estimated expense for the coming year is determined, it is divided by one-half, and the figure thus derived is established as the year end accrued liability. Based on these procedures, it would seem that the amount shown as expense might not be particularly accurate for any one year. However, examination of the amounts actually paid revealed that the error never exceeded 8 per cent in any of the three years and that in 1966 the error was less than 2 per cent.

As vacation payments are made, they are debited to the accrued liability account. These entries are made only to the workers paid on an hourly basis. Vacation payments to monthly salaried employees are not segregated as such nor is any provision made for such payments. That is, vacation payments made to salaried personnel are retained in the respective salaries expense accounts.

In addition to the vacation payments to hourly workers, the account is also charged for holiday wage payments to hourly workers as they are incurred. No provisions are made in the accrual account for this item. Holiday pay is given only to those workers employed on the date of the respective holidays. In general, eight paid holidays are recognized each year.

Vacation payments are made almost entirely during the month of August. This is a result of the company's policy of stopping production for two weeks during August. All disbursements are made as of this time except for a few of the maintenance personnel that are needed during this two-weeks' period. Workers that are not entitled to a full two-weeks' paid vacation, because of inadequate employment longevity, are used for maintenance work during the period when production is stopped.

In order to utilize correlation for estimating the annual expense for vacation payments, it would be necessary to develop considerable information that is not readily determinable from existing company records. The necessary information would be similar to the following:

A. Average number of employees each month having X number of years of service, by groups.

B. Average hourly wage rate for each group in (A) above.

C. A probability factor or turnover rate for each group of employees indicating the likelihood of employment at the start of the annual vacation period.

The foregoing information would provide an excellent basis for developing an estimating equation of considerable accuracy. Yet, to develop such data from present records would be an extremely onerous task. Even without the foregoing information, correlation might be appropriate under the present system if eight or more years of data were available. But a considerable risk would be involved if this approach were employed, because of the high likelihood of heterogeneous data. Therefore, correlation analysis will not be employed for estimating the vacation payments.

In view of the company's procedures for handling vacation and holiday pay, it appears that correlation is not a particularly relevant estimating technique for the vacation payments account, in total. But if the payments for vacation pay were segregated from payments for holidays, the latter item might be amenable to correlation techniques.

Employment of correlation for estimating the annual expense for holiday payments presents mechanical problems entirely different from those encountered in estimating the annual vacation payments. To estimate this expense, one would merely need to know: (a) the number of holidays in the coming year, (b) the estimated number of employees on each date, and (c) the average straight-line hourly rate. No problems as to the length of employment service or the likelihood of termination before a certain date, as was the case in the vacation payments estimate, would appear. The problems involved in securing the above information are not particularly extensive. For example, the number of employees would be reasonably determinable based on production forecasts, and the average hourly rate could be determined approximately by reviewing past experience, with appropriate adjustments for expected future changes that will occur in the rate scale. This is the same information that would be needed for conventional methods. The point is that this expense could probably be estimated just as accurately with conventional methods as with correlation. In fact, conventional estimates might be more valid because of the strictly linear nature of the expense and the possible existence of a pure constant in a previously determined estimating equation. An estimating equation, based on correlation, will not be employed in this case since the lack for its need in effect implies impracticability.

Payroll Taxes--63

Charges are made to this account for the employee's share of the F.I.C.A., and for the state and federal unemployment compensation premiums paid on all factory employees. This account obviously represents a classic example of an expense that is incurred and determined primarily by external factors. Conventionally, budget estimates of this expense are prepared in annual terms and are expressed as a percentage of the expected annual payroll. No effort is normally made to predict or isolate this expense by months or quarters because of the nature of the charges. The estimating problems that exist using conventional techniques continue to exist and perhaps become even more complex, when correlation is employed. The general question which must now be posed is whether or not the use of correlation would significantly improve the accuracy of budget estimates for this type of expense. This question may only be answered by examining possible procedures and problems in estimating this expense.

The conventional approach would be to determine the percentage of payroll tax expense in previous years to the gross payroll. This percentage would then be adjusted proportionally to reflect known future changes in either the tax rate or base. The next step would be to determine the expected gross payroll. These computations could be based on labor standards for planned production, to determine the direct labor costs, and on the estimating equations for indirect labor, etc., assuming correlation was used for these items. The percentage derived would then be applied and the annual budget estimate computed. This is obviously one of the simplest methods possible, but it is reasonably accurate and widely employed by industry.

At first glance, the foregoing methodology appears to have three major deficiencies. First, it appears that employee turnover is not considered, although it is a definite factor in determining the annual expense total. However, such is not the case because the arithmetic average does implicity reflect turnover even though this fact is not readily obvious, but it does not reflect future turnover rates. Correlation, per se, would not significantly assist in this regard since the turnover is a function of many factors, some of which are external to the firm. Abnormal and extreme increases in the turnover rate, similar to those that occurred within the Grace Company during 1966, would render a correlation estimating equation invalid. Such would be the case even though a variable representing the prior year's monthly turnover rate had been included in the estimating equation.

A second apparent shortcoming is that the total annual expense figure derived is an absolute dollar amount and conventional estimating procedures do not provide a probability range within which this total figure may fluctuate. An estimate derived using correlation would provide such a range through its related statistic, the standard error of the estimate. However, because of the exogenous stochastic factors affecting the turnover rate it is possible that the range of error would be too narrow to be technically valid for planning needs.

Finally, the timing of expense incurrence, by months or quarters, is not reflected by the conventional procedures. This deficiency could be relevant for planning cash flows or for comparative monthly budgets. Correlation could be useful in establishing patterns for the timing of expense incurrence. But refinement of the conventional data could provide similar estimates, though not so statistically valid. Here again, the externally stochastic nature of the turnover rate impugns the validity of both correlation estimates and conventional estimates.

In summary, it appears that correlation may offer additional and more valid estimates of payroll tax expense if, and only if, employment is stable. The existence of stability in the turnover rate is critical. If stability does not exist, correlation does not provide budget estimates significantly more valid than conventional procedures, as related to this particular account. Further, the need for monthly or quarterly expense patterns is in this case, questionable. This latter consideration alone implies that correlation techniques may not be practical and that a conventional tool which is much less complex is probably appropriate.

Summary of the Manufacturing Service Department Costs Analyses

In the preceding sections, it was determined that the four-manufacturing-service-department-reporting structure employed by the Grace Company was inappropriate for correlation analysis. Therefore, these four departments were separated into a series of 36 accounts as reported by a secondary system within the Grace Company. This step permitted an analysis of each account and thus provided a much more valid basis for developing individual estimating equations. Each of the 36 accounts was then superficially examined to determine the general nature and type of charges each contained. This broad review provided a basis for their initial classification into one of several groups.

The first group classification was that of fixed costs. The fixed costs classification was further segregated into a committed cost group and a programmed cost group. Committed costs were defined as costs of a sunk nature over which the operating management had little or no control in the short-run. Four accounts were placed in this group: Machine Rental--60, and depreciation of, Cars--76, Machinery--78, and Buildings--80.

