

EST.03

An Introduction to Parametric Estimating

Mr. Larry R. Dysert, CCC

AACE International describes *cost estimating* as the “predictive process used to quantify, cost, and price the resources required by the scope of an asset investment option, activity, or project ^[1].” The methods and techniques used to prepare a cost estimate will typically vary based on the level of project definition available at the time the estimate is prepared ^[2, 3]. Early in a project’s lifecycle, there is a need to prepare conceptual cost estimates to support timely decision-making such as to make a “go or no-go” decision to continue development of a project proposal or to select between alternative designs.

This paper is aimed at those wanting to explore the concepts of parametric estimating. Parametric cost estimating models are useful tools for preparing early conceptual estimates when there is little technical detail to provide the basis to support using more detailed estimating methods. This paper will introduce the concept of parametric estimating and discuss the steps involved in the creation of a parametric estimating model.

PARAMETRIC ESTIMATING

A parametric estimating model is a mathematical representation of cost relationships that provide a logical and predictable correlation between the physical or functional characteristics of a project (plant, process system, etc.) and its resultant cost ^[4]. A parametric estimate is comprised of cost estimating relationships and other parametric estimating functions that provide logical and repeatable relationships between independent variables (such as design parameters or physical characteristics) and the dependent variable (cost). The independent variables are known as cost drivers, and typically may be physical, performance, or operational characteristics associated with the project to be estimated.

Capacity factor and equipment factors estimates are simple examples of parametric estimates; however sophisticated parametric models typically involve several independent variables or cost drivers. Similar to other conceptual estimating methods, parametric estimating is reliant on the collection and analysis of previous project cost data in order to develop the cost estimating relationships (CERs). An underlying assumption of parametric estimating is that the historical framework on which the parametric model is based is applicable to the new project (i.e., the technology has not radically changed).

Parametric estimating provides several advantages as an estimating technique. Parametric estimates are:

- **Efficient:** They not only allowing estimates to be prepared in much less time than required by more detailed techniques, but require less engineering and level of project definition to support the estimate.
- **Objective:** Parametric models require quantitative inputs that are linked to algorithms providing quantitative outputs. All costs are traceable.
- **Consistent:** If two estimators input the same values for parameters, they will get the same resulting cost. Parametric models also provide a consistent estimate format and estimate documentation.
- **Flexible:** Parametric models provide costs for a range of input values, extrapolating to derive costs for projects of a different size or nature than you may have history for. The models can be easily adjusted to provide cost sensitivity analysis for proposed design changes.

- Defensible: The models highlight the design parameters used, and can provide key statistical relationships and metrics for comparison with other projects. Statistical measures, such as R^2 , and t and F statistics for individual CERs can provide validity to the model.

The development of a parametric estimating model can appear to be a daunting task; however the use of modern computer technology (including popular spreadsheet programs) can make the process tolerable, and much easier than it would have been years ago. The process of developing a parametric model should generally involve the following steps^[5, 6]:

- Cost model scope determination
- Data collection
- Data normalization
- Data analysis
- Data application
- Testing
- Documentation

Determining the scope of the parametric model

The first step in developing a parametric model is to establish its scope. This includes defining the end use of the model, the physical characteristics of the model, the cost basis of the model, and the critical components and cost drivers. The end use of the model is typically to prepare conceptual estimates for a process plant or system. The type of process to be covered by the model, the type of costs to be estimated by the model (TIC, TFC, etc.), the intended accuracy range of the model, etc. should all be determined as part of the end use definition.

The model should be based on actual costs from complete projects, and reflect your organization's engineering practices and technology. The model should generate current year costs or have the ability to escalate to current year costs. The model should be based on key design parameters that can be defined with reasonable accuracy early in the project scope development, and provide the capability for the estimator to easily adjust the derived costs for specific complexity or other factors affecting a particular project. The parameter ranges that the model will be applicable to should be defined (i.e., the model should not be used for projects where the parameters are far outside the range of parameters used to develop the model).

