Volume 11 | Issue 1 | April 2023
ISSN: 2160-4746

Journal of Cost Analysis and Parametrics

Editor in Chief: David L. Peeler, Jr., CCEA®

ICEAA
International Cost Estimating & Analysis Association
www.iceaaonline.com

ICEAA

Journal of Cost Analysis and Parametrics

Volume 11 | Issue 1 | April 2023

ISSN: 2160-4746

Editor's Note

David L. Peeler, Jr., CCEA®

Welcome to the fourth issue of our revived Journal of Cost Analysis and Parametrics (JCAP). As we conclude year two, we thank all who have taken time to write, submit, and revise pieces for publication within these pages, as well as those who have taken the time and mental energy to peer review all submissions received. This issue contains articles range from the psychology of cost estimating to the tactics of simplified function point calculations.

The intellectual pursuits shared by these authors are the results of work papers, school studies, and pieces written for and presented to our annual ICEAA professional development and training workshop. Among the discussion topics leading up to resumption of journal publication was widening the aperture to include an opening for qualitative pieces. One such paper that shouldn't be overlooked, is **Andy Prince**'s *The Psychology of Cost Estimating* – winner of the 2015 ICEAA best paper award. The article discusses the problems of cost overruns, the biases that lead to them, and rational approaches to overcome said biases. Having introduced this paper to numerous students, I highly encourage you to read, or re-read, this piece; it has value and insight for us all.

The second article is from Captain **Kyrie Rojo**, summarizing her graduate work: *Cost Estimating Relationships for Recurring T100 Flyaway Costs*. This paper received the 2023 ICEAA Outstanding Air Force Institute of Technology Thesis Award. Her research applies regression techniques to create two cost estimating relationships predicting recurring T100 flyaway costs, in the largest such aircraft regression study to date. The results provide a useful cross-check in early estimation of aircraft costs.

Article three is a data-rich piece addressing learning curve applications in second source facilities. Packed with interesting historical as well as contemporary insights, *Second Source Manufacturing: Lessons from the Second World War* was the best overall paper from last year's ICEAA Workshop, making it the third conservative overall win for **Brent M. Johnstone**. This piece is a riveting read, incorporating WWII data forward to inform modern production operations.

Given numerous attempts – over decades – to improve cost and schedule outcomes for major defense acquisition programs, a critical look at these outcomes is worthy of examination. Our fourth article does just that. In *Schedule and Cost Estimations Through the Decades: Are They Improving?*, Captain **Sammantha Jones** uses both descriptive and inferential techniques to investigate schedule and cost trends from the 1970s to 2010s. Her findings present a mixture of positive and negative results.

Douglas K. Howarth provides our fifth contribution, *8D Cost Trades with Entanglement*. This article proposes a construct with elements more dynamic than simple two-dimensional interactions. He uses eight dimensions to describe how jets and their engines can work in tandem to enhance sales; as well as the need for suppliers and the offerors of the final products to work together toward common goals.

In *The BS in BoeS: Oh, the Games That Are Played*, **Sandy Burney** tells a story about the basis of estimates. He emphasizes estimates tell a story and should be constructed in story form, with a prologue, chapters, and an epilogue. Further, this story must clearly present an acceptable estimating methodology and supporting data. Possibly presented as a sales brochure, enticing the buyer to purchase.

The seventh and final article addresses *Simplifying Software Sizing with Simple Function Points*. **Carol Dekkers** and **Dan French** introduce simple function point methodology, demonstrate its use, and highlight both the challenges and the opportunities for software cost estimators. They provide a method to estimate software size from high level software requirements. Additionally, the paper explores the key differences between simple function points and the globally accepted function point approach by the International Function Point Users Group.

We hope you have found this overview helpful in focusing your reading choices. Of course, we'd love for you to read all the articles, but understand time constraints and interests. Enjoy your choices. May you find something in these pages applicable to your efforts and helpful to your professional and/or personal pursuits. Thank you for your interest, attention, and support.

Colonel **David Peeler**
US Air Force Retired
Editor-in-Chief

The Psychology of Cost Estimating

Andy Prince

Introduction

Cost estimation for large (and even not so large) government programs is a challenge. The number and magnitude of cost overruns associated with large Department of Defense (DoD) and National Aeronautics and Space Administration (NASA) programs highlight the difficulties in developing and promulgating accurate cost estimates. These overruns can be the result of inadequate technology readiness or requirements definition, the whims of politicians or government bureaucrats, or even as failures of the cost estimating profession itself. However, there may be another reason for cost overruns that is right in front of us, but only recently have we begun to grasp it: the fact that cost estimators and their customers are human.

