Volume 10 | Issue 1 | October 2021

ISSN: 2160-4746

Journal of Cost Analysis and Parametrics

ICEAA
International Cost Estimating & Analysis Association

www.iceaaonline.com

ICEAA

Journal of Cost Analysis and Parametrics

Editor's Note
David L. Peeler, Jr., CCEA 2

How Green Was My Labor: The Cost Impacts of Manufacturing Personnel Changes
Brent M. Johnstone 4

Does Cost Team Leadership Matter?
Christina N. Snyder, CCEA 18

Parametric Joint Confidence Level Analysis: A Practical Cost and Schedule Risk Management Approach 29
Sara Jardine
Christian B. Smart, Ph.D., CCEA
Kimberly Roye, CCEA

Reasons a Cost Estimate Could Go Wrong During a Concurrent Engineering Study (and How to Avoid Them) 43
Andy Braukhane

In Search of the Production Steady State: Mission Impossible? 60
Patrick McCarthy, CCEA

Using Dummy Variables in CER Development 75
Dr. **Shu-Ping Hu**
Alfred Smith, CCEA

Improvements on the Development of Correlated Input Variables for Monte Carlo Simulation 90
Douglas Henke

Volume 10 | Issue 1 | October 2021 ISSN: 2160-4746

Editor's Note

David L. Peeler, Jr., CCEA

Welcome to the reboot/restart/rebirth/ resurrection of the Journal *of Cost Analysis and Parametrics* (JCAP). After a five-year hiatus, we have pulled the publication in-house and ceased paper production as we reinstate our journal. The journal will be published on the ICEAA website, at least at first. We are also exploring options and avenues for traditional library access as well as social media promotion.

After much discussion over the past couple of years, the ICEAA Board of Directors, upon the recommendation of the JCAP editors, have decided to broaden the scope of our journal. As you'll see in this issue, we include qualitative as well as the previously preferred quantitatively rigorous articles. The new direction is calculated to make the journal of more interest to the totality of our membership. We plan to publish discerning work of significance that speaks to the membership and advances the intellectual pursuits within the cost arena *writ large*.

In this issue the reader will find the three best paper award winners from the 2021 ICEAA Workshop, as well as best-in-track award winning papers from 2020. Two additional articles are from prior submissions.

Brent Johnstone provides a pertinent and oft forgotten look at broken learning. In *How Green Was My Labor: The Cost Impacts of Manufacturing Personnel Changes,* he provides a model for estimating the cost of personnel changes on a program, adding instruction where existing learning curve literature falls short.

Christina Snyder addresses the question *Does Cost Leadership Matter?* in a novel way. She applies cost estimator survey responses regarding leadership impact on program/product success and then uses the ten behaviors identified by Google's Project Oxygen seeking an understand of the necessary skills for successful cost leadership.

Rounding out the three 2021 ICEAA Workshop best papers is *Parametric Joint Confidence Levels: A Practical Cost and Schedule Risk Management Approach* by **Sara Jardine**, **Kimberly Roye**, and Dr. **Christian B. Smart**. They present a parametric and machine learning approach, developing mathematical models for cost and schedule risk, with the application of machine learning models, developing higher-fidelity schedule models.

From the 2020 conference, we feature **Andy Braukhane**'s *13 Reasons a Cost Estimate could go wrong during a Concurrent Engineering Study (and How to Avoid Them)*, wherein he highlights and addresses the difficulties of concurrent engineering on the cost estimator.

Patrick McCarthy explores the culmination of learning curve knowledge *In Search of the Production Steady State: Mission Impossible?* He adds to learning curve application by seeking the prediction of a production steady state and, given the unique nature of defense acquisition, whether one really exists.

The final two articles are legacy submissions to the journal, yet they retain currency and advance

the understanding of simulation and cost estimating relationship development. *Using Dummy Variables in CER Development* advances our understanding in this area, as two Frank Freiman Lifetime Achievement Award recipients, Dr. **Shu-Ping Hu** and **Alfred Smith** clarify when the use of dummy variables is appropriate. The authors propose specific guidelines for appropriate use of dummy variables as well as common errors analysts experience in application.

In *Improvements on the Development of Correlated Input Variables for Monte Carlo Simulation*, **Douglas Henke** adopts methods of rank correlation and multivariate input for modeling by Monte Carlo simulation to provide insights into algorithm mechanics, yielding a richer understanding of the process and a more accurate reflection of desired correlations for both symmetric and highly skewed distributions.

This issue returns the Journal *of Cost Analysis and Parametrics* to the ICEAA membership. The expansion of content should serve us well and be a resource for the whole community. We think there is something in these pages for the wide array of practicing cost/risk/schedule estimators/analysts/leaders, etcetera. We hope you enjoy the articles and apply the gained knowledge to your professional efforts. Thank you to all who have encouraged and supported the return of our professional journal; and most importantly, thank you for reading.

David Peeler
JCAP Editor

How Green Was My Labor:
The Cost Impacts of Manufacturing Personnel Changes

Brent M. Johnstone

Abstract: Estimators are frequently confronted with manufacturing personnel increases or decreases and asked to calculate shop performance impacts. However, existing learning curve literature offers little guidance how to do so. This paper identifies issues associated with both new hires (so-called "green labor") and workforce reductions and offers an analytical format. Based on a study of a large workforce expansion on a mature aircraft program, a model to analyze future personnel changes is presented as well as example cases.

Introduction

In the ideal world of theoretical learning curves, product configurations, production rates and the quantity and skill level of employees remain constant throughout time, thus allowing us to plot smooth, continuous reductions in unit cost over time. In the real world, this is far from the case. All three variables – configuration, production rates, and employees – are in constant flux.

The impact of configuration changes on the learning curve and on manufacturing unit costs are well-understood. The impact of production rates on unit cost has also been extensively studied, albeit with contradictory opinions whether changes in production rates have significant or insignificant long-run impacts on manufacturing hours. (Johnstone, 2017.) However, the published learning curve literature is largely silent on how changes in the number or skill level of manufacturing employees affect cost. While this issue impacts all long-cycle manufacturing operations, typically the literature only addresses the subject in the context of production gaps and the subsequent loss of learning. (Anderlohr, 1969.) However, the impact of production gaps is so deleterious that planners and schedulers go out of their way to avoid them, and therefore they occur only infrequently. Far more common are increases or decreases in workforce levels due to fluctuations in production deliveries, and yet little is written down to guide the estimator.

This paper hopes to fill in the gaps. Estimators are frequently confronted with workforce increases or decreases and asked to calculate the impacts on shop performance. This paper identifies issues associated with both new hires and workforce reductions and offers an analytical format. Based on a study of a large workforce expansion on a mature aircraft program, a model to analyze future workforce changes is presented as well as an example case.

Overview of the Problem

Questions about workforce changes really ask: What is the impact of assigning new work to people? In turn, assignments of new work are usually driven by changes in delivery rates.

If everyone was equally proficient and skilled – or if the work was simple – this would not be an issue. Manufacturing jobs, however, require a high degree of proficiency and product-specific knowledge which are not easily transferable. This

is particularly acute in the shipbuilding industry. According to industry sources, it takes three to five years for a new hire off the street to be trained and developed into a journeyman employee. It takes an average of eight years to become a fully certified nuclear pipefitter. (Cuccias, 2018.) Estimates of shipbuilding costs require careful consideration of the productivity levels of so-called "green labor" (new hires) and "seasoned labor" (experienced employees). A RAND study on the shipbuilding industry summarizes the issue:

> "Workers with some experience are generally more productive than inexperienced workers. Thus, for a workforce with a higher proportion of inexperienced workers, additional effort is needed to complete an identical task. This additional work might be done using temporary workers, overtime, additional full-time employees, or even lengthening the ship production schedule." (Arena, 2004.)

But this issue is hardly restricted to shipbuilders. The short-term cost impact of hiring new, untrained employees is found in the aerospace industry as well. Commercial aircraft build provides several examples. In the late 1990s, Boeing attempted to significantly increase its 737 and 747 production rates by hiring thousands of new workers. Boeing's 1997 annual report laments: "In pushing to double production rate to meet heavy demands of a booming market, we experienced serious cost and schedule problems." (Boeing, 1997.) A front page story from The New York Times that same year describes this further:

> "In early October, overwhelmed by thousands of foul-ups, Boeing temporarily halted production of the 747 as well as the smaller 737….Boeing had to scramble to find people to build its airplanes, hiring 32,000 workers in the last 18 months. Despite what they describe as an aggressive training program, with five weeks of instruction before starting work, Boeing executives conceded that many new workers were still not fully prepared. 'We have incurred the penalty of these people learning' on the job, said Gary R. Scott, the vice president in charge of producing the 737 and 757." (Zuckermann, 1997.)

Interestingly, Boeing experienced similar issues on the same 747 production line 30 years earlier:

> "At the time production was starting on the 747, Boeing could not find enough workers in the Seattle area and was forced to recruit intensively. Of the workers hired, less than half developed into normally productive workers. Labor hours per aircraft increased as production rate and cumulative quantity increased, i.e., the learning curve had a positive instead of a negative slope." (Large, 1974.)

Yet another case comes from McDonnell Douglas during the same time period, as it struggled to keep up with demand for a stretched DC-8 as well as an increase in DC-9 production. (Large, 1974.)

These impacts are not permanent. Eventually the new employees reach a level of proficiency and their productivity approaches that of experienced employees. But even temporary impacts can create painful damage on production schedules and company profits.

Disruption due to changes in staffing levels is not restricted to new hires brought on to accommodate a production rate increase. Delivery rate *decreases* which result in workforce r*eductions* can create similar problems, as we will see. Even maintaining the same number of employees in an area can be problematic, if shop management is forced to reassign roles and responsibilities due to a reduction in another department.

Types of Workforce Changes

As mentioned earlier, there is very little discussion of this potential disruption in learning curve literature. The only text which deals with it at any length is E. B. Cochran's *Planning Production Costs: Using the Improvement Curve*. (Cochran, 1968.) The first section of this paper draws extensively from Cochran's exposition.

Suppose a new customer is signed up and delivery rates increase. Small increases in delivery rates can sometimes be satisfied by greater efficiencies in the learning curve as hours per unit decrease over time. However, beyond a certain point, additional employees will need to be hired. There is no change to the aircraft configuration (that would be a design change) but we simply need more bodies to produce the additional aircraft. The consequences are, as Cochran notes:

- Some tasks are continued by workers experienced performing them
- Some tasks are assigned to workers who have no experience performing them
- Some tasks are removed from personnel already performing them for reassignment to either the new workers or members of the original crew

A graphical illustration of a workforce addition due to an increase in production rates is shown in Figure 1. In the current state, it takes a crew of three mechanics to deliver an end-item in seven days. But in order to accommodate a production increase, the production line must be sped up to deliver an end-item in five days. To make that happen, a fourth crew member is added.

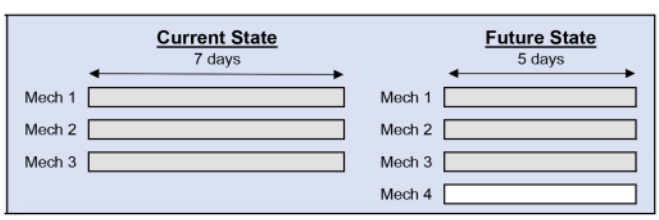


Figure 1. Workforce Addition

This will require work to be reassigned. That reassignment will obviously impact the fourth mechanic, who finds himself doing unfamiliar work. It will also impact the three original crew members. At a minimum, they will have work removed from them. But it is also possible that work will be reassigned among themselves as well.

To express this mathematically, Cochran suggests we use the proportion of new workers added as an index of the tasks which are new to the revised crew. To develop a "new man ratio," we define p_1 and p_2 as the number of people required before and after the change respectively. For a workforce addition, the "new man ratio" becomes:

$$t_a = (p_2 - p_1) / p_2$$

For a crew of 15, an increase to 20 would mean $t_a = (20 - 15) / 20 = 25\%$. We can interpret t_a as the *minimum* proportion of workers who must perform tasks which are new to them. In this case, at least a quarter of the crew members will have tasks which are new. Due to the reassignment of previous crew workers, the *actual* proportion of workers performing task new to them may be somewhat higher than t_a, but it cannot be less.

Workforce Reduction

Now suppose the delivery rate decreases due to a reduction in customer sales. As deliveries slow and production intervals increase, there is no longer a need for as many mechanics in the shop. To keep costs economical, some will be transferred off the program to perform work on other projects, or perhaps temporarily furloughed or even released by the company altogether.

Many aerospace shops are unionized. The basic agreement between labor and management typically regulates how workforce reductions carried out. In most cases, workforce reductions are carried out by seniority. If a reduction in force is required, the employees with the least experience will be laid off first. This happens

across the entire shop regardless of which program drives the workforce reduction – a reduction in the delivery rate for Program A may drive a layoff of employees in unaffected Program B, if they have less seniority. This potentially creates gaps in crews across the shop floor. As less-senior employees are laid off, the remaining employees are "bumped" into different work areas and sometimes different programs to accomplish the necessary reductions.

There are two major cost impacts of these moves:

- Some employees will have all-new tasks due to being "bumped" into a new area with associated learning loss.
- The remaining employees will have an expanded scope of work as span times increase. Employees must be trained to handle additional scope; a percentage of everyone's work will be new to them.

A graphical illustration of a workforce reduction due to a decrease in production rates is shown in Figure 2. In the current state, we have four crew members working to deliver an end-item every five days. To accommodate the slowdown, we only need three crew members to deliver a product every seven days. The fourth mechanic may be moved to a different area, a different program, or leave the company altogether; but his work will have to be reassigned to the remaining three mechanics, who now find themselves performing tasks with which they are unfamiliar.

Not surprisingly, this necessary realignment of work will create a temporary disruption which will result in higher hours per unit while personnel become accustomed to their new tasks.

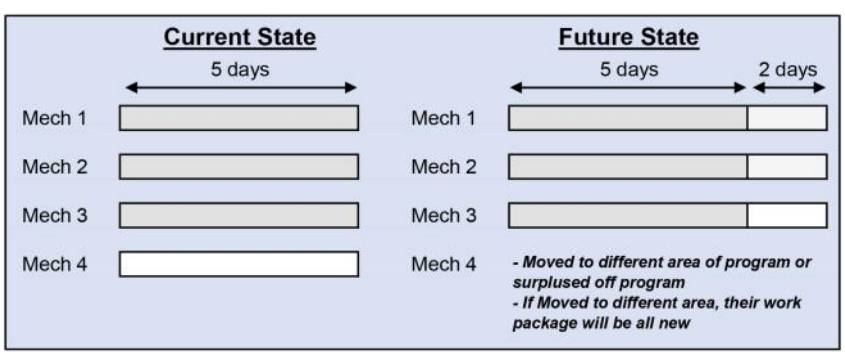


Figure 2. Workforce Reduction

As defined by Cochran, the "new man ratio" for a workforce reduction is:

$$t_d = (p_1 - p_2) / p_1$$

For a crew of 15 which is reduced to 10 mechanics, $t_d = (15 - 10) / 15 = 33\%$. At least one third of the work must be reassigned to mechanics for whom it is new. Note that our denominator is different for a workforce reduction. Whether we are dealing with a workforce increase or decrease, we always measure the change against the larger of the two numbers, p_1 or p_2.

Turnover

Turnover occurs when a certain number of mechanics are replaced by an equivalent number, but the total workforce count does not change. It frequently occurs as an extension of a workforce decrease. In the previous example, Program A experienced a delivery rate requiring a workforce reduction. Program B continues to build at the same delivery rate as before and requires no change in headcount. But because of the "bumping" of employees across the shop floor, Program B now finds itself with employees who formerly worked on Program A and are unfamiliar with the unique requirements and processes of Program B. This too will create some temporary disruption.

Designating p as the total number of employees and d as the number removed, Cochran defines the "new man ratio" for a task turnover as:

$$t_t = d/p$$

For a crew of 15, assume five mechanics are removed and their places taken by new ones. This yields $t_t = 5/15 = 33\%$. So at least

a third of the mechanics will be performing tasks which are new to them.

Estimating the Impacts of Changes

Calculating the "new man ratio" for a workforce change does not, however, tell us the cost impact of a workforce change. We cannot assume that a 25% "new man ratio" translates to a 25% cost increase. If we think back to Anderlohr's five elements of learning improvement – production personnel, supervision, continuity of production, tooling, and methods – we can see that only the first and second elements are impacted by a workforce change. (Anderlohr, 1969.) Assuming there is no change to the production configuration, tooling, or the production process itself, those contributors to learning should not see an impact.

It is also probable that minor workforce changes do not impact cost. In any large organization, there is a certain level of turnover – hires, firings, retirements – which occurs as an ordinary part of the business. "[I]t appears," writes Cochran, "that the new manpower effect must exceed a certain 'threshold' level before its cost effects need be taken into account." (Cochran, 1968.)

Nonetheless, it seems reasonable to make four assumptions about the impact of workforce changes:

a) Employees new to a task will initially perform less efficiently than experienced employees.
b) Over time the performance of new employees will improve relative to experienced employees.
c) At some point the performance of new employees should converge with that of experienced employees.
d) How long it takes to fully integrate a new employee varies depending on how much prior experience that employee has – with the industry he is working in, with the specific company he works for, with the production program that employs him, and with his or her specific work assignment. The more familiar an employee already is with Program A, for example, the faster his performance in a new job will approach the other Program A employees already performing that job.[1]

These assumptions can be illustrated graphically, as seen in Figure 3. If we use employee performance to earned standard as our baseline for efficiency, we can see that employees performing a task new to them initially perform at a higher variance factor (measured as actual hours divided by earned standards) -- which is to say, they will be less efficient relative to their more experienced peers. This delta cost premium will continue for some length of time until eventually the performance levels converge, and our new hires are no longer "new."

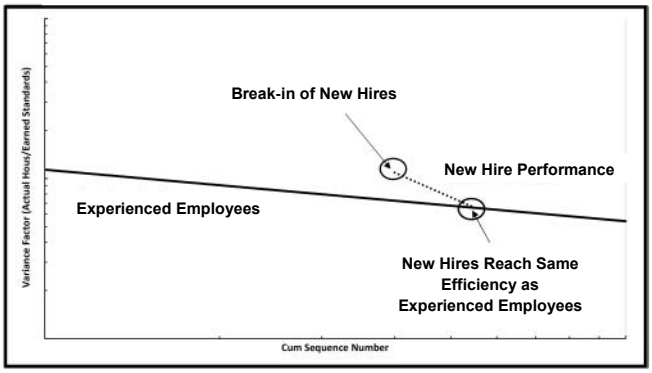


Figure 3. Theoretical Performance of New/Experienced Employees Over Time

Students of the learning curve may note that this graph looks like the cost impact of a product design or configuration change. Certainly, the two have similarities. "In both cases the work is new to the operator, the penalty is larger for events occurring further out in the production sequence, and it shrinks rapidly as production continues," writes Cochran. "However, a workforce change is less severe than a design change because supervision, tooling, support personnel, and other crew members are left unaltered." (Cochran, 1968.)

The graph leaves us with two unanswered questions, however. First, how much loss of

learning occurs at the point an employee begins a task new to him? Second, how long does it take for the performance of new employees and experienced employees to converge?

Naturally at this point, we turn to Cochran for guidance on the next steps. But here the usually reliable author fails us. He attempts to translate the workforce change into a "task turnover ratio," and then use the ratio to develop the cost of the workforce change. But the "task turnover ratio" is constructed *a priori* without any reference to data and is consequently impossible to duplicate or verify. Moreover, his solution for converting the "task turnover ratio" into a cost impact is clumsy and difficult to incorporate into a model. (Cochran, 1968.) Clearly a different approach is required.

Our Study

Earlier in the paper, we identified four assumptions about the cost impact of manufacturing personnel changes. Can we verify these assumptions from actual data?

The ideal dataset would allow us to examine performance data by employee with sufficient visibility to identify if an employee was new to an area or already experienced in the work. The dataset would cover also a reasonably long period of time, allowing us to examine a "before" and "after" situation related to a major production rate change involving either a sizeable increase or decrease in workforce levels.

Employee-level data is valuable because it allows us to correlate change in cost to changes in workforce levels more easily. High-level hours per unit (HPU) data runs the risk of being contaminated by other factors that influence cost – part shortages, quality problems, etc. While we might observe increases in manufacturing hours as workforce levels changed, we cannot be certain using HPU data if the increased hours were driven by workforce, or by other factors. Employee-level data, on the other hand, allows us to see how the performance of employees new to an area varied from that of experienced employees, giving us a greater certainty that we have really captured the cost delta associated with workforce changes only.

It was also important to identify a point in an aircraft program where the configuration as well as the planning, tooling, and build processes had stabilized. There is always a large influx of personnel at the beginning of a program as the initial aircraft are built, but it is impossible to distinguish the cost impacts related to new personnel becoming acquainted with their jobs from the seismic shifts in HPU caused by managing engineering changes, correcting planning and tooling for so-called "make-it-work" changes, the untangling of part shortages driven by late engineering, and general chaos of an aircraft development program. Likewise, a major change in the aircraft configuration creates similar confusion, albeit on a smaller scale. Only during a period of program maturity can we successfully analyze the unique impacts of a workforce change.

Fortunately, a situation arose on an Aeronautics production program which involved a substantial increase in production rates and a corresponding change in shop floor personnel headcounts. For proprietary reasons, this program will not be named in this paper, but simply referred to as Program D.

Production Intervals and Workforce Changes

Before discussing how data for Program D was collected and analyzed, a quick discussion of production intervals (PI) is in order. To measure the schedule impacts of increases or decreases in delivery rates, production management frequently refers to "production interval," "line move rate" or "takt time." All three terms mean the same thing: the number of workdays between product deliveries.

The PI is directly tied to delivery rates. Assume that a typical work-month has 21 planned workdays (excluding Saturdays, Sundays and holidays). To support a delivery rate of four per month, we must deliver aircraft approximately every five days. (This is calculated as 21 planned workdays per month / 4 deliveries per month = 5.25 days between line moves. If the delivery rate slows to three per month, then the line move rate increases to seven days. (This is calculated as 21 planned workdays per month / 3 deliveries per month = 7 days between line moves.) The PI and delivery rate per month are inversely related: as delivery rates increase, the PI decreases; as delivery rates decrease, the PI increases.

Program D required a significant increase in production rates to be carried out over a two-year period. For example, in the mate thru delivery area, the PI was scheduled to decrease from 4 to 2.7 to 2 over three lots, yielding a doubling of delivery rates. Other components of the aircraft, while working at different PIs due to coproduction, had similar rates of change.

Overall, the number of employees touching the aircraft during build also had to double over a two -year period. The timing of the hiring waves varied by component depending on their position in the build sequence (component assembly was first, followed by mate and final assembly, and finally flight line). The increase in personnel was not accomplished all at once but timed with the planned decrease in PI. This meant that a given area experienced two and sometimes three distinct hiring waves.

The Method

We began by pulling weekly performance data by employee for the components under study over a 42-month period. In order not to release too much data about Program D, we will avoid identifying specific calendar years and refer to these periods as Year Zero, Year One, Year Two and Year Three. The six-month period prior to the workforce increase will be designated Year Zero. Based on staffing plans, we identified January of Year One as the beginning of the planned workforce increases. Weekly data was then accumulated through December of Year Three. The point of "full staffing" was reached earlier – in March of Year Three – but going beyond that point gave us the ability to see how long it took for performance between new and experienced employees to stabilize.

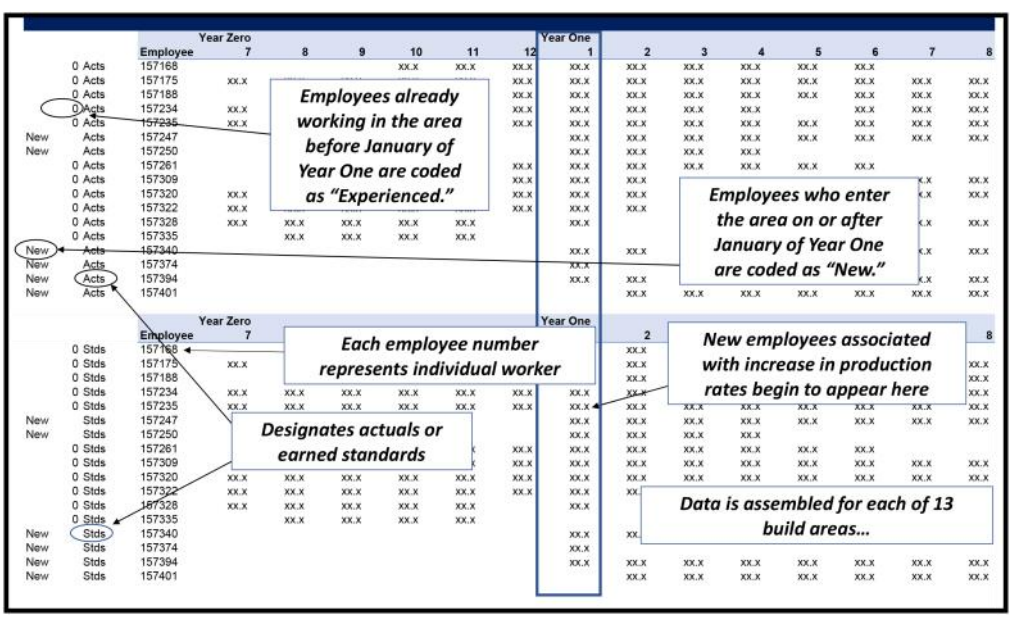


Figure 4. Example of Data Collection by Employee (Subset)

Status	Catg	Employee	1	2	3	4	5	6	7	8	9	10	11	12
New	Stds	168775	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168775	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Stds	168785	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168785	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Stds	168786	xx.x	xx.x	xx.x									
New	Acts	168786	xx.x	xx.x	xx.x									
New	Stds	168788	xx.x	xx.x	xx.x									
New	Acts	168788	xx.x	xx.x	xx.x									
New	Stds	168789	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168789	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Stds	168792	xx.x	xx.x										
New	Acts	168792	xx.x	xx.x										
New	Stds	168794	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168794	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Stds	168795	xx.x	xx.x	xx.x									
New	Acts	168795	xx.x	xx.x	xx.x									
New	Stds	168879	xx.x	xx.x	xx.x									xx.x
New	Acts	168879	xx.x	xx.x	xx.x									xx.x
New	Stds	168885	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168885	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Stds	168886	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x
New	Acts	168886	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x	xx.x

Represents the first month's performance for a given employee, once introduced into the area

Successive months represent month 2 in the area, month 3 and so forth....

Figure 5. Example of Data Stratification over Time (Subset)

The data gave us actual hours and earned standards by employee number for each component. We limited the study to assembly areas since these planned standards largely represent an engineered standard, typically constructed from predetermined time systems such as Methods-Time Measurement (MTM) data. In the old MIL-STD-1567A vocabulary, these would be Type I standards. (MIL-STD-1567A, 1983.) The fabrication areas, on the other hand, use standards developed from prior actuals (Type II standards per MIL-STD-1567A) and while, sufficiently accurate for fabrication shop management, do not demonstrate enough fidelity over time to provide us with a meaningful analysis.

Using January of Year One as a baseline, any employee who was charging to an area during the six months prior (during Year Zero) was labeled as an "experienced employee." An examination of staffing profiles for the prior two years showed a very stable headcount by area, giving us confidence that we could safely designate these employees as "experienced" without researching everyone's situation. Correspondingly, employees which began charging to an area after January of Year One was designated as "new." These could be existing employees which were transferred to a different area, or employees which were new to the company altogether. Figure 4 shows a tabular example of how the data was arranged.

The weekly data by individual build area was rolled up by month by employee and segregated by "new" and "experienced" employees. The variance factors (actual hours spent divided by earned standard hours) was used as the point of comparison.

One of the problems in the data was that each build area experienced two or three waves of new employees, making it difficult to analyze how long it took for a wave of employees new to an area to reach the same performance levels as their peers. To solve this problem, the performance data of new employees was stratified across time. We then added up the actual hours and earned standards for the first month of performance by employees and calculated a collective variance factor. We then did the same thing for month two, month three, etc. We then calculated the average variance factor for employees who were in place prior to Year One across all months to calculate a baseline for comparison. Figure 5 shows an example of the data stratification.

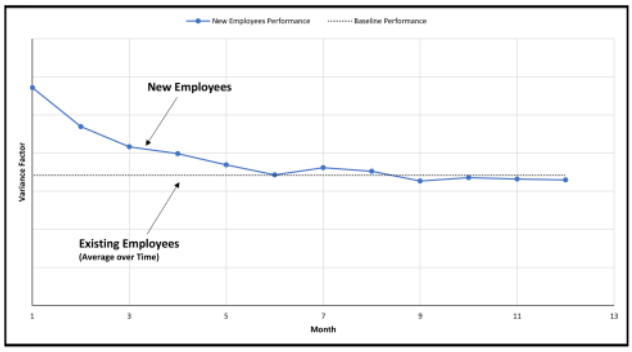


Figure 6. Comparison of New/Experienced Employee Performance by Month Since Introduction - Example Build Area

This allowed us to construct charts by build area like Figure 6.

In their first month after introduction to a new area, the performance of employees was worse than the baseline performance of their experienced peers. This improved over the succeeding months until – in this case – the performance of the two groups converged at month six.

Across all build areas, the performance of employees new to an area was initially worse than that of experienced employees. Most areas (10 of the 13 build areas) showed convergence or near-convergence (defined as achieving a less than 10 percent difference) to the baseline over varying lengths of time. The median length of time for convergence was nine months.

This gives us the opportunity to validate our four assumptions:

- New employees had worse performance initially than their experienced peers.
- New employees improved their performance relative to more experienced employees.
- In most cases, there was convergence of new employee performance back to the baseline.
- The median timeframe of that convergence took nine months.

Using this finding, the data labels were reconfigured. For the first nine months after introduction to an area, individual employees were classified as "new." For month 10 and on, those same employees were now considered "experienced."

This allows us to show performance by new and experienced employees over time and see the overall performance delta over time. Figure 7 shows the variance factors for a sample build area plotted on the first Y axis, with the number of employees counted as "new" (subject to the nine-month time frame) plotted against the second Y axis.

By calculating this performance delta in terms of performance to standard, we can also calculate the estimated hours impact of introducing new employees. The cost of integrating new employees for a given month is calculated using the following formula:

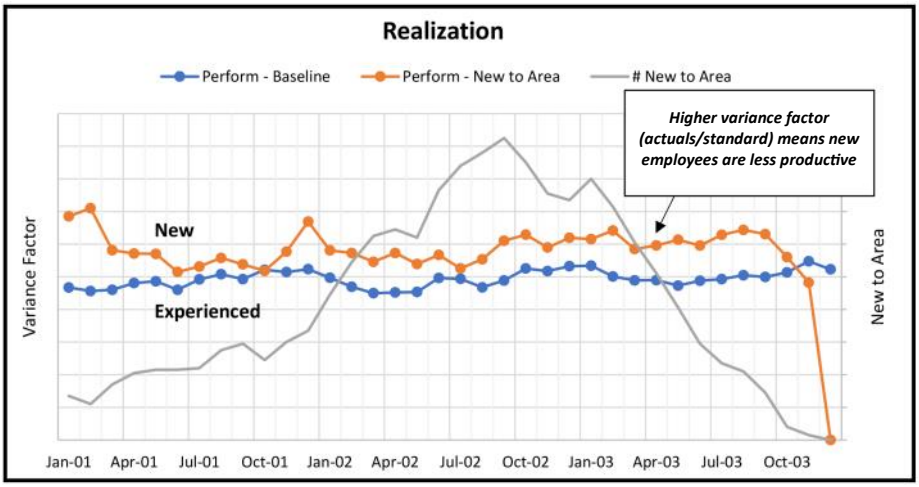


Figure 7. Comparison of New/Experienced Employees (Subject to Nine-Month Limit) Over Time – Sample Build Area

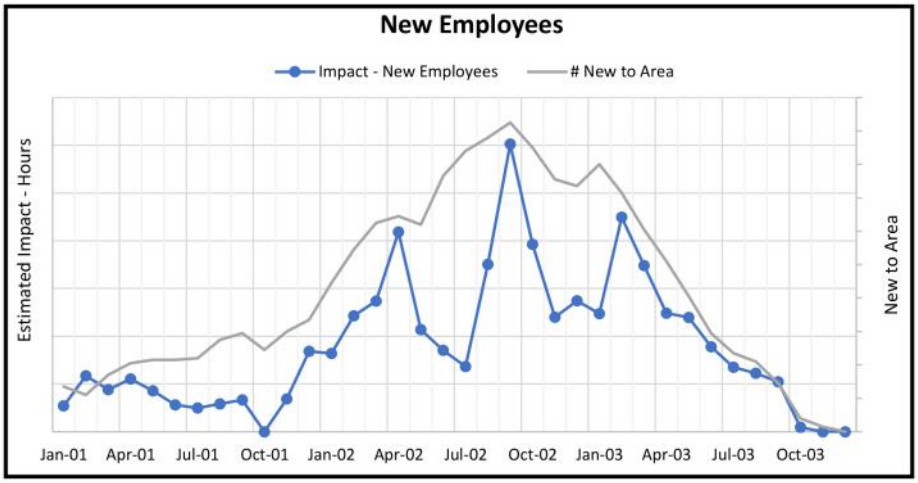


Figure 8. Cost Delta for New Employees Correlated to Number of Employees New to Area – Sample Build Area

$$H_d = H_s \times (V_n - V_e)$$

where H_d = delta cost (hours) of integrating new employees for a given month, H_s = total earned standards by new employees for a given month, V_n = monthly variance factor for new employees and V_e = monthly variance factor for experienced employees.

We can now look for the correlation between the hours delta attributable to introducing new employees against the number of employees considered new to the area at any given point in time. An example of this is given in Figure 8.

Our goal now is to calculate some useful relationship between the level of workforce changes and the resulting cost impact. To do this, all the data was aggregated across the build areas by month. For each month, the following data was calculated:

- Percent of new employees (number of new employees divided by the number of total build employees)
- Percent cost delta (delta hours associated with new employees divided by the total number of hours charged)

This allowed us to relate the two variables and see the relationship plotted on a log-log space in Figure 9.

For proprietary reasons, the value of the intercept is omitted. However, since the value of the logarithmic coefficient is close to unity, the relationship approaches a linear relationship. The R-square fit is excellent, and the model explains virtually all the observed variation.

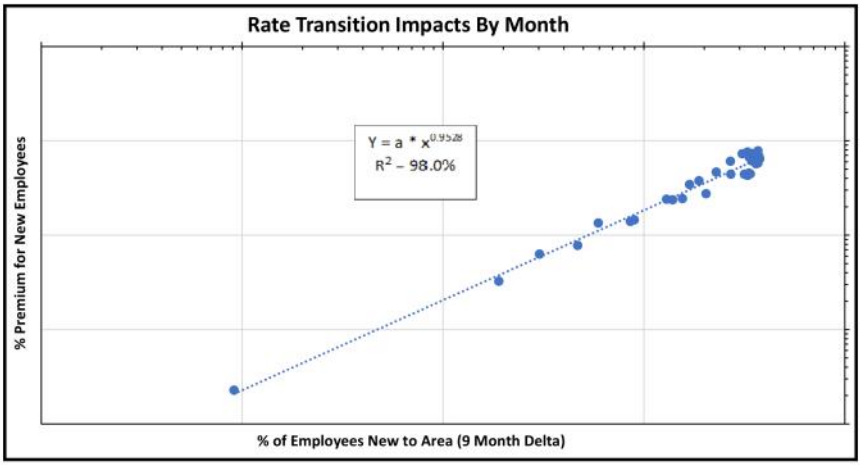


Figure 9. Observed Relationship Between New Employees and Associated Cost Premium

Length of the Recovery Period

This study represents, in many cases, a worst-case scenario. First, Program D's workforce doubled over an approximately two-year period. Most workforce increases or decreases are not so severe. Second, this workforce increase was accomplished in successive, overlapping hiring waves. Most workforce increases or decreases occur at as one-time discrete events. Finally, many of the new employees were not simply new to the program or the company, but new to the aerospace industry in general.

This suggests that the nine-month recovery period could be shorter under other rate change scenarios. For example, a smaller workforce increase could transfer workers already in place on the shop floor but currently working on a different program. It is logical to suppose that those transferred employees will not take as long to acclimate themselves to their new jobs. In the case of a work reduction, the remaining employees are already on the program, but they may have to learn some new tasks. Theoretically this recovery period would be even shorter.

Application of the Model and Approach

Now that we have built this model, how is it applied?

The ideal situation would allow us to work from a detailed staffing forecast. That forecast would tell us how many heads need to be added or deleted and the timing of those impacts. However, this data is not always available – for example, our company may not have detailed forecasts longer than a year out.

However, using production interval and estimated hours per unit, we can approximate headcount levels in order to use the model. Note that the model depends on relative changes in headcount, and not absolute values. If we can correctly approximate the magnitude of the change, the answers that are returned should be good.

Table 1 provides an example of how we can calculate a relative headcount change using this information. We assume at unit 600 a production interval change takes place which will create a workforce reduction. Based on this information, we can calculate a relative headcount change and the associated ratio of new/reassigned task shown in Table 1.

Two caveats should be noted. First, this calculation will provide a value of full-time equivalent heads. In fact, due to absences (vacation, sick leave), overtime, or time charged to indirect efforts, the actual yield rate per employee could be more or less than a simple workdays per month multiplied by hours per day. However, since our interest is in the relative change in headcount, these refinements should not significantly alter the values or cause difficulties. Second, this approach also ignores the possibility at very low production rates, a minimum staffing level must be maintained in order to preserve critical skills. Where such a

Unit Break-in of Change	600
Estimated Hours per Unit (HPU) at Break-in	5,000
Average Work Days per Month	21
Number of Hours per Day	8

	Production Interval (PI) (Days)	Delivery Rate per Month
Before Change	10.5	2.00
After Change	14	1.50

	HPU		Delivery Rate per Month		Hours per Month		Work Days per Month		Hrs per Day		Operators
Before Change	5,000	x	2.00	=	10,000	/ [	21	x	8	] =	60
After Change	5,000	x	1.50	=	7,500	/ [	21	x	8	] =	45
					Headcount Reduction						15

* Delivery rate calculated as: Average Work Days per Month / Production Interval

Task Reassigned to Remaining Operators:

Operators Reduced	15
Total Operators before Reduction	60
Percent Reassigned Task	25.0%

* 25.0% reduction in operators, so 25.0% of task will be new to remaining operators.

Table 1. Assembly Workforce Reduction

potential exists, the estimator should consult with his Industrial Engineering department to establish such a minimum skill level and adjust his calculations appropriately.

Having calculated a 25% reassignment ratio, how do we translate into a cost impact? First, we must assume how long the disruption will last. We might imagine three general separate scenarios which could occur:

Scenario A, Workforce Reduction: The reduction is accomplished within the program, with work reassigned to existing employees. While these workers must learn some new tasks, they are already familiar with techniques peculiar to the program. This scenario will have the shortest recovery period.

Scenario B, Workforce Increase: The increase is accomplished by transferring workers from other company programs. While familiar with the aircraft industry as well as the company "way of doing business," these workers will not be familiar with techniques peculiar to the program. In addition, some existing workers will be reassigned to different roles in order to help balance crews and optimize workflow. This scenario will have a recovery period somewhere between Scenario A and C.

Scenario C, Workforce Increase with Outside Hires: The increase is largely accomplished by hiring workers from outside the company. These workers are not only unfamiliar with the unique program peculiarities, they are also unfamiliar with the company's way of doing business. They may not even have prior experience in the aircraft industry. This scenario will have the longest recovery period.

Our previous empirical study falls squarely into Scenario C. Therefore, the median nine-month recovery period would represent the "worst case." While it follows logically that Scenario A will have the shortest recovery period, with Scenario B falling somewhere between A and C, it is difficult to establish *a priori* exactly how long these periods will last. The estimator should consult with Production management team and the Industrial Engineering department to help him make the best determination.

For the purposes of our example, we have assumed a four-month period of recovery on the grounds that our situation corresponds most closely with Scenario A. Table 2 shows the calculations. It is first necessary to calculate the undisrupted HPU and spread those hours across time. Although our PI change is scheduled to take place in January effective unit 600, it is important to note that any component in work in January, even ones started prior to unit 600, can be impacted by the personnel change. That is because the "bumping" of personnel to accommodate the workforce reduction is likely to reassign workers across stations, potentially affecting any unit currently in work.

For proprietary reasons, the actual cost equation

							Percent Reassigned Task	25.0%	25.0%	25.0%	25.0%	
							Percent Cost of Reassignment	x%	x%	x%	x%	
	Hours		Spread of Hours By Month				Percent Cost of Reassignment				Disruption
Unit	per Unit	Start Date	Jan	Feb	Mar	Apr	Jan	Feb	Mar	Apr	Hours
595	5,010	15-Oct	5.6%				x%				zzz
596	5,008	1-Nov	22.2%				x%				zzz
597	5,006	15-Nov	44.4%	5.6%			x%	x%			zzz
598	5,004	1-Dec	55.6%	22.2%			x%	x%			zzz
599	5,002	15-Dec	44.4%	44.4%	5.6%		x%	x%	x%		zzz
600	5,000	1-Jan	22.2%	55.6%	22.2%		x%	x%	x%		zzz
601	4,998	22-Jan	5.6%	44.4%	44.4%	5.6%	x%	x%	x%	x%	zzz
602	4,996	12-Feb		12.5%	52.8%	33.3%		x%	x%	x%	zzz
603	4,994	4-Mar			22.2%	55.6%			x%	x%	zzz
604	4,992	25-Mar			1.4%	33.3%			x%	x%	zzz
605	4,990	15-Apr				5.6%				x%	zzz
											z,zzz

For ship #600, disruption hours = 5,000 hrs x [(22.2% * x%) + (55.6% * x%) + (22.2% * x%)] = zzz hours

* Date of PI Change: 1 January
* Assumed Months of Recovery: 4 months (January through April)
* Total Build Span: 3 months
* Percent Cost of Reassignment: Not shown due to proprietary reasons. This is percent cost delta correlated with Percent Reassigned Task.