Programmed costs were defined as costs of a fixed nature over which the operating management had a limited measure of control in the short-run, but which were not particularly responsive to variations in the volume of output, etc. The programmed costs were further subdivided into two groups. The first group contained costs over which management has the least control and included: Insurance--81 and 82, Property Taxes--84, and Employee Pension Costs--71. The second subdivision of the programmed cost group included accounts, the balance of which consisted, for the most part, of salaries and a minimum of supplies, etc. These accounts were the Indirect Supervision--62, Purchasing Department Expenses--67, Planning Department Expenses--66, Factory Engineering Expenses--65, Time Study Department Expenses--63, Personnel Department Expenses--68, and Plant Security Salaries--83. In regard to these expenses, it was noted that each account demonstrated insignificant variability and that, in general, each had a slightly positive trend. In this case correlation techniques, per se, offer no significant assistance in budgeting such expense items. This might not be the case in other firms, depending on the nature and behavioral patterns of the individual accounts.

A second group classification was employed to deal with several accounts that were not amenable to correlation analysis or for which correlation analysis was not practical. This group was further subdivided into three segments, although the segments were not entirely unrelated. The first two segments were comprised of accounts that were exempted from correlation analysis because of their stochastic nature, their relative lack of materiality or on both grounds. Accounts included in these segments were: Time Lost from Injuries--57, Safety and First Aid--70, Repairs to Company Cars--77, Telephone and Telegraph Expense--73, and General Factory Travel Expense--74.

The third segment, non-operating costs, contains only one account, Building Repairs--79. Non-operating costs were excluded from correlation analysis on the basis of the nature of the causation factors. The factors are not related, to a determinable extent, to production or other activity levels. Therefore, correlation techniques cannot be meaningfully employed.

The third group classification was variable costs. Accounts were placed in this group based strictly on a superficial review of their individual behavior in terms of variability, magnitude, and presumed causation factors. This third group was further subdivided into three broad categories. The first category contained those accounts which appeared to be most closely related to production activities. Presumably the costs in Category I should have been functions of one or more of the primary independent variables, such as the direct-labor-hours, or dollars-of-material-usage series. Accounts placed in this category were: Trucking--53, Inspection--54, Rework--55, Time-Study Allowances--56, and Stockroom Expenses--69.

The Category II costs were defined as those costs which were responsive to secondary independent explanatory variables, derivatives of the primary explanatory series, or to some type of related synthetic explanatory variable series. Accounts placed in this classification were: Machinery Repair--51, Indirect Supplies--52, Utilities Expense--61, Scrap Recovery --86, Direct Supervision--50, Overtime Premium--58, and Night-Shift Premium--59.

All the accounts contained in both Categories I and II were subjected to correlation analysis, except for the Direct Supervision--50 account. An additional change was made in the classifications when the Overtime Premium--58 and the Night-Shift Premium--59 accounts were combined for purposes of correlation analysis.

The third category within the variable costs classification contains those cost accounts incurred only because other direct or overhead costs had been previously incurred. These same accounts are also dependent upon external or exogenous factors. Accounts placed in this category were: Employee Group Insurance--72, Vacation Payments--75, and Payroll Taxes--63. None of these accounts was subjected to correlation analysis.

It should be noted that it was not necessary or practical to conduct correlation analyses of all the accounts within the manufacturing service department sector. Similar circumstances would probably prevail in other companies. A summary of the estimating equations developed for the several accounts analyzed follows:

Trucking Expenses--53

$$X_{c1.2^3 2^2 34} = +\$552.67 + \$.32542X_2^3 - \$11.6965X_2^2 + \$.0049641X_3 + \$585.289X_4 \quad \text{(Eq. 19)}$$

$$R^2_{1.2^3 2^2 34} = .7807$$

$$S_{y1.2^3 2^2 34} = \$1,393.14$$

Inspection Expenses--54

$$X_{c1.23^3 3^2 3} = -\$4,179.58 - \$103.49X_2 + \$.0000000001133X_3^3 - \$.00001544X_3^2 + \$.7363X_3 \quad \text{(Eq. 20)}$$

$$R^2_{1.23^3 3^2 3} = .8013$$

$$S_{y1.23^3 3^2 3} = \$708.27$$

Time-Study Allowances--56

$$X_{c1.2^2 23^2} = +\$2,385.77 - \$6.9647X_2^2 + \$217.873X_2 + \$.0000002331616X_3^2 \quad \text{(Eq. 22.2)}$$

$$R^2_{1.2^2 23^2} = .9125$$

$$S_{y1.2^2 23^2} = \$715.58$$

Stockroom Expenses--69

$$X_{c1.2^3 2^2 234} = -\$582.85 + \$.2510797X_2^3 - \$15.6343X_2^2 + \$261.725X_2 + \$.0163458X_3 + \$273.577X_4 \quad \text{(Eq. 24.1)}$$

$$R^2_{1.2^3 2^2 234} = .750$$

$$S_{y1.2^3 2^2 234} = \$464.88$$

Rework Costs--55

$$X_{c1.2^{3.4} 3^2 3} = -\$7,981.66 + \$.45661X_2^{3.4} - \$.000000041093X_3^2 + \$.00782479X_3 \quad \text{(Eq. 25.2)}$$

$$R^2_{1.2^{3.4} 3^2 3} = .9979$$

$$S_{y1.2^{3.4} 3^2 3} = \$340.34$$

Machinery Repair Expense--51

$$X_{c1.2^3 2^2 23} = +\$679,545.91 - \$.6627857X_2^3 + \$204.60596X_2^2 - \$20,619.497X_2 + \$1,491.2635X_3 \quad \text{(Eq. 26)}$$

$$R^2_{1.2^3 2^2 23} = .8855$$

$$S_{y1.2^3 2^2 23} = \$1,660.14$$

Indirect Supplies--52

$$X_{c1.23^{1.4} 3} = +\$13,699.24 + \$89.3544X_2 + \$.012017X_3^{1.4} - \$.93773X_3 \quad \text{(Eq. 27.2)}$$

$$R^2_{1.23^{1.4} 3} = .8461$$

$$S_{y1.23^{1.4} 3} = \$1,616.86$$

Scrap Recovery--86

$$X_{c1.2^2 3} = -\$1,218.20 + \$65.638X_2^2 + \$.0126507X_3 \quad \text{(Eq. 28)}$$

$$R^2_{1.2^2 3} = .9422$$

$$S_{y1.2^2 3} = \$1,919.57$$

Utilities Expense--61

$$X_{c1.2^3 2^2 234} = +\$3,149.35 + \$.1752695X_2^3 - \$9.655598X_2^2 + \$149.0858X_2$$
$$+ \$.001014126X_3 + \$18.03887X_4 \qquad \text{(Eq. 29)}$$

$$R^2_{1.2^3 2^2 234} = .8234$$

$$S_{y1.2^3 2^2 234} = \$226.67$$

CHAPTER VIII

APPLICABILITY OF REGRESSION AND CORRELATION ANALYSIS IN THE BUDGETING FUNCTION

Introduction

Management is generally considered to have three broad functions: planning, organizing, and controlling. Budgets play a vital role in assisting management to fulfill its functions. The budget, a basic mode of communication within the organization, represents a quantification of much of management's plans. Because it communicates to all affected personnel the company's goals and the resources provided to achieve these goals, the budget assists in control. Finally, the budget provides a basis for evaluating performance by comparing the actual results of a period to those that were planned. Without question, the budget has been accepted and utilized by most manufacturing firms as an important quantitative technique.