The last 70+ years of research into human psychology and behavioral economics have yielded amazing findings into how we humans process and use information to make judgments and decisions. What these scientists have uncovered is surprising: humans are often irrational and illogical beings, making decisions based on factors such as emotion and perception, rather than facts and data. These built-in biases to our thinking directly affect how we develop our cost estimates and how those cost estimates are used.

We cost estimators can use this knowledge of biases to improve our cost estimates and also to improve how we communicate and work with our customers. By understanding how our customers think, and more importantly, why they think the way they do, we can have more productive relationships and greater influence. By using psychology to our advantage, we can more effectively help the decision maker and our organizations make fact-based decisions.

This paper is structured into three parts. Part 1 provides a discussion of the problem of cost overruns in the aerospace industry and some of the findings from traditional cost overrun studies. Part 2 talks about human irrationality and how that opens the door to biasing the estimate. Findings from psychology and behavioral economics are used to describe how biases and faulty logic on the part of the estimator and the customer can lead to poor cost estimates and a failure on the part of the cost estimator to add value. In Part 3 I discuss practical techniques and approaches that can drive rationality into the estimate, thus changing the conversion to one that is productive and influential, leading ultimately to better decisions and fewer cost overruns.

One final note, my experiences and observations are drawn from almost 31 years working as a NASA cost estimator. While I know that cost overruns are a common problem across the Federal Government and industry, everyone has a unique perspective. All I ask is that as you read this paper you keep an open mind. If you don't think what I have to say is useful or applicable to your situation, that's fine. As a wise instructor of mine once said, "I am offering you a buffet, feel free to take what speaks to you and leave the rest."

Part 1: The Problem

Cost estimators face a challenging environment when trying to predict the cost of large, technologically sophisticated government programs. Early in the design process, requirements are often poorly defined or understood. Incorporation of new technologies creates estimating uncertainties. The underlying industrial base (at least for NASA) is small and highly specialized, creating a "use it or lose it" mentality which leads to a focus on maintaining the industrial base at the expense of program efficiency.

The overarching business environment has its own set of issues. Program funding, schedules, and requirements are driven by political and budgetary considerations. Large bureaucracies in government and industry tend to focus on process versus outcome –making hard decisions difficult and creating programmatic inertia. The culture within government agencies can be to strive for consensus, even if that consensus comes at the expense of healthy conflict and an honest discussion of what something will truly cost.

The estimating profession itself has challenges brought about by small, noisy data sets and models which are sometimes mysterious (PRICE & SEER) or may not be adequately validated. We in the community have few models that can claim to account for the physics of the systems they are estimating or the underlying industrial processes for developing and producing the hardware.

A result of this challenging estimating environment is that cost overruns have become institutionalized within the Federal Government. As can be seen in Figure 1, a cost growth history of 156 NASA projects shows that 84% have some level of cost growth, and almost 30% have cost growth of 50% or more. While NASA is making a serious attempt (and showing early progress) to control costs through the use of the joint cost schedule confidence level (JCL) analysis, cost growth continues to be a significant issue as illustrated by projects such as the James Webb Space Telescope (JWST).

Cost overruns create obvious problems for government agencies. In the current environment of flat or declining budgets, cost overruns in existing projects mean that new projects will be starved of funds (leading to possible cost growth later) or will unable to begin development. Delaying the development of capabilities needed by warfighters or scientists is not in anyone's best interest. Nor is it in the best interest of government agencies to disappoint Congress or the Office of Management and Budget (OMB). Excessive cost growth invites the scrutiny of Congress, the Government Accountability Office