Collecting data to support model development

Data collection and development for a parametric estimating model requires a significant effort. The quality of the resulting model can be no better than the quality of the data it is based upon. Both cost and scope information must be identified and collected. The level at which the cost data is collected will affect the level at which the model can generate costs, and may affect the derivation of the CER's.

It is best to collect cost data at a fairly low level of detail^[7]. The cost data can always be summarized later if an aggregate level of cost information provides a better model. It is obviously important to include the year for the cost data in order to normalize costs later. Location may also be important to support cost normalization.

The scope information should include all proposed design parameters or key cost drivers for the model, as well as any other information that may affect costs (level of complexity, schedule type, etc.). The type of data to be collected is usually decided upon in cooperation with the engineering and project control communities. It is recommended to create a formal data collection form that can be consistently used, and revised if necessary.

Normalizing data to support model development

After the data has been collected, the next step in the process of developing a parametric model is to normalize the data before the data analysis stage. Normalizing the data refers to making adjustments to the base cost data to account for the differences between the actual basis of the data for each project, and a desired standard basis of data to be used for the parametric model.

Typically, data normalization implies making adjustments for escalation, location, cost scope, site conditions, and system specifications. Costs from the historical projects should be adjusted to a common time frame and a common location serving as the standard basis for the model. Normalizing for cost scope implies making adjustments (i.e.,

deleting costs) for the unique scope that a specific historical project incurred but that is not included in the standard basis for the parametric model. A comments section on the data collection forms may be required to note such occurrences.

Analyzing data to support model development

Data analysis is the next step in the development of a parametric model. There are many diverse methods and techniques that can be employed in data analysis, and are too complex to delve into in this paper. Typically, data analysis consists of performing regression analysis of costs versus selected design parameters to determine the key cost drivers for the model. You may need to use reasoned hypotheses or expert opinion to initially identify and test the best cost drivers.

Most spreadsheet applications now provide regression analysis and simulation functions that are reasonably simple to use. The more advanced statistical and regression programs have goal seeking capabilities, which can also make the process easier.

Generally, a series of regression analysis cases (linear and non-linear) will be run against the data to determine the best algorithms that will eventually comprise the parametric model. The algorithms will usually take one of the following forms:

$$\$ = a + bV_1 + cV_2 + \dots$$

Linear Relationship

$$\$ = a + bV_1^x + cV_2^y + \dots$$

Non-Linear Relationship

where V_1 and V_2 are input variables; a, b, and c are constants derived from regression; and x and y are exponents derived from regression. Note that there are many forms of non-linear relationships that may apply.

The various relationships (cost versus design parameters) are first examined for “best-fit” by looking for the highest “R-Squared” value. R^2 has the technical sounding name of “coefficient of determination”, and is commonly used as a measure of the goodness of fit for a regression equation. In simple terms, it is one measure of how well the equation explains the variability of the data. The resulting algorithms from the regression analysis are then applied to the input data sets to determine on a project-by-project basis how well the regression algorithm predicts the actual cost.

Regression analysis can be a time consuming process (especially with the simple regression tools of a spreadsheet program), as iterative experiments are made to discover the best-fit algorithms. As an algorithm is discovered that appears to provide good results, it must be tested to ensure that it properly explains the data. Advanced statistical tools can quicken the process but can be more difficult to use. Sometimes, you will find that erratic or outlying data points will need to be removed from the input data in order to avoid distortions in the results. It’s also very important to realize that many costs relationships are non-linear, and therefore one or more of the input variables will be raised to a power (as in the equation above). You will need to experiment both with the variables you are testing against, and the exponential powers used for the variables. Regression analysis tends to be a continuing trial-and-error process until the proper results are obtained that appears to explain the data. Several individual algorithms may be generated and then later combined into a complete parametric model.