Although manufacturing firms prepare both financial and operating budgets, considerable variation appears among companies in the amount of effort expended in developing budgets. Often a firm will have only one person who prepares the budget during spare hours. Conversely, other firms may have a complete budgeting department charged with the responsibility of preparing the budgets and continually up-dating current budgets. Regardless of the number of personnel involved or the relative sophistication of the techniques employed, budget personnel are faced with the common problem of developing accurate forecasts. It is with the problem of accuracy that this chapter will deal.

In six of the seven preceding chapters the writer has been concerned with the conceptual and mechanical difficulties involved in actually computing estimating equations for various costs and expenses incurred by the Grace Company. A number of estimating equations were developed, many of which were analyzed or discussed in depth. In this chapter the experience gained in previous chapters and several of the estimating equations developed will be employed in examining the possible relevance of regression and correlation techniques for improving the accuracy of forecasts.

Forecasts with Regression and Correlation

In essence, budgets are quantified forecasts of future activities couched within the framework of the company's goals. These goals are a function of the company's desires for the future and its present and expected capabilities to accomplish the prescribed activities. Therefore, the budget executive, when developing the budget, must consider the company's present financial and production capabilities along with its predetermined objectives. The general approach is to commence with the company's present financial position, develop an operating budget for the budget period, and then determine the expected financial position as of the end of the budget period. Although regression and correlation could be of direct assistance in preparing the financial budget, this aspect will not be discussed in the following sections. Instead, the discussion will be limited to the applicability of correlation in developing the operating budget.

Determination of an Operating Budget

For purposes of discussion, it shall be assumed that the sales forecast has been determined; thus the models and quantities of product units to be handled and produced are given. The task now is to estimate, or forecast, all the cost elements.

Adopting the Grace Company as a hypothetical firm, the budget or cost model for the manufacturing cost sector may now be developed. Since the number of unit models to be produced is known, the direct materials cost and direct labor costs may be computed using the respective standard costs per unit. Thus, the estimated total annual cost for direct materials and direct labor may be conveniently determined. But such is not the case for the overhead cost element.

The conventional methods of estimating total overhead should now be examined. The purpose here is to determine a basis for comparing the assumption of conventional methods with those of multiple regression analysis. In all conventional approaches the first step is to segregate the strictly fixed costs from the variable and semi-variable overhead costs. These fixed costs would include both committed and programmed cost elements. Of course, this is also a necessary step if regression analysis is to be employed.

The variable and semi-variable overhead costs might then be estimated by aggregating all such accounts and computing an experience ratio. This approach assumes, in this case, that the overhead costs are not departmentalized or related to the individual production departments. The experience ratio would be computed as follows:

$$\frac{\text{Total variable and semi-variable overhead cost } t_1}{\text{Total Direct Labor Dollars } t_1} = \text{E.R.}t_1$$

The experience ratio of time period one (t_1) would be multiplied by the estimated direct labor dollars for time period two (t_2). This method is pure extrapolation and in effect, is a manifestation of seat-of-pants intuition. Added refinements might use a different denominator or include in the t_2 estimate a "fudge factor" for price increases and other possible changes. Further, a separate experience ratio might be computed for each overhead element and different denominators might be used, or the direct overhead accounts might be related to the respective production departments before the experience ratio was computed.

Assume this latter sophistication was added, wherein the direct variable and semi-variable overhead costs are associated with the respective production and service departments before the experience ratio is computed. Then it seems several assumptions are being made that may not be valid. These assumptions are: (a) that a perfectly linear relationship exists at all levels of activity, and (b) that the relationship is constant over time.

The first assumption is not valid in that practically all overhead costs have some fixed element if the firm is to continue operating. The second assumption would be valid only under extremely stable conditions that probably exist in only rare instances. Other deficiencies in the method are that no quantifiable error estimate is determined and that no measure of the relationship between the variables is computed. The regression and correlation techniques employed in the cost models do not involve making either of these two assumptions and an estimate of the probable range of error is provided along with a measure of their relationship.

A slightly more sophisticated approach, representing an attempt to recognize the existence of a fixed element within each overhead cost, is the "high-low points method." Aside from the fact that it does provide a fixed element, it suffers from the same deficiencies as the experience ratio. But it may be an even less desirable method than the experience ratio because it presumably develops an authoritative fixed cost element. It is likely that the fixed cost element established is absolutely invalid because the previously stated underlying assumptions are normally not valid.

The next step, in terms of relative sophistication, is the scatter chart. This approach requires that a free-hand curve be fitted to a series of plotted observations. Properly drawn, the curve will provide a fixed cost element for variable and semi-variable overhead costs, and it will indicate the existence of a possible curvilinear relationship. But this method assumes a constant relationship between the variables over time, which is a highly unlikely condition in most real-life situations. In addition, the method does not automatically provide an estimated range of error.

Finally, a fourth approach that may be employed is generally considered as having the highest degree of sophistication of the conventional methods. Although the simple or linear least squares approach has been classified among the conventional methods, it is likely that many budget personnel would not regard it as such. This approach involves the computation of a linear estimating equation. It will provide a fixed cost element that may be valid if a linear relationship does exist in the data, and if the data employed cover a sufficient range of activities. The primary improvement in forecasting offered by this method is that an estimated range of error is computed in the form of a standard error of the estimate, and a coefficient of determination is provided indicating the per cent of variation in the overhead cost explained by the independent variable of direct labor dollars, etc. Note that the simple least-squares technique does not remove the assumption of a constant relationship over time.

Aside from the four fundamental deficiencies mentioned previously, a fifth and quite critical deficiency should be brought out. That is, each of the four methods depends entirely on the explanatory ability of a single independent variable in only its linear form. In conducting the regression analyses in earlier chapters, it was found that after the effects of time were eliminated the use of a single explanatory variable was satisfactory in many cases. But this was true only after the time relationship was considered, and in no case, where a single explanatory variable was employed was the most efficient equation form linear. Curvilinear relationships similar to those developed may not hold true in other situations or under different conditions, but in this study a linear function was usually less efficient.

