CHAPTER 19

DATA COLLECTION AND NORMALIZATION FOR THE DEVELOPMENT OF COST ESTIMATING RELATIONSHIPS

1. INTRODUCTION

Cost estimating relationships or parametric equations are mathematical statements that indicate that the cost is proportional to a physical commodity. Parametric estimating requires that statistical analysis be performed on data points to correlate the cost drivers and other system parameters. The basis of the data points is collected from databases or is developed by building a model of a project or scenario. The data collection effort is the first step in the development of CERs. Once the data is collected, it must be adjusted so that comparable relationships can be developed. This chapter discusses considerations for data collection and normalizations.

2. DATA COLLECTION

A minimum data requirement exists for any given job, but before data collection begins, the analyst must consider the scope of the problem, define in general terms what is to be accomplished, and decide how to approach the problem. In both construction and remediation projects, many different technologies and methods exist. To obtain data necessary to develop CERs, it is important to identify common or similar procedures among projects. It is also important to remember that both the cost and duration for each project are affected by site specific conditions.

A. Examining the Historical Data for Selection

The data required to estimate long-range maintenance costs at a DOE facility can be substantially different from the data required to develop short-term cleanup costs at the same facility. In the former, equipment upgrade and replacement costs must be considered, but in the latter, these items may not be a big factor. For major items, this means that a functional breakout (e.g., direct labor, materials, engineering, and installation) must be done. One can postulate problems requiring even a greater amount of detail. Suppose, for example, that two similar cleanup projects that are being evaluated have substantially different costs. Only by examining the cost detail can this difference be explained.

In performing this initial appraisal of the job, the analyst will be aided by a thorough knowledge of the kind of project being evaluated, its characteristics, the state of its technology, and the available information. With this knowledge the analyst can determine the kinds of data that are required compared to what are available, where the data are located, and the kinds of adjustments that are required to make the collected data base consistent and comparable.

Only after the problem has been given this general consideration should the task of data collection begin. All too often large amounts of data are collected with little thought about use. The result is that some portion may be unnecessary, unusable, or not completely understood, and other data that was necessary was not collected. Data collection is generally the most troublesome and time-consuming part of developing a CER. Consequently, careful planning in this phase of the overall effort is well worthwhile.

B. Sources for Historical Data

When developing a CER from historical data, it is important to consider the different sources available. Some examples of sources include published reports; adopted and draft regulations; local, commercial, and DOE databases; past and current estimates; and bid documents. The information extracted from these sources will provide the estimator an understanding of the steps that are necessary to perform the work for a project so the cost drivers can be identified.

For example, when considering the Uranium Mill Tailings Remedial Action (UMTRA) Project, data collection might include the following:

- searching a DOE energy database;
- reviewing the DOE Uranium Mill Tailing Remedial Action Project (UMTRAP) Status Report, Project Plan, Project Management Plan, Project Schedule and Cost Estimate Report;
- reviewing the remedial action plans (RAPs), site conceptual design and information for bidders for various UMTRAP sites;
- reviewing the Final Rule for Radon 222 Emissions from Licensed Uranium Mill Tailings, U. S. Environmental Protection Agency (EPA) document 52011-86-009, August 1986; and
- reviewing the report titled *Cost Components of Low Level Waste Remedial Action*, PEI Associates, Inc. Contract No. PE-AC01-85MA00205, PN 3685-9.

The information obtained in this historical data search identified the following information.

PURPOSE: Title I of the Uranium Mill Tailings Radiation Control Act (UMTRCA) of 1978, Public Law (PL) 95-604, authorized the DOE, in cooperation with affected State government and Indian tribes, to develop and provide a program to stabilize and control the tailings and other residual radioactive materials located at inactive uranium processing sites.

DATA COLLECTED: A map of the 24 designated processing sites (22 locations) was developed.

- The stabilization method for each location was defined. The processing sites, their priority, and the estimated amount of materials to be handled were identical.
- The current status and estimated cost and completion dates for the work elements at each site were found.