Table2. Calculating Cost of Reassignment

cannot be released. We will assume that a 25% reassignment ratio relates to a cost premium of x%. Therefore, for the month of January, the total HPU for a given unit number is multiplied by the percent of task completed that month times x%. The result provides us the disruption expected for January. Similar calculations are performed for the month of February, March and April. Since we have assumed the impact ceases after four months, our calculations cease after April.

In the end, our disruption will impact 11 units (unit numbers 595 through 605). It can be inferred that the number of aircraft impacted will increase at higher production rates (or longer production span times) since there will be more work in process at any given time, and therefore more opportunity for additional aircraft to be disrupted by workforce changes. For our example, the total disruption will look something like Figure 10.

This distribution is different from the theoretical construct supposed in Figure 3. In that case, the units prior to the break-in aircraft for the rate change would be unaffected. But this theoretical construct does not really account for the "bumping" effect, which is likely to spread the disruption across any unit in work at the time of the workforce reduction.

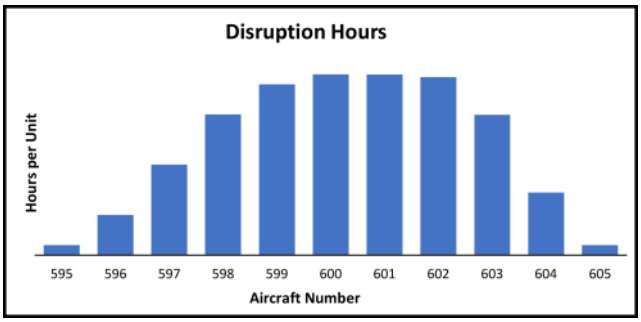


Figure 10. Disruption Hours by Unit

It could be argued that if new or reassigned workers become more proficient over time that the value of x will also change over time – that it should be higher in the first month of disruption and decrease over the four-month period until reaching the termination point, following the pattern we observed in Figure 6. This adjustment will potentially alter the shape of the distribution of disruption hours by unit. Such a feature could be incorporated into the model, but for simplicity of explanation a flat percentage has been assumed for each of the four months of impact. Depending on the estimator's specific needs, precision at the individual unit may be traded off for simplicity of presentation and calculation.

Next Steps

As always, further research remains. Testing of the model in future instances of production rate increases or decreases will provide insight to its accuracy and the need for further refinements. In addition, data collection surrounding smaller -scale workforce increases or decreases will provide further insight into the length of the recovery period. Our empirical research has established the outer bound of the time between initial disruption and eventual recovery to the underlying performance trend for a large workforce expansion. But we have supposed that this is the worst case, and for workforce reductions, cases of pure turnover, or smaller workforce increases, we are left to judgmentally decide how long the recovery period should be. Further research will reduce that element of judgment and provide the estimator more precise guidance, as well as verification that the model provides accurate predictions in those specific scenarios.

Conclusion

In the idealistic world of learning curve literature, fluctuations in delivery rates or the potential disruption created by "green labor" are frequently ignored or else assumed away. Its silence suggests that these matters are not worthy of much consideration.

Charles E. Ferguson, an economics professor at Texas A&M University, facetiously defended himself to students who demanded to know how theoretical economics related to the problems of the real world by saying: "The real world is a special case -- and not a very interesting one at that." The academic can strike such a pose. But cost estimators must live in – and produce estimates for – the real world.

Changes in production rates and the associated impacts on workforce numbers and experience levels remain one of the production estimator's greatest challenges. This paper aims to bring these issues to light and provide an analytical framework for these changes. It also provides a model for forecasting the impact of these changes based on analysis of a real-life case study of a workforce expansion. This analysis is presented not as the final word on the subject. As John Maynard Keynes reminds us, "It is better to be roughly right than precisely wrong." It is hoped that this paper will stimulate further discussion in the estimating community on practical solutions to handle one of the common production issues.

Footnotes

[1] Compare these to the assumptions of RAND's 2004 model to assess changes in shipbuilding labor: "(1) It takes three years to become fully proficient at a trade; (2) worker productivity improves linearly with experience to a fully qualified status, beyond which no further productivity is modeled; (3) a worker with no experience has a productivity of two-thirds that of a fully proficient worker; and (4) the changes of hiring a worker at any experience level are identical." (Arena, 2004.)

References

Anderlohr, George (1969). *What Production Breaks Cost*, Industrial Engineering, September 1969, pgs. 34-36.

Arena, Mark V.; Schank, John F.; Abbott, Megan. *The Shipbuilding and Force Structure Analysis Tool: A User's Guide*. RAND. 2004

Cochran, E. B. (1968). *Planning Production Costs: Using the Improvement Curve.* San Francisco, CA: Chandler Publishing Company.

Cuccias, Brian (2017). *Statement of Brian Cuccias, President, Ingalls Shipbuilding, Huntington Ingalls Industries, Subcommittee on Sea Power, Senate Armed Services Committee*, May 24, 2017. Retrieved from https://www.armed-services.senate.gov/imo/media/doc/Cuccias_05-24-17.pdf

Johnstone, Brent M. (2017). *Do Production Rates Really Matter?* Retrieved from https://www.iceaaonline.com/pdx17papers/#DA03

Large, Joseph P.; Hoffmayer, Karl; Kontrovich, Frank (1974). *Production Rate and Production Cost.* RAND, R-1609-PA&E, December 1974.

MIL-STD-1567A (1983). Work Measurement. Washington, DC: Department of Defense. March 11, 1983.

The Boeing Company (1997). 1997 Annual Report.

Zuckermann, Laurence (1997). *Stacked Up At Boeing; Plane Giant Tries to Recover From Its Snarls in Production.* New York Times, November 14, 1997.

Brent Johnstone *is a Lockheed Martin Fellow and production air vehicle cost estimator at Lockheed Martin Aeronautics Company in Fort Worth, Texas. He has 33 years' experience in the military aircraft industry, including 30 years as a cost estimator. He has worked on the F-16 program and has been most recently the lead Production Operations cost estimator for the F-35 program. He has a Master of Science from Texas A&M University and a Bachelor of Arts from the University of Texas at Austin.*

Does Cost Team Leadership Matter?

Christina N. Snyder, CCEA

Abstract: Does cost team leadership matter? An anonymous survey of 150+ cost analysts unanimously reported that a cost-team lead's effectiveness ultimately impacts the team's products. However, there has been minimal guidance as to what defines good leadership. Using the ten behaviors identified by Google's Project Oxygen, this paper seeks to understand what skills are necessary for successful cost leadership. The findings lead to a simple conclusion that mirrors that of Project Oxygen: improving our soft skills will improve cost leader efficacy.

Introduction

The International Cost Estimating and Analysis Association (ICEAA) is a cost estimating organization built by cost estimators for cost estimators. While having a cost-specific emphasis does facilitate focus on state-of-the-art estimating techniques, it also leads our community to potentially be unaware of peripheral opportunities for professional growth. This paper seeks to capitalize on one of these peripheral knowledge areas, hypothesizing that by using lessons learned from Google's Project Oxygen we can identify leadership behaviors that will result in positive cost team outcomes. If there is consensus from our community on the behaviors of good leadership, it would establish the groundwork for future training on cost leader improvement.

This paper initiates a similar approach to survey, reflect, and update training; beginning with the question "does cost team leadership matter?" Given a lack of literature and training materials on leadership skills in the cost community, could the behaviors identified in Project Oxygen as being related to good managers translate to better cost leadership? This paper builds a foundation, expands upon, and provides context to the ten attributes that characterize great managers at Google. It shows the relevancy of Google's research to the cost estimating community and, in replicating their study, highlights the important attributes of cost leaders. It also explores the differences between the opinions of leaders and non-leaders and how the findings from our cost community compares to the Project Oxygen management ranking. The results of these findings will provide examples of how the top perceived attributes can be put into action within our work. The identified attributes should also fuel further exploration into the relationship between leadership skills and team efficacy, ideally creating new training for those growing into leadership roles.

Background - Google's Project Oxygen

A case study was conducted by Professor David A. Garvin, Alison Berkley Wagonfeld, Executive Director of the HBS California Research Center, and Senior Researcher Liz Kind for the Harvard Business Review in 2013 highlighting the behavior measurement of Google's management, why managers matter and what the best managers do1. Known for a culture of consistent improvement, Google requested their personnel analytics team internally identify opportunities for team enhancement; they thought outside the box and questioned "Do Managers Matter?" Beginning in 2009 with the Google People and Innovation Lab (PiLab), they called the effort Project Oxygen and hypothesized that a very flat organizational hierarchy like Google's "of engineers for engineers" was ideal, and that managerial roles had very little impact on

Figure 1 - Manager Quality vs. Performance
From https://www.youtube.com/watch?v=JattR1uoX7g

performance. To determine if managers matter, they wanted data to see who the highest performing managers were, who were the lowest performing, and whether it impacts the team. Figure 1 shows their scatterplot of manager performance and the team's view of the manager. In the next step they studied the quantifiable differences between the most effective and least effective managers. Teams with managers in the most effective quadrant consistently had better team morale, less turnover, and greater employee satisfaction than those with less effective managers in the bottom left quadrant. For example, retention had a stronger correlation to manager quality versus other employee metrics like seniority, performance, tenure, or promotions. To the surprise of the researchers, the data suggested that not only did mangers matter, but that good managers had a significant impact on job satisfaction, employee retention, and performance.

With sufficient data to prove the correlation between manager quality and team performance, the researchers asked the next logical question – "What do the best managers do?". They sought to understand the qualities demonstrated by top versus the lower scoring managers. During the summer of 2009, Google conducted company-wide double-blind interviews with managers to identify the skills that correlate to manager efficacy and later compared that coded data with the manager performance. After several months of data analysis, they came up with eight behaviors that were common among high-scoring managers throughout the company. In 2018, behaviors 9 and 10 were added to make ten total actionable behaviors that improve manager performance.

These behaviors are listed in order of frequency with which the behavior was mentioned during the interview and analysis process. The results of the data analysis were shocking to many at Google. The company that had been built by engineers and typically promoted people based on their technical expertise seemed to value "soft skills" like being a good coach, creating an inclusive team, and caring about the team members more than their technical knowledge and abilities. Laszlo Bock, senior vice president of people operations, commented, "It turns out that

Figure 2 - re:Work Google Manager Behaviors

[technical skills] that's absolutely the least important thing. It's important but pales in comparison. Much more important is just making that connection and being accessible."6

Diving into these behaviors a little deeper, this is how Google defines each behavior:

1. ***Is a good coach*** – Agree on development priorities and check in with employee regularly
2. ***Empowers the team*** – Does not micromanage
3. ***Express interest for team personally –*** Show your team you care
4. ***Is productive and results-oriented*** – Focus on priority results and deliverables
5. ***Is a good communicator*** – Set the stage for two -way dialogue
6. ***Support career development*** - Help your team grow skills for their professional development
7. ***Has a clear vision –*** Develop and share your vision for the team
8. ***Has the technical skills to advise*** – Have the expertise and technical skills to advise team
9. ***Collaborates across Google*** – Create stronger, more deliberate connections across teams
10. ***Is a strong decision maker*** - Provide guidance and act swiftly

Google uses the Project Oxygen findings to revise their annual feedback surveys, curriculum, tools, and programs to improve manager quality but not to penalize or demote. "Project Oxygen was always meant to be a developmental tool, not a performance metric" and we're pleased to see that after coaching them on these skills, "the least effective managers improve the most over time."[1 I]

Methodology

For the purposes of this effort to replicate Google's study, the method of collecting data was an anonymous survey shared with ICEAA members and the cost estimating community. The primary purpose of the survey was to determine if the cost community agrees that team leadership matters and what behaviors are perceived to be the most important; it uses the ten identified Google Manager Behaviors and tried to establish their ranking in regard to cost team leadership. The survey was intentionally designed to be brief to maximize participation. It also included an open-ended response field to allow respondents to include any additional behaviors that should be added in the future. The study also sought to determine if there were any differences in the perceived importance of these behaviors between leaders and non-leaders.

The survey was completed by 163 cost analysts with cost estimating experience ranging from one to 40+ years. Approximately 80% of respondents had more than 5 years of cost estimating experience and 46.6% would traditionally be considered "senior cost estimator/analysts" with over 15 years of cost estimating experience.

To explore possible differences between leaders and non-leaders with regards to behavior ranking, a self-reported assessment of their past and current roles as cost estimating team leadership is provided in the table below. An overwhelming majority - 83.4% - of respondents have at one time in their career served as the leader of a cost team with one of more analysts reporting to them,

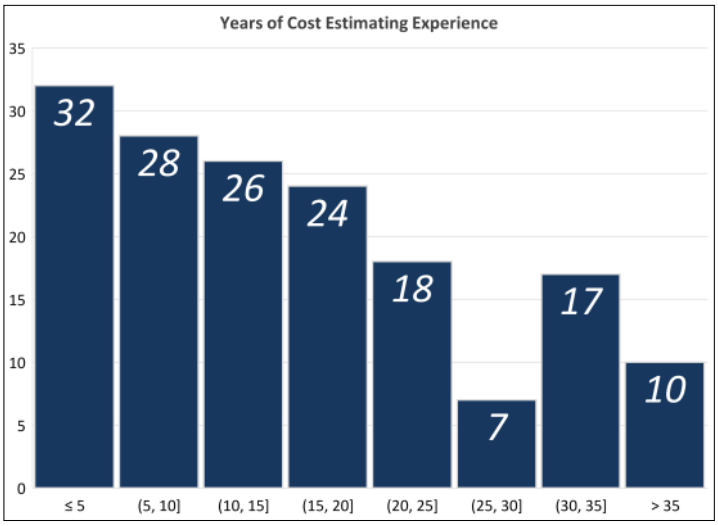


Figure 3 - Experience Histogram

	Yes	No	Yes %	No%
Have you ever served as the leader of a cost team, as in, leading the effort for a cost product with one or more analysts reporting to you?	136	27	83.4%	16.6%
Within your current company/organization are you considered leadership or management?	89	74	54.6%	45.4%

and currently 54.6% are considered leadership/ management.

With regards to the importance of leadership in cost estimating, 100% of respondents responded "True" - a cost team lead's effectiveness has an impact on the cost products generated by the team. Given this unanimous response, defining which behaviors leadership should exhibit was the next step.

The survey then asked "The following are qualities that you may value in a cost team lead. Using a Likert scale of (1-5), rate these qualities from important (1) to not important (5)". The respondents were then presented with attributes of the ten behaviors but did not specifically reference the behaviors themselves. This is similar to how Google surveys their own employees and helps by defining a specific attribute to try to reduce vagueness or different understanding the definition of the behaviors. Based on their response regarding their leadership role, respondents were presented with one of two versions of the survey: one version was

	Non-Leaders	Leaders
Q1	My team lead assigns stretch opportunities to help me develop in my career.	Leadership assigns stretch opportunities to help team develop in their careers.
Q2	My team lead communicates clear goals for the team.	Leadership communicates clear goals for the team.
Q3	My team lead gives actionable feedback on a regular basis.	Leadership gives actionable feedback on a regular basis.
Q4	My team lead provides the autonomy needed to do individual jobs (i.e., does not get involved in details that should be handled at other levels).	Leadership provides the autonomy needed to do individual jobs (i.e., does not get involved in details that should be handled at other levels).
Q5	My team lead consistently shows consideration for me as a person.	Leadership consistently shows consideration for team as people.
Q6	My team lead keeps the team focused on priorities, even when it's difficult (e.g., declining or deprioritizing other projects).	Leadership keeps the team focused on priorities, even when its difficult (e.g., declining or deprioritizing other projects).
Q7	My team lead has the technical expertise needed to review my work.	Leadership has the technical expertise to review the team's work.
Q8	The actions of my team lead show they value different perspectives brought to the team, even if it is different from their own.	The actions of leadership show they value different perspectives brought to the team, even if it is different from their own.
Q9	My team lead makes tough decisions effectively (e.g., decisions involving multiple teams, competing priorities).	Leadership makes tough decisions effectively (e.g., decisions involving multiple teams, competing priorities).
Q10	My team lead effectively collaborates across boundaries (e.g., team, organizational).	Leadership effectively collaborates across boundaries (e.g., team, organizational).

presented to those who responded that they are currently in a leadership role, while the other was presented to non-leaders.

The survey then presented the ten behaviors and had the respondents pick the top five that seemed most important to them, finally narrowing those down to the number one most important behavior. They were then presented with the five unselected behaviors and asked them to pick the least important behavior. The final question was a free response to allow respondents provide any skills not mentioned that are important qualities of cost team leaders.

Results

Initial analysis compared the survey results of leaders and non-leaders in the cost community to the Project Oxygen ranking. By calculating the average Likert scale response for each attribute, an ordinal ranking of most to least important was established for the leader and non-leader groups. Figure 4 shows not only the discrepancies between the cost community's responses and the original Google study but the differences in reactions to the attributes between leaders and non-leaders within the cost community.

While both cost groups value the importance of *clear communication of goals*, opinions of leaders and non-leaders differ in several notable actionable attributes, such as non-leaders assigning a much higher importance to leadership *having the technical expertise to review the team's work* and *giving actionalbe feedback on a regular basis* more than leaders do. Meanwhile, self-identified leaders seem to give higher significance to *leadership consistently shows consideration for*

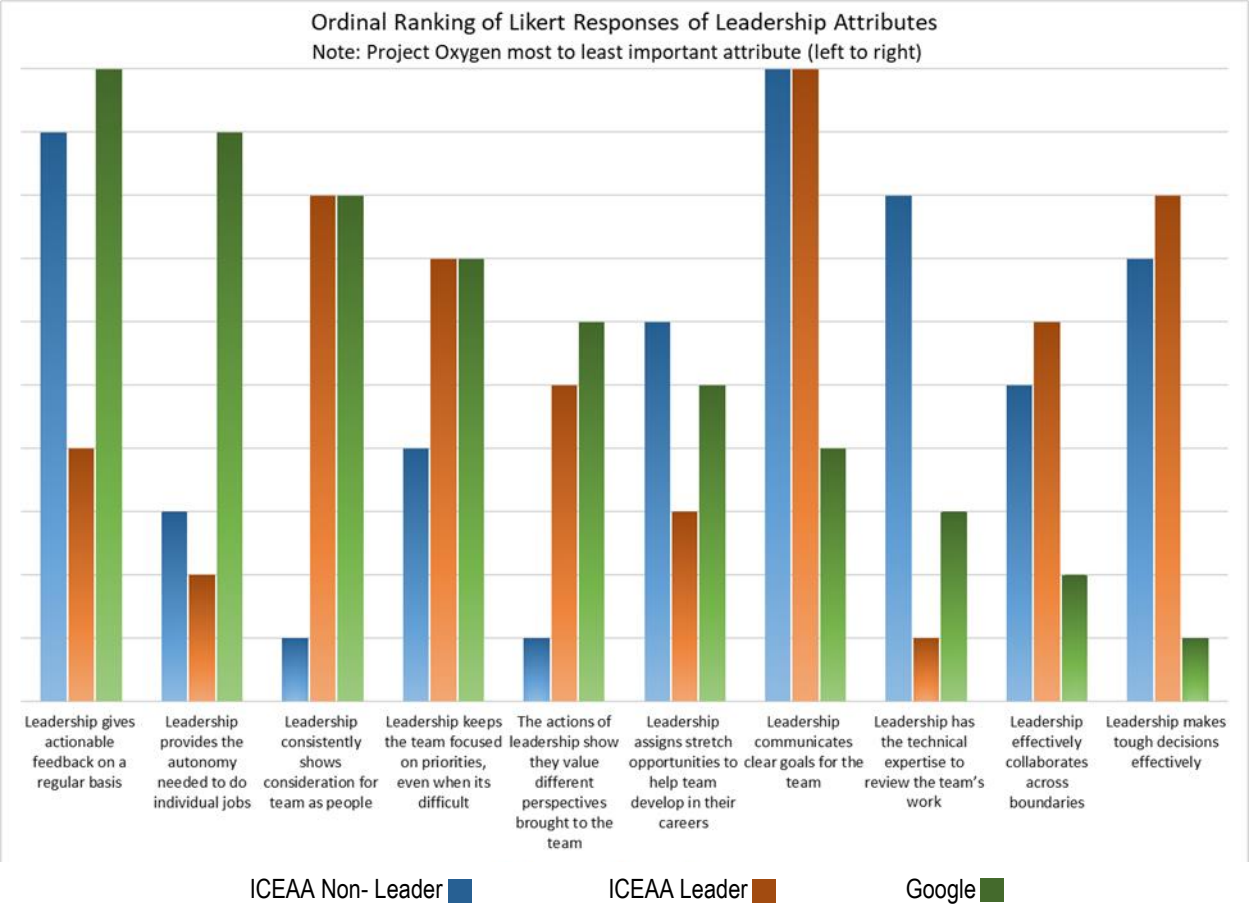


Figure 4 - Ranking of Leadership Attributes

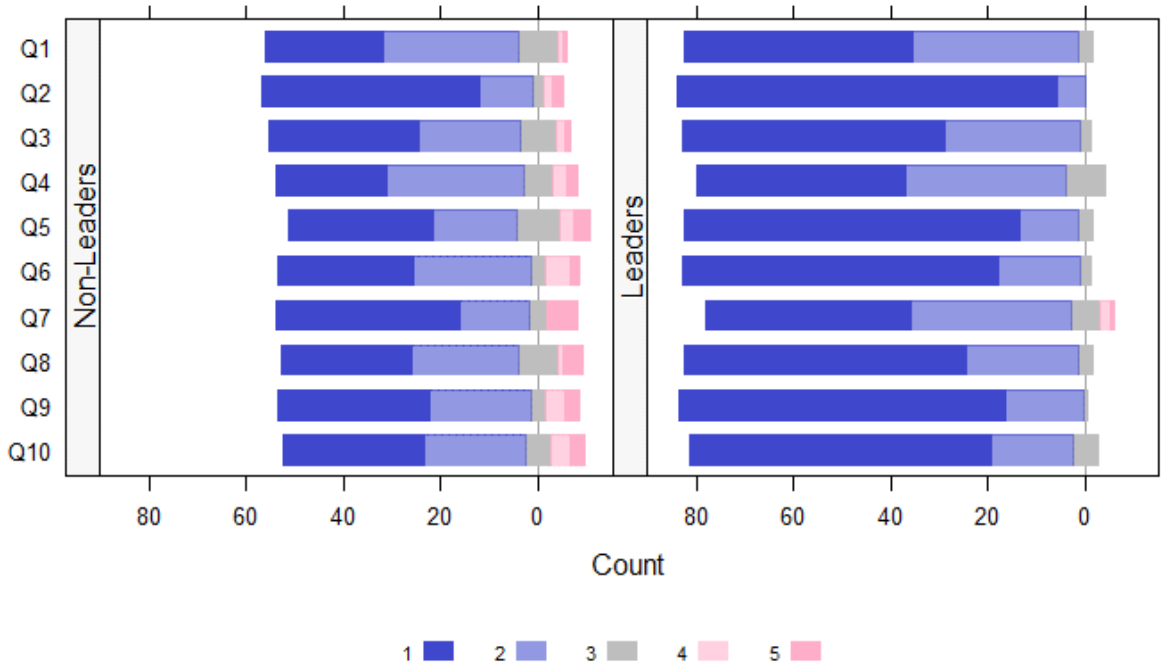


Figure 5 -Leadership Attribute Divergent Bar Chart

team as people and *the actions of leadership show they value different perspectives brought to the* team much more than the non-leaders do. Analyzing the Likert scale responses to the attributes using a divergent bar chart (Figure 5), it is apparent the only skill that leaders believed was less than neutral was *Leadership has the technical expertise to review the team's work*. Leaders otherwise seemed more likely to rank skills as a important (1) while non-leaders were much more likely to give neutral or not important (5) responses.

To address what respondents indicated was the most important attribute, *communicating clear goals,* the guidance can come from what Google provides to train their own managers. In their own rework training site, "Google's high-scoring managers are clear, concise, and honest in their verbal and written communications. But being a good communicator also means being an effective listener. Google encourages managers to be available for their teams and to encourage open dialogue and honest feedback."[7] The results suggest the cost estimating community would benefit from creating and implementing training that teaches rising leaders best practices in written and verbal communication as well as active and effective listening skills. This training could likely also address how to best give actionable feedback.

Figure 6 shows how many respondents chose behaviors as one of the top five responses that they deemed most important when presented with all ten Google Manager Behaviors. The data has been normalized to account for the percentage of each type of survey respondent as more leaders responded to this cost community survey than non-leaders.

The three behaviors chosen most frequently were: *is a good communicator*, *empowers the team*, and *uses the technical skills to advise*. Self-identified leaders overwhelmingly chose *being a good communicator* within their top five most important behaviors. Non-leaders valued the ability of leadership to use their *technical skills to*

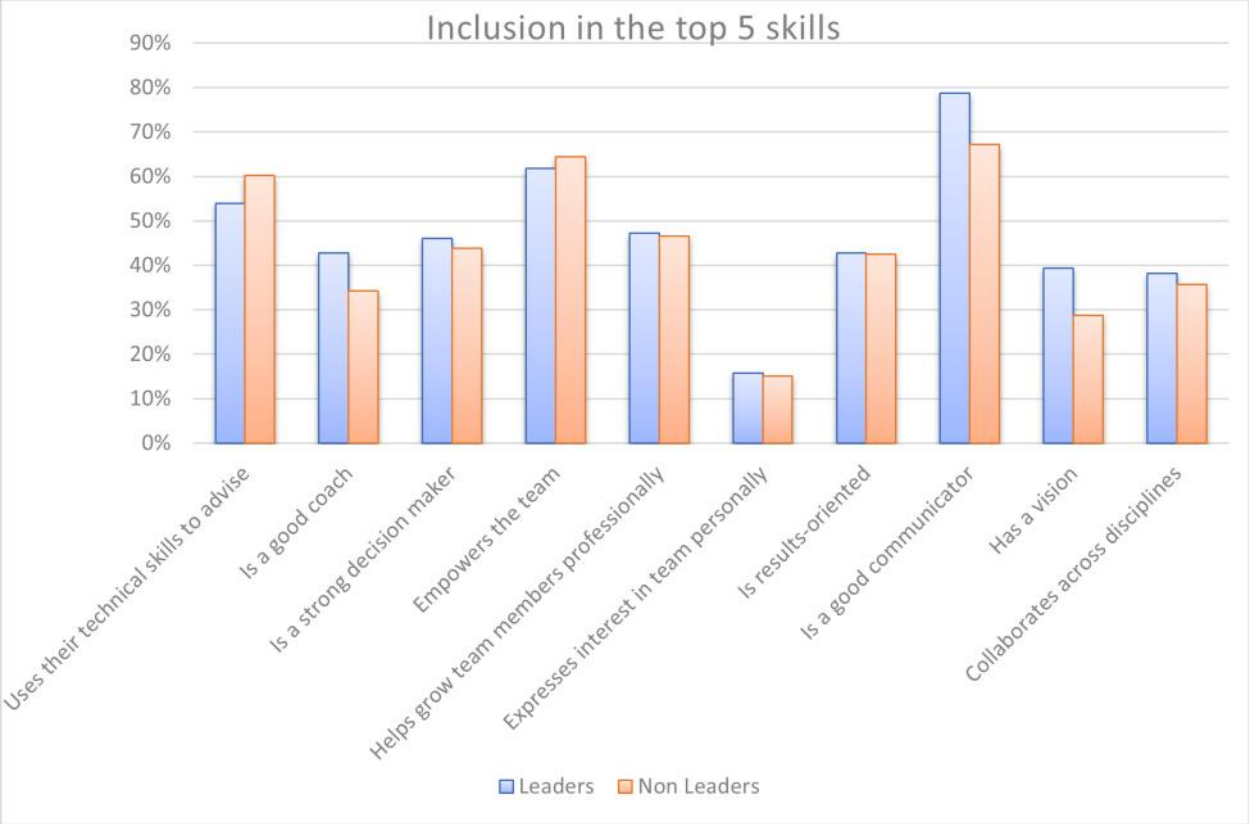


Figure 6 - Top 5 skills

advise and their ability to *empower the team* slightly more than the self-identified leaders, but otherwise the responses appear very similar between the two populations.

After selecting their top five important behaviors, they then are asked to narrow it down to the most important and least important skill. From this data, another interesting point to note is the different perceptions within the leader and non-leader groups. This difference is best highlighted by looking at the skills that have similar amounts of respondents that say it is the most important as those that say it is the least important. For examples with leaders, while ten chose *using their technical skills to advise* as the most important skill, nine believed it was the least important. Five leaders believe that *a strong decision maker* was the most important while five leaders felt it was the least important.

Non-leaders have noticeable discrepancies both in the importance of leadership being *a good coach* and being *results-oriented.* These discrepancies may be caused by different interpretations of the meaning of the behaviors and the bias of the respondent.

Overwhelmingly, the cost community believes that *expressing interest in the team personally* is the least important skill, followed by *having a vision* and *collaborating across disciplines*. For the next iteration of the survey *expressing interest in the team personally* should not be listed as a specific behavior on its own but included as an attribute into other behaviors like *being a good coach*.

Given the nature of this exploratory survey, it was also imperative to ask if respondents believed additional important behaviors that should be included the next time. These responses are shown in full in Appendix 1 and depicted visually

in Figure 8. Of the 163 respondents, 51 mentioned additional behaviors, skills, or attributes that should be considered in the next iteration of this survey; proving there is a lot of room for "soft skills" training at ICEAA's Workshops and within ICEAA's curricula. One notable comment shows the importance of this paper and hopefully value of potential ICEAA training courses: "Skills outside of Cost. Cost team leaders who are too narrow into the field are stifling to innovation. Must be progressive and willing to deviate from the "guides" and "training" which are beyond dated (or even wrong from the start)."

Limitations and Future Work

This paper scratched the surface of an element of cost estimating that previously received very little attention from the community. However, given that all 163 respondents agreed that a cost team lead's effectiveness has an impact on the cost products generated by the team, it seems that refining the work from this initial study and implementing the findings would be worth further endeavors. Ideally, future studies would include correlating cost product/team metrics to a team's rating of their cost team leader on aspects such as delivery time, accuracy, team productivity, product credibility, or team satisfaction. This addition would provide data similar to that produced by the Project Oxygen study, to prove that the team leader quantifiably affects the team and determine the behaviors that have the most positive impact. Realistically, all cost organizations and agencies could do their own internal experimentation using the Project Oxygen method to gain their own unique insights into their leadership. In the meantime, using the feedback from this exploratory survey, training could be developed that will result in more effective leadership and therefore improve the products delivered and team satisfaction.

Though ICEAA has over 1,000 cost estimators worldwide, this survey was delivered in English and filled out primarily by North American cost analysts. ICEAA's wide variety of international support could expand the reach of a future survey. Also, with over 1,000 members of ICEAA the survey participation rate was only between 10-15% of known cost estimators. Although this was a good response rate for the initial work, hopefully future work will have a greater participation.

Though the survey asked for years of cost experience, that does not necessarily reflect the respondent's age, especially in the lower numbers.

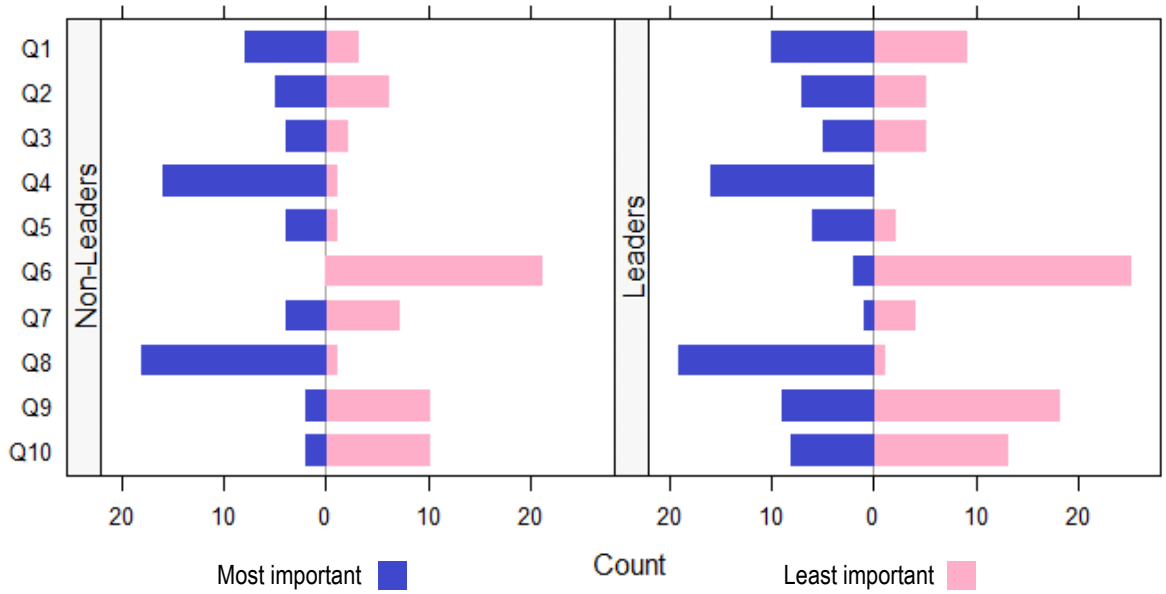


Figure 7 - Least and Most Important Skills

	Non-Leaders		Leaders	
	Most important	Least Important	Most important	Least Important
Uses their technical skills to advise	8	3	10	9
Is a good coach	5	6	7	5
Is a strong decision maker	4	2	5	5
Empowers the team	16	1	16	0
Helps grow team members professionally	4	1	6	2
Expresses interest in team personally	0	21	2	25
Is results-oriented	4	7	1	4
Is a good communicator	18	1	19	1
Has a vision	2	10	9	18
Collaborates across disciplines	2	10	8	13

If a future survey were also to ask the respondent's age, it could be determined whether younger cost estimators have the same expectations of their leadership as the more senior estimators. Similarly, "A survey by Virtuali found that 83% of millennials want fewer layers of management. This means they want managers who are easily approachable and willing to take their opinions into account."[4] Seeing if one's expectations of cost leadership is different based on age would be insightful when trying to lead and motivate a team with a diverse age set.

Figure 8 - Leadership Skills Word Cloud

Conclusion

In a survey of cost analysts, the community unanimously agreed that leadership is important to the end cost product. Even with consensus on the impact of the cost lead, the community at large does not provide guidance on or appear to value the behaviors that make an effective leader. Using the ten behaviors identified by Google's Project Oxygen, our study, even accounting for its discussed limitations, found the same results. "Soft skills" like being a good communicator, though proven to have a positive impact on team and leadership performance, have been largely overlooked by our profession. This paper established a foundation to determine the most important qualities of effective cost team leadership. While additional and more refined studies will be valuable, the responsibility is ours to now emphasize the importance of these skills and develop training/best practices for effective communication skills in leadership, a proven metric for cost team success.

Appendix 1 – Are there any skills not mentioned above that you think are important qualities of cost team leaders?

1. Telling the truth and doing the right thing whether it benefits you or not
2. Ability to handle conflict within the team and ability to guide client
3. Ability to include diverse perspectives
4. Ability to influence and participate in hiring of team
5. Ability to manage to scope and limit scope creep
6. Able to clearly define the requirements of the project
7. Act with personal integrity
8. An understanding of mathematics
9. Approachability. Humility. Brand new analysts have really good ideas, it's important that they have a voice. Also, important to demonstrate that it's OK to not always know the answer, and work together to find one. Being a "strong decider" often means, stubborn and not receptive, so I value the opposite and try to welcome input for the improvement of the team and our processes.
10. At our Cost Department, a CTL is the hardest job. Because we are matrixed to our program offices that have two bosses and a team handed to them. The Captain is demanding time, supervisors has requests and trainees need daily attention. Balancing every need is important and on the job training
11. Be supportive to your team members, make sure they know you have their back
12. Calmness under pressure
13. Candor and trust
14. a skill related to interfacing with the customer or translating things into requirements. I think that's a very important skill
15. Collaboration within the cost team
16. Combination of business and technical savvy
17. Creative problem solving, innovative, exceptional time management, ability to define scope and ask the hard questions of program/technical leadership, and a network of SMEs in all disciplines to provide reach-back support as necessary
18. Empathy
19. Empathy
20. Encourages open minded techniques that allow team members to think outside the box
21. Flexibility Resilience Political Savvy Influencing/ Negotiation Integrity/Character
22. Flexibility to adapt to changing circumstances. It is inevitable that the assumptions at the beginning of the estimate are changed and a leader needs to not get frustrated and keep the team from being frustrated or distracted
23. Has a backbone to stand up to pressure to change an estimate
24. Has a clear plan for achieving the team's goals
25. Having experience as a member of a cost team under multiple Cost Team Leads
26. Ownership. A team lead owns the team and the outcomes, both good and bad, of the team
27. Innovation, creative problem solving, critical thinking, curiosity
28. Integrity and Responsibility
29. Know your customer
30. Knows and balances the strengths and weaknesses of the team members
31. Leaders should treat the team with respect
32. Manages time well
33. Mentoring
34. Organized, approachable
35. Planning. The team lead needs to be able to backwards plan and work the plan in order to complete the task at quality an on time.
36. Positive attitude
37. project management
38. Providing top-cover and standing up for their team
39. Remain neutral
40. Sets individual goals for each team member.
41. Skills outside of Cost. Cost team leaders who are too narrow into the field are stifling to innovation. Must be progressive and willing to deviate from the "guides" and "training" which are beyond dated (or even wrong from the start)

42. Strong and productive relationship with the customer
43. Support team members when they fail
44. Teaches
45. team leaders need to know our business
46. Technical expertise is important, but the leadership ability through emotional intelligence is the most important.
47. The ability to effectively multi-task
48. The team leader needs to advocate for the cost team in organizations that do not place high importance or regard on the cost team. I would be happy to discuss in more detail what I have experienced in this regard
49. Transparency
50. Trusting the team to finish the activities and should be able to delegate
51. Well, the ability to communicate is mentioned, but should also include working with the customer to truly understand their goals. This could be part of the "vision" but feels a bit different

References

1. Garvin, D. A., Wagonfeld, A. B., & Kind, L. (2013). Google's Project Oxygen: Do Managers Matter? Harvard Business School publishing corporation.
2. Tran, S. K. (2017). GOOGLE: a reflection of culture, leader, and management. *International Journal of Corporate Social Responsibility*, *2*(1), 2–10. https://doi.org/10.1186/s40991-017-0021-0
3. Kalliamvakou, E., Bird, C., Zimmermann, T., Begel, A., DeLine, R., & German, D. M. (2019). What Makes a Great Manager of Software Engineers? *IEEE Transactions on Software Engineering*, *45*(1), 87–106. https://doi.org/10.1109/tse.2017.2768368
4. Impraise. (2020, October 14). Project Oxygen: 8 ways Google resuscitated management. Impraise. https://www.impraise.com/blog/project-oxygen-8-ways-google-resuscitated-management
5. re:Work. (2021). Re:Work. https://rework.withgoogle.com/
6. Google Got Rid of Its Bosses -- And Then Brought Them Back For These 10 Reasons. (2019, February 6). INC. https://www.inc.com/michael-schneider/google-didnt-always-appreciate-its-managers-now-it-relies-on-them-for-these-10-things.html
7. re:Work - Guide: Set and communicate a team vision. (2021). Re:Work. https://rework.withgoogle.com/guides/managers-set-and-communicate-a-team-vision/steps/introduction/
8. *Academy on Air: Project Oxygen: Why Managers Matter*. (2017, February 2). [Video]. YouTube. https://www.youtube.com/watch?v=JattR1uoX7g&t=922s

*For over 16 years, **Christina N. Snyder** has been a noteworthy leader in the field of cost estimating; supporting multiple DOD programs with various aspects of cost estimating, project management, and strategic planning. She received her CCE/A® certification in 2012 and is passionate about translating cost data into insights that resonate with decision makers. In addition to her daily work, she has been recognized for her extraordinary commitment to the ICEAA organization and is currently the Executive Vice President. Mrs. Snyder graduated from Virginia Tech in 2005 with a B.S. in Applied Computational Mathematics and currently resides outside of Tampa, Florida with her husband, 2 young sons, and basset hound.*

Parametric Joint Confidence Level Analysis: A Practical Cost and Schedule Risk Management Approach

Sara Jardine

Christian B. Smart, Ph.D., CCEA

Kimberly Roye, CCEA

Abstract: The use of Joint Confidence Level (JCL) analysis at NASA has proven to be a successful policy. Bottom-up resource-loaded schedules are the most common method for jointly analyzing cost and schedule risk. However, high-level parametrics and machine learning for JCL have been used successfully by one of the authors. This approach has some advantages over the more detailed method. In this paper, we discuss the use of parametrics and machine learning methods, especially as they apply to JCL analysis. The parametric and machine learning approach involves the development of mathematical models for cost and schedule risk. Parametric methods for cost typically use linear and nonlinear regression analysis. These methods applied to schedule often do not provide the high R-squared values seen in cost models. We discuss the application of machine learning models, such as regression trees, to develop higher-fidelity schedule models. We then introduce a bivariate model to combine the results of the cost and schedule risk analyses, along with correlation, to create a JCL using models for cost and schedule as inputs. We provide a previous case study of the successful use of this approach for a completed spacecraft mission and apply the approach to a large data set of cost, schedule, and technical information for software projects.

