Virginia Commonwealth University

**Introduction**

Considerable research has been conducted over the past 30 years on both descriptive and prescriptive models for competitive bidding. The majority of that work, a bibliography of which was given in Stark and Rothkopf [17], is concerned with the development of optimal bidding strategy models.

Dixie [6] and Fuerst [10] point out that an optimization model for bidding has three basic inputs: A cost estimate, information on competitors’ bids, and an actual cost. The models that have been developed tend to use game theory or decision theory concepts, developing a probability distribution for a successful (i.e., “won”) bid based on likely competitor actions.

However, as is pointed out by Grinyer and Whittaker [11], and in a somewhat different context by Brown [2], profits on fixed-price contracts tend to be lower than expected since the bids tend to be awarded to those who underestimate the costs the most.

Regardless of what specific bid development approach is being used, in order to bid intelligently, it is necessary to develop a picture of the possible and likely financial consequences of obtaining and carrying out the contract. It has been proposed by a number of authors (Bjornsson [1], Finch and Postula [7], Finch, Postula and Perry [8], Finch and Young [9], Kraemer [14], Spooner [16]) that risk analysis is an appropriate model for developing such a picture.

This article describes the development of a risk analysis model for estimating the probability distribution of project cost for a manufacturer of sophisticated electro-mechanical control systems, using statistical analysis to develop the necessary model structures and parameters. The use of the model for estimating the project cost probability distribution will be illustrated for an example project.

**Project Cost Structure**

The accounting system used by this company divides the costs charged to a particular project into two categories: Category I) Direct charges are identifiable as being associated with a particular project—standard labor, project labor, material, installation, engineering salaries, and miscellaneous; Category II) Allocated costs are charged to the project as the result of applying various formulas to the direct charges—standard overhead, other manufacturing costs, engineering overhead, and general sales and administrative expenses. These individual cost areas will be referred to as components in the remainder of this article.

To develop a total project cost estimate, estimates are made by appropriate company personnel of the costs to be incurred for each of the direct components over the project’s anticipated life. These direct cost estimates are then used as the bases for applying the formulas to estimate the indirect or allocated costs.

As the project runs its course, costs are periodically charged against it to these ten components, either on the basis of directly attributable use of resources (Category I) or by applying the relevant formulas to the direct Category I charges (Category II). 1

When the project is complete one can compare the original cost estimates of these ten components with their final charged amounts. If the total of the final charges differs from the sum of the original estimates, the profit on this project will be different from that originally expected (although contribution to profit and overhead may be the same as originally forecast).

The probability distributions of the final charged costs, both for individual components and for total cost for the project, and the resulting profit implications are what was referred to earlier as developing a financial picture of the project as a necessary input for intelligent bid determination.

**Risk Analysis Concept**

The general concept behind risk analysis is the combination of probability distributions for components of a system or project to develop an overall probability distribution for some characteristic of the whole system or project. One approach to this is analytic, combining the component probability distributions in a theoretical fashion. This is the approach taken by Spooner [16J. A second approach, which might be called semi-analytic, is more numeric, combining the component distributions by calculating probabilities for combinations of representative values. This is the approach followed by Finch and his colleagues [7] [8] [9]. A third approach, using simulated or Monte Carlo sampling, was articulated originally by Hertz [13]. It has been applied in a variety of areas, including the construction industry, as discussed by Bjornsson [1], Connor [4], Deshmukh [5], Hamlin and Jones [12], and Kraemer [14], and is the approach discussed in this article. This approach uses random numbers to sample values from the relevant system or project component probability distributions. These values are combined to give a single sample observation of the project characteristic of interest, in this case cost. The process is repeated many times and the cost values generated are aggregated to develop a probability distribution of project cost.

The simplest way of doing this is to assume that all relevant project cost components are independent of one another and then to estimate, either on the basis of historical records or management judgment (or both), probability distributions for those costs. In this case that requires estimating independent probability distributions for the six direct cost components. Since the mean and variance of a sum of independent random variables are, respectively, the sums of the means and variances of the individual variables, the mean and variance of total direct project cost can be derived directly.

