# COLLECTION AND NORMALIZATION OF PARAMETRIC DATA

**Tomiţă Vasile**
*Universitatea din Craiova Facultatea de Economie şi Administrarea Afacerilor Drobeta Turnu Severin, Str. R. Calomfirescu, nr. 9, Bl.C5, Sc.1, Ap.11 e-mail: vasiletomita@yahoo.com tel. 0747-323076*
**Stuparu Dragoş**
*Universitatea din Craiova Facultatea de Economie şi Administrarea Afacerilor Drobeta Turnu Severin, Str. Topolniţii, nr. 11, Bl.2, Sc.2, Ap.5 e-mail: d_stuparu@yahoo.com tel. 0745-835764*
**Dăniasă Cora-Ionela**
*Universitatea din Craiova Facultatea de Economie şi Administrarea Afacerilor Drobeta Turnu Severin, Str. Adrian, nr. 42 tel. 0741-043125*

*Historical cost data are required as a basis for cost estimating, and parametric estimating is no exception. There are two types of data: Primary data which is obtained from the original source and the Secondary data which it is not obtained directly from the source, that it was derived from the original data. Once data has been collected there are normalized and calibrated. Normalization is the process to adjust the cost data for effects like inflation, anomalies, seasonal patterns, technology changes, accounting system changes, reorganizations, etc. Calibration is, in terms of a cost model, a technique used to allow application of a general model to a specific set of data. This is accomplished by calculating adjustment factor(s) to compensate for differences between the referenced historical costs and the costs predicted by the cost model using default values.*

*Key words: parametric data, data normalization process, calibration*

*JEL classification: C: Mathematical and Quantitative Methods; C5: Econometric Modeling; C51: Model Construction and Estimation*

## 1. INTRODUCTION

A universal format for collecting technical and cost information is the Work Breakdown Structure (WBS). The WBS provides for uniform definitions and collection of cost and technical information.
Historical cost and labor hours data are required as a basis for cost estimating, and parametric estimating is no exception. Data should be collected and maintained in a manner that provides a complete audit trail, and expenditure dates should be recorded so that dollar valued costs can be adjusted for inflation.
Also required is Technical Non-Cost Data that describes the physical, performance and engineering characteristics of a system, sub-system or individual item. For instance, weight is a common non-cost variable used in CER's (Cost Estimating Relationships) and parametric estimating models. (Other typical examples of cost driver variables include horsepower, materials of construction, watts, thrust, length, etc.)
A fundamental requirement for the inclusion of a non-cost variable in a CER is that it be a statistically significant predictor of cost (that is, it should be a cost driver).
Relevant program data including development and production schedules, quantities produced, production rates, equivalent units, breaks in production, significant design changes, and anomalies such as strikes, explosions, and other natural disasters are also necessary to fully explain any perturbations in historical data. Such perturbations may exhibit themselves in a profile of monthly cost accounting data as the labor hour charging may show an unusual "spike" or "depression" in the level of charged hours. Such historical information comes from knowledgeable program personnel or program records (also known as program "memory").
The collecting point for cost and labor hours data is, in most instances, called the general ledger or a company accounting system. All cost and labor hours data, used in parametric CER's or cost

models, must be consistent with, and traceable back to, the original collecting point (the source). The data should also be consistent with accounting procedures and cost accounting standards.

Technical non-cost data comes from engineering drawings, engineering specifications, certification documents, or direct experience (i.e., weighing an item). Schedule, quantity and equivalent units, and similar information comes from Industrial Engineering, Operations Departments, program files or other program intelligence.

Inflation indices normally combine external and internal information. Examples of external information used in these determinations include the Consumer Price Index (CPI), Producer Price Index (PPI), Commodity Price Indices and other forecasts of inflation from various econometric models.

There are other external sources of data including databases containing pooled and normalized information from various places (other companies or public record information). Although such information can often be useful, weaknesses of these sources include:

(1) The inability of the user to have knowledge of the procedures (i.e., accounting) used by the other contributors.

(2) The treatment of anomalies (how they were handled) in the original data.

