APPENDIX E

Table of Contents

Paragraph *Page*

E-000 Graphic And Computational Analysis Techniques

E-001 Scope of Appendix E1

E-100 Section 1 --- Correlation and Use of Scatter Diagrams

E-101 Introduction E1
E-102 Audit Objective E1
E-103 Terminology E1
E-104 The Scatter Diagram E2
E-104.1 Purpose E2
E-104.2 Construction – Scatter Diagram E2
E-104.3 Interpretation E2
Figures E-1-1 through E-1-3 Scatter Diagrams E4
Figures E-1-4 through E-1-9 Scatter Diagrams E5
E-104.4 Selection of Appropriate Scales E6
Figure E-1-10 Scatter Diagrams-Illustration of Poor Choice of Scales E6
Figure E-1-11 Scatter Diagrams-Illustration of Correct Choice of Scales E7
E-104.5 Time Relationships E7
Figure E-1-12 Scatter Diagrams-Illustration of Joining Points by Time Sequence E8
Figure E-1-13 Estimating or Evaluating an Estimate of the Dependent (Y) Value From a Known Independent (X) Value E9
E-104.6 Use and Limitations E10

E-200 Section 2 --- Correlation and Regression Analysis

E-201 Introduction E11
E-202 Simple Linear Regression Analysis E11
E-202.1 Definition and Concept E11
Figure E-2-1 Graph of Simple Linear Regression Line E12
E-202.2 Example of Computerized Regression E13
Figure E-2-2 EZ-Quant Regression E14
E-203 Curvilinear Regression Analysis E16
E-204 Multiple Regression Analysis E16
E-205 Correlation Analysis E17

Paragraph *Page*

E-205.1 Interpretation of Coefficient of Determination E17
Figure E-2-3 Graph Illustrating Computation of Index of Determination (R2). E18
E-205.2 Determining the Existence of Correlation..... E18
E-205.3 Evaluating the Extent of Correlation..... E19
E-205.4 Comparison of Two Analyses E20
Table E-2-1 Table of Significant Values of the Coefficient of Determination (R2) E21
E-206 Prediction Intervals..... E22
E-207 Adjustments for Economic Factors in Regression Analysis..... E22
E-208 Other Considerations in Using Regression Analysis..... E23

E-300 Section 3 --- Time Series Charts

E-301 Introduction E25
E-302 Definition and Concept..... E25
Figure E-3-1 Number of Indirect Employees by Quarters E25
E-303 Chance and Unusual Variations E26
E-304 Secular Trends..... E26
E-305 Seasonal Variations E27
E-306 Business Economic Cycle E28
E-307 The Moving Average..... E28
E-307.1 Description E28
E-307.2 Construction – Moving Average E29
Figure E-3-2 Three-Month Moving Average..... E30
E-307.3 General E30
E-308 Adjustment of Data..... E31
E-309 Length of the Analysis Period E32
E-310 Interpretation and Use E32
E-310.1 Comparison of Concepts E32
E-310.2 Application..... E33
E-311 Illustration of Computation of the Least-Squares Line for a Time Series..... E33
Figure E-3-3 Computation of the Least Squares Line for a Time Series E34

E-400 Section 4 --- Audit Applications of Graphic and Computational Techniques

E-401 Introduction E36
E-402 Identification of Departures from Historical Cost Patterns E36

Paragraph *Page*

E-402.1 General E36
E-402.2 Compilation of Data E36
E-402.3 Analyzing the Data E37
E-403 Evaluation of Proposed Overhead Costs E38
E-404 Evaluation of Allocation Bases E39
E-405 Evaluation of Proposed Labor Rates E39
E-405.1 General E39
E-405.2 Illustration of the Use of Analysis in the Evaluation of Labor Rates E40
Table E-4-1 Average Monthly Direct Labor Hourly Rates and Number of Direct Workers E41
E-405.3 Alternative Methods E43
E-406 Other Applications E43
Figure E-4-2 Correlation of Hourly Labor Rates and Number of Employees ... E44

APPENDIX E

E-000 Graphic And Computational Analysis Techniques

E-001 Scope of Appendix

a. This appendix and Appendix F provide guidance for the auditor in the use of computational methods, charts, and graphs for analyzing contract costs and associated production data. The presentation is directed toward special problems of cost evaluation in the field of contract auditing and is not intended as a general discussion of graphic and computational analysis techniques. Each of the procedures presented should be considered from this limited viewpoint.

b. The techniques discussed in this appendix are not by themselves means of evaluating costs and cost estimates. They are, under appropriate conditions and combined with proper auditing procedures, valuable aids or tools which the auditor may use in establishing a basis of facts from which sound conclusions on the reasonableness and acceptability of the contractor's cost statements may be formed.

c. The discussion of improvement (or learning) curves, which also involves consideration of graphic and computational analysis techniques, has been placed for convenience in a separate appendix, Appendix F.

E-100 Section 1 --- Correlation and Use of Scatter Diagrams

E-101 Introduction

This section presents information on the objectives of correlation techniques and the preparation of scatter diagrams.

E-102 Audit Objective

a. The auditor's objective when using the techniques described for Scatter Diagrams and Regression Analysis (E-200) is to evaluate the relationships and interdependencies that may exist between two or more variables. The auditor is concerned with determining, measuring, and describing the manner in which the level of a particular cost or group of costs is influenced by changes in other factors.

b. One of the principal audit uses of these techniques is for identifying current departures from historical cost patterns. The graphic presentations and computations described in the following paragraphs provide a ready means of focusing attention on those costs which are deviating from experienced trends and which, therefore, may be assumed to require some degree of special examination. By thus isolating cost factors needing special examination, the auditor is also provided with a means for improving the control and planning of the audit.

c. Another important use is for predicting costs. Such applications entail establishing a mathematical relationship between a cost the auditor wants to predict and one or more additional variables that can be predicted. For example, if the auditor determines that a contractor's overhead costs are related to direct labor hours, he can use this information to predict overhead rates. Such predictions assume that relationships observed in the past will continue in the future. They are most likely to be reliable when they are within the general relevant range of the historical data. If changed conditions or circumstances, such as operating efficiencies, changes in plant location, etc., which will affect overhead costs are anticipated, projections based on historical data should be adjusted to reflect the related change in costs anticipated.

d. Other applications of these techniques include the evaluation of how closely costs in an overhead pool have been related to various possible bases of allocation.

E-103 Terminology

a. A variable is a quantity that assumes different values at different times or for different units. A contractor's direct labor-hours, average labor rates, and expense lev-

els are examples of variables that assume different values for different periods or points in time. The wage rate and period of service for individual employees are examples of variables that assume different values for different members of a population.

b. Two variables are said to be correlated when there is a measurable tendency for a change in one to be accompanied by a change in the other. If a variable (for example, direct labor hours) is considered to cause changes in the other, it is called the independent (or *x*) variable. The other variable (for example, indirect labor hours) is called the dependent (or *y*) variable. The nature of this relationship may vary from one of a high degree of dependence, where a change in the independent variable directly causes a change in the other, to one of merely a casual association. For example, some indirect labor is required to supervise and support direct labor. However, changes in other categories of indirect labor are not directly caused by changes in direct labor requirements. Rather, both are the result of changes in production requirements.

c. Unless otherwise stated, "average" signifies the "arithmetic mean" usually referred to simply as the "mean." In order to avoid any confusion with other types of averages, the word "mean" has been used throughout this appendix where appropriate. The word "average" is used either because it is an adjective or verb or because statistical usage requires it, as in the discussion of "semi-average" and "moving averages." The words "formula" and "equation" are generally used interchangeably. The latter term is preferred when the text pertains to the mathematics involved or where it is necessary to conform to statistical usage as in "simultaneous equations."

E-104 The Scatter Diagram

E-104.1 Purpose

The scatter diagram, sometimes called the scattergram, is the basic device for graphically displaying the relationship between two variables. Ordinarily the initial step in simple correlation analysis should be the drawing of such a diagram, either manually or by using computer software such as Microsoft's Excel scatter chart, or EZ-Quant Regression's Least-Squares Multiple Linear Regression program. If only a rough notion of the degree of relationship is required, the scatter diagram alone may yield adequate results. In other cases, further analysis may be advisable. Nevertheless, the scatter diagram is useful for exploratory purposes in any analysis of quantitative or trend relationships and can serve as a guide to further evaluation and audit work.

E-104.2 Construction – Scatter Diagram

A scatter diagram is plotted manually on graph paper or by computer software using a horizontal *x* scale to measure values of the independent *x* values and a vertical *y* scale to measure values of the dependent *y* values. The method of plotting should be clear from observation of Figure E-1-1, and it is evident that a scatter of points is obtained. Other illustrations of scatter diagrams are given in Figures E-1-2 to E-1-9. The size of the plot points in these graphs has been exaggerated for the purpose of illustration. In practice, accuracy requires more precise plotting.

E-104.3 Interpretation

a. Direction, Position, and Linearity of Pattern.

(1) The points of a scatter diagram normally form a pattern having a definite \direction in relation to the axes of the graph. It is inherent in the construction of diagrams of this type, as illustrated in Figure E-1-1, that when the general path of the pattern is from the lower left to the upper right of the graph, the dependent *y* values increase with increases in the independent *x* factor. The correlation in these cases is said to be direct or positive. When the path is from the upper left to the lower right (Figure E-1-2), the dependent *y* values decrease as the independent *x* values increase, and the correlation is said to be inverse or negative.

(2) The general path of the scatter of points also indicates whether the correlation is linear or curvilinear. If the central path through the pattern of points from the lowest to the highest *x* value is approxi-

mately a straight line as shown in Figures E-1-1 and E-1-2, the correlation is linear and there is said to be a constant or straight-line relationship between the variables. If the central path through the pattern is curved, as illustrated in Figure E-1-3, the correlation is curvilinear and a curved-line relationship is said to exist between the variables. It is important to distinguish between linear and curvilinear relationships because different computational techniques are required for analysis of the data. Serious errors may result from assuming a straight-line relationship when the true relationship is curvilinear. An important reason for constructing a scatter diagram prior to computational analysis is to determine from the path of the points the type of analysis that should be used. However, the determination of a functional relationship should not be made exclusively from a scatter diagram; the auditor should also consider the degree of logic in relating the variables concerned.

b. Degree of correlation.

(1) In the ideal case, all the points would fall precisely on a slanted straight line or smooth curve, as in Figures E-1-4 and E-1-5. In such a situation, seldom found in practical work, there is perfect correlation, and the value of one variable may be accounted for and computed from the value of the other variable. When this line is straight and lies at or near a 45 degree angle to the axis, assuming *x* and *y* scales of equal magnitude, a change in one variable would be associated with a proportionate change in the other variable. However, as the pattern approaches either the horizontal or vertical position, changes in one of the variables are associated with either much smaller or much larger changes in the other variable. When the plotted points fall in a line that is either horizontal or vertical, as in Figures E-1-6 and E-1-7, there is no correlation between the two sets of data, since one and only one value of one of the variables is associated with all values of the other variable. A scatter diagram of fixed expenses and direct labor hours, for example, would show this relationship.

(2) Ordinarily the points do not fall in a line but tend to form a band-like pattern as in Figure E-1-1. The width of this pattern (the scatter of the points), together with its slope as indicated by its line of central tendency, is a direct indication of the degree of correlation or the closeness of the relationship between the two series of data. The degree of correlation indicates the confidence that can be placed on the historical validity of the relationship, but not on its continuance. A very narrow sloping pattern, approximating a line, indicates a high degree of correlation. The wider the pattern and the more that a pattern of a given width approaches the horizontal or vertical, the lower the degree of correlation. If the pattern formed by the points is very wide, as shown in Figure E-1-8, little correlation is present as there is little tendency for one variable to change in consonance with the other.

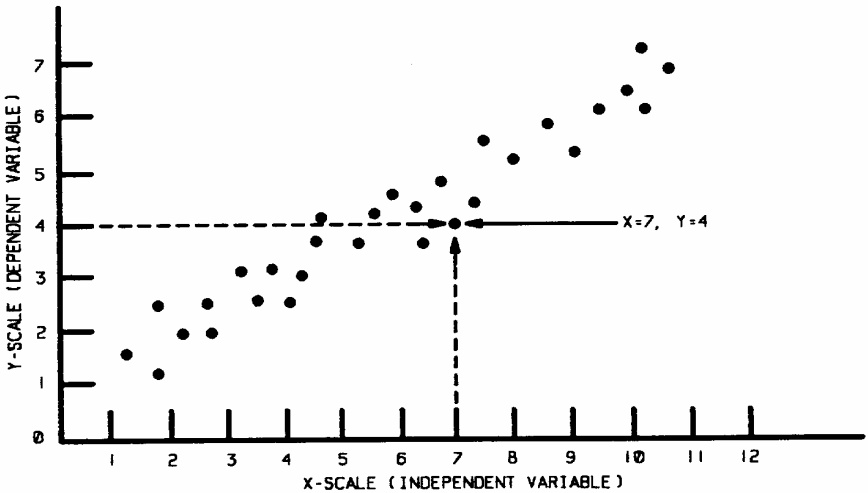
(3) A commonly encountered distribution or scatter of points is illustrated in Figure E-1-9. While most of the plotted points are located within an area that forms a cohesive generalized pattern, some of them lie outside this area, as at (A) in Figure E-1-9. Other points, (B) in Figure E-1-9, while adjacent to or a part of the general pattern area, may be so located as to produce a significant distortion of an otherwise smooth pattern. The values represented by these out-of-pattern points have the effect of reducing the degree of correlation between the two variables; and the more these values deviate from the general pattern of the rest of the data, the greater their adverse effect on the correlation. A material variation in a few values from an otherwise uniform pattern suggests that unusual circumstances might have caused the extreme variation. The presence and nature of these circumstances, however, cannot be assumed. If examination indicates that such variations are characteristic of the data being analyzed and, therefore, pertinent to the analysis, they must be considered in any evaluation and in any estimate or computation. However, if the examination indicates that they are not pertinent to the analysis (for example, they might result from nonrecurring expenditures that have no bearing on the analysis), proper adjustment of the data and the diagram must be made before any further analysis of the data is undertaken or the data evaluated.

(4) Techniques for mathematically evaluating the extent of correlation are described in E-205.

