A Note on the Calibration of COCOMO II to Project Environments With Limited Historical Data

Jeff Singer
Software Strategies

Background

This paper describes the author’s experience calibrating COCOMO in a number of commercial product development groups. The term “commercial” here means products intended for sale in a general market and competing with other similar products for market share. This is distinguished from commercial software developed under contract, often for a single customer.

The central challenge in these groups has been the achievement of accurate project estimates when very little reliable historical data is available. Typically, these organizations have no systems for collecting labor accounting data and do not keep records of formal project milestones. As a result, these types of data must be gathered from interviews with project personnel and manually cross-checked for consistency. Data on code size is generally more accurate, since almost all development groups now use some type of automated source code control system.

The calibration problem is made more difficult because there are always so few historical projects available to serve as a calibration base. This is partly because many of these organizations are relatively new with little prior history and may have only one or two projects they have completed. A second reason is, that in groups that have longer histories (often more than ten years), the “horizon of recollection” of people in the group virtually never extends beyond one previous project. Since the effort and schedule data depends almost completely on these people’s memory, the number of projects that can be reconstructed is usually one to three even in quite large groups. The major consequence of this small number of available projects is that even one project which is inaccurately characterized by the data gathered for it, or which is in some way not representative, seriously distorts the calibration.

The most significant errors we have experienced in doing calibrations are in the estimation of COCOMO Cost Driver and Scale Factor parameters for historical projects. These errors seem to result from three primary factors that we have observed.
Key people have left the company or organization and are unavailable for data collection. This is often the project leader who usually has the most complete set of information.

People remember selectively and with bias. This is impossible to quantify, however cross-checking recollections among different project personnel on common data has revealed a high incidence of inconsistency. This raises the question of the accuracy of unchecked data.

People are asked to quantify project characteristics that require the application of a "calibrated opinion" when they may not have any other reference point. One person's "many" may be another's "some". With so few project personnel available, there is often a lack of a consensus based opinion.

The net result of all of these effects is that the calibration process in these environments is difficult at best and has required us to develop somewhat unorthodox techniques to achieve acceptable results.

Methodology

The uncertainty present in the project based data we have been required to work with has led us to look for additional methods and tools outside the COCOMO environment that can be used to detect and in some cases help to resolve this uncertainty. In this regard, we have developed straightforward techniques for applying two parameters developed by Putnam [1] for characterizing project environments. These are: MBP (Manpower Buildup Parameter) and PP (Productivity Parameter). Conceptually, each of these parameters is defined as follows:

The Manpower Buildup Parameter calculates a measure of how quickly effort is expended in a project and is based on the finding by Putnam that the staffing rate of software development projects is approximated by a Rayleigh curve[1]. The value of this parameter is given by the equation:

\[ MBP = 875.2 \times \frac{Effort_{(x-m)}}{E.F.D. \times (Sched_{(med)})^3} \]  

Eqn. 1

Effort and Schedule are the total effort and total development schedule of the project and E.F.D. is an adjustment for the effort expended on the functional design phase of the
project. Its value is dependent on the code size of the product and varies from 1.6 for projects smaller than 15 KLOC to 1.2 for projects at or above 100 KLOC.

The PP parameter is a measure of aggregate project productivity and is based on Putnam’s original "Software Equation" developed empirically from a set of historical project data. Its value is given by the equation:

\[ PP = 92,304 \times \frac{KLOC}{\left( \frac{\text{Effort}_{(z-m)}}{E.F.D \times B} \right)^{d} \times (\text{Sched}_{(mos.)})^{d}} \]  

Eqn. 2

Where B is a “Special Skills Factor” which ranges in value from 0.16 at a program size of 5 KLOC to 0.39 for program sizes at or above 70 KLOC. KLOC is the size (or equivalent size) of the executable code in thousands of lines of code in the project and effort and schedule are the same as in the MBP expression.

The practical value of these parameters stems from their use in the SLIM estimation tool in the software development marketplace on a large set of projects (more than 6000) of widely varying types. This wide use has demonstrated characteristics of these parameters that can be confidently applied in other project environments. Specifically:

- The MBP parameter is constant or nearly so, in any given development environment, where environment means roughly a single engineering group with common processes and practices. While data supporting this assertion has not been published by Putnam (to the author’s knowledge), our experience in a number of different development groups has consistently demonstrated its validity.