A more sophisticated and, where applicable, a more useful approach is to develop a model that, in addition to using probability distributions for the direct cost components, specifies those probability distributions on the basis of significant relationships between project characteristics and the cost components and among the cost components themselves. By taking advantage of such dependencies, where they exist, a more accurate probability distribution of total project cost can be estimated.

The statistical analysis discussed in the next section forms the basis for developing such a dependent risk analysis model. It is based upon 33 company projects completed during a three and one-half year period.

Given the wide variation in the size of these projects, ranging from under $100,000 to several million dollars, it is more appropriate to focus on the ratio of the final charged cost to the original cost estimate than on the actual deviation of the final cost from the original estimate. In addition, since both the original estimates and final charged costs in Category II (allocated) are derived from those in Category I (direct) by the application of formulas, the analysis is restricted to the six direct cost components, although the Category II charging must be incorporated into the total cost risk analysis model.

**Statistical Analysis**

**Procedures Used**

There are many statistical procedures available for uncovering the significant inter-component and characteristic-component relationships that would be useful for developing the risk analysis model. The techniques used in this study are all standard ones covered in most first-level statistics courses in business and engineering programs.2 They follow the suggestions of Connor [4] and Uriegas-Torres [18] for the statistical exploration of a company’s historical data to uncover useful relationships.

Using the ratios of the final costs to the original estimates on these 33 projects as the variables to be predicted or explained, two types of analyses were conducted:

1. Attempts were made to relate the final/original cost ratios for the six Category I cost components to the following project characteristics:

a. Project type – These were four Types, designated for the reasons of confidentiality as I, II, III, and IV;

b. Original total project cost estimate;

c. Project length in months; and

d. Original estimate of cost for that component.

2. Attempts were made to relate the final/original ratio of one cost component to that of another, considering all 15 pairs of the six components in Category I.

The tests performed included: 1) analysis of variance to test if there were differences in the ratios by project type; 2) simple and multiple linear regression analysis to test for the significance of the other project characteristics listed above (project cost estimate, project length, and component cost estimate); and 3) correlation analysis to test for the strength of the relationships among the ratios for the different cost components. Summaries of the results of those tests are shown in Tables 1 and 2. Their significance is discussed next.

**Results**

The **Standard Labor** ratio is relatively unpredictable, being correlated only with the Installation ratio (Table 2).

The **Project Labor** ratio is highly predictable from the project length, type, and total cost estimate (Table 1). This last relationship is very reasonable as both are functions of the amount of custom designing in the project.

The **Material** ratio is highly predictable as a function of the project length, type, and cost estimate (Table 1). It is not particularly correlated with any other cost component ratio other than Miscellaneous (Table 2). The negative sign in this relationship is explainable in terms of the company’s accounting practice. As changes are made during the project’s execution, parts that were originally assumed in Material (made) may be switched to Miscellaneous (bought), and vice versa.

**Table 1** Summary of Results of Analysis of Variance & Regression Analysis Tests (Table entry is a level at which test is significant.)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Direct Cost Element** | | | | | | |
| **Explanatory Variable(s)** | **Standard Labor** | **Project Labor** | **Material** | **Installation** | **Engineering** | **Miscellaneous** |
| Project Cost Estimate | - | .05 | .20 | - | .10 | - |
| Project Length | - | .01 | .10 | - | .02 | .20 |
| Project Type | - | .10 | .05 | - | - | - |
| Element Cost Estimate | - | - | - | - | - | - |
| Project Length & Type | - | .10 | .05 | - | .025 | - |

**Table 2**  Correlation Coefficients Between Final/Original Cost Ratios for Direct Cost Elements

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Project Labor** | **Material** | **Installation** | **Engineering** | **Miscellaneous** |
| Standard Labor | .016 | .064 | .723 | -.117 | -.194 |
| Project Labor |  | .166 | .185 | .565(.01) | .146 |
| Material |  |  | .270 | .102 | -.259(.20) |
| Installation |  |  |  | .603(.15) | -.067 |
| Engineering |  |  |  |  | -.239(.20) |

The number in ( ) indicates the level of statistical significance.