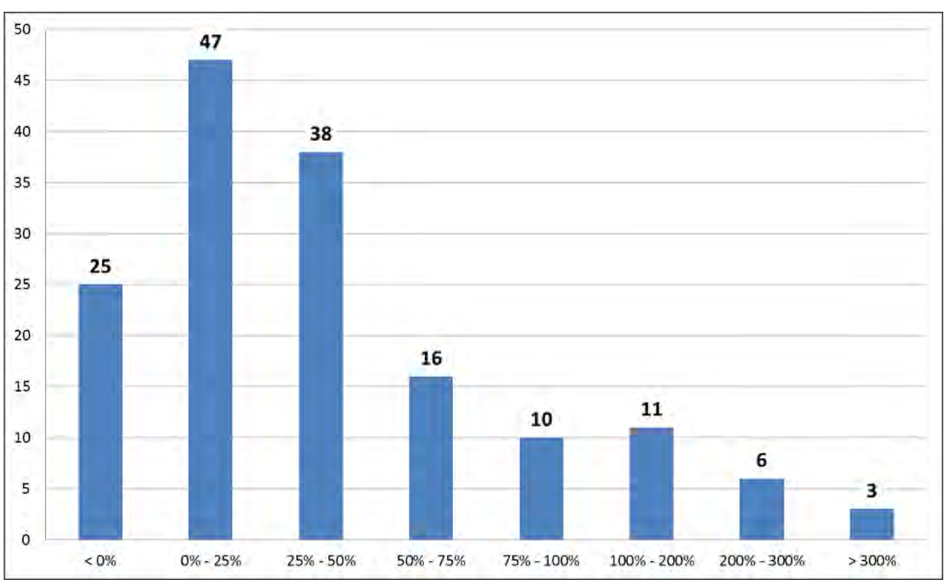


Figure 1. NASA Cost Growth History.

(GAO), and the press. Perceptions of management incompetence brought about by program overruns can lead to project terminations and agency funding cuts.

Causes of Cost Overruns

Numerous studies have been performed and reports have been written on the causes of cost growth in government programs. The causes identified in these reports tend to be depressingly consistent. For example, a National Academes report from 2010 titled "Controlling Cost Growth of NASA Earth and Space Science Missions" lists the following four commonly identified factors in cost growth:

- Overly optimistic and unrealistic initial cost estimates
- Project instability and funding issues
- Problems with the development of instruments and other spacecraft technology
- Launch service issues

A 2002 Booz Allen Hamilton presentation titled "Space Systems Development Growth Analysis" looked at Air Force space mission cost growth. Among its findings were:

- Initial cost estimates are based on inadequate data and do not capture total program cost
- Evolving technical requirements
- External program funding adjustments are frequent
- Optimistic acquisition strategies
- Overemphasis on bottom line cost versus cost realism
- Cost estimating capabilities that had been allowed to atrophy

Finally, a NASA Project Management Study dated January, 1981 identified the following four contributors to cost growth (obviously, not much has changed):

- Technical complexity of projects
- Inadequate definition prior to NASA's budget decision and external commitment
- Effect of NASA's tendency to select based on basis of bid price and the contractor low bids
- Poor tracking of contractor accomplishments against approved plans in a timely fashion

While I could list many more sources (such as the GAO) and their findings, in the interest of brevity I will proceed to a discussion of cost growth causes using a taxonomy that summarizes several factors into four broad categories or groups.

The first group that could be blamed for cost overruns is us, the professional cost estimator. If we are developing poor cost estimates; perhaps as a result of bad models, inadequate data, or poorly trained cost estimators, then our estimates could mislead management into believing that they have sufficient resources for successful project execution. While I certainly believe ample opportunity exists to improve government cost estimating, I don't believe that the cost estimators, the tools, or the data is the root cause of the problem.

The second category of factors causing cost overruns are those related to technical or managerial actions. Almost all cost overrun studies cite inadequate or poorly defined requirements or underestimated technical complexity or overestimated technology readiness as root causes. Several studies also mention poor project management and acquisition practices. And the impact of funding cuts on cost growth is a known fact. However, while these factors are a documented source of cost growth, I believe there is a more fundamental issue that is behind the cost overrun problem.

The third group that can be blamed is the corporate management community. In a paper written for the Journal of Parametrics ("What are Quality Cost Estimates or the 260 Hz Estimate," Spring 2007), Joe Hamaker, then the lead for cost estimating at NASA, explicitly relates the importance of management culture on the quality of the cost estimate.

> "I believe that actually, the first enabler for quality cost estimates in any organization is the state wherein the management of that organization actually wants to know at ATP what the projects are likely to cost. This is somewhat important. If management doesn't really want to know the truth, that fact flows down to the estimating community pretty quickly in the form that quality estimates aren't an important product anymore."

Obviously, a corporate culture that does not see value in cost estimating as means to improve performance is not going to support an estimating culture that provides realistic cost estimates.

Related to management culture is an organization's governance process. Governance here refers to the various boards and committees necessary to running a large organization. If the governance process is broken and does not get non-advocate cost estimates to right people at the right time, decision makers are making decisions with inadequate data. However, I believe that an unsupportive management culture or a broken governance process are symptoms, rather than the cause of cost overruns.