- The PP parameter has been related to software of different types (e.g. system software, telecommunications, firmware, etc.) over the same large set of projects. This data has been published by Putnam [1], [2] in summary form and is frequently useful as a rough indicator of the validity of historical project data. A convenient feature of this parameter is that it is calculated directly from historical values of size, effort and schedule and does not require additional parameter calibration.

The techniques developed by the author for applying the Putnam parameters have evolved pragmatically from a number of client situations with limited amounts of data of uncertain accuracy. These techniques are described below:
In environments where there are a small number of historical projects available for calibration, the values of MBP can be used to detect projects that are probably unrepresentative of the development group’s typical performance and that should be eliminated from the calibration process. For example, in all of our client data, the value of MBP does not fall below 7. Thus, when we encounter a historical project with an MBP value of 2 or 3, we are immediately suspicious of its validity. Typically, in these cases, on closer inspection, the project turns out to have had a major flaw such as a design that had to be re-done part way through coding.

In planning new projects with a COCOMO calibration established from local historical data, the value of MBP for the new project can be calculated from the COCOMO estimates for effort and schedule and should be at, or very close to the established historical value for the group. We have seen a number of client projects planned at an MBP value considerably higher than the established value and have watched as the projects were replanned during their execution until the historically established MBP value was reached. In one case, the project was begun at an MBP value of 45.5 when the established historical value was roughly 10. The project was replanned three times over the next eight months (because planned staffing rates could not be achieved) at successive MBP values of 26.0, 14.7 and 8.0 before finally being abandoned because of lateness.

The PP parameter has two primary uses in helping to validate COCOMO calibrations. First, because there are average values for different types of software that have been published for the parameter [1], it can often be used to detect non-representative projects. For example, if a historical systems software project has a PP value of 1,400, which is typical of firmware projects, and the PP value for all other projects in the calibration was in the range of 8,000 to 12,000, one would immediately suspect that the project with the low value was flawed in some serious way. Generally, a closer inspection of the project’s history will reveal the causes.

The second use for the PP parameter is in establishing a reference point for COCOMO calibrations which act as an aid in detecting historical Effort Adjustment Factors that are significantly in error. This has often proven to be a critical step in achieving an accurate calibration with only one or two historical projects. This reference
point is generated by calculating the value of the PP parameter for the standard "out of the box" COCOMO calibration at the appropriate historical project size and with the Effort Adjustment Factor and Scale Factor Sum set to their nominal values. For example, in the COCOMO II 98 calibration where the multiplicative constant values are $C_1=2.94$ and $C_2=3.67$ and the nominal exponents are 1.10 for effort and 0.318 for schedule, for a project of 100 KLOC, the corresponding value of the PP parameter is 12,069. This parameter can then be used as a reference point for generating effort and schedule estimates which in turn, are used to confirm historically generated values of $C_1$ and $C_2$.

Case Study Examples

The application of the techniques described above are best demonstrated by presenting case studies of actual client situations from the author’s consulting practice. In the two examples below, the calibrations of the COCOMO model both involved very small numbers of historical data points, some of which were not representative. In addition, the project data was developed largely from the personal recollection of project members and was significantly in error in some cases. The techniques used illustrate ways in which these errors can be detected and successfully dealt with.

The Penknife Project

This project was the sole historical data point of a two year old group that was developing driver software for high performance custom graphics chips targeted at the mass PC market. The group was about to begin a successor project "Scimitar", which was needed at a specific point in time in order to meet the "Christmas buy" window of the PC market. The central goal of the calibration therefore, was to estimate the Scimitar project accurately enough to determine whether it could be brought to market in the needed time frame.