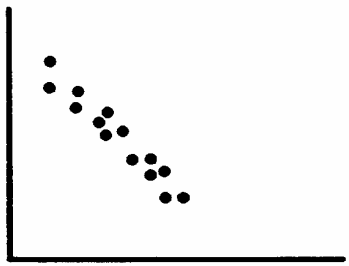
Figures E-1-1 through E-1-3
Scatter Diagrams

Close, Positive (Direct), Linear Correlation

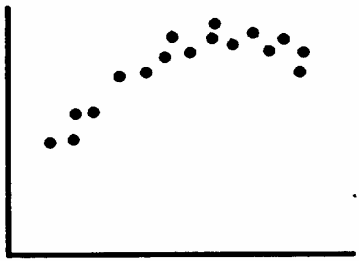
FIGURE E-1-1



CLOSE, NEGATIVE (INVERSE), LINEAR CORRELATION
FIGURE E-1-2



CLOSE, CURVILINEAR CORRELATION
FIGURE E-1-3

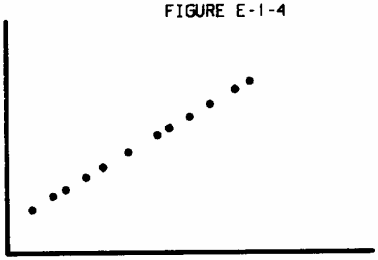


Figures E-1-4 through E-1-9
Scatter Diagrams

Close, Positive (Direct), Linear Correlation

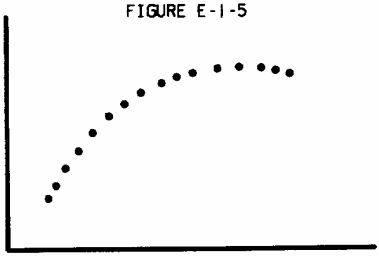
PERFECT, POSITIVE, LINEAR CORRELATION

FIGURE E-1-4



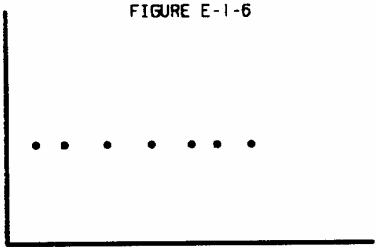
PERFECT, CURVILINEAR CORRELATION

FIGURE E-1-5



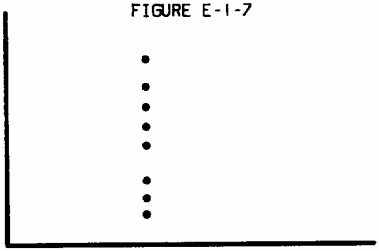
NO CORRELATION. (ONE VALUE OF Y IS ASSOCIATED WITH ALL VALUES OF X)

FIGURE E-1-6



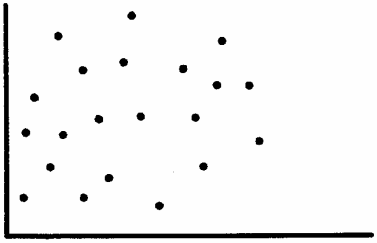
NO CORRELATION. (ONE VALUE OF X IS ASSOCIATED WITH ALL VALUES OF Y)

FIGURE E-1-7



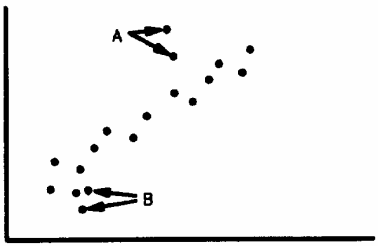
LITTLE OR NO CORRELATION. (NO UNIQUE ASSOCIATION BETWEEN VALUES OF X AND Y)

FIGURE E-1-8



DEGREE OF CORRELATION MATERIALLY REDUCED BY A FEW EXTREME VARIATIONS (A) AND (B)

FIGURE E-1-9



E-104.4 Selection of Appropriate Scales

a. Failure to prepare and to use scatter diagrams in an appropriate manner can lead to formulation of unsound conclusions as to the pattern of the plot points and the degree of correlation between the variables. Some of the relevant factors are discussed below.

b. Ordinarily, arithmetic scales should be used for both the *x* and *y* axes in most cases when constructing scatter diagrams. In some instances, it is desirable to emphasize the rate of change rather than the amount of change. This can be accomplished by using graph paper having a logarithmic scale on one axis (semilog paper) to prepare the scatter diagrams.

c. The scales selected must have a wide enough range to accommodate the lowest and highest values of the data to be plotted. It is good practice to organize the graph so that the plot points cover a fairly wide area. This will facilitate a visual evaluation of whether the points display a trend. Compression of the points into a small space may not convey a correct interpretation of the relationship between the variables. An example of a poor choice of scales is illustrated in Figure E-1-10. The restriction of the plot points to a narrow rectangular area at the bottom of the paper results in a diagram that is difficult to interpret. Figure E-1-11, where the same data have been plotted using an expanded scale for the *y* axis, provides a much better picture of the degree of scatter.

Figure E-1-10
Scatter Diagrams-Illustration of Poor Choice of Scales

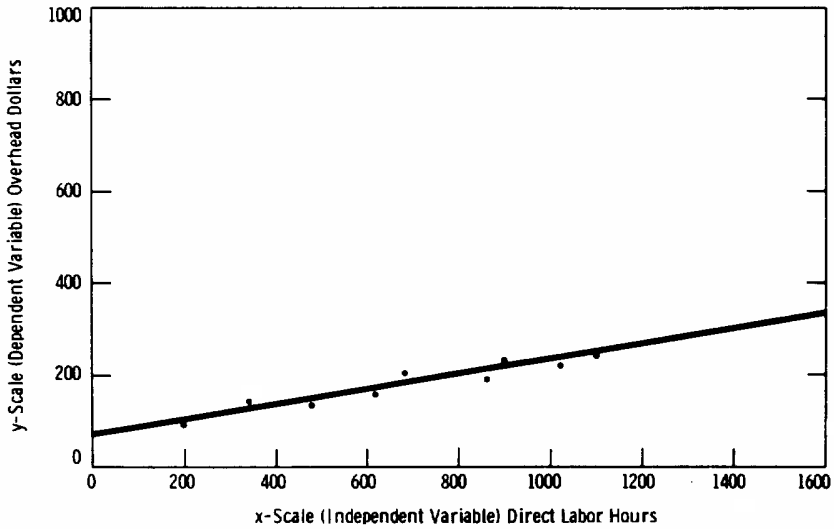
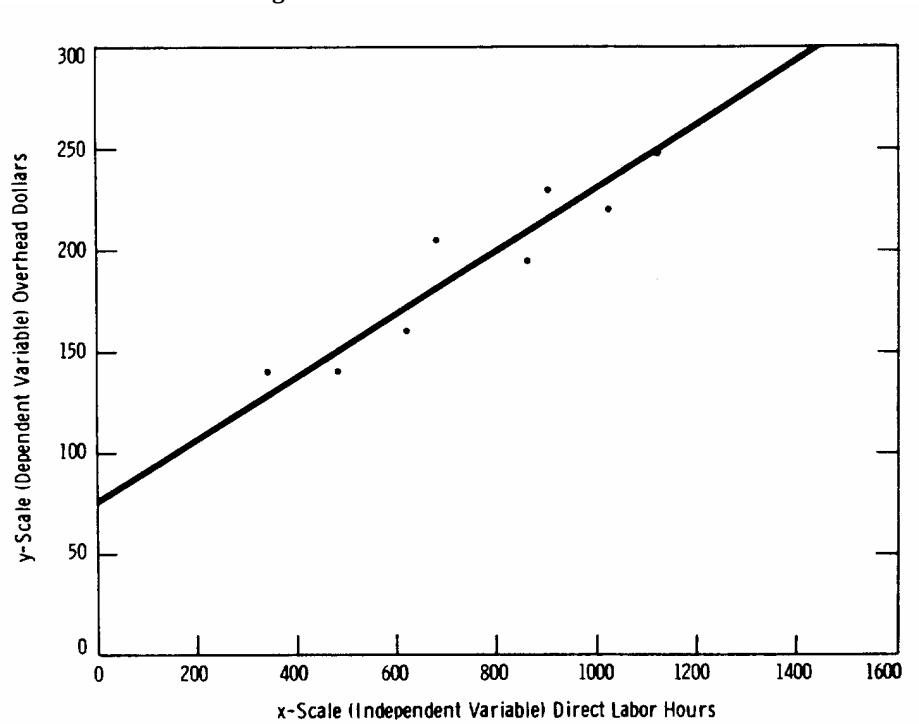


Figure E-1-11
Scatter Diagrams-Illustration of Correct Choice of Scales



d. In some cases, a better picture of the pattern and spread of the plot points can be obtained if the smaller values on one or both scales are not plotted. This will, however, result in a distortion of the range data relative to the total possible range. In addition, where the unit of measure (e.g. dollars) is the same for both the *x* and *y* values, the use of different scales will distort the relative magnitude of the variable. These distortions are not important when the sole purpose of the diagram is to evaluate the pattern of the plot points and degree of correlation of the two variables.

E-104.5 Time Relationships

a. Although no time scale is shown on a scatter diagram, the time element may be introduced by joining, numbering, coding, or otherwise identifying the successive points.

b. Figure E-1-12 is a chart correlating general and administrative expenses to the base, cost input, for a period of 13 quarters. The points have been joined according to their sequence in time. This line shows that the period was one of generally rising production except for two quarters between B and C. During this period of declining business, G&A expenses remained at a higher level relative to cost input than the auditor might expect from observation of the trend established during the first four quarters, A to B. When production again resumed an upward direction, the trend of G&A expenses was parallel to but at some distance above the previous trend. The auditor would want to know why G&A expenses did not decline during the period B to C and why the correlation of these expenses to cost input established a new trend at a significantly higher level than in the earlier months.

c. When used with discretion, indicating time on the diagram can be useful for detecting changes in relationships due to inflation or some other factor that occurs over a period of time and for identifying portions of the data that may require special examination. Because a least-squares line fitted to the same data may not reveal this type of information, it is advisable to use both techniques.

Figure E-1-12
Scatter Diagrams-Illustration of Joining Points by Time Sequence

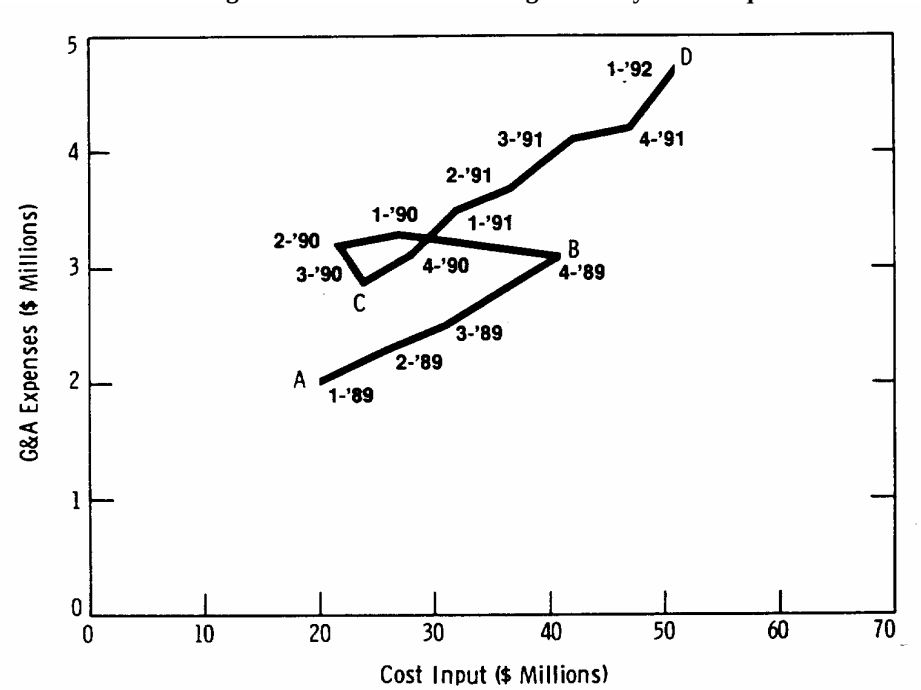
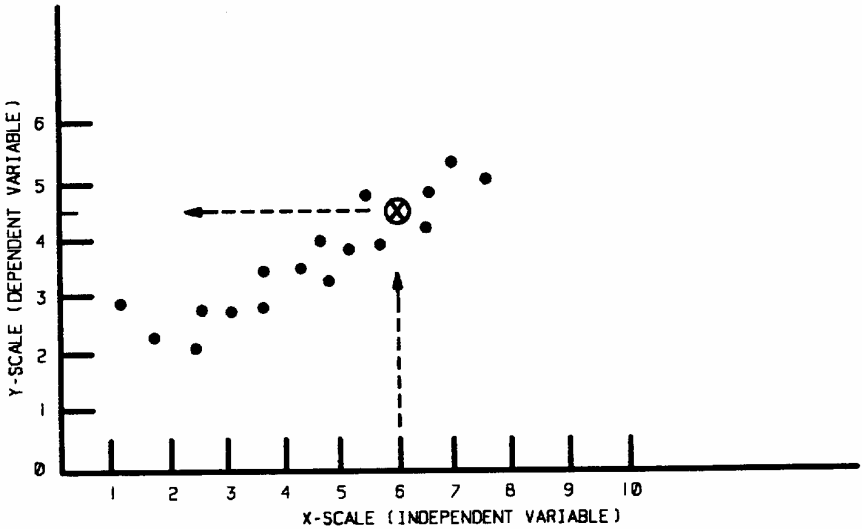


Figure E-1-13
Estimating or Evaluating an Estimate of the Dependent (Y) Value
From a Known Independent (X) Value



GIVEN: AN INDEPENDENT VARIABLE (X) EQUAL TO 6, MOVE VERTICALLY TO THE CENTER OF PATTERN OF POINTS REPRESENTED BY ⊗. FROM ⊗ MOVE HORIZONTALLY TO Y-SCALE AND READ THE SCALE VALUE OF 4½

THUS FOR X=6, Y=4½

E-104.6 Use and Limitations

a. The scatter diagram, without the refinements of computational analysis discussed in subsequent paragraphs, may be adequate when the auditor's purpose is to identify current departures from historical patterns, as discussed in E-402. It also can provide a quick appraisal of the degree of relationship between two variables and the general pattern of this relationship.

b. A rough estimate of the amount of a dependent *y* variable from a given value of the independent *x* variable, or the evaluation of an estimation, may be made directly from the scatter diagram. To accomplish this, first locate, as in Figure E-1-13, the point on the horizontal *x* scale which represents a given value of the independent *x* variable. From this point, move vertically and parallel to the *y*-scale until the approximate center line of the pattern is reached. From this point move horizontally to the *y*-scale and read the indicated value as an estimated value of the dependent *y* variable. Alternatively, a line that follows the central path through the plotted points can be drawn judgmentally through the scatter diagram and used to estimate the amount of the dependent variable corresponding to any given value of the independent variable. The line should be drawn in such a manner that the sum of the vertical distances between the line and the points above the line is approximately equal to the sum of the vertical distances between the line and the points below the line. If a straight line best fits the pattern, this may be accomplished by drawing the line through the plotted mean values of the two variables. The first step is to average separately the *x* values and the *y* values and plot the point corresponding to these averages. The line may then be drawn based on visual observation by placing a ruler through the scatter of points and rotating it on the plot point of the mean values until it falls along a path that appears to provide the best fit.

c. While in many instances the scatter diagram and the approximated line of best fit described in b. above furnish sufficient information for preliminary determinations, their usefulness is limited by the auditor's inability, using only the diagram, to define precisely the location of the path of central tendency or to determine objectively the closeness of the association between the variables. Furthermore, a straight line fitted judgmentally to a scattergram will differ according to the opinion and experience of each individual and can be subject to disagreement as to whether it is properly drawn. E-200 discusses how a unique line of best fit can be fitted to historical data based on a mathematical equation which has been judgmentally determined to express a logical cause and effect relationship.

