EST.04

A Methodology for Estimating Engineering Details

Mr. James D. Whiteside II, PE

W hile historical data supports the idea that engineering is a percent of TIC, the impact of factored engineering to a project is unclear to project managers. Too many project managers leave this clause unchallenged and expose the project to unrealized-growth risks in cost and schedule. Factored engineering can:

- Bury a significant amount of contingency;
- Hide behind growing scope and construction costs;
- Leverage project profit to the front of the project; and
- Avoid discussing the true cost of engineering services.

There is a strong correlation between engineering hours and construction hours. This correlation allows a project team to accurately forecast final project costs and schedule before the final invoice to the engineering contractor is paid and before the project experiences the first project over-run forecast. Engineering is solely responsible for controlling the scope of the project. Since construction hours grow three-fold for every additional hour of engineering, there is no acceptable explanation for project over-run being forecasted when construction is halfway completed. The only explanation for a late over-run forecast is that the correlation between engineering and construction is being misunderstood or ignored.

This paper introduces a concept called an estimate triangle. Estimate performance can be improved through better estimating of engineering, by directly calculating engineering hours from material quantities, and by cross-checking with a correlation from direct labor hours. The focus of this paper will be on calculating engineering hours from material quantities.

STUDY BASIS

The techniques used for engineering analysis and model development are discussed in detail in "Developing Estimating Models" [1]. The data for this paper is drawn from over one hundred completed refining and petrochemical projects ranging from $100,000 to $500,000,000. Absolutely no cost data was used. The estimating methodology is completely based on material quantities and labor hours. Only discipline engineering hours such as those shown in table 1 are analyzed in this paper. Discipline engi-

neering hours are discussed as they relate to direct labor hours. Direct labor does not include construction management. It only includes the hours that directly contribute to the asset under construction.

There should be no expectation that all of the engineering disciplines can be modeled by a single equation. A lot of work is necessary to take a data set and apply several data analysis techniques until something begins to deliver consistent results. A relational database removes much of this drudgery.

FILTERING DATA

Of the hundred projects that comprise the engineering model, hundreds more were discarded. There are a couple of approaches to filtering data. One approach is to manually discard data that is out-of-bounds. For example, if a similar group of projects all have about 20,000 hours of engineering and one has 500 hours, then the low value should be initially removed from the data analysis. The removed data point may be part of a different analysis, or it may be an anomaly.

Another approach is to statistically discard data. In the case of piping, study results have determined that direct-labor hour per pound is a strong correlation. Using statistics, data is discarded that does not fall within one standard deviation of the average hour per pound. Similarly, data filters can be applied to the remaining accounts.

IDENTIFYING THE PROBLEM

Figure 1,"Percent of Total Installed Cost vs. Project Size in Millions of Dollars," shows the result of applying a simple statistical average across the entire project data population. In any given project size, engineering is close to both the industry average and the expected 18 percent of TIC for refining projects.

Figure 2 is the same data set showing the variation of the data for each project size. The ninetieth percentile data point is represented in P90. Ninety percent of the data is below this point. The other percentile marks are handled similarly. The column for projects less than one million dollars appears to depict an engi-

Table 1—Engineering Disciplines.