Three of the conventional methods discussed provide a figure for the fixed element contained in most overhead costs. For clarification it it should be noted that this fixed cost element is a function of the range of variation in the data being analyzed, and it is not a fixed cost in the true sense. Therefore, a fixed cost element derived by these methods is probably of questionable validity because it is based on imperfect information.

In summary, four of the most widely used conventional methods of determining the characteristics of variable and semi-variable overhead costs have been discussed. The objective of these methods is to develop valid cost relationships for forecasting expected costs. It has been pointed out that each of the four methods has at least one or more deficiencies that may impugn the accuracy and thus validity of forecasting based on any one of these approaches. However, each method represents, in general, a step forward in the development of the art. In the case of the first three methods at least, they meet two important criteria established by most firms. That is, they are: (a) practical in the sense that most budgeting personnel have the technical know-how for their application, and (b) they are economical in the sense that a minimal amount of time is required for their calculation and implementation. But in terms of accuracy their value is difficult to assess.

The estimating equations developed in the preceding chapters do not possess the characteristics criticized in the conventional methods. The basis of their development was multiple linear and multiple curvilinear correlation, which does not employ the assumptions underlying the more conventional approaches. For example, the first assumption discussed was of a perfectly linear relationship. This assumption was eliminated by the introduction of curvilinear functions. The assumption of a constant relationship over time was eliminated by the introduction of a time variable. Conceptually, if the two or more series being correlated maintained exactly constant relationships over time, the regression coefficient for the time variable would be zero. This condition did not exist in any of the equations computed. In fact, in most instances the time relationship was curvilinear. Therefore, it seems that multiple and curvilinear regression and correlation analysis provides considerable flexibility not found in the conventional methods. Further, because several explanatory variables may be employed to describe the behavior of a specific cost, the method provides vastly more sensitivity to the cost environment and changes in this environment than do the conventional methods.

Multiple or multiple curvilinear correlation also provides a measure of the total error which may occur in the budgeted figures. For example, assume the budget period is one year and that monthly data covering one year were used to compute the estimating equation. If none of the assumptions underlying the tool itself were not violated and if no alterations will occur in the dependent and independent variable relationships, then the estimated annual sum of the dependent cost variable may be expected to approximate closely the actual costs incurred. In fact, the sum of the $X_1 - X_{c1}$ differences should approach zero. This occurs because the monthly differences should off-set one another over the course of a year. This property of the least-squares method should be of tremendous value to both the firm's budgeting and operating personnel.

In light of the foregoing discussion, it seems evident that multiple and multiple curvilinear correlation represents another step forward in the evolutionary development of cost estimating techniques.

A generalized approach to developing an operating budget for the Grace Company follows. This approach employs the standard cost system developed by the company and the estimating equations developed in the preceding chapters. In essence the following framework represents a cost model.

<u>Operating Budget</u>

Manufacturing costs:
- Standard direct labor costs
- Standard materials costs
- Overhead costs
  - Committed costs
  - Programmed costs
  - Variable and semi-variable costs

Trucking expenses—53 = $+\$552.67 + \$.32542X_2^3 - \$11.6965X_2^2 + \$.0049641X_3 + \$585.289X_4$

Inspection expenses—54 = $-\$4,179.58 - \$103.49X_2 + \$.0000000001133X_3^3 - \$.00001544X_3^2 + \$.7363X_3$

Time-study allowances—56 = $+\$2,385.77 - \$6.9647X_2^2 + \$217.873X_2 + \$.0000002331616X_3^2$

Stockroom expenses—69 = $-\$582.85 + \$.2510797X_2^3 - \$15.6343X_2^2 + \$261.725X_2 + \$.0163458X_3 + \$273.577X_4$

Rework costs—55 = $-\$7,981.66 + \$.45661X_2^{3.4} - \$.000000041093X_3^2 + \$.00782479X_3$

Machinery repair expense—51 = $+\$679,545.91 - \$.6627857X_2^3 + \$204.60596X_2^2 - \$20,619.497X_2 + \$1,491.2635X_3$

Indirect supplies—52 = $+\$13,699.24 + \$89.3544X_2 + \$.012017X_3^{1.4}$

$- \$.93773X_3$

Scrap recovery—86 = $-\$1,218.20 + \$65.638X_2^2 + \$.0126507X_3$

Utilities expense—61 = $+\$3,149.35 + \$.1752695X_2^3 - \$9.655598X_2^2$

$+ \$149.0858X_2 + \$.001014126X_3 + \$18.03887X_4$

Shipping and warehousing costs:
Committed costs
Programmed costs
Variable and semi-variable costs = $+\$1,259.85 + \$77.887X_2 + \$.5775X_3$

$- \$13.372X_4^2 + \$4,126.70X_4$

Sales department costs:
Committed costs
Programmed costs
Variable and semi-variable costs
Class A costs = $+\$18,345.00 + \$362.19X_2 + \$.0105X_3 + \$1,300.13X_4$

Class B costs = $+\$5,743.76 + \$375.75X_2 + \$.000001521X_3^2$

$- \$.0260X_3 + \$2,488.20X_4$

Class C costs = $-\$1,904.96 + \$.5363X_2^2 - \$11.731X_2 + \$.00019765X_3$

$+ \$257.0258X_4 + \$577.169X_5$

Class D costs = $-\$28.437.00 + \$18.23X_2 + \$.526X_3 - \$422.23X_4^2$

$+ \$7,283.81X_4$

If the budgeting personnel of the Grace Company were to commence computations in the order the foregoing cost model is presented, a complete annual operating budget could be developed. Monthly budget data would be developed for direct materials, direct labor, and all variable and semi-variable costs in each cost segment. Almost all the data needed for the numerous explanatory variable series in the estimating equations could be taken from the forecasted sales data and the production data, which have been assumed as given, or would be generated by the direct materials and direct labor computations. Inputs for the time variables would merely be continuations of the series employed in each equation.

The preceding budget model represents an effort to synthesize as complete and reliable a cost model as possible for the variable and semi-variable costs in the factory overhead sector, in the Sales Department, and in the Shipping and Warehousing Department. Programming this model on a computer would provide a basis for forecasting or budgeting that would be more reliable, more flexible, more realistic, and just as practical as any of the contemporary methods of a lesser degree of sophistication.

Cost Control with Regression and Correlation

Control of overhead and other indirect costs represents one of the most difficult functions of management. Problems of considerable complexity are encountered when attempting to establish the budgeted costs for the coming year. Several of the conventional approaches for establishing these budget costs were previously discussed. But once the budget is established using one of these methods, is it a valid control basis? Management often goes to great lengths to convince the operating supervisors that such budgeted costs are valid as a basis for cost control. One approach taken to accomplish this is to utilize the opinions of the operating supervisors in initially developing the budgeted figures. The attitudes of operating supervisors toward the validity of the budgeting costs does have a significant bearing on morale and their willingness to cooperate in not exceeding the budgeted costs. In this regard the following observation seems appropriate.