E-200 Section 2 --- Correlation and Regression Analysis

E-201 Introduction

This section presents a brief discussion of the mathematical procedures used to express relationships between two or more variables, the use of EZ-Quant for Windows (EZ-Quant)to perform the necessary calculations, and other aspects of regression and correlation analysis applicable to the evaluation of contract costs. The mathematics of regression analysis is beyond the scope of this appendix. Auditors who wish to understand the underlying computations should refer to a standard statistics textbook or the EZ-Quant, "Help" documentation for Regression Analysis.

E-202 Simple Linear Regression Analysis

E-202.1 Definition and Concept

a. Paragraph E-104.3a.(2) explained that it is possible to fit either a straight or curved line to a set of data. Different computations are required for each type of line and the type should be selected to express a logical relationship between the variables and follows the central path, or trend pattern, of the data. Many of the relationships, which are of interest to the contract auditor, tend to follow a straight line within the relevant range. This paragraph is limited to fitting a straight line to the data. To apply the described techniques to curvilinear relationships or situations involving more than two variables, see E-203 and E-204.

b. If two variables x and y are associated by a straight-line relationship, the equation expressing that relationship is commonly designated as:

$$y = a + bx$$

where a is the value or point at which the line if it were extended would intersect the vertical y axis and b is the slope of the line, that is, the ratio of the change in the dependent y variable that is associated with a given change in the independent x variable. In other words, b tells how much y changes for a unit change in the value of x. When the sign of b is positive, the line slopes upward, when negative, the line has a downward slope.

The values in a regression equation, which are calculated from the observed data, are referred to as "coefficients" (also termed "parameters"). Thus, a and b are the coefficients of this regression equation because they are calculated based on the observed values of x and y.

c. As indicated in E-104.6c, a freehand straight line, which is judgmentally fitted to a scatter diagram, will vary from individual to individual; different auditors using the same data would draw slightly different lines. However, there can be only one mathematically computed straight line of best fit. Such a line is not influenced by the auditor's judgment. A line so determined is commonly referred to as the simple linear regression line, the straight line of least-squares, the least-squares straight line, the straight line of average relationships, and the straight line of best fit. All of these names are descriptive of the characteristics of the unique line which satisfies the following requirement: the sum of the squares of the vertical distances from each point to the line is less than any other straight line. In order to obtain the equation of the line which meets this requirement, it is necessary to calculate values for the coefficients a and b so that the sum of the squares of the differences between the actual observed values of the independent variable y and the corresponding values calculated from the equation shown in paragraph b. is minimized.

d. The use of the least-squares principle is based on complex statistical concepts. Essentially, the principle produces a mathematical line of best fit that minimizes the sum of the squares of the differences between a series of numbers and their mean. Figures E-2-1 and E-2-3 show simple linear regression lines fitted to data points. Figure E-2-3 is particularly helpful in visualizing the following characteristics of such a line:

(1) The sum of the squares of the vertical distances between the regression line and the points is less than for any other straight line.

(2) The sum of the distances above the line equals the sum of the distances below the line.

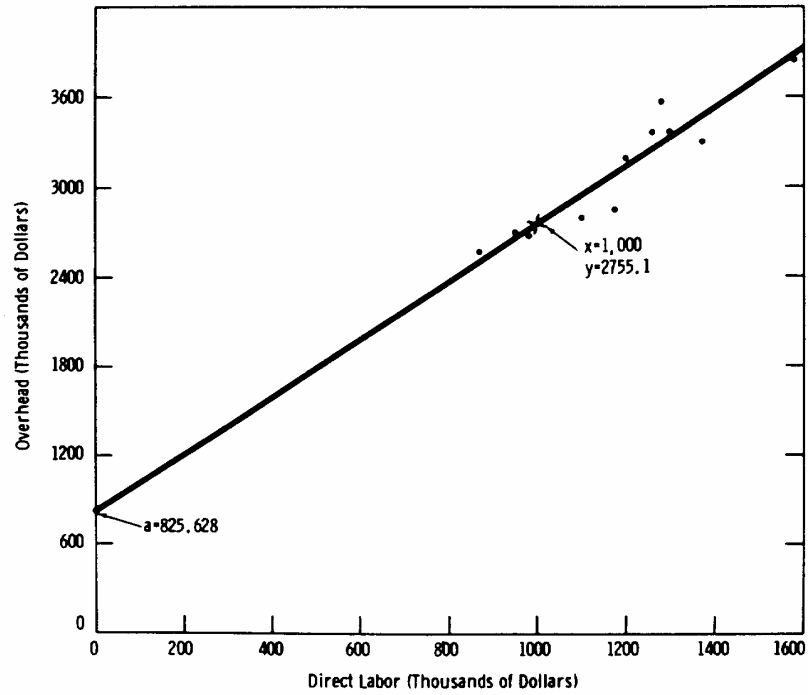
(3) The line passes through the point corresponding to the mean of the observed values of the *x* variable and the mean of the observed values of the *y* variable. The least-squares equations for the coefficients *a* and *b* are referenced in EZ-Quant, Help documentation for Regression Analysis.

e. In addition to the values of *a* and *b,* it is necessary to calculate other statistics which measure the degree of closeness of the fit of the calculated relationship between the dependent and independent variables. The statistics generally used for this purpose are the coefficient (or index) of correlation, denoted as *r*, and the coefficient (or index) of determination, denoted as r^2. Equations for both the coefficient of correlation and coefficient of determination are referenced in EZ-Quant, Help documentation for Regression Analysis.

f. The coefficient of determination (r^2) may be calculated by simply squaring the coefficient of correlation (*r*). The possible values of *r* range from -1 (perfect negative correlation) to 1 (perfect positive correlation) and the values of r^2 range from 0 (no correlation) to 1 (perfect correlation). Further information on the interpretation of these statistics is contained in E-205.

g. A considerable savings in auditor time, together with a greater assurance of accuracy, can be realized via the use of Agency software, such as EZ-Quant, to perform the calculations described in the preceding subparagraphs. Also, the use of Agency software provides the auditor with the ability to specifically document and explain the calculation methodologies employed. Further information on EZ-Quant is contained in E-202.2.

Figure E-2-1
Graph of Simple Linear Regression Line



E-202.2 Example of Computerized Regression

a. The EZ-Quant software is available to perform the regression analysis computations. EZ-Quant is described, operationally and with some theoretical treatment, in its "Help" documentation.

b. Figure E-2-2 contains the data used by the regression software to fit a line to the data shown below and plotted in Figure E-2-1. Figure E-2-2 shows only the most prominent output features of the EZ-Quant linear regression option.

Year	Quarter	Direct Labor X	Overhead Expense Y
			(in thousands)
1997	1	$1,203	$3,212
	2	1,304	3,367
	3	872	2,573
	4	1,104	2,804
1998	1	1,176	2,873
	2	1,258	3,376
	3	983	2,701
	4	1,283	3,572
1999	1	1,576	3,862
	2	1,372	3,303
	3	956	2,678
	4	957	2,684
Σ or	Total	$14,044	$37,005

c. The first part of Figure E-2-2 shows how the data file looks when it is ready for regression processing. The rest of the figure shows the regression output from the linear equation regression option. The first output table, Data Summary and Correlation Coefficients, depicts the linear correlation between pairs of variables.

d. The second output table, The Fitted Equation, lists the coefficients that were estimated by the regression analysis, the associated "*t*" value for the *b* coefficient, the "inclusion assurance" for the *b* coefficient, and the "comparison assurance" for the regression equation. The *t* values are computed for each independent variable coefficient to answer the question, "Is the appearance of a relationship (the value of the coefficient) simply a random event, or is there reason to believe that a relationship does exist between the dependent variable and this independent variable?" In this example, the computed "*t*" value corresponds to an inclusion assurance of 99.9 percent for the *b* coefficient. For each independent variable coefficient, the inclusion assurance is the assurance (or confidence) that the regression equation is a better predictor with (rather than without) that particular variable. The Help documentation for Regression Analysis in EZ-Quant provides additional guidance on acceptable levels for inclusion assurance values. The other calculated statistic, comparison assurance, is discussed in E-205.2c.

e. The third output table, Actual vs. Calculated Dependent Variable Values, compares the calculated and actual values for the dependent variable for each case (or observation). This table can assist the auditor in identifying out-of-pattern plot points which should be analyzed in detail with greater scrutiny. It also assists in identifying significant "runs" of points above and below the regression line, as described in E-208f.

f. The last output table, Projecting with the Regression Equation, shows how projections can be made using the regression equation. A projected value of the dependent variable (DVAR) will be calculated for each independent variable (IVAR) value entered by the auditor.

g. To manually plot the regression line, the value of *a* (i.e., 825.6276 which corresponds to $x = 0$) and any pair of *x* and calculated *y* values (for example, $x = 1203$ and $y = 3146.78$) are sufficient to determine the position of the regression line. Alternatively, for ease in plotting, the auditor may prefer to use a rounded value for the *x* variable, such as 1000. In Figure E-2-1 the auditor followed this procedure and found that when *x* is 1000, the projected *y* value is 2755.10. This pair of values as shown in Figure E-2-2 is then used for plotting the regression line. It is good practice to check the accuracy of a manually plotted line by verifying that a third pair of *x* and calculated *y* values falls on the line. When using computer software to generate a line of best fit, verifying the accuracy of the input data is sufficient to validate the line's accuracy.

Figure E-2-2
EZ-Quant Regression

Regression Data

Overhead (1000s of $)	Direct Labor (1000s of $)
3,212	1,203
3,367	1,304
2,573	872
2,804	1,104
2,873	1,176
3,376	1,258
2,701	983
3,572	1,283
3,862	1,576
3,303	1,372
2,678	956
2,684	957

Data Summary and Correlation Coefficients

Data Summary				
12	Number of Observations			
	Average Value	Minimum Value	Maximum Value	Standard Deviation
Overhead (1000s of $)	3,083.75	2,573.00	3,862.00	418.731
Direct Labor (1000s of $)	1,170.33	872	1,576.00	205.499
Correlation Coefficients Between Variable Pairs				
	Dependent Variable (Y)	Independent Variable 1		
	Overhead (1000s of $)	Direct Labor (1000s of $)		
Overhead (1000s of $)	1			
Direct Labor (1000s of $)	0.946917	1		

The Fitted Equation
$Y = a + b(X_1)$

Coefficients Estimated by Regression:				
	Variable	Coefficient Value	Computed T-Value	Inclusion Assurance (%)
a	Intercept	825.6276174	3.3585	99.3
b	X_1 is Direct Labor (1000s of $)	1.92946942	9.3145	99.9
Summary Statistics:				
	R squared	Regression Std Error		
	0.896651	141.183604		
	Comparison Assurance (%)	F Statistic	Regression Degrees of Freedom	Residual Degrees of Freedom
	99.9	86.7597	1	10

Actual vs. Calculated Dependent Variable Values for
$Y = a + b(X_1)$

Data Item	Actual Value	Calculated Value	Difference (Actual-Calc'd)	% Diff (Diff/Actual)
1	3,212.00	3,146.78	65.22	2
2	3,367.00	3,341.66	25.34	0.8
3	2,573.00	2,508.12	64.88	2.5
4	2,804.00	2,955.76	-151.76	-5.4
5	2,873.00	3,094.68	-221.68	-7.7
6	3,376.00	3,252.90	123.1	3.6
7	2,701.00	2,722.30	-21.3	-0.8
8	3,572.00	3,301.14	270.86	7.6
9	3,862.00	3,866.47	-4.47	-0.1
10	3,303.00	3,472.86	-169.86	-5.1
11	2,678.00	2,670.20	7.8	0.3
12	2,684.00	2,672.13	11.87	0.4

Projecting With the Regression Equation $Y = a + b(X_1)$

	Independent Variable	Coeficient Value	Independent Variable Value	Contribution to Projection
a	Intercept	825.62762		825.63
b	X_1 is Direct Labor (1000s of $)	1.92947	1,000.00	1,929.47
	Projection:			
	Y is Overhead (1000s of $)			2,755.10

E-203 Curvilinear Regression Analysis

a. The same least-squares criterion which is used to obtain a straight line of best fit, as described in E-202, can be applied to more complicated equations. The equation used should be the one which has been judgmentally determined to reflect a logical relationship between the variables. Since these equations can be represented by a curved line on a scattergram, the process of mathematically fitting this line to data is referred to as curvilinear regression analysis. There are an infinite number of non-linear equations; however, only a few of them have been found to have application to contract cost audits.

b. The curvilinear models with greatest applicability to contract audit are improvement curves. Comprehensive guidance on the use of improvement curves is provided in Appendix F.

c. Another curve with applicability to certain business data is expressed by the simple exponential or compound interest equation.

$$y = a\,(1+r)^x$$

Several types of business data tend to follow such a curve. For example:

(1) where y is the principal plus interest of an amount a, compounded x times at a fixed rate r, or

(2) where y is the remaining book value of an asset purchased for an amount a, and depreciated x years by the declining balance method with rate r, or

(3) where y is a price level or wage level which has increased (or is expected to increase) from a level a, at a constant rate r for x periods of time, or

(4) where y is the price of a new product which has decreased (or is expected to decrease) from a level a, at a constant rate r for x periods.

It should be noted that in fitting a least-squares line to a time series, as suggested by (3) and (4), a thorough understanding of the caveats discussed in E-310 is essential.

d. An exponential curve appears as a straight line on semi-logarithmic graph paper. The exponential curve type in the regression analysis option of EZ-Quant can be used to fit this type of curve to historical data. An application of this software for curvilinear analysis is discussed in E-311. Figure E-3-3 shows the important output features of the regression analysis option of EZ-Quant. In this figure, the equation $y = a\cdot\cdot b^x$ at first glance appears to be unlike the expression given above in E-203.c. That expression, when written in the form shown on the output, would be $y = a\cdot\cdot(1+r)^x$. Since b is equal to $1 + r$, the expressions are functionally the same.

E-204 Multiple Regression Analysis

a. Multiple regression analysis is concerned with evaluating the relationship between a dependent variable and two or more independent variables, and is used in those applications where the regression of y on a single independent variable is found to be inadequate. It employs the least-squares method to determine the combined effect of the independent variables on the dependent variable. Since more than two variables are involved, this relationship cannot be plotted graphically; however, the same principles and techniques are involved as in simple regression analysis.

b. The equation used in most calculations of multiple regression analyses takes the form of

$$y = a + bx_1 + cx_2\ .$$

where y is the dependent variable,

x_1 is the first independent variable,

x_2 is the second, and so on.

The letters a, b, c . . . stand for constants (coefficients) which minimize the sum of the squares of differences between the actual values of y and values calculated from the equation. In addition to calculating the values of these coefficients, it is necessary to calculate a statistic which measures the closeness of the relationship between the dependent variable and the independent variables. This statistic, called the coefficient of determination (r^2), is discussed in E-205.

c. The computations required in multiple regression analysis are so laborious that manually fitting just one least-squares equation can be prohibitively time-consuming. However, the computer performs the computations so rapidly that the auditor can experiment with various combinations of independent variables until he or she finds the combination which seems to provide the best prediction of the

dependent variable in which he is interested.

d. The regression analysis option of EZ-Quant will perform a multiple regression on input data files that have two or more independent variables. Additional discussion of the multiple regression option is available in the EZ-Quant, Help documentation for Regression Analysis.