Differences in significance level are due to varying sample sizes, as not all projects have all cost elements.

The **Installation** ratio is not especially predictable other than the previously noted correlation with the Standard Labor ratio and a less significant correlation with the Engineering ratio (Table 2).

The **Engineering** ratio is reasonably predictable as a function of project length and type, and, to a lesser extent, project cost (Table 1). Its correlations are those previously mentioned with the Project Labor and Installation ratios and a considerably less significant and somewhat inexplicable negative correlation with the Miscellaneous ratio (Table 2).

The **Miscellaneous** ratio is not especially predictable, either in terms of project characteristics, although there is some relationship with project length, or the other cost component ratios, other than the weak negative correlations previously mentioned with the Material and Engineering ratios (Table 2).

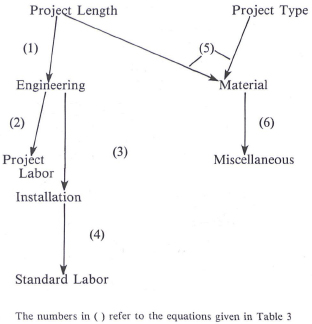
Considering all of these relationships, it appears that there is some predictability of the various component final/original cost ratios, both in terms of the project characteristics, especially project length total cost estimate, and through the relationships between the component ratios. This suggests that it is possible to develop a model exploiting these relationships to predict cost performance on a project. Given that the total project cost estimate is a less significant predictor than the project length and that total project cost and length are themselves correlated, total project cost will be dropped from further consideration as a predictive variable.

**Risk Analysis Model**

As discussed earlier, the development of a dependent risk analysis model requires two things. First is the determination of useful relationships, specifying which project characteristics and which component pairings are to be used in developing a total project cost predictability distribution through the risk analysis procedure. Second is the specification of the forms and parameters of the equations to be used for estimating the final cost component ratios. The statistical analysis just discussed provides both of these.

Using the relationships discussed in the previous section and summarized in Tables 1 and 2, the tree diagram in Figure 1 was developed to depict the relationships selected as most appropriate for relating the project characteristics to the component cost ratios for Engineering and Material and then using these ratios, through chains of relationships, to predict the remaining component ratios — Project Labor, Installation, Standard Labor, and Miscellaneous — through the regression equations estimated from the data base and summarized in Table 3 or as modified by management judgment.

**Figure 1** Tree Diagram Showing Relationships for Estimation of Final/Original Cost Ratios

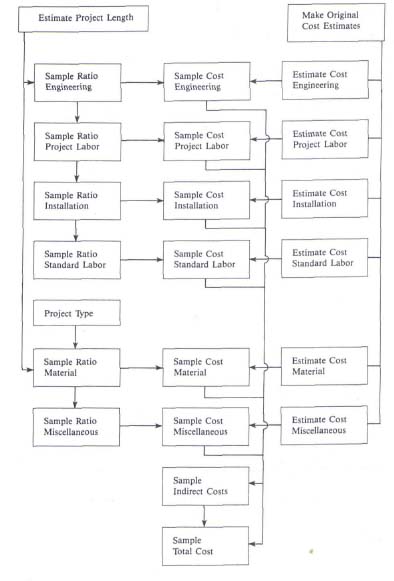


**Table 3**  Equations for Estimating Cost Ratio Relationships Diagrammed in Figure 1

|  |  |
| --- | --- |
| (1) Engineering Ratio | = .793 + .0078 (Project Length) |
| (2) Project Labor Ratio | = -.251 + 1.452 (Engineering |
|  | ration |
| (3) Installation Ratio | = .247 + .576 (Engineering Ratio) |
| (4) Standard Labor Ratio | = .891 + .278 (Installation Ratio) |
| (5) Material Ratio | = 1.112 + .0015 (Project Length)     -.034 if Type I Project     -.209 if Type II Project     -.166 if Type III Project     -.174 if Type IV Project |
| (6) Miscellaneous Ratio | = 2.448 - 1.210 (Material Ratio) |