The Penknife project had produced 62.4 KLOC of executable code over 23.0 months, expending 240.0 staff-months of effort. At the time the estimation effort was begun, the project leader and all of the senior development engineers who had been assigned to the project were still in the group. The development of the COCOMO historical E.A.F. and S.F. data was done by conducting a number of (approximately eight) structured interviews with these project personnel. The data from these interviews established the Penknife E.A.F. at 1.16 and the sum of the Scale Factors ($\sum S.F.$) at 14.7. This yielded effort and schedule exponents of 1.057 and 0.309. (Note that this work was originally done using the COCOMO II v 1.0 calibration and converted here to COCOMO II 98 for
The values for the Putnam parameters were computed and were: MBP=14.9 and PP=10,688. This data is summarized below:

<table>
<thead>
<tr>
<th>Size (KLOC)</th>
<th>62.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effort (s-m)</td>
<td>240.0</td>
</tr>
<tr>
<td>Sched. (mos.)</td>
<td>23.0</td>
</tr>
<tr>
<td>E.A.F.</td>
<td>1.16</td>
</tr>
<tr>
<td>Σ S.F.</td>
<td>14.7</td>
</tr>
<tr>
<td>MBP</td>
<td>14.9</td>
</tr>
<tr>
<td>PP</td>
<td>10,688</td>
</tr>
</tbody>
</table>

Table 1 Penknife Project

The values of PP and MBP as well as E.A.F. and Σ S.F. all appear to be completely reasonable. The PP value of 10,688 is quite typical of system software projects and the MBP value of 14.9 is characterized as “slow” in Putnam’s data and is compatible with a project who’s staffing rate is dependent in part on the rate at which the associated hardware design is completed. The values for the multiplicative constants calculated from the historical data were: C₁=2.58 and C₂=4.23. As a double check on the E.A.F. value (which is close to nominal), the reference value of PP was calculated and the constants calculated from it. This was done as follows:

1.) The COCOMO effort and schedule equations were used to generate effort and schedule values with E.A.F. and Σ S.F. set to nominal (1.0 and 18.97 respectively). Size was set to 62.4 KLOC. This yielded 277.36 staff-months of effort and a schedule of 21.96 months.

2.) The effort and schedule values from the step above were entered in the Putnam equations to generate corresponding MBP and PP values. These were MBP=19.8 and PP=10,833. This establishes 10,833 as the reference PP value for COCOMO at this size (62.4 KLOC).

3.) An assumption is made that as a general rule, the value of PP does not change significantly as MBP is changed. This is a largely unverified assumption as the author has little real world project data to test it against. A check of this was made on the data published by Kemerer [3] for 15 projects. These projects were all done at much higher values of PP than the author’s client data, however it showed approximately a 22% rise in PP over the full range of MBP values. Given that the typical change in MBP that is involved in the author’s cross-checking is one-quarter or less of the full MBP range, the assumption seems reasonable when confined to cross-checking calibrations.
4.) Using the reference value for the PP parameter established in Step 2 above and the historical value of MBP (14.9) an estimate of effort and schedule is made at the project size. This is done by solving equations 1 and 2 above for effort and schedule as a function of MBP and PP. These equations have the following form:

\[ \text{Effort}_{t} = 50,420 \times \left( B^2 \times E.F.D. \times \frac{MBP^4}{PP^2} \right) \times KLOC^2 \]  
Eqn. 3

\[ \text{Sched.}_{\text{months}} = \left( 353.4 \times \left( \frac{B}{MBP} \right)^4 \times \left( \frac{1}{PP} \right)^2 \right) \times (KLOC)^2 \]  
Eqn. 4

These values for effort and schedule are the values that COCOMO is predicted to generate for the historical MBP value at nominal settings of E.A.F. and S.F. In the case of the Penknife project, these values (at MBP=14.9) are: Effort=236.8 staff-months and Sched.=22.9 months.

5.) The effort and schedule estimates from Step 4 can now be used to calculate \( C_1 \) and \( C_2 \) using the COCOMO master equations. These values are: \( C_1=2.50 \) and \( C_2=4.02 \). These compare very favorably with the values calculated directly from the historical project data. (\( C_1=2.58 \) and \( C_2=4.23 \)). In this case, the E.A.F. and S.F. values from the historical data collection project were used with the increased confidence provided by the cross-checking process.

As a final note on the outcome of the Scimitar project, the COCOMO based estimate showed that the project could not meet the required ship date and the group was required to cancel the project and replan its entire product line. While this initially created a good deal of displeasure with the engineers in the group, with the passing of time it is now seen as a very sound decision as they have moved on to more profitable ventures.