(3) Knowledge of the manufacturing processes used and how they compare to the current scenario being estimated.

(4) The inability to accurately forecast future indices.

Internal contractor information includes analyses such as private corporate inflation studies, or "market basket" analyses. Such interval information provides data specific to a company's product line(s) (i.e., radar products) that could be relevant to a generic segment of the economy as a whole (i.e., electronics); etc. Such specific analyses would normally be prepared as part of an exercise to benchmark government provided indices (the CPI), and to compare corporate performance to broader standards.

It is important to realize that sources of data can be almost unlimited, and all relevant information should be considered in a parametric analysis, if practical. Although major sources are described above, data sources should not be constrained to a specific list.

Any data included in calculating parametric parameters will vary between model developers. However, the way in which parametric models are calculated from historical data and the way they are applied in the estimating process should be consistent within individual estimating systems.

## 2. SIGNIFICANT ADJUSTMENTS TO PARAMETRIC DATA

What follows below are some of the more significant adjustments that may have to be made to historical parametric cost data.

### Consistent Scope

Adjustments are appropriate for differences in program or product scope between the historical data and the estimate being made. For example, if the systems engineering department made a comparison of five similar programs and then realized that only two of the five had design to cost (DTC) requirements. To normalize the data, the DTC hours were deleted from the two programs to create a consistent systems scope and definition for CER development.

### Anomalies

Historical cost data should be adjusted for anomalies (unusual events), prior to CER analysis, when it is not reasonable to expect these unusual costs to be present in the new projects. The adjustments and judgments used in preparing the historical data for analysis should be fully documented. For example, a comparison has been made to compare the development test program from five similar programs and then certain observations are made (from history and interviews) that one of the programs experienced a major test failure (e.g., qualification, ground

test, flight test). A considerable amount of labor resources were required to fact find and then determine the root cause of and develop an action plan for a solution. Should the hours be left in or deleted?

**Improved Technology**

Cost changes, due to changes in technology, are a matter of judgment and analysis. All bases for such adjustments should be documented and disclosed. For example, electronic circuitry was originally designed with discreet components, but now the electronics are ASIC technology. A hardware enclosure once was made from aluminum and now is made, for weight constraints, of magnesium. What is the impact on the hours? Perfect historical data may not exist, but judgment and analysis should supply reasonable results.

A careful analysis should be performed on the data to determine why it behaved the way it did. There may have been a strike, or possibly an unusual and serious engineering problem impacted production costs. In any event, careful analysis is important.

**Inflation**

There are no fixed ways to establish universal inflation indices (past, present or future) that fit all possible situations. Inflation indices are influenced by internal considerations as well as external inflation rates. Therefore, while generalized inflation indices may be used, it may also be possible to tailor and negotiate indices used on an individual basis to specific labor rate agreements and the actual materials used on the project. Inflation indices should be based on the cost of materials and labor on a unit basis (piece, pounds, hour) and should not include other considerations like changes in manpower loading or the amount of materials used per unit of production. The key to inflation adjustments is consistency. If cost is adjusted to a fixed reference date for calibration purposes, the same type of inflation index must be used in escalating the cost forward or backwards, from the reference date, and then to the date of the estimate.

## 3. DATA NORMALIZATION PROCESS

Specifying an estimating methodology is an important early step in the estimating process. The basic estimating methodologies (analogy, catalog prices, extrapolation, factors/ratios, grassroots and parametric) are all data-driven. To use any of these methodologies, credible and timely data inputs are required. If data required for a specific approach is not available, then that methodology cannot be used.

Given that all methodologies are data-driven, it is critical that the estimator know the best data sources. Here are nine basic sources of data and a description of what specific data can be obtained from each source. Definitions of the differences between primary and secondary sources of data are provided. Finally, there is a review of the type of information that should be available from an accounting system, and a description of how to collect and analyze data is also given.

The information presented will help the collection and analysis of the two data types (primary and secondary) required to specify, and apply a parametric estimating methodology. Remember - any data needs to be available, reliable and convincing before an estimating methodology can be chosen that utilizes the foundation data. The two types of data are:

1. Primary data is obtained from the original source. Primary data is considered the best in quality, and ultimately the most useful.
2. Secondary data is derived (possibly "sanitized") from primary data. It is not obtained directly from the source. Since it was derived (actually changed) from the original data, it may be of lower overall quality and usefulness.