Creating the parametric model application

The data application stage of the development process involves establishing the user interface and presentation form for the parametric cost model. Using the mathematical and statistical algorithms developed in the data analysis stage, the various inputs to the cost model are identified; and an interface is developed to provide the estimator with an easy and straightforward way in which to enter this information. Electronic spreadsheets such as Excel provide an excellent mechanism to accept estimator input, calculate costs based upon algorithms, and display the resulting output.

Testing the parametric model

One of the most important steps in developing a cost model is to test its accuracy and validity. As mentioned previously, one of the key indicators of how well a regression equation explains the data is the R^2 value. It provides a measure of how well variability in the underlying data is explained by the model, and is commonly used as a proxy for how well the algorithm predicts the calculated costs. However, a high R^2 value by itself does not imply that the relationships between the parameter inputs and the resulting cost are statistically significant.

Once you have performed the regression analysis, and obtained an algorithm with a reasonably high R^2 value, you still need to examine the algorithm to ensure that it makes common sense. In other words, perform a cursory examination of the model to look for the obvious relationships that you expect to see. If the relationships from the model appear to be reasonable, then you can run additional tests for statistical significance (t-test and F-test), and to verify that the model is providing results within an acceptable range of error. Briefly, the t-statistic helps us identify whether each individual independent variable is a good predictor of costs, and the F-statistic tells us whether the regression as a whole is a good parametric model.

One of the quick checks to run is to test the regression results directly against the input data to see the percent error for each of the inputs. This lets you quickly determine the range of error, and interpreting the results can help you to determine problems and refine the algorithms. After all of the individual algorithms have been developed and assembled into a complete parametric cost model, it is important to test the model as a whole against new data (data not used in the development of the model). You should consult statistical texts for more information about testing regression results and cost models.

Documenting the parametric model

Lastly, the resulting cost model and parametric estimating application must be documented thoroughly. A user manual should be prepared showing the steps involved in preparing an estimate using the cost model, and describing clearly the required inputs to the cost model. The data used to create the model should be documented, including a discussion on how the data was adjusted or normalized for use in the data analysis stage. It is usually desirable to make available the actual regression datasets, and the resulting regression equations and test results. All assumptions and allowances designed into the cost model should be documented, as should any exclusions. The range of applicable input values, and the limitations of the model's algorithms should also be explained.

PARAMETRIC MODEL EXAMPLE

As an example of developing a parametric estimating model, we will examine the costs and design parameters of Induced Draft Cooling Towers. These units are typically used in industrial facilities to provide a recycle cooling water loop. The units are generally prefabricated, and installed on a subcontract or turnkey basis by the vendor. Key design parameters that appear to affect the costs of cooling towers are the cooling range, approach, and flow rate. The cooling range is the difference in temperature between the hot water entering the cooling tower and the cold water leaving the tower. The approach is the difference in the cold water leaving the tower and the design wet bulb temperature of the ambient air; and the flow rate measures the desired cooling capacity of the tower.

Table 1 provides the actual costs and design parameters of six recently completed cooling towers. The costs have been normalized (adjusted for location and time) to a Northeast U.S., Year 2000 timeframe.

2008 AACE INTERNATIONAL TRANSACTIONS

Induced Draft Cooling Tower Costs and Design Parameters

Cooling Range (Deg F)	Approach (Deg F)	Flow Rate (1000 GPM)	Actual Cost
30	15	50	\$1,040,200
30	15	40	\$787,100
40	15	50	\$1,129,550
40	20	50	\$868,200
25	10	30	\$926,400
35	8	35	\$1,332,400

Table 1 – Cost and Design Information for Recent Cooling Tower Projects

This data provides the input to the data analysis steps of running a series of regression analyses to determine a sufficiently accurate algorithm for estimating costs. After much trial and error, the following cost estimating algorithm was developed:

$$\text{Cost} = \$86,600 + \$84,500(\text{Cooling Range in Deg F})^{.65} - \$68,600(\text{Approach in Deg F}) + \$76,700(\text{Flow Rate in 1000GPM})^{.7}$$