RESULTS: From this data CERs were developed, and an estimate could be developed for UMTRA with the cubic yards or acres of tailings being the only information available. The estimates could be a total cost or an individual component, such as Planning and Design Development.

C. Developing Data from Model Estimates

Sometimes an analyst will be required to develop estimates and CERs pertaining to projects for which there is no historical data. In this situation, the analyst can develop a conceptual design of the project (a model) and can estimate the cost of the model. This estimate is a more comprehensive effort than ordinary estimates. The project must be designed and then estimated for three to five cases or sizes differing from one another with respect to be parameter expected to drive the CER.

Costs for many of the required activities can be obtained from standard cost references and published reports. For example, remediation scenario costs would use references such as—

- R.S. Means Company Building Construction Cost Data Manuals;
- “Cost of Remedial Action,” Version 3.0, a computer cost program prepared by CH_2M Hill under Contract No. 68-01-7090, for the United States Environmental Protection Agency, Office of Solid Waste and Emergency Response;

- *Cost Components of Low Level Waste Remedial Action*, prepared by PEI Associates, Inc., under Contract No. DE-AC01-85MA00205 for the United States Department of Energy;

- *Guide for Decontaminating Buildings, Structures, and Equipment at Superfund Sites*, prepared by PEI Associates, Inc., and Battelle Columbus Laboratories under Contract No. 68-03-3190, for the United States Environmental Protection Agency, Office of Research and Development, Hazardous Waste Engineering Research Laboratory.

D. Historical Data Versus Model-Developed Cost Estimating Relationships

The advantage of a model-developed CER is that the user knows exactly what went into the CER. The assumptions and design that the estimator used to develop the CER are available to the user. The elements of a CER based on historical data are usually less well-defined. The advantage of the CER based on historical data is that the costs were produced by actual projects. Factors the estimator did not think to include in the model CER would be included in the CER based on historical data.

3. DATA NORMALIZATION

The historical data collection and a thorough understanding of the elements of a project are both important in developing CERs. Knowing the different elements that go into building the total project helps the estimator to normalize the data. Two projects may look similar on the surface, but if they are analyzed in more detail, it frequently becomes apparent that unique problems were encountered in each of the projects.

To be useful to the cost analyst, data must be consistent and comparable, and in most cases the data as collected are neither. Hence, before estimating procedures can be started, adjustments must be made for definitional differences, scope differences, etc. The more common adjustments are examined in this section. It is by no means an exhaustive treatment of the subject. The list of possible adjustments is long and frequently they are project-specific. Also, evidence on certain types of adjustments (for contractor efficiency, for contract type, for program stretch-out) can consist largely of opinion rather than hard data. While the cost analyst may allude to such adjustments, the research necessary to treat them in some definitive way has not yet been done.

A. Accounting Differences

Different contractor accounting practices require adjustment of the basic cost. Companies record their costs in different ways. Often they are required to report costs to the Government by categories that differ from those used internally. Also, Government categories change periodically. Because of these definitional differences, one of the first steps in cost analysis is to state the definitions that are being used and to adjust all data to these definitions.

B. Physical and Performance Considerations

A problem that resembles the one discussed above is the need for consistency in definitions of physical and performance characteristics. For example, remediation requirements may be referenced in many ways: remediation required by the regulations, remediation required by a specific contract, or remediation necessary for facility operations. All of these defining terms differ in exact meaning and value. The remediation required by a specific contract may be more than the remediation required by the regulations. The remediation required to place a facility into operation may not be exactly the same as the remediation required in the regulations. Differences such as these can lead an analyst unfamiliar with remediation to use inconsistent or varying values inadvertently. When data are being collected from a variety of sources, an understanding of the terms used to describe physical and performance characteristics is necessary to understand the content of the various cost elements.

C. Nonrecurring and Recurring Costs

Another problem that involves questions of definition concerns nonrecurring and recurring costs. Recurring costs are a function of the number of items produced; nonrecurring costs are not. Thus, for estimating purposes it is useful to distinguish between the two. Unfortunately, historical cost data frequently show such cost elements as nonrecurring and recurring engineering hours as an accumulated item in the initial contract. Various analytical techniques have been developed for dividing the total into its two components synthetically, but it is not yet known whether the nonrecurring costs that are obtained by these methods will be accurate.