> A criticism of multiple regression analysis is that it is complicated, and so would be difficult to "sell" to lower management and supervisory personnel. However, the method allows for a more complete specification of "reality" than do simple regression or the fixed-variable dichotomy. Studies have shown that supervisors tend to disregard data that they believe are "unrealistic," such as those based on the simplification that costs incurred are a function of units of output only. Therefore, multiple regression analysis should prove more acceptable to supervisors than procedures that require gross simplification of reality.[1]

A second feature of multiple regression that aids in cost control relates to the coefficients of multiple and partial determination and the regression coefficients. These statistics readily indicate the extent of the relationship between the dependent variable and the various explanatory variables. They also assist management and supervisors to understand better what elements and degrees of relationship are present in a real-life cost situation. Because of this improved understanding, management is in a better position to evaluate variations and performance as a whole.

A third feature of multiple linear and multiple curvilinear regression that assists in cost control relates to the standard error of the estimate. Once a valid estimating equation has been developed for a particular overhead cost, the associated standard error of the estimate may be used as a measure of the monthly $X_1 - X_{c1}$ variations for control purposes. Of course, it is necessary that all the previously stated assumptions be fulfilled before this procedure may be employed. If the assumptions are fulfilled, then management may prescribe a range within which monthly variations could occur and still be considered acceptable. Employing the Indirect Supplies

---

[1]Benston, _loc_. _cit_., p. 659.

--52 equation and its associated standard error of the estimate of $1,616.86 for purposes of illustration, management could establish as acceptable, a variation range equal to 1.96($1,616.86). Management would know there is a probability of .05 that actual monthly costs will differ from their average value by more than $3,169. If a difference occurred of more than $3,169 it would be a signal for management to review this cost and perhaps quiz appropriate personnel as to its cause.

Use of the standard error of the estimate for this purpose provides a marvelous control technique that lends significant validity to the principle of management by exception. This technique not only provides a measured range of variation, but it also attaches a probability of its occurrence. Thus, management is not forced to establish arbitrary cost control ranges, as in the case of less sophisticated methods. Further, supervisors are fully aware of the range of permissible variations.

It was stated earlier that the variations should be both positive and negative over the budget period and because of this behavior the sum of the variations should approach zero over the entire budget period. Thus, it would seem that a variation pattern of positive, negative, positive, etc. is to be expected. Although this may not always be the case, the monthly variations over the period may well sum to approximately zero. Some indication, of the variation pattern to expect, might be determinable from an examination of the variation pattern demonstrated in prior periods. In any case, the month-to-month variation pattern may provide early warning to management that actual annual costs are not going to approximate closely the budgeted costs, even though not one of the individual monthly variations exceeds the predetermined limits.

Limitations of Regression and Correlation for Forecasting and Cost Control

If accurate forecasts that are valid for cost control purposes are to be developed by employing multiple and multiple curvilinear regression and correlation analysis, then it is obvious that the assumptions underlying regression and correlation must not be violated. If these assumptions are not fulfilled then an apparently valid estimating equation may be just an authoritative facade. In conducting analyses the analyst must always be alert for possible conditions that might completely invalidate any estimating equations to be computed.

Criticism is often leveled against these tools because the estimating equations are developed from historical data. Presumably, the implication is that what has happened in the past is not an adequate indication of the future. This is a valid criticism, but then it also applies equally well to the conventional methods employed by most firms. However, if the forecasts prepared from the estimating equations are to be accurate, the same patterns and relationships on which they were computed must continue in the future.

It is extremely dangerous to prepare forecasts employing data that are not within the range of experience from which the estimating equations were computed. This danger of fallacious accuracy may be somewhat mitigated by the employment of only current data when computing the estimating equations and by the inclusion of a time variable that itself provides an indication of the changing time relationships. Further, a process may be employed whereby the estimating equations are continually up-dated by including the experience of each new month in the estimating equations. Thus, in effect, a revised budget would be prepared each month. Another procedure that might be employed, when it is known that specific changes will occur in the future, involves restating the historical data in light of the anticipated changes and then developing revised estimating equations. Occasionally similar results may be obtained by including an additional variable in the estimating equation representing the expected change. However, these last two procedures should be carried out circumspectly.

If one is preparing forecasts for a one year period and requires that the sum of the $X_1 - X_{c1}$ variations over the year should approach zero, then it is necessary that a one-year period be used to compute the estimating equations. If monthly data are employed, then only 12 observations would be available. Unless it is possible to fit an extremely efficient estimating equation with not more than two explanatory variables, it is advisable that more observations be included. In general, when multiple curvilinear regression is employed, the inclusion of only 12 observations will not provide many degrees of freedom. Thus, management's confidence in the estimating equation's forecasting accuracy might be reduced. Increasing the number of observations to include 18 to 36 months might provide a better basis for forecasting. But if such a procedure were followed the sum of the variances over the next 12 months might not $\longrightarrow 0$.

Managers and supervisory personnel, unless properly informed, may have a tendency to view the pure constant term in the estimating equations as a fixed cost element. Such an interpretation might lead to incorrect conclusions and decisions pretaining to cost control. It has been repeatedly emphasized in previous chapters that the pure constant can not be interpreted as being the fixed cost element, "...unless the linear relationship found in the range of observations obtains back to zero.... Thus, the value of the constant term... is not the cost that would be expected if there were no output; it is only the value that is calculated as a result of the regression line computed from the available data."[2] The same conditions apply in the curvilinear case.

In the preceding chapters a number of estimating equations were developed. These analyses clearly revealed that the analyst must sift and organize the data in depth in order to determine the proper explanatory variables. Determining causation and valid relationships is a complex task requiring knowledge of the statistical tools and a complete knowledge of the costs and activities being related. It would seem that the knowledge needed and the perceptiveness required could come only from experience and study. Library computer programs make it possible to churn out any number of estimating equations, but without an analyst with the foregoing characteristics the validity of such estimating equations would be highly questionable.

One final observation that may be construed as a limitation in the applicability of regression and correlation for forecasting and cost control remains. It was observed that a number of possible estimating equations exist for any set of data. The problem lies in determining the best estimating equation. It was clearly stated throughout the analyses that the

---

[2]<u>Ibid.</u>, p. 661.

estimating equation accepted as a cost model for each cost was the most efficient of those computed, but there was no assurance that it was the best obtainable. For each set of data an estimating equation can be developed for which the sum of the deviations will equal zero. But there is only one equation that will produce a minimum standard error of the estimate, given a specified set of variables. The problem is to determine that one equation. The approach was iteration and at each step the coefficients of determination and standard error of the estimate were evaluated. It is believed this approach will lead to an estimating equation that at least approximates the one best equation. Of course, that this type of analysis requires time and manpower points up the economic maximization rule that the marginal cost should not exceed the marginal revenue generated by the activity. Therefore, "once a significant correlation is obtained, the judgment as to whether or not it is good enough will depend upon the experienced accuracy with other methods, the standard for accuracy that has been set, and the time available for further refinement."[3]

Summary

The objective in this chapter was to examine the applicability of regression and correlation in the budgeting function. In this effort budgeting was declared in essence a forecast that must serve as a framework for cost control. In accord with previous analyses and discussions, committed and programmed cost elements were excluded from the realm of applicability of regression and correlation. Four conventional methods for analyzing, forecasting, and establishing cost control bases for variable

---

[3]Knapp, loc. cit., p. 16

and semi-variable costs were then presented. The assumptions underlying each of these four methods were discussed and it was pointed out that these assumptions rendered each approach, at the least, deficient. However, it was shown that these assumptions were not relevant when regression and correlation analysis was employed. The advantages of correlation for forecasting and cost control were presented along with a complete cost model for the Grace Company, which employed the estimating equations developed in the preceding chapters.