Background

For over fifty years, the cost analysis community has applied uncertainty analysis methods using univariate probability theory in risk analysis to generate separate distributions of a program's estimated cost and schedule (*Garvey*, 2000). In the schedule analysis and broader project management professional communities, the use of the schedule risk analysis has also been around for even longer and dates back to the Project Evaluation and Review Technique (*Hulett*, 2009). The interdependency between cost and schedule has long been recognized, but NASA is one of the few government agencies that has established official policy to conduct integrated cost and schedule risk analysis, which they call "joint confidence level analysis." We will use the term **joint confidence level** and its common abbreviation **JCL** throughout this paper.

The use of joint cost and schedule risk analysis has largely been limited to resource-loaded schedule analyses. While providing a great deal of insight into a project, resource-loaded schedules are labor-intensive. They also suffer from a drawback common to most bottom-up methods, which is the underestimation of the true amount of cost and schedule risk for a program. Parametric models can be developed much quicker and can provide a more comprehensive picture of program risk. Despite the development of such methods more than 20 years ago (*Garvey*, 2000), little has been adopted from multivariate theory to combine or develop conditional cost and schedule probability distributions to present to decision-makers.

Introduction

This paper reintroduces the top-down parametric approach to conducting JCL analysis. This technique is less cumbersome yet just as accurate

in the quantification of top-level cost and schedule risk as the familiar bottom-up resource loaded JCL method. We enhance the practice of the top-down parametric method with the consideration of machine learning techniques in addition to the use of traditional parametric regression analysis. We introduce the application of optimization methods to develop Cost Estimating Relationships (CER). We present regression trees as a means to develop better Schedule Estimating Relationships (SER), since it is more difficult to use traditional regression methods to derive meaningful trendlines using historical schedule data. Using the results of the individual cost and schedule analysis, uncertainty analysis is applied separately to compute the means and variances, which are used to specify the parameters of a bivariate probability model for a given program. Dr. Christian Smart has developed a standalone MS Excel spreadsheet to compute a bivariate probability model. Using the means and variances from the Cost Risk Analysis (CRA) and Schedule Risk Analysis (SRA) along with the program's target budget and schedule values, the calculator will produce the JCL and associated iso-curves at various joint confidence levels.

In this paper, the following topics are addressed:

- Benefits of JCL within Project Management
- JCL Methods: Bottoms-Up and Top-Down Parametric
- Parametric Machine Learning Techniques: Optimization and Regression Trees
- Top-Down Parametric Method Case Study: NASA MAVEN spacecraft program

In summary, this paper highlights the benefits of JCL analysis and offers a quicker top-down parametric JCL method to be used by the cost community. The JCL provides a more holistic view of uncertainty so that decision-makers can make more informed decisions. We provide a comparison of the top-down and more well-known bottom-up JCL approaches, provide an in-depth process for the top-down JCL method using a software program example, and demonstrate a real-life successful NASA spacecraft program that used the top-down parametric JCL approach.

Joint Confidence Level Benefits to Risk Management

Projects of all types frequently experience cost growth and schedule delays. Projects that do not suffer from one or both maladies are the rare exception, rather than rule. In addition to being common, these phenomena are often extreme, especially for cost. Indeed, the cost for approximately 1 in 6 defense and NASA missions doubles or more from the initial plan to the final actual. Defense and NASA projects are comparable to other industries, as shown in Table 1. These issues are long-standing and have shown no signs of improving over the last several decades.

	Olympics	Software/ IT	Dams	NASA/ DoD	Rail	Bridges/ Tunnels	Roads
Average Cost Growth	156%	43–56%	24–96%	52%	45%	34%	20%
Frequency of Occurrence	10/10	8/10	8/10	8/10	9/10	9/10	9/10
Frequency of Doubling	1 in 2	1 in 4	1 in 5	1 in 6	1 in 12	1 in 12	1 in 50
Average Schedule Delay	0%	63–84%	27–44%	27–52%	45%	23%	38%
Frequency of Schedule Delay	0/10	9/10	7/10	9/10	8/10	7/10	7/10

Table 1. Comparison of Cost Growth and Schedule Delays Across Several Industries. (Source: Solving for Project Risk Management, Christian Smart, McGraw-Hill, 2020).

The extent and the frequency of cost increases and schedule slips is prima facie evidence that these programs have a significant amount of resource risk and that this risk has not been managed well. The resource risks for these projects have also not been analyzed with accuracy, as exhibited by the track record for cost and schedule risk analysis. For cost analysis, see Table 2 for a comparison of the 90% confidence levels (90th percentile of the CDF or S-curve) with the actual costs.

Project	Cost Growth	Ratio of Actual Cost to 90% Confidence Level
1	0%	0.6
2	19%	1.1
3	31%	1.0
4	32%	1.1
5	greater than 45%	greater than 1.0
6	52%	1.5
7	84%	1.7
8	93%	1.6
9	121%	2.0
10	280%	2.2

Table 2. Cost Growth and Ratio of Actual Cost to 90% Confidence Level for 10 Historical Projects (Source: Solving for Project Risk Management, Christian Smart, McGraw-Hill, 2020).

The projects in Table 2 are from a variety of applications. JCLs were conducted for at least two of the missions. For 5 of the 10 missions, the actual cost was at least one and a half times the 90% confidence level, and for 2 it was double or more. Two of the missions listed in the table were cancelled. If they had not been cancelled, the cost growth would have been higher. The term "90% confidence level" for these analyses is grossly erroneous. Even so, 90% confidence levels should have been high enough to capture these variations. However, the actual cost was greater than the 90% confidence level for 8 of the 10 projects. This dismal result is even worse than it appears. A more in depth discussion for projects one and five is provided below.

- **Project 1.** One of the authors conducted a cost and schedule risk analyses using the top-down parametric method for project 1, which was a relatively rare mission that did not experience cost growth. The estimate of the 50% confidence level was within 1% of the actual cost. The project also completed on time, in line with the 50% confidence level for schedule. This kind of outcome is the exception rather than the rule. As can be seen from the table, all the other missions experienced significant cost growth. This provides evidence that the parametric JCL approach may be better at capturing the full extent of resource risk.
- **Project 5.** This project experienced such significant growth from one phase to the next that it exceeded the 90% confidence level well before completion.

The National Aeronautics and Space Administration (NASA) is one of the few government agencies that requires a JCL analysis be conducted for programs and projects. A JCL analysis is a process that combines a program or project's cost, schedule, and risk into an integrated picture. It represents the probability that a program cost will be equal to or less than the targeted cost, and that the schedule will be equal to or less than the targeted finish date. According to the most recent NASA JCL policy, by providing a confidence level that integrates cost and schedule, the JCL helps inform management of the likelihood of a program's programmatic success. Implementing JCL requirements for NASA programs has proven to be an effective forcing function to help program managers integrate stove-piped work products such as an Integrated Master Schedule (IMS), resource management, and risk management (*NASA JCL Requirements Update Memo*, 2019).

A program manager's decision space encompasses cost, schedule, and performance of a program. Risk analysis is needed when the expectations in any of these domains limit what is feasible. Therefore, managing risk is to manage the conflicts that exists within each domain and interdependencies across all three (*Garvey*, 1993). Generating a joint probability distribution supports the estimation of a program's cost and schedule, which simultaneously have a specified probability of not being exceeded. Because it is a more stringent requirement, the JCL is almost always higher than either the cost or schedule confidence level when developed separately. The JCL provides program managers with an assessment of the likelihood of achieving a budget for a given schedule, which aids the creation and management of credible project plans. Depending on the agency's JCL goal, the amount of cost reserves and additional schedule

can be determined and provided to decision-makers. Project management can then more effectively manage scope, cost reserves and schedule reserves of the project to mitigate the risk.

Joint Confidence Level Methods

There are two proven processes to calculate a JCL: the bottom-up resource-loaded schedule method and the top-down parametric method. Although the intention of this paper is to encourage the use of the top-down parametric as a more practical approach in the cost estimating field, we will briefly discuss the bottom-up method for the purpose of comparing it to the top-down method.

Bottom-Up Method

The bottom-up JCL method starts with a robust cost estimate and is mapped to a resource-loaded Integrated Master Schedule (IMS). A risk list is incorporated in the joint cost and schedule model at the lowest WBS element level and schedule and cost uncertainty is assigned. Although the bottom-up method is popular and can successfully calculate a JCL, it has its disadvantages.

Shortcomings of the bottom-up JCL approach include being resource intensive and time-consuming. As with any bottom-up estimating approach, it is easier to inadvertently miss the accounting for uncertainty of lower-level risk elements and thus, underestimate risk of the overall program. It is also difficult to justify uncertainty probability distributions on lower-level elements since data is scarcer and is typically not available at a low level. The bottom-up method also ignores unknown-unknowns, which are largely covered in the historical parametric data used in the top-down approach. While unknown-unknowns cannot be predicted in advanced, their existence in the aggregate can be used in the quantification of cost and schedule risk with just as much confidence as actuaries place in the quantification of insurance risk (Augustine 1983). While they are impossible to predict in advance, they dominate the bulk of cost and schedule risk, so their inclusion is imperative in conducting realistic risk assessments. The inclusion of unknown-unknowns is largely captured by the standard error and prediction intervals derived from the parametric cost and schedule equations.

The 2014 Joint Agency Cost Schedule Risk and Uncertainty Handbook (JA CSRUH) highlights the Fully Integrated Cost and Schedule Method (FICSM) as a bottom-up JCL approach. To provide a general understanding of the time-intensive bottom-up process, the FICSM approach is illustrated in Figure 1 below. This method can be applied using Joint Analysis of Cost and Schedule (JACS) in the ACEIT software suite and MS Project.

Top-Down Method

The top-down parametric JCL approach is less resource intensive than the bottom-up approach.

Figure 1. FICSM Process (Source: Joint Agency Cost Schedule Risk and Uncertainty Handbook 2014).

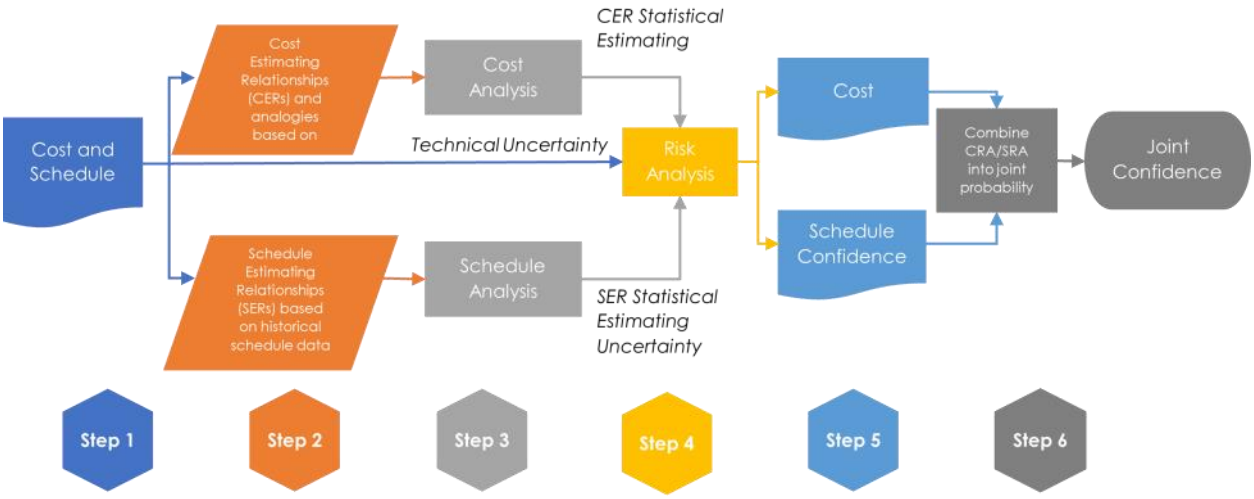


Figure 2. Top-Down Parametric JCL Process.

The reference to understand and explain the top-down parametric JCL approach was adopted from "*A Family of Joint Probability Models for Cost and Schedule Uncertainties*" (*Garvey*, 1993). To begin the discussion, an illustration of the top-down parametric process is illustrated in Figure 2.

A description for each of the six steps will be provided, while a more in-depth approach will be discussed for Step 2, where cost and schedule analyses are developed independently. During this step, if traditional parametric regression approaches do not result in any viable statistically significant estimating relationships, machine learning techniques can be used to predict estimating relationships. Throughout the steps, we will use a hypothetical software program example to demonstrate the top-down parametric JCL process.

Step 1: Cost and Schedule Data Collection. To begin, the analyst should collect a schedule and cost dataset separately that meets the criteria for performing parametric analysis to test the statistical significance of a cost and schedule estimating relationship. Data collection for the dataset would include historical analogous programs.

In the software program example, the cost dataset included hours as the dependent variable and peak staff and Equivalent Source Lines of Code (ESLOC) as the independent variables. The schedule dataset included duration in months as the dependent variable and potential schedule drivers such as new code, peak staff, and total development hours.

If data are not available, there are a variety of off-the-shelf parametric estimating tools that can be used including SEER-H, SEER-SEM, and SEER-Space.

Step 2: Cost and Schedule Regression Analysis. Perform regression analysis on the cost and schedule datasets separately using linear and nonlinear models. Test the statistical significance of regression equations and determine if any viable regression equations result. Different statistical software tools can be used to perform regression analysis during this step, including MS Excel, CO$TAT, or JMP. If traditional regression analysis does not result in any CERs or SERs, machine learning techniques should be considered.

Parametric techniques are within the scope of machine learning and can be applied to determine relationships between cost and

schedule and their drivers. These machine learning techniques include optimization to produce the "best" coefficients for a regression equation and regression trees. We introduce the discussion of regression trees in parametric estimating of schedules due to the fact that SERs are more difficult to estimate using traditional regression methods. The range of schedules typically has a smaller spread than cost, making trendlines less statistically significant. However, program technical data often includes a considerable amount of categorical data, which lends itself well to the use of regression trees. In a later section of this paper, we will provide a more in-depth discussion on the use optimization and regression trees for Step 2 of the top-down parametric approach.

In the software program example, optimization was applied using MS Excel Solver to develop a CER where peak staff and ESLOC were the independent variables driving hours. Since the example software program dataset was large (e.g., more than 50 data points), Maximum Likelihood Estimation Regression for Log Normal Error (MRLN or "Merlin") regression method was used (Smart 2017). MRLN will be further discussed in the next section to demonstrate how to apply optimization to determine the optimal coefficients, $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}, \boldsymbol{\beta_2}$, for the regression equation. With a Pearson's R^2 equal to 74%, the resulting CER had the following nonlinear power equation:

$$Total\ Hours = \boldsymbol{\beta_0}(Peak\ Staff)^{\beta_1}(ESLOC)^{\beta_2}$$

In the software program example, the schedule dataset did not result in any statistically significant SERs. With a significant amount of categorical data such as development process type (e.g., waterfall, incremental, agile, evolutionary, etc.), operating environment, and application domain, a regression tree with a Pearson's R^2 equal to 50% was developed using the *R* statistical programming platform.

Step 3: Cost and Schedule Analysis. This step represents the parametric results of the cost and schedule analyses developed in Step 2.

Step 4: Cost and Schedule Risk Analysis. Conduct a cost and schedule risk analysis on the cost and schedule estimate results, respectively. To achieve this step, a brief discussion of how to apply uncertainty analysis to regression equations is necessary. Regression equations have two forms of uncertainty that need to be accounted for: **input** and **estimating**.

Input uncertainty represents variability in the independent variables in a CER/SER regression equation. One approach to computing input uncertainty, *X*, is to assume a triangular distribution on input variables and run low (*L*), most likely (*ML*), and high (*H*) values through the CER/SER to obtain *L*, *ML*, and *H* estimates. Calculate the mean, $\boldsymbol{\mu_x}$, and standard deviation, $\boldsymbol{\sigma_x}$, of the triangular distribution. The calculations of the mean and standard deviation are:

$$\boldsymbol{\mu_x} = \frac{L+ML+H}{3}$$

$$\boldsymbol{\sigma_x} = \sqrt{\frac{L^2 + ML^2 + H^2 - L*ML - L*H - ML*H}{18}} \quad (1)$$

Estimating uncertainty is inherent to regression equations because, regardless of the parametric method used, even if the independent variables are known precisely, the CER/SER equation will return a result that is not certain. The error of the regression equation scales with the CER/SER result, making multiplicative error terms the preferred approach to modeling CER/SER estimating uncertainty. Regression estimating uncertainty represents uncertainty about the estimate's residual ε, (e.g., $Y = aX^b\varepsilon$). The farther the input variable is from the center of mass data used to derive the CER/SER, the greater the uncertainty of the CER/SER. The prediction interval or standard error provided by the

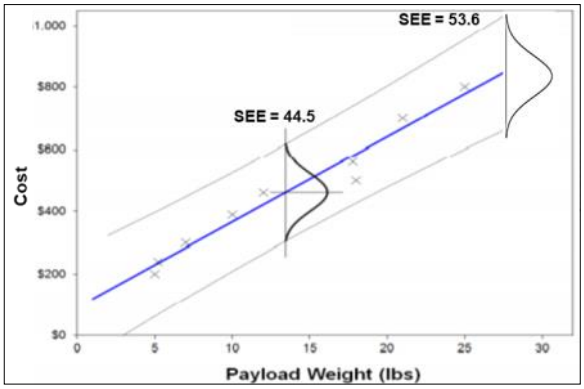


Figure 3. CER/SER Uncertainty Bounds (Source: Joint Agency Cost Schedule Risk and Uncertainty Handbook, 2014).

regression analysis can be used to determine the CER/SER uncertainty bounds. The Standard Error of the Estimate (SEE) converts to a prediction interval to account for the distance of the estimate from the center of the CER/SER dataset. Figure 3 shows a CER example of cost as a function of weight where uncertainty increases (standard deviation gets larger) as the point estimate moves towards the data boundaries (*JA CSRUH*, 2014).

One approach to computing estimating uncertainty, *Y*, is to treat uncertainty as a lognormal distribution and calculate the mean and standard deviation. Compute the mean and standard deviation in log space and then convert the values to unit space. The formulas to convert the mean, $\boldsymbol{\mu_Y}$, and standard deviation, $\boldsymbol{\sigma_Y}$, from log to unit space are shown below:

$$\boldsymbol{\mu}_y = e^{\mu+\frac{1}{2}\sigma^2}$$

$$\boldsymbol{\sigma}_y = \sqrt{\left(e^{\sigma^2}-1\right)e^{2\mu+\sigma^2}} \qquad (2)$$

To compute the total uncertainty of the regression equation, input and estimating uncertainty can be combined using propagation of errors. Assuming input and estimating uncertainty are independent and the residuals are multiplicative, the total uncertainty is obtained by multiplying the means and standard deviations of the input and estimating uncertainty calculated in (1) and (2). The formulas to combine input, *X*, and estimating, *Y*, are shown below:

$$\boldsymbol{\mu}(X*Y) = \mu_X * \mu_Y$$

$$\boldsymbol{\sigma}(X*Y) = \sqrt{\sigma_x^2*\sigma_y^2+\sigma_x^2\mu_y^2+\sigma_y^2\mu_x^2} \qquad (3)$$

Figure 4 shows a graphical representation of combining input and estimating uncertainty of a CER/SER calculated in (3) referenced from the 2014 *JA CSRUH*.

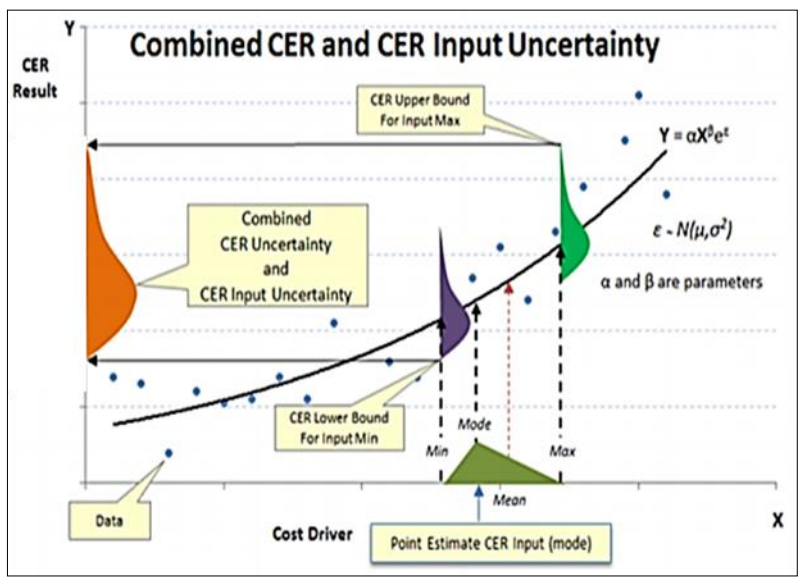


Figure 4. Combining Input and Estimating Uncertainty of a CER/SER. (Source: Joint Agency Cost Schedule Risk and Uncertainty Handbook, 2014).

If a regression tree was the method employed in Step 2 to derive a parametric relationship, as with the SER analysis done for the software program example, uncertainty analysis should be conducted on the regression tree. Input

uncertainty is modeled the same way as a regression equation using a triangular distribution on the input variables. To derive regression tree input and estimating uncertainty, assume they are independent. Depending on the data, the errors may be additive or multiplicative. In examining the data, we found the residuals best fit an additive model with a Gaussian/normal distribution. The input variables such as peak staff and software development hours are varied and simulated. For each trial, the simulation also sampled from the Gaussian for the regression tree residuals. For each of the 1,000 trials, the results from the varied input variables, and the estimation uncertainty from the residuals were added to yield total uncertainty.

Step 5: Cost and Schedule Confidence Levels. This step represents the results of the cost and schedule risk analyses (CRA/SRA) developed in Step 4. The results would reflect separate cumulative probability distributions or S-Curve results from the cost and schedule risk analyses.

Step: 6: Joint Confidence Level. The final step is to combine the CRA and SRA developed in Step 5 into a joint probability distribution to calculate the JCL. The reasons being are because they directly incorporate correlation between cost and schedule for programs and these distributions provide at least some probability of a cost or schedule overrun (lognormal distribution having a larger skew to the right while the normal distribution is not skewed). In accordance with Paul Garvey's method to combine cost and schedule as a joint probability model, we provide the following distributions to model the behavior of program cost and schedule: bivariate normal, bivariate lognormal, and bivariate normal-lognormal distributions. Figure 5 provides graphical depictions of a normal distribution.

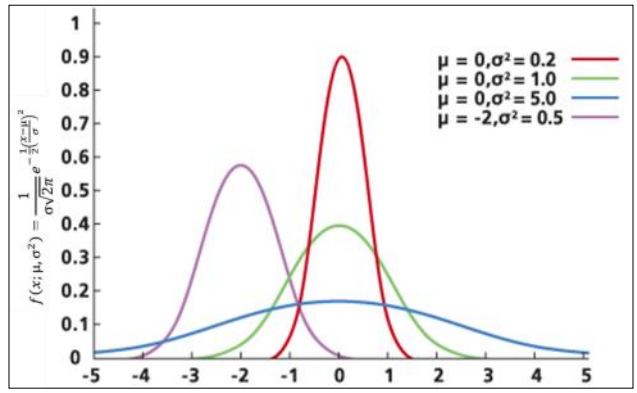


Figure 5. Gaussian/Normal Distribution.

In the authors' experience and results from the Air Force, Cost Risk Uncertainty Analysis Metrics Manual (CRUAMM), cost uncertainty is rarely normally distributed. When it comes to cost estimating the so-called normal is anything but normal! Cost estimating uncertainty is typically best modeled with a lognormal distribution. Schedule distribution uncertainty is typically lognormal, but in some instances, as in our example, the normal distribution is a good fit. The lognormal distribution is a skewed distribution. The lower bound is never less than zero meaning the cost and schedule cannot become negative and has an upper bound of infinity. The probability is skewed right providing at least some probability of a large cost or schedule overrun. These characteristics make the lognormal appealing for cost modeling and a best choice in the absence of better information (*JA CSRUH*, 2014). Figure 6 provides graphical depictions of lognormal distributions.

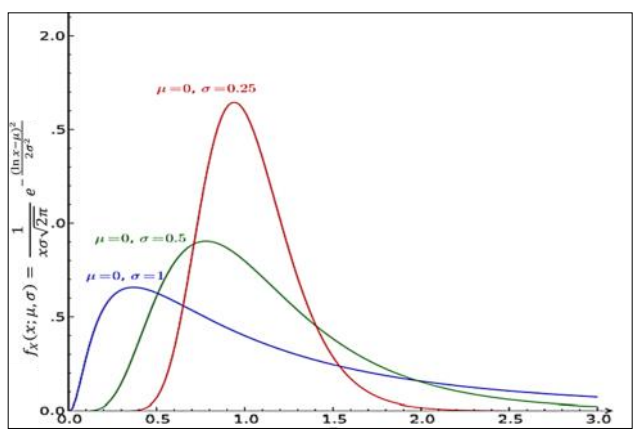


Figure 6. Lognormal Distribution.

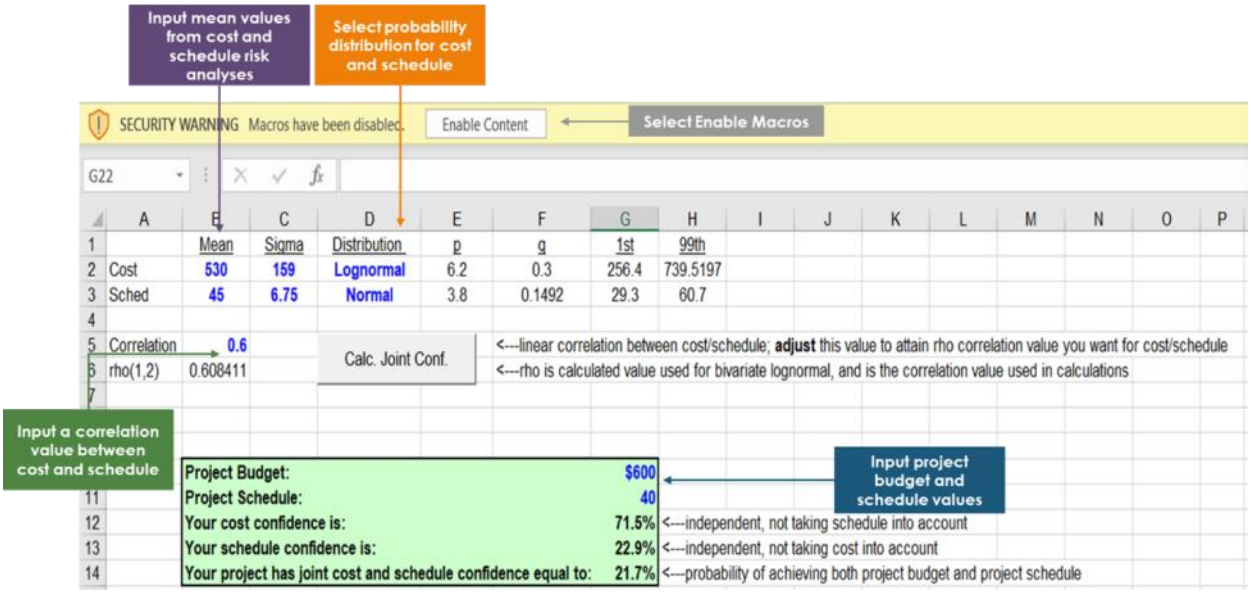


Figure 7. MS Excel JCL Calculator (Top-Down Parametric Method).

The bivariate normal-lognormal has two different marginal distributions. One marginal is normal, and the other is lognormal. Situations may arise when normal and lognormal distributions characterize a program's cost and schedule distributions.

To calculate a joint confidence, assume lognormal or normal risk distributions on cost and schedule using the mean and standard deviation as the parameters derived from the cost and schedule analyses in Step 5. Assume a positive linear correlation value between cost and schedule (e.g., correlation value equal to 0.6 or 0.7). Figure 7 is a screenshot of the MS Excel JCL calculator developed by one of the authors, Dr. Christian Smart, to take values derived from the top-down parametric approach and provide a JCL. The calculator uses a macro that numerically approximates the bivariate probability distribution, aka JCL, values.

In the notional example provided in Figure 7, the target budget of the given program is $600M and target schedule is 40 months. Using the results of the CRA and SRA, the mean and standard deviation is $530M and $159M respectively for cost, and 45 months and 6.75 months respectively for schedule. Cost is assumed to be lognormally distributed while schedule normally distributed. The correlation value between cost and schedule was selected to be 0.6. Based on the author's experience and data analysis, this is a reasonable value. The resulting JCL is 21.7%, meaning there is a 21.7% chance that the program cost will be equal to or less than $600M and that the schedule will be equal to or less than 40 months. If schedule was not

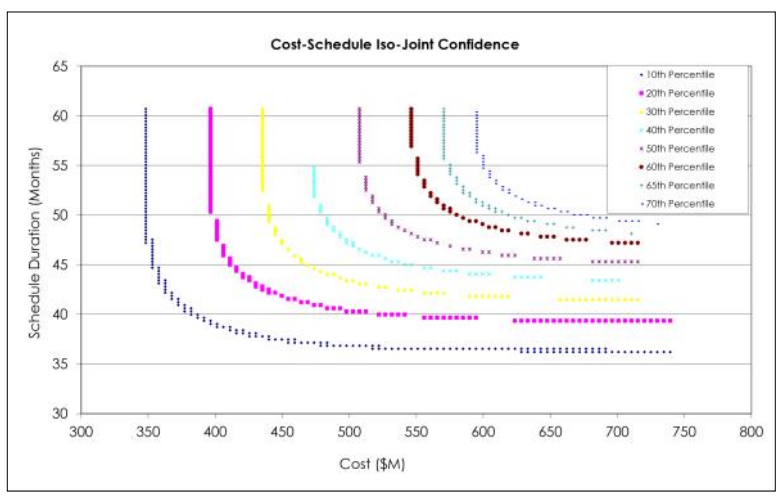


Figure 8. Example JCL Iso-Curves.

considered, the cost confidence level was 71.5% and if cost was not considered, the schedule confidence level was 22.9%.

Figure 8 shows the resulting JCL iso-curves with cost on the x-axis and schedule on the y-axis. As "iso" is a prefix meaning "equal", each cost and schedule joint iso-curve in the graph represents a specific JCL confidence level percentile. You can determine the current JCL level of your project by looking at the position of the budget and project planned schedule.

Figure 9 shows another example of a JCL output that would be presented to management.

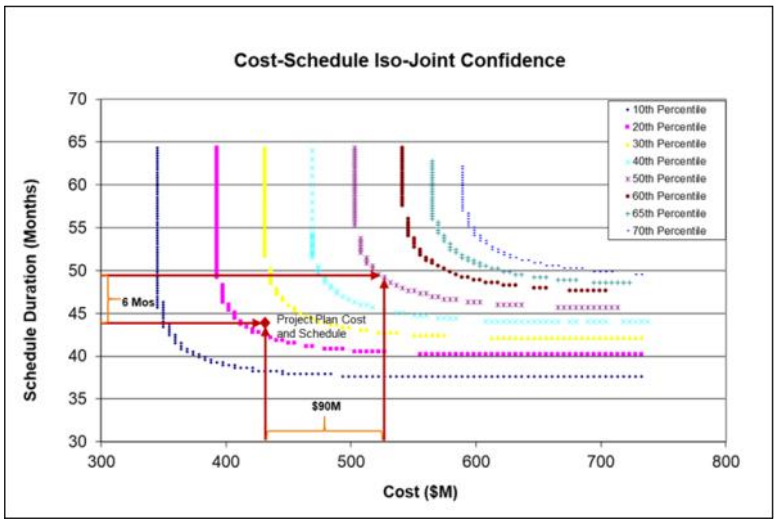


Figure 9. Example JCL Result.

The example program has a current project plan of a $435M budget and a 43-month schedule. Looking at the graph, this project plan position is at the 24% JCL. If management were interested in how much funding and schedule was needed to achieve the 50% JCL, you would look at the 50th percentile JCL iso-curve and see that an additional $90M and 6 months would be required. As there are multiple pairs of cost and schedule on each iso -curve, depending on the relative importance of schedule versus cost, an analyst can determine the amount of cost reserves and additional schedule duration needed to achieve the agency's JCL goal.

Machine Learning Techniques for Parametric Estimating

Machine learning methods can be a powerful mechanism to determine estimating relationships in a dataset when conducting the top-down parametric JCL. Machine learning is a collection of mathematical methods and computer algorithms for prediction and classification that represent a more modern way of conducting analysis on datasets that incorporates the use of computer programming with statistical analysis. Modern machine learning methods include decision trees, deep learning, and text analytics. As mentioned, machine learning techniques can be applied when developing a cost and schedule analysis during the top-down parametric JCL approach when traditional regression methods do not provide meaningful results, such as a regression equation with a low R^2, for example. For the purposes of this paper, we will focus on how to apply optimization and regression trees to develop cost and schedule estimating relationships when conducting a top-down parametric JCL analysis.

Optimization Technique for Parametric Estimating

Regression analysis as performed in Step 2 of the top-down JCL method, is a form of optimization. Optimization is a collection of mathematical principles and methods used for solving quantitative problems. The goal is to minimize or maximize a function in pursuit of finding the "best" solution. As previously mentioned, we will discuss the application of maximum likelihood as a regression approach to develop unbiased, optimal estimates of the mean when the errors are lognormally distributed. MRLN was developed by one of the authors, Dr. Christian Smart (*Smart*, 2017).

Let $a_1,\ldots, a_n$ represent the observed data and $x_1, \ldots,x_n$ represent random variables where a_i results from observing the random variable x_i. The likelihood function, which represents the likelihood of obtaining the sample data, is:

$$L(\theta) = \prod_{i=1}^{n} Pr(X_i = A_i|\theta)$$

The vector, θ, maximizes the likelihood function in the likelihood function. This consistent and efficient method is advantageous because maximizing the likelihood of finding the true underlying parameters of this distribution is exactly what we hope to accomplish in developing a CER. Other advantages of maximum likelihood are that it is always available to use, and it uses all the available data, where other methods such as percentile matching and method of moments do not.

Recall in the software program example, we estimated the following CER power equation model form:

$$Y = \beta_0 X_1^{\beta_1} \ldots X_p^{\beta_p}$$

The goal for MRLN is to maximize the function:

$$l(\beta_0, \beta_1, \ldots, \beta_p, \theta) = -\frac{n}{2}\, ln(\theta) - \frac{1}{2\,\theta}\sum_{i=1}^{n}\left(ln(y_i) - ln(\beta_0) - \sum_{j=1}^{p}\beta_j\, ln(X_{ij}) + \frac{\theta}{2}\right)^2$$

Using the MRLN method, MS Excel Solver can be used to find an optimal value in a cell. Decision variables are used to compute the formulas defined in the objective to converge on a solution that maximizes values for $\beta_0, \beta_1, \ldots, \beta_p$ to form the power equation, $Y = \beta_0 X_1^{\beta_1} \ldots X_p^{\beta_p}$. When using Excel Solver to optimize the coefficients in the software program using the MRLN regression method, recall that it resulted in the following CER with a Pearson's R^2 equal to 74%:

$$Total\ Hours = \boldsymbol{\beta_0}(Peak\ Staff)^{\beta_1}(ESLOC)^{\beta_2}$$

Regression Tree Technique for Parametric Estimating

Regression trees are an effective way to visualize the relationships between features within datasets, particularly when there is a large amount of categorical data such as historical schedule datasets. Regression trees can be used in preliminary data exploration to understand the most significant variables within a dataset. Regression trees can also be used to show the relationships within a dataset in Step 2 of the top-down JCL method when traditional regression analysis does not produce any good results. Pairwise analysis combined with regression trees can help shorten the time running regression models in search of significant relationships. Two of the authors, Kimberly Roye and Dr. Christian Smart, provided an overview of regression trees in a 2019 ICEAA presentation (*Roye and Smart,* 2019).

In a regression tree, the data are split into homogenous groups, and the graphs present splits with the use of branches (called decision nodes) and leaves (terminal nodes). The goal of a regression tree is to partition data into smaller regions where interactions are more manageable. They are useful when there is a non-linear and complex relationship between dependent and independent variables that cannot otherwise be represented by a regression equation. Figure 10 illustrates the structure of a regression tree.

The root node represents sample dataset that is being analyzed. The method asks its first yes or no question and splits the data into two groups based on the answer. The decision nodes represent the first set of homogenous groups discovered within the dataset. On the left, another yes or no question is asked, and the group splits into two nodes: one terminal and one decision node. The criterion for splitting is the choice that reduces the sum of squared errors by the biggest

amount. This process is recursively applied to each of the subsets produced until the reduction in error is smaller than a pre-specified limit, such as 1x10^-5. When a decision node be can split no further, the branch ends in a leaf, or terminal node. Each terminal node is a subset of the data set, and the estimate at each terminal node is the average of the data points in that subset.

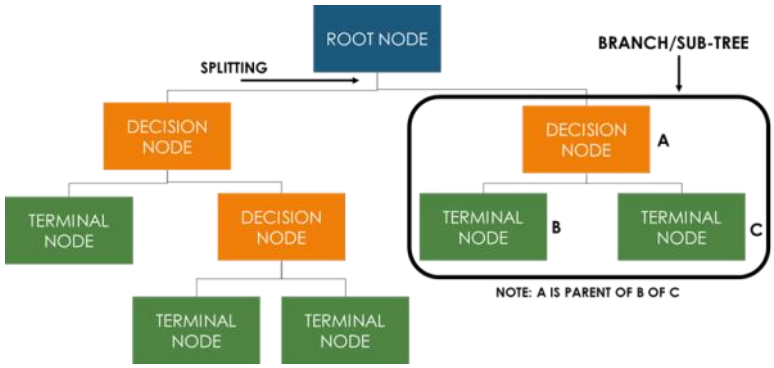


Figure 10. Regression Tree Layout.

In our software program example, since no SERs were significant in Step 2 of the top-down parametric approach, the schedule dataset was used to develop a regression tree using the *R* statistical programming platform. In the software program schedule dataset, total software development hours proved to be the most important factor. With a Pearson's R^2 approximately equal to 50%, Figure 11 shows an abbreviated version of the resulting schedule regression tree for the software program.

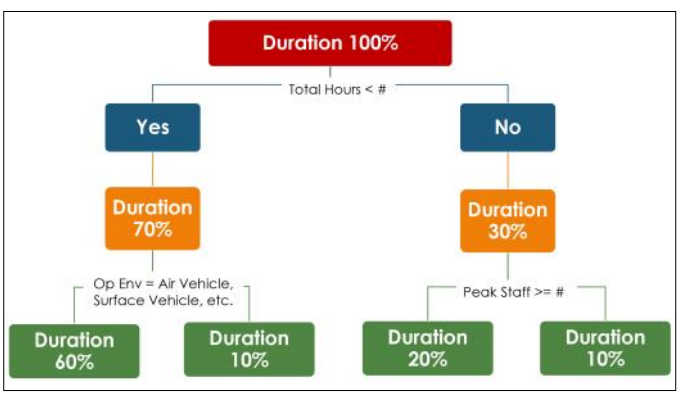


Figure 11. Software Program Example Regression Tree.

To explain this software program regression tree, we start with the total (100%) schedule dataset. Next, we ask if the total number of development hours is more than or less than a duration value. If the answer is yes, then the data is split into a branch to the left and if the answer is no, then the data is split into the branch on the right. In this example, 70% of the data satisfied the condition for number of hours and 30% did not. Total software development hours best minimize the squared error when estimating schedule duration. There are three decision nodes that ask questions about the value of total software development hours. Based on the value of total software development hours, we end at one of the terminal nodes of the tree. The estimate at the terminal, or leaf, node is the average duration of the subset included in that node. Each split is labeled with a condition and the branches between them are labeled with the average duration for that dataset split. An example interpretation of the first decision in the tree is, "if the total software development hours is more than 50,000 hours, my estimate is 60 months, otherwise it is 40 months." The regression tree produces a point estimate. Just like with traditional regression analysis, the regression tree uncertainty analysis is conducted by assessing the residuals, fitting a distribution, and combining this with parameter uncertainty, which provides an overall uncertainty distribution for the parametric schedule estimate.

For each node in the tree, the regression tree split is chosen by the algorithm to minimize the sum of squared errors. The algorithm chooses the variable and the associated value based on what reduces the sum of squared errors the most.

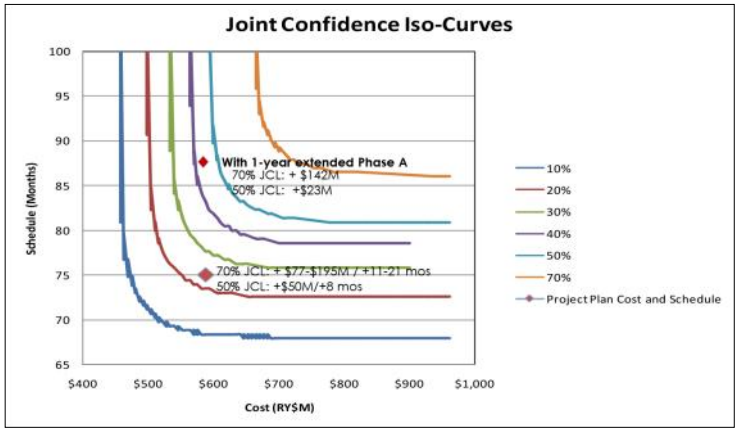


Figure 12. NASA Maven Spacecraft Program JCL Estimate.