| Engineering Discipline	Responsibility
Process	Heat and Material Balance, Simulations
Mechanical	All Equipment Specifications
Piping	Piping, Paint, Insulation
Electrical	Power, Cabling and Infrastructure for Instrumentation and Power
Instrumentation	Control Systems, Safety Systems, Relief Devices, Instrument Devices
Steel	Steel, Fireproofing
Civil	Concrete, Foundations, Earthwork
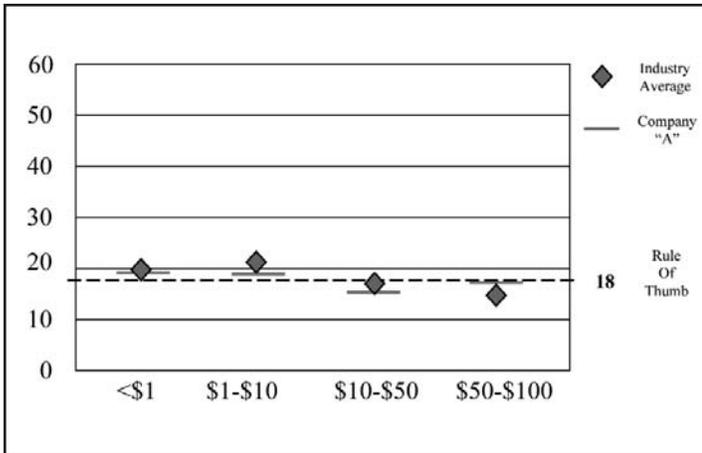
Figure 1—Percent of Total Installed Cost Vs. Project Size in Millions of Dollars.
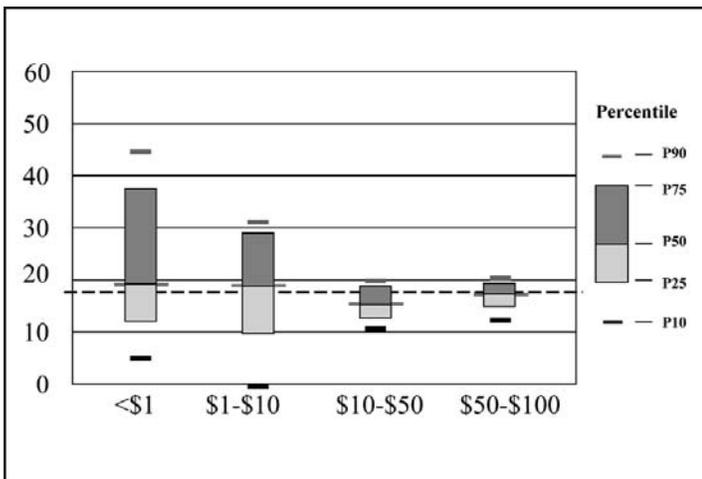

Figure 2—Percent of Total Installed Cost Vs. Project Size in Millions of Dollars (Data Variation Display).

neering group that is out-of-control, thus resulting in a wide variance in the engineering hours required for a project.

Figure 3, "Engineering Hours vs Direct Construction Labor Hours," is a graph of one hundred and 50 projects where each wavelet is a project size class. The graph is an expansion of the first column (Projects < $1MM) of figure 2 and demonstrates that there is a direct correlation between direct labor hours and disci-pline engineering hours. Each wavelet (1-5) demonstrates that organizations improve efficiency as projects grow in size. Organization growth in staffing occurs in step changes, represented by each wavelet. For example, staffing a procurement specialist will also require part-time staffing of contract administrators, expediters, and warehouse personnel. Most organizations can not arbitrarily use individuals for a few hours. Once a position is staffed, the cost is incurred by the project whether or not there is enough work for that person. This is inefficient.

The French mathematician Joseph Fourier (1786-1830) demonstrated that the most general form of periodic waves could be built as a summation of simpler harmonic waves. A Fourier program was used to correlate total engineering hours to total direct labor hours. Applying the Fourier principle in data analysis, as in figure 3, demonstrates that there is a strong behavioral pattern between engineering and direct labor hours. The overall analysis indicates that there are strong correlations in the lesser accounts.

ESTIMATE TRIANGLE

This paper introduces a concept called an estimate triangle. The triangle shows the balance between materials, direct labor and engineering in the estimate. If a person is given any one account, the other two accounts should be easily calculated. The basis of the data analysis makes use of the strong and direct correlation between three accounts:

- Material to direct labor hours;
- Direct labor hours to engineering hours; and
- Material to engineering hours.

Given the plentiful labor data from completed projects, direct labor hours are easily calculated from material quantities and labor productivity (figure 4, side A). In examining the detailed data for projects that perform poorly, the balance in the estimate triangle is still satisfied. However, compared to the funding estimate, poor performance is due to inadequate estimating techniques in the engineering and construction accounts. The estimate details for each side of the estimate triangle must be as robust as the other two sides.

Estimating has traditionally had a very robust side A, an adequate side B, but a poor side C. Estimating engineering using a percent of TIC makes side C collapse. Estimate performance can be improved through better estimating of engineering and by cross-checking between two sides of the triangle.

There is another benefit to developing an estimate of engineering in detail. Since direct labor and engineering hours have a strong correlation, the efficiency between direct labor and engineering can be analyzed.

Engineering can now be analyzed in the same manner that productivity is analyzed for direct labor. Engineering productivity has historically been elusive because there were no readily published data. If collected project data can be transformed into engineering models, then a set of engineering efficiencies and pro-

ductivity indices can also be formulated. In order to reduce engineering costs, the project scope must also be reduced.

PROBLEM SOLVING METHODOLOGY

Engineering is a complex and daunting system to explain. Fortunately, engineers are organized, methodical, slow to change, and constantly evolving (improving) successful systems. Therefore, they are highly predictable. Estimating engineering hours is a quantification of the activity hours. It is not the why or how of what engineers do. Once a mathematical model is successful, it will be stable for a long time. The model can be evolved along with changes in engineering.