Although the beneficial characteristics of multiple linear and multiple curvilinear regression and correlation were presented, it is hoped that their limitations relating to forecasting and cost control capabilities were as clearly discussed. In the final analysis, it seems the following statement aptly synthesizes the discussions within this chapter.

> Correlation analysis does not, of course, supplant an investigative procedure whereby the forecaster looks through the eyes of the people responsible for the area of operations to be forecast. Only through such a process can impending developments without historical precedent be factored into any prediction. Correlation analysis does provide the most concise expression of the historical interrelationships of the selected variables that can be derived. It offers no guarantee that the relationships will hold. But, used in combination with knowledge of operating plans and judgment of future business conditions, it can eliminate a large percentage of the error inherent in other forecasting methods.[4]

[4] Ibid., p. 14.

CHAPTER IX

SUMMARY AND CONCLUSIONS

Summary

The principal objective of this study has been to determine the practical applicability of multiple linear and multiple curvilinear regression and correlation for forecasting and control of manufacturing overhead and other indirect costs. In the main, it appears that these tools are practical for the purposes described. In order to achieve the principal objective and formulate a positive conclusion, it was necessary: (1) to examine the underlying assumptions of the tools themselves, (2) to determine conditions under which the assumptions may not be fulfilled and the resulting implications, (3) to review present day techniques of developing cost information, (4) to apply the tools in a real-life environment, and (5) to compare the information generated by these statistical tools to that generated by traditional methods.

The concepts of the "least-squares method" of regression and correlation require five broad assumptions which were discussed in the first section of Chapter II. These assumptions permit calculation of an estimating equation offering an unbiased estimator of the true population parameters. However, if certain conditions in the data invalidate these assumptions, then the resulting estimating equation will be biased. A number of biasing conditions were discussed, such as multicollinearity, autocorrelation, and measurement errors. For each condition remedial or at least mitigative procedures were discussed. The purpose here was to determine techniques that would: (a) indicate the presence of such conditions and (b) measure or eliminate their possible effects. It is not possible to declare merely that if biasing conditions do exist they will be eliminated. This is too broad a statement and is made without regard to the type of error being introduced. Recognition, measurement, and possible elimination of biasing conditions may be discussed only within the context of a specific set of data.

A discussion of standard cost systems was presented for the purpose of establishing an accounting framework and of indicating how regression and correlation might supplement this technique in the overhead costs sectors. Thus in Chapter II the conceptual aspects of regression and correlation were perused, some of the complications were examined, and a cost accounting environment was established.

The most critical and by far the most complex aspect of this study was commenced in Chapter III. At this point actual analyses were started for the various cost sectors of the Grace Company. This phase of the study continued through Chapter VII, where the last of the manufacturing service department costs were analyzed. In these chapters selected empirical cost data of the Grace Company were subjected to regression and correlation analysis. If valid estimating equations could not have been developed for the various costs, then in this case the relative applicability of regression and correlation for forecasting and cost control purposes would have been suspect.

The mathematical manipulation required to develop a single multiple regression equation is almost overwhelming if done by hand or on a desk calculator. Therefore, the computer facilities at the University of Texas at Arlington were employed for all calculations. The program employed was card program 6.0.002 for the IBM 1620 computer from the IBM users' library.

The objective in Chapters III through VII was to develop valid and efficient estimating equations for selected cost functions. The general methodology, similar for each cost group, is summarized as follows:

1. Determine the functions of the department.
2. Determine the accounts employed in the department to collect and summarize costs.
3. Review each cost account as to its variability and clerical accuracy. Make appropriate corrections for clerical and classification errors.
4. Classify each cost account as variable or fixed.
5. Each cost account classified as fixed should then be re-examined to verify its committed or programmed nature. If the account is confirmed as a fixed cost, it is excluded from further analysis.
6. Determine a series of possible explanatory variable candidates. This should be done in light of the functions carried out in the department and in light of the nature, variability, and other characteristics of the costs to be explained.

7. Evaluate each of the explanatory variables series as to availability, accuracy, and logical causation. These procedures involve first a determination that the data are available. If it is available it must be reviewed for possible clerical and classification errors. Obviously, the explanatory variable series should cover the same time periods as does the proposed dependent series and it must be in the same time increments, e.g., months, weeks, days.
8. The explanatory variables series remaining after step 7 was completed were then charted with time plotted on the X axis.
9. The variable cost accounts were then examined in depth. Concurrently they were charted and compared to the several explanatory series. All of the foregoing steps are vital as they assist in eliminating bias attributable to clerical and classification errors and provide an understanding of the relationships between the variables to be correlated. In many cases these procedures also provided a basis for consolidating several of the cost accounts.
10. Compute a linear estimating equation using a time variable and one other explanatory variable. The time variable was included in the equation for the purposes of: (a) accounting for variations between the series due to time and (b) mitigating or eliminating the effects of autocorrelation if it exists in the data.
11. For the equation computed, evaluate the coefficient of determination, the standard error of the estimate, and the standard

error of the regression coefficient. If the coefficient of determination was statistically significant, even though the standard error of the estimate was larger than desired, further steps were then taken to determine if that equation form provided the best fit.

12. This involved carrying out a series of iterations. Normally an effort was made to determine first the most efficient form of the time variable and second the most efficient form of the other explanatory variable. Determination of which equation form was most efficient was based on a combined judgment of the coefficient of multiple determination, the standard error of the regression coefficients, the standard error of the estimates, and the logic of the indicated relationships.

13. If it was deemed necessary or desirable, additional explanatory variables were included in the estimating equation. But they were added only one at a time. In each case the same evaluation procedures were carried out as before, except that particular alertness was needed to determine if significant multicollinearity had been introduced into the equation by the inclusion of a new variable. Inclusion of a new explanatory variable should have the effect of significantly increasing the coefficient of determination, decreasing the standard error of the estimate, and often of decreasing the standard error of the regression coefficients. When multicollinearity was also being introduced, it would normally increase the standard error of the regression

coefficient for the explanatory variable(s) that was correlated with the new variable being introduced.