To use the Risk Analysis Model, estimates of the six direct cost categories would be made in the usual manner. These six estimates, along with the project type, are the constant basic inputs to the repeated computer simulations required for estimating the Total Project Cost probability distribution. By either sampling from a managerially estimated project length distribution or using the single best estimate for length, a single complete sampling of all direct and indirect costs can be performed. Using the relationships summarized in Figure 1 and Table 3, ratios are developed for each of the direct costs. These ratios are applied to the original direct cost category estimates to get sample final direct costs which are then used, via the predetermined company indirect cost formulas, to generate sample indirect costs. In this way a complete set of sample project costs is created. This process can be repeated as many times as desired to generate a Total Project Cost probability distribution. The process of creating a sample total cost value is diagrammed in Figure 2.

**Figure 2** Diagram Representing Generation of a Sample Value for Total Project Cost



Once this Total Project cost probability distribution has been estimated, it can be used as one of the inputs to the bidding process. Having this distribution, management is in a better position to assess the financial implications of any competitive strategy and has a better picture of the probability of achieving any particular profit level.

**An Illustrative Example**

The application and impact of the proposed model will be illustrated with a hypothetical example consistent with the company’s actual projects.

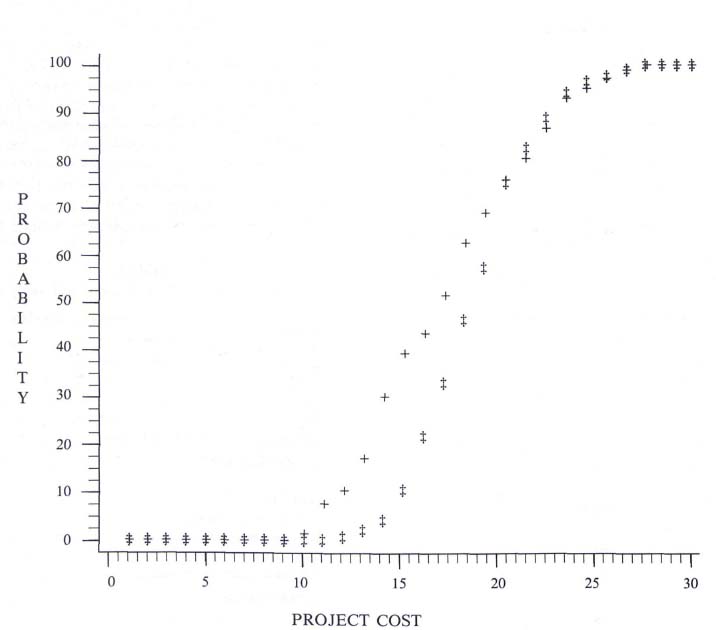
Suppose the company is preparing a bid on a project with estimated direct costs as in Table 4. Category II costs are estimated as follows: Engineering Overhead is 100 percent of Engineering Salaries; Standard Overhead is 50 percent of the sum of Material and Standard Labor; Other Manufacturing is 20 percent of the combined total of Project Labor, Standard Labor, Material, and Miscellaneous; and Sales and General Administration is 25 percent of total direct costs. Application of these percentages gives the estimated Category II costs in Table 4. The other information needed for the risk analysis is that this is a Type II project with an expected length of 20 months.

To develop a probability distribution of Total Project Cost, the sampling process discussed in the previous section was repeated 100 times. Sample values for the component cost ratios were determined by selecting randomly from normal distributions with means given by applying the equations in Table 3 and standard deviations as determined from the regression equations used to fit those equations. The results of the sample are summarized in Table 5 **(With Model)** and the estimated cumulative probability distribution is shown in Figure 3.