Case Study

To advocate for and demonstrate the effectiveness of the top-down parametric approach, we highlight a success story for the NASA MAVEN spacecraft program. In 2009, one of the authors developed a JCL using the top-down parametric approach. At the time, with the project plan cost and schedule, the JCL was estimated at 23% and if a year was added to the development schedule, the JCL was estimated at 44%. With the current project plan, to achieve a JCL of 50%, an additional $50M and eight months would be needed, while to achieve a 70% JCL, an additional $77M-195M and 11-21 months would be needed. In 2013, the actuals for cost and schedule for the Maven program came in at the 50% JCL that was estimated in 2009. This is one of the few programs to show no cost growth, demonstrating an estimate that actually "hit the mark" when funded to the predicted 50% JCL. Table 3 summarizes the JCL results estimated in 2009 and the actual results in 2013.

	2009 (Estimate)			2013 (Actuals)
	Project Plan	50% JCL	70% JCL	
Cost ($M)	~$590	$640	$667-$785	$640
Schedule (months)	~75	83	86-96	83

Table 3. MAVEN Program JCL Results.

Figure 12 shows the iso-curves calculated for the MAVEN program.

MAVEN was the mission in Table 2 presented earlier for which the actual cost was below the 90th percentile of the cost risk analysis. While only one data point, it provides evidence that parametric JCLs can help ensure credible risk ranges.

Conclusion

NASA is the only known government agency that currently has a JCL policy. Very few organizations perform JCL analysis routinely as part of project management decision making. The more informed and holistic cost and schedule risk analysis results of a JCL should be considered by the cost analysis community and project managers when making decisions about programs.

Traditional bottom-up joint confidence level analysis can be cumbersome and resource intensive. This paper offers a proven top-down parametric JCL approach as a more manageable approach for cost analysts, while just as accurate as a bottom-up JCL approach based on the author's experience.

Machine learning techniques such as optimization and regression trees provide an analytical method to develop cost and schedule estimating relationships when traditional regression methods do not provide significant results.

References

Garvey, P.R., 1993, "A Family of Joint Probability Models for Cost and Schedule Uncertainties," presented at the 26th Annual Department of Defense Cost Analysis Symposium, September 1992, Williamsburg, VA.

Garvey, P.R., *Probability Methods for Cost Uncertainty Analysis: A Systems Engineering Perspective*, Marcel Dekker, New York, 2000.

Hulett, D., *Practical Schedule Risk Analysis*, Gower, Burlington, VT, 2009.

Naval Center for Cost Analysis, *Joint Agency Cost Schedule Risk and Uncertainty Handbook*, 2014, Washington, D.C.

Norman R. Augustine, *Augustine's Laws*, American Institute of Aeronautics and Astronautics, 1983, Reston, VA,1983, page 50.

Roye, K. and C.B. Smart, "Beyond Regression: Applying Machine Learning to Parametrics," presented at the 2019 ICEAA Professional Development & Training Workshop, Tampa, Florida, May 14-17, 2019.

Smart, C.B., "Maximum Likelihood Estimation for Regression of Log Normal Error," presented at the 2017 ICEAA Professional Development & Training Workshop, Portland Oregon, June 6-9, 2017.

Smart, C.B., *Solving for Project Risk Management: Understanding the Critical Role of Uncertainty in Project Management*, McGraw-Hill, New York, 2020.

Sara Jardine *is a Senior Cost Analyst for Galorath Federal with over 16 years of financial management experience. She has served a broad variety of federal agencies including the Army, Navy, OUSD AT&L, DAU, Veterans Affairs, and the Department of Homeland Security. She is skilled in Cost Analysis and course development, Project Management, Requirements Analysis, Contract Management, and Budget Management. Sara earned a MS in Project Management from George Washington University and a BS in Mathematics from the University of Michigan.*

Kimberly Roye *is a Senior Data Scientist for Galorath Federal. Starting her career as a Mathematical Statistician for the US Census Bureau, Kimberly transitioned to a career in Cost Analysis over 10 years ago. She has supported several Department of Defense hardware, software and vehicle programs, as well as NASA and the Department of Homeland Security (DHS). She is currently a lead developer of Machine Learning training for the Army and DHS. Kimberly earned a MS in Applied Statistics from Rochester Institute of Technology and a dual BS in Mathematics/Statistics from the University of Georgia.*

Dr. Christian Smart *is the Chief Data Scientist with Galorath Federal. He is author of the book Solving for Project Risk Management: Understanding the Critical Role of Uncertainty in Project Management. Dr. Smart is the VP for Professional Development with ICEAA. He regularly presents at conferences and has won several best paper awards. Dr. Smart received an Exceptional Public Service Medal from NASA in 2010 and has a PhD in Applied Mathematics.*

13 Reasons a Cost Estimate Could Go Wrong During a Concurrent Engineering Study (and How to Avoid Them)

Andy Braukhane

Abstract: During early phase spacecraft design, the concurrent engineering (CE) approach is proven to be very efficient. But the compressed and iterative nature of CE sessions can make life difficult for a cost estimator due to immature data, many design changes, and an intense workflow, among other issues. This work discusses 13 problem areas that have been encountered or observed mainly during one-week-long, interdisciplinary space system design studies at the German Aerospace Center. It provides practical examples on how to tackle them, e.g. how to deal with rapid data changes, false expectations and a heterogeneous engineering team.

Introduction

Concurrent Engineering (CE) is an efficient Systems Engineering approach which is increasingly applied in early phase spacecraft (S/C) design due to the involvement of all relevant disciplines, including the customer, and is often supported by data models and tools as well as by a communication fostering infrastructure.

During several moderated sessions, the latest results and problems are shared with the entire team, which supports the convergence towards a common solution. This exchanged information is a key input for the cost estimator and provides guidance on what to further discuss, to research, or how cost models should be used or adapted. But the data is constantly changing due to the iterative approach. Moreover, the space sector is not famous for public data, making research and comparisons often difficult. With predominantly technical people in the room, the cost estimate may also be perceived disconnected.

Based on the study context, managers expect either rough order of magnitude (ROM) cost or a detailed estimate following an elaborated work breakdown structure (WBS). These and other reasons why cost estimation could go into the wrong direction are discussed within the paper, based on experience and observations related to systems, concurrent and cost engineering. It includes real-world examples, ideas for solutions and some anecdotes which shall round off this lesson learnt compilation.

This paper has been prepared to discuss and raise awareness about particularly significant stumbling stones which can be encountered during otherwise very efficient and recommendable Concurrent Engineering activities for space missions and systems in the early phase (here so-called CE studies). We use the German Aerospace Center (Deutsches Zentrum für Luft- und Raumfahrt) Concurrent Engineering (DLR CE) approach as our example.

The potential problems mentioned are not exclusively applicable for CE, nor for the cost domain. As for CE studies in general, the approach for Cost Engineering in such an environment varies amongst different institutions. This relates to tools, data available, time available and likely even the objectives and expected outcomes.

This work is based on experiences and observations gathered during several DLR CE studies, during which a particular approach is applied, but also common rules and practices are followed. Please note that throughout the entire paper, the term CE is exclusively used for Concurrent Engineering and not for Cost Engineering.

Concurrent Engineering at DLR

The German Aerospace Center (DLR) is the national aeronautics and space research center. It performs extensive research and development (R&D) activities related to aeronautics, space, energy, transport, security and digitalization. Furthermore, DLR contains the German Space Administration, acting on behalf of the Federal Government, which is responsible for the implementation of Germany's Space Program, on national and international level. In 2007 the Institute of Space Systems was inaugurated within the Space R&D branch, with the objective to perform analysis, design, development, testing, integration and management of space systems, including e.g. satellites, probes, habitats and launch vehicles.

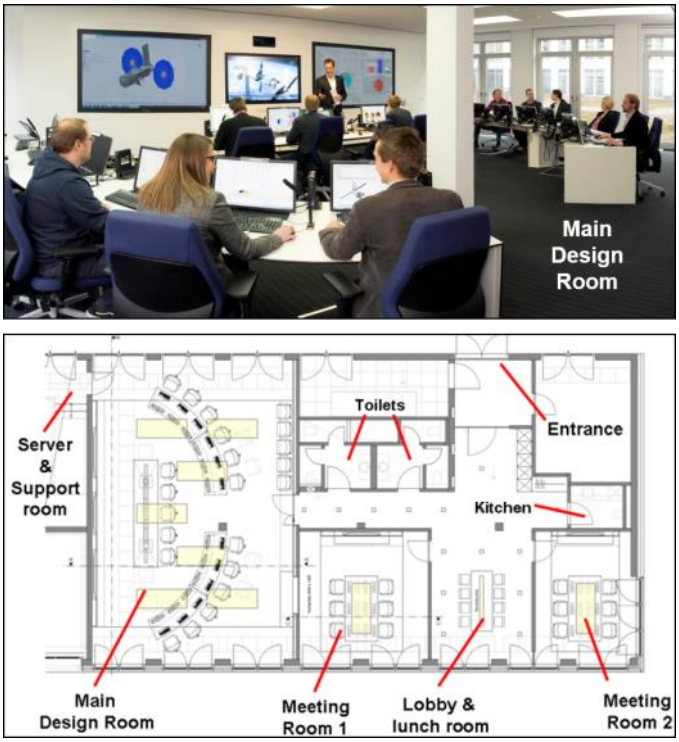


Figure 1: DLR Concurrent Engineering Facility (CEF)

In order to conduct efficient feasibility and preliminary design studies for internal and external space missions and systems, the DLR Concurrent Engineering Facility (CEF) has been established as part of the Institute build-up [1]. It is shown in Figure 1.

According to a definition from the European Space Agency (ESA), Concurrent Engineering is a systematic approach to integrated product development that emphasizes the response to customer expectations. It embodies team values of co-operation, trust and sharing in such a manner that decision-making is by consensus, involving all perspectives in parallel, from the beginning of the product life-cycle [2].

The major elements of CE, as it is applied in the space sector, are a guided and structured process, an infrastructure which fosters communication and collaborative working, a central data model to enable instant and simultaneous data exchange, as well as a team representing all relevant disciplines, including the customer [3].

CE in space has been applied already in the U.S. for more than 20 years, initially by the Aerospace Corporation and NASA's Team-X. ESA also implemented this approach in 1998. It clearly proved the efficiency and high quality for early space system and mission studies. Nowadays, many international organizations apply CE in one way or another as part of their Systems Engineering activities. These organizations include agencies (e.g. NASA, ESA, DLR), system integrators (e.g. Airbus), private or governmental organizations (e.g. Aerospace Corporation, NRO) and universities (e.g. Utah State University, ISU Strasbourg) [3]. More details on the general CE approach and existing facilities can be found for example in [1], [2] and [3].

With initial support of the ESA Concurrent Design Facility team, the DLR CEF adapted the Concurrent Engineering process and all related elements such as the actual infrastructure, required data models and software tools, and also the team (regarding size and compilation) to their own needs. With currently more than 70 studies completed, the CE process is already well established, but also continuously improving due to the on-going challenges of new customers,

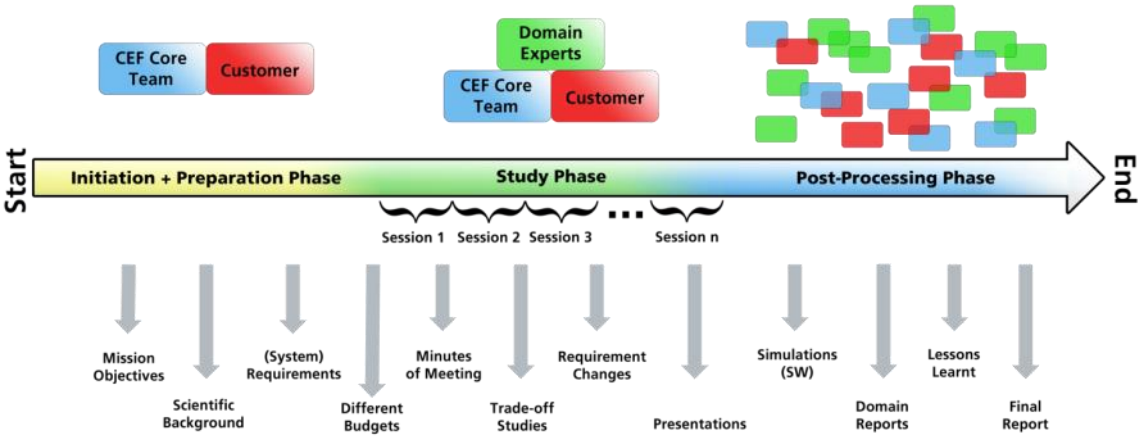


Figure 2: DLR Concurrent Engineering study, overall timeline

study topics, support technologies or team members.

Whereas the overall study timeline including initiation, preparation and also post-processing phases can last several weeks, the actual CE study phase at DLR typically lasts one full week [4]. Figure 2 shows the overall timeline including the different parties involved and information products generated. Figure 3 presents a typical schedule for the actual one-week study phase. In this phase, there is a mixture of moderated (indicated in red) and non-moderated sessions (blue), in which either general and system-relevant or more specific trades and tasks are carried out.

As a different example, ESA organizes their sessions over several weeks with only one or two moderated sessions per week [2], while Team-X at NASA Jet Propulsion Laboratory compresses all study sessions into less than one week, as indicated amongst others in [5].

A common set of domains and their representatives covers the moderator, the

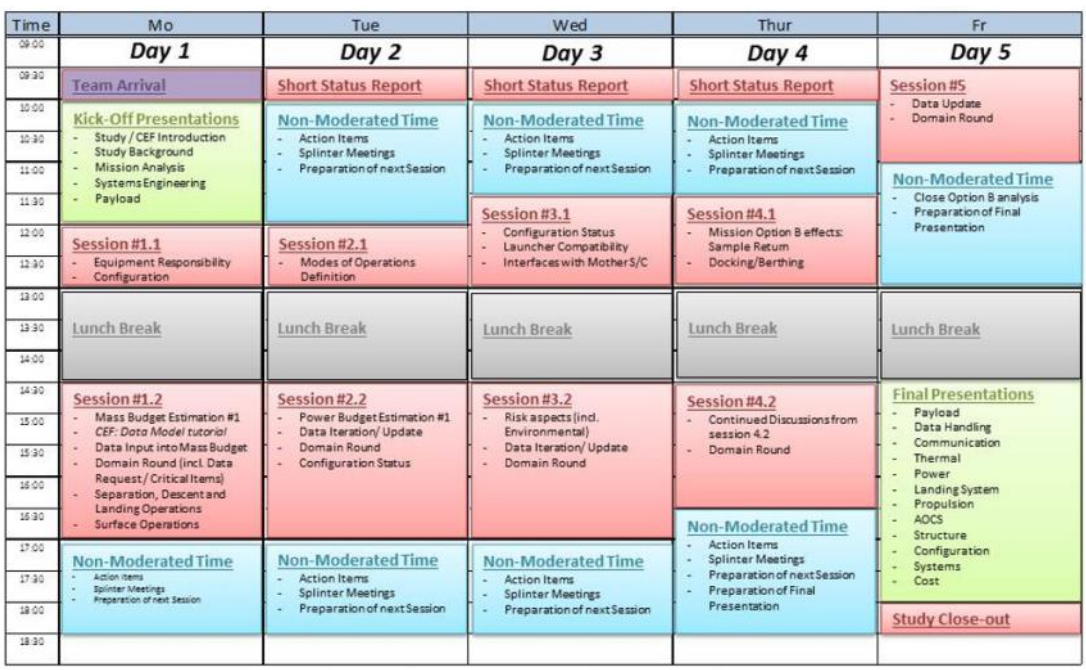


Figure 3: DLR Concurrent Engineering study phase schedule example (one-week approach)

customer, Science/Payload Engineering, Systems Engineering, Mission Analysis, subject matter expertise for Structure, Thermal, Power, Command and Data Handling, Telecommunication and Telecommand, Attitude and Orbit control, Propulsion, Accommodation, Mission Operations, Risk/Product Assurance and also Cost Engineering/Analysis.

At the end of the preparation phase, the CE study organizers distribute a study scope document to the entire team to create a common foundation. In the beginning of the actual study week, when everybody comes together in the CEF, the key information is presented again to the team. Afterwards, the work starts immediately with discussing the impact of the top-level requirements for the mission and system design, and with initial definitions of the product tree, preliminary subsystem (S/S) sizing and operational modes. That is when the fun part for all participants including the cost estimators begins.

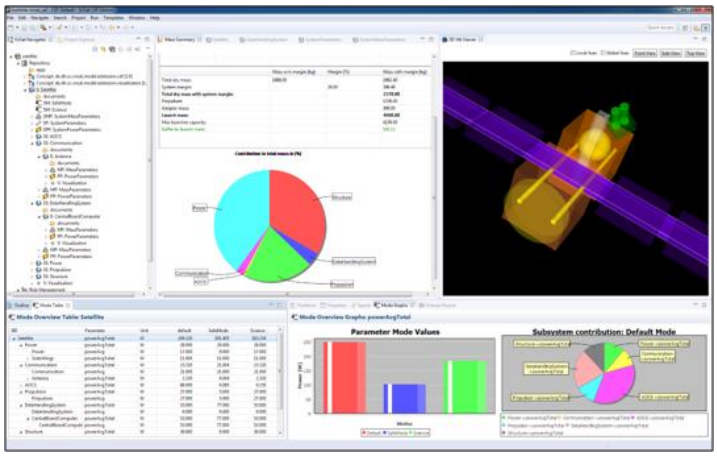


Figure 4: DLR Virtual Satellite data model

Cost Estimation in a Concurrent Engineering Environment

In not so serious terms, one has to imagine to have a counter on the desk with 32 to 40 hours counting backwards, approximately 20 people in one single room (thereof at least 15 purely technical experts), a challenging mission statement displayed on the screen, and the only one who is interested in things like fiscal year or full accounting cost is the cost estimator. This is cost engineering within a CE environment in a nutshell.

In more serious terms, depending on the level of preparation, time and experience of the cost estimator, a typical set of activities during such kind of CE study looks like this:

- gather project-related data to establish technical and programmatic baseline,
- identify similar missions (if data available) and derive analogy-based specific ROM cost values as starting point,
- check what methods and tools should be further used, and discuss this with project manager and customer,
- use available data, perform estimates, iterate as the data becomes more mature,
- support the technical team and managers with cost expertise during system trades,
- compare and cross-check estimates amongst different methodologies and tools, if possible, and
- identify and present what cost have been elaborated in detail, and which are estimated with more simple rules of thumbs, or even have not been included at all.

Methods and tools used in the CEF cover amongst others the Small Satellite Cost Model (SSCM) 2014, TransCost, internal Excel tools based on Cost Estimation Relationships (CERs), WBSs and/ or the T1 Equivalent Units approach [6], and formerly also the Unmanned Space Vehicle Cost Model. Central data models used at DLR include mainly the Virtual Satellite (VirSat) shown in Figure 4, but also the ESA Open Concurrent Design Tool and the former ESA Integrated Design Model workbooks as complementary and optional models [7].

13 Reasons Why…

In the following, the selected 13 reasons why a cost estimate during a Concurrent Engineering study could go wrong are discussed. For sure, there are plenty of others which could lead to tough work or even wrong results, but these are most prominent reasons according to the author's experience. Moreover, most of them are intertwined and also not exclusively applicable during CE studies but also in any other cost estimation activity, some are even very obvious, but these selected reasons may increase the level of impact when they come true. For each of the aspects there are some ideas, lessons learnt or recommendations provided on how problems could be reduced or even avoided.

Wrong expectations (#1)

Customers in a CE study at DLR come from totally different areas. They could be project managers, department/group/directorate heads in charge of a space program, Principal Investigators or entire science teams. Depending on the type and number of stakeholders, their background and interests as well as their expectations with respect to the cost estimation results may extremely vary from study to study but also amongst the estimator and the customer within one particular study.

The CE approach is very suitable for early design activities. That is why these multi-disciplinary studies take place most often in Phase 0 or A of a project. This results in a certain granularity of the estimate, with cost usually presented on segment or subsystem level.

However, often it is expected to provide already a bottom-up estimate on work package level, showing even labor cost, material cost, facility and operational cost (see also problem area #11). Other customers rather want to see a split between non-recurring development and recurring production cost. Most of them expect the results (without knowing them in advance of course) to meet their available budget, which is often fix and constant per year, within the available time. Almost all study customers would like to get a single, final number at the end of week which they can take home and which is rarely considered a subject for further correction or increase afterwards.

The expected level of detail is often not consistent with the time available to provide the results, nor with what the estimator believes should and could be done at this early stage. Moreover, it might not be understood that even the cost estimation tools available can barely be applied to all of the missions to be studied, particularly for new designs.

To avoid bad surprises at the end, the cost estimator needs to iterate with the study leader and customer the expectations already during the preparation phase, i.e. prior to the first study session. Part of such discussion should be how the standard cost estimation process, as described for example within the *NASA Cost Estimating Handbook* [8] and shown in Figure 5, could be tailored. It has to be agreed on what the most relevant and possible cost breakdown might be. For instance, if and how production and

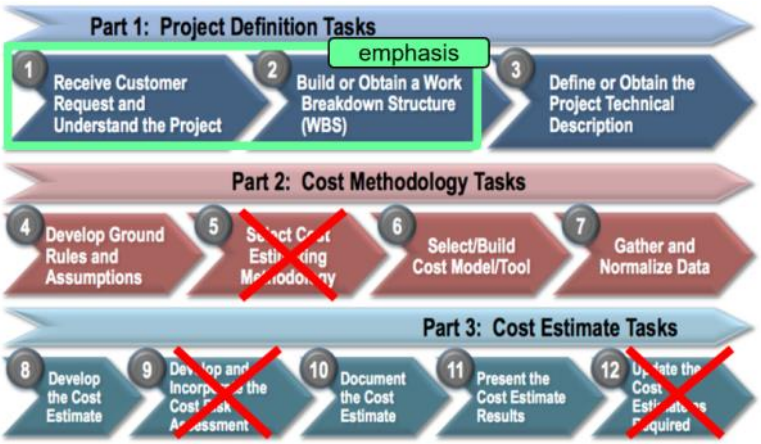


Figure 5: Random example of tailoring the NASA cost estimating process, adapted from [8]

development cost, systems and subsystem engineering, labor and material cost, investments and facilities are broken down. Moreover, it shall be clarified, if the focus is set on space segment cost or e.g. operations cost. These discussions support the decision which methods and tools could be an option for the estimator, and the identification on how the final format for the representation of the cost can be set up in a most suitable way.

International and multi-disciplinary Team (#2)

The CE study team is not only multi-disciplinary but also very international, particular in European entities such as ESA or DLR. Various nationalities are working together in one room, which brings in different cultures, different ways of thinking, working and communication, as well as different languages and levels of English. This is a very powerful basis to boost creativity and it also provides a vast range of knowledge due to the different educational backgrounds and maybe previous international company experiences. On the other hand, for a one-week CE study this compiled team has to be harmonized somehow, which is a challenge for all domains and subject matter experts (and not only cost).

In addition to the team working on the design within the CEF, if the CE study is part of a bid preparation, the potential industry consortium planned for implementation may significantly affect the labor rate or productivity assumptions to be considered. This is true for parametric and other estimation methodologies.

During one study there was an engineer who considered the involvement of Greek institutions for building a formation of CubeSats. Although the currency for most countries in Europe is euro (€), labor rates can be completely different when comparing e.g. northern with southern European countries. In this study case it was required to decide which work packages (or S/S) should be assessed with a labor rate of 200,000 € per work-year and which ones with 100,000 €.

Prior to the study, even if not a single detail is available for the technical baseline, a cost-internal stakeholder analysis should be performed. It needs to cover all aspects related to the different team members and maybe their different attitude in terms of supporting a cost estimate. It also shall identify how different international contributions for the mission could affect the estimation process, and what elements (e.g. labor) might need adjustments. This exercise last only minutes but it can save a lot of time, hassle and last-minute corrections during the study sessions.

Tools not available or applicable (#3)

Concurrent Engineering follows an iterative approach which requires rapid assessments and analyses, quick engineering tools, intensive communication, the ability to think out of the box, but also a systematic way of performing the tasks as much in parallel as it could be. But space missions are often characterized by unique designs. Space system cost and also technical data is barely available, especially if the own company does not have a large record of building space systems itself.

The time do develop dedicated CERs, maybe even based on a poor data set, is often simply not given during a CE study (see also #7 and #8). Therefore, the remaining solution is typically to use an already established tool which supports the estimate with historical underlying data and CERs gathered and developed by others that are not transparent to the end user.

There are some tools out there which are accessible for everyone on no cost. In many cases they cover a special mission or system type, for example the Small Satellite Cost Model [9], which covers the S/C bus cost (which is the full satellite without the payload), roughly in the 100-1000 kg range. More detailed or more powerful tools can be in-house developments (such as several NASA ones) or commercially available. Unfortunately,

not all institutions are able to afford commercial tools or even to invest in extensive internal developments.

However, sometimes the space mission to be designed and analyzed is so special that even no tool is applicable. This leads to a lot of modelling during the dense set of CE sessions by the cost estimator, which is already challenging. By using as a starting point freely available CERs, a parametric tool, or specific ROM cost factors from former missions for the basis of estimate (BoE), still a lot of adjustments have to be made. Most cost data are captured in US dollar ($), and might force the estimator to convert the results into the required currency. Then the question still remains which inflation scheme should be applied, the NASA inflation index, European annual average inflation or a national one. Compared to the overall uncertainties, especially for the very specific space missions, such aspects could potentially be neglected. Furthermore, the desired cost breakdown is not fully possible, or the tools/CERs do not capture the latest technologies, or some parameters are out of range.

The lack of full applicability could be compensated by following an amalgamation approach as described in [10] and substitute e.g. certain parametric estimates with dedicated analogy or bottom-up estimates on S/S or unit level, or by performing benchmarking [11] and combing cost references from different other missions where some elements are similar in one, and some elements are similar in another mission (or system). Ultimately, the decision has to be made whether the available support tools are fully or partly applicable, whether they can be made applicable or not. If the latter is the case, then do not use it.

Specific / ROM cost (#4)

CE studies could be hectic events from time to time. The fact that one can hardly compensate with working over-hours, given the short and intense study phase, may lead to too quick and hence too dirty assessments. For example, in order to have an initial feeling on the overall cost, the cost estimator could do a quick ROM cost assessment using a simple analogy estimate or specific cost factors from literature, such as cost per S/C mass (e.g. in k€/kg). However, due to lack of time, data clarity, understanding or precision, both the estimator and the customer could simply have wrong interpretations of such a factor, which was identified or given.

Specific costs are often not equipped with a fiscal year, which should be carefully considered if the developed/found value is old. But even more important are the correct contextual assumptions for the mass and cost contributors. If they are unclear, following situation could occur: Imagine a mission with a S/C dry mass of 250 kg and launch mass of 350 kg, with a cost of 50,000 k€ for the S/C itself and 100,000 k€ for the entire project lifecycle (incl. launch and operations). If it is not completely clear what the specific cost value in k€/kg is referring to, this can lead to significant differences up to a factor of 2.8 in our example (i.e. 400/143), as can be seen in Table 1. Additionally, the term S/C is sometimes used for the service segment (bus) only, but sometimes for the full satellite including the payload (P/L).

Specific cost options [k€/kg] (FY 2020)		S/C bus cost [k€]	Project cost [k€]
		50000	100000
S/C dry mass [kg]	250	200	400
S/C launch mass [kg]	350	143	286

Table 1: Different interpretations of specific cost in k€/kg

The estimator needs to make sure what values shall be taken, and explain this in front of the entire study team. And if someone else is arguing during the study that the specific cost number from another source is different, first it has to be agreed on the correct interpretation of this initial -quick-look reference number.

Use of margins and contingencies (#5)

In the early design phase, there is still a lot of uncertainty carried along, and therefore, a proper margin and contingency philosophy has to be applied. There are several standards and guidelines, for instance the *Concurrent Design Facility Studies Standard Margin Philosophy Description* compiled by the European Space Agency [12].

Options: =>	Option (1)		Option (2)	Option (3)
Cost Item	mass [kg]	Design Maturity Margin (DMM)	mass + DMM [kg]	mass + DMM + SM [kg]
Payload	85	10%	86.10	103.32
S/C Bus S/S total	206	17%	240.17	288.20
Structure	70	19%	83.30	99.96
Thermal	10	20%	12.00	14.40
Power (EPS)	40	12%	44.80	53.76
AOCS/GNC	35	15%	40.25	48.30
Propulsion	10	20%	12.00	14.40
TT&C	23	14%	26.22	31.46
C&DH	18	20%	21.60	25.92
Total dry	291		326.27	n/a
Systemmargin (SM)		20%	65.25	
Total dry + SM			391.52	391.52
Propellant			30.00	30.00
Total wet			421.52	421.52
Launch Adapter			5.00	5.00
Total launch			426.52	426.52

Table 2: Satellite mass budget example on subsystem level, showing 3 options for mass values to be used as potential inputs for parametric cost estimation tools

In CE studies there is an interdisciplinary and multi-cultural team (as for most projects in general) which has been called in to support the present study. And this team is not necessarily used to work together. This means that the systems engineer and team leader have to make sure that everyone has the same understanding related to the application of contingencies and margins. This is to avoid double-counting or forgetting them, or piling them up in an unfortunate way, as shown e.g. in [13]. Furthermore, for using the technical parameters and requirements as input for parametric cost models, it must be clear exactly what values are to be taken. When using for instance mass-based CERs, there are in principal three major options.

Table 2 presents a mass budget on subsystem level for a small satellite, where the masses are the sum of the respective equipment, with and without margins. Out of these three options listed in the following, it has to be decided which mass values should be used:

1) S/S mass (as sum of the equipment masses) without any margin, i.e. best guess only, shown in the 2nd column from the left,
2) S/S mass including design maturity margins (DMM), displayed in the 4th column from the left,
3) S/S mass including DMM, plus the system margin portion on top, i.e. the values from the 4th column and additionally 20% extra for each S/S.

Usually, the tools and CERs are primarily based on actual data. Furthermore, mass growth is a typical phenomenon in space system development which eats up contingencies and margins throughout the phases. Hence, it is recommended to use option (3) if nothing else is explicitly requested, which is the S/S mass including DMM and the system margin portion on top. Alternatively, option (2) could be used, but the additional uncertainty shall be clearly reflected in the cost-risk analysis or at least within the documentation of the results.

Depending on the data model used for the CE study (or project in general), the S/S mass values may need to be recalculated at some stage, e.g. with factor 1.2 in our case. Looking at our example, this means that the Thermal S/S mass to be taken for the CER or tool is not 10 kg, nor 12 kg, but 14.4 kg. Please see also reason #9 (rapid data changes) for further discussions on data model value utilization.

During the tool selection process, which should take place prior to the actual study phase, the use and application of technical margins not only for

mass but also for other parameters should be clear, documented and agreed on. During the rapid and iterative estimation loops within the CE environment these details may be easily overlooked.

Heritage & Complexity (#6)

As for any other study or project, the cost estimation has to consider factors for heritage and complexity adjustments. Particularly for parametric estimates, which are based primarily on CERs with mass as independent variable, the results would not capture how much of the design and test effort and models could be saved or needed due to heritage, nor how complex either the design, assembly and integration or control of the space system could be. In an early phase CE study, the team has likely an understanding whether they design something new or just a derivation of an existing system. But for the cost estimator the question remains, how strong this would affect the results. Some CERs and tools account for one or both factors already. Some do not consider them at all. Moreover, there are big differences on how heritage and complexity are addressed within these tools.

The estimator has to make sure whether the data, tools, models or CERs account for this already, or if these factors have to be applied on top of the given outcomes. The key assumptions in the *SSCM 2014 User's Manual* [9] for example state an average amount of heritage and an average level of technological complexity, stressing the fact that a proper cost-risk assessment is required. Alternatively, a certain percentage, a linear or an exponential factor could be used as done for several CERs. However, this has to be selected and defined with care. These factors can vary from only a few additional percentages to doubling or tripling of cost when comparing an average heritage (e.g. 50%) to a completely new development. The same is true for similarly subjective assessments of complexity.

Such adjustments, either manually or as a part of a tool, should be factored in at the very end of the study, when most technical data are available. During the CE study itself the team or at least systems engineer will usually strive for highest possible heritage and lowest complexity. As an estimator, keep an eye on it, try to support the discussions and trades along the way, but work this out in detail as late as possible. If possible, this exercise should be done on S/S-level to reflect a potential high or low re-use and complexity per S/S of the space system, compared to others.

Lack of time (#7)

This is a major, but self-explaining issue, probably partly also a self-made problem of the DLR CE approach or institutions with similarly dense study timelines. Although this approach is very efficient, the absolute time for analysis and potential re-work is short. First, within one week plus maybe some days before and afterwards, one cannot perform the complete cost estimation process as stated e.g. within the *NASA Cost Estimating Handbook* [8] in full detail, simply due to the lack of resources and the early stage of most studies.

The lack of time is a central reason for potential cost estimation errors or incompletion. It is critical for all domains, but the cost domain is heavily dependent on the outputs from others, which are used as input for the cost analyses, and hence the estimator is rather busy during later design iterations.

Therefore, it is imperative to use a tool, calculation, CER, or a model template the estimator is familiar with. There won't be much time for experimenting. Implementing a proper process and adapting the tools for it, standardizing them and connecting them to a data base could turn the problem into an opportunity, and enable a very efficient design process and cost estimation. This is the case for example during NASA Team-X studies, where costing at the speed of light [5] is commonly performed.

Lack of data (#8)

Cost estimation relies heavily on data. This includes technical data to establish the technical baseline for an estimate, as well as cost data from previous missions, designs or equipment selected. Often the estimator is lacking both, due to the technical immaturity of the present mission/system at that stage, and also due to low (or no) comparability to former missions or simply lack of access to previous mission data. Unfortunately, in Europe there is no public database available such as CADRe or ONCE [14] in the United States.

This is again one of the reasons why parametric tools with a few technical input parameters are essential and of great help during this early stage of mission design. CERs and related tools making use of them (if available and applicable) contain already a large set of data points, which do not have to be researched again. If there are technological or operational differences apparent between the CERs used and the spacecraft to be designed for instance, effort shall be made to replace or adjust the cost of particular subsystems which differ most. This can be done by using e.g. benchmarks from other subsystems of more suitable space missions where cost may be known, as also proposed in [11] (see also #3). At least the unknowns have to be known and clearly documented in any case.

Rapid data changes (#9)

Concurrent Engineering and its highly iterative nature involving every discipline early on in the project is a big advantage. However, the rapid evolution of data leads to a couple of challenges.

During one week, the total launch mass may change dramatically after each session. We look at following example: There is a requirement for a small satellite mission with a maximum launch mass of 300 kg. At the end of study day one, with an initial version of the product tree available, the preliminary mass budget indicates a launch mass of 225 kg. However, not every engineer adds the relevant data into the data model already in the beginning, so maybe the structural mass is still missing entirely, the harness mass is not yet considered, and the propellant mass is completely unknown. In the course of the second day, subject matter experts close some design gaps, discuss and re-iterate with comfortable contingencies. This results in a total launch mass of 410 kg. During day 3, the team identifies that the P/L and the S/C bus both included an optical bench and Star Trackers within their budgets, that the operational modes are not fully consistent, and that there is no need for an X-band system anymore. This leads to an updated launch mass of 340 kg. Day 4 is typically the day reserved for refinements. The amount of data needed as input for an e.g. parametric cost

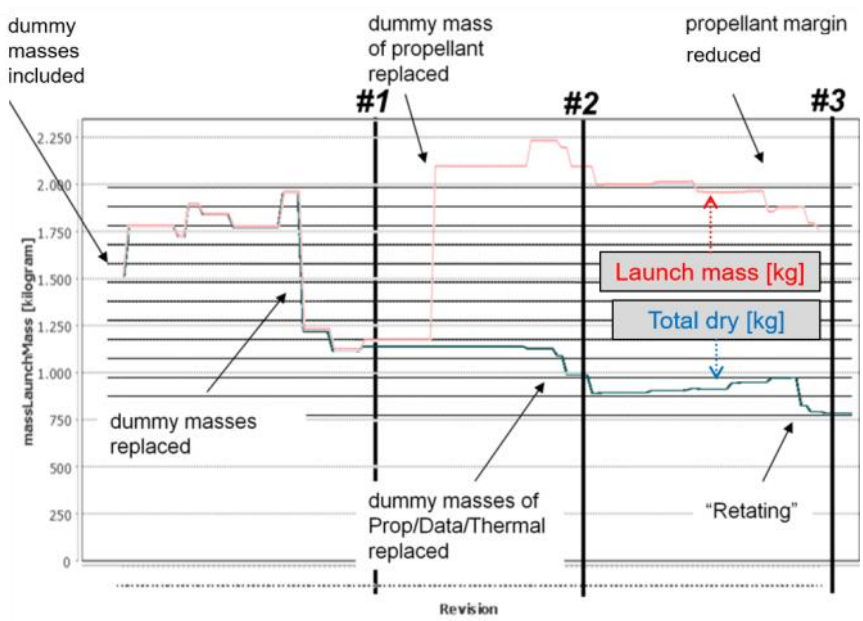


Figure 6: Example of mass variations over time during a Concurrent Engineering study, taken out of the data model history

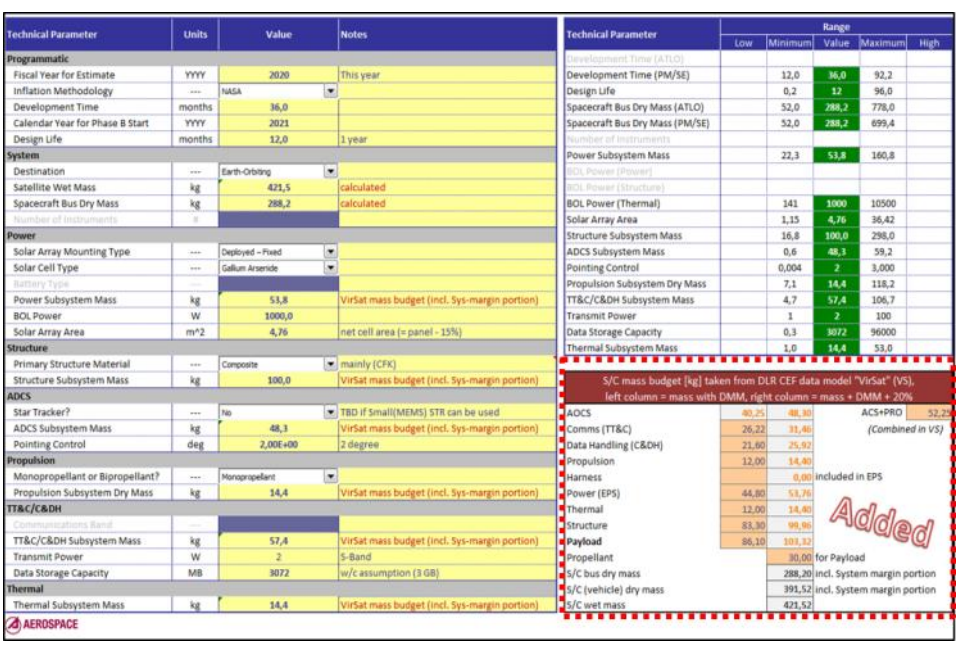
estimation model is mostly complete, and – in our example – the total launch mass deceased to 290 kg, which is compliant to the requirement.

However, during the final presentation session on day 5, one engineer figures out that the redundancy scheme for the avionics is not compliant to the failure-tolerance requirement for this mission. Now the mass increases up to 320 kg again, which won't be a show stopper at this stage, but shall indicate that there is always changes to be expected. Another example of these changes is presented in Figure 6. In order to constantly build up and update the cost estimate, by e.g. using amongst others the SSCM for this S/C size, the available cost model needs to be updated easily without mixing up numbers or forgetting something.

As for many other problems, preparation is also the key here. The cost estimator needs a good understanding of the potential cost drivers already prior to the study, make first and robust assumptions for the technical baseline, and perform initial sensitivity analyses. Furthermore, the selected tools should be usable for such a series of iterations. As an example, Figure 7 shows the SSCM 2014, where an input sheet has been modified accordingly. On the left side is the original input area for the tool, while on the top right the technical parameter values are checked to see whether they are in the permissible range or not. Manually added, there is a box on the lower right side, in which the mass budget on S/S level can directly be taken from the CEF data model.

In Figure 7, the S/S masses are converted to masses including design maturity margins plus the system margin portion (as discussed in #5), and then linked to the actual input area. Moreover, the system masses (dry, wet, launch) are organized in such a way that a quick comparison with the actual system mass budget is easily possible to identify gaps or overlaps.

It would be even better, however, if such an adaptation effort would not be necessary, but unfortunately most cost tools or calculations are difficult or inconvenient to connect to the central, multi-accessible data model, and vice versa. This brings us to the next reason why a cost estimate in a CE environment could go wrong.