From this equation, we can see that the cooling range and flow rates affect costs in a non-linear fashion (i.e., they are raised to an exponential power), while the approach affects costs in a linear manner. In addition, the approach is negatively correlated with costs. Increasing the approach will result in a less costly cooling tower (as it increases the efficiency of the heat transfer taking place). These appear to be reasonable assumptions. In addition, the regression analysis resulted in an R² value of .96, which indicates the equation is a “good-fit” for explaining the variability in the data; and the F-Test shows statistical significance between the input data and the resulting costs.

In Table 2, the design parameters are displayed as used in the model (raised to a power where needed) and shown against the actual costs and the predicted costs from the estimating algorithm. In addition, the amount of the error (the difference between the actual and predicted costs), and the error as a percent of actual costs are shown. The percentage of error varies from -4.4% to 7.1% for the data used to develop the model.

Induced Draft Cooling Tower Predicted Costs from Parametric Estimating Algorithm

Cooling Range (Deg F)^{.65}	Approach (Deg F)	Flow Rate (1000 GPM)^{.7}	Actual Cost	Predicted Cost	Error	% Error
9.12	15	15.46	\$1,040,200	\$1,014,000	-\$26,200	-2.5%
9.12	15	13.23	\$787,100	\$843,000	\$55,900	7.1%
11.00	15	15.46	\$1,129,550	\$1,173,000	\$43,450	3.8%
11.00	20	15.46	\$868,200	\$830,000	-\$38,200	-4.4%
8.10	10	10.81	\$926,400	\$914,000	-\$12,400	-1.3%
10.08	8	12.05	\$1,332,400	\$1,314,000	-\$18,400	-1.4%

Table 2 – Predicted Costs for Cooling Tower Parametric Estimating Example

Using the estimating algorithm developed from regression analysis, we can develop tables of costs versus design parameters (Table 3), and plot this information on graphs (Figure 1).

2008 AACE INTERNATIONAL TRANSACTIONS

Induced Draft Cooling Tower Costs Based On Parametric Model

Cooling Range (Deg F)	Approach (Deg F)	Flow Rate (1000 GPM)	Predicted Cost
30	15	25	\$559,000
30	15	30	\$658,000
30	15	35	\$752,000
30	15	40	\$843,000
30	15	45	\$930,000
30	15	50	\$1,014,000
30	15	55	\$1,096,000
30	15	60	\$1,176,000
30	15	65	\$1,254,000
30	15	70	\$1,329,000
30	15	75	\$1,404,000
40	15	25	\$717,000
40	15	30	\$816,000
40	15	35	\$911,000
40	15	40	\$1,001,000
40	15	45	\$1,089,000
40	15	50	\$1,173,000
40	15	55	\$1,255,000
40	15	60	\$1,334,000
40	15	65	\$1,412,000
40	15	70	\$1,488,000
40	15	75	\$1,562,000

Table 3 – Data for Cost Graph Based on Parametric Estimating Example

This information can then be rapidly used to prepare estimates for future cooling towers. It would also be very easy to develop a simple spreadsheet model that will accept the design parameters as input variables, and calculate the costs based on the parametric estimating algorithm.