There is no shortcut to decide which analysis tool works best. This is left to trial and error and the experience of the analyst. The procedure used to break down complex systems has three major steps. These steps are designed to produce quicker regression results and to break down complex problems. Understanding simpler things may provide insight to more difficult issues. These are the steps that determined the order of processing data and how to correlate engineering hours to material quantities.

I. Divide the problem into simpler ones.

1. Find the accounts that have the strongest correlations with the least amount of analysis work.
2. Evaluate data at the highest level possible before working down to the next level.
3. Evaluate the accounts with the most data.
4. Correlate quantities to direct labor hours.
5. Correlate direct labor hours to engineering hours.
6. Correlate quantities to engineering hours.

II. De-couple the easy issues from the complex ones.

1. Find a different correlation if a correlation do not yield at least a 0.75 R2 (goodness-of-fit) value.
2. Subdivide the data set into ranges, groups, etc.
3. Examine accounts for interdependencies (process/equipment, steel/concrete, electrical/instrumentation).

III. Evaluate the predictive model.

1. Build objectivity by dividing data into two sets, one for regression and one for testing.
2. Check for convergence because all three correlations must converge to the same R2 goodness-of-fit..
3. Test for predictability. Success is defined when the predictive model is within ten percent of actual data across the entire regression data set and within fifteen percent of test data for any given actual data point.
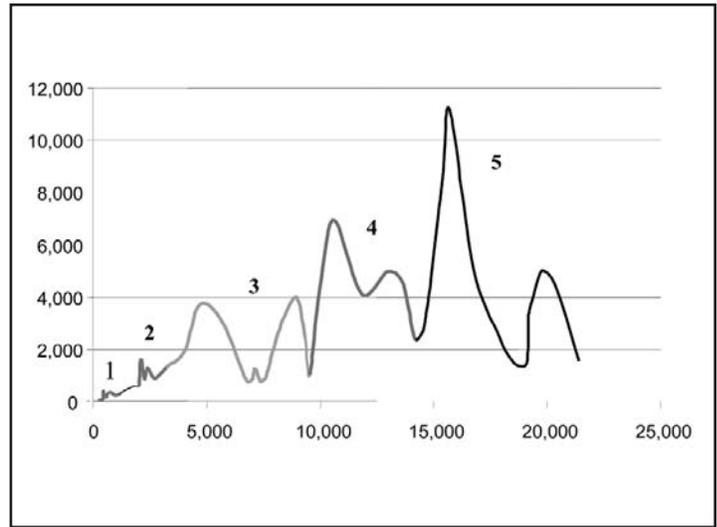

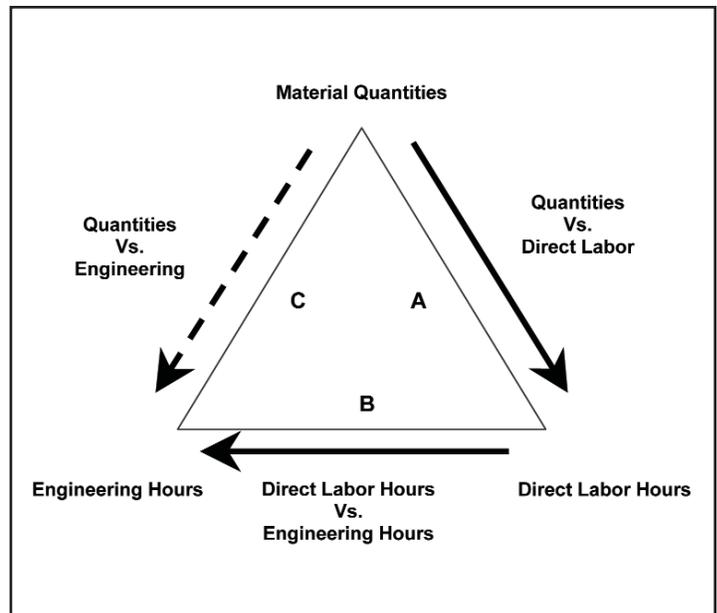Figure 3—Engineering Hours Vs. Direct Construction Labor Hours.

Figure 4—Estimate Triangle.

CONVERGENCE

For an equation to be accepted as model for estimating engineering there has to be three equations for each relationship on the estimate triangle. Rather than publish equations for each accounts, which would be highly proprietary, the chart of correlation results will be provided.