Technical Parameter	Units	Value	Notes
Programmatic			
Fiscal Year for Estimate	YYYY	2020	This year
Inflation Methodology	---	NASA	
Development Time	months	36,0	
Calendar Year for Phase B Start	YYYY	2021	
Design Life	months	12,0	1 year
System			
Destination	---	Earth-Orbiting	
Satellite Wet Mass	kg	421,5	calculated
Spacecraft Bus Dry Mass	kg	288,2	calculated
Number of Instruments	#		
Power			
Solar Array Mounting Type	---	Deployed – Fixed	
Solar Cell Type	---	Gallun Arsenide	
Battery Type	---		
Power Subsystem Mass	kg	53,8	VirSat mass budget (incl. Sys-margin portion)
BOL Power	W	1000,0	
Solar Array Area	m^2	4,76	net cell area (= panel - 15%)
Structure			
Primary Structure Material	---	Composite	mainly (CFK)
Structure Subsystem Mass	kg	100,0	VirSat mass budget (incl. Sys-margin portion)
ADCS			
Star Tracker?	---	No	TBD if Small(MEMS) STR can be used
ADCS Subsystem Mass	kg	48,3	VirSat mass budget (incl. Sys-margin portion)
Pointing Control	deg	2,00E+00	2 degree
Propulsion			
Monopropellant or Bipropellant?	---	Monopropellant	
Propulsion Subsystem Dry Mass	kg	14,4	VirSat mass budget (incl. Sys-margin portion)
TT&C/C&DH			
Communications Band	---		
TT&C/C&DH Subsystem Mass	kg	57,4	VirSat mass budget (incl. Sys-margin portion)
Transmit Power	W	2	S-Band
Data Storage Capacity	MB	3072	w/c assumption (3 GB)
Thermal			
Thermal Subsystem Mass	kg	14,4	VirSat mass budget (incl. Sys-margin portion)

Technical Parameter	Range: Low	Minimum	Value	Maximum	High
Development Time (ATLO)					
Development Time (PM/SE)		12,0	36,0	92,2	
Design Life		0,2	12	96,0	
Spacecraft Bus Dry Mass (ATLO)		52,0	288,2	778,0	
Spacecraft Bus Dry Mass (PM/SE)		52,0	288,2	699,4	
Number of Instruments					
Power Subsystem Mass		22,3	53,8	160,8	
BOL Power (Power)					
BOL Power (Structure)					
BOL Power (Thermal)		141	1000	10500	
Solar Array Area		1,15	4,76	36,42	
Structure Subsystem Mass		16,8	100,0	298,0	
ADCS Subsystem Mass		0,6	48,3	59,2	
Pointing Control		0,004	2	3,000	
Propulsion Subsystem Dry Mass		7,1	14,4	118,2	
TT&C/C&DH Subsystem Mass		4,7	57,4	106,7	
Transmit Power		1	2	100	
Data Storage Capacity		0,3	3072	96000	
Thermal Subsystem Mass		1,0	14,4	53,0	

S/C mass budget [kg] taken from DLR CEF data model "VirSat" (VS), left column = mass with DMM, right column = mass + DMM + 20%

AOCS	40,25	48,30	ACS+PRO 52,25
Comms (TT&C)	26,22	31,46	(Combined in VS)
Data Handling (C&DH)	21,60	25,92	
Propulsion	12,00	14,40	
Harness		0,00	included in EPS
Power (EPS)	44,80	53,76	
Thermal	12,00	14,40	
Structure	83,30	99,96	
Payload	86,10	103,32	
Propellant		30,00	for Payload
S/c bus dry mass		288,20	incl. System margin portion
S/C (vehicle) dry mass		391,52	incl. System margin portion
S/C wet mass		421,52	

Added

AEROSPACE

Figure 7: Screenshot of an adapted SSCM 2014 [9] input sheet

Disconnection to central data model (#10)

Using a central data model, which acts as a single source of truth is great. It can be used in any project, but a CE study is a good event for which the data model could be initiated or initially be prepared for. In principle there is nothing negative but only positive: a bit of consistency is better than no consistency, thus we are talking about a luxury problem. But the cost domain is typically not included in these data models. This is also true in many cases for some domains, which use powerful commercial software, such as computer-aided design (CAD) or orbital simulations tools. There are attempts to interface these tools to the central data model but this is still not very common.

For the cost estimator this means that an effort could be made to somehow link the estimation templates (e.g. spreadsheets), CERs, or own databases to such a model, if confidentiality or other non-technical aspects allow it. Since rapid data changes occur (see #9), it is mandatory to make robust, well forecasted assumptions for premature technical input data. One needs to keep an eye on the data model results and organize the relevant model outputs, which are of interest for the cost estimation as good and efficient as possible. Cost Engineering as part of Model-based Systems Engineering (MBSE) is definitely an underestimated issue, which provides a lot of opportunities for further research.

Figure 4 showed the Virtual Satellite central data model used at DLR. It is an eclipse-based and open source tool enabling multiple-access (with dedicated role management). It uses Subversion for version control. It includes features such as a product tree, prepared mass budgets, power budgets and modes, a preliminary distributed CAD functionality, functional diagrams, a calculation mask and an Excel interface, but no dedicated cost estimation feature. This is just one example, which indicates that the concept of cost estimation has still not fully arrived in the MBSE world.

DLR is working on this topic and welcomes any other activities going into the same direction, which seems to be the case by looking for instance into the presentation list of the ICEAA 2020 workshop [15].

Bottom-up estimates during a CE study (#11)

CE studies are most suitable for Phase 0/A studies, as mentioned already. This means that the primary cost estimation methodologies are parametric or based on analogies. However, similar space missions or systems are barely available, either because something comparable has never been designed or the data is simply not available, which makes analogy assessments sometimes difficult. The parametric approach on the other hand is not well understood by many engineers and sometimes not even accepted (see also #1 and #13). This is particularly true when a tool or CER is used, which does not really reflect the way of computing cost for a certain type of mission or for a certain institution or culture. As a result, much effort is spent defending the methodology selection and respective results, instead of improving the estimate itself.

Besides the managers or customers who want a super-detailed cost estimate already in a Phase 0 study, although it is still not even clear if for instance a Propulsion system is needed or not (again, see #1), many engineers tend to feel more comfortable discussing materials and labor cost than to trust a number which is spit out of a parametric tool. Unfortunately, the power of parametric estimation is not always understood. During trade studies, where the estimator could easily assess with their CERs the financial impact of using e.g. a Star Tracker or not, or the pointing accuracy cost sensitivity, many engineers do not trust this statistics-based approach.

Consequently, during many CE studies, a preliminary bottom-up estimate shall and has been made. The advantage is that the estimate makes use of the engineer's experience in terms of materials and labor cost. However, the former

may not be properly linked to the model philosophies, test and ground equipment. Especially the spacecraft operation is often drastically under or overestimated, which is due to the short time available and the pressure to continue iterating rather on the technical parameters. As a result, within a CE environment which follows more condensed approach of days instead of weeks, the disadvantage of bottom-up estimates in early phases becomes very apparent.

One lesson learnt is to have, based e.g. on parametric studies, a rough cost distribution per S/S at hand, and a preliminary assessment of how much additional effort is needed for system wraps, such as management or product assurance. It could be decided on a case by case basis whether or not the domain experts should be confronted with these historical and average values upfront, to get an idea on the ballpark values for their more detailed cost contributions. If specific cost factors (e.g. in k€/kg) are available and well understood (see #4), they are helpful for sanity checks, too.

Moreover, for a bottom-up estimate there has to be a common attitude and set of assumptions amongst all contributors, which include the subject matter experts, and maybe their superiors. It makes a huge difference if someone tends to provide a very conservative number to already claim a certain work package budget and to prepare for upcoming negotiations, or if someone does rather the opposite and estimates rather at the lower end, with realistic cost distributions over time, to ensure that the project is more likely to be funded. If bottom-up estimates are really necessary or desired, the cost breakdown and approach need to be clear to everyone (see also #1).

Optimizing in the wrong place (#12)

A space mission consists of different segments, such as the space system (including bus and payload), the launch vehicle, and the ground segment including operations. Most CERs and tools are available for the space system, some with, some without payload. Moreover, the majority of CE study team members each represent one S/C subsystems. This might support a more detailed cost estimate on S/C bus level, no matter which estimation methodology is applied, compared to the other segments.

There are also holistic tools out there, such as the parametric QuickCost tool [16]. The S/C bus and P/L cost in version 6 of this tool are estimated using CERs. Launch cost are entered directly (if desired) while all other NASA WBS elements are covered by adding various and suggested percentages to the sum of the S/C bus and P/L costs. Using the average values shown in [16], the space segment is dominating the total project cost with approx. 60-80%, depending on the launch cost. However, especially for long-duration science and exploration missions, the operations cost can increase significantly. But this can also be the case for more regular Earth Observation missions if standard components are used, or low complexity and a strong heritage approach is followed.

The key message is that while detailing one part of the project life-cycle cost, it could be easily underestimated that there is significant cost, or uncertainties associated to other parts as well. Focus should be set on the cost drivers. Discussions on 100 k$ level should be saved for later, and a rough mission cost breakdown has to be prepared, based on the most suitable references and most driving requirements. For the most likely used CERs, the sensitivity and slopes need to be known in order to know better on which updated values to focus, and where the estimate can survive with rougher assumptions (since the cost differences may not be significant).

Lack of acceptance or perceived relevance (#13)

As indicated already, the non-technical participants of early space mission studies are

the absolute minority. Focus is often set on the science case and technical feasibility. However, without an initial assessment on the cost, no statement regarding a potential implementation of this mission can be made. Only studies of commercial systems (e.g. for a new communication satellite) might be different, since they do not only include the cost but also business model considerations. On the other hand, commercial systems usually do not require a pre-Phase A analysis, since there is very likely a reference S/C platform available within the industry, and the mission-related aspects are comparably simple.

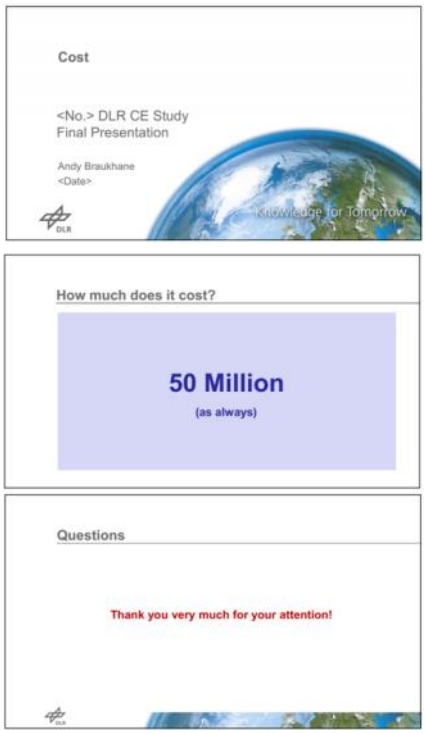


Figure 8: Set of concluding cost presentation slides at the end of a study, which raised attention before true content was shown

However, technical models and design processes amongst engineers are understood, even if one does not exactly know how to design another subsystem for example. For example, an electrical engineer developing a power subsystem has an idea of the steps needed to design an on-board computer architecture, and should also be able to properly assess the risk and potential mitigation strategies. The work package leader also might have some cost numbers at hand and can provide an estimate of the required labor throughout the development (with a very big uncertainty for fancy missions analyzed in a very early phase). However, if less technical terms like confidence levels or fiscal years are presented, most engineers often cannot, or do not want to understand why this is even important, or do not pay attention at all until the final magic cost number is shown.

Experience shows, that the cost estimate presented is subject to intensive discussion, much more than the maximum power demand during an orbit raising maneuver. There is also sometimes the tendency, rather from management than engineering side, to quickly re-assess and oversimplify the cost on a napkin, with the aim to show that the estimate is still too high.

Having in mind that most of the described problem areas in this paper are also applicable to some other technical domains, there simply might not be the time left to talk extensively to the engineers for proper cost and cost-risk assessments, since they need (or want) to focus on their design tasks.

As for many other things, it is important to properly explain all assumptions, processes and steps to make them transparent. Educating others, and to make aware that a decision made by someone affects the design of someone else is imperative, is one of the strengths of the CE methodology.

For example, in the course of a mission selection campaign at DLR, several 3-day CE studies have been conducted, with the aim to investigate missions and science cases to be realized with a small satellite. During the final presentation session, cost is usually one of the last talks (maybe this should be changed one day). Probably due to the above discussed aspects or the fact that long days were behind the team, almost no one paid attention. For one of the later designed missions, another lesson learnt was to shock them a bit, with an extreme simplified (i.e. very easy to digest) content and presenting solely the maximum possible cost which had been assessed.

Figure 8 shows the three slides that were presented as a first shot, before the joke was confessed and the actual presentation was given with all the assumptions and the approach. As a result, everyone was awake, paying attention and

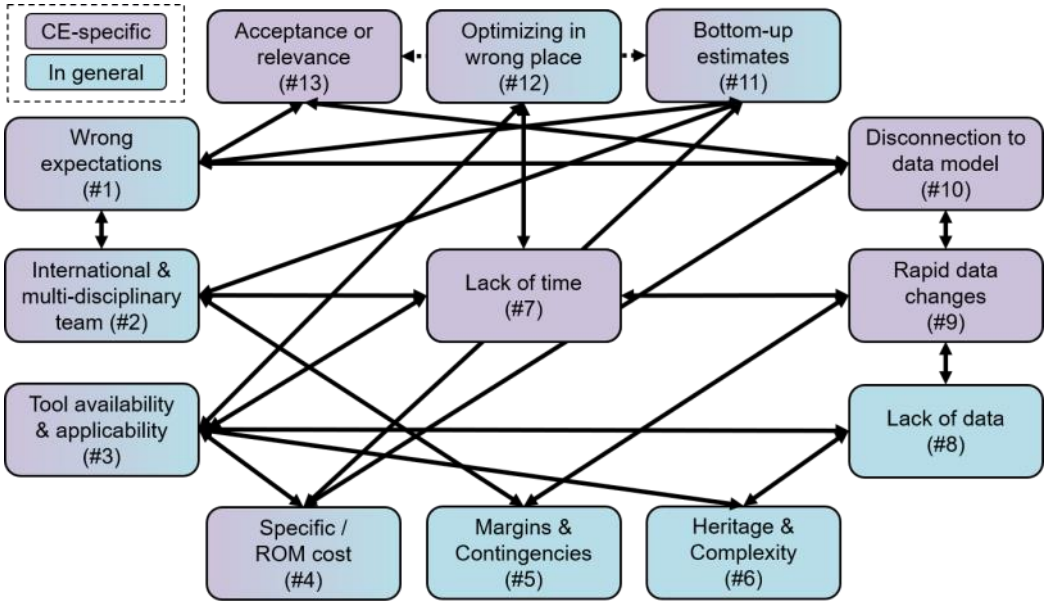


Figure 9: Mutual influences of discussed problem areas

ultimately well understood how the numbers were identified, adjusted and how they could be compared with the other missions.

Summary and Conclusion

Concurrent Engineering is a very efficient approach, well suitable for early phase studies in the space domain. It reduces time, cost and risks while increasing quality and mutual understanding. However, it is not perfect and also has some dark sides as discussed in [17], depending on the implementation and application.

The presented work discusses 13 problem areas and reasons why a cost estimate, which is performed in a CE environment, could go wrong. The focus was set on the DLR approach to Concurrent Engineering.

As stated, many of these reasons are not exclusively limited to the cost domain or even CE, but also for early phase projects and collaborative efforts in general. They are also not self-standing but closely linked to each other, and the list is not exhaustive at all.

Mutual influences

As indicated within the previous subchapters, most of these reasons are linked, mutually influenced and even dependent on each other. Some are more CE-specific, some apply to the cost engineering process basically within all projects. Some are more DLR-specific, some relate to all similar processes.

The mutual influences presented in Figure 9 are an attempt to highlight what are the most dominant reasons, which potentially could create or amplify other reasons why cost estimation in a CE study could go wrong. The more connections, the stronger might be the direct influence on other factors. However, this does not relate to the actual impact on the cost estimate but shall only indicate what should be kept in mind first in order to maintain full control over the cost estimate performed during a CE study.

Lessons learnt

Derived from the discussions above, a set of lessons learnt is compiled in the following. It focused on four main categories, which are: Awareness, Preparation, Communication and Documentation. These categories are further broken down into twelve recommendations to fight against the 13 discussed problem areas.

Awareness

- Check who is involved
- Understand potential problems, prioritize
- Accept to make compromises, be flexible

Preparation

- Check all available data, tools, methods
- Adjust, to be fast
- Tailor, to be in-line with expectations

Communication

- Clarify and harmonize inconsistencies and assumptions
- Explain what the estimator/analyst wants and can do
- Educate how the estimate is done, and shake (or shock) the team if needed

Documentation

- Agree on what has been discussed and decided by consensus
- Make transparent what the estimator/analyst assumes and is able to provide
- Try to connect cost data to common data set/ model

One promising approach to address several of the above-mentioned aspects is to use a top-level all-in-one tool, such as the S-chart [18] used at NASA Jet Propulsion Laboratory for rapid, comprehensive mission architecting at Team-X. The aim is to provide a simultaneous view of all major mission considerations, such as the programmatic constraints, technical performances, capabilities and margins, science performances, high-level system descriptions and also cost. Such chart, or something along those lines, can be permanently displayed in the CE environment to keep everyone informed about the latest status. If it is already embedded within a central data model, this would be even better.

References

[1] Romberg, O., Braukhane, A., Schumann, H., Status of the Concurrent Engineering Facility at DLR Bremen, in: German Aerospace Congress (DLRK) 2008, Darmstadt, Germany

[2] Bandecchi, M., Melton, B., Gardini, B., Ongaro, F., The ESA/ESTEC Concurrent Design Facility, in Proceedings of the 2nd Concurrent Engineering Conference (EuSEC), Munich, Germany, September 2000

[3] Martelo Gomez, A., Quantius, D., CE facility and process at DLR, Airbus Engineering Dialogue 2019, 19.10.2019, Bremen, Germany

[4] Braukhane, A., Romberg, O., Lessons Learned from One-Week Concurrent Engineering Study Approach, International Conference on Concurrent Enterprising (ICE), Aachen, Germany, 20-22 June 2011

[5] Hihn, J., et al., Costing at the Speed of Light: How a Concurrent Engineering Team Can Bootstrap Your Estimation Capabilities, presentation at the 8th International SECESA, 27 September 2018, Glasgow, UK

[6] Reinbold, G., Successful Cost Estimation with T1 Equivalents, Proceedings of 2016 International Training Symposium, Bristol, UK, October 2016

[7] Schaus, V., Braukhane, A., Model-based Systems Engineering applied in Concurrent Engineering, Workshop at the ISU Space Studies Program (SSP) 2013, Straßbourg, France, July 17, 2013

[8] NASA, Cost Estimating Handbook version 4.0 (NASA CEH v4.0), February 2015

[9] Mahr, E.M., Small Satellite Cost Model 2014 (SSCM14) User's Manual, Aerospace, Report No. ATR-2014-03752, 30 September 2014

[10] Trivailo, O., Sippel, M., Sekercioglu, Y.A., Review of hardware cost estimation methods, models and tools applied to early phases of space mission planning, Progress in Aerospace Sciences 53, 2012, pp. 1-17, March 28, 2012

[11] Van Pelt, M., Phase 0 Space Mission Estimates, Proceedings of the 2019 ICEAA Professional Development & Training Workshop, Tampa, FL, USA, May 2019

[12] ESA, ESA Concurrent Design Facility Studies Standard Margin Philosophy Description, Issue 1.2, ESA-TECSYE-RS-006510, August 14, 2017

[13] Kaparti, G., et al., Resource management and Contingencies in Aerospace Concurrent Engineering, NASA GSFC, 2012

[14] NASA, Office of Evaluation, Brochure: Cost Analysis Data Requirement (CADRe), One NASA Cost Engineering (ONCE), NP-2012-02-840-HQ

[15] ICEAA, 2020 W/S paper presentation summary, https://www.iceaaonline.com/sat20papers/, last access on February 18, 2020

[16] Hunt, B., et al., QuickCost 6.0, presentation at the 2016 ICEAA Training Symposium, Bristol, UK, October 2016

[17] Braukhane, A., Bieler, T., The Dark Side of Concurrent Design: A Story of Improvisations, Workarounds, Nonsense and Success, 6th international SECESA, Stuttgart, Germany, 8-10 October 2014

[18] Nash, A., Rapid, Comprehensive, Mission Architecting at the Jet Propulsion Laboratory, presentation at the 8th International SECESA, 27 September 2018, Glasgow, UK

Andy Braukhane *is a Systems and Cost Engineer at the German Aerospace Center (DLR) Concurrent Engineering Facility and Space System Analysis department. He holds degrees in Aerospace Engineering and Space Systems Engineering. Since 2007, Andy is working on early phase mission design. He worked in project management roles, as an engineer and the last years mainly as cost estimator, often within a Concurrent Engineering environment. Furthermore, he is interested in human factors and engineering process improvements.*

In Search of the Production Steady State: Mission Impossible?

Patrick McCarthy, CCEA

Abstract: Learning Curves are a vital tool for cost estimators when predicting the number of direct labor hours required for a production run. One challenge of utilizing learning curves is predicting when no additional improvement can be expected, otherwise known as the steady state of the production run. This paper will address the formal definition of a learning curve, the different types of learning that impact production systems and why the steady state plays such a critical role in cost estimates. The steady state concept, as well as its importance and impact will be explored. Interpretation of data and causes of the steady state, both genuine and artificial, will also be addressed. A sample estimate will be developed that utilizes historical data to identify an anticipated steady state and predict direct labor requirements for a new system. Lastly, the unique nature of Department of Defense (DoD) acquisition and its impact on production environments will help us determine whether the steady state truly exists or not.

Introduction

Learning curve theory is one of the most common cost estimating and analysis techniques. It is critical to estimating direct labor requirements and can have substantial impacts on costs that are derivatives of direct labor requirements, including facility/space requirements and support labor staffing.

As long as learning curve theory has been used in cost estimating, a key question asked is:

> *"Even though the mathematical model indicates that learning will continue indefinitely, is that really the case and, if not, when does the point where learning stops occur?"*

The state of the process when learning ceases, or is mathematically negligible from unit-to-unit, is called the steady state. It is critical that cost estimators understand how to analyze historical production data to determine when a system enters the steady state and utilizing that determination for estimating future system requirements. Not accounting for a steady state could result in underestimating direct labor requirements. Alternatively, predicting a steady state will occur too early could result in an overestimation of direct labor requirements. Before addressing these scenarios and answering the question as to whether the steady state even exists, a brief recap of learning curve theory is warranted.

Learning Curve Theory Recap

The universally agreed upon definition of learning curve theory is that it is a measure of efficiency gained by the act of repetition in a *constant* system over time. A critical component of this definition, and ultimately our search for the existence of a steady state is the phrase "constant system". In manufacturing, learning curve analysis in its truest form entails tracking the rate of reduction with regards to resources required (e.g. labor hours) over a period of time with the following variables remaining the same throughout:

- Production rate or throughput
- The employees performing the work
- The facility, tools and equipment used
- The scope of work being performed (including the materials and sub-assemblies used)
- Quality requirements
- Safety Requirements
- Labor Laws

Albeit with slightly different techniques, Wright[1] and Crawford[2] both sought to capture this improvement mathematically by theorizing that as the quantity of items produced or tasks completed double there will be constant rate of reduction in terms of resources required. Their techniques reflect the mathematical representations presented below.

Crawford's Unit Improvement Curve Theory

$Y = aX^b$, where:

Y = Cost of the Xth unit

a = Theoretical cost (T1) of the first unit in the production run

X = Sequential unit number of unit being calculated

$b = \log_2(LCS)$, a constant reflecting the rate of cost decrease from unit to unit

LCS = Learning Curve Slope

Wright's Cumulative Average Curve Theory

$Y = aX^b$, where:

Y = Cumulative average cost of X units

a = Theoretical cost (T1) of the first unit in the production run

X = Sequential unit number of unit being calculated

$b = \log_2(LCS)$, a constant reflecting the rate of cost decrease from unit to unit

LCS = Learning Curve Slope

Both theories address "learning" in terms of the reduction of resources required. However, Konz[3] points out that in production environments, there are actually two distinctly different types of learning that take place. This concept can have a substantial impact on how we utilize learning curve theory in the search of the steady state, so we address each learning type below.

Individual Learning vs. Organizational Learning

Konz defines individual learning as the improvement demonstrated by an individual worker or entire workforce while utilizing a "constant product design and constant tools and equipment". In contrast, Konz defines organizational learning as the learning attributed to modifying the product design, tools and equipment. Individual learning clearly echoes our definition of learning in the previous section. However, and as discussed later in this paper, organizational learning must be considered in determining the existence and timing of the steady state in a specific production environment.

Individual Learning

Individual learning can be represented by two distinctly different scenarios:

1. Suppose a manufacturer wins a U.S. Army contract that will require the company to build 1,000 units of a particular ground vehicle system. The manufacturer typically builds commercial items, so it is starting up a separate assembly line specifically for this weapon system that will have ten dedicated workstations. The manufacturer does not want to disrupt its commercial business, so it hires brand new staff and purchases all new tooling, machines and fixtures in order to deliver the Army vehicles. The Army has indicated that the delivery schedule is somewhat flexible, so the manufacturer decides that it will hire 100 workers who will start on the first day of the project and work in one, 8 hour shift per day to accomplish the work. As time passes and the workers become more experienced, improvement will be achieved in the number of hours required

to assemble and deliver each unit. As the delivery schedule will not be firm, improvement will also be achieved in the number of vehicles completed in a single day (i.e. the production rate will be variable as a function of individual learning).

2. Konz provides another scenario that only involves a single person to help further demonstrate individual learning. Suppose a novice golfer decides to learn by playing one hundred rounds this year using only a driver, 5-iron and putter. The golfer will play the exact same course at noon each day and use the exact same type of golf balls for each round. For the first round, the golfer takes 135 strokes to complete the round. The second round, he takes 127 strokes. Over the course of the year he sees his stroke total starting to level out around 100, plus or minus a few strokes each round.

In both scenarios, the environment and resources available to those performing the work remain constant.

Organizational Learning

Konz introduces the idea of organizational learning by defining it as improvement that results from "changing product design, changing tools and equipment, and changing work methods". We again use the two scenarios from above to demonstrate organizational learning.

1. Returning again to the new contract for 1,000 Army vehicles, suppose that instead of hiring all 100 workers on the first day, the workforce increases ten employees at a time over the first several weeks. Also, assume that after completing the first 100 units, the tooling and equipment purchased to complete this effort is not optimal. Then new equipment to increase efficiency is purchased. In addition, after 500 units are completed and delivered, the Army notifies the manufacturer of some design changes that will be incorporated into the assembly in order to improve survivability. Each of these changes represent the potential for organizational learning to occur as it would be anticipated that these changes would impact the hours required to build the end item when compared to the system when production commenced.

2. Konz introduces organizational learning in the golfer example by proposing that during the year, the golfer decides to add additional clubs to his bag (e.g. a 7 iron and sand wedge). The golfer may also decide to switch the brand of balls he is using and also move his tee time to 8:00 AM because he found it to be too hot playing at noon and he would become fatigued.

In both scenarios, substantial changes were made to the "systems" while they were active which more than likely altered the performance of the system and, subsequently, the measurable output or results. This is a very important concept as you will recall that one of the major tenets of learning curve theory is that the system, and the parameters that define the system, remain constant.

The Steady State Defined

Now that we have revisited learning curve theory and explored the two different types of learning, we will focus on what the steady state should look like and how we can test whether the system has truly reached that state. Gagniuc[4] provides a general definition of a steady state by stating that if the variables which define the behavior of the system are unchanging over time, the system has reached a steady state. In continuous time, this means that for those properties p of the system that we are interested in measuring or analyzing (e.g. performance), the partial derivative with respect to time (t) is zero and remains so:

$$\frac{\delta p}{\delta t} = 0,$$

for all present and future t.

In discrete time, it means that the first difference of each property is zero and remains so:

$$p_t - p_{t-1} = 0,$$

for all present and future t.

The term steady state is used in several fields and can mean many different things to many different individuals, organizations and environments. We will attempt to define what a steady state means in a DoD production environment. In doing so, we will also consider examples in non-DoD environments in order to demonstrate how and why the steady state occurs in other walks of life.

Production Steady State Causes

While there are several variables and influences within production systems that could cause individual learning to level off, we consider three of the most common.

1. Time/Repetition

This is the most easily understood cause of the production steady state because we all experience this phenomenon in various aspects of our lives. For example, consider commuting to work. Given constant system parameters, we all eventually reach a best case commute time. Assuming we travel to work by car, our system parameters would be as follows:

- Home and workplace location
- Car functionality
- Speed limits
- Lack of construction
- Stop signs/Traffic lights
- Traffic patterns
- Time of day

Assuming these parameters are held constant, the learning we experience would come in the form of identifying the fastest route to take and the improvement is measured in the time it takes us to commute to work form day-to-day. Over time, the best route will be identified and the improvement will eventually cease.

2. Achievement of Quality Thresholds

Another steady state forcing function within production systems is the influence of quality control on the behavior of the system. Up to this point, our discussion has focused solely on the measurement and reduction of direct labor hours from unit-to-unit relative to a defined delivery schedule. However, the majority of projects are also concerned with the quality of the end-items being produced. Quality thresholds and standards can be a major forcing function. When they are not met, cost can increase and schedule can be delayed. Because of this, quality receives quite a bit of attention (and deservedly so).

When production quality standards are not being met, the end-item is often "re-worked". This additional work can either occur at the station where the work content being corrected initially occurred, or, there can be a station at the end of the assembly line where all rework is performed. Either way, additional hours are incurred and recorded for each unit that required rework. As learning and quality increase, the amount of rework decreases and hours required per unit tend to level off. If management sees that the quality standards are being met, less emphasis may be placed on the need to improve efficiency.

3. Physical Space Limitations

The third forcing function for reaching the steady state in a production environment is the limitation of physical space to complete the work. A production manager may decide that if each employee is responsible for completing less work content for each unit, they are likely to increase

their rate of individual learning and cost savings will be realized earlier in the production run. In addition, if there are more employees completing less work content per unit, throughput can be increased.

However, there is certainly an upper bound to this strategy. For instance, management might consider analyzing a station on an assembly line that requires 20 hours of work content per unit that is currently being performed by 5 workers over an 8 hour work day and has a throughput of 2 units per day. The manager might then say, if my 5 workers are each performing 4 hours of work content apiece per unit and I doubled my staff to 10, then I could have them each do 2 hours of work content per unit and double my throughput for the 8 hour shift. This thought process could continue by adding staff and even having multiple shifts. However, the station might eventually get to a point where there is physically not enough room for workers to effectively maneuver and complete their processes without getting in each other's way, effectively slowing the process back down.

The Production Steady State Defined

Hopp and Spearman[5] address the concept of steady state in manufacturing, production or assembly environments using intriguing terminology. First and foremost, they define the steady state as just that – a concept. Secondly, they use the following, two part statement to define steady state that will be impactful to us as cost estimators:

> *"For a system to be in steady state, the parameters of the system must never change and the system must have been operating long enough that the initial conditions no longer matter."*

Production steady state is the point during a production run when the difference between the labor hours required from unit-to-unit is zero and remains unchanged until the end of the production run. It also means that at a certain point the starting parameters of the system no longer matter. Given the mathematical construct of Wright's Cumulative Average theory and its reliance on all data points on the curve until it ends (1 through *n*), a steady state could never truly commence as the cumulative average would always rely on the behavior of the system when it began. Because of this, we will utilize the Crawford's unit curve theory throughout the remainder of this paper.

Now, anyone who has spent a substantial amount of time in production facilities with a low-to-moderate production rate (a typical situation for DoD weapon systems) knows that finding a point where labor hour requirements remain exactly constant until the end of production is next to impossible. This impossibility exists not so much from individual learning ceasing and then starting again, but from the seemingly endless number of variables that can impact low-to-moderate rate environments. Below we identify just a fraction of the issues that can occur at any point of a production run:

- Facility/Equipment/Tooling Issues
- Staffing Irregularities (sick, vacation, etc.)
- Supplier Quality Defects

Instead, we will modify the definition of production steady state to account for the unique nature of the defense production environment:

> *"In weapon system production environments, the steady state commences at unit n when the probability of unit n+1's hours being higher than those required for unit n are equal to the probability of unit n+1's hours being lower than those required for unit n".*

For this to be true, both of these probabilities would be 50%. We define these as follows:

$P_{n+1,h} = P_{n+1,l} = 0.5$, for:

$P_{n+1,h}$ = Probability of Unit n+1 requiring the *same amount or more* direct labor hours than unit n

$P_{n+1,l}$ = Probability of Unit n+1 requiring *same amount or less* direct labor hours than unit n

This definition is critical to us as estimators when attempting to identify and confirm the steady state. Below we look at a plot of direct labor hour requirements for a commercial ground vehicle program (Figure 1) to get a better idea of what a steady state typically looks like.

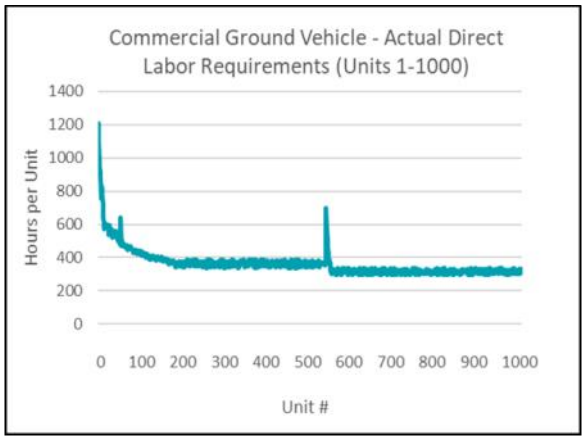


Figure 1

Note how the curve begins to level off at unit 200, albeit with a reasonable amount of variation still occurring from unit-to-unit until we get out past unit 500. Figure 2 is presented to help us explain what is occurring between unit 200 and the point around unit 550 (it is actually unit 539) where the curve spikes back up.

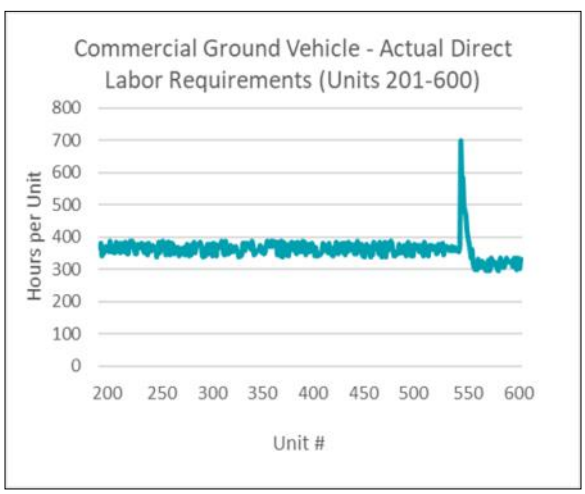


Figure 2

The plot seems to indicate that we are in steady state for these 338 data points. However, it is important to perform statistical analysis and testing to help confirm that observation.

Statistical Analysis and Stationarity Testing

Descriptive statistics for the data (Table 1) tell us that the mean of the 338 data points is 364.16 hours per unit. However, we still see some variance within the data (albeit not much since the coefficient of variation is only 0.039), so we remain uncertain about this being the steady state.

Descriptive Statistics for Units 201-538	
Mean	364.16
Standard Error	0.77
Median	362.50
Mode	363.75
Standard Deviation	14.22
Sample Variance	202.28
Kurtosis	-1.19
Skewness	0.07
Range	50.43
Minimum	339.80
Maximum	390.24
Sum	123085.96
Count	338.00
Confidence Level(95.0%)	1.52

Table 1

Figure 3 gives us a much better graphical representation of how the data is behaving for these 338 units, in revealing that the system appears to be behaving as a stationary process. A stationary process, or system, consists of time-series data that does not have any upward or downward trend or seasonal effects, if applicable. Consequently, the statistical properties of the system, such as mean and variance, also do not change over time.

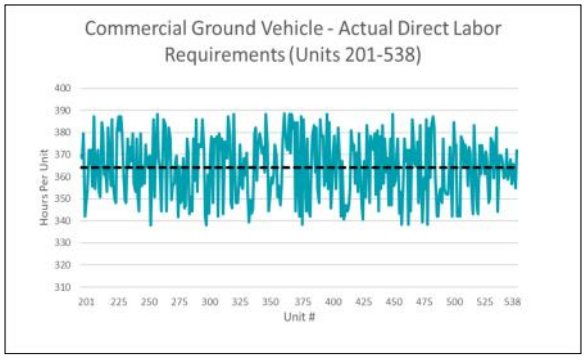


Figure 3

Before getting to a formal statistical test, we can also perform a quick sanity check of the data to see if it meets the definition of a stationary process as defined above. We can quickly check to see if metrics such as the mean and variance stay relatively constant by dividing the dataset into bins. In Table 2 we break the data up into ten (almost) equal sized bins and calculate the mean and variance for each sub-set of data.

While the mean stays relatively constant, we do still notice a fair amount of change in the variance across the bins. So we turn to statistical testing to further support our observation that the system is stationary. One statistical test that can help us determine whether or not the system is stationary, and subsequently whether our production system is in steady state, is the Dickey -Fuller test. The Dickey-Fuller test considers a stochastic process (y_n):

$y_n = fy_{n-1} + e_n$,

where $|f| \leq 1$ and e_n is white noise. If $|f| = 1$, we have what is called a unit root. In particular, if $f = 1$, we have a random walk (without drift), which is not stationary. In fact, if $|f| = 1$, the process is not stationary, while if $|f| < 1$, the process is stationary. We will not consider the case where $|f| > 1$ further since in this case the process is called explosive and increases over time. The null hypothesis for the Dickey-Fuller test is that a unit root is present in a time series sample. The more negative the Dickey-Fuller statistic is, the stronger the rejection of the hypothesis that there is a unit root and the system is stationary:

Null Hypothesis (H_0): If accepted, it suggests the time series has a unit root, meaning it is non-stationary and has some time dependent structure.

Alternative Hypothesis (H_1): The null hypothesis is rejected; it suggests the time series does not have a unit root, meaning it is stationary.

The first step in applying the Dickey-Fuller test is calculating the difference for consecutive data points ($Dy = y_n - y_{n-1}$).

We can use the usual linear regression approach to calculate our Dickey Fuller statistic, except that when the null hypothesis holds, the t coefficient doesn't follow a normal distribution and so we can't use the usual t test, and subsequently, the t tables. Instead, this coefficient follows a tau distribution. Therefore, we are testing to determine whether the tau statistic τ (which is equivalent to the usual t statistic) is less than τ_{crit} based on a table of critical tau statistics values shown in the Dickey-Fuller Table (Table 3).

If the calculated tau value is less than the critical value in the table of critical values, then we have a significant result. Otherwise we accept the null hypothesis that there is a unit root and the time series is not stationary.

	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6	Bin 7	Bin 8	Bin 9	Bin 10
Qty	34	34	34	34	34	34	34	34	34	32
Mean	367.02	364.50	361.69	363.42	368.00	365.32	363.43	363.10	361.69	363.35
Variance	187.22	182.90	211.49	228.38	203.02	258.02	197.80	252.73	212.64	96.73

Table 2

Dickey Fuller Table

N \ α	0.01	0.025	0.05	0.10
25	-3.724	-3.318	-2.986	-2.633
50	-3.568	-3.213	-2.921	-2.599
100	-3.498	-3.164	-2.891	-2.582
250	-3.457	-3.136	-2.873	-2.573
500	-3.443	-3.127	-2.867	-2.570
>500	-3.434	-3.120	-2.863	-2.568

Table 3

We perform regression analysis on the following data set in Excel (Table 4) to determine the t statistic for our test:

$Dy = y_n - y_{n-1}$, for n = 202-538

From Table 4, we see that the t statistic for the coefficient is -18.1263. Comparing this with the tau critical values in Table 3, we can reject the null hypothesis and safely conclude with a high degree of confidence that the system is stationary and in steady state, beginning with unit 201. Before moving on, we end with a couple of notes:

1. One parameter of the analogous system that was not explored was the production schedule and the rate that was needed to fulfill delivery requirements. For simplicity purposes, we assume that the analogous system had a comparable delivery schedule and rate. However, if the rate for the analogous system was substantially different than the future system, it may impact the suitability of utilizing the conclusion that the steady state starts at the 201st unit for future, similar systems.
2. The high level of variance occurring within the system could be driven by something occurring on the assembly line that is driving the peaks and valleys. For instance, there could be one or multiple bottlenecks in the system that are causing disruptions and/or reassignment of resources to keep the line moving. Below we address how finding the steady state can help us in addressing issues such as this.

Regression Statistics	
Multiple R	0.703669
R Square	0.49515
Adjusted R Square	0.493643
Standard Error	14.26107
Observations	337

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	66822.47991	66822.48	328.5629	1.17536E-51
Residual	335	68131.63537	203.378		
Total	336	134954.1153			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	360.689	19.91345053	18.11284	1.33E-51
X Variable 1	-0.99052	0.054645335	-18.1263	1.18E-51

Table 4

Why Should We Care About the Production Steady State?