E-205 Correlation Analysis

As explained in the preceding paragraphs of this section, regression analysis is concerned with obtaining equations which express functional relationships among variables. Correlation analysis is concerned with evaluating how closely the variables are related.

E-205.1 Interpretation of Coefficient of Determination

a. The statistic most widely available and commonly used to evaluate how well an equation fits available data is the coefficient of determination, denoted r^2. The coefficient of determination quantifies how much of the change in the dependent variable is "explained" by the change in the independent variable(s). In simple regression, where there is only one independent variable, the r^2 indicates how much of the change in y is due to the change in x. In multiple regression, where there can be multiple independent variables (up to five in EZ-Quant), the r^2 indicates how much of the change in y is due to the changes in x_1,x_2, x_3, x_4, and x_5.

b. In simple linear regression, the coefficient of determination measures how much closer the plot points (historical values of y) are to the regression line than they are to a line drawn horizontally through the average of the y values. This is illustrated graphically in Figure E-2-3.

c. The coefficient of determination is used to evaluate the relative efficiency of predicting future values of y based on the regression equation, as opposed to predicting that future values of y will equal the average value of y in the historical data. The value of r^2 is not indicative of how much better (or worse) predictions based on the regression line will be compared with predictions based on, for example, a judgmental evaluation or the contractor's most recent experience.

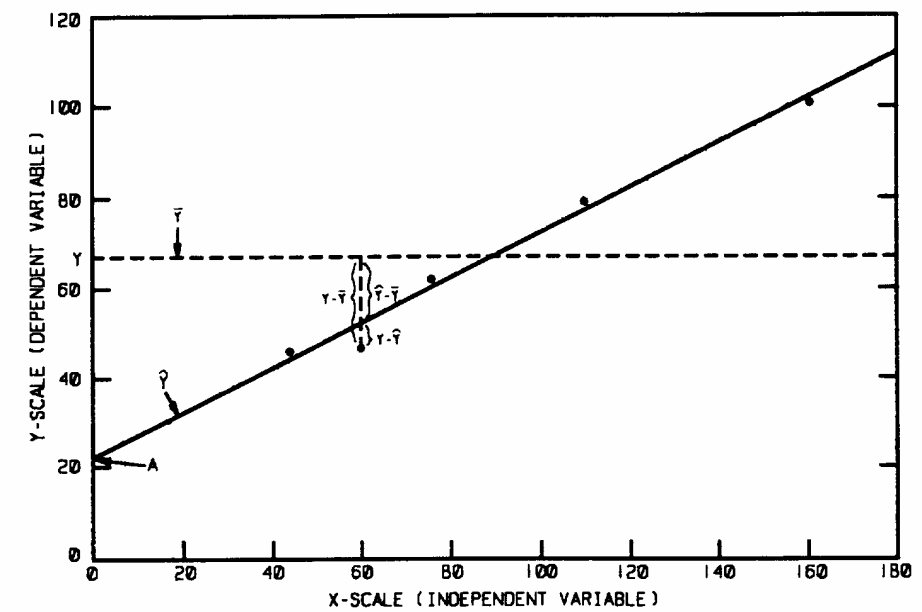
d. Even if the historical observations vary significantly about the regression line, a high coefficient of determination will be obtained if the variations about the mean are much wider. On the other hand, if the independent variable has remained fairly constant during the period covered by the observations, a low coefficient of determination can be obtained even though all observations are close to the regression line.

e. If all of the observed values of y fall on the regression line, r^2 will equal one, indicating perfect correlation. This result will always be obtained if the number of observations equals the number of coefficients in the regression equation (two in the case of simple linear regression); however, it will rarely occur otherwise. On the other hand, if there is no better fit to the historical data than the mean of the y values, then r^2 will equal zero. This result, which would produce a perfectly horizontal regression line in simple linear regression, is also very unlikely, regardless of whether or not the variables are correlated.

f. The principal problem encountered in the use of r^2 derives from the fact that if the number of observations is small, regression analysis may produce a high coefficient of determination even if the variables are unrelated. Instances have been noted where auditors have cited a high coefficient of determination as evidence of a good fit when, in fact, the coefficient of determination provides no evidence of correlation and could easily have occurred by chance. When there are only a few plot points on a scatter diagram, it will usually be possible to draw a straight line at an angle which fits the points much better than a horizontal line. Consequently, a high coefficient of determination is nearly always obtained and the only use that can be made of the coefficient of determination is to test whether it is high enough to provide statistical support for an assumption that the variables are actually correlated (E-205.2). On the other hand, the larger the number of observations used, the more remote is the possibility of a high coefficient of determi-

nation occurring by chance; the coefficient of determination can then be regarded as a measure of the extent to which changes in the *y* values can be explained by changes in the *x* values (E-205.3).

Figure E-2-3
Graph Illustrating Computation of Index of Determination (R2)



E-205.2 Determining the Existence of Correlation

a. Because a coefficient of determination (r^2) may be calculated for any data set, it is important that the value of r^2 is large enough to make it probable that the changes in the dependent variable are caused by changes in the independent variable. The reliability of a regression equation for making predictions must be evaluated by the auditor. Table E-2-1 or the "comparison assurance" reported by EZ-Quant can be used to evaluate the calculated coefficients of determination.

b. Table E-2-1 gives significant values of r^2 at the 75 percent and 95 percent confidence levels. To test the significance of the value of r^2 obtained from a regression analysis, the auditor selects the column corresponding to the number of coefficients in the regression equation and the row corresponding to the number of observations in the historical data. In simple linear regression, there are two coefficients (*a* and *b*). Consequently, in order to test the significance of the .9825 value of r^2 obtained in a simple regression of data with five observations at the 95 percent confidence level, the auditor would compare it with .772 obtained from Table E-2-1. The table indicates that if the *x* and *y* variables were unrelated, there would be a 95 percent probability that the coefficient of determination would in turn be less than .772 and a 5 percent probability that it would, by chance, be greater than .772. Since .9825 is significantly larger than .772, the auditor could conclude that the probability of the *x* and *y* values being correlated is much greater than 95 percent.

c. Multiple regression analysis employs equations with three or more coefficients. For example, the model $y = a + bx_1 + cx_2 + dx_3$ has four coefficients (a, b, c, and d). If there were 16 observations, the appropriate value from Table E-2-1 for testing the coefficient of determination would be .466 at a 95 percent confident level. If the coefficient of determination exceeds .466, the auditor could conclude that there is more than 95 percent probability that the three independent variables (x_1, x_2, and x_3), as a group, could influence the values obtained for y.

d. EZ-Quant reports a statistic, the comparison assurance, for all of the regression analysis curve options and for all improvement curve regression options (Appendix F). The comparison assurance is the level of confidence (assurance) associated with the value of the coefficient of determination (r^2). Its function is similar to that of Table E-2-1 which helps the auditor decide how much reliance to place on the estimated regression equation. The difference between the two reference values is that Table E-2-1 shows the r^2 value needed to achieve a specific level of confidence while the comparison assurance indicates the level of confidence achieved by the r^2 value produced by the regression. A further discussion of comparison assurance is available in the EZ-Quant, Help documentation for Regression Analysis.

e. It should be clearly understood that simply because a regression analysis passes the test described in this section, it does not follow that the regression equation will provide the best possible estimates of future y values, or even that the estimates will be reasonably close. It merely means that a statistical analysis of the historical data identifies the probability that the equation will provide better estimates than could be obtained by merely averaging the historical values of y.

f. The auditor should also bear in mind that the test considers only the mathematical relationship between the variables. It does not take into account the audit significance of the relationship expressed in the regression equation. In this connection, the logic of the relationship is most important. If the regression equation expresses a relationship which can be strongly defended by logical arguments, and only a few observations are available, some significance can be attached to the analysis even if the test does not provide conclusive statistical evidence that the values are correlated. However, a special effort should be made in such cases to determine why the correlation is not close and whether the factors causing the observed variations or any other unusual conditions are likely to be present and materially affect the costs during the forecast period.

E-205.3 Evaluating the Extent of Correlation

a. The larger the number of observations that are included in a regression analysis, the smaller the significant values shown in Table E-2-1 become and the more remote the chance of a high coefficient of determination occurring by accident becomes. Consequently, the coefficient of determination can be regarded as a measure of the proportion or percentage of the total variance in the dependent variable(s). The larger the coefficient of determination, the smaller the proportion of the variance that is attributed to other influences. For instance, if the coefficient of determination obtained from fitting a regression equation is .92, it may be concluded that approximately 92 percent of the variance in the dependent variable is associated with changes in the independent variable(s) and 8 percent of the variance is associated with chance or other influences.

b. Even though the coefficient of determination may be large enough to provide ample assurance that the variables are related, it may be too small for the auditor's purposes. Assume, for example, that a simple linear regression analysis based on 20 observations produced a coefficient of determination of .3. Reference to Table E-2-1 indicates that this coefficient of determination provides substantially greater than 95 percent assurance that the variables are related. However, since such a small proportion of the variation in the y variable is explained by variation in the x variable, the auditor would generally seek a more reliable basis for predicting y. He or she may be able to identify the reasons for substantial deviations from the regression line and improve

the fit by appropriate adjustments to the data. Alternatively, he may be able to identify and add to the regression equation another independent variable which is influencing the dependent variable, or he may consider methods of prediction other than regression analysis.

E-205.4 Comparison of Two Analyses

a. The coefficient of determination can also be used to determine which of two different equations (or independent variables) provides the better fit to historical data. A direct comparison between the values of r^2 obtained in two different regression analyses is valid only if both analyses employ the same observations and the same number of coefficients. Such a comparison provides an objective means of choosing between two equally logical equations. The greater the number of observations used and the more substantial the differences between the values of r^2, the more reliance it is possible to place on such comparisons.

b. In assessing the relative efficiency of two regression equations in explaining changes in the value of a dependent variable, the ratio of the values for $1 - r^2$ is more important than the difference between the values obtained for r^2. The value of $1 - r^2$, called the "coefficient of nondetermination," measures the extent to which variations in the historical values of the dependent variables are not explained by a regression equation. The greater the ratio of the larger value of $1 - r^2$ to the smaller value of $1 - r^2$, the greater is the confidence that can be placed on the superiority of the analysis which produced the larger value of r^2 (and hence the smaller value of $1 - r^2$). However, when the number of observations is small, the value of $1 - r^2$ obtained from one analysis can be several times larger than the value obtained from another analysis, without providing conclusive evidence of the superiority of the second analysis.

Table E-2-1
Table of Significant Values of the
Coefficient of Determination (R2)

NUMBER OF OBSER-VATIONS	AT THE 95% CONFIDENCE LEVEL					NUMBER OF OBSERVATIONS	AT THE 75% CONFIDENCE LEVEL				
	NUMBER OF COEFFICIENTS *						NUMBER OF COEFFICIENTS *				
	2	3	4	5	6		2	3	4	5	6
3	.994					3	.854				
4	.903	.998				4	.562	.882			
5	.772	.951	.999			5	.403	.667	.891		
6	.659	.865	.967	.999		6	.311	.533	.703	.896	
7	.57	.777	.903	.975	1	7	.253	.444	.586	.721	.898
8	.5	.699	.832	.925	.98	8	.213	.382	.506	.614	.732
9	.445	.632	.765	.865	.938	9	.184	.335	.447	.541	.632
10	.4	.576	.705	.806	.887	10	.161	.298	.401	.486	.564
11	.363	.528	.651	.752	.835	11	.144	.269	.364	.443	.513
12	.332	.487	.605	.702	.786	12	.13	.245	.334	.407	.472
13	.306	.451	.563	.658	.74	13	.118	.225	.308	.377	.437
14	.284	.42	.527	.618	.698	14	.109	.208	.286	.351	.409
15	.265	.394	.495	.582	.66	15	.1	.193	.267	.329	.383
16	.248	.37	.466	.55	.625	16	.093	.181	.251	.31	.362
17	.233	.349	.441	.521	.593	17	.087	.17	.236	.292	.342
18	.22	.33	.418	.495	.565	18	.082	.16	.223	.277	.325
19	.208	.313	.397	.471	.539	19	.077	.151	.212	.263	.309
20	.197	.297	.378	.45	.514	20	.073	.143	.201	.251	.295
21	.188	.283	.361	.43	.492	21	.069	.136	.192	.24	.282
22	.179	.271	.345	.411	.472	22	.066	.13	.183	.229	.27
23	.171	.259	.331	.395	.453	23	.062	.124	.175	.22	.26
24	.164	.249	.318	.38	.435	24	.06	.119	.168	.211	.25
25	.157	.239	.305	.365	.419	25	.057	.114	.161	.203	.241
26	.151	.23	.294	.352	.404	26	.055	.109	.155	.196	.232
27	.146	.221	.284	.339	.39	27	.053	.105	.15	.189	.224
28	.14	.213	.274	.328	.377	28	.051	.101	.144	.182	.217
29	.135	.206	.265	.317	.365	29	.049	.098	.139	.176	.21
30	.131	.199	.256	.307	.354	30	.047	.094	.135	.171	.203
31	.127	.193	.248	.297	.343	31	.045	.091	.131	.165	.197
32	.123	.187	.241	.288	.333	32	.044	.088	.126	.16	.191

* Note: The Number Of Coefficients column to use is the number of independent variables plus one, if the intercept was computed (as is typically the case).

E-206 Prediction Intervals

a. In addition to the coefficient of determination, several other statistics are sometimes used to measure the extent of the historical correlation between the variables and the reliability which can be placed on projected values computed from the regression estimate. One such statistic is the "prediction interval." A confidence interval at any specified confidence level can be calculated for a prediction of the independent variable obtained from a regression equation. This interval is included in the output of many standard regression software packages. The mathematical equation used to calculate the interval assumes that (1) the relationships between the variables which existed in the period covered by the historical data will continue into the future and (2) the predicted values of the independent variable are accurate. For the reasons discussed below, prediction intervals have limited application to contract audit work.

b. Prediction intervals and confidence intervals (discussed in B-204 and B-205) are similar, but there are two important differences between them. First, the confidence intervals computed in sampling applications relate to the precision of the single estimate of the mean or, when the mean is projected, the corresponding single estimate of the total value of all items in the universe. A prediction interval in regression analysis relates to the precision of the regression estimate of the value of the dependent variable that is associated with given values for each of the independent variables. Since the independent variable values can vary without limit, there are an unlimited number of values that the dependent variable might take, hence the number of prediction intervals is unlimited. Each prediction interval is unique. Second, the auditor can reduce the range of a confidence interval by increasing the sample size, but cannot reduce a prediction interval if all pertinent historical data has been properly used. As a result, while a confidence interval can help the auditor determine the adequacy of the sample size, a prediction interval in regression analysis does not serve a similar purpose.