Mechanical, electrical, instrumentation and concrete accounts responded well to simple trend lines. Very little out-of-bound data was discarded. Piping and steel accounts required spectral analysis. Process account required building a modified yield calculation. Mechanical and "Steel > 30 Tons" are examples (table 2) that did not converge. The subjects will be covered in ascending order of complexity.

Table 2—Convergence Results.

Account	"Goodness of Fit" (R^2) Value			
	Quantity Vs. Direct Labor Hours	Direct Labor Hours Vs. Engineering Hours	Quantity Vs. Engineering Hours	Length Vs. Engineering Hours
Mechanical 1	0.588	0.374	0.798	
Mechanical 2	0.731	0.966	0.942	
Concrete	0.942	0.878	0.901	
Instrumentation	0.717	0.763	0.657	
Electrical	0.4633	0.507	0.526	
Steel < 20 Equip	0.794	0.754	0.600	
Steel > 20 Equip	0.846	0.464	0.896	
Pipe 1"	0.942	0.984	0.938	0.954
Pipe 2"	0.967	0.981	0.935	0.960
Pipe 3"	0.871	0.742	0.747	0.683
Pipe 4"	0.710	0.630	0.839	0.568
Pipe 5"	0.744	0.958	0.694	0.633
Pipe 6"	0.948	0.948	0.924	0.521
Pipe 7"	0.958	0.851	0.942	0.896
Pipe 8"	0.980	0.877	0.932	0.896
Pipe 9"	0.921	0.933	0.956	0.941
Pipe 10" +	0.997	0.933	0.956	0.890
Pipe All Sizes	0.638	0.783	0.630	0.644

Hours Vs. Weight
(All Pipe Diameters)
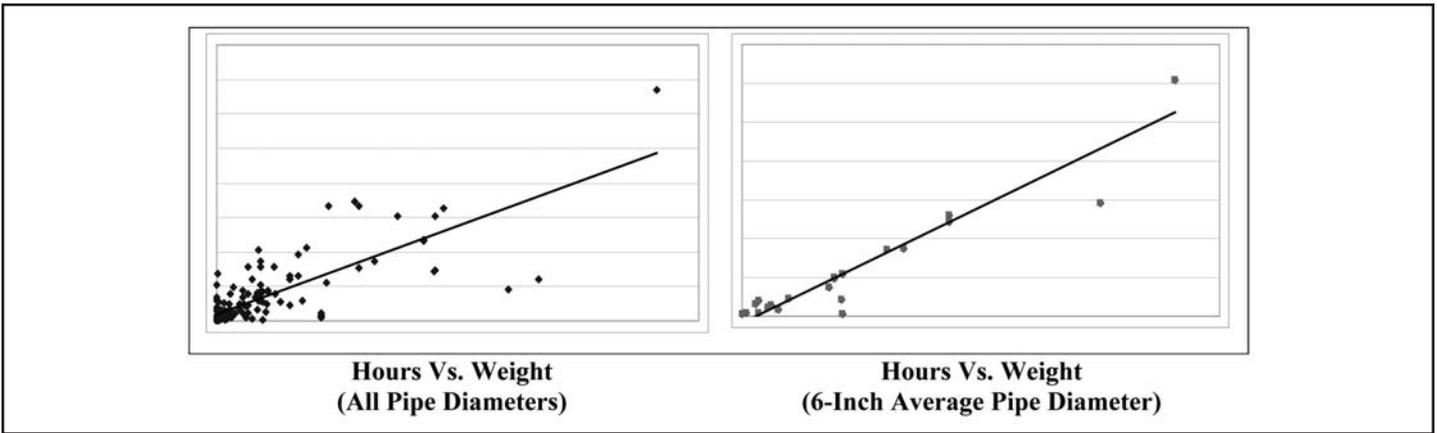
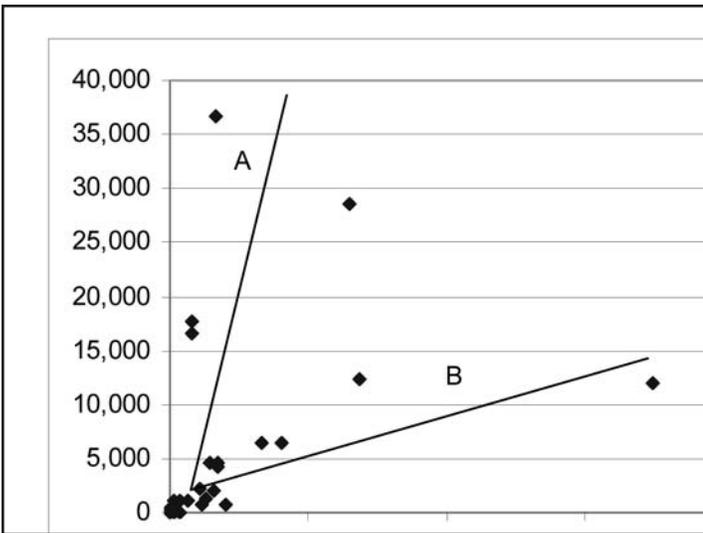Hours Vs. Weight
(6-Inch Average Pipe Diameter)