The fourth category are those factors that have nothing to do with cost estimating capabilities, technical definition, funding shortfalls, or management culture. Rather, these factors apply to all of us: estimators, technical experts, managers, and executives. This cause of cost overruns is the fact that everyone involved in the cost estimating process is human.

The remainder of this paper will explain how our own built-in human biases lead to poor, or poorly used cost estimates. These biases apply to all of us, no one is immune. But knowledge is power and knowledge of how these biases affect our estimates gives us, the estimating community, the power to add value even in the face of management resistance.

Part 2: The Irrational Human

Psychologists have discovered that human thought processes are surprisingly irrational. I use the word "surprisingly" because for many years, economists (and others) assumed that humans behaved in ways that are rational and predictable. Among some of the things these researchers have found are:

- We are unfailingly optimistic in our outlook on life
- We are overconfident in our abilities
- Our thinking is often shallow and colored by our current emotional state
- We prefer stories and anecdotes over facts and data
- Statistics are non-intuitive, so we either discount them or misuse them
- We have problems accepting randomness – everything must be explained
- We fear losses more than we value gains
- Personal experience and knowledge trumps everything

Psychologists have also learned that our irrational behavior is also predictable. In his book Predictably Irrational, Dan Ariely writes

> "…we are really far less rational than the standard economic theory assumes. Moreover, these irrational behaviors of ours are neither random nor senseless. They are systematic, and since we repeat them again and again, predictable."

So why do we behave in such an irrational manner? According to psychologists and others, it is a coping mechanism that allows us to adapt to our environment. Douglas Hubbard, in his book How to Measure Anything, explains the how this coping mechanism works.

> "It's no revelation that the human mind is not a purely rational calculating machine. It is a complex system that seems to comprehend and adapt to its environment with an array of simplifying rules. Nearly all of these rules prefer simplicity over rationality. Those that are not quite rational but perhaps not a bad rule of thumb are called "heuristics." Those that fly in the face of reason are called "fallacies.""

Thus we are not thinking machines, processing information in a computer-like manner. Rather, our decision making process is much more complex and biased by extraneous information, beliefs, and unconscious thoughts. Our human thinking mechanism leads to irrational (or suboptimal) decisions. These irrational decisions affect cost estimates because cost estimates are predictions, and predictions are the result of decisions about how what is known today is going to lead to an outcome at some point in the future.

In his book Thinking, Fast and Slow, Nobel Laureate Daniel Kahneman explains how a common human thought process call "substitution" leads to biased predictions.

> "…the prediction of the future is not distinguished from an evaluation of the current evidence – prediction matches evaluation. This is perhaps the best evidence we have for the role of substitution. People are asked for a prediction but they substitute an evaluation of the evidence, without noticing that the question they answer is not the one they were asked. This process is guaranteed to generate predictions that are systematically biased; they completely ignore regression to the mean."

In other words, Kahneman is pointing out how this one flawed thought process can cause us to make decisions that lead to unsupported deviations from trends identified by the historical record. Our cost estimates are led astray because we believe the story we are told about the system we are estimating rather than what the historical data tells us the system should cost!

Biases

Psychologists have identified a number of common biases in how we think that affect all of us, or at least those of us who are human. These biases influence our decisions because they act on our subconscious and thus we are unaware that our decisions are being affected. These biases are comfortable because they cater to our self-image or they appeal to our desire to create an orderly, explainable world. All of these biases are well-known and understood by psychologists and the fact that we have them will probably not be a surprise to the observant reader.

Optimism/Overconfidence*:* Most of us are optimistic by nature. We like to believe that to at least some extent, we are masters of our fate. In fact, research has shown that an absence of optimism can lead to depression. However, optimism is a form of self-delusion. Too much optimism leads to overconfidence and overconfidence leads to a poor understanding of risk and the underestimation of the probability of failure. Overconfidence affects estimators and managers alike. While we often accuse managers and study leads of being too optimistic about design complexities or technology, we estimators can also be overconfident in our ability to predict costs.

Anchoring*:* Sometimes referred to as relativity, anchoring is the ability of a number to influence an analytical (or estimated) outcome. For example, if a cost estimator is told that the cost is expected to be $100M, the estimate will be significantly lower than if the estimator had been told the expected cost is $500M. Anchoring also works when we are told that what we are estimating is similar to a previous project, or that we are inheriting hardware, software, management team, etc. from a previous project.