E-207 Adjustments for Economic Factors in Regression Analysis

Many contract audit applications of regression analysis include variables which are affected by changes in wage and price levels. When economic changes have significantly affected any of the variables during the period covered by the historical data, the regression analysis applied to the raw data will not produce reliable results. In such cases it is necessary to (1) include a measure of economic change as a separate explanatory variable in multiple regression or (2) adjust the data to eliminate the effects of the economic changes.

a. One method is to use multiple regression with an economic index as one of the independent variables. For example, to measure the effect of direct labor volume on labor costs, changes in costs resulting from economic factors (such as cost-of-living allowances) can be considered by applying the following multiple regression equation:

$$y = a + bx_1 + cx_2$$

where y is the average labor rate; x_1 is the number of employees; x_2 is an economic wage index; and a, b, and c are coefficients (constants) which minimize the sum of the squares of the differences between the actual values of y and values calculated from the equation.

(1) Economic indexes that can be used as an independent variable in the multiple regression are discussed in DCAAP 7641.74, Use of Economic Indexes in Contract Audits. Salary and wage changes generally include both inflation (cost-of-living) as well as the effects of other economic changes. The most precise cost-of-living adjustments can be made using wage-related economic changes specific to a particular contractor --- union agreements or other labor records. However, if such agreements or records do not exist or it is too time consuming to construct the data, economic indexes should be used to make the adjustments. If wages are to be adjusted to account for inflation alone, the Consumer Price Index, Urban Wage Earners and Clerical Workers (CPIW), should be used. However, if labor rates are adjusted for all wage-related economic changes

(including inflation), wage indexes should be used. DCAAP 7641.74 provides a discussion of the types of indexes available to make such adjustments.

(2) EZ-Quant reports several statistics that are used to judge the adequacy of the regression equation that was fitted to unadjusted data. When accepted by the auditor, the multiple regression equation can be used to predict future costs.

b. A second way to perform a valid regression analysis is to adjust the cost data to eliminate the effects of economic changes. For example, to measure the effect of direct labor volume on labor costs, changes in costs resulting from economic factors (such as cost-of-living allowances) must be removed from the data. This is accomplished by removing the effects of the changes from all data during the historical period, thus placing all amounts on an economic level comparable to the earliest observation in the historical data. Alternatively, all of the data may be updated to the current economic level or adjusted to some other base period. Because this example has one independent variable (i.e., number of employees), the simple (two-variable) regression curve options of EZ-Quant can be used to fit a regression equation to the adjusted cost data.

(1) When adjusting an individual cost element or a homogenous grouping of similar costs, an economic index (commodity, industry, or category) that is the most appropriate disaggregate index available should be used. When adjusting a group of costs, either a composite index should be developed for the group, or the most representative disaggregate index of the group of costs should be used. The method for placing costs on a common base is discussed in DCAAP 7641.74.

(2) When the regression equation is used to predict future costs, the anticipated effects of economic changes between the base period and the prediction period must be considered. Economic data are used to adjust the results of the regression so that realistic future costs can be estimated. Predictions of future economic conditions, particularly long-range, can best be made by qualified economists. DCAAP 7641.74, provides information on the forecasting services purchased by DCAA, including the list of the indexes readily available to all auditors through DCAA's Intranet by selecting File Libraries on the Home page.

E-208 Other Considerations in Using Regression Analysis

a. Changes in facilities, production methods, and accounting procedures generally limit the period of time over which consistent cost data can be obtained without extensive adjustments. Accordingly, the historical data included in a regression analysis should normally cover this entire period. In addition, the use of quarterly or monthly data in performing regression analyses is usually favored over the use of annual data, in order to provide more plot points. The auditor's selection of an appropriate period for the development of historical cost data will, however, be governed to some extent by the contractor's operations and accounting methods. It may be necessary to use annual data if extensive adjustments to quarterly or monthly data would be required to compensate for accruals and seasonal fluctuations in activity or costs. When periods of a month or less are used, it may be necessary to adjust the data for the number of working days or hours, as discussed in E-309.

b. It is usually necessary to purify historical costs and bases to some extent. Adjustments for economic factors are described in E-207. Other adjustments may be required to eliminate unusual or nonrecurring costs (e.g., costs incurred during interruptions in normal activity because of strikes, floods, or fires), to compensate for accounting system changes, and to effect transfers of costs recorded in a period to which they do not apply.

c. It is essential that the equation used in a regression analysis represent the relationship between the variables. When, for example, the auditor uses the equation $y = a + bx$, he or she must first establish that a straight line best describes the relationship between the two variables. This step will require a scrutiny of the scatter diagram, consideration of how changes in the x variable can logically be expected to influence the value of the y variable, and determination if a logical meaning can be ascribed to each of the a and b coefficients.

d. Sometimes the sign of a coefficient obtained from regression analysis is the opposite from that which could logically be expected. For example, in using the equation $y = a + bx$ to evaluate the relationship between direct labor x and overhead y, it would normally be expected that both the a and b coefficients would be positive. A negative value for b would indicate that increases in direct labor are accompanied by decreases in overhead. More commonly, a negative value of a would indicate a negative amount of fixed overhead. Under these conditions the auditor should reexamine the logic of the association between the variables and determine that the equation used expresses a valid relationship between the x and y values. The auditor should satisfy him or herself as to the acceptability of the basic data used in the correlation. In some instances the data may require further purification as described in paragraph E-208b. When the data appear to be correct but result in an illogical negative a value, consideration should be given to forecasting the value of y by some other method, such as the average or most recent ratio of the y values to the x values.

e. In the application of statistical methods to regression analysis, there is a fundamental assumption that the distance of each y value from the underlying regression line is independent of the other y values. If the y plot points are, for example, moving averages or cumulative averages, this assumption is not met. Consequently, a value of r^2 . obtained in fitting a least-squares line to such data has little or no meaning and cannot be tested for significance as described in E-205.2.

f. The auditor should be aware of the significance of a "run of points" in the data. A run consisting of a long series of points which are all above or below the regression line may occur when the historical data are arranged chronologically or in order of increasing values of the independent variable. The existence of such runs may indicate that (1) some factor not considered in the regression equation is influencing the dependent variable, (2) the equation being used in the analysis does not fully describe all the underlying relationships between the variables, and/or (3) the data do not satisfy the assumption of independence cited in paragraph e. Suppose, for example, that all of the earlier values of the dependent variable are below the regression line and all of the later values are above the line. This condition indicates that a general rise in price levels, or some other factor not considered in the regression equation, may have caused the dependent variable to increase over the period of time covered by the analysis. As an example of the second condition, suppose that all of the small and large values of the independent variable are below the straight line of best fit, while those for intermediate values are above the line. This condition could indicate the existence of a curvilinear relationship between the two variables.

g. Predictions made from a regression equation will be most reliable (precise) when the independent variable values are close to their average values in the data. In the case of simple linear regression, for example, the regression line passes through the point corresponding to the average observed values of the x and y variables. Small errors in the slope of the line, caused by random fluctuations in the data, can result in large errors in predicted values far away from this point. When the predicted values of the independent variable(s) are outside the range of the historical data, particular care must be exercised. A regression equation which does not represent the true underlying relationship between the variables can provide a reasonably good fit to historical data within a limited relevant range, but result in grossly erroneous predictions when the independent variable is outside that range.

E-300 Section 3 --- Time Series Charts

E-301 Introduction

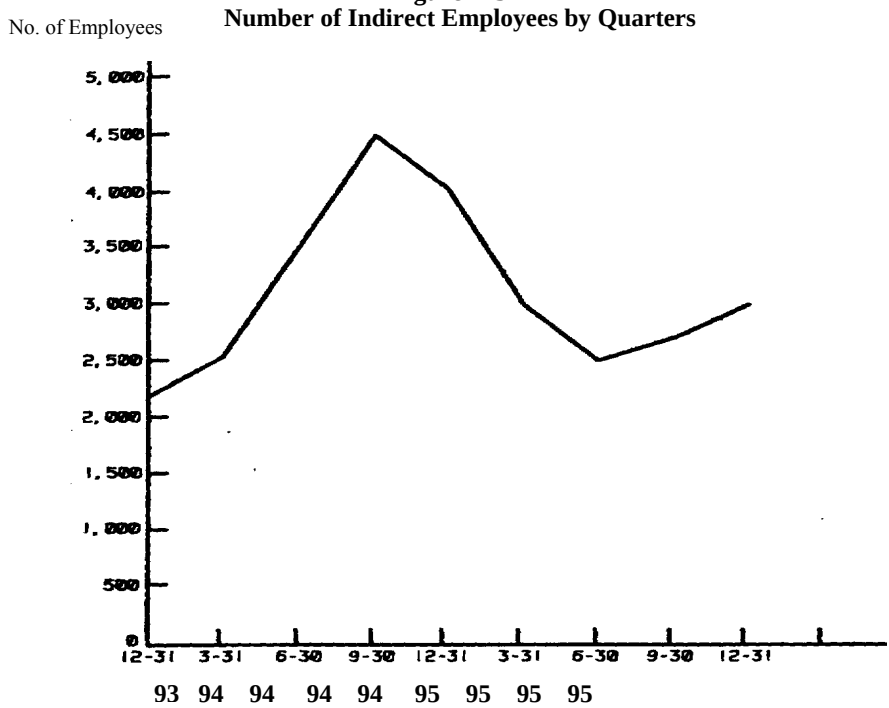
This section presents a discussion of the use of time series analysis to evaluate contract costs.

E-302 Definition and Concept

a. Data which are collected and classified by successive time intervals so as to reflect changes occurring with the passage of time form what is generally called a "time series." The graph of a time series is in effect a scatter diagram in which the independent variable has been replaced by a time element, such as years, months, or weeks. An illustration of indirect labor plotted as a time series is given in Figure E-3-1. As shown on the chart, successive plotted points are usually connected by a series of straight lines to show more clearly the variations in the data. This usually is not done in a scatter diagram because the plotted points are seldom in chronological order. A type of situation in which it would be informative to join the points is discussed in E-104.5.

b. The graphs of few time series approximate straight lines, a few others look like rough curves, but the majority gives the impression of a series of hills and valleys. For this reason, time series analysis is used to bring some order into the irregular pattern and seemingly erratic appearance of a time series. A variety of circumstances and conditions acting simultaneously and with their effects superimposed give the time series its irregular appearance. It is customary to divide the fluctuations of a time series into the following four general classifications: (1) chance and unusual variations, (2) secular or long-term industry and company trends, (3) seasonal variations, and (4) business or economic cycles.

Figure E-3-1
Number of Indirect Employees by Quarters



E-303 Chance and Unusual Variations

Chance and unusual variations of time series are those changes which, being caused by unusual events such as strikes, floods, and fires, do not follow the normal pattern of business operations. To the extent that variations are caused by chance, there is little that can be done to predict the behavior of the time series. However, changes in time series which are due to unusual events may be correlated with the events, simply by inspecting the data. The primary reason for the discovery and isolation of fluctuations due to rare events is that adjustment can subsequently be made for these variations in the investigation of secular trends and seasonal patterns.

E-304 Secular Trends

The term "secular trend" is used in this appendix to refer to any long term industry or company trend as distinguished from long term variations in the business data caused by the business or economic cycle.

a. Secular trends depict the gradual and consistent pattern of long term changes in the business operations. Whether a long term trend is secular or a phase of a longer term economic cycle is often difficult to determine; generally, however, it will represent a combination of both. It is for this reason that, as long as the data are comparable, it is advisable to have data covering as long a period of time as practicable.

b. In simple analyses and evaluations of time series, secular trends are usually thought of as straight lines fitted to the series indicating the gradual growth or decline of the variable. For example, if wage rates increase from year to year, we say there is an upward trend. Time series trends are usually shown as straight lines on regular or semi-logarithmic graph paper. A straight line on semi-logarithmic graph paper represents an exponential curve. This type of curve is applicable to a variable which tends to increase or decrease by a fixed percentage from one period to the next. Few time series are actually of such a simple nature that they closely follow a straight line when plotted on either regular or semi-logarithmic graph paper. For example, national indexes of wage and price levels generally have not closely followed either a simple linear or exponential curve. For reasons of simplicity, straight line trends are the most commonly used and this discussion, also for simplicity, will be limited to their use. Nevertheless, straight lines should be used only if they reasonably represent the trend for the desired period. When a straight line does not reasonably approximate the trend, the analysis requirements for cost auditing purposes can usually be met by drawing a moving average line as described in E-307.

c. As in the case of a scatter diagram, a convenient means for obtaining an objectively fitted trend line to time series data is by the method of least-squares described in E-202. An illustration of the computation of such a trend line using the quantitative methods software, EZ-Quant, is given in E-311. This approach is useful in studying the behavior of historical data when there is a reasonable degree of correlation between the passage of time and the growth of, or decline of, a variable. Such correlation is demonstrated when there are few changes in the general direction of the slope and if the approximate trend, when plotted on arithmetic paper, is a straight line. However, business data tend to result from the synthesis of many variable factors whose effects cannot be isolated or measured, resulting in a high degree of variability in the accumulated data. This variability may make the use of the mathematical precision of the least-squares line rather pointless for the analysis of time series; hence, when only a rough idea of the trend is needed, a freehand line based solely on judgment and sight is sufficiently accurate. Extreme caution should be exercised in using trend lines fitted to time series data as a forecasting technique. Use and limitations of such trend lines for forecasting purposes are discussed in E-310.

d. The precision of the semi-averages method lies between the method of least-squares and the freehand drawing of a line.

This method, which also may be applied to several partial averages, requires the following steps:

(1) divide the data into two or more parts; normally, two equal parts are used; but, if it would be more appropriate, especially if a very long time period is used, the data can be divided into three or more parts,

(2) calculate the individual average (mean) for each part of the data,

(3) plot these averages at the midpoints of each part of the data, and

(4) extend a smooth line through the points.

(a) To illustrate the method of semi-averages, consider the following data:

Year	Month	Monthly Value	6-Month's Total	Monthly Average
1995	Jan	3,800		
	Feb	3,900		
	Mar	4,400	< Plot Point 22,800	3,800
	Apr	3,600		
	May	3,100		
	Jun	4,000		
	Jul	5,500		
	Aug	4,400		
	Sep	3,900	< Plot Point 29,400	4,900
	Oct	4,300		
	Nov	6,600		
	Dec	4,700		

(b) In the above table the six-month's totals and the corresponding monthly averages are listed at the midpoint of each averaging period; that is between March and April and between September and October. The monthly values from this table are then plotted at the midpoint of each month; and the two averages (3,800 and 4,900) are plotted at the midpoints of each six-month period. The line drawn through these two points is the semi-average trend line.

E-305 Seasonal Variations

a. The type of variation in a time series easiest to understand is seasonal variation, which consists of regularly repeating patterns of moderate or short duration in the contractor's operation. Although the name of this type of variation implies a connection with the seasons of the year, it is used to indicate any kind of variation which is periodic in nature, provided it is also of short duration.

b. While as a rule, it is easy to determine the length or period of seasonal cycles, the description of their characteristics is more involved. There are a number of techniques for describing them, including some which involve rather tedious calculation. A relatively simple method of measuring seasonal patterns, which may be called the simple average method, consists of constructing a seasonal index. This is a descriptive measure that compares, by a series of ratios, the value of each month with the overall monthly average for the entire period.