Figure 5

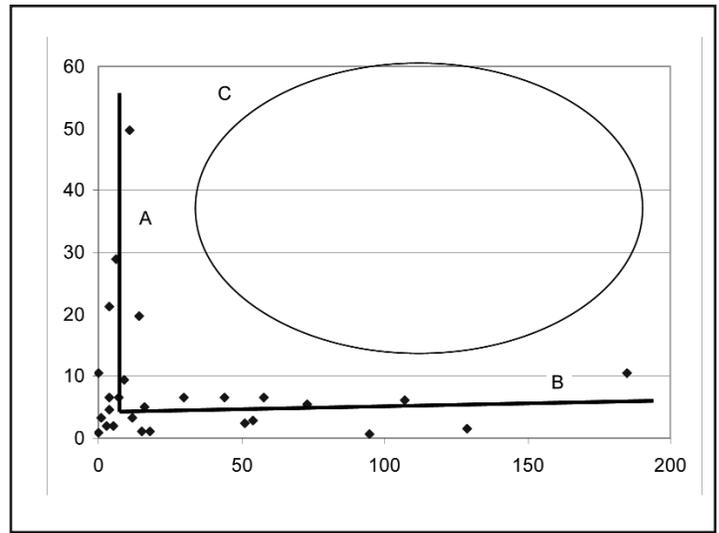Figure 6—Engineering Hours Vs. Steel Tonnage.

Figure 7—Delineation Factor (DF) TechniqueDirect Labor: Hours / Engineering Hours Vs. Total Equipment.

MECHANICAL, CONCRETE, ELECTRICAL, INSTRUMENTATION ENGINEERING

For mechanical, instrumentation, electrical, and concrete data, a simple trend line produced a reasonable average. Correlating hours to the total mechanical equipment count did not produce a convergence (table 2, Mechanical 1). Mechanical 2 shows improvement to the correlation when the mechanical account was correlated separately to various equipment classes, pumps, exchangers, vessels, etc. It would seem reasonable that the electrical account would have a better correlation on an hour to weight basis. In this group, a goodness of fit (R^2) between 0.5 and 0.7 is not significant enough to detract from using the simple regressions as averages to calculate engineering hours. Missing any one of these accounts by thirty percent will not change the total engineering hours by more than five percent. Based on the collected data, the correlations for these accounts are strong enough to determine that each side of the estimate triangle is satisfied.

PIPING ENGINEERING

Analyzing piping over all sizes produced a satisfactory hour to weight correlation. There is, however, a significant improvement in the correlation when piping is broken into various average pipe sizes. Figure 5A, "Hours Vs. Weight (All Pipe Diameters)" and figure 5B "Hours Vs. Weight (6-inch Average Pipe Diameter)," compare the difference between the piping regression of all diameters and of six-inch average pipe diameter.

Analyzing piping based on length produced a larger standard deviation. The "Length Vs Engineering Hours" column in table 2 shows the correlation for piping. In nearly all cases, the weight basis (columns A and C) produced better results than the length basis. One reason the weight basis works better is that a joint of pipe is installed at approximately the same rate per pound as the same joint of pipe with a valve attached. For example, a valve weighs as much as a joint of pipe and is about twenty times shorter. The joint of pipe may take 1.6 hours per foot and the valve may take 11 hours to completely install. Using the length basis, the valve adds less than a foot to the length of pipe or less than 1.6 hours to the installation calculation. However, the pipe and the valve both are installed at approximately weight per hour.

STEEL ENGINEERING

Steel presented a problem during analysis. At first there appeared to be no correlation between direct labor and engineering hours. It is unclear in figure 6, "Engineering Hours Vs. Steel Tonnage," if this is a scattered data plot or there are two different population trends (A and B) in the data. All simple correlation combinations of labor hours, engineering hours, equipment count, and steel tonnage produced similar plots.

All the study data is collected in short tons. Since a short ton is 0.907 metric tons (1,000 kg), the results and figures are essentially the same.