A more subtle problem arises when nonrecurring costs on one product are combined with recurring costs on another (i.e., when the contract is allowed to fund development work on new products by charging it off as an operating expense against current production). Separation of the nonrecurring and recurring costs means an adjustment of the production costs shown in contract or audit documents to exclude any amortization of development. The nonrecurring expense that has been amortized can then be attributed to the item for which it was incurred. Such an adjustment can only be accomplished in cooperation with the accounting department of the companies that are involved.

D. Price-Level Changes

Changes in the average hourly earnings of workers must be considered. Wage rates fluctuate from year to year. Also, the location of the workers must be considered. Wage rates differ in different areas of the country.

E. Cost-Quantity Adjustments

The cost-quantity relationships must be considered. Costs are usually a function of quantity. Typically, as the total quantity of items produced increases, the cost per item decreases. If this principle is applied to remediation projects, it becomes apparent that as the amount of replacement material increases, the cost of replacement per unit decreases. Thus, in speaking of cost, it is essential that a given quantity be associated with that cost. A replacement cost might be $3.00 per square foot or $3.50 per square foot for the same material, depending on the total number of square feet replaced.

F. Escalation

Data will be collected from several projects. Typically, they do not all occur at the same time. Thus, the cost data must be normalized to the same base year prior to developing the CER. The data should be adjusted by using the escalation indices guidelines produced by the Office of Infrastructure Acquisition (FM-50).

G. Regional Differences

The same type of project may have been built all over the United States. The cost data may be for the same activities, but it is from several different regions. There are regional cost differences, and they must be considered when using the data.

H. Other Possible Cost Normalizations

The lack of a way to adjust cost data for productivity changes over time is illustrative of the current situation in which more kinds of cost adjustments have been theorized than have been quantified. For example, it has been suggested that adjustment may be required because of differences in contract type (fixed-price, fixed-price-incentive, cost-plus-fixed-fee contracts) or differences in the type of procurement (competitive bidding or sole source). The hypothesis is that the type of contract or procurement procedure will bias costs up or down, but this hypothesis is difficult to substantiate.

Another question concerns changes in techniques and available equipment. A related question concerns the efficiency of the contractor. It may be surmised that Contractor A has been a lower cost producer than Contractor B on similar items, but this is extremely difficult to prove. A low-cost producer may be one who, because of geographical location, pays lower labor rates.

The cost of delays can also skew the data. For example, when comparing two similar projects, the estimator may learn that one project was delayed for several months because of regulatory problems while the other project proceeded smoothly.

In order to normalize these two projects so they can be compared, the cost of the delay should be deleted from the project that experienced these problems.

Prior to using historical data in a CER, it should be checked to ensure it will not be used out of context. This is particularly important when the data come from a project with special considerations, such as a discount that will not apply to a project being estimated.

4. DEVELOPING COST ESTIMATING RELATIONSHIPS

Once the data have been collected and normalized, a set of data points is developed. These data points are used to build the CER.

A. Simple Averages

Many estimating relationships are simple statements that indicate that the cost of a commodity is directly proportional to the weight, area, volume, or other physical characteristics of that commodity. These estimating relationships are simple averages. They are useful in a variety of situations and, because of their simplicity, they require little explanation.

B. Detail of Cost Estimating Relationships

The estimator will sometimes want to build CERs for each step of a project. These CERs can then be summed to produce a CER for the total project if the steps are independent. This additional detail makes it easier to apply the CER to a new project. If the CER predicts a total cost that is significantly different from the existing estimated cost, the ability to use CERs to estimate the cost of different parts of the project allows the estimator to analyze those parts and identify where in the project the cost variances occur.

C. Enhanced Cost Estimating Relationship Program

The derivation of more complex relationships (i.e, equations that are able to reflect the influence of more than one cost variable) must be developed by using statistical analysis. A computerized software package, called the Enhanced Cost Estimating Relationship (ECER) Program, was designed for DOE for the development of estimating relationships and is available from FM-50.