***Availability**:* Psychologists have proven through experimentation that we will assign a higher probability of occurrence to information that is easy to retrieve. Therefore, what we are most familiar with will have the greatest impact on our cost estimates. For example, if we are told that Project X is going to use an inherited design, and we go and search our historical data and find several examples of projects that realized cost savings by using an inherited design, we are more likely to accept the proposition that the cost for Project X will be less than the historical average. Note that what we have done is merely confirm that an inherited design saves money, it says nothing about the ability of Project X to actually use an inherited design. Availability can also affect our customers. By repeating the same stories over and over to themselves ("we are TRL-6 or greater") they come to strongly believe what they want to believe.

***What you see is all there is (WYSIATI)**:* WYSIATI is a phrase coined by Daniel Kahneman to describe how our minds can quickly develop a coherent story out of limited information. Two surprising facts emerge from WYSIATI. First, the less information we have the more confident we are in our coherent story. Second, the coherent stories that we build often ignore probability and statistics. The danger with WYSIATI is that we will be overconfident in our knowledge, thinking we know that answer when in fact we are relying on a story that is plausible, supported by what our minds can readily recall, and is consistent with our worldview.

***Halo/Horns Effect**:* The Halo/Horns effect (also known as the confirmation bias) is our tendency to emphasize data that agrees with our belief or intuitive assessment, and to discount information that disagrees with our position. The Halo/Horns effect can also cause us to look for (or be more open to accepting) data that confirms our position or opinion. Obviously, the danger with this bias is that we will overlook or discount important information that is inconsistent with the desired outcome.

***Plausibility Effect**:* When we believe the more plausible outcome over the more probable outcome, we are falling victim to the plausibility effect. The Plausibility Effect occurs because we like explanations that address all of the facts, even if those facts are suspect or spurious. Cost estimators fall victim to the Plausibility Effect whenever we confuse a good story with a probable outcome (see WYSIATI).

***Bandwagon Bias**:* Humans have a strong need to conform. In a group setting it is not uncommon for the most vocal and outspoken members of the group to dominate the conversation. Typically, these vocal members will eventually bring the others around to their point of view. Cost estimators are human, too. We can easily be influenced by a strong, vocal project manager (are there any other kind?) and a project team that is already on-board with their leader's opinions.

***Attractiveness**:* Appearances matter. Psychologists have known for years that people assign more favorable characteristics to attractive people or products. We are also more likely to believe a good presenter over a poor presenter. Attractiveness and the Plausibility Effect and the Confirmation Bias are interrelated. We like a good (attractive) story that makes sense and explains all the data, especially if it confirms a previously held belief or opinion.

These biases and thought patterns do not operate independently. Rather, they interact with and reinforce each other, leading to poor decisions on the part of executives, managers, project leads, and cost estimators. For example, if the project team tells us a good story that is logical and plausible, if they all obviously believe it, if the data all lines up, then we are more likely than not to go along with the project and produce a cost estimate that is consistent with the project's expectations. In addition, we will probably feel very confident in the outcome. We all want to be good people and get along with everyone and be part of the team. The danger for the cost estimating profession is this: if we are not adding value by increasing the probability that the project can be performed on budget, then we are at risk for losing support within our organizations.

Part 3: Antidotes to Biases

You, or your customer, cannot avoid being biased any more than you can avoid breathing. These biases are hardwired into our psyche and can only be overcome with great mental effort. What we can do, however; is to engage in actions that force us to approach cost estimating as analytically as possible. These actions can help us improve our estimates and add value to the products we provide our customers. They also give us the tools by which we can influence our customers into making better decisions. The six actions (or antidotes) I have identified are listed below.

- Have a good process
- Inject a healthy dose of reality
- Validate your results
- Embrace the uncertainty in your estimates
- Be the cost expert
- Build and tell your story

The following sections address in detail each one of these antidotes. However, let me say up-front that the common theme running through each of these actions is the desire to bring more data and information to bear on the estimate at hand. As Kahneman states on page 201 of "Thinking, Fast and Slow," ignorance creates the fertile ground for biases, and as humans we fail to appreciate how ignorant we can be.

> "At work here is that powerful WYSIATI rule. You cannot help dealing with the limited information you have as if it were all there is to know. You build the best possible story from the information available to you, and if it is a good story, you believe it. Paradoxically, it is easier to construct a coherent story when you know little, when there are fewer pieces to fit into the puzzle. Our comforting conviction that the world makes sense rests on a secure foundation: our almost unlimited ability to ignore our ignorance."

The other thing that Kahneman is telling us is the importance of stories. Everything done for a cost estimate can and should contribute towards the cost estimator telling his or her own story. We are more influenced by stories than we are by data, and the cost estimator should use that to their advantage.