A delineation factor (DF) technique was applied to decide if the data bifurcates or if there are parallel data trends. Bifurcation is a discipline of chaos theory that deals with nonlinear phenomena. It means that the system splits from one state into two possible states. Systems that increase exponentially in complexity, like steel, tend to bifurcate. For example, small tonnages of steel are designed piece by piece. On the high end, like open bay, multiple floor structures, steel is designed by advanced engineering computing models.

A DF technique exaggerates discontinuities in a data set. The point of discontinuity is where the data should be divided into separate data sets. A plot is made of a simple value against a complex value, the delineation factor. The complex value (DF) is a function of a principle value and an associated value(s). In this case, the DF was obtained by dividing the associated value (direct labor hours for steel) by the principal value (engineering hours for steel) and plotted against simple value (total equipment).

The intersection of lines A and B in figure 7, "Delineation Factor (DF) Technique," indicate that the division in data population is approximately 10 tons of steel. The clear field represented by the oval (C) in figure 7 provides further graphical proof that there are two different states of steel design. If there had been data in field, then another approach to analyzing steel would have had to be taken. Clearly, the steel data represents two different populations for this data sample.

Apply the delineation technique to different equipment subsets until all the major nodes are found. The next equipment subset would be to drop data with less than ten pieces of equipment. The result is shown in figure 8, "Equipment to Steel Complexity Branches (Engineering Hours Vs. Steel Tonnage)." Each complexity of steel to equipment produces good convergence and correlation. Steel accounts seem to split at 5, 10, 20, 75 and 100 pieces of equipment. Equipment to steel complexity differences are represented by each branch in Figure 8.

PROCESS ENGINEERING

This is the most difficult category to estimate. Examining 95 completed projects uncovered six drivers to process engineering hours. The six drivers that are significant to the calculation of process hours are pipe weight, exchangers, pumps, compressors, vessels and furnaces. Process engineering for piping does not seem that it should be included as one of the six drivers. However, most engineering contractors assign the hydraulics to process engineers because the software that is used for process simulation can easily perform hydraulic calculations. The hours are accounted in the process engineering account regardless of who performs the calculations.

Of the entire project data set, it became obvious that there was no single universal regression to cover all equipment configurations for all size projects. The data population was therefore subdivided according to project characteristics.

Project Characteristics

There are three types of projects. Large capital projects have at least forty pieces of equipment, and all equipment groups are typically present. Re-vamps, work on existing process units that may or may not be running during construction, are divided into two groups: 20 to 40 pieces of equipment and 5 to 20 pieces of equipment. Small projects with a half dozen pieces of equipment and some equipment groups present may not have equipment.

For a given project type (large, revamp, small), a correlation was developed for each equipment type as it related to the total hours in process engineering. Miscellaneous equipment count is added to the equipment account total that closely resembles the miscellaneous equipment's principle configuration. For example, the process specification for an exchanger is developed by process engineers, but the mechanical design is typically performed by the manufacturer. If the miscellaneous equipment is similar to an exchanger, then add it to the exchanger total.

Weighted Calculation

Figure 9, "Equipment Count Vs. Process Hours," is an example of where the equipment count to process hours was a simple straight-line regression. For any project, process hours are the summation of five equipment equations and one piping equation. Each of these products is weighted by the R^2 value for the equipment account divided by the total R^2 of all the equipment accounts.

$$\Pr ocess\ hours = \sum f(n) \times \left(R^2(n) \div \sum R^2\right) for\ n = 1\ to\ 6$$

Table 3, "Weighted Calculation," shows that for any given project, the process hours are calculated within ten percent of the actual hours. The predicted process hours are ten percent higher than the actual hours for nearly the entire set of projects. The ranges in the delta column are important inputs to running a Monte Carlo risk analysis. Table 3 also demonstrates that not all of the accounts need to be populated to calculate accurate estimates of process hours. This is the benefit of a weighted calculation.

CALCULATING ENGINEERING

The process of estimating engineering begins by developing a study to produce the regression equations for all sides of the Estimating Triangle. This means that there will be at least thirty

Table 3—Weighted Calculation.

Project	n=1 Exchanger	n=2 Pump	n=3 Compressor	n=4 Vessel	n=5 Furnace	n=6 Pipe Wt	Predict	Actual	Delta
1	13	4	2	22	3	470,470	13,968	13,937	1.00
2	14	18	0	12	1		10,667	11,582	1.09
3	18	15	0	14	1	0	8,359	7,419	0.89
4	22	13	6	15	1		19,778	19,648	0.99
5	26	16	1	22	2	1,484,423	21,734	18,815	0.87
6	35	24	0	21	1	0	14,476	13,694	0.95
7	36	24	0	24	3	0	26,093	24,288	0.93
8	45	26	5	33	2	2,810,780	38,494	31,893	0.83
9	2	0	2	2	0	106,300	2,819	1,548	0.55
						Sum	156,389	142,824	0.91
						Average	17,377	15,869	0.90

equations to describe the data. Once this is done, program automation can be used.