**The SixFou Project**

This project was being started in a client group that develops the software components for a line of high-end, mission critical hardware/software systems. At the time the SixFour project was started, the group had been developing products for 15 years and
had a large established code base which they often modified as successive generations of new hardware processors and architectures became available in the general market. Despite this long experience, they had little data collected from previous projects and the COCOMO calibration process involved numerous detailed interviews with development personnel as well as the collection of fragmented written data on code size, effort and project milestones. This collection was made more difficult by the high personnel turnover rate which meant that the "recollection horizon" in some areas covered less than one full project. There were three historical projects examined as part of the COCOMO calibration process: NetStack, a protocol stack project, WinOne, a large new operating system development and UpRev, a major revision of an existing product to support a new processor chip. The basic size, effort and schedule data for these projects is summarized in Table 2 below.

<table>
<thead>
<tr>
<th>Project</th>
<th>Size (KLOC)</th>
<th>Effort (s-m)</th>
<th>Sched. (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetStack</td>
<td>30.0</td>
<td>258.0</td>
<td>42.0</td>
</tr>
<tr>
<td>WinOne</td>
<td>225.0</td>
<td>500.0</td>
<td>33.5</td>
</tr>
<tr>
<td>UpRev</td>
<td>74.0</td>
<td>450.8</td>
<td>33.5</td>
</tr>
</tbody>
</table>

Table 2
SixFour Project Calibration Base

As a first step in the calibration process, the Putnam parameters were calculated for each project. They were:

<table>
<thead>
<tr>
<th>Project</th>
<th>MBP</th>
<th>PP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetStack</td>
<td>2.4</td>
<td>2,116</td>
</tr>
<tr>
<td>WinOne</td>
<td>9.6</td>
<td>18,849</td>
</tr>
<tr>
<td>UpRev</td>
<td>8.7</td>
<td>6,404</td>
</tr>
</tbody>
</table>

Table 3
Putnam Parameters for the SixFour Calibration Projects

The most noticeable feature of these parameters are the extremely low values of MBP and PP for the NetStack project. These values are found in the Putnam data typically in projects such as microcode where design times are often very long and the total amount of code being produced is frequently very small. The E.A.F. and ∑S.F. values however, were close to nominal at 1.11 and 19.35 respectively. The low values of the NetStack
Putnam parameters led to a more detailed analysis of the project which in turn revealed a number of serious problems including major design errors and unanticipated difficulties in acquiring test resources. After analyzing this additional data, it was concluded that the NetStack project was not representative of the environment and it was eliminated from the calibration. The remaining two projects showed the expected agreement in their MBP values and their PP values were consistent with the type of development being done in each project. The WinOne project had been done largely as a new code set with a highly capable group of engineers while the UpRev project was largely a maintenance effort done by a group that had experienced the departure of a high percentage of their senior personnel.

The values of the E.A.F. and $\sum S.F.$ parameters for these two projects are shown in Table 4 below. These were gathered in a number of interviews with project personnel over a period of roughly three months.

<table>
<thead>
<tr>
<th>Project</th>
<th>E.A.F.</th>
<th>$\sum S.F.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WinOne</td>
<td>2.21</td>
<td>25.8</td>
</tr>
<tr>
<td>UpRev</td>
<td>1.16</td>
<td>22.3</td>
</tr>
</tbody>
</table>

Table 4

A comparison of the values in Table 3 with those in Table 4 immediately reveals an inconsistency. Given the nearly equal value of MBP for the two projects and the fact that the productivity parameter for WinOne is much higher than for UpRev, the COCOMO parameter values from the data collection process would be expected to be smaller for the WinOne project. This is not the case however, since both the E.A.F and $\sum S.F.$ values for WinOne are higher than those of the UpRev project. The conclusion that must be drawn is that the COCOMO E.A.F. and S.F. parameters for either one or both of the projects are significantly in error. When the values for the multiplicative constants $C_1$ and $C_2$ for each project are calculated, the large difference in the values of $C_1$ help illustrate the problem. These are shown in Table 5 below.