14. When an efficient estimating equation had been developed it was then used to restate the dependent variable series. The variations between the computed and actual historical values were then computed. This procedure indicates the existence of extreme autocorrelation in the data and the variance pattern that might be expected in future periods. Most importantly it enhances or conversely detracts from the creditability of the normality assumption. Therefore, if the variations were not approximately evenly distributed above and below the computed curve and if their dispersion was not approximately normal, the estimating equation was not accepted as a model of the dependent variable cost.

Steps one through eight were carried out for each sector of the Grace Company. Steps nine through fourteen were carried out for each variable cost account or group of accounts. The writer believes that this methodology did provide a valid series of estimating equations.

Since it was shown that valid cost estimating equations could be determined in the specified business environment, it was then appropriate that the characteristics of these equations relating to forecasting and controlling costs be examined. This phase of the study was conducted in Chapter VIII. Because there are no established "ideal" forecasting or control techniques, it was necessary to evaluate the capabilities of the estimating equations in view of the traditional methods employed. These comparisons established that regression and correlation estimating equations do represent a vastly more sophisticated and accurate approach to forecasting and controlling costs than do traditional methods.

Conclusions

Multiple linear and multiple curvilinear regression and correlation are practical techniques for improving the accuracy of cost forecasting and for increasing management's capability to control overhead and other indirect costs in a manufacturing firm having multiple production and service departments. This conclusion is based on the following premises.

1. Given the empirical overhead and other indirect costs of the Grace Company, a series of valid and efficient cost estimating equations was developed. These estimating equations are valid and efficient in accordance with predetermined criteria prescribing the degree of efficiency required. Of primary importance was the criterion that each computed estimating equation must be capable of restating the dependent variable series within acceptable limits and in acceptable variation patterns.
2. The estimating equations computed, in general, explain the relationships between the dependent cost series and the several explanatory variable series so that management through employing this technique is in a better position to determine and evaluate activities in terms of their cost consequences. The inclusion of several explanatory variables provides flexibility and a degree of realism that is not even remotely attainable when traditional methods are employed.

3. The standard error of the estimate, assuming a normal distribution within the population, provides a measured range of variation and the probability of occurrence. Therefore, management has a statistically valid quantification of the expected monthly, weekly, etc., variations. None of the conventional methods of forecasting costs possesses these attributes. Further, the estimating equations developed in this study do, on the whole, provide variation patterns that are in conformity with the normality assumption. This surely lends credibility.
4. For the estimating equations computed and for all valid estimating equations, the sum of the $X_1 - X_{c1}$ differences, over the period used for calculation of the estimating equation, will approach zero. Therefore, providing the same relationships were to hold in the forecast period as were present in the twelve-month computational period, management could justifiably expect the sum of the twelve monthly variations in the forecast year to approach zero. That is, the annual budgeted costs should closely approximate the actual costs incurred.
5. Of all the various cost accounts contained within the three cost sectors of the Grace Company, and for which regression and correlation were a valid tool of analysis, only one account failed to yield satisfactorily to analysis. This account was the Overtime Premium Expense. It was determined that the incompatibility of the time increments creating the dependent series as opposed to those within which the several explanatory variables were couched

was the reason why such a condition existed. Thus, the tool itself, through the statistical measures it provides, assists the analysts to develop proper cost and explanatory variable relationships.

6. The fact that only one of the accounts failed to yield to regression analysis belies the common presumption that most firms do not have accounting systems that generate valid data for regression analysis. It would seem that contemporary accounting systems do yield a great many series of data that are amenable to regression analysis. Given that intermediate and larger manufacturing firms do employ computerized accounting systems, it would seem that the data necessary for regression and correlation analysis could be routinely and economically generated.

In view of the foregoing experiences and observations it appears the conclusion drawn is justified. However, it must be noted that the regression and correlation technique was not employed to analyze any of the direct costs of manufacturing. Thus inferences may not be made regarding such costs and the relevance of regression and correlation. Nor should any parallels be drawn in connection with the financial elements of the firm.

In the process of conducting the numerous regression analyses contained in this study, the writer encountered a number of problems that aptly pointed up limitations of regression and correlation for business data of the type being analyzed. One of the most common problems encountered was multicollinearity. When this condition exists to the point that the standard errors of the regression coefficients are unduly large or even dominate their associated regression coefficients, then the latter measures are not properly interpreted as marginal costs. Further, the coefficients of partial determination are not an appropriate measure of the variation explained by the correlated explanatory variables. On the other hand, elimination of the explanatory variable that introduced the multicollinearity condition might well increase the magnitude of a valid standard error of the estimate. Thus, the analyst is in a dilemma. Should he retain the standard error of the affected regression coefficient? The answer probably depends on the purpose for which the equation was being developed. In any case, it would certainly be helpful if the effects of multicollinearity could be isolated and measured. No doubt future statisticans will develop satisfactory techniques for combating this problem.

A second problem of considerable magnitude is accuracy of the data. The data employed in this study were often cluttered with classification errors, clerical errors, errors due to untimely recording of expenses, and inclusion of irrelevant items. Considerable time was spent in correcting these data deficiencies, time that the practicing accountant might not have available. The answer lies in developing an accurate cost accounting system.

Regression and correlation techniques of the types employed in this study require large amounts of computer operating time. This may or may not be a limitation for many companies. In addition, the numerous and complex subtleties of the tools require that close attention be given to all the statistical measures of fit generated. It is not sufficient to accept blindly that equation having the highest coefficient of determination. The analyst must have an understanding of the tool itself as well as a thorough knowledge of the cost system from which the data is being abstracted.

In conclusion, it is believed that multiple regression and correlation will gain wider acceptance among managerial accountants as an extremely useful tool and that eventually it will become a standard tool. It is hoped that this study will contribute some measure of light to the evolutionary development of managerial accounting.

BIBLIOGRAPHY

Books

Anton, Hector R., et al. Contemporary Issues in Cost Accounting. Edited by Hector R. Anton and Peter A. Firmin. Boston: Houghton Mifflin Co., 1966.

Brennan, Michael J. Preface to Econometrics. 2nd ed. Cincinnati: South-Western Publishing Company, 1965.

Comiskey, Eugene E. "A Study of Loan Cost Behavior in Consumer Finance Companies." Unpublished Ph. D. dissertation, Michigan State University, 1965.

Davies, Owen L., et al. The Design and Analysis of Industrial Experiments. Edited by Owen L. Davies. New York: Hafner Publishing Co., 1960.

Ezekiel, Mordecai. Methods of Correlation Analysis. 2nd ed. 8th printing. London: Chapman & Hall, Limited, 1963.

Ferber, Robert. Statistical Techniques in Market Research. New York: McGraw-Hill Book Company, Inc., 1949.

Horngren, Charles T. Cost Accounting: A Managerial Emphasis. Englewood Cliffs: Prentice-Hall, Inc., 1964.

Johnston, J. Econometric Methods. New York: McGraw-Hill Book Company, Inc., 1963.