A final theme not explicitly expressed in the bulleted list above, but one that I believe is very important to the work we do, is the opportunity to add value beyond the cost estimate. By adding value, I am saying that we, as cost estimators, need to look for opportunities to contribute to the success of whatever endeavor we are estimating. Contributing to success is providing analytical products and data that can increase the probability of the project accomplishing its cost, technical, and programmatic objectives. Contributing to success *is not* producing a cost estimate that meets the project's expectation so that everyone is happy.

The Cost Estimating Process

A good cost estimating process can improve the quality of an estimate and work to minimize biases. Every organization should develop a process or adopt a process from a known and recognized source, such as the GAO Cost Estimating and Assessment Guide. A standard process provides traceability and repeatability, keeping the estimator focused on the task at hand and creating a documented basis of estimate. The process should capture best practices and include activities that counteract biases and keep the estimate as objective as possible. Finally, the process forms the foundation of the cost estimator's own story: a story that will describe how the estimator has used the facts, data, and subjective assessments to build a credible estimate.

Injecting Reality

The cost estimating process provides a means to achieve an outcome, but the process may not lead the estimator to consider all the possible dimensions of the problem at hand. One can take a very narrow focus of the cost estimating problem: the customer wants an estimate – I will deliver an estimate. Or, the estimator/analyst can take a broader point of view. This broader point of view considers not only the data at hand (WYSIATI) but enables the estimator to seek out and consider data that exists beyond what is needed to perform the job. By looking at the bigger picture the estimator is forced to incorporate a reality that is larger and more complex. By increasing the breadth and depth of the information used in the analysis, the effect of biases can be reduced simply by broadening the estimator's perspective.

I have identified four general sources of information that can inject a dose of reality into an analysis. The first of these, historical data, is worthy of a more detailed discussion and thus, will be covered comprehensively in the next section. The second source of reality is the technical and programmatic experts who are supporting the project or study. These individuals often have extremely useful information, information that may not be shared in team meetings or short sidebar conversations. I strongly suggest that you meet with these individuals one-on-one. One-on-one meetings minimize the effect of group think and may encourage a more open conversation. However, experts are subject to the same biases discussed earlier, so take what you learn and compare it to data from history, cost experts, and other technical and programmatic experts. Remember that facts are unbiased but that the context within which those facts are communicated may be biased. Try to understand their motivations and factor that into how you use their input.

All of us who work in the field know and understand the value of discussing our estimates and analyses with other cost professionals. I simply mention it here for the sake of completeness. However, I do want to emphasize that having another cost professional, preferably someone who is considered the expert in your organization, review and critique your estimate is an excellent way to get an outside point of view. An outside point of view can be invaluable for finding areas of questionable judgment and methodology. A good cost expert can also identify other analyses, data, and techniques that can improve the estimate.

Events that take place outside our organizations can have a meaningful impact on our estimates. Our various organizations do not exist in a vacuum. Actions by national and international leaders directly affect the perceptions and decisions of our political leaders. Broader economic trends directly affect decisions by companies. A prime example of this is what happened after Russia's annexation of Crimea and fomentation of instability in Ukraine. Russia's actions highlighted our dependence on the Russian RD-180 rocket engine used in the Atlas V launch vehicle and our dependence on the Russian Soyuz for astronaut transportation to the International Space Station (ISS). Congress, unhappy with Russia's actions, has chosen to address these dependencies by encouraging the development of a domestic counterpart to the Russian RD-180 and providing greater support for the NASA Commercial Crew Program. United Launch Alliance (ULA), the company responsible for the Atlas V, has responded by partnering with a company called Blue Origin to develop an RD-180 replacement that uses different fuel. How all these actions will impact future cost estimates is unclear at this time. However, the potential for impacts must be recognized and addressed.

When gathering additional information, the analyst needs to be aware that a couple of biases can creep in. As we began to incorporate outside information into our analysis we cannot help but evaluate how this information fits into the story we are telling ourselves about our estimate. If the information fits, we readily accept it. However, if the information is counter to our story, we will tend to reject or minimize it. This is the confirmation bias in action. Therefore, be open minded to information that does not "fit" and aware of actions that lead to the elimination or minimization of inconvenient data.