**Table 4**  Estimated Costs for Illustrative Example

|  |  |
| --- | --- |
| **Category I** |  |
| Engineering Salaries | $ 100,000 |
| Project Labor | 200,000 |
| Installation | 300,000 |
| Standard Labor | 100,000 |
| Material | 250,000 |
| Miscellaneous | 50,000 |
| Total Category I | $1,000,000 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| **Category II** |  |
| Engineering Overhead | $ 100,000 |
| Standard Overhead | 175,000 |
| Other Manufacturing | 120,000 |
| Sales and General Administration | 250,000 |
| Total Category II | $ 645,000 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| Total Estimated Cost | $1,645,000 |

**Figure 3**  
**Cumulative Probability of Cost**  
Probability **With Model** - Symbol is +  
Probability **Without Model** - Symbol is ‡



|  |  |  |
| --- | --- | --- |
| **Measure** | **With Model** | **Without Model** |
| Mean | $1,677,560 | $1,808,340 |
| Standard Deviation | 398,050 | 277,590 |
| Range | 927,530 ― | 1,176,150 ― |
|  | 2,696,450 | 2,524,960 |

Summary Measures of Risk Analysis Model for Illustrative Example **Table 5**

For comparison purposes, to illustrate the effect of using the regression equations to estimate the cost ratios rather than simply assuming independence and using the historical relative frequency distributions of the ratio of final to original estimate for each direct cost component, a second sample, using the same random numbers, was taken. The results of that sample are also given in Figure 3 and Table 5, designated as **Without Model.**

To check whether the use of the model has a significant effect on the estimation process, a paired comparison t-test3 was done on the differences in the 100 sample outcomes. The t-ratio for the hypothesis that the two procedures, With Model and Without Model, gave the same mean Total Project Cost as 6.08, is highly significant for a sample of 100. It is apparent that the use of the regression equations in the Risk Analysis Model does make a difference. Whether it leads to a better estimate of the probability distribution is another matter. That question cannot be answered without applying the model to a substantial number of actual projects over a period of time.

**Conclusions**

A Risk Analysis Model was developed to estimate the probability distribution of Total Project Cost for use as an input in the competitive project bidding process.

By statistically analyzing a data base of completed projects, the relationships between project characteristics and the ratios of final to originally estimated costs for the different direct cost components and among these ratios were uncovered. These relationships form the basis for the development of a Risk Analysis Model to generate a Total Project Cost probability distribution which can be used to assess the financial implications of any competitive bidding strategy.

Applying the model developed for a particular company to a typical example project shows that the use of the regression-based Risk Analysis Model leads to a significantly different estimated probability distribution for Total Project Cost than assuming independence of component cost ratios.

The question of improved bidding performance through the use of this type of model, however, must wait for testing over a period of time on actual projects.

**Appendix**

Four types of statistical analyses and their associated testing procedures were used in this study. Brief descriptions are given of each. The interested reader is referred to any standard text on statistics, such as Canavos [3], for a more complete discussion and the tables needed to actually perform the tests.

When a statistical test is performed, it is a check on a particular hypothesis about a population parameter (such as the mean) or the relationship between the parameters of two or more populations (e.g., are two or more population means equal). The usual hypothesis is that the parameter (if there is one population) has particular value. If there are two or more populations whose parameters are being compared, we generally test if they are equal. The alternative would be that the single population parameter does not have the specified value or that two or more populations do not have the same parameter value.

When we say that a test has a certain significance level, we are saying how strongly we can accept the alternate hypothesis. The lower the value of the significance level, the stronger the result. This is because the significance level measures the probability of rejecting a true hypothesis of equality. For example, a significance level of .10 means that there is only a ten percent chance that the test based on a sample would say two parameters are different when they are really the same.