Estimating engineering starts with estimating the number of pieces of equipment (by major account) and material quantities. Estimating quantities from an analogous project is a good start for conceptual estimates. This data is simply entered into the study equations to calculate engineering hours by discipline. This is good for immediate and near-term answers. As new data is acquired into the database, the study equations will updated and replace by newer ones.

The best solution is to use automation to create new regressions based on the techniques documented in this study. Newly acquired data is fed into a program that performs the tasks of: retrieving, dividing, regressing, filtering and evaluating to produce new engineering equations. These steps can be performed manually or by a program.

Retrieve

Export account data into a table using the queries from the study that developed the first set of study regressions. This could include allowing data retrieval by project characteristics. Though this study covered large, revamp and small projects, there are other project type variations,contract type, location, facility type, time of year, etc., that could make the data more project specific.

Divide

Divide the data set into two sets: one for regression and one for acceptance testing. If there are fewer than a dozen points, skip this step. Fewer than a dozen data points does not render an adequate picture of the data characteristics.

Regress New Equations

Using the data set for regression, new equations will need to be developed to determine which equation best fits the data points. Programming does not allow for dynamic updates to the
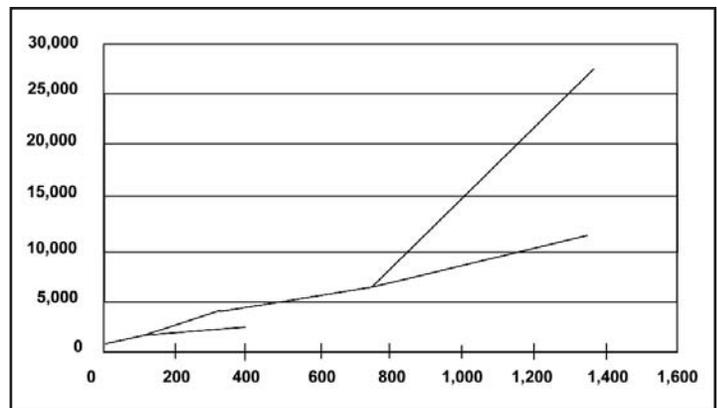

Figure 8—Equipment to Steel Complexity Branches (Engineering Hours Vs. Steel Tonnage).

source code. Some creativeness will need to be made to process basic equation formats by storing coefficients in a table for the program to use later.

Filter

Remove data which does not fall into one standard deviation. Another pass may need to be made to provide a narrower bandwidth, depending on the application. A funding estimate may need a data bandwidth that is much tighter than a conceptual estimate.

Evaluate

Calculate engineering hours by discipline against the projects in the test group. If the regressions for the Estimate Triangle converge and are acceptable, then calculate the engineering hours for the project being estimated.
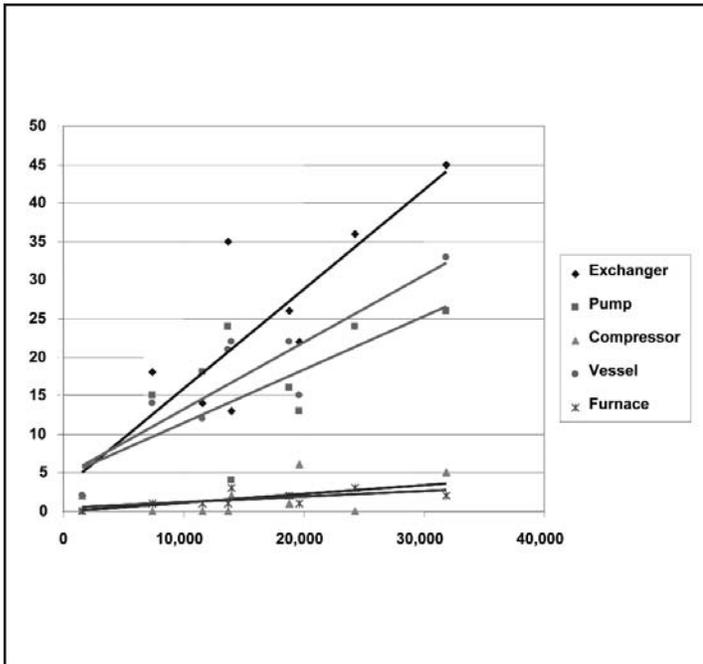
Figure 9—Equipment Count Vs. Process Hours.