The second bias is actually a very interesting paradox that psychologists have observed. If we know little about a subject area, our predictive ability in that subject area is very low and our confidence in our predictions is correspondingly low. As we gain more knowledge and data, our predictions improve and so does our confidence. However, at some point, additional knowledge does not improve our predictions, but it does increase our confidence in our predictions. We therefore become stronger advocates for our estimate without improving the quality of our estimate. As we gather more information and broaden our perspective we must be aware of the potential for this "overconfidence bias" and take care not to place more confidence in our estimate than what is directly supported by the data.

History

In my opinion the cost community's greatest asset is our historical data and perspective. We have access to information on all types of projects that document the what, when, why, who, and how. And because of the internet, we are no longer limited to the information we have in our libraries and data bases, or what we can learn through our personal contacts. The amount of information available to the professional cost estimator has never been greater, and the volume grows daily. One word of caution. As you study history be aware of your mind's attempts to confirm preconceived notions (confirmation bias) and a growing sense of rightness as your knowledge increases (overconfidence).

So what can be learned from studying history? Well for starters, we can learn how projects are managed and systems are developed. We can learn about typical problems and issues. We can learn about the challenges that projects have faced and how they have achieved success despite these challenges. We can also learn why projects fail, what were the root causes, what did and did not work. All of this information provides background and context. Background and context broaden our perspective and strengthen the foundation of our own story.

Studying history brings a dose of reality to our estimates. By examining specific technical and programmatic analogies to our estimating problem we anchor ourselves in actual results. We also gain knowledge about realistic boundary conditions for subjective assessments – highly useful for performing sensitivity and uncertainty analyses. History can provide data for developing and supporting ground rules and assumptions (it was done like this in the past, so we can assume it will be done like this in the future).

One of the most important benefits of studying history is the ability to use real cost data to establish base rates. Base rates are simply prior knowledge, such as the knowledge of what it cost to develop analogous systems, calculated values based on a sample such as average cost per pound for developing new spacecraft, or factors or ratios (for example, project management is 10% of the hardware cost) based on actual data. Base rates are extremely useful for quick sanity checks to tell if an estimate is reasonable and thus as a check against rampant optimism. Base rates can also be used to quickly provide decision makers with information that enables real-time decisions, eliminating the need for more time-consuming detailed analyses.

Historical project data will tell you what worked, what did not work, and what unforeseen problems were encountered. This information can be helpful in finding ways to be useful and add value beyond the cost estimate. So how does this work? Let's say, for example, that you are doing an estimate for a project that is using a certain type of detector. You know from looking at your historical data that this type of detector proved to be a manufacturing challenge on a previous project, causing schedule slips and cost overruns. You take this knowledge to the detector expert on the study team, using it to engage in a conversation around the current state of the art and any changes to the technology since the previous project. What you learn can then be used to inform and substantiate your cost estimate, providing a more credible, supportable, and defendable result. You are adding value because you are putting forth the effort identify and understand the technical challenges and, by sharing this information with the study team, you are making everyone aware of the issue and creating the opportunity to address and solve a specific problem.

Validation

How do you answer the question "is my estimate reasonable?" One way to determine the reasonableness of any estimate is to validate the estimate relative to historical experience. Figure 2 shows an estimate for a science mission plotted against weight. Also on the plot are costs for several completed science missions as well as a trend line.

The simple comparison shown in Figure 2 can be used both as a check and as a communications tool to demonstrate the validity of the estimate. By using a plot like this you can determine if the estimate is consistent with actual costs for similar projects or graphically compare the costs to close analogies. If significant deviations are present you can evaluate the rationale for the deviations and determine if they are credible. Finally, you can demonstrate to the customer the reasonableness of the estimate relative to similar past projects.

Obviously, you should not limit yourself to basic cost versus weight plots when doing validation. What is important is that you compare your cost estimate to the cost of historical analogs. If your estimate seems reasonable when compared to these analogs, then you can validate, for yourself and for your customer, that you have a credible estimate.

Validation by comparison to historical experience is an excellent way to determine the