**Analysis of Variance (ANOVA)**

A standard problem in statistics is the comparison of several different populations, such as the groups of projects identified by type, to see whether they can be considered to all have the same mean or not. In this case the hypothesis tested was that all four project types had the same mean cost ratio for a given cost component. For example, one of the tests run was on the hypothesis that the cost ratio for Standard Labor was the same for all four project types. This hypothesis was accepted (See Table 1). Another hypothesis tested was that the cost ratio for Project Labor was the same for all project types. This was rejected at a significance level of .01, indicating that it was extremely unlikely that all project types were the same in this respect. That does not mean they are all different, but it does mean that it is reasonable to assume that at least one differs from the rest.

**Regression Analysis**

Regression analysis is a way of fitting a predictive model, linear in this case, to explain the differences between values of one variable (a specific cost ratio) as a function of other variables measured on the same test subjects (projects). In a simple regression model we predict the variable of interest as a function of one other. In multiple regression we predict it as a function of several others. Table 3 shows the regression equations developed for this model. With the exception of Equation (5), they are all simple regressions. To give an example of their use, consider using Equation (5) for a Type II project with a projected length of 20 months. The predicted cost ratio for Material would be:

Ratio = 1.112 + .0015(20) - .209 = .933

We would expect a Material cost ratio of .933.

There are a number of tests that can be done to check whether a model has coefficients (the .0015 and the -.209 in this case) that are significantly different from zero. All those shown are at one significance level or another.

**Paired t-Test**

A common problem is the comparison of two different procedures or methodologies. A way of testing if they have different effects is to use both under conditions that are as identical as possible, thus reducing the effects of differences in the experimental setting (“noise” in the system). The differences between the observations are recorded and tested to see if they are, on average, significantly different from zero. The differences between the observations are recorded and tested to see if they are, on average, significantly different from zero. The pairing of observations was done here by using the same sets of random numbers to sample from the distributions with and without the dependent model.

**Correlation Coefficient Test**

A correlation coefficient measures the strength of the linear relationship between variables. If a graph is plotted with the value of one variable (one cost ratio) plotted against the value of another variable (a second cost ratio) for the same test subject (a project), the correlation coefficient measures how closely the points lie to a straight line. A correlation coefficient close to + 1 or -1 indicates a very good fit. One close to zero indicates very little linear relationship. The sample correlation coefficient can be tested to see if it is significantly different from zero using a t-test, calculating

img

where r is the correlation coefficient for the sample of size n. This can be compared to the critical value for a t distribution at the desired significance level.

**References**

 1.   Bjornsson, H.C. Risk Analysis of Construction Cost Estimates. Transactions, American Association of Cost Engineers, 1981.

 2.   Brown, K.C. A Note on Optimal Fixed-Price Bidding with Uncertain Product Cost. Bell Journal of Economics, Autumn, 1975.

 3.   Canavos, G.C. Applied Probability and Statistical Methods. Boston: Little, Brown and Co., 1984.

 4.   Connor, H.P. What’s The Hang-up with Computers in Construction Cost Estimating? Transactions, American Association of Cost Engineers, 1981.

 5.   Deshmukh, S.S. Risk Analysis. Transactions, American Association of Cost Engineers, 1976.

 6.   Dixie, J.M. Bidding Models – The Final Resolution of a Controversy. American Society of Civil Engineers, Journal of the Construction Division, September, 1974.

 7.   Finch, W.C. & Postula, F.D. Benefits of Probabilistic Cost Estimating for Nuclear Power Plant Projects. Transactions, American Association of Cost Engineers, 1979.

 8.   Finch, W.C., Postula, F.D. & Perry, L.W. Probabilistic Cost Estimating of Nuclear Power Plant Construction Projects. Transactions, American Association of Cost Engineers, 1978.

 9.   Finch, W.C. & Young, K.A. The Application of Probabilistic Cost Estimating to an Advanced Energy System. Transactions, American Association of Cost Engineers, 1981.

10.  Fuerst, M. Theory for Competitive Bidding. American Society of Civil Engineers, Journal of the Construction Division, March, 1977.