Table 4—Engineering Input Parameters.

Description	Quantity
Exchangers	13
Pumps	2
Vessels	9
Furnaces	1
Total Equipment	25
Pipe	357,276 Lb 4.7" Avg. Diameter 10,840 Ft
Electrical	16,830 Feet
Instruments	366
Steel	192 Tons
Concrete	2,171 Cubic Yards

Table 5—Engineering Evaluation.

Engineering Discipline	Prediction	Actual	Efficiency
Process	999	322	3.10
Piping	8,742	15,951	0.54
Electrical	2,000	1,010	1.98
Instrumentation	8,364	3,259	2.26
Steel	1,600	2,375	0.67
Concrete	8,684	5,836	1.48
Mechanical	1,047	1,865	0.56
Total	31,436	30,618	1.02

Table 6—Engineering Evaluation Adjusted.

Engineering Discipline	Prediction	Contractor Estimate	Efficiency
Process	249	322	0.77
Piping	14,342	15,951	0.89
Electrical	2,000	1,010	1.98
Instrumentation	4,364	3,259	1.33
Concrete, Steel	8,684	8,211	1.05
Mechanical	1,047	1,865	0.56
Total	30,683	30,618	1.00

ACCOUNT ADJUSTING

There are still account adjustments that need to be made depending on the contractor. The hours in the database are stored in only one way. Each contractor has a different accounting system used to store their hours. Once the hours are calculated, they need to be translated into the contractor's format.

Using table 4, "Engineering Input Parameters," produces the output found in table 5, "Engineering Evaluation." Any account efficiency problems in the adjusted engineering evaluation must be discussed with the contractor.

Table 5 shows the difference between the total engineering hours predicted and the actual contractor hours. During estimate reviews, the actual column will be the estimated hours for evaluation purposes. The example in table 5 shows that the total difference between the two is less than ten percent. If funding estimates are to be within ten percent of the estimate, this account is correct

and contingency could be added to fund another 3,000 hours. This would bring the estimate for engineering to a fifty percent probability of success depending on company practices.

The sub accounts seem to indicate that something has gone wrong in the predicted calculations. Remember, the regressions are focused on returning average values and there could be an accounting difference from the way the database stores data and the contractor's accounting system.

For example, the contractor may have sent work to piping for structural support issues. This may shift 1,600 hours to piping. The contractor may have combined civil and steel to one account.

There may have been process engineering study work that was not included in the estimate. The prediction value includes front-end loading work. Generally, three quarters of the process development work occurs during front-end loading. In this case that would be about 750 hours, leaving 249 for the estimate.

Instrumentation in the prediction column includes "typical" instrumentation. The contractor may have decided that pressure relief devices are specified by the instrumentation group, but the rest of the work falls into piping. Pressure gauges, thermowells, etc. may also have been transferred to piping for design and installation. In this case, that could shift 4,000 hours to piping.

Table 6, "Engineering Evaluation Adjusted," represents adjustments for accounting issues that may affect the project. The total hours did not change appreciably, and the other accounts now look more reasonable. This table now provides the project team with a starting point to discuss how to trim the project budget. Since the hours appear reasonable, the only way to trim the project budget is to cut project scope which defines the quantities in table 4.

Application

The engineering hours represented eleven percent of TIC for the actual project from which tables 4 through 6 are based. As a result of using this estimating triangle, about 25,000 hours were removed from the initial estimate supplied by the engineering contractor.

*U*sing a relational database and populating detailed data can produce very accurate estimates of engineering hours based on material quantities in the same manner as construction estimates are built. Engineering and direct labor hours have a strong correlation. All sides of the estimating triangle can be satisfied.

This means that lump sum bids can be accurately evaluated based on a plot plan, sized equipment list, and a few hours time. Being able to estimate engineering can assist in project negotiations and facilitate project scope reduction as early as the first conceptual estimate.

Since engineering can be calculated in detail, using the "eighteen percent of TIC" rule can now be challenged. Estimate scope can be cross-checked between materials, direct labor and engineering. For those who now take the time to calculate engineering hours will have a competitive advantage.

REFERENCES

1. Whiteside, James D. II. *Developing Estimating Models.* **2004 Cost Engineering**. (vol. 46 2004): 23-30.

Mr. James D. Whiteside II, PE
Conoco Phillips
306 Huckleberry
Lake Jackson, TX 77566

E-mail: jim.d.whiteside@conocophillips.com