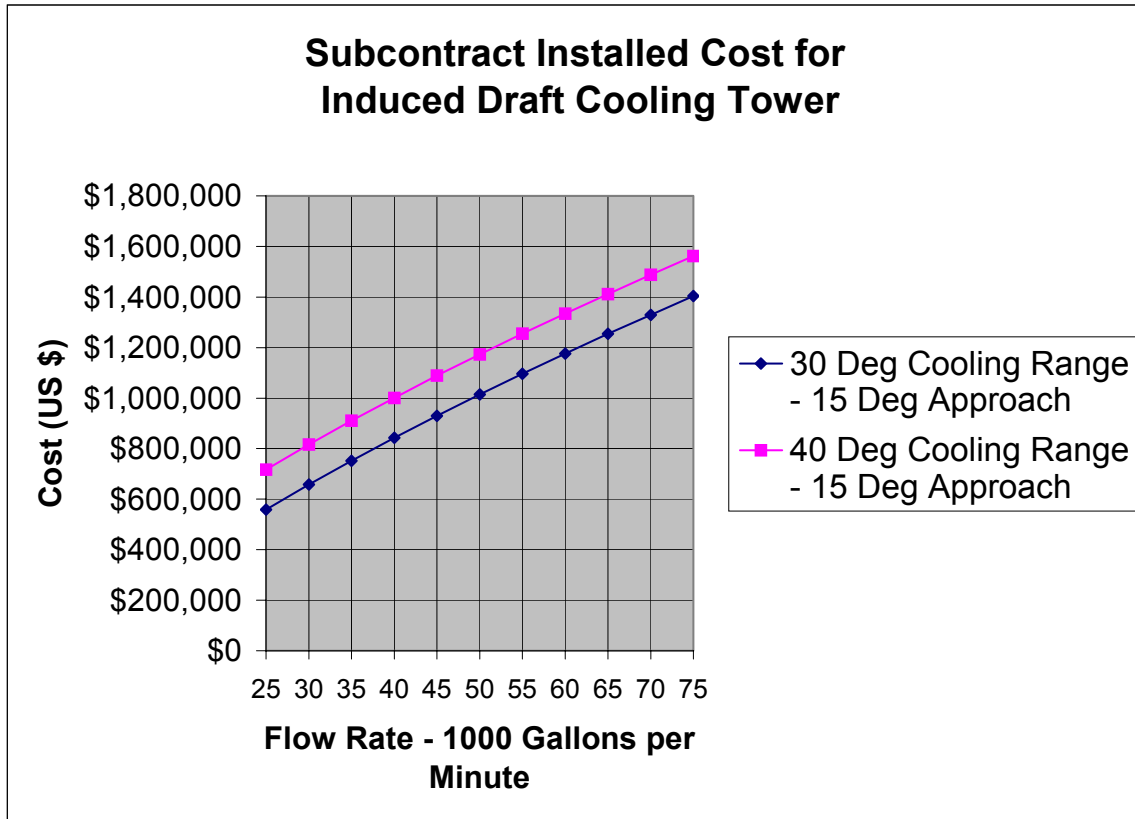


Figure 1 – Graph of Cooling Tower Costs Based on Parametric Model

CONCLUSION

Parametric cost models can be a valuable resource in preparing early conceptual estimates. They are often used during both the concept screening and feasibility stages of a project. Parametric models can be surprisingly accurate for predicting the costs of even complex process systems. Parametric estimating models can be developed using basic skills in estimating, mathematics, statistics, and spreadsheet software. It is important to understand that the quality of results can be no better than the quality of the input data, and great care should be taken during the data collection stage to gather appropriate and accurate project scope and cost data.

REFERENCES

- [1] Recommended Practice 10S-90, “Cost Engineering Terminology”, AACE International
- [2] Recommended Practice 17R-97, “Cost Estimate Classification”, AACE International
- [3] Recommended Practice 18R-97, “Cost Estimate Classification – As Applied in Engineering, Procurement, and Construction for the Process Industries”, AACE International
- [4] Parametric Cost Estimating Handbook, NASA
- [5] Black, Dr. J. H., “Application of Parametric Estimating to Cost Engineering”, 1984 AACE Transactions, AACE International, 1984
- [6] Dysert, L. R., “Developing a Parametric Model for Estimating Process Control Costs”, 1999 AACE Transactions, AACE International, 1999
- [7] Rose, A., “An Organized Approach to Parametric Estimating”, Transactions of the Seventh International Cost Engineering Congress, 1982

Mr. Larry R. Dysert, CCC
 Conquest Consulting Group
 ldysert@ccg-estimating.com