SEASONAL INDEX OF DIRECT LABOR HOURS

Year	Month	Direct Labor Hours	Seasonal Index
1994	Jul	38,414	88
	Aug	38,610	89
	Sep	45,203	104
	Oct	36,734	84
	Nov	31,368	72
	Dec	41,415	95
1995	Jan	55,556	127
	Feb	44,158	101
	Mar	34,545	79
	Apr	41,214	95
	May	68,252	157
	Jun	47,460	109
Total		522,929	
Monthly Average (522,929 ÷ 12)		43,577	

The seasonal index is computed by dividing the labor hours for the month by the monthly average labor hours and multiplying by 100. For example, the seasonal index for the month of July, rounded to 88 in the table, is computed as follows:

$$\frac{38,414}{43,577} X 100 = 88.15$$

c. The principal use of the seasonal index in contract auditing is in the evaluation of the relative level of an estimate for an interim period. To illustrate this use, assume that the seasonal labor hour index for the month of May is 157, as shown in the above table, and that this value is considered typical. This would mean that based on the seasonal pattern the number of labor hours for May could be expected to be approximately 157% of the monthly average. If the total for the year were estimated at 600,000, the average monthly estimated hours would be 50,000 (600,000 ÷ 12) and the estimated hours for May would be 78,500 (50,000 x 157%).

E-306 Business Economic Cycle

a. Generally speaking, the economic cycle consists of those up and down movements of a time series which extend over long periods of time and originate from and are associated with general economic conditions such as prosperity, inflation, and depression. That is, the economic cycle for a particular business would be those variations in a time series for that business which would remain if the trend, the seasonal variation, and the chance or unusual fluctuations were eliminated.

b. The general business economic cycle may be analyzed into a number of different types of cycles associated with various theories which endeavor to explain their occurrence. However, since most cost evaluations for contract auditing purposes cover relatively short periods of time, it is not usually feasible to apply the refined techniques necessary to isolate the business economic cycle.

E-307 The Moving Average

E-307.1 Description

When the time series trend is not a straight line and when in evaluating costs the only interest is in the general movement of the series, it is customary to study the behavior of the series by means of a so-called "moving average." A moving average time series is an artificially constructed series composed of overlapping averages in which the value for each period is replaced by the average (mean) value for two or more adjacent periods. In computing the

second and subsequent averages, the earliest value of the preceding computation is dropped and the next new value in point of time is picked up to yield the new overlapping average. In constructing a graph of a moving average time series each average value is plotted at the midpoint of its respective averaging period. The moving average has the effect of smoothing the minor fluctuations in a time series. In fact, a perfectly regular, periodic pattern can be eliminated from the series by the use of the appropriate moving average. The moving average method, however, will yield a smooth trend only if the variations to be eliminated are essentially stable in both duration and amplitude and if the trend is basically linear. By averaging the effects of the seasonal and other short term variations, the resultant average line will primarily express the approximate general trend of the longer term variations.

E-307.2 Construction – Moving Average

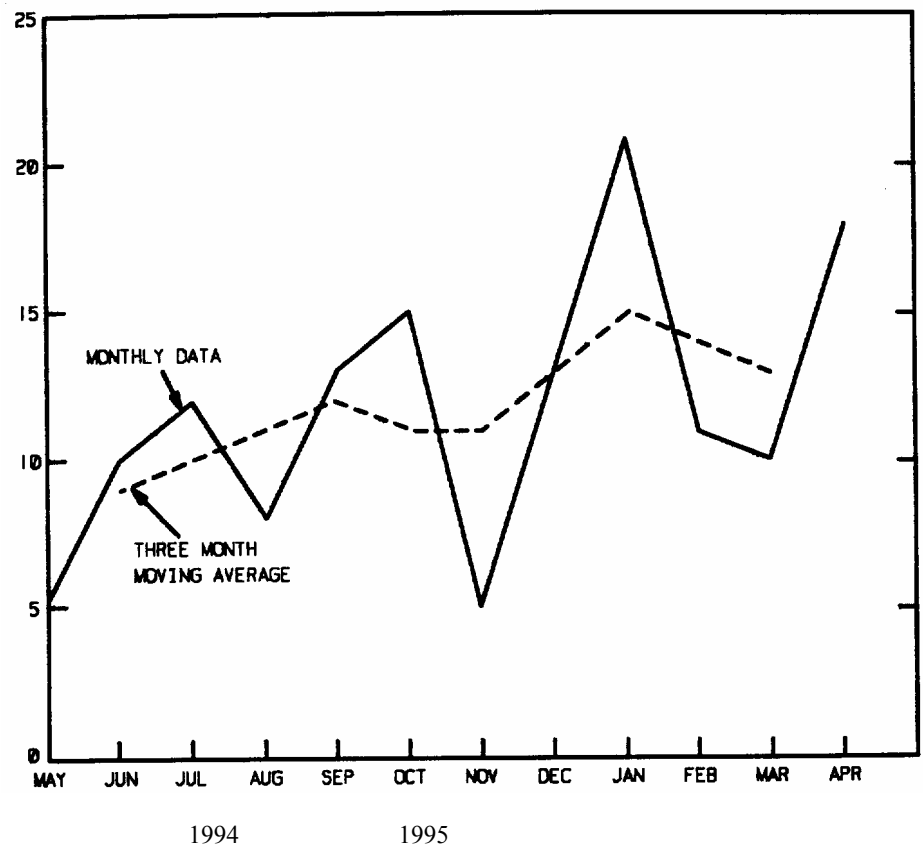
a. The first step in the construction of a moving average time series is the selection of the proper averaging period. As a general rule, it should correspond to the average length of the cycle of the variations to be eliminated. If a shorter period is selected, then influences pertinent to the average but occurring before or after the averaging period will be disregarded and, as a result, part of the cycle will remain in the moving average. On the other hand, if the period is too long, occurrences which in point of time are not pertinent to the average will be included in its computation.

b. The computation of the moving average is illustrated in the following table; a graph showing both the monthly and moving average data is given in Figure E-3-2. For ease and clarity of presentation, an averaging period of three months was selected and data for only one year has been shown. The moving total, column 4, was computed by first adding the values for the first three months and entering the amount opposite the midpoint of the period, the second month. Next a similar total is computed by adding the values for the second, third, and fourth months, and entering the total after the third month. At each subsequent shift of the moving total a new total is computed dropping the earliest of the months previously used and substituting the value for the next succeeding month. The final total (for February, March, and April 1995) is entered opposite March 1995. The moving average values, column 5, are computed by dividing the moving total by the number of terms in the average period; in this case, three.

COMPUTATION OF A THREE-MONTH MOVING AVERAGE FOR FIGURE E-3-2

Year	Month	Monthly Values	Three-Month Moving Total	Three-Month Moving Average
1994	May	5		
	Jun	10	27	9
	Jul	12	30	10
	Aug	8	33	11
	Sep	13	36	12
	Oct	15	33	11
	Nov	5	33	11
	Dec	13	39	13
1995	Jan	21	45	15
	Feb	11	42	14
	Mar	10	39	13
	Apr	18		

Figure E-3-2
Three-Month Moving Average



E-307.3 General

a. As shown above, both the moving total and moving average are entered at the midpoint of the averaging period. The reason for this is that an actual figure or a calculated figure, referring to a given period of time, refers equally to all portions of that period; and it is usually best to show the figure as applicable to the middle of the period. A yearly figure is referenced to the middle of the year, July first, a monthly figure to the fifteenth of the month, and a figure applying to an odd number of terms (weeks, months, or years) is referenced to the midpoint of the middle period. From this reasoning an average for an even number of terms would be located between the data for the two middle periods, and it may be so shown. However, it is usually desirable to show the average in its customary position, as applying to the middle rather than the end of a period; this requires one more computation to shift the average one half term. It may be accomplished by adding in turn each two successive moving averages and dividing by two. This latter average is then centered at the midpoint of the combined periods. This process, known as "centering the moving average," is applicable to any situation where the moving average period includes an even number of terms.

b. While the moving average will indicate the general direction and extent of a trend, it will closely portray only a linear trend. When the basic trend is curvilinear, the moving average line will follow a path which is inside the curve of data points. In effect, in a curvilinear situation the moving average values do not reflect the full swing of the trend variation, but tend to somewhat reduce it. Another limitation on the use of moving averages is that they cannot be computed for time periods nearer to the end of the series than approximately one-half of the averaging period.

E-308 Adjustment of Data

It is essential that the data be comparable item by item throughout the analysis period. Because this is not always true of the raw data, it is frequently necessary to adjust the figures before an analysis is undertaken. Two common types of adjustments are discussed below:

a. Adjusting for Changes in Classification Methods.

Since the methods of classifying and apportioning costs may be changed from time to time, the contents of the accounts to be analyzed may also change during the analysis period. Occasionally, these changes are not significant to the analysis and may be disregarded. This, however, is not usually true; and, to the extent that the changes are significant, the data accumulated before the change are not comparable to the data accumulated after the change. For evaluation of estimates, the contract auditor is interested in the analysis of only that data which were developed under current procedures. If a significant change in procedures is made, the auditor would have a choice of using a shorter analysis period, or adjusting the data to place it on a comparable basis, preferably adjusting the first part of the data. Adjustment of the data is the better method if the analysis period is relatively short. In this case, any further shortening of the period to avoid adjustment of the data could result in a period so short that the full pattern of the variations might not be reflected.

b. Adjusting for Periods of Various Length.

(1) Most of the data which the contract auditor will use are available in monthly, weekly, or daily forms; and the auditor must recognize that the work month, work week, and even the work day are variable units of time. The shortest month is about 10% less than the longest; and the variation may be even greater if holidays are taken into consideration. The work week may vary from one to seven days; and the work day may vary from a few hours to one containing considerable overtime or to one composed of three eight-hour shifts. The solution to the problem, usually, is to put the data on a "per standard working day" basis by dividing the total for each period by the approximate number of standard working days or hours. When there is considerable overtime, more than one shift, or an abnormally low work load, the reduced efficiency of these abnormal operations could also affect the comparability of the data, depending on the purpose of the analysis. For example, if the object of the analysis is to determine the normal costs for the operation of a production center at normal levels of production, reduction of the data to a standard working day basis without recognition of the extra costs due to the lowered efficiency of extended operations could distort the results of the analysis.

(2) The adjustment for the number of working days is necessary for any series, like that of monthly production and cost totals, where each individual working day means a more or less proportionate increase in the data values. There are, however, two kinds of data which do not require adjustment to a working day or similar basis: those involving values as of certain dates and those stated as averages. For example, figures showing the number of workers in departments are usually based on the number on the payroll on a certain day of the month; and inventory quantities and values are also stated as of a specific day. On the other hand, monthly average labor rates, and percentages, and index numbers are obviously values which do not depend in any systematic way on the length of the period to which they are applied.

E-309 Length of the Analysis Period

a. The length of the selected analysis period can greatly affect the outcome of a trend analysis. Although no simple rules can be given for determining the proper length, the period normally should be sufficiently long to reflect the full pattern of changes that may be pertinent to the purpose of the audit. When this is not possible and a trend analysis is prepared on the basis of a few observations, the auditor should bear in mind the resulting limitations.

b. In selecting the period for fitting a trend line to cyclical data, the series should normally start and end in about the same phase of the cycle. For example, if it is to end during a period of high costs, then it normally should also start during a period of high costs. If it starts and ends at different points in the cycle, then the slope of the trend line is apt to be steeper or more shallow than conditions warrant. On the other hand, changes that suddenly reverse a trend are not uncommon; and, if there is a distinct or abrupt change in the series, it is usually best to break the series into two parts and fit separate trend lines.

E-310 Interpretation and Use

The time series chart and trend lines have been used for a number of years in analyzing costs and in preparing cost estimates. While in most instances these devices and techniques have been properly and successfully used, a number of cases reflect a lack of understanding of the purpose, nature, and limitations of time series analysis in the evaluation and estimation of costs. Much of the misuse appears to stem from a confusion of the devices and methods of time series analysis with those of correlation analysis. Because of the similarity between these two techniques, the problems involved in the interpretation and use of a time series from the contract auditing standpoint can probably be best visualized in contrast with those of correlation analysis.

E-310.1 Comparison of Concepts

a. In the evaluation of contract costs and cost estimates, it is frequently important to know and to measure the interrelationships between the various costs and cost factors; to know whether, with what probability, and under what conditions a change in one cost parameter will be accompanied by a change in another cost parameter. These facts about the cost parameters can only be determined by studying their interrelationships independently of all other parameters, including that of time. The discovery and measurement of these interrelationships, without reference to the sequence of the events in which these relationships occur, is the objective of correlation analysis.

b. Although few occurrences in business are functions of time, the amount of cost that will be incurred in any operation is directly dependent on the nature and extent of the contractor's facilities and the skill and experience of its workers. These capabilities are dynamic; they are constantly changing and expanding with time as new facilities are acquired and experience is gained. Any evaluation of incurred or forecast costs must take into consideration this changing pattern of capability; therefore, any analysis of changes in costs or cost parameters must also take their sequential relationship into account. This is the function of time series analysis; to discover, analyze by type, and depict in sequence the changing values of a variable during a selected period of time.

c. Correlation analysis and time series analysis are supplemental, not alternative procedures. For example, the growth and decline of indirect expense may be related to the changes in the direct labor by correlation analysis, but their relationship can be fully understood only if the changes in these variables are also considered in relationship to time. A time series chart displaying the costs of both direct labor and indirect expense might be used for this latter purpose; and these in turn could be correlated to the changes in plant capacity, such as the use of extra shifts and extra production lines.

d. A time series is a succession of periodically measured values from a sequence of individual events. The time series line which joins the otherwise disconnected points serves only to make fluctuations in the measured values and their sequential

relationship more readily apparent. As a result, the segments of the line between the data points have no meaning in regard to intermediate values. Changes in the measured values during a period of time are associated, except in a few instances, with the changing events and not with the passage of time per se. Consequently, the trend line of a time series does not portray a mathematical relationship whereby the value of one variable may be computed from the known or assumed value of another variable. Instead, it indicates the general direction of the variable values during the total time period being analyzed.