In order to stress the importance of predicting when the steady state will occur on an estimate, we return to our example involving 1,000 Army vehicles. Based on analysis of production data for five commercial vehicles, we determine that the typical learning rate is approximately 85% and assume this slope for the new vehicle. The data for some of these vehicles indicates the steady state starts around 50 for some and 1000 for others. We decide to analyze how impactful the prediction of our steady state could be in increments between the units of 50 and 1,000. For the purposes of exhibiting the significance, we assume a theoretical first unit value (T1) of 1,000 hours.

We begin by plotting this curve for all 1,000 units with no steady state being reached (Figure 4).

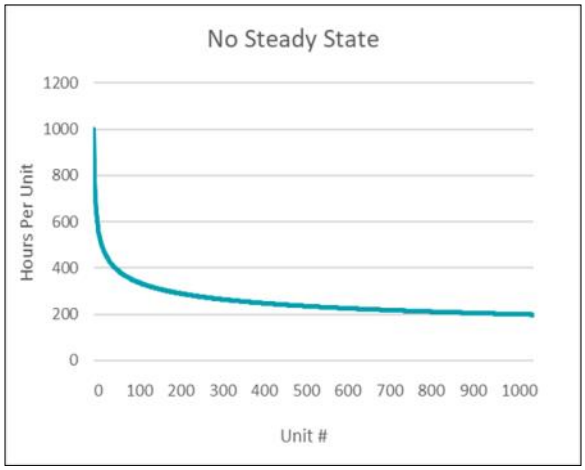


Figure 4

The resulting total hours required for all 1,000 units would be 215,978. We then decide to look at the other extreme – what if our system were to reach a steady state at the 50th unit as is true for at least one of our commercial items? We compare this curve with our curve from Figure 4 in Figure 5:

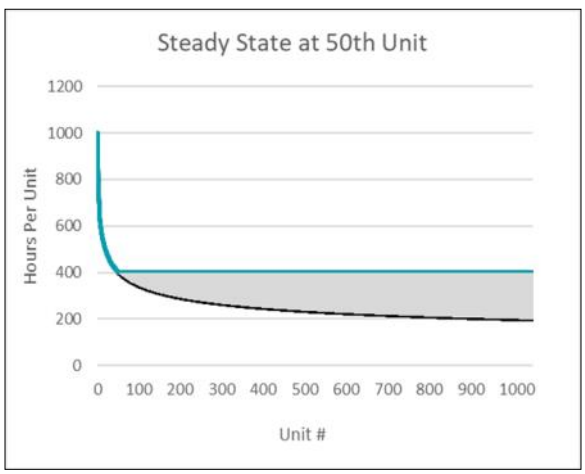


Figure 5

If we assume steady state begins at the 50th unit, our total hours required would increase to 405,155. The gray shaded area in Figure 5 depicts this 57.1% increase. Table 5 provides the sensitivity of total hours to changes in the steady state starting unit.

Clearly, when the steady state is estimated to begin can have a big impact on the direct labor estimate as a whole. If the learning curve slope is estimated to be lower (i.e. our curve is steeper), this impact becomes even more significant.

In addition to impacting the amount of direct hours that are estimated, identifying when the learning curve will happen and at what the direct labor hours will be at that point can provide substantial benefits with regards to how we predict the system will behave. As Hopp and Spearman[3] point out, analyzing a system in steady state, or one that we will assume to be in steady state, can help us in analyzing other key metrics of the system including cycle time, work in process (WIP), bottleneck rates and also help in optimizing the design and layout of the system.

In addition, McCarthy[6] introduced the concept of utilizing the steady state to enhance the analysis and increase the quality of estimates in integrated production environments (i.e. environments where two or more products with at least some common work content are being produced concurrently with the same resources). The concepts presented in that research utilized the identification of the point where the steady state commences to recognize commonality across all end items or any subsets of end items being produced in the integrated environment. The commonality identification and subsequent extraction of common work content enabled inter -product learning curves to be developed and more accurately depict how learning would occur in the environment. By analyzing work content

Steady State Starting Unit	1,000	750	500	250	100	50
Total Hours Required	257,918	259,754	267,905	294,340	349,466	405,155
Difference in Hours Required	N/A	1,836	9,987	36,422	91,548	147,237
% Increase in Hours	N/A	0.7%	3.9%	14.1%	35.5%	57.1%

Table 5

from a static perspective, which is what the steady state provides, the direct labor requirements that were deemed to be duplicative for two or more end items could be extracted and analyzed for anticipated rates of learning separate from end-item unique work content.

A Sample Estimate with the Steady State

Now that we have established the importance of identifying the steady state, we return to the 1,000 Army vehicles described in the sections above on individual and organizational learning. When defining individual learning, we held the number of employees constant and let their rate of learning dictate the delivery schedule. As this is almost never the case, we introduce the following monthly delivery schedule requested by the Army (Table 6).

Month	Units		Month	Units
1	3		13	50
2	5		14	50
3	10		15	50
4	25		16	50
5	35		17	50
6	45		18	50
7	50		19	50
8	50		20	50
9	50		21	50
10	50		22	50
11	50		23	45
12	50		24	32

Table 6

The delivery schedule indicates production ramps up to 50 units per month and stays there from months 7-22. As mentioned above, we will assume that the commercial item used to identify a steady state point of the 201st unit had a comparable schedule and rate. Before estimating direct labor hour requirements we must identify some more characteristics about our system, including:

- Learning Curve Slope
- Budgeted Work Standards

Learning Curve Slope

The learning curve slope for a production environment can easily be estimated by looking at actual data for an analogous system produced the same environment with more or less the same parameters (e.g. workforce, material, and tooling). We again return to the commercial system and, as shown in Figure 6, use the first 200 units of our system (i.e., where it was clear learning was taking place) to identify a representative rate of learning:

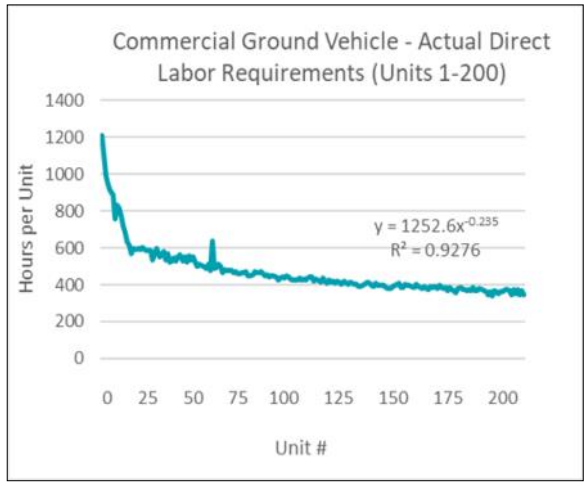


Figure 6

Fitting a power model trend line to the data results in an R^2 value of 0.9276 and model equation of $1252.6x^{-0.235}$. For the purposes of predicting the rate at which we can expect future systems with comparable parameters to learn, we now know that our learning curve slope is $2^{-0.235}$, or, 85.0%.

Budgeted Work Standards

Developing budgeted work standards can be a very beneficial tool in managing a facility and help cost estimators predict future costs. The true definition of what a standard hour means varies

by industry. Some industries set the standard to be "the lower bound" amount of time that an operation should take to complete. Others define the standard as the time an operation *should* take to complete, but operator's performing at greater than 100% efficiency can perform it in less time. Either definition is acceptable, but must be consistently applied. Labor and time standards can be developed using a variety of methods:

- Time and motion studies can be used to develop work standards by measuring how long it takes an operator to complete a specified task or series of tasks. The person performing the time study can then "rate" the operator in terms of the level of efficiency achieved. Multiplying these values and then normalizing for established personal fatigue and delay allowances provides us with the standard.
- Industry established, pre-determined time measurements, such as Methods Time Measurement (MTM) or Maynard Operation Sequence Technique (MOST), break down work content into very specific, measurable motions that have specific times associated with them that are then adjusted for other parameters (e.g. weight lifted, degrees the body will turn during a movement).

Regardless of how budgeted work standards are developed, they can often be re-used from system -to-system based on commonality. However, it is critical that the standards be updated as production proceeds for the new system. For our commercial item in the section above, if our BWS for that system was 330.0 hours per unit and the mean hour requirement in steady state was 364.16, we can infer that our steady state efficiency was 90.6%. For our new system, we have established a BWS of 258.75 total hours per unit for assembly, paint, test and delivery of the new system. Assuming the same steady state efficiency for the DoD environment means we will require 285.6 hours per unit.

Developing the Estimate

Based on the information we have gained from our commercial item data, we can now estimate our direct labor requirements for a system that we expect to reach steady state at the 201st unit and have a direct labor requirement of 285.6 hours per unit from units 201-1,000. For units 1-200, we assume learning will take place at a rate of 85.0%, culminating in the 201st unit requiring 285.6 hours. We compute for our theoretical first unit hours as follows:

$285.6 = T1*201^{(\ln(0.85)/\ln(2))}$,

T1 = 990.3 hours

The resulting learning curve for predicting total direct labor hour requirements (302,543 total hours for 1,000 units) is shown in Figure 7.

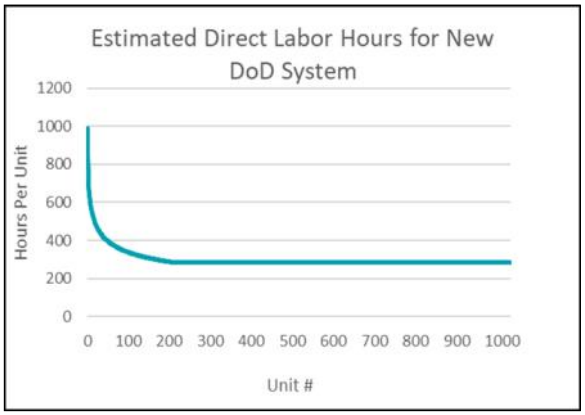


Figure 7

Beware of False Alarms: the Impact of Organizational Learning on the Steady State

Recall from Figures 1 and 2 the large spike that occurred in labor requirements at unit 539 before returning to what appears to be another steady state unit after unit 550. The lead manufacturing engineer for that system indicated that a new machine was integrated into the assembly line that enabled increased throughput at one of the highly staffed stations. The same engineer explained that it took the staff a few days to learn how to operate the machine (hence the spike in

hours), but thereafter less staff were needed at the station due to the new machine's capability. This explains why the system was able to return to a steady state so quickly and why less hours were required. This is a perfect example of production data alerting us to explore the root cause of the data's behavior. A lot of times, this alert is not so evident.

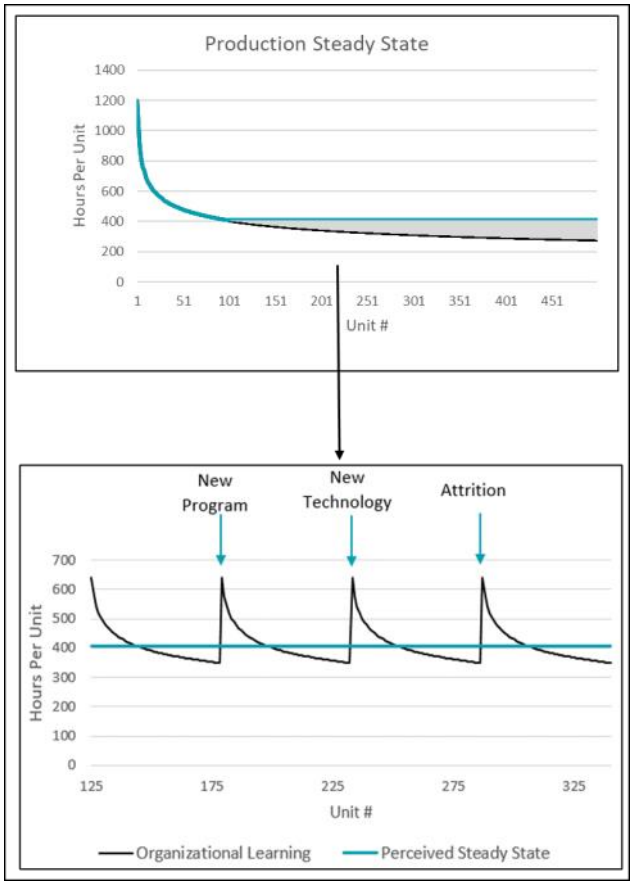


Figure 8

As cost analysts and estimators, we are trained to collect, normalize and analyze data in helping us make sound decisions or develop reliable estimates. However, analysis of direct labor data can pose a unique challenge. Manufacturing and assembly facilities are complex, dynamic environments with many variables at play that can impact our data and potentially mislead or misinform us. These variables can lead us to believe that a production system or environment is behaving one way and that is truly not the case at all. Figure 8 depicts a system that appears to be in steady state. However, the individual learning that is still taking place is being offset by a series of changes impacting the system parameters, leaving the system in a unique state of equilibrium.

Below we discuss several scenarios that can alter the parameters of our system and leave our system experiencing what amounts to a false alarm (i.e. believing that we are in steady state when we are not). The majority of the scenarios relate to what was defined as organizational learning earlier. Whether these scenarios occur by themselves or in conjunction with each other, they can have a substantial impact on what is occurring in a system and, more importantly, impact the data that is recorded for the system.

1. Modifications to Scope

Rarely, if ever, does the configuration of a particular weapon system remain the same during a production run, much less its lifecycle. As the needs of the user for the end item evolve, so too will the configuration of the end item and subsequently the scope and effort required to produce it. Depending on the modification, work content and the direct labor requirements can either increase or decrease. More often than not, the work content will increase due to something that has been learned about the performance, safety, reliability or maintainability of the system.

2. Variable Production Rates

The rate at which end items are built generally varies over a production run. Once a production contract is awarded, a manufacturer will typically start out with a Low Rate Initial Production (LRIP) phase to help the staff ease into the production process in order to track lessons learned and not overload the system with too much staff too early. As more staff become increasingly familiar with the work content they are responsible for, and as the production

process becomes more defined, the amount of expected throughput will increase. In order to do this and meet delivery schedule requirements, the manufacturer will be required to add staff. So long as new staff is being added, there will be individual learning taking place.

3. Business Base Additions/Subtractions

As McCarthy[4] addressed, when dealing with integrated production environments, parameter modifications to other systems could subsequently impact our system or end-item of interest. For example, it is not uncommon for a DoD system to be produced on the same line as other DoD systems or even commercial items that have common work content or operations. Variations in the delivery schedules, and subsequently rates, for other systems could then impact the performance of our system of interest by influencing the number of times an operator accomplishes a certain task where there is commonality. Additionally, if new systems/end-items are introduced to the assembly line or even the facility, the impact could be felt by management reassigning members of our staff to the new program, either for experience or capability purposes, leaving our system parameters modified.

4. New Technology

As production runs evolve, we often learn quite a bit about our system. We learn which workers are most efficient, we learn how to re-order operations in order to enable higher efficiency/ maximize throughput and we also learn about alternative tools, equipment and technology that can improve our system's performance. These upgrades could be the result of either new technology being developed during our run or perhaps the result of cost benefit analysis being performed during our run (i.e. an upgrade to a piece of machinery may initially require training and additional individual learning, but it will eventually double throughput through efficiency gains experienced by the employees or the capability of the machinery itself). Regardless of what inspires management to invest in new technology, the performance and subsequent output of operations impacted by that technology could experience significant variance in data reported.

5. Attrition

Organizations rarely, if ever, experience a production run with the exact same staff from start to finish. Team members get promoted, retire, rotate and leave the organization constantly throughout a production run. Depending on the size of the organization and resource requirements needed for a particular end-item, the impact of staff churn may be negligible, but it may also be quite significant. Simply put, for every person that leaves an organization, so does their individual learning. It is possible that an equivalent amount of learning that has been lost via attrition must be gained by a replacement.

Another phenomenon that occurs in production organizations is bumping, a process used by companies to retain high-valued or longer tenured staff members when downsizing. Typically, the employee being retained "bumps" another employee from their position. Ironically, despite the seniority of the retained employee and their experience within the organization at large, their new assignment may require substantial individual learning. In some cases, the employee doing the bumping may be getting moved to a new role with which they have no familiarity. Small scale bumping likely does not have a large impact. However, mass bumping prompted by a variety of factors (e.g. other programs ending, contracts not being won) would likely have a substantial impact on the performance of a particular production run.

Potential Remedies

As the last section demonstrated, organizational learning can (and will!) occur in DoD production environments. This begs the question - Is it reasonable to assume that individual learning will continue, unimpeded by various organizational learning impacts, long enough to reach a steady state? The short answer to this is yes, but not always. Delivery schedule and production rate is usually the best place to look for this answer. If a new system had a rate of 1.0 unit per day, the chance of organizational learning impacting the system prior to the steady state being reached is much higher than for a rate of 20.0 units per day. To fully explore the reasonableness of a steady state being reached in a future system, analysis of how often and when various cases of organizational learning occurred in analogous systems should be performed.

In order to accomplish this analysis, communication with key team members with direct experience in the analogous systems is critical. For instance, we could talk to the following organizations regarding the type of organizational learning listed:

1. Human Resources: Attrition statistics, including labor category/level of expertise and dates that people left, as well as any bumping due to down-sizing.
2. Industrial Engineering/Production Management: Production rate data, including staffing levels and efficiency reports relative to the BWS at particular times.
3. Manufacturing Engineering: New technology and modifications to scope. As manufacturing engineers typically develop and update work instructions, they represent the most reliable resources in terms of identifying when scope and/or technology took place.
4. Program Management: Business base changes. Plant management will be aware of all programs occurring at a particular facility and to what extent resources were shared between systems.

Conclusions & Recommendations

Throughout this paper, we have explored several facets of the learning and improvement that occur in production environments. We have also identified the significance of the impact that comes from estimating when a system will enter into steady state as well as the criticality of predicting the steady state will occur too early or not at all. Unfortunately, the volatility that occurs within and around the system parameters for DoD production environments makes the likelihood of a system remaining in such a state for an extended period highly unlikely. Moreover, even though we know that parameters are going to change, it will still be next to impossible to predict when those parameters will change and what the subsequent impact on the system will be.

Despite these challenges, all estimators are strongly advised to study the behavior of analogous systems and attempt to identify when a steady state will occur for a particular production environment. Simply assuming that organizational learning will continuously impact individual learning and negate the presence of a steady state can lead to direct labor hours being drastically underestimated.

Our analysis of the commercial system in Figure 1 led us to a three step approach for identifying whether a system is in steady state:

1. Analyze a visual display of the data
2. Divide the data into bins and check for low variance in system parameters across bins
3. Statistical Testing (i.e. Dickey-Fuller Test)

In analyzing the analogous systems, we must stress the importance of not solely relying on production data to determine how future systems will perform. Only by performing root-cause analysis on key system parameters in conjunction with the data analysis will we be able to distinguish system improvement caused by individual learning from improvement driven by

organizational learning. As discussed, a system operating under a constant set of parameters will eventually reach a steady state as a result of individual learning due to either time/repetition, quality thresholds, facility constraints or any combination of these forcing functions.

In order to identify when steady states have commenced in analogous systems, it is critical to account for modifications to system parameters whenever feasible. As cost estimators, it is imperative that we conduct the necessary research and go beyond just data collection and analysis in doing so. By digging deeper into these parameters, we gain an enhanced understanding of how the manufacturing system is behaving, what causes it to behave that way and how that behavior impacts the data. This additional research is what can help us take a good manufacturing cost estimate and turn it into a great one.

References

1. Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the aeronautical sciences*, 3(4), 122-128.
2. Crawford, J. R. (1944). Estimating, Budgeting and Scheduling. Lockheed Aircraft Co., Burbank, California.
3. Konz, Stephan (1995). *Work Design: Industrial Ergonomics (4th ed.)*(pp. 447-460). Scottsdale, AZ: Holcomb Hathaway.
4. *Gagniuc, Paul A. (2017). Markov Chains: From Theory to Implementation and Experimentation. USA, NJ: John Wiley & Sons. pp. 46–59.* ISBN 978-1-119-38755-8.
5. Hopp, Wallace J., Spearman, Mark L. (2000). *Factory Physics (2nd Ed.)*. New York, NY: McGraw-Hill.
6. McCarthy, Patrick. "Don't Get Caught in the Learning Curve Vacuum: Exploring the Impact of Product Commonality on Cost Estimates", presented at the International Cost Estimating and Analysis Association (ICEAA) Professional Development and Training Workshop, Portland, OR, USA, June 6-9, 2017

***Pat McCarthy**, a Program Manager at Technomics' Troy, MI office, has over 19 years of experience in Federal and private industry performing cost analysis and Industrial Engineering studies. He has previously worked as a Cost/Price Analyst and Team Lead with Booz Allen Hamilton and the Army Contracting Command and as a Senior Industrial Engineer and Team Leader with General Dynamics Land Systems. Pat is CCEA and PMP certified and holds B.S. and M.S. degrees in Industrial Engineering from Purdue University.*

Using Dummy Variables in CER Development

Dr Shu-Ping Hu

Alfred Smith, CCEA

Abstract: Dummy variables (also referred to as indicator variables) are commonly used in regression analysis to stratify data into distinct categories. The dummy variable t-test validates the assumption that distinct categories in the data set share the same sensitivity (rate of change) for the ordinary independent variable; the only difference is in the response level. However, many analysts ignore this test when specifying dummy variables in their cost estimating relationships (CER). Consequently, the fit statistics can be misleading. This paper explains when to use dummy variables and how to use them properly when deriving CERs. Specific guidelines are proposed to help analysts determine if dummy variables are appropriate for their data set and common errors analysts experience when applying dummy variables to real examples are explored. The paper also explains how to use the Chow test and dummy variable *t*-test to validate the CER and discusses using dummy variables in splines (to derive the fitted equation as well as the intersection).

Introduction

Background

A dummy variable is used to capture a characteristic that is not directly quantifiable, but exerts an important influence on the behavior of the dependent variable. For example, the cost of high-power amplifiers may vary because some are airborne while others are ground based. For another example, data may be collected by different analysts, or arise from different factories. In such a case, a continuous scale cannot be assigned to the qualitative variable "analyst" or "factory." In other words, within a class of items there may be an attribute that explains the separate effects on the response. These effects can be captured in a regression model by the use of a dummy variable. The dummy variable is simply another variable in the regression except that it can only take on discrete values. In the case of amplifiers that are either airborne or ground based, the values of the dummy variable would only take on one of two values: a zero for airborne amplifiers and a one for ground-based amplifiers or vice versa.

Purpose

The objectives of this paper are threefold. 1) Explain the purpose of using dummy variables and their properties in a regression equation. 2) Identify several common mistakes when using dummy variables in an equation. 3) Describe the Chow test and dummy variable *t*-test, which are used to validate the application of dummy variables. Some general cautionary notes are also recommended. These objectives are illustrated in several examples.

Before specifying dummy variables in a regression equation, a brief review of additive and multiplicative error models is provided.

Additive Error Model

An additive error model can be stated as follows:

$$y_i = f(x_i, \beta) + \epsilon_i = f_i + \epsilon_i \quad (for\ i = 1, \dots, n) \qquad (1)$$

where:

y_i = the observed dependent variable of the ith data point, i = 1 to n

$f(x_i, \beta)$ = f_i = the value of the hypothesized equation at the ith data vector

x_i = the ith data vector of the independent variables

β = the vector of unknown parameters to be estimated by the regression equation

$\mathcal{E}_i$ = the error term with a mean of 0 and a variance σ^2 (assumed to be independent of the explanatory variables)

n = the sample size

Multiplicative Error Model

A multiplicative error term is preferred in the cost analysis field because the error of an individual cost observation is generally proportional to the magnitude of the hypothetical function. A multiplicative error model can be specified as follows:

$$\begin{aligned} y_i &= f(x_i, \beta) * \epsilon_i \\ &= f_i * \epsilon_i \ \ (for\ i = 1, \ldots, n) \end{aligned} \tag{2}$$

The definitions of y_i, $f(x_i, \beta)$, etc. are the same as given in Equation 1. Unlike the additive error model (Equation 1), the standard deviation of the *dependent variable* (e.g., cost) in Equation 2 is proportional to the size of the hypothetical function rather than some fixed amount across the entire data range.

There are three popular methods to fit multiplicative error models: Log-Error, Minimum -Unbiased-Percentage-Error (MUPE) and Minimum-Percentage Error Regression under Zero-Percentage Bias (ZMPE) methods. Both MUPE and ZMPE methods model the CER where the multiplicative error term e is assumed to have a mean of one and a variance s^2. The MUPE method is an Iteratively Reweighted Least Squares (IRLS) regression technique (Hu, 2001; Seber & Wild, 1989; Weisberg 1985; Wedderburn 1974). For a detailed explanation of the ZMPE method, see Book and Lao (1999).

Log-Error Model. If the multiplicative error term (ε_i) in Equation 2 is assumed to follow a log-normal distribution with a mean of zero and a variance of Σ^2 in log space, then the error can be measured by the following:

$$e_i = \ln(\varepsilon_i) = \ln(Y_i) - \ln(f(x_i, \beta)) \tag{3}$$

where ln is the natural logarithm function. In this situation, the objective is to minimize the sum of squared $\mathcal{E}_{is}$ (i.e., $(\Sigma(\ln(e_i))^2)$. If the transformed function is linear in log space, then ordinary least squares (OLS) can be applied in log space to derive a solution for β. In this situation, the CER is termed a log space OLS equation (LOLS) or a log-linear CER. If not, a non-linear regression technique should be applied to derive a solution.

Model Form with a Single Dummy Variable

Linear Model

Consider a linear model using one ordinary independent variable X and one dummy variable D:

$$\begin{aligned} Y &= \alpha + \beta X + \delta D + \theta DX \\ &= \alpha + \beta X + D(\delta + \theta X) \end{aligned} \tag{4}$$

where:

D = 1 if observation n_i is from category #1

D = 0 if observation n_i is from category #2

α, β, δ, θ = coefficients to be estimated by the regression equation

Equation 4 is equivalent to fitting two separate linear equations to the two categories. This specification allows regression of both categories simultaneously. The estimated coefficients derived by this regression model (Equation 4) will be precisely the same as when the two equations are fitted separately. If all the coefficients in Equation 4 are significant at a certain significance level (say 5%), then this implies that the two populations (with and

without the attribute D) behave totally different and they should be estimated by two separate regression equations.

If a regression analysis indicates the coefficient θ is insignificant, then a reduced model can be considered:

$$Y = \alpha + \beta X + \delta D \tag{5}$$

Equation 5 is the usual form when applying a dummy variable. It indicates that these two populations exhibit only a difference in the response level, but share the same sensitivity (rate of change) for the independent variable X.

If coefficient δ is insignificant in Equation 4, a reduced model is given by:

$$Y = \alpha + \beta X + \theta DX = \alpha + X(\beta + \theta D) \tag{6}$$

Equation 6 indicates that two populations have different sensitivity reactions to the relative change in the independent variable X, but share the same fixed cost, which would not be of great interest to us. In other words, if θ is significantly different from zero in Equation 4, then the two populations are statistically different and should be analyzed separately.

Log-Linear Model

The respective log-linear equation form using one ordinary independent variable X and one dummy variable D is given by:

$$\begin{aligned} Y &= \alpha X^{\beta} \delta^{D} X^{\theta D} \\ &= \alpha X^{\beta} e^{\lambda D} X^{\theta D} \end{aligned} \tag{7}$$

Similarly, if a regression analysis indicates the coefficient θ is insignificant, then a reduced model can be considered:

$$Y = \alpha X^{\beta} (\delta)^{D} \tag{8}$$

Similar to Equation 5, Equation 8 is the usual form of applying a dummy variable for log-linear models. It indicates that these two populations exhibit a difference in response levels only. They share the same sensitivity in the exponent for the independent variable X.

However, if the coefficient λ is found to be insignificant in Equation 7 (i.e., δ is not significantly different from one), a reduced model is then given by:

$$Y = \alpha X^{\beta} X^{\theta D} = \alpha X^{\beta + \theta D} \tag{9}$$

Equation 9 indicates that the two populations have a different sensitivity reaction towards the relative change in the independent variable X, but share the same cost at unit one. Just like Equation 6, Equation 9 is also not of great interest to us. Similar to Equation 4, if θ is significantly different from zero in Equation 9, then the two populations are statistically different and should be analyzed separately.

Model Form with Multiple Dummy Variables

The method of Equation 4, as well as Equation 7, can be extended to include more than one dummy variable in the equations. First, ensure the dummy variables are not linearly related among themselves; otherwise, it will result in a singular design matrix. Handle m different responses levels by introducing (m-1) dummy variables. Create the basic allocation pattern for m dummy variables by writing down an (m-1) x (m-1) identity matrix, I_{m-1}, and then adding a row of (m-1) zeros as a comparison baseline:

$$\begin{cases} \begin{array}{ccccc} D_1 & D_2 & D_3 & \dots & D_{m-1} \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{array} & \begin{array}{l} \\ If\ item\ is\ from\ category\ \#1 \\ If\ item\ is\ from\ category\ \#2 \\ If\ item\ is\ from\ category\ \#3 \\ \\ If\ item\ is\ from\ category\ \#m-1 \\ If\ item\ is\ from\ category\ \#m \end{array} \end{cases} \tag{10}$$

See Draper and Smith (1981) for details.
Note that the dummy variable's representation is not unique. There are different ways of choosing dummy variables for a given regression situation.

One common mistake when specifying m different levels is specifying the relative distance between the levels using a discrete variable, e.g., D = 1, 2, …, m, rather than letting the regression equation estimate the separations. The following example demonstrates this common error.

Consider three stratification dummy variables to identify different guidance mechanisms in missile programs:

$$D1 = \begin{cases} 1 & \text{Active radar, but no midcourse (MC) guidance} \\ 0 & \text{Otherwise} \end{cases}$$

$$D2 = \begin{cases} 1 & \text{MC guidance, but no active radar} \\ 0 & \text{Otherwise} \end{cases}$$

$$D3 = \begin{cases} 1 & \text{Both MC guidance and active radar} \\ 0 & \text{Otherwise} \end{cases}$$

Listed below is a basic representation using the above-defined dummy variables:

$$\begin{cases} D_1 & D_2 & D_3 & \\ 1 & 0 & 0 & \text{For active radar} \\ 0 & 1 & 0 & \text{For MC Guidance} \\ 0 & 0 & 1 & \text{For both active radar \& MC guidance} \\ 0 & 0 & 0 & \text{Otherwise} \end{cases} \tag{11}$$

However, the following representation is not the same as the representation given above:

$$\begin{cases} D_1 & D_2 & \\ 1 & 0 & \text{For active radar} \\ 0 & 1 & \text{For MC Guidance} \\ 1 & 1 & \text{For both active radar \& MC guidance} \\ 0 & 0 & \text{Otherwise} \end{cases} \tag{12}$$

Equation 12, which is a common practice for applying dummy variables, does not let the regression equation freely estimate the true level of the response from the category D_3 = 1 (both active radar and MC guidance). It simply assumes the level of D_3 is the product of the levels of D_1 and D_2. It is difficult to evaluate the validity of using dummy variables in Equation 12 and the fit statistics could be misleadingly significant. See McDowell (2012) for illustrative examples of using two dummy variables.

In summary, the representation of dummy variables should:

- account for different levels of responses
- use the regression equation (rather than an assumption) to derive the different levels of response (compare Equation 11 with Equation 12)
- make sure the design matrix is not singular

Chow Test and Dummy Variable *t*-Test

Although most analysts are familiar with the F-test, the Chow test is not as well-known. The Chow test is used for testing the significance of the overall model that includes dummy variables. The F-test and the related F-Statistic are introduced before explaining the Chow test.

F Test for the Overall Model

Consider a linear model with an intercept where the dependent variable Y can be estimated by k independent variables; namely, $X_1, X_2, \ldots, X_k$:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

for i = 1, 2, …, n

This model can be written using matrix notation:

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{13}$$

where:

$\boldsymbol{Y}$ is the n by 1 vector of observations (i.e., the dependent variable),

$\boldsymbol{X}$ is the n by (k+1) design matrix, which consists of the independent variables,

$\boldsymbol{\beta}$ is the (k+1) by 1 vector of unknown coefficients, i.e., $\boldsymbol{\beta} = (\beta_0, \beta_0, \ldots, \beta_k)^t$

$\boldsymbol{\mathcal{E}}$ is the n-by-1 vector of error terms with a variance matrix, $\mathrm{Var}(\boldsymbol{\mathcal{E}}) = \mathbf{V}[\sigma^2]$,

$\mathbf{V}$ is an n-by-n diagonal matrix with the non-negative value v_i in the diagonals (for i = 1, …, n) and zeros elsewhere,

[$\boldsymbol{\sigma}^2$] is used to denote a diagonal matrix where its diagonal element is σ^2, and

n is the sample size.

Note that the matrix $\boldsymbol{V}$ is assumed to be an identity matrix **I** for OLS. The discussion in this paper can be applied to weighted least squares (WLS). OLS is used to demonstrate the use of dummy variables.

The F-Statistic (F-Stat) is used in a hypothesis test to determine whether the overall regression model is significant. It is defined as the ratio of the regression sum of squares to the error sum of squares adjusted by their own degrees of freedom (DF) in the fit space:

$$F - Stat = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE} \tag{14}$$

where *SSR* is the sum of squares due to regression, *SSE* is the error sum of squares, and *k* is the total number of independent variables, not including the intercept. *MSR* is the mean squares due to regression, while *MSE* is the mean squares due to error.

To check the significance of the overall model, the null hypothesis (H_o) is tested against the alternative hypothesis (H_a):

$$H_o: \beta_1 = \beta_2 = \cdots = \beta_k = 0 \quad \text{vs.} \quad H_a: \beta_i \neq 0$$

for at least one slope parameter

Using the vector notations, it is given by:

H_o: $\boldsymbol{\beta} = \mathbf{0}$ vs. H_a: $\boldsymbol{\beta} \neq \mathbf{0}$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots \beta_k)$ not including the intercept.

If H_o is true, the two statistics *SSR* and *SSE* are independent and the F-Stat follows an F distribution with *k* and *n-k-1* DF, respectively, i.e., F-Stat ~ $F(k, n\text{-}k\text{-}1)$. Intuitively, if the model is adequate (i.e., H_o can be rejected), then *SSE* will be small and F-Stat will be large. Therefore, if the F-Stat is greater than or equal to $F_\alpha(k, n\text{-}k\text{-}1)$, it is concluded that there is a significant relationship between the dependent variable and independent variables at a (100α)% significance level. Note that $F_\alpha(k, n\text{-}k\text{-}1)$ denotes the upper (100α)% cut-off point of an F distribution with *k* and *n-k-1* DF, respectively. For a no-intercept model, compare the F-Stat with $F_\alpha(k, n\text{-}k)$ instead of $F_\alpha(k, n\text{-}k\text{-}1)$. The decision rules are summarized below.

Reject H_o:
Model with Intercept: if F-Stat ≥ F_α(k, n-k-1)
Model wo Intercept: if F-Stat ≥ F_α (k, n-k)

Alternatively, the *p*-value for the F-Stat can be used to test the null hypothesis H_o versus H_α:

Reject H_o:
if *p*-value for the F-Stat ≤ α (the significance level of the test)

Chow Test (F Test) for the Overall Model

Given a simple linear model $Y = X\beta + \mathcal{E}$ (see Equation 13), if there are two groups, (A) and (B), in which the parameters are not necessarily the same, the linear model can be rewritten as follows:

$$\begin{cases} Y = X\theta + \varepsilon_1 & \text{for Group (A) with } n_1 \text{ observations} \\ Y = X\gamma + \varepsilon_2 & \text{for Group (B) with } n_2 \text{ observations} \end{cases} \tag{15}$$

Now test the null hypothesis (H_o) against the alternative hypothesis (H_a):

H_o: θ = γ vs. H_a: θ ≠ γ

If the null hypothesis H_o is false, then analyze two regression equations separately as given in Equation 15. Their error sums of squares are denoted by SSE_1 and SSE_2 for Group (A) and Group (B), respectively. The "unrestricted" sum of squares due to error (*USSE*) for Equation 15 is then given by:

$$USSE = SSE_1 + SSE_2 \tag{16}$$

Let *p* denotes the total number of estimated parameters (coefficients) in the equation. If there

are n_1 observations in Group (A) and n_2 observations in Group (B), then the total number of observations is $n = n_1 + n_2$ and *USSE* has $(n_1 - p) + (n_2 - p) = (n - 2p)$ DF.

But if the null hypothesis H_o is true, use a single equation (i.e., Equation 13) to model all the data points. In this case, the *SSE* for Equation 13 is termed the "restricted" sum of squares due to error (*RSSE*), which has $(n - p)$ DF. Intuitively, if the null hypothesis is true, there should **not** be any significant difference between *USSE* and *RSSE*. Consequently, an F statistic for the Chow test is formulated below:

$$F_{Chow\ Test} = \frac{\frac{RSSE - USSE}{p}}{\frac{USSE}{n-2p}} \sim F(p, n-2p) \text{ if } H_o \text{ is true.} \tag{17}$$

The decision rule is as follows:

if $F_{ChowTest} < F_\alpha(p, n - 2p)$, then there is no sample evidence to reject the null hypothesis. On the other hand, if $F_{ChowTest} \geq F_\alpha(p, n - 2p)$, then it is concluded that Groups (A) and (B) respond differently to the relative change in the independent variable *X* at a (100α)% significance level. Note that $p = k + 1$ if there is an intercept in the model; otherwise, $p = k$, where k stands for the number of independent variables.

Dummy Variable t-Test, Individual Parameters

A dummy variable *t*-test is used for testing the significance of individual parameters. Here is an alternative approach to test the following model:

$$Y_i = X_i\beta + D_i X_i \delta + \varepsilon_i \tag{18}$$

where the dummy variable *D* is given by:

$$D_i = \begin{cases} 1 & if\ i \in Group(A) \\ 0 & if\ i \in Group(B) \end{cases}$$

The hypothesis H_o: $\theta = \gamma$ for Equation 15 is the same as the hypothesis Ho: $\delta = 0$ for Equation 18. Since both tests lead to the same conclusion, use either Equation 15 or Equation 18 to test the validity of pooling data from various categories to analyze them together. However, the Chow test (an F-test) is used for testing the significance of the overall model. If the Chow test result is significant, it does not indicate which parameters between the two groups are significantly different. The dummy variable t-test can further examine which specific parameters in both groups are statistically different. As a result, the dummy variable *t*-test (e.g., Equation 18) provides more detailed information than the Chow test.

If there are *m* different groups in the data set, use the F-stat given by Equation 17 to test the null hypothesis with the following:

$$\left.\begin{aligned} &n = \textstyle\sum_{i=1}^{m} n_i \\ &USSE = \textstyle\sum_{i=1}^{m} SSE_i \\ &\text{DF for } USSE = n - m(k+1) \\ &\text{DF for } RSSE = n - (k+1) \end{aligned}\right\} \tag{19}$$

where n_i is the sample size and SSE_i is the error sum of squares for each group, respectively ($i = 1, \ldots, m$). Based upon Equations 17 and 19, an F test statistic for the Chow test is derived accordingly.

The alternative approach (*t*-test) can also be applied to test m different groups in a given data set by including $(m - 1)$ dummy variables. The process is a generalization of Equation 18. See the example section below for using dummy variable *t*-test in a CER.

General Cautions and Statistical Tests When Using Dummy Variables

Some general guidelines and cautionary notes to consider before adding dummy variables to an equation are provided in this section.

Analyze individual groups first.

Examine whether different categories (or groups) should be analyzed by separate regression equations before pooling them together using

dummy variables. Specifically, analyze separate regression equations (by Equation 4 or 7) before choosing a parallel relationship (e.g., Equation 5).

At least three data points for each category

If there are only one or two data points left in a particular category (indicated by a dummy variable, *D*), the *t*-statistic associated with the dummy variable *D* tends to be artificially large and hence misleading. The general rule is to have *at least three* data points in a particular category before using a dummy variable.

Do not use many dummy variables to answer yes/no questions

If there are five categories in the data set, an analyst can create four (4 = 5 – 1) dummy variables to capture the five categories (see Equation 10). However, if a CER contains four dummy variables to answer yes/no questions about the data points, there are actually 16 possible combinations of the four yes/no answers (2^4 = 16). In other words, it creates 16 different categories in the CER. The number of categories can grow rapidly as the number of yes/no questions grows. For example, five dummy variables create 32 (=2^5) categories in a CER; six dummy variables create 64 (=2^6) categories, etc. Analysts should make sure that they have enough observations for the respective regression analysis.

Do not single out specific program.

Dummy variables should not be abused. There can be a temptation to use several dummy variables to account for various aspects of a class of systems to the point where there are no (or few) degrees of freedom left in the overall regression equation. *If a dummy variable is used to capture a single data point in a different level, the regression result is the same as when that point is left out*. Hence, a category of one point is the same as eliminating the point. The general rule is to do *data plotting* and data analyses before using dummy variables.

Examine if all groups have the same variance

The last caution is to ensure that data associated with a particular attribute act no differently from those without it. In other words, the noise term associated with the dependent variable (i.e., cost) should be the same for all items with or without the attributes. F and χ2 tests can be used for testing the equality of the variances of different categories.

If there is only one dummy variable hypothesized in the model, then a simple F-test comparing the mean squared errors (*MSE*) of these two separate regression lines will be adequate

Test H_o: $\sigma_1 = \sigma_2$ vs. H_a: $\sigma_1 \neq \sigma_2$

Test Stat: $F = \frac{MSE_1}{MSE_2}$ $if\ MSE_1 > MSE_2$

Decision Rule:

Reject H_o if $F \geq F_\alpha(df_1, df_2)$ (20)

where $F_\alpha(df_1, df_2)$ indicates the upper (100α)% cut-off point of an F distribution with DF df_1 and df_2, respectively, while df_1 and df_2 are the DF associated with the corresponding *MSE*.