When preparing a cost estimate, look for all credible data sources. If at all possible, use primary sources of data.

There are nine main sources of data and they are listed in the chart below:

| | SOURCES OF DATA | TYPE |
|---|---|---|
| 1. | Basic Accounting Records | Primary |
| 2. | Cost Reports | Either (Primary or Secondary) |
| 3. | Historical Databases | Either |
| 4. | Functional Specialist | Either |
| 5. | Other Organizations | Either |
| 6. | Technical Databases | Either |
| 7. | Other Information Systems | Either |
| 8. | Contracts | Secondary |
| 9. | Cost Proposals | Secondary |

Collecting the data to produce an estimate, and evaluating the data for reasonableness, is a very critical and time-consuming step of the estimating process.

When collecting the data needed to integrate cost, schedule, and technical information for an estimate, it is important to obtain cost information, and also the technical and schedule information. The technical and schedule characteristics of programs are important because they drive cost. They provide the basis for the final cost.

For example, assume the cost of another program is available and a program engineer has been asked to relate the cost of the program to that of some other program. If the engineer is not provided with specific technical and schedule information that defines the similar program, the engineer is not going to be able to accurately compare the programs, nor is he or she going to be able to respond to questions a cost estimator may have regarding the product being estimated vis-a-vis the historical data.

The bottom line is that the cost analysts and estimators are not solely concerned with cost data. They need to have technical and schedule data available in order to adjust, interpret, and lend credence to the cost data being used for estimating purposes.

A cost estimator has to know the standard sources where historical cost data exists. This knowledge comes from experience and from those people, the so-called local experts, that are available to answer key questions.

A cost analyst or estimator should be constantly searching out new sources of data. A new source might keep cost and technical data on some item of importance to the current estimate. Do not hesitate to ask anyone who might know or be able to help, since it is critical to have relevant cost, schedule and technical information at all times.

The chart below summarizes important points about data collection and evaluation.

| DATA COLLECTION, EVALUATION AND NORMALIZATION |
|---|
| - Very Critical, Time Consuming Step<br>- Need Actual Historical Cost, Schedule, and Technical Information<br>- Know Standard Sources<br>- Search Out New Sources<br>- Capture Historical Data |

In order to develop a parametric model, a necessary requirement is to possess historical cost, schedule and technical data on a set of data points. The idea here is that generally more data is

better than less. It is necessary to know what trends exist, and to understand why the trends are as they are. Some models have been found to be based on the opinions of experts instead of historical data. Although the opinions of experts may be germane, sound historical data is preferable for model development, audit and analysis.
In addition to the historical data points, information on the cost, technical and quantity drivers needs to be examined to determine which does the best job of predicting cost. A statistical analysis on the data is accomplished to determine the strongest predictor(s) or driver(s) of cost, that is, the independent variable(s).
It is very important to note that when performing a statistical analysis, be sure that functional specialists can provide realistic and reliable parameters for independent variables, given the stage of the program being estimated. Illustrating this point, suppose a statistical relationship is developed that has very strong correlation, and a potential cost driver has been discovered. However, data for the same independent variable for the estimate is not available. The parametric model would not then help with the estimate.
Finally, knowledge of basic statistics, modeling skills and an understanding of analytical techniques is necessary to develop parametric estimating relationships.
The above information is summarized on the chart below:

| **TYPE OF INFORMATION NEEDED TO DEVELOP A PARAMETRIC MODEL** |
|---|
| - Reliable Historical Cost, Schedule, and Technical Data on a Set of Data Points |
| - Work Breakdown Structure (WBS), WBS Dictionary & Product Tree |
| - Analysis to Determine Significant Cost Drivers |
| - Knowledge of Basic Statistics, Modeling Skills and CER Development |
| - Analysis Techniques |