Johnston, J. Statistical Cost Analysis. New York: McGraw-Hill Book Company, Inc., 1960.

Matz, Adolph, Curry, Othel J., and Frank, George W. Cost Accounting. Cincinnati: South-Western Publishing Co., 1962.

McFarland, Walter B. Concepts for Management Accounting. New York: National Association of Accountants, 1966.

Welsch, Glenn A. Budgeting Profit Planning and Control. 2nd ed. Englewood Cliffs: Prentice-Hall, Inc., 1964.

Articles

Amicucci, Donald J. "Budget Variance Trend Reports," N.A.A., Vol. 46 (July, 1965), pp. 9-14.

Bedford, Norton M., and Onsi, Mohamed, "Measuring the Value of Information - An Information Theory Approach," Management Services, Vol. 3. No. 1 (Jan.-Feb. 1966), pp. 15-22.

Benninger, L. J. "Accounting Theory and Cost Accounting," The Accounting Review, Vol. 40, No. 3 (July, 1965), pp. 547-557.

Benston, George J., "Multiple Regression Analysis of Cost Behavior," The Accounting Review, Vol. 41, No. 4 (October, 1966), pp. 657-672.

Brightly, Donald S., "A Statistical Tool for the Cost Accountant," Management Services, Vol. 2, No. 5 (Sept.-Oct., 1965), pp. 13-17.

Carlisle, Howard M., "Cost Accounting for Advanced Technology Programs," The Accounting Review, Vol. 41, No. 1 (January, 1966), pp. 115-120.

Charnes, A., Cooper, W. W., and Ijiri, Y. "Breakeven Budgeting and Programming to Goals," Journal of Accounting Research, Vol. 1, No. 1 (Spring, 1963), pp. 16-41.

Chiu, John S. and De Coster, Don T., "Multiple Product Costing by Multiple Correlation Analysis," The Accounting Review, Vol. 41, No. 4 (October, 1966), pp. 673-680.

Comiskey, Eugene E., "Better Budgeting with Multiple Regression Analysis," Cost and Management, Vol. 41, No. 1 (Jan., 1967), pp. 14-17.

Comiskey, Eugene E., "Cost Control by Regression Analysis," The Accounting Review, Vol. 41, No. 2 (April, 1966), pp. 235-238.

Copeland, Ben R., "Analyzing Burden Variance for Profit Planning and Control," Management Services, Vol. 2, No. 1 (Jan.-Feb., 1965), pp. 34-41.

Corcoran, A. Wayne, "A Proposal for Condensing Diverse Accounting Procedures," Management Services, Vol. 3, No. 6 (Nov.-Dec., 1966) pp. 15-24.

Curry, Richard A., "Fixed-Variable Cost Separation--Basic and Difficult," N.A.A., Vol. 44, No. 7 (March, 1963), pp. 50.

Duffey, Daniel J., A Control Chart Approach to Manufacturing Expense," Journal of Industrial Engineering, Vol. 11, No. 6 (Nov.-Dec., 1960), pp. 451-458.

Earley, James S., "Business Budgeting and the Theory of the Firm," Journal of Industrial Economics, Vol. 9, No. 1 (Nov., 1960) pp. 23-42.

Earley, James S. and Carleton, Willard T., "Budgeting and the Theory of the Firm: New Findings," Journal of Industrial Economics, Vol. 10, No. 1 (Nov., 1961), pp. 165-173.

Ferrara, William L., "Responsibility Accounting--A Basic Control Concept," N.A.A., Vol. 46 (September, 1964), pp. 11-19.

Foley, Eugene P. "Reducing Indirect Costs--A Case Study," N.A.A., Vol. 46 (June, 1965), pp. 19-24.

Graese, C. E. "Responsibility Reporting to Management," The Accounting Review, Vol. 39, No. 2 (April, 1964), pp. 387-391.

Gynther, R. S. "Improving Separation of Fixed and Variable Expenses," N.A.A., Vol. 44, No. 10 (June, 1963), pp. 29-38.

Hanson, Ernest I. "The Budgetory Control Function," The Accounting Review, Vol. 41 (April, 1966), pp. 239-342.

Henrici, Stanley B. "New Views on Standards," N.A.A., Vol. 46 (July, 1965), pp. 3-8.

Horngren, Charles T. "Choosing Accounting Practices for Reporting to Management," N.A.A., Vol. 44, No. 1 (September, 1962), pp. 3-15.

Jensen, Robert E. "A Multiple Regression Model for Cost Control--Assumption and Limitations," The Accounting Review, Vol. 42, No. 2 (April, 1967), pp. 265-273.

Kennon, Walter. "Simulation--Management's Laboratory," N.A.A. Vol. 45 (April, 1964), pp. 19-26.

Khowry, E. N. and Nelson, H. Wayne, "Simulation in Financial Planning," Management Services, Vol. 2, No. 2 (March-April, 1965), pp. 13-21.

Knapp, R. A., "Forecasting and Measuring with Correlation Analysis," Financial Executive, Vol. 31 (May, 1963), pp. 13-19.

Mattessich, Richard, "Budgeting Models and System Simulation," The Accounting Review, Vol. 36 (1961), pp. 384-397.

McClenon, Paul R., "Cost Finding Through Multiple Correlation Analysis," The Accounting Review, Vol. 38, No. 3 (July, 1963), pp. 540-547.

McLaughlin, Robert L., "The Breakthrough in Sales Forecasting," Journal of Marketing, Vol. 27, No. 2 (April, 1963), pp. 46-54.

Manes, Rene Pierre, "Using Computers to Improve Distribution of Service Costs," The Journal of Accountancy, Vol. 115, No. 3 (March, 1963), pp. 57-60.

Marple, Raymond P. "The Relative Contribution Approach to Management Reporting," N.A.A., Vol. 44, No. 10 (June, 1963), pp. 3-14.

Newman, Maurice S., "The Essence of Budgetary Control," Management Services, Vol. 2, No. 1 (Jan.-Feb., 1965), pp. 19-26.

Onsi, Mohamed. "Quantitative Models for Accounting Control," The Accounting Review, Vol. 42, No. 2 (April, 1967), pp. 321-330.

Peirce, Richard F., "The Importance of the Distinction Between Fixed and Variable Costs," N.A.A., Vol. 45 (May, 1964), pp. 19-26.

Troxel, Richard B., "Variable Budgets Through Correlation Analysis: A Simplified Approach," N.A.A., Vol. 46 (February, 1965), pp. 48-55.

Walton, Horace C., "Profit Control and Measurement Through Statistical Correlation," The Controller, Vol. 27, No. 9 (September, 1959), pp. 410-411, 430-431.

Zannetos, Zenon S., "Standard Costs as a First Step to Probabilistic Control: A Theoretical Justification and Extension and Implications," The Accounting Review, Vol. 39, No. 2 (April, 1964), pp. 296-304.