11.  Grinyer, P.H. & Whittaker, J.D. Managerial Judgment in a Competitive Bidding Model. Operational Research Quarterly, 1974.

12.  Hamlin, J.L. & Jones, G.R. Profitable Probability. Transactions, American Association of Cost Engineers, 1977.

13.  Hertz, D.B. Risk Analysis in Capital Investment. Harvard Business Review, January, 1964.

14.  Kraemer, G.T. Meaningful Risk Assessment. Transactions, American Association of Cost Engineers, 1976.

15.  Krieg, T.E. Selecting from Candidate Parametric Estimating Relationships. Transactions, American Association of Cost Engineers, 1980.

16.  Spooner, J.E. Probabilistic Estimating. American Society of Civil Engineers, Journal of the Construction Division, March, 1974.

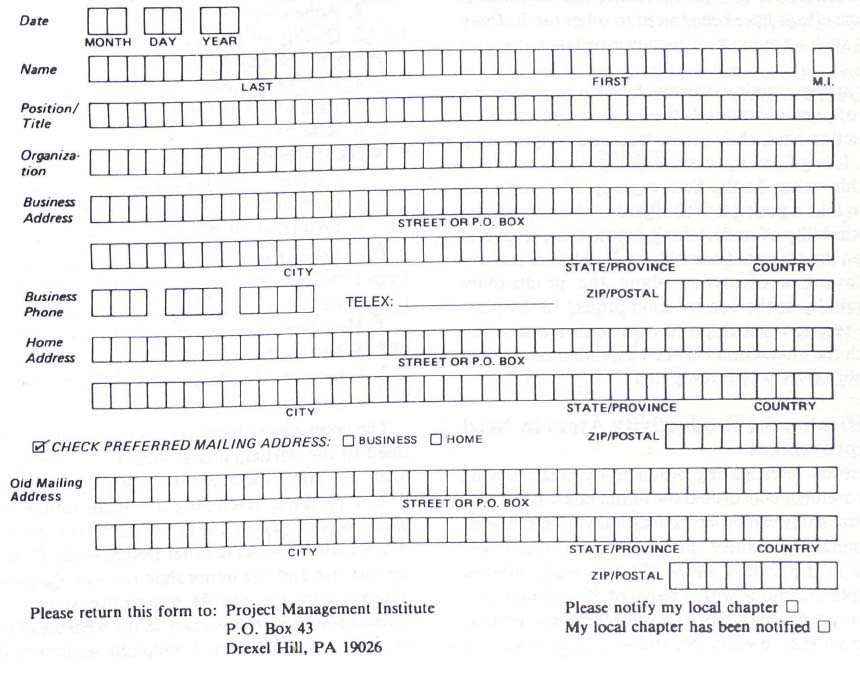
17.  Stark, R.M. & Rothkopf, M.H. Competitive Bidding: A Comprehensive Bibliography. Operations Research, March, 1979.

18.  Uriegas-Torres, C. Statistical Techniques in Cost Engineering. Transactions, American Association of Cost Engineers, 1981.

Dr. David W. Pentico is Associate Professor of Management Science at the School of Business, Virginia Commonwealth University.

\*\*\*

**PROJECT MANAGEMENT INSTITUTE  
NOTICE OF CHANGE OF PERSONAL INFORMATION  
(ENTER ONLY CHANGES TO CURRENT INFORMATION ON RECORD)**



1 It must be noted that the formulas used for applying Category II costs are not necessarily the same ones as those used in making the estimates since overhead rates change annually to reflect overall corporate operating budget changes. This is one source of deviation of final charged costs from the model and thus is part of the “unexplained variation” for the model developed.

2 The techniques used in this study are described briefly in the Appendix. A reader interested in a more complete description is referred to any standard statistics text, such as Cana-vos [3].

3 This test is described briefly in the Appendix.

This material has been reproduced with the permission of the copyright owner. Unauthorized reproduction of this material is strictly prohibited. For permission to reproduce this material, please contact PMI.