To use a parametric model, the model needs to be well-documented. The documentation of a parametric model should include the source of data used to derive the parameters, and the size and range of the database. Additional information that should be included in the documentation of a parametric model are: how the parameters were derived, what the model's limitations are, the time frame of the database and, how well the parametric tool/model estimates its own database..
All of this information should be located in the source document of a parametric top functional experts knowledgeable about the program you are estimating to identify most-likely range for cost drivers to confirm applicability of parametric from technical perspective.
A parametric estimating methodology can be used at any stage of a program's life cycle. For example, a general parametric model may be utilized in the early, conceptual phase of a program, although the same parametric model could be inappropriate to use in the follow-on production phase of a program. However, a detailed parametric model used in production estimating that is based on the experience and actual historical data of two or three previous production lots, could yield excellent validity.
Hence, a parametric methodology can be used at any stage of a program's life cycle as long as the parametric model is based on the level and type of information available at that stage.
The methodology can be used for any WBS element, not just hardware and software. Parametrics can be successfully applied to Systems Engineering/Program Management, Test, Training and Data, etc., provided that historical data points are available to develop solid, statistical relationships that provide reliable estimates of independent variables.

## 4. CALIBRATION AND VALIDATION OF COST MODELS

Once data has been collected and normalized, cost models can be developed. There are two general types of cost models: internal (contractor developed) and commercially available. Internal, contractor developed models are derived from unique contractor data and generally do

not require calibration since they have been calibrated in a defacto manner. On the other hand, commercial models are based on more universal data, and almost always need some form of calibration to be useful.
The cost driver equation(s) utilized in a commercial cost model are based on a database external to the specific data being used to support the current estimate. Calibration, then, is the process of computing a multiplied(s), to be used with the general purpose equation(s), such that the combined equation(s) will predict the cost as reflected by the technical and programmatic data being used to support the estimate.
Specialized (Internal) cost models are based directly on the data being used to support the estimate. Since the CER's are derived directly from the supporting data, the model is, by definition, calibrated.
The result of calibrating an item, in a commercial model, is a calibration factor which is used in the commercial model's equations, such that the equations are then made to calculate the value of the item.
Cost models need to be calibrated and validated for acceptance. The validation of a cost model is a process which usually includes the following steps:
(1) Calibrate the model to historical cost data.
(2) Estimate the cost of past completed projects.
(3) Compare the estimates with actual costs to demonstrate acceptable accuracy.
It is the combined use of the model with the estimating process that must achieve acceptable results to provide a basis for the validation of the model. It may also involve disclosure of how the model works so that the effects of scaling and heuristic analysis can be evaluated by management, customers or auditors.
Validation implies that interested parties have agreed that the model is a valid and acceptable estimating tool, and predicts costs within a reasonable range of accuracy.

## 5. PITFALLS TO THE USE OF A PARAMETRIC MODEL

When a parametric model is applied to values outside its database range, the credibility of the resulting estimate becomes questionable. In cost estimating, one rarely finds large, directly applicable databases, and the source document has to be evaluated to determine if the parametric can be applied to the current estimate. However, it is possible to develop parametric tools that relate cost based on generic complexity values or tables. Such generalized parameters, can be related to the task at hand by an experienced modeler that results in a good cost model, but a parametric model always needs to make sense for the present estimate.
Additionally and before using, one should validate models based on expert opinion. This is accomplished first by obtaining some actual, historical data points (technical, schedule, and cost) on completed programs similar to the current program. With this data in hand, apply the model to the actual technical and schedule information and see how well the parametric model.

## REFERENCES

1. Stuparu, D., Vasile, T. – „*Matematici aplicate în economie*”, Editura Şcoala Mehedinţiului, Drobeta Turnu Severin, 2002,
2. Vasile, T. – “*Metode statistice în managementul afacerilor*”, Editura Sitech, Craiova, 2008,
3. Vasilescu, N., Costescu, M., Ionaşcu, C., Babucea, G., Vasile, T., Stuparu, D. – „*Statistică*”, Editura Universitaria, Craiova, 2003,
4. Wonnacott, T., H., Wonnacott, R., J. – „*Statistique*”, Editure Economica, Paris, 1991.