<table>
<thead>
<tr>
<th>Project</th>
<th>$C_1$</th>
<th>$C_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WinOne</td>
<td>0.40</td>
<td>4.26</td>
</tr>
<tr>
<td>UpRev</td>
<td>2.99</td>
<td>4.61</td>
</tr>
</tbody>
</table>

Table 5
The dilemma then, is deciding how to select the best parameter values for COCOMO use in the group's new project estimation. (It should be noted that with only two projects available in the calibration, the base exponents, 0.91 and 0.28 in COCOMO II '98, cannot sensibly be recalculated and the standard calibration values are used.)

Following the example outlined in the Penknife project above, the first step in generating a "cross-check" set of COCOMO constants was to calculate a reference point from the Putnam parameters. To do this, the average size of the two projects was used in the standard COCOMO "out of the box" equations to generate effort and schedule values. Using the average size of 149.5 KLOC, and setting E.A.F. and $\sum S.F.$ to nominal values, COCOMO generated estimates of 725.2 staff-months and 29.8 months for effort and schedule respectively. Using equations 1 and 2 above, the corresponding Putnam parameters can be calculated. These are: MBP=20.0 and PP=12,901. Note that this MBP value is different enough from the value of the calibration projects (average=9.2) so that a new reference point is needed. Using equations 3 and 4 above and the average MBP value of the two projects, the effort and schedule corresponding to MBP=9.2, PP=12,901 (the reference point value) and size=149.5 were calculated. These were: 464.8 staff-months and 33.3 months. Using these two values in the COCOMO master equations to solve for $C_1$ and $C_2$ yields: $C_1$=1.88 and $C_2$=4.75. Note that this value of $C_1$ is roughly midway between the values calculated for this parameter in each historical project, suggesting that the value in the WinOne project is too low and conversely that it is too high in the UpRev project.

Given the large discrepancy in the values of the $C_1$ constants in the calibration projects, it was decided to briefly revisit the E.A.F. calibration process for each project. This proved to be unproductive as no additional data surfaced in talking with project personnel. It was decided therefore to proceed with the new SixFour project using the values of $C_1$ and $C_2$ generated in the cross-checking process. As estimates of the project size, effort and schedule were developed by the project leaders using their traditional techniques of "educated intuitive guessing", the COCOMO calibration proved to be quite consistent with these. The project estimates at the point where a corporate schedule commitment was made, were: 213 KLOC, 775 staff-months and 39 months. These values correspond to COCOMO parameters of 1.13 for E.A.F. and 19.0 for $\sum S.F.$ using the cross-check generated constants. The values for the E.A.F. and $\sum S.F.$ parameters that were generated from interviewing project personnel using the COCOMO II '97 questionnaire were: E.A.F. ranging from 1.01 to 1.06 and $\sum S.F.$ between 20.6 and 22.2. These values generate effort estimates between 753 and 861 staff-months and schedule values between 39.9 and 42.5 months. The three estimates are summarized in Table 6 below:
The project decided to proceed with the estimate generated by the project leaders as it was in close agreement with the “COCOMO Low” estimate. This optimistic outlook seems to be typical of new projects in the author’s consulting experience and often needs revision as the project runs into unplanned difficulties.

**Summary**

The problem of developing an accurate COCOMO calibration in environments with a small number of historical projects is often an exercise in guesswork. In the author’s experience, the level of error in the COCOMO parameters generated from questionnaires can be substantial. By applying the two Putnam parameters to the data, a cross-check can be made on the COCOMO parameter values. In some cases, data which is in error can be spotted by relating the Putnam parameter values to published ranges of these parameters. In others, a reference point can be established which can both help gauge the level of error present in the questionnaire generated COCOMO parameters as well as establish approximate values for COCOMO parameters.

<table>
<thead>
<tr>
<th>Estimate</th>
<th>E.A.F.</th>
<th>$\Sigma$ S.F.</th>
<th>Effort (s-m)</th>
<th>Sched. (mos.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Ldrs.</td>
<td>1.13</td>
<td>19.0</td>
<td>775</td>
<td>39</td>
</tr>
<tr>
<td>COCOMO Low</td>
<td>1.01</td>
<td>20.6</td>
<td>753</td>
<td>39.9</td>
</tr>
<tr>
<td>COCOMO High</td>
<td>1.06</td>
<td>22.2</td>
<td>861</td>
<td>42.5</td>
</tr>
</tbody>
</table>

Table 6
SixFour Project Estimates
References