E-310.2 Application

a. The fact that the trend line generally cannot be used as a regression line does not mean that it cannot be used in making or evaluating cost forecasts. The trend line and the moving average line may be projected into the immediate future to the extent that the future values will develop in a predictable manner based on the current operating position. Their usefulness, however, diminishes rapidly the farther they are extended; and, if this extension is carried beyond the point where influences may be evaluated, unknown factors intervene so that the connection from the current to the future operating status becomes tenuous. The extended lines then lose their value, and their use in making or evaluating an estimate is hazardous. The validity of using the extended trend line or moving average line in connection with cost forecasts is closely associated with the lead time of production and production planning, with industrial and economic conditions, and on the forecaster's or evaluator's knowledge and understanding of the cost parameters which will determine the trend of the costs from the present through the forecast period.

b. The time series readily reveals limits within the historical operations, such as the highest, lowest, and normal values. If similar factors can be expected to operate in the future these limits may be used as a guide to the audit evaluation of forecasted values. This use of time series must be based entirely upon judgment. The coefficients of correlation and determination cannot be statistically evaluated as described in E-205.2. The reason for this is that a time series has only one variable whose values are sequentially related and not two (or more) variables whose values tend to be related to one another, as in a correlation.

c. Arranging data in the form of a time series facilitates comparisons between data occurring at different points in time and thereby gives meaning to observations that otherwise would signify virtually nothing. For example, it is possible to ascertain whether the data reveals a trend in the movement of the values or whether their occurrences are erratic. If there is a trend, and it disappears and reoccurs over a period of time, it is usually indicative of a cyclic operation, and it would be desirable to determine its nature and causes. A study of the interrelationships implied by coincident troughs or peaks for several cost factors should lead to a better understanding of the effects of changes in production methods and policies. On the other hand, lead time is frequently indicated by changes in one cost parameter regularly preceding those of another parameter. This fact will often enable the auditor to anticipate changes before they appear as historical data.

E-311 Illustration of Computation of the Least-Squares Line for a Time Series

a. The values required for computation of the least-squares line for a time series study of material prices are given in the following table. In this example, a manufacturer of electronic components publishes quarterly price lists. A newly-developed component first appears in the list at $10 and nine quarters later has declined in price to $4.

Year	Quarter	Unit Price
1994	1st	$10.00
	2nd	9.00
	3rd	8.00
	4th	7.00
1995	1st	7.00
	2nd	5.50
	3rd	5.50
	4th	4.50
1996	1st	4.50
	2nd	4.00

b. Price levels often tend to follow an exponential equation more closely than a linear equation. In preparing input to the program the periods may simply be numbered in sequence beginning with one. Figure E-3-3 illustrates the output of EZ-Quant, which used the exponential curve type to obtain an estimate of the price of the component in the 2nd quarter of 1997. A $2.60 estimate was accomplished by typing "14" in response to the computer inquiry for "*x*." The number 14 was used because the price list for the 2nd quarter of 1997 will be the 14th in the series.

Figure E-3-3
Computation of the Least Squares Line for a Time Series

Regression Data

Unit Price ($)	Time (in Quarters)
10	1
9	2
8	3
7	4
7	5
5.5	6
5.5	7
4.5	8
4.5	9
4	10

Data Summary and Correlation Coefficients

Data Summary				
10	Number of Observations			
	Average Value	Minimum Value	Maximum Value	Standard Deviation
Unit Price ($)	6.5	4	10	2.041
Time (in Quarters)	5.5	1	10	3.028
Correlation Coefficients Between Variable Pairs				
	Dependent Variable (Y)	Independent Variable 1		
	Unit Price ($)	Time (in Quarters)		
Unit Price ($)	1			
Time (in Quarters)	-0.979837	1		

The Fitted Equation

$Y = a * b\text{^}(X_1)$

Coefficients estimated by Regression:				
	Variable	Coefficient Value	Computed T-Value	Inclusion Assurance (%)
a	Intercept	10.93722613	73.1707	99.9
b	X_1 is Time (in Quarters)	0.9024271	-19.4853	99.9
Summary Statistics:				
	R squared	Regression Std Error	Regression Std Error (log-based)	
	0.979364	0.278486	0.047858	
	Comparison Assurance (%)	F Statistic	Regression Degrees of Freedom	Residual Degrees of Freedom
	99.9	379.6781	1	8

Actual vs. Calculated Dependent Variable Values for
$Y = a * b\text{^}(X_1)$

Data Item	Actual Value	Calculated Value	Difference (Actual-Calc'd)	% Diff (Diff/Actual)
1	10	9.87	0.13	1.3
2	9	8.91	0.09	1
3	8	8.04	-0.04	-0.5
4	7	7.25	-0.25	-3.6
5	7	6.55	0.45	6.5
6	5.5	5.91	-0.41	-7.4
7	5.5	5.33	0.17	3.1
8	4.5	4.81	-0.31	-6.9
9	4.5	4.34	0.16	3.5
10	4	3.92	0.08	2.1

Projecting With the Regression Equation $Y = a * b\text{^}(X_1)$

	Independent Variable	Coefficient Value	Independent Variable Value	Contribution to Projection
a	Intercept	10.93723		10.94
b	X_1 is Time (in Quarters)	0.90243	14	0.24
	Projection:			
	Y is Unit Price ($)			2.6

E-400 Section 4 --- Audit Applications of Graphic and Computational Techniques

E-401 Introduction

This section presents illustrations of the application of graphic and computational analysis to specific problems of cost evaluation in the field of contract cost auditing.

E-402 Identification of Departures from Historical Cost Patterns

E-402.1 General

a. Both time series analysis and scatter diagrams are easily used and especially helpful devices in focusing attention on costs which deviate from historical patterns. Regression analysis can also be used for this purpose. Whether or not a specific historical trend may be applicable to a current situation can be determined by a direct comparison of past and current data. The auditor should be aware of the cost patterns which are normal to the contractor's operation, and should look for, recognize, and investigate significant departures from normal patterns. Special attention should be given to the impact of any significant changes in products or production methods.

b. The analyses can and should be subjective since the auditor's goal is to identify those accounts which have deviated significantly from experienced trends and thus require a more detailed analysis. If, for example, there are significant changes in overhead expenses or rates or in direct labor costs which cannot be explained by a corresponding change in the level of production, the auditor will want to know the reason. Such changes may signal lack of management control over cost levels or a change in accounting procedures. This does not mean that costs falling within the normal pattern are automatically acceptable and need not be examined. The auditor should analyze individual expenditures which are so large that they have a material impact on Government contract costs. In addition, samples of items in cost categories which account for large proportions of the costs being audited should be examined. Appendix B provides guidance on the determination of appropriate sample sizes for this purpose.

c. By pinpointing costs that need special examination, by establishing a pattern within which costs may be considered generally acceptable and subject to only a minimum examination, and by indicating the trend of these costs in relation to some base of recognized acceptability, graphic and computational analysis techniques provide the auditor with an additional and quickly usable tool for determining both the direction and extent of required audit effort.

E-402.2 Compilation of Data

The data to be collected can be updated on a monthly or quarterly basis from the contractor's accounts and operating statements. Contractors frequently prepare graphs and analyses to identify cost trends and departures from historical patterns for attention by management. The auditor should attempt to identify and gain access to these documents in order to avoid duplication of effort. The following is a listing of basic variables, accounts and groups of accounts that may be useful for analysis purposes. In addition to collecting and analyzing current data for such variables, valuable information can be gained from tracking budgeted or forecasted amounts.

a. Overhead expenses, by pool, in total and segregated by account or homogeneous groups of accounts, such as:
(1) Indirect labor
(2) Employee benefits and payroll taxes
(3) Repairs and maintenance
(4) Marketing
(5) Bid and proposal expense
(6) Allocations
(7) Other variable overhead expenses
(8) Fixed expenses (depreciation, rent, etc.)

b. Overhead base for each pool.

c. Direct costs and other variables related to volume, such as:
(1) Direct material dollars
(2) Direct labor dollars and/or hours
(3) Other direct charges (tooling, travel, etc.)
(4) Sales

(5) Cost of sales or cost input
(6) Employees hired
(7) Number of employees, direct and indirect
(8) Square feet of plant space

E-402.3 Analyzing the Data

Some examples of the types of graphs that the auditor may wish to consider in identifying departures from normal cost patterns follow. The more points of data included in an analysis, the more useful the analysis will be for identifying deviations from historical patterns.

a. Time-Series Analysis.

(1) Overhead rates by month. Both the actual monthly and year-to-date rates should be plotted on a time-series graph. The current estimating rate should also be plotted for comparison. These graphs will enable the auditor to become aware of within-the-year trends and evaluate whether the current estimating rate is acceptable for the remainder of the year. If rates from prior years are also plotted this will enable the auditor to identify months which deviate from experienced trends.

(2) Relationship of estimated amounts of key variables to actual amounts. The objective of these analyses is to determine if there is any pattern of under- or overestimation by the contractor. For example, does the contractor consistently underestimate sales, the number of employees, or the amount of the overhead base? Does the contractor overestimate certain expenses which are not particularly susceptible to audit? Over a period of time, such patterns, if they exist, will show up and the auditor can consider them in his audit of future estimating rates. The analyses can be accomplished by plotting the estimates on graphs used to track the actual data.

b. Scatter Diagrams.

(1) Relationship of overhead dollars to the base of allocation. The total overhead dollars are plotted, preferably monthly or quarterly, on a scatter diagram on which overhead dollars are scaled on the *y* (vertical) axis and the base is scaled on the *x* (horizontal) axis. This type of graph can also be used to compare individual accounts or groups of accounts with the base. Such graphs will assist the auditor in determining when overhead costs for a period are out of pattern. In addition, it will give the auditor some insight into whether the overhead costs are dependent on the base of allocation or whether they are controlled by other factors.

(2) Relationship of overhead expenses to other variables. For example, scatter diagrams relating total overhead expenses to sales, cost of sales, or direct labor will enable the auditor to evaluate whether overhead costs are increasing or decreasing consistently with changes in business volume. A scatter diagram relating indirect labor dollars, hours or employees to direct labor dollars, hours or employees will indicate whether indirect labor is increasing or decreasing disproportionately. A graph relating rent, depreciation and/or maintenance expense to floor space may be helpful in determining the reasonableness of these expenses. A graph relating the expenses of the payroll and/or personnel operations to the number of employees is helpful in evaluating the reasonableness of the expenses of these operations. Other potentially useful graphs are the relationship of allocated expenses to the appropriate allocation bases; the relationship of employee benefit costs to the total labor dollars; the relationship of recruitment and personnel costs to the number of employees hired; and the relationship of material handling costs to material costs.

c. Regression Analysis.

(1) Scatter diagrams are adequate to detect significant departures from historical relationships between two variables. Consequently, regression analysis should be used for this purpose in applications which are sufficiently computerized to avoid the expenditure of excessive auditor time. EZ-Quant can be used for this purpose. Where it is important to evaluate the reasonableness of a cost in terms of more than one independent variable, multiple regression will be required.

(2) If possible, obtain the contractor's historical data in an electronic file format, preferably as an ASCII or Excel file. EZ-Quant can import data files that were saved in a proper format using Excel. The Help facility in EZ-Quant contains guidance related to "saving an Excel file as a differ-

ent file type." Otherwise, the EZ-Quant Regression Analysis Wizard can be used to manually input data into a file and to complete the regression analysis steps.

E-403 Evaluation of Proposed Overhead Costs

a. One of the most important applications of regression analysis is analytical testing the reasonableness of estimated overhead costs. Overhead rates vary in response to many causes but because many overhead costs are fixed or semi-fixed, the level of operations is one of the most important factors. Overhead rates generally are lowest when a plant is operating at capacity and increase substantially when production levels are reduced. This relationship becomes increasingly significant in evaluating bid proposals for large contracts which will involve production over extended periods of time. In these cases it is necessary to predict overhead rates many years in advance on the basis of operating levels projected for these years.

b. The simplest way of using regression analysis to evaluate proposed overhead costs is to fit a least-squares line to the historical amounts of an overhead pool (the dependent variable) and the base of allocation or some other predictable measure of the level of operations (the independent variable). A computerized example is presented in E-202.2. This procedure assumes that the overhead pool consists of (1) fixed expenses which remain approximately the same at all operating levels, (2) variable expenses which tend to be directly proportionate to the base and (3) semi-fixed expenses which cannot be reduced below a certain level while the plant is in operation, and which increase above this level in proportion to the base. The value obtained for "*a*" will then represent the average value of the fixed expense plus the fixed portion of semi-fixed expenses during the period covered by the historical data. The value obtained for "*b*" will represent the average ratio of the variable expenses plus the variable portion of the semi-fixed expenses to the base. Estimates of the a and b parameters can also be obtained by judgmentally segregating fixed and variable expenses based on an evaluation of each account in the overhead pool. However, many overhead accounts consist of semi-fixed expenses and it is difficult to judgmentally determine the amount of fixed and variable expense in each such account. Consequently, regression analysis may be an easier means of estimating the total amount of fixed and variable expenses in an overhead pool. For reasons discussed in E-208g, caution should be exercised in the use of the *a* value obtained from a regression analysis as the total amount of fixed expense. However, in many contract audit applications, the use of simple linear regression as described in this paragraph has been found to provide sufficiently accurate and reliable predictions of overhead costs. The following paragraphs describe refinements to this basic procedure which may be needed to improve predictions of overhead. The extent to which these refinements are appropriate in any given audit situation depends on the extent to which departures from the assumptions cited earlier in this paragraph can be expected to affect Government contract prices.

c. If the variables included in the regression analysis have been significantly affected by changes in major economic factors (such as inflation or wage levels), adjustments of the type described in E-207 will be required. As discussed in E-208a. and b., adjustments may also be required to eliminate unusual or nonrecurring costs and to compensate for accounting system changes.

d. Some overhead expenses can often be predicted more accurately by methods other than regression analysis. For example, future depreciation expense may be obtained from depreciation schedules for existing facilities and planned acquisitions. Likewise, future rental expenses may be predicted based on existing leases. Acceptable research and development and bid and proposal expenses may be predictable based on advance agreements. Changes in union agreements, management policies and payroll tax laws can cause significant deviations from past patterns of employee benefit costs. Consequently, it may be necessary

to predict employee benefits based on a judgmental consideration of these factors in conjunction with anticipated employment levels.

e. In some situations management may be planning to institute changes which will result in greater operating efficiencies and lower costs. The auditor should adjust projections based on historical data accordingly.

f. Some overhead expenses may vary in response to a parameter other than the base of allocation. If the parameter is predictable, better forecasts of the related costs may be obtained by applying regression analysis to the historical amounts of these costs and the parameter. In some cases, more than one parameter can be identified as influencing certain expenses. For example, maintenance costs may vary in response to both floor space and direct labor. Multiple regression analysis should be considered as a means of forecasting such expenses.

g. The use of regression analysis assumes that overhead costs will be the same in the future as the past except for predictable changes in the independent variable(s). If a change in the basic pattern of an expense is anticipated, regression analysis applied to historical data will obviously not provide useful results and some other method of prediction should be employed. When significant changes in production capacity are planned, it generally will not be reasonable to assume that past patterns will continue. However, if no drastic changes in operating methods are planned, it may be possible to compensate for the changes by adjusting the value of "*a*" obtained from the application of regression analysis to historical data.

h. In some cases, it may not be practical to adjust prior data to compensate for the creation of new overhead pools and transfers of functions among pools. In such cases, the application of regression analysis to the combined expenses in two or more pools should be considered. Of course, when this is done, some means of breaking out predictions of overhead costs among pools must be devised.