If several dummy variables are used in a regression model, a joint hypothesis of the *equality of several variances* should be considered in addition to the simple F-test (Mood et al., 1974). Dummy variable analysis will be valid when these tests are insignificant.

Demonstration of Dummy Variables in a Spline

In mathematics, a spline is a numeric function that is piecewise-defined by functions such as polynomials (see Wikipedia). In many practical situations, dummy variables can be used to account for two distinct trends occurring in the response data, i.e., segmented lines and splines.

The application of splines can be classified into two categories: (1) it is known which data points lie on which trends and (2) it is not known. This paper only addresses category (1).

Observations	Y	X	Z_1	Z_2
1	y_1	x_1	x_1	0
2	y_2	x_2	x_2	0
…	.	.	.	.
m	y_m	x_m	x_m	0
m+1	y_{m+1}	x_{m+1}	x_0	$x_{m+1} - x_0$
m+2	y_{m+2}	x_{m+2}	x_0	$x_{m+2} - x_0$
…	.	.	.	.
n-1	y_{n-1}	x_{n-1}	x_0	$x_{n-1} - x_0$
n	y_n	x_n	x_0	$x_n - x_0$

Table 1: Dummy Variables Z_1 and Z_2 for Spline (Case 1a)

It is known which data points lie on which trends

If data points (x_1, y_1), (x_2, y_2), ..., and (x_m, y_m) are in one straight line, while data points (x_m+1, y_m+1), ..., and (x_n, y_n) are in another, discuss two subcases: (1a) the intersection of these two lines is a given number between x_m and x_m+1, say x_o, and (1b) the intersection of the two lines is not known and the regression is used to estimate the intersection.

(1a) The intersection of the two lines is at x_0 ($x_m < x0 < x_{m+1}$). In this case, set up two dummy variables Z_1 and Z_2 to take account of the specifications (see Table 1).

Consider the following equation:

$$Y = \beta_o + \beta_1 Z_1 + \beta_2 Z_2 \qquad (21)$$

The regressed estimates should have the following properties:

$\widehat{\beta_o}$ = intercept of line 1

$\widehat{\beta_1}$ = slope of line 1

$\widehat{\beta_2}$ = slope of line 2

(1b) The intersection of the two lines is somewhere between x_m and x_{m+1}. In this case, a third dummy variable D (in addition to Z_1 and Z_2) is created to take care of the unknown point of intersection (see Table 2).

Given a regression line as follows:

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \beta_3 D \qquad (22)$$

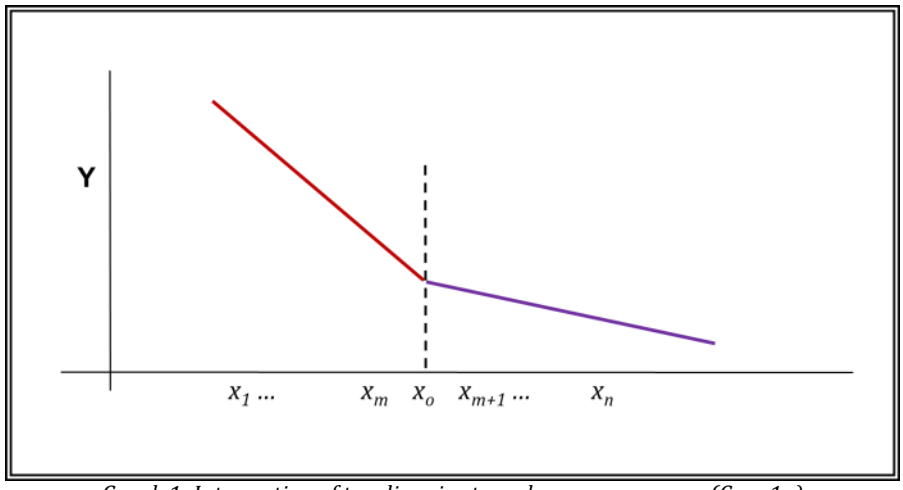
The estimated parameters will have the following interpretations:

$\widehat{\beta_o}$ = intercept of line 1 (same as above)

$\widehat{\beta_1}$ = slope of line 1 (same as above)

$\widehat{\beta_2}$ = slope of line 2 (same as above)

$\widehat{\beta_3}$ = the vertical distance between line 1 and line 2 at the $(m+1)^{th}$ observation



Graph 1: Intersection of two lines is at x_0 where $x_m < x_0 < x_{m+1}$ (Case 1a)

Observations	Y	X	Z_1	Z_2	D
1	y_1	x_1	x_1	0	0
2	y_2	x_2	x_2	0	0
…	.	.	.	.	.
m	y_m	x_m	x_m	0	0
m+1	y_{m+1}	x_{m+1}	x_{m+1}	$x_{m+1} - x_{m+1}$	1
m+2	y_{m+2}	x_{m+2}	x_{m+1}	$x_{m+2} - x_{m+1}$	1
…	.	.	.	.	1
n-1	y_{n-1}	x_{n-1}	x_{m+1}	$x_{n-1} - x_{m+1}$	1
n	y_n	x_n	x_{m+1}	$x_n - x_{m+1}$	1

Table 2: Dummy Variables Z1, Z2 and D for Spline (Case 1b)

The point of intersection can be found by writing both lines in terms of the Z_1 scale. The first fitted line is given by:

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1} Z_1 \tag{23}$$

The second fitted line is given by:

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}(x_{m+1}) + \widehat{\beta_2} Z_2 + \widehat{\beta_3} Z_3 \tag{24}$$

Since $Z_2 = 0$ when $Z_1 = x_{m+1}$,
substitute $Z_2 = Z_1 - x_{m+1}$ into Equation (24):

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}(x_{m+1}) + \widehat{\beta_3} + \widehat{\beta_2}(Z_1 - x_{m+1}) \tag{25}$$

The intersection of the x-axis is then derived using both Equations 23 and 25:

$$Z_1 = (x_{m+1}) + \frac{\widehat{\beta_3}}{\widehat{\beta_1} - \widehat{\beta_2}} \tag{26}$$

For more information about splines, see Ahlberg et al. (1967); Bacon & Watts (1971); Beckman & Cook (1979); Bellman & Roth (1969); Ertel & Fowlkes (1976); Greville (1969).

Example Section

Two sample data sets are used in this section. Several examples are derived using these two data sets to demonstrate some common errors when applying dummy variables in CER development. For illustration purposes, all CERs are generated by the LOLS method so the test results can be easily verified in Excel.

Rocket Propulsion CER

The database is given in Appendix A. Below is a log-linear CER to predict the cumulative average cost for a solid rocket motor:

$$CAC(Q) = 53.27 Q^{-.019} NWlbs^{0.60} NNZ^{0.41} (2.091^{D1})(1.261^{D2}) \tag{27}$$

where:

CAC(Q) = cumulative average unit cost of Q units, FY17$K, no fee

NWlbs = weight of nozzles and thrust vector control hardware

NNZ = number of nozzles

D1, D2 = stratification dummy variables for motor case material, where

$$\begin{cases} D_1 = 1 & D_2 = 0 & \text{if case material is kevlar} \\ D_1 = 0 & D_2 = 1 & \text{if case material is glass} \\ D_1 = 0 & D_2 = 0 & \text{if case material is steel} \end{cases}$$

Note that Equation 27 is fit in **log** space. Equation 27 can be interpreted as a cost improvement curve (CIC) under the disjoint theory. It can also be viewed as a rate curve using the production quantity as the surrogate for rate. The cost improvement (CI) slope (or the rate slope) for Equation 27 is **87.6**% (i.e., $2^{-0.19}$), which is very significant (see the regression output below for details).

Since there are three levels of the motor case material, two dummy variables (D_1 and D_2) are adequate to account for the different levels of

response. As shown by Equation 27, a solid rocket motor made of glass at a given specification (quantity, nozzle weight, number of nozzles) costs 26% more than a rocket motor made of steel at the same specification. Similarly, a rocket motor made of Kevlar on the average costs 109% more than a rocket motor made of steel. Analysts should verify whether these factors are reasonable by engineer's logic. If the regressed coefficients are nonsensical, the fitted equation cannot be accepted regardless of the statistical measures.

Regression Output. Detailed regression outputs for the fit measures, along with the summary predictive measures, are given in Table 3.

Based upon the fit measures, all the regressed coefficients are significant at the 5% significance level (all the *p*-values are less than 0.05). This equation does not have the problem of multicollinearity; no outliers are identified in the report. This CER appears to be a very solid equation.

Average Actual (Avg Act)	1337.80
Standard Error (SE)	372.2712
Root Mean Square (RMS) of % Errors	17.18%
Mean Absolute Deviation (Mad) of % Errors	12.39%
Coef of Variation based on Std Error (SE/Avg Act)	27.83%
Coef of Variation based on MAD Res (MAD Res/Avg Act)	13.28%
Pearson's Correlation Coefficient between Act & Pred	96.97%
Adjusted R-Squared in Unit Space	91.69%

Table 4: Summary of Predictive Measures for Equation 27

However, there is a downside of using dummy variables in this CER. If the data points are analyzed *separately* by their individual material types, the motors made of steel have very little cost improvement (CI) with quantity. Their CI slope is 97% (3% decrease in cost each time the quantity doubles). The motors made of glass have a moderate CI, with a slope of 93%. Most of the CI is, in fact, coming from the **five** motors made of Kevlar and their CI slope is at 61%. This finding demands further investigation (61% slope is rather unusual). Note: this example is simply used to point out the danger of combining different categories by using dummy variables without first analyzing their separate regression equations.

Coefficients Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value
Intercept	3.9753	0.5413		7.3436	0.0000
Qty	-0.1908	0.0654	-0.2636	-2.9152	0.0101
NZ_Wt	0.5978	0.0538	0.6553	11.1215	0.0000
NNZ	0.4139	0.0811	0.3363	5.1020	0.0001
EXP_D1	0.7377	0.1719	0.4083	4.2912	0.0006
EXP_D2	0.2320	0.0980	0.1506	2.3668	0.0308

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.1901	95.42%	93.98%	0.9768

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	P-Value
Regression	5	12.0355	2.4071	66.6130	0.0000
Residual (Error)	16	0.5782	0.0361		
Total	21	12.6137			

Table 3: Fit Measures for Equation 27

Receiver CER

This hypothetical CER is derived from a suite-level Unmanned Space Vehicle Cost Model, Ninth Edition (USCM9) database (Nguyen et al., 2010), but sanitized to retain the desired behaviors while protecting the source of the data. (See Appendix B for the "fake" data set.)

Listed below is a suite-level recurring CER for receivers using two dummy variables:

$$T_1 = 68.65X^{0.83}\,1.48^{EHF}\,1.96^{Gov} \quad (28)$$

where:

T_1 = first unit cost

X = receiver suite weight in pounds

EHF = a dummy variable to indicate if the receiver is operating at Ka-band (*EHF*) or higher

Gov = 1 for government programs, 0 for commercial programs

At first glance, this CER appears to be a solid equation since it is derived by 51 data points with a standard error (SE) in log space of 33%. All the regressed coefficients are significant and the factors for the two dummy variables are also reasonable. Additionally, its Adjusted R^2 is 84% (evaluated in log space), while the Pearson's correlation coefficient between the actual and the predicted value is 0.87 (evaluated in unit space).

As shown by Appendix B, however, there are four categories in this data set: *Gov* = 1, *EHF* = 1; *Gov* = 1, *EHF* = 0; *Gov* = 0, *EHF* = 1; *Gov* = 0, *EHF* = 0. Be sure to use *three* (not two) dummy variables to identify these four categories. Furthermore, four different CERs are given below when analyzing them by their individual categories:

Gov = 1, *EHF* = 1:
$T1 = 608.93X^{0.660}$
(n = 9; SE = 0.28; R^2_{Adj} =0.89) (29)

Gov = 0, *EHF* = 1:
$T1 = 245.3X^{0.678}$
(n =11; SE =0.15; R^2_{Adj} =0.84) (30)

Gov = 1, *EHF* = 0:
$T1 = 69.43X^{0.938}$
(n =13; SE =0.33; R^2_{Adj} =0.90) (31)

Gov = 0, *EHF* = 0:
$T1 = 35.77X^{0.944}$
(n =18; SE =0.32; R^2_{Adj} =0.55) (32)

According to the above equations, there seem to be two different levels of the weight exponent for these four categories: one is at 0.67, versus the other at around 0.94. (The weight exponent 0.83 in Equation 28 behaves like an average of these weight exponents.) In fact, the dummy variable *t*-test shows these two weight exponents to be significantly different. Consequently, this data set should be grouped by the *EHF* dummy variable: one group for *EHF* = 0; the other for *EHF* = 1. In each group, the Gov dummy variable is significant and the CER meets the requirement of using a dummy variable by the *t*-test.

EHF = 1:
$T1 = 271.2X^{0.6634}\,2.206^{Gov}$
(n = 20; SE = 0.21; R^2_{Adj} = 0.88) (33)

EHF = 0:
$T_1 = 36.98\,X^{0.9389}\,1.869^{Gov}$
(n = 31; SE = 0.32; R^2_{Adj} = 0.88) (34)

Chow test and Dummy Variable *t*-test. This receiver data set is used to demonstrate how to use the Chow test and dummy variable *t*-test. Listed below are the *USSE* numbers and sample sizes for the two unrestricted CERs, Equations 29 and 30:

Gov = 1, *EHF* = 1 (Equation 29):
$USSE_1 = 0.5395$; $n_1 = 9$ (35)

Gov = 0, *EHF* = 1 (Equation 30):
$USSE_2 = 0.1953$; $n_2 = 11$ (36)

If Equations 29 and 30 are combined into a restricted model, Equation 37 is derived:

EHF = 1:
$T1 = 1642.54X^{0.4275}$
($RSSE$ = 2.5145, R^2_{Adj} = 0.61) (37)

Equation 38 is derived when using the Gov dummy variable to combine Equations 29 and 30 into one CER:

EHF = 1:
$T1 = 271.16X^{0.663}\,2.206^{Gov}$
($RSSE$ = 0.7355, R^2_{Adj} = 0.88) (38)

The test statistic for the Chow test is then given by:

$$F_{Chow\ Test} = \frac{\frac{RSSE - USSE}{p}}{\frac{USSE}{n-2p}} = \frac{\frac{2.5145 - 0.5395 - 0.1953}{2}}{\frac{0.5395 + 0.1953}{20-4}} = 19.4 \quad (39)$$

Since the test statistic $F_{ChowTest}$ is greater than $F_{0.01}$ (2, 16) = 6.23 , it is concluded that there is a significant difference between the government and commercial programs at the 1% level. However, the Chow test (an F-test) does not indicate which parameters (slope, scale, or both) are significantly different between these two groups.

On the other hand, the dummy variable *t*-test can be used to further examine whether some specific parameters (coefficients) in both groups are statistically different. Given below is a full model using the dummy variable on both the scale and exponent coefficients:

EHF = 1:

$$T1 = 245.3X^{0.678}X)^{-0.018Gov}\ 2.482^{Gov} \quad (40)$$

Based upon the dummy variable *t*-test, the exponent -0.018 (which captures the weight difference between the government and commercial programs) is not significant because its *t*-ratio is only -0.12.

Since no significant difference is found between the weight exponents of these two groups, use the Gov dummy variable to combine Equations 29 and 30 into one equation (i.e., Equation 38). Note that the Coefficient 2.206 in Equation 38 is significant.

Similarly, for the government programs (*Gov* = 1), it can be shown that both the exponent and scale parameters associated with the *EHF* variable are significant using the dummy variable *t*-test (as their *p*-values are less than 0.05):

Gov = 1:

$$T1 = 69.43X^{0.938}\ X^{-0.278EHF}\ 8.77^{EHF} \quad (41)$$

Consequently, the two groups, *EHF* = 1 and *EHF* = 0, should be analyzed separately; namely, they should not be pooled together using a dummy variable.

Conclusions

Analysts should consider general guidelines before adding dummy variables to an equation. The main purpose of using dummy variables is to conserve DF for small sample analysis. However, the full model hypothesis should be tested before using the reduced model. Besides checking the fit measures of the regressed coefficients, analysts should run appropriate tests first to determine the relevance of applying dummy variables to their equations. Listed below are a few basic rules for using dummy variables in CER development:

1. Analyze individual groups first. Examine whether different groups (or categories) should be analyzed by separate regression equations before pooling them together using dummy variables. To be more specific, analyze separate regression equations (e.g., Equations 4 and 7) before choosing a reduced model (e.g., Equations 5 and 8).

2. Use Chow test and dummy variable *t*-test to determine whether a reduced model is appropriate.

3. Use (*m*-1) dummy variables to specify *m* different groups. In addition, do not specify the relative distance between the group levels using a discrete variable, e.g., *D* = 1, 2,…, *m*. Instead, let the regression equation estimate the separations.

4. Use the rule of three points. If there are only one or two data points left in a particular category (indicated by a dummy variable, *D*), the *t* -statistic on the slope or exponent coefficient of the dummy variable *D* tends to be artificially large and hence misleading. The general rule is to have at least three data points in a particular category before using a dummy variable.

5. Do not single out a specific program. It can be tempting to use several dummy variables to account for various aspects of a class of systems to the point where there are no (or few) degrees of freedom left in the overall regression equation. If a dummy variable is used to capture a single data point at a different level, the regression result is the same as when that point is left out.

6. Check whether all groups have the same variance to ensure that data associated with a particular attribute act no differently from those without it. In other words, the noise term associated with the dependent variable (i.e., cost) should be the same for all items with or without the attributes. F and χ^2 tests can be used to check the equality of the noise band (i.e., variance) of the dependent variable (Mood et al., 1974).

7. Select dummy variables by engineer's logic. Dummy variables based upon sound logic and solid technical grounds are more likely to have merit. For example, the dummy variables chosen in USCM9, such as "communication mission" (yes or no), "agency type" (1 = government program, 0 = commercial program), etc. are based upon engineer's logic, so they have practical meaning. Selecting dummy variables by engineer's judgement is as important as the statistical considerations in CER development.

Finally, dummy variables can be used to find the intersection between two lines (splines). This can be a useful application in cost improvement curve (CIC) analysis. For example, in a CIC data set, if the first few data points appear to follow one CIC slope, while the remainder follows another CIC slope, use dummy variables to model the two distinct trends.

Appendix A: Solid Rocket Motor Data Set

Data Point	CAC$K	Quantity	Nozzle Weight	Number of Nozzles	D_1	D_2
Obs 1	1,411.7	2,249	948.0	4	0	0
Obs 2	951.7	925	390.0	4	0	0
Obs 3	1,025.4	1,324	350.0	4	0	1
Obs 4	670.7	1,547	169.0	4	0	1
Obs 5	520.0	698	227.0	1	0	1
Obs 6	1,241.8	350	604.0	4	0	0
Obs 7	1,077.5	350	309.0	4	0	1
Obs 8	1,802.6	667	1,440.0	4	0	1
Obs 9	901.9	667	172.0	4	0	1
Obs 10	993.6	547	761.0	1	0	1
Obs 11	957.4	547	424.0	1	0	1
Obs 12	4,248.1	71	1,535.0	1	1	0
Obs 13	5,084.4	103	1,485.0	2	1	0
Obs 14	3,693.8	71	479.0	2	1	0
Obs 15	635.6	85	176.0	1	0	1
Obs 16	209.4	524	92.5	1	0	0
Obs 17	286.2	546	114.0	1	0	0
Obs 18	733.7	184	157.2	1	1	0
Obs 19	603.0	184	151.0	1	1	0
Obs 20	734.1	1500	520.0	2	0	0
Obs 21	1,112.5	1230	750.0	3	0	0
Obs 22	536.6	1680	256.0	2	0	0

Appendix B: Receiver Data Set

Observation	T1	X (Weight)	EHF	Gov
Obs 1	6,600.21	254.37	0	1
Obs 2	1,424.00	28.26	0	1
Obs 3	25,364.46	782.09	0	0
Obs 4	28,902.57	685.42	0	0
Obs 5	11,084.69	737.25	0	0
Obs 6	17,456.22	628.53	0	0
Obs 7	18,174.66	791.46	0	0
Obs 8	24,701.53	358.18	0	1
Obs 9	5,320.50	122.18	0	1
Obs 10	7,826.23	204.68	0	1
Obs 11	2,764.87	43.69	0	1
Obs 12	45,021.55	1,184.43	0	0
Obs 13	19,083.38	652.19	0	0
Obs 14	8,172.09	39.39	1	1
Obs 15	57,801.60	621.18	1	1
Obs 16	1,957.13	29.80	0	1
Obs 17	23,130.17	359.39	0	1
Obs 18	18,262.27	345.47	0	1
Obs 19	26,415.75	348.59	0	1
Obs 20	7,993.50	120.96	0	1
Obs 21	16,727.47	791.46	0	0
Obs 22	63,784.22	2,410.84	0	0
Obs 23	9,289.77	654.11	0	0
Obs 24	25,737.49	1,162.01	0	0
Obs 25	17,697.46	1,067.34	0	0
Obs 26	15,631.43	934.49	0	0
Obs 27	2,251.56	49.04	0	1
Obs 28	20,497.51	637.93	0	1
Obs 29	22,645.97	888.16	0	0
Obs 30	25,812.86	920.00	0	0
Obs 31	16,975.38	533.64	1	0
Obs 32	36,001.45	1,676.22	1	0
Obs 33	21,145.31	618.80	1	0
Obs 34	7,677.11	38.36	1	1
Obs 35	12,051.18	359.50	0	0
Obs 36	15,607.81	737.75	0	0
Obs 37	11,138.75	209.80	1	1
Obs 38	38,767.66	548.44	1	1
Obs 39	41,176.09	566.80	1	1
Obs 40	11,228.76	93.08	1	1
Obs 41	33,248.99	1,228.50	1	0
Obs 42	28,903.69	1,035.00	0	0
Obs 43	20,381.97	957.30	1	0
Obs 44	50,546.40	2,539.59	1	0
Obs 45	27,160.39	713.67	1	0
Obs 46	13,891.36	522.49	1	0
Obs 47	20,687.47	680.32	1	0
Obs 48	18,438.14	173.89	1	1
Obs 49	51,652.59	752.67	1	1
Obs 50	20,834.76	752.22	1	0
Obs 51	22,756.41	678.87	1	0

References

1. Ahlberg, J. H., Nilson, E. N., & Walsh, J. L. (1967). *The Theory of Splines and Their Application*. New York: Academic Press.
2. Bacon, D. W., & Watts, D. G. (1971). *Estimating the Transition between Two Intersecting Straight Lines*. Biometrika, pages 58, 525-54.
3. Beckman, R. J., & Cook, R. D. (1979). *Testing for Two-Phase Regression*. Technometrics, pages 21, 65-69.
4. Bellman, R., & Roth, R. (1969). *Curve Fitting by Segmented Straight Lines*. J. Am. Statist. Assoc., pages 64, 1079 -1084.
5. Book, S. A., & Lao, N. Y. (1999, November). *Minimum-percentage-error regression under zero-bias constraints*. Proceedings of the 4th Annual U.S. Army Conference on Applied Statistics, U.S. Army Research Laboratory, Report No. ARL-SR-84, pp. 47–56.
6. Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York, NY: John Wiley & Sons.
7. Ertel, J. E., & Fowlkes, E. B. (1976). *Some Algorithms for Linear Spline and Piecewise Multiple Linear Regression*. J. Am. Statist. Assoc., pages 71, 640-648.
8. Greville, T. N. E., (1969). *Theory and Applications of Spline Functions*. New York, NY: Academic Press.
9. Hu, S. (2001, June 12-15). *The minimum-unbiased-percentage-error (MUPE) method in CER development*. Paper presented at the 3rd Joint Annual ISPA/SCEA International Conference, Vienna, VA.
10. McDowell, J. (2012, September 17-19). *Pooled Regression and Dummy Variables in CO$TAT*. Paper presented at the 2012 ACEIT User Workshop, McLean, VA.
11. Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. McGraw-Hill.
12. Nguyen, P., Kwok, B. et al., (2010, August). *Unmanned Spacecraft Cost Model, Ninth Edition*. U. S. Air Force Space and Missile Systems Center (SMC/FMC), Los Angeles AFB, CA.
13. Seber, G. A. F., & Wild, C. J. (1989). *Nonlinear Regression*. New York, NY: John Wiley & Sons, pages 37, 46, 86-88.
14. Wedderburn, R. W. M. (1974, December). *Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method*. Biometrika, Vol. 61, Issue 3, pages 439-447.
15. Weisberg, S. (1985). *Applied Linear Regression* (2nd ed.). New York, NY: John Wiley & Sons, pages 87-88.

Ph.D. in Mathematics, emphasis in Statistics. ***Shu-Ping Hu*** *retired as the Chief Statistician of Tecolote Research, Inc. She has over 30 years cost analysis experience and has published many technical papers, including the PING Factor (adjust log-linear CER result to reflect the mean), adjusted R-square for MUPE and ZMPE CERs, and Generalized Degrees of Freedom for the constrained minimization process. Shu-Ping is a recipient of the Frank Freiman Lifetime Achievement Award.*

Bachelor in Mechanical Engineering, Master's of Science in Naval Architecture. ***Alfred Smith*** *retired as a General Manager with Tecolote Research, Inc. He managed the development, distribution, and training of tools such as ACE, CO$TAT, RI$K, and JACS. He was the contractor lead for handbooks, including Joint Agency Cost Schedule Risk and Uncertainty (JA CSRUH), JA CER Development, and DoD Cost Estimating Guide. Alfred is a recipient of the Frank Freiman Lifetime Achievement Award.*

Improvements on the Development of Correlated Input Variables for Monte Carlo Simulation

Douglas Henke

Abstract: In this paper, the development of model inputs with specified correlation is explored. Input variable correlation is an influential driver of final cost risk distributions; especially so for highly positively or negatively correlated inputs. If not captured appropriately between inputs, significant errors in resultant cost risk distribution will occur. In general, the further away from a 50% confidence level cost value, the greater the error will be when input data does not reflect accurate correlation. The widely adopted Iman-Conover (IC) Method for inducing desired rank correlation on a multivariate input for modeling by Monte Carlo simulation is reviewed. The IC method culminates in the re-ordering of the values of each input vector such that the resultant correlation of the vectors is close to the desired correlation. This paper provides insights into how the IC Method, devised as a method for inducing a desired rank correlation, can be equally if not more powerful for inducing desired Pearson product-moment (linear) correlation on inputs. Spearman's rank correlation and linear correlation values that result from the IC Method are compared to the desired correlation values ranging from -1 to 1. Insights into the mechanics of the algorithm are presented in order to provide a richer understanding of the process and to inform aspects of work when the algorithm is employed. Extending this IC method to an iterative process described in this paper shows that the resulting set of variates would more accurately reflect the desired correlation in all cases for the calculated linear correlation. Conversely, for highly skewed distributions, the iteration process resulted in increasing the error of the calculated Spearman's rank correlation. The iteration process is explained with examples to illustrate improved linear correlation accuracy for both symmetric and highly skewed distributions.

Most cost risk and statistical software platforms on the market today permit the user to define the distribution type of each input variable as well as the correlation between these variables. Common forms of quantifying correlation include Pearson product-moment (linear) correlation and Spearman's rank correlation. Using linear correlation with input variables that have outlier or clustered data; or have unusual forms of distribution may not appropriately quantify the relationship between variables (Tamhane, 2000). For those input relationships that are not monotonic or which depart significantly from a linear relationship, Spearman's rank correlation metric may be the more appropriate measure. Accurately quantifying the correlation of input variables of the phenomenon, stochastic process or estimate being modeled is a well-recognized and necessary element of increasing the accuracy and realism of resultant risk distributions and cost estimating results. Inaccurate correlation values exacerbate cost risk distribution errors that are further away from 50% confidence level. In the simplest terms, the final cost risk variance is inflated or contracted with inaccurate input correlation. It is important to note that the IC method culminates in the reordering of the original set of input variates based on the rank order of a resultant linear transformation they undergo. This linear transformation is based on the desired correlation matrix of the inputs; which may be defined as either linear or rank correlation. In either case, the efficacy of the IC method is considered in this paper by how close the resultant linear and rank correlation is to the specified desired correlation. The ability to

perform this work independent of embedded software features can be useful; as well as beneficial to understanding any limitations of the process that apply in practice.

Richard L. Iman and W. J. Conover published a paper in 1982 entitled *A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables*. Consider, for example, M input vectors to represent M variables in a cost model. Each of the M vectors contain N values for a model run to be performed with N iterations. Assume the correlation between all M input vectors is known. Iman and Conover devised a powerful and effective methodology that culminates in the reordering of values of all but the first input vectors such that the resulting correlation of these reordered vectors is very close to the desired correlation. Once reordered, these vectors are then used as appropriate inputs.

The following summarizes the steps of the IC Method. Of note, Stephan J. Mildenhall provides an expanded treatment of the IC Method in his paper, *Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method* presented in 2005.

Assume a model requires M input variables; each with N values for a model run of N iterations. Let [A] represent the set (or matrix) of input variables of size N (rows) by M (columns). For illustrating the IC method, the distribution type of each input is not significant. In this paper, the inverse of the cumulative distribution function of each input variable is known. This is used to determine the elements of each input vector as van der Waerden scores where the ith element of the jth variable in [A] is initially determined as follows.

$$A_{i,j} = \Phi_j^{-1}\left(\frac{i}{N+1}\right)$$

The desired correlation matrix must be positive semi-definite to enable a tractable decomposition. Define the desired input vector correlation matrix as [S]. It is of dimension M by M. Zero-mean and scale each of the input vectors such that the variance is 1. This transforms [A] to what is now defined as [X]. At this point, each vector is ordered from lowest to highest rank by virtue of their derivation as van der Waerden scores. The IC method requires linear independence of the input vectors. To invoke this linear independence, randomly permute each of the input vectors. This is not necessary for the first vector as it is unaffected by the IC method. Now define a new matrix [E] as the covariance of [X].

$$[E] = \frac{1}{N} \cdot [X]^T [X]$$

Since the vectors of [X] have zero mean and variance of 1, [E] represents the linear correlation matrix of [X] and should have low absolute values off-diagonal due to the prior random permutation of the vectors that induced linear independence. The next step is to apply a Cholesky Decomposition to [E].

$$[E] = [F]^T [F]$$

[F] is an upper triangular matrix and represents the square root of [E]. As a square symmetric matrix, [E] can be transformed to $[L][D][L]^T$ where $[L][D^{(.5)}]$ is the lower triangular matrix of the decomposition and equals $[F]^T$. Cholesky decomposition would fail with any of the diagonals of [D] less than zero (i.e. Not positive semi-definite). A significant facet of this decomposition can be shown as follows:

$$[F]^{-T}[F]^T[F][F]^{-1} = [I]$$

$$\frac{1}{N} \cdot [F]^{-T}[X]^T[X][F]^{-1} = [I]$$

This shows that pre-multiplying the covariance matrix of [X] by the inverse of the $[F]^T$ and post-multiplying by the inverse of [F] give the identity matrix.

To incorporate the desired correlation, a similar Cholesky decomposition is performed on the desired correlation matrix [S] where

$$[S] = [C]^T[C]$$

Now consider inserting the identity matrix between the two [C] matrices without effect.

$$[S] = [C]^T[I][C]$$

It was shown above that pre and post multiplying $[X]^T[X]/N$ by the inverses of the Cholesky decomposition of [X] results in the identity matrix, [I]. Replacing [I] in the equation above with that identity gives the following.

$$[S] = \frac{1}{N} \cdot [C]^T[F]^{-T}[X]^T[X][F]^{-1}[C]$$

This shows us that multiplying the randomly permuted starting vectors of [X] by $[F]^{-1}[C]$ results in a set of vectors with a calculated linear correlation that exactly matches the desired correlation, [S].

For convenience, the new transform matrix [T] is defined as follows.

$[T] = [F]^{-1}[C]$ and define $[X'] = [X][T]$, then

$$\frac{1}{N} \cdot [T]^T[X]^T[X][T] = \frac{1}{N} \cdot [X']^T[X'] = [S]$$

However, since each vector (column) of [X'] is a linear combination of the vectors of [X], the original distributions are no longer preserved. That is, the desired correlation is achieved, but no longer with the original values that comprised the vectors of [X]. It is noteworthy that the variance of each vector of [X'] is 1 since the covariance of the [X'] matrix is [S].

At this point, Iman and Conover capitalize on the connected relationship between Spearman's Rank correlation and linear correlation where the rank correlation is the linear correlation of the ranks of the values that comprise the vectors. As the next step in the IC method, the elements of each of the variates of [X] are reordered to have the same rank ordering of the corresponding vectors of [X']. This generally results in a correlation of the re-ordered vectors that is close to the desired correlation prescribed in [S]. Since these data were zero-meaned and rescaled to have a variance of 1 for the IC method, one needs to simply reverse the process to regain the original distribution that is now re-ordered for the desired correlation.

By way of example, consider a matrix [A] that contains three input vectors of 30 elements each (not shown). Each are sampled as van der Waerden scores from a normal distribution with zero mean and are normalized to have a variance of 1. With a zero mean and variance of 1, [A] becomes [X], and is of dimension 30 X 3. In this case, each input vector of [X] is identically distributed.

The second and third vector are randomly permuted so that columns of [X] are linearly independent. Once accomplished, determine the linear correlation matrix of [X], defined as [E].

Permuted $[X]^T[X]/N$ = Correlation = [E]		
1	0.014629	-0.13703
0.014629	1	0.144805
-0.13703	0.144805	1

The upper Cholesky decomposition of this correlation matrix, [E], and its inverse are as follows.

[F]		
1	0.014629	-0.13703
0	0.999893	0.146825
0	0	0.979624

$[F]^{-1}$		
1	-0.01463	0.1420773
0	1.000107	-0.149895
0	0	1.0207994

A positive-definite desired correlation matrix, [S], is developed for this example and its upper Cholesky decomposition are shown as follows.

[S]		
1	-0.7	0.5
-0.7	1	-0.2
0.5	-0.2	1

[C]		
1	-0.7	0.5
0	0.7141428	0.210042
0	0	0.8401681

We now have the information to produce the transform matrix [T] where $[T] = [F]^{-1}[C]$.

[T]		
1	-0.71045	0.616296
0	0.714219	0.084128
0	0	0.857643

The next step is to multiply [X] by [T]; resulting in [X']. The resulting covariance matrix of [X'] ($1/N * [X']^{T}[X']$) now equates to [S] precisely.

$[X']^{T}[X'] /N = [S]$		
1	-0.7	0.5
-0.7	1	-0.2
0.5	-0.2	1

As discussed before, [X'] is a linear combination of the vectors of [X] and all but the first vectors of [X'] now have a different distribution than that of the [X] vectors. Per the IC method, each vector of [X] is reordered to have the same rank order as the corresponding vector of [X'].

The following shows the Resultant linear and Spearman's rank correlation matrices based on the reordered vectors of [X]:

Spearman Rank Correlation		
1	-0.74	0.44
-0.74	1	-0.25
0.44	-0.25	1

Pearson Linear Correlation		
1	-0.72	0.53
-0.72	1	-0.23
0.53	-0.23	1

The resultant rank correlation of [X'], and by process, the rank correlation of the reordered vectors of [X] are close to, but do not match the desired correlation, [S], precisely. Similarly, the linear correlation of the reordered elements of [X] is close to, but does not match [S] precisely. As will be seen, larger values of N, which are typically employed in practice, result in much closer alignment with the desired correlation matrix.

Exploring the IC Method

In the following, the IC method is applied to two-vector input matrices (i.e. two input variables) of various sizes (N). This facilitates a more tractable analysis and the concepts apply to M dimensional input matrices. The difference between the desired correlation as prescribed in [S] and resultant correlation (both linear and rank correlation) from the reordered elements of [X] will be considered; as well as means to reduce this difference.

A 4 X 2 input matrix is considered first. From a Euclidean space construct, [X] spans 4 dimensions (N) and is of rank 2 (M=2); provided the vectors are linearly independent. There are 4 factorial (4! or 24) different permutations of the second vector that is reordered in the process. When the values of each input vector are viewed as coordinates, each of the permutations of the second vector occupies a discrete point in N-dimensional space and each also has its own

correlation value with the first vector. In this case of input vectors, the first is derived from a normal distribution and second is a skewed distribution. There are 24 unique linear correlation values and 11 unique rank correlation values with the first vector from all possible permutations of the second vector. These permutations represent a constellation, as it were, of 24 discrete points at a Euclidean distance of the square root of N from the origin due to a variance of 1.

Consider a randomly permuted second vector of [X] and the notion that the IC method can be applied for any desired correlation value between the vectors of [X]. Recall that [X'] = [X][T] gives the exact desired correlation, but no longer possesses the original distribution of values. There is a continuum of the second vectors of [X'] associated with each value within the range of desired correlation values from -1 to 1 since. The second vector of [X'] is a linear combination of the two vectors of [X] as determined by [T]. This continuum of [X'] spans a plane (or arc) of the continuous N dimensional space subtended by the vectors of [X]. This subspace is referred to herein as the **[X] subspace**. In the last step of the IC method, the re-ordered second vector of [X] becomes one of the 24 possible permutations (and one of the 24 possible linear correlations) that is near the [X] subspace because the re-ordering is based on the rank of the second vector of [X']. This does not allow for a permutation that may result in a correlation value closer to the desired correlation but is in a region away from the arc of the [X] subspace. Accordingly, the initial random permutation of the vectors of [X] pre-determine the [X] subspace and possible resultant correlation values.

Similar to the vectors of [X] and [X'], consider now all other possible points in the continuum of N dimensional space that are a distance from the origin of square root of N and whose mean is zero. This continuous collection of points is comprised of bounded and connected regions of all possible coordinate rank orders. Since there are N factorial possible permutations of a vector (or set of coordinates) of dimension N, there are N factorial regions whose coordinates have the same rank order. Each of these regions is referred to as a **rank order region**. Each rank order region is defined by the rank order of the values in the vector that map to that region. For example, where N = 4, the vector comprised of the following values in the order shown is 1423, {-0.780, 1.724, -0.575, -0.439}. Provided that all values are unique in a vector of [X], each permutation resides as a discrete point within each of these regions. The IC method determines in which regions the re-ordered vectors of [X] reside by virtue of the rank order of vectors in [X'].

To illustrate these concepts, principal component analysis was employed for the dimensional reduction of all possible permutations of the input vector that spans 4 (N) dimensional space. The points were mapped to 3 dimensions and are shown below from a perspective angle with the background walls and floors aligning with the Cartesian coordinates. There was no loss of information in the dimensional reduction because each vector has a mean of zero and variance of 1. Hence, each vector represented as point in 3-D space is equidistant from the origin. Further, all possible vectors of N equals 4, of zero mean and variance of 1 occupy the surface of a sphere of radius 2 in this 3-D space. In general, all possible real numbered vectors of dimension N with zero mean and constant variance will occupy an N-2 dimensional subspace.

Figure 1 below is a graphic that shows the boundaries of all 24 rank order regions. Figure 2 shows only the front facing rank order regions for illustrative purposes. The points shown represent permutations of a single vector. In this example, the vector is {-0.710, -0.575, -0.439, 1.724}. Of note, these permutations derived from a skewed distribution are located near a corner of each region. Each of the 24 permutations occupies one point in each of the rank order regions.