E-404 Evaluation of Allocation Bases

a. While perfect correlation between any overhead pool and any available base of allocation is unlikely, some bases of allocation will provide a greater degree of correlation than others. The higher the correlation, the more likely it is that a base will provide an equitable distribution of overhead. One of the principal arguments that can be presented to support an audit recommendation for a change in a contractor's base of allocation is an analysis which shows that variations in the overhead pool have historically been more closely related to variations in the base recommended by the auditor than that being used by the contractor. An objective evaluation of the relative efficiency of alternative bases of allocation can be obtained by comparing the coefficients of determination obtained as a result of fitting a least-squares line to the pool and each base. A discussion of such comparisons is presented in E-205.4. In order to maintain comparability between the coefficients, data for the same periods must be included in each analysis. The relationship between each base and the overhead pool should be tested over a sufficient period of time to establish its historic stability. Both the conditions tending to produce stability and the causes of any material deviation should be evaluated.

b. As noted in E-403f., better predictions of some expenses in an overhead pool may be obtained by correlating them with some parameter other than the base of allocation. When this situation is encountered, consideration should be given to recommending that these expenses be included in a separate pool for allocation based on this parameter.

E-405 Evaluation of Proposed Labor Rates

E-405.1 General

a. In recent years labor rates have generally increased from one period to the next. However, it must be remembered that this pattern of increase applies only to individuals with the same qualifications doing the same job. In analyzing labor rates projected for a Government contract, the audi-

tor must be sure that the types of employees whose wages are being projected are the same as those to be assigned to the contract. It would not be proper, for example, to project engineering department labor costs based on average departmental rates, if a proposed contract will require a disproportionate amount of time by lower paid draftsmen. In addition to the category of work required, the auditor must consider changes in the projected level of production, e.g., if new workers are to be hired, they will probably start at wage rates lower than the current average.

b. In the usual application of regression analysis to the prediction of labor rates, average historical rates are first adjusted to eliminate the effect of changes in wage-related economic factors, as discussed in E-207b. The adjusted wage rates are then correlated with the plant population or the number of labor hours. The logic underlying this procedure is that as the number of employees increases, new employees will be hired at rates below the plant average, thereby reducing the average labor rate. Conversely, when the plant population is reduced, the employees with the least seniority and experience will be released first, thereby increasing the average labor rate. This procedure has produced excellent results in many contract audit applications. However, it does have limitations. It is most appropriate when the historical data included in the regression analysis is relatively recent, so that the auditor can be reasonably sure that manufacturing techniques and conditions have not changed substantially during the period of time covered by the data. It is also best when the projected plant population falls within, or close to, the range of population levels included in the data. A possible complication is that although a plant's population may vary at times, average seniority and hence labor rates may tend to rise during extended periods (say several years) of stable plant population. Under such circumstances, an evaluation based on estimates of the number of employees to be hired and separated and the average wage rates of the employees in each category may be required. In any projection of labor rates, the impact of expected changes in economic factors and employee mix must be considered along with the impact of expected changes in the number of employees.

c. Cautions set forth in E-310.1 in regard to limitations of the use of time series analysis are pertinent to the forecast of labor rates using multiple regression analysis. In other words, wherever time is not appropriate as an independent variable in a simple regression analysis, it is equally not appropriate when combined with another independent variable in a multiple regression analysis.

E-405.2 Illustration of the Use of Analysis in the Evaluation of Labor Rates

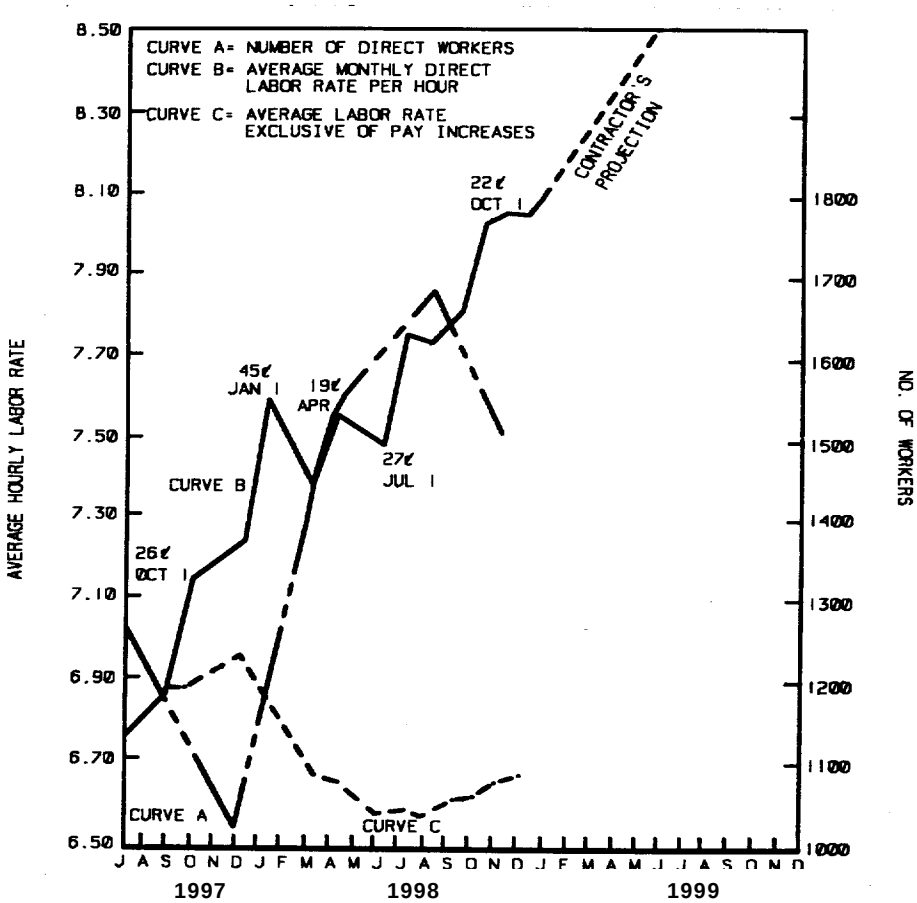
a. The time series chart in Figure E-4-1 graphically portrays hypothetical data on the number and average hourly rate of a contractor's direct manufacturing employees for an 18-month period starting July 1, 1997 and ending December 31, 1998. The data are shown in Table E-4-1.

Table E-4-1
Average Monthly Direct Labor Hourly Rates
and Number of Direct Workers
Curves A, B, and C
July 1997 to December 1998

Month		Average Hourly Labor Rate	Cumulative Rate Increases	Adjusted Hourly Labor Rate	Number of Employees
1997	Jul	$ 6.76	$ -	$ 6.76	1,289
	Aug	6.80	-	6.80	1,227
	Sep	6.88	-	6.88	1,143
	Oct	7.15	.26	6.89	1,128
	Nov	7.20	.26	6.94	1,053
	Dec	7.22	.26	6.96	1,022
1998	Jan	7.60	.71	6.89	1,117
	Feb	7.52	.71	6.81	1,244
	Mar	7.39	.71	6.68	1,419
	Apr	7.56	.90	6.66	1,532
	May	7.50	.90	6.60	1,564
	Jun	7.48	.90	6.58	1,617
	Jul	7.76	1.17	6.59	1,652
	Aug	7.74	1.17	6.57	1,682
	Sep	7.78	1.17	6.61	1,613
`	Oct	8.01	1.39	6.62	1,568
	Nov	8.05	1.39	6.66	1,513
	Dec	8.06	1.39	6.67	1,452

b. Three lines and two scales are shown on the chart (Figure E-4-1). The right-hand scale applies to Curve A, which indicates the number of workers employed each month, while the left-hand scale is used with Curve B, which shows the average direct labor rate per hour exclusive of premium pay. The dotted line, Curve C, represents a replotting of Curve B to eliminate the effect of general pay increases granted during the period. The example assumes that the union contract provides for quarterly cost-of-living adjustments to wage rates and that a general wage increase was negotiated as part of a 2-year contract effective in January 1998. These pay increases, the effective dates of which are indicated on the graph by arrows, total $1.39. The peak month of employment occurred in August 1998 when there were 1,682 workers. The cumulative pay increases up to that time totaled $1.17. Adding this increase to the July 1997 starting rate of $6.76 would make a prospective average rate of $7.93 ($6.76 + $1.17) as of August, 1998. The average rate excluding overtime at that time was however, only $7.74. The decline of 19 cents ($7.93 = $7.74) resulted from other causes, primarily the increase in the number of direct workers from 1,289 in July 1997 to 1,682 in August 1998. During the following four months when the number of workers declined to 1,452, the trend was reversed and the wage rate increased 32 cents, of which only 22 cents was accounted for by wage increases. A projection line has been constructed indicating an expected increase in the average hourly labor cost over the next year of 8 cents per month. This represents a projection made by the contractor and corresponds to the average experience for the last 18 months. Two defects in this projection are apparent: (1) the projected rate of increase reflects the general wage increase granted in January, 1998 although only cost-of-living increases are provided under the union contract during the period covered by the projection and (2) it gives no consideration to any anticipated changes in the number of direct employees.

Figure E-4-1
Average Monthly Direct Labor Rates and Number of Direct Workers



c. It is apparent from a comparison of curves A and C that changes in the number of employees have had an impact on hourly rates. However, the precise relationship between these two variables cannot be determined from the graph. Figure E-4-2 illustrates the use of the hyperbolic curve type, that was selected using the regression analysis option of EZ-Quant, to estimate the relationship between the number of employees and the average labor rate. Based on the analysis in Figure E-4-2 it is concluded that the adjusted labor rate tends to follow the following equation:

Rate = $5.95717 + (1033.5/Number of workers)

d. The following computation illustrates how the foregoing analysis might be used in the evaluation of proposed direct labor costs. It is assumed that the contract will be performed in the second, third and fourth quarters of 1999 with employment reaching a peak in the third quarter.

	Year 1999		
	2nd Qtr	3rd Qtr	4th Qtr
Estimated Number of Employees	1,500	1,650	1,450
Estimated average rates based on regression analysis in Figure E-4-2 (base Jul 1997)	$6.65	$6.58	$6.67
Estimated cumulative increases since Jul 1997 ($1.39 in 4th Qtr 1998, plus $.27, $.30 and $.31 for each respective Qtr thereafter)	$1.66	$1.96	$2.27
Estimated hourly labor rate	$8.31	$8.54	$8.94
Estimated direct labor hours for proposed contract	62,000	119,000	31,500
Estimated total direct labor cost for proposed contract	$515,220	$1,016,260	$281,610

E-405.3 Alternative Methods

a. In the preceding example, the adjustment of historical data for pay increases was simplified by the fact that the union contract provided for uniform increases for all workers. If the increases had not been uniform, a more detailed computation based on the amount of the raise granted to each class of worker and the number of workers in each class would have been necessary. When the number of workers in each class is not readily available, estimates may be based on samples of the workers.

b. At some contractor locations, union pay increases are granted to non-union workers. In such cases, the union agreement can be used to adjust the rates paid to non-union personnel. Even at some plants which are not unionized, management grants across-the-board increases to employees to compensate for changes in economic levels. Adjustment of historical data for such increases can be accomplished in the same manner as for union increases.

E-406 Other Applications

a. An application of graphic and computational analysis techniques to the evaluation of proposed material prices is illustrated in E-311.

b. Correlation techniques may also be used in evaluating standards, such as:

(1) material handling costs,
(2) scrap, rework, and spoilage,
(3) export packaging costs, and
(4) field service warranty or guarantee expense.

As indicated in Chapter 9, contractors frequently use loading factors based on such standards in developing cost estimates. Before accepting such a standard, the auditor should know:

- the relationship between the costs included in the standard and the costs to which the standard is applied and
- the causes of any material deviations from this relationship.

Figure E-4-2
Correlation of Hourly Labor Rates and Number of Employees

EZ-Quant Regression
Regression Data

No. of Employees	Ave. Hourly Labor Rate ($)
1,289	6.67
1,227	6.8
1,143	6.88
1,128	6.89
1,053	6.94
1,022	6.96
1,117	6.89
1,244	6.81
1,419	6.68
1,532	6.66
1,564	6.6
1,617	6.58
1,652	6.59
1,682	6.57
1,613	6.61
1,568	6.62
1,513	6.66
1,452	6.67

Data Summary and Correlation Coefficients

Data Summary				
18	Number of Observations			
	Average Value	Minimum Value	Maximum Value	Standard Deviation
Average Hourly Labor Rate ($)	6.727	6.57	6.96	0.136
No. of Employees	1,379.72	1,022.00	1,682.00	225.762
Correlation Coefficients Between Variable Pairs				
	Dependent Variable (Y)	Independent Variable 1		
	Average Hourly Labor Rate ($)	No. of Employees		
Average Hourly Labor Rate ($)	1			
No. of Employees	-0.972254	1		

The Fitted Equation

$Y = a + b / (X_1)$

Coefficients Estimated by Regression:				
	Variable	Coefficient Value	Computed T-Value	Inclusion Assurance (%)
a	Intercept	5.95717063	158.5538	99.9
b	X_1 is No. of Employees	1,033.50	20.768	99.9
Summary Statistics:				
	R squared	Regression Std Error		
	0.964231	0.026427		
	Comparison Assurance (%)	F Statistic	Regression Degrees of Freedom	Residual Degrees of Freedom
	99.9	431.3104	1	16

Actual vs. Calculated Dependent Variable Values for $Y = a + b / (X_1)$

Data Item	Actual Value	Calculated Value	Difference (Actual-Calc'd)	% Difference (Diff/Actual)
1	6.67	6.76	-0.09	-1.3
2	6.8	6.8	0	0
3	6.88	6.86	0.02	0.3
4	6.89	6.87	0.02	0.2
5	6.94	6.94	0	0
6	6.96	6.97	-0.01	-0.1
7	6.89	6.88	0.01	0.1
8	6.81	6.79	0.02	0.3
9	6.68	6.69	-0.01	-0.1
10	6.66	6.63	0.03	0.4
11	6.6	6.62	-0.02	-0.3
12	6.58	6.6	-0.02	-0.2
13	6.59	6.58	0.01	0.1
14	6.57	6.57	0	0
15	6.61	6.6	0.01	0.2
16	6.62	6.62	0	0.1
17	6.66	6.64	0.02	0.3
18	6.67	6.67	0	0

Figure E-4-2

Projecting With the Regression Equation Y = *a* + *b* / (X_1)

	Independent Variable	Coefficient Value	Independent Variable Value	Contribution to Projection
a	Intercept	5.95717		5.96
b	X_1 is No. of Employees	1,033.50	1,500.00	0.69
	Projection:			
	Y is Ave. Hourly Labor Rate ($)			6.65

Projecting With the Regression Equation Y = *a* + *b* / (X_1)

	Independent Variable	Coefficient Value	Independent Variable Value	Contribution to Projection
a	Intercept	5.95717		5.96
b	X_1 is No. of Employees	1,033.50	1,650.00	0.63
	Projection:			
	Y is Ave. Hourly Labor Rate ($)			6.58

Projecting With the Regression Equation Y = *a* + *b* / (X_1)

	Independent Variable	Coefficient Value	Independent Variable Value	Contribution to Projection
a	Intercept	5.95717		5.96
b	X_1 is No. of Employees	1,033.50	1,450.00	0.71
	Projection:			
	Y is Ave. Hourly Labor Rate ($)			6.67