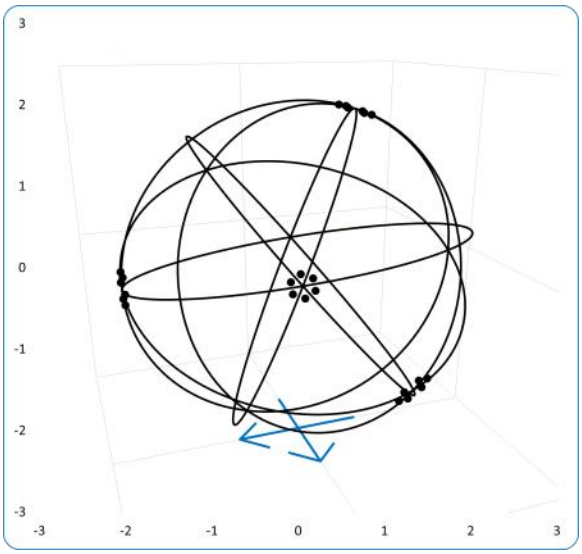


Figure 1

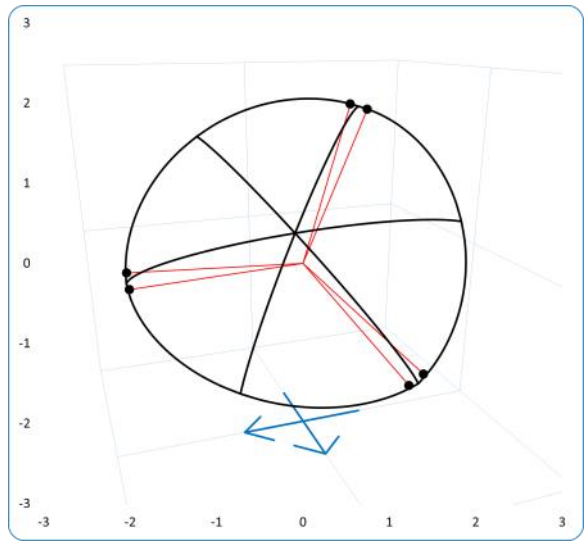


Figure 2

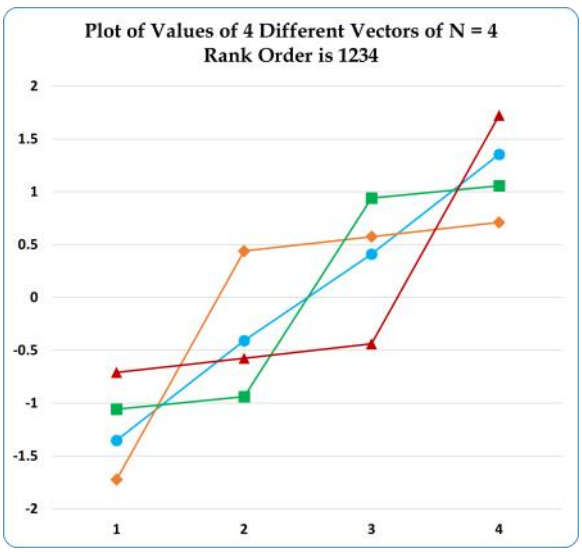


Figure 3

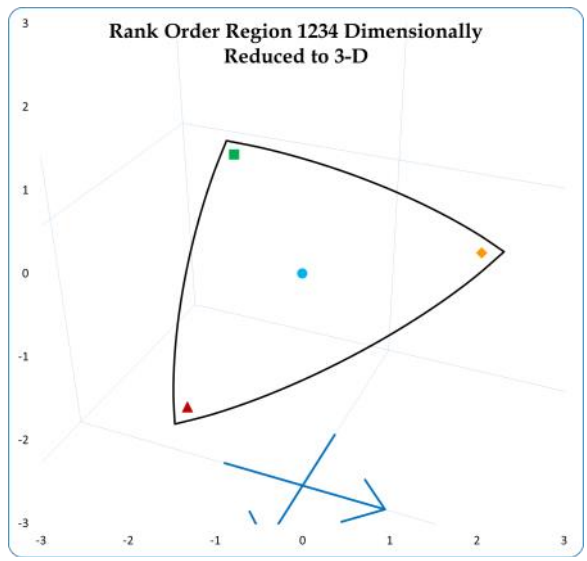


Figure 4

Figure 3 is a plot of four different vectors of N = 4 with rank order 1234. Figure 4 shows where each of these map to a point in the rank order region 1234. It is then shown where each of these map to a point in the rank order region 1234. Of note: the point in the center is the geometric center of the corners of the rank order region. It was determined that the four values of the vector associated with this point are very close to those values derived as van der Waerden scores from a normal distribution.

Using the construct developed above, the following is a graphical illustration of the IC method starting with [X] comprised of 2 vectors of N = 4. The first vector is derived from van der Waerden scores from a normal distribution and the second vector, as shown above, is comprised of {-0.439, 1.724, -0.575, -0.780} with starting rank order 3421. The desired correlation is -0.2225. The linear correlation of the starting vectors of [X] is -0.326; reflecting an absolute error from desired correlation of .103.

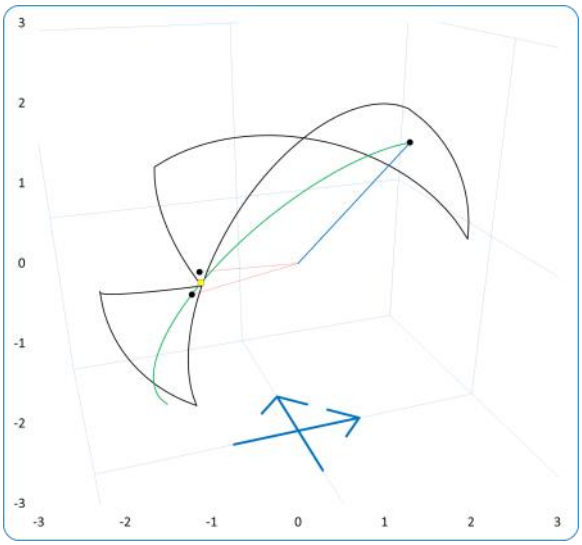


Figure 5

- The blue vector emanating from the center represents the first vector in [X]. As a vector derived from van der Waerden scores from a normal distribution, this point is in the center of the region with rank order 1234.
- The second vector of [X] is represented as the lower of the two red vectors. It occupies the lower left rank order region shown of rank order 3421.
- The green curve is the locus of all points mapped from the second vector of [X'] through correlation values ranging from -1 to 1. Recall that [X'] = [X][T] in the IC method where the second vector of [X'] is a linear combination of the vectors of [X]. This is the [X] subspace.
- The yellow marker is the point on the locus of points that satisfies the desired correlation value of -0.2225 exactly. This solution of the second vector of [X'] is {-0.607, 1.732, -0.546, -0.578} and is of rank order 1432. This point lies in a different rank order region than the second vector of [X].
- In the IC method, the second vector of [X] is then reordered to have that same rank order as the second vector [X']. This is depicted as the upper of the two red vectors and is merely the permutation of the second vector of [X] with the new desired rank order.
- The resultant correlation of the reordered second vector of [X] with the first vector is -0.175; an absolute error from desired correlation of 0.048.
- The following shows a clearer view of the region of interest.

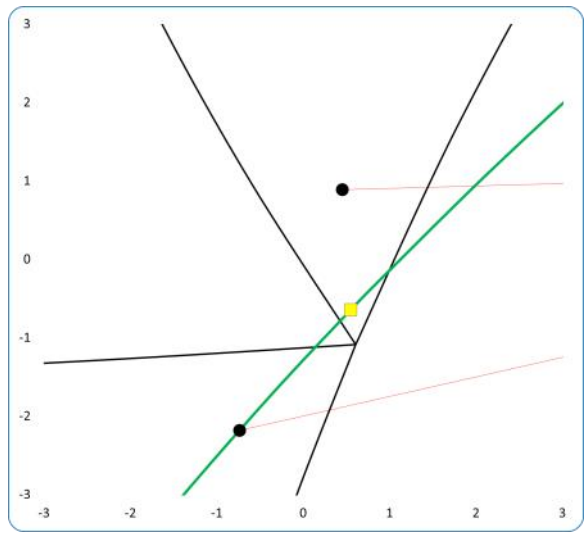


Figure 6

[X] Subspace

To demonstrate the constraint of the reordered second vector or [X] to points near the [X] subspace, the IC method was performed on the 4x2 input matrix over a range of desired correlation values from -1 to 1 in increments of .00125 (1/800). The first and second vectors of [X] applied here are the same as those used in the above illustration of the IC method. Now, define the **resultant correlation error** as the difference between the resultant correlation value and the desired correlation value (input to the [S] correlation matrix). The following graphs show both the resultant correlation and the resultant correlation error for both linear and Spearman's rank correlation measures.

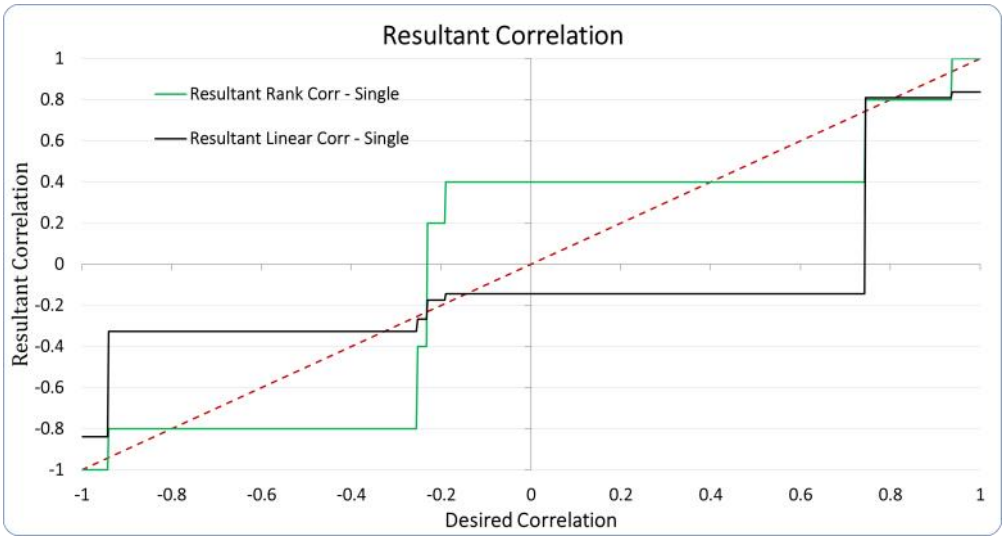


Figure 7

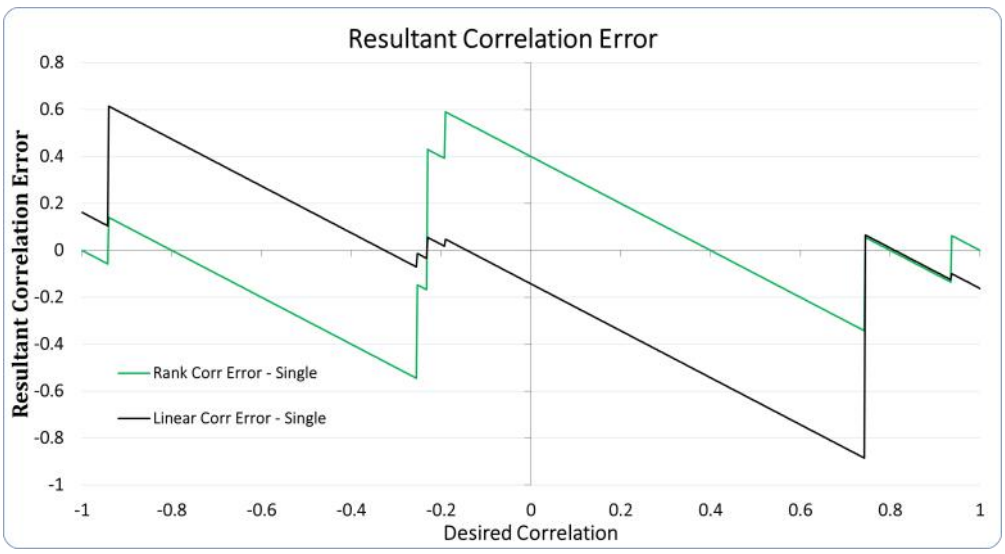


Figure 8

With only 4 values in each vector, a large error across much of the correlation range from -1 to 1 would not be unexpected. However, more noteworthy as it relates to the [X] subspace constraint is that of the possible 24 linear and 11 rank correlation values associated with all possible permutations of the second vector, there were only 7 distinct correlation values that resulted from the desired correlation values evaluated between -1 and 1.

Correlation with N = 10, 30, 100 and 1000

Resultant linear and rank correlation were evaluated in four other cases where M=2 and the number of values in each vector was 10, 30, 100 and 1000. As before, the desired correlation ranged from -1 to 1 by .00125. These starting vectors were also derived as van der Waerden scores from a normal distribution. The starting vectors of [X] remained the same through the range of desired correlation values. The following shows the resultant correlation and resultant correlation error for N = 100. The results from

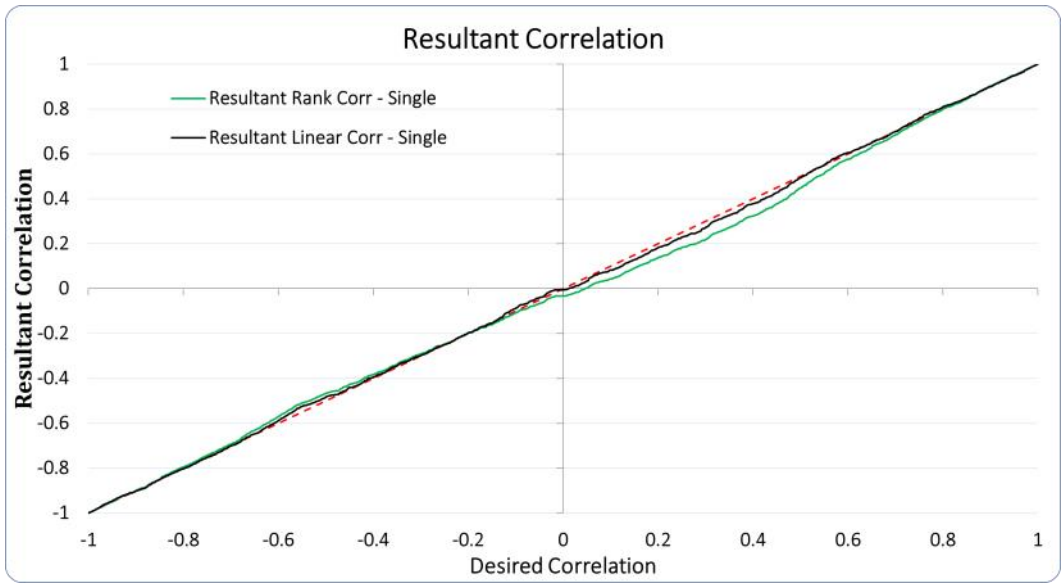


Figure 9

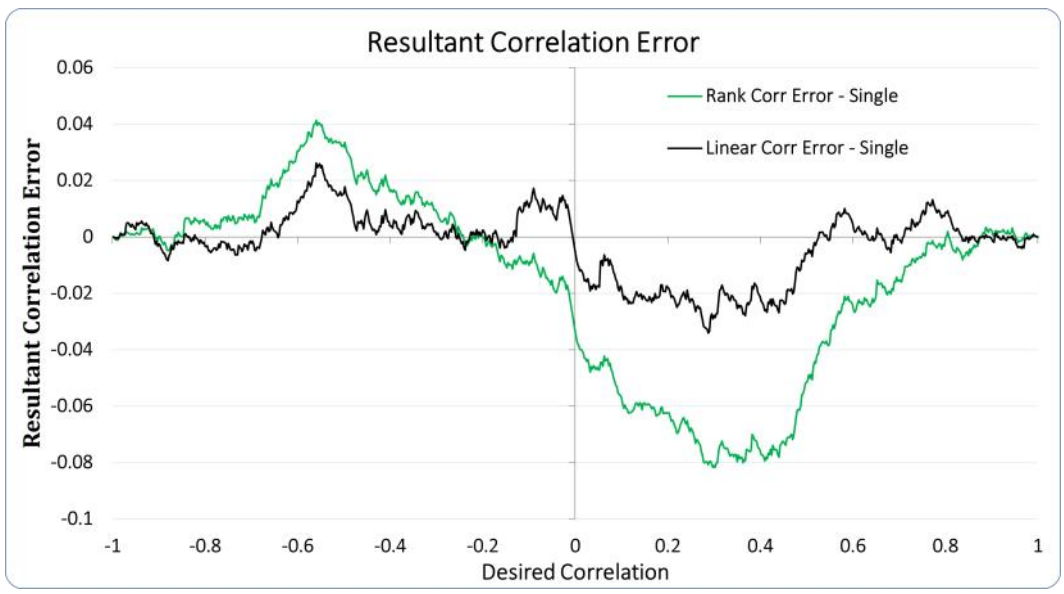


Figure 10

the other cases are available as supplemental material for this paper.

Observations:

- In the case of N = 10, there are 45 distinct resultant correlation values that are derived by the IC method. This is based on the same second vector in [X] for all desired correlations between -1 and 1 (evaluated at increments of 1/800). There are 10 factorial (3.63 million) possible permutations of the second vector of [X]. Each has a linear correlation value with the first vector. In this case, there are hundreds of thousands of unique linear correlation values associated with the 3.63 million different permutations of the second vector; yet only 45 resultant correlation values are revealed across the range of correlations evaluated with the IC method in this example; emphasizing the resultant correlation's constraint to the [X] subspace.

- A **global error** is defined as the root mean square of each of the 801 error values for the linear and rank resultant correlation. The following table shows those results. The linear correlation global error is less than rank correlation global error for each case of N. Heuristically, the notion that there are more possible linear correlation values associated with all possible permutations than there are rank correlation values would support a lower global error with linear correlation. Notably the ratio of global errors (i.e. Rank divided by linear global error) increases significantly with N.

	Global Error	
N	Linear Correlation	Rank Correlation
5	0.13913	0.14355
10	0.06457	0.07311
30	0.03749	0.05139
100	0.01217	0.03622
1000	0.00098	0.01411

Table 1

Improvement over the [X] Subspace Constraint.

For a given desired correlation, the reordered vectors of [X] can be considered to subtend a new [X] subspace and may be used as the starting point for another iteration. Recall [T] transforms the vectors of [X] into a set of vectors, [X'], which have the exact linear correlation prescribed in the desired correlation matrix [S]. Re-ordering the vectors of [X] based on the rank order of the vectors of [X'] is equivalent to selecting the permutations of the vectors of [X] that have the closest possible alignment to the vectors of [X']. That is, the closest possible alignment to a set of vectors whose correlation is exactly [S] ("closest possible alignment" implies maximum inner product of the vector of [X'] and the corresponding re-ordered vector of [X]). Conversely, any permutation of a particular [X] vector that does not lie in the rank order region of the corresponding [X'] vector would be less aligned (lower inner product) with the [X'] vector; resulting in linear correlation values further from the desired correlation.

Consider an input matrix of N X 2. After performing the IC method, the reordered second vector of [X] resides in a rank order region that results in a correlation value close to that prescribed in [S]. With a new [X] subspace defined by the reordered second vector of [X], the IC method is applied once more. Two alternatives may occur:

- The rank order of the resulting second vector of [X'] remains unchanged and so there would be no change to the rank ordering of the second vector of [X]; or,
- Based on the new [X] subspace, the rank order of the second vector of [X'] changes. The first iteration resulted in a good solution. However, since the second vector of [X] no longer has the same rank order of the recalculated second vector of [X'], it is less correlated with the revised exact solution second vector of [X']. Once reordered, it becomes better correlated with the second vector of [X']. As a result, the revised permutation of the second vector in [X] has a resultant linear correlation value even closer to the desired correlation. Any other permutation (particularly the previous one) would be less linearly correlated with the second vector of the revised [X']. Hence, the iteration results in a reduction in resultant linear correlation error.

For NX2 input matrices, iteration until convergence always yields resultant linear correlation values closer to desired with each iteration (when more than one iteration is necessary for convergence). However, it was found that for input matrices of N X M, M>2, improvements with each iteration for each resultant correlation matrix value does not always hold. It will be shown that iterating until convergence reflects improvements for resultant linear correlation; but not necessarily for resultant rank correlation.

The iterative process is described as follows; where the subscript denotes the iteration number:

$[X_1][T_1] = [X_1']$, $[X_1]$ is re-ordered based on the ranks of $[X_1']$ and becomes $[X_2]$.

$[T_2]$ is recalculated from $[X_2]$ and $[S]$.

$[X_2][T_2] = [X_2']$, . $[X_2]$ is re-ordered based on the ranks of $[X_2']$ and becomes $[X_3]$. If there is no change in resultant correlation of the vectors of $[X_3]$, stop. Otherwise…
$[T_i]$ is recalculated based on $[X_i]$ and $[S]$.

$[X_i][T_i] = [X_i']$, $[X_i]$ is re-ordered based on the ranks of $[X_i']$ and becomes $[X_{i+1}]$. Continue iterating until there is no change in resultant correlation of the vectors.

To illustrate this notion, the previous example illustrated in figure 5 serves as the starting point for this iterative process.

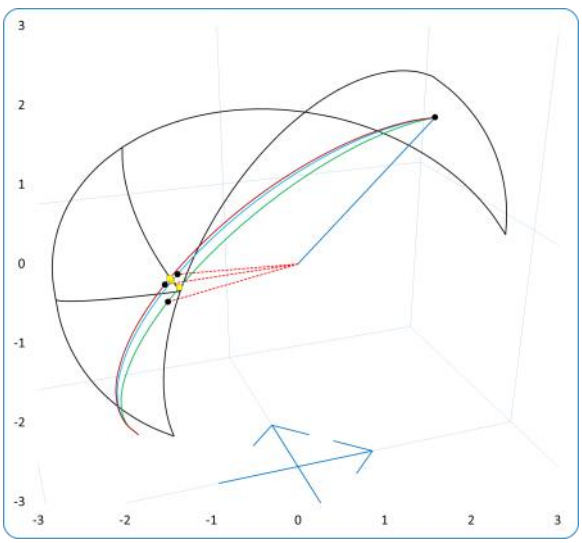


Figure 11

Referring to figure 11 , the blue vector emanating from the center represents the first vector of [X] and its point lies in rank order region 1234. The second vector of $[X_1]$ is the represented as the lower of the 3 red vectors and is mapped to a point in rank order region 3421. The green arc represents the linear combinations of the first and second vectors of $[X_1]$ as determined by [T] through the range of desired correlation values from -1 to 1. The desired correlation remains -.2225.

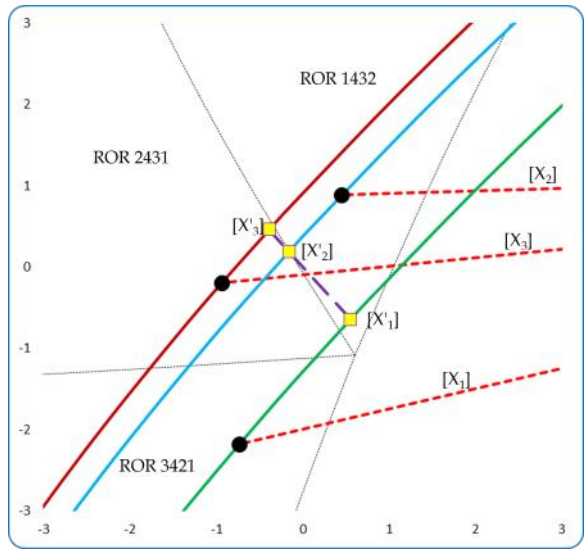


Figure 12

The figure 12 provides a closer view of the area of interest.

- The correlation between the starting vectors of $[X_1]$ is -0.326; yielding a absolute correlation error of 0.103.
- The far right yellow marker represents second vector of $[X'_1]$ whose correlation with the first vector is exactly -0.2225.
- The second vector of $[X'_1]$ lies in a different rank order region (1432). Thus the second vector of [X] is reordered to have the same rank order. This solution is depicted by the upper three red vectors, $[X_2]$. The correlation of the vectors of $[X_2]$ is -0.175; yielding an improved absolute resultant correlation error of 0.048
- Once again, $[X_2]$ is used to calculate $[X'_2]$ where the correlation between the first vector of [X] and $[X'_2]$ is exactly -.2225. In this case, the second vector of $[X'_2]$ has a rank order of 2431; different than that of the second vector of $[X_2]$.
- Thus, the second vector of [X] is reordered to that of the second vector of $[X'_2]$. This is depicted as $[X_3]$. The resultant correlation with this iteration is -.2663; yielding an absolute resultant correlation error of 0.0438; lower than the previous iteration error.

- Another iteration yields [X'$_3$]; whose second vector's rank order remains unchanged.
- Hence the final solution after iteration, [X$_3$], represents the optimal solution for linear correlation. Importantly, this is predicated on the starting vectors of [X$_1$] where possible solutions are confined to this subspace. However with iteration, regions beyond the initial [X] subspace may be encountered as was the case in this highly simplified, but wholly representative instance.

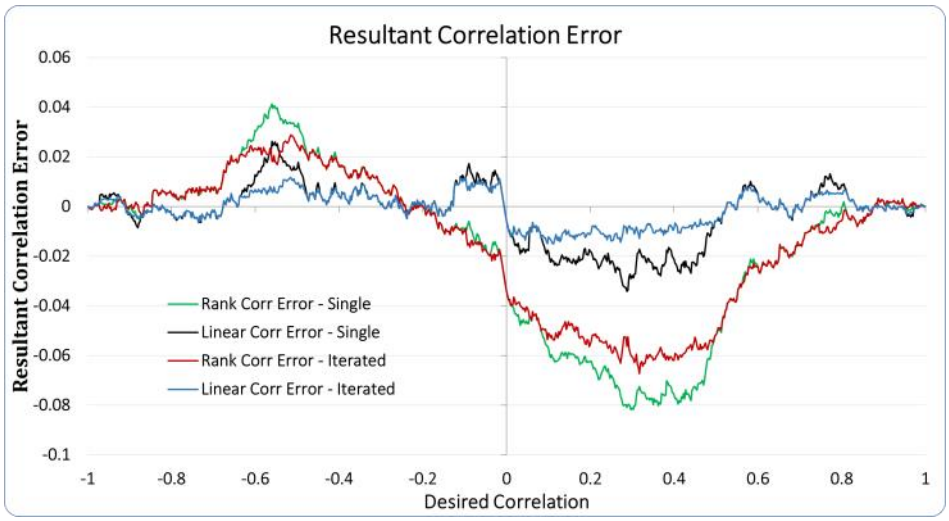


Figure 13: Results for N = 100, Correlation Error Combined

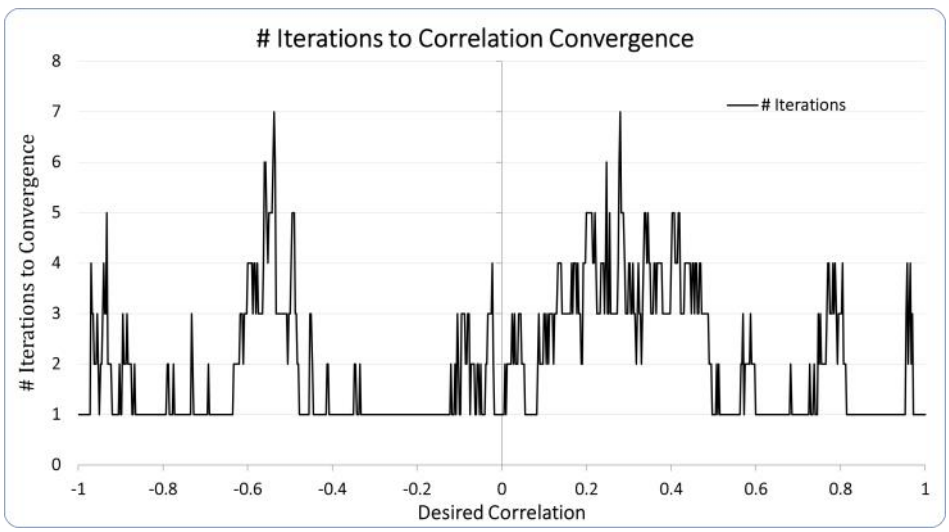


Figure 14: Results for N = 100, Number of Iterations

Correlation with N = 10, 30, 100 and 1000 and Iterating

Similar to before, resultant correlation and resultant correlation error was evaluated for desired correlation values ranging from -1 to 1 in increments of .00125. This now includes results of the iteration process described above for the same four cases of N=10, 30, 100, 1000. Each case used the same starting [X] to equitably compare the single iteration IC method to the iterative process. The following shows the resultant correlation error and number of iterations to convergence for N = 100. The results from the other cases are available as supplemental material for this paper. The chart legend indicates "Single" for applying the IC method once, or "Iterated" which indicates iteration until convergence.

It was found that for all cases where more than one iteration was performed to achieve convergence of the reordered vector, an improvement in linear correlation error resulted. There are, however, ranges of desired correlation where rank correlation error (absolute value implied) worsened (e.g. -.1, .75). The following table shows global error, as previously defined, for all cases and the change in global error from the single IC method to the iterative process discussed here.

	Global Error			
	Single IC Method		Iterating Until Convergence	
N	Linear Correlation	Rank Correlation	Linear Correlation	Rank Correlation
10	0.06457	0.07311	0.05649	0.06464
30	0.03749	0.05139	0.02186	0.03557
100	0.01217	0.03622	0.00651	0.03028
1000	0.00098	0.01411	0.00056	0.01389

Table 2

	Global Error	
	% Change from Single to Iterative Result	
N	Linear Correlation	Rank Correlation
10	-12.52%	-11.59%
30	-41.68%	-30.79%
100	-46.50%	-16.39%
1000	-42.75%	-1.57%

Table 3

The Iteration Method with Skewed Distributions

To stress test the iterative process, three highly skewed input vectors were used as the second vector of [X] for N = 10, 30 and 100. These distributions have a population skewness of 2.62, 3.14 and 8.36; respectively. As before, the first input vector values were derived from the normal distribution. The following graph is a plot of values of the rank ordered second versus the first vector of [X]. Linear correlation of the rank ordered vectors is shown in the legend; quantifying the dissimilarity of the distributions.

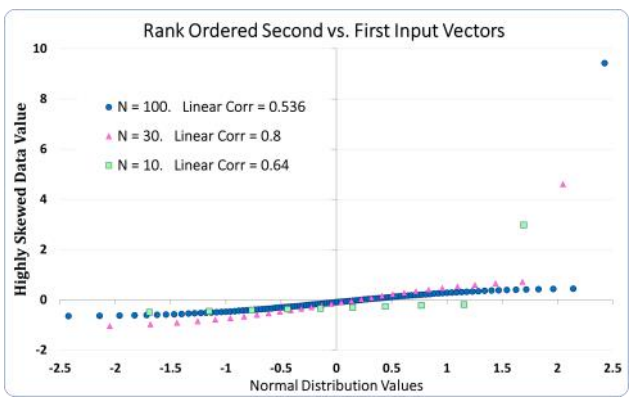


Figure 15

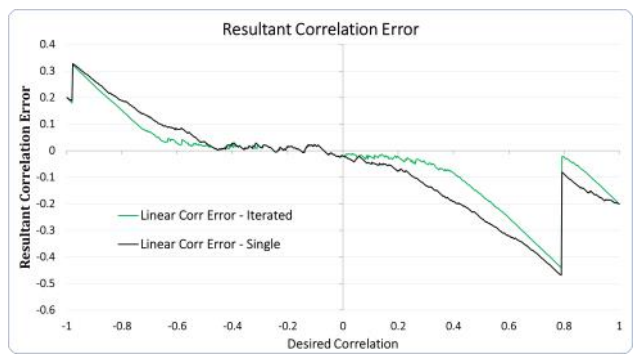


Figure 16: Results for N = 30, Linear Correlation

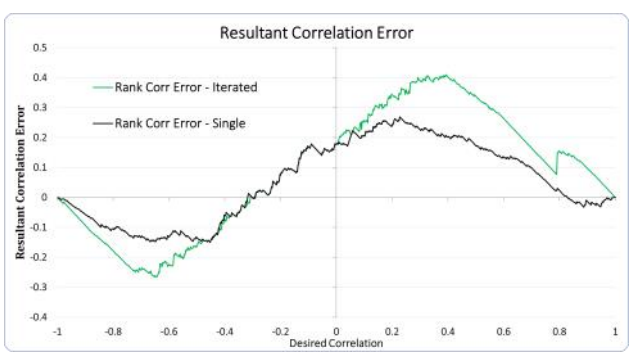


Figure 17: Results for N = 30, Rank Correlation

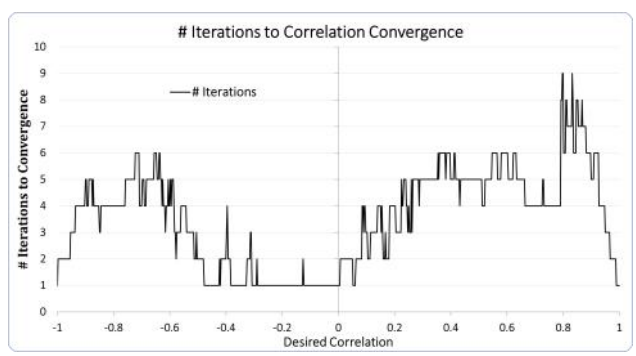


Figure 18: Results for N = 30, Number of Iterations

Results are shown for N = 30. The results from the other cases are available as supplemental online material for this paper.

Observations:

- In all instances, resultant correlation error (absolute value implied) for linear correlation was reduced with more than one iteration until convergence.
- With very few exceptions, resultant correlation error for rank correlation increased with more than one iteration until convergence.
- There is significant error with linear correlation toward the ends of the desired correlation range because of the dissimilarity between the distributions of the vectors of [X].

Higher Dimensional [X] Subspace: Input matrices with 2 variables have been used for illustrating improved linear correlation results of the IC method with iteration. The following considers an input matrix with 16 vectors (i.e. M = 16) for cases of N = 30 and 100. Once again, values are derived from a normal distribution as van der Waerden scores. By virtue of the transform matrix [T] being of upper triangular form, the vectors of [X'] are linear combinations of the corresponding vector of [X] and those to the left. (e.g. the 11th vector of [X'] is a linear combination of the first 11 vectors of [X].) The [X] subspace has expanded to a higher order subspace for input vectors further to the right in the [X]; potentially enabling reduced resultant correlation error. The following steps were taken to assess the effect of the higher dimensional [X] subspace.

- The desired correlation matrix, [S], has all values in the first column and first row the same. That is, all vectors will be seeking the same correlation value with the first vector in order to assess any effect of the higher dimensional [X] subspace. The remaining values of the desired correlation matrix were chosen to ensure positive definiteness.
- Linear correlation error was evaluated with single iteration and with iteration to convergence for desired correlation values ranging from -1 to 1 by .00125. From these results, global error was evaluated for each vector.
- In order to ensure linear independence of the vectors of [X], the 2nd through 16th vectors of [X] were randomly permuted for each desired correlation value.
- Evaluating linear correlation results of the first vector with all others for desired correlation values ranging from -1 to 1 was performed 30 times to attain data with a degree of statistical significance.
- The average and sample standard deviation of the global error from each of the 30 runs, and for each of the 2nd through 16th vectors was evaluated.

The following graphs show the global error results for the 16 vector input matrix [X] with N=30 and N = 100. Also plotted are plus and

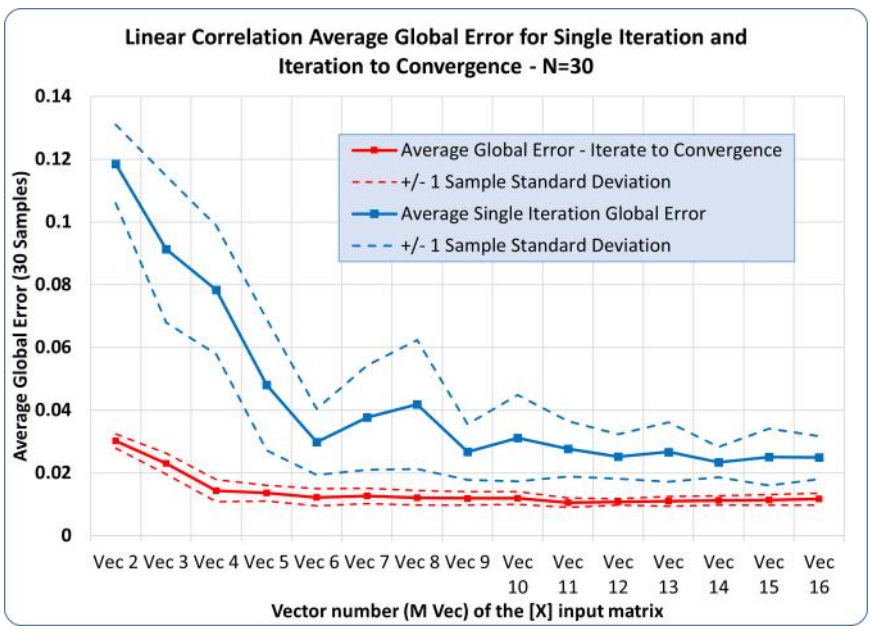


Figure 19

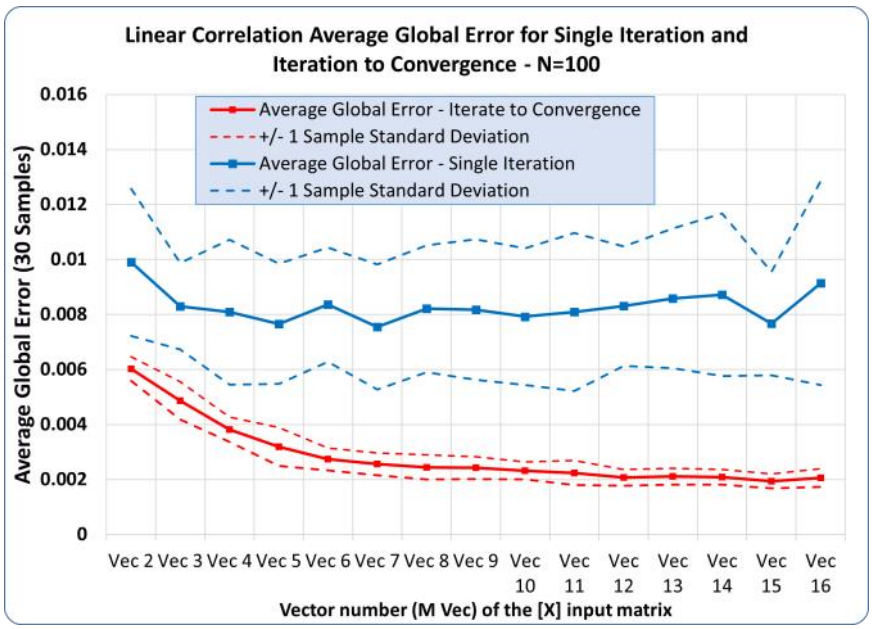


Figure 20

minus one sample standard deviation from the sample mean for the single iteration global error results and for the iteration to convergence global error results based on the 30 runs for desired correlation from -1 to 1 by .00125.

Observations:

- As shown previously in the case of M=2, linear correlation global error is reduced with iteration to convergence.
- For both cases of iteration to convergence for N=30 and N=100, the global error of the 16th vector is approximately one third of that of the 2nd vector. This suggests that iterating within a higher order subspace of N dimensions enables reduced global error. In other words, vectors more to the right in [X] are more likely to have less resultant correlation error than preceding vectors with the iteration method.
- It was observed that while the root mean square of all linear correlation errors reduced with each iteration, some individual correlation errors (i.e. correlation of vector i with vector j; j ≠ i) increased during the iteration process.

Summary

The IC method of developing a multivariate input variable with prescribed correlation is first described and demonstrated. The IC method is then applied for the case of two input variables where the correlation between the two input vectors is prescribed in [S] and the resultant linear and rank correlation of the reordered vectors is calculated. IC method results are evaluated for desired (prescribed) correlation values between -1 and 1 by increments 1/800th. This assessment is performed for input vectors with the number of values (N) ranging from 5 to 1000. A global error is defined as the root mean square of the difference between the desired correlation and that calculated from the IC method for all 801 instances of desired correlation between -1 and 1.

There are N factorial possible permutations of an input vector which span N dimensional space. With all permutations viewed as a constellation of discreet points in N dimensional space, each of these points has a correlation value (linear and rank) with the other input vectors. The IC method constrains possible outcomes to those near the

subspace in N dimensional space subtended by the linear combination of the vectors of [X]. It is possible that there are values in the constellation of possible resultant correlations that are closer to the desired correlation.

An iterative process is presented where the resultant re-ordered vectors of [X] are used as the starting point to once again apply the IC method. This is repeated until convergence of the resultant correlation is achieved. Using the global error previously described as a comprehensive metric, the iterative method shows marked reductions in linear correlation error (i.e. absolute value of the difference between prescribed and resultant correlation) from that of the single step IC method. For highly skewed and dissimilar distributions of input variates in [X], it is shown that rank correlation error often worsens with iteration while linear correlation error improves.

The single iteration Iman Conover method is a powerful technique that is likely more than adequate given the confidence limits of prescribed correlation values. As such the practitioner should consider the value of improving the correlation accuracy of the input variables by iteration.

Having a practical understanding of the methods by which correlated multivariate input variables are developed is useful when the software platform does not provide what is needed. The ability to perform this manually, to understand the nature of its limitations and to experiment with various distribution types may be useful and is certainly instructive for the practitioner. There may be special circumstances where increased accuracy of the correlation of a set of input variables is needed. The author notes that the improvement in accuracy with the application of the iterative IC method described herein should be considered in the context of the confidence limits of the desired correlation coefficient value derived from sampled data where the Fisher r to z transformation has applicability.

References

1. Iman, R. L., Conover, W. J., *A Distribution-Free Approach to inducing Rank Correlation Among Input Variables*. Commun. Statist. Simula. Computa. 11(3) 311-334 (1982)
2. Mildenhall, S., *Correlation and Aggregate Loss Distributions With An Emphasis On The Iman-Conover Method*. Casualty Actuarial Society Forum, Winter 2005. (4) 42-55
3. Tamhane, A. , Dunlop, D. 2000, *Statistics and Data Analysis*, Upper Saddle River (NJ), Prentice Hall Inc. (10) 380-385

Douglas Henke *served in the U. S. Coast Guard from 1980 to 2004. His service included tours as Chief Engineer on the Coast Guard Cutter CAMPBELL, Associate Professor in the Coast Guard Academy's Engineering Department and final tour as Program Manager for recapitalizing the Coast Guard's aging fleet of ships and boats. Since retirement in 2004, he's been a consultant to Program Managers in the Coast Guard's Acquisition Directorate; providing a variety of services including financial and statistical analysis and modeling.*

The International Cost Estimating and Analysis Association is a 501(c)(6) international non-profit organization dedicated to advancing, encouraging, promoting and enhancing the profession of cost estimating and analysis, through the use of parametrics and other data-driven techniques.

www.iceaaonline.com

Submissions:

Prior to writing or sending your manuscripts to us, please reference the JCAP submission guidelines found at

www.iceaaonline.com/publications/jcap-submission

Kindly send your submissions and/or any correspondence to JCAP.Editor@gmail.com

International Cost Estimating & Analysis Association

4115 Annandale Road, Suite 306 | Annandale, VA 22003

703-642-3090 | iceaa@iceaaonline.org