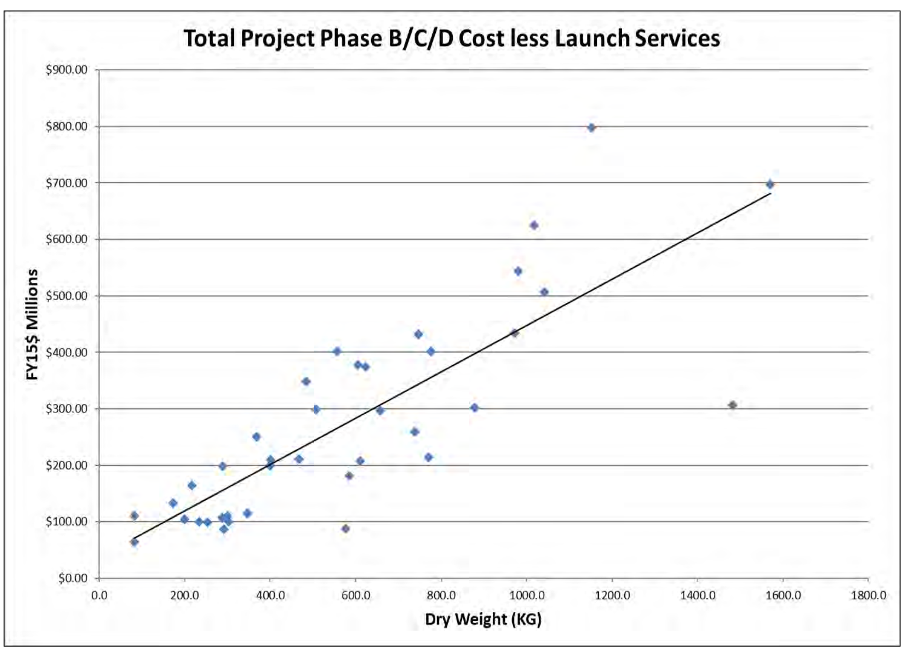


Figure 2. Example of Graphical Estimate Validation.

reasonableness of a cost estimate, but requires that the estimator have access to either a large data set of similar projects or a few close analogies or both. However, we often find ourselves in the difficult position of doing a cost estimate for a project for which we have little or no relevant historical experience. How can the estimator handle this problem?

The first step is to study the data you have. Look for parallels and similarities, if not in whole then at least in part. For example, the systems engineering processes for all large research and development programs should generally be the same. Another example, while the overall system may be different from anything done in the past, certain subsystems or even components might be similar to past efforts.

If you are presented an estimate from another source (remember Anchoring?), an approach that I have used in the past is one I call calibrate and evaluate. In calibrate and evaluate you take the existing estimate along with its technical and programmatic data and reproduce it using a known cost model. You could use PRICE TruePlanning or SEER-H or an in-house model. What is important is that the model gives you the ability to adjust complexity or new design parameters to make the model match the estimate. Once you have calibrated the model to the estimate, you can evaluate the model settings for reasonableness relative to the project's known challenges. You can share this information with the customer. This approach also gives you a model for developing your cost estimate and a mechanism for explaining your estimate relative to previous estimate.

Another approach that we have used is called disaggregation. In disaggregation you take your cost estimate and decompose it into function elements based on historical data. The end product might look something like Figure 3.

Once you have the disaggregated cost you can review the estimate with functional experts, which can be very enlightening. They may be able to validate the results or you can use their knowledge and experience to modify your functional estimate. In either case you will have valuable input that you can use to demonstrate credibility.

One final approach to validating an estimate is to use a Bayesian approach. Bayes Theorem is based on conditional probability, the chance that something is true given a prior state or condition. At the heart of the Bayesian approach is the concept of that there is existing knowledge. Another term for this existing knowledge is the base rate. Now, in the Bayesian world, base rate is used to refer to a probability or a proportion for a population (such as, 30 out of 100 people will get cancer over their lifetime). The way that I am defining base rate for this paper is different. I consider a base rate to be any cost data, model, average, or factor that can be used to quickly evaluate an estimate. For example, the simple linear CER in Figure 2 could be a base rate, or an average cost per pound for a certain type of hardware, or the ratio of project management/

	OCM WBS	Booster Motor	First Stage	Booster Hardware	-	Upper Stage Engine	First Stage Engine	-	-	-	Integ Vehicle	Cost Element TOTAL
V	VEHICLE SEGMENT	**$ 338**	**$ 248**	**$ 112**	**$ -**	**$ 256**	**$ 399**	**$ -**	**$ -**	**$ -**	**$ -**	**$ 1,353**
V1	Expendable Hardware Mfg	$ 39	$ 188	$ 14	$ -	$ 98	$ 199	$ -	$ -	$ -		$ 538
V2	Reusable Hdwe Refurb	$ 200	$ -	$ 44	$ -	$ -	$ -	$ -	$ -	$ -		$ 243
V3	Vehicle Spares	$ 13	$ -	$ 8	$ -	$ -	$ -	$ -	$ -	$ -		$ 21
V4	Vehicle Overhauls	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -	$ -		$ -
V5	Acceptance Test	$ -	$ -	$ -	$ -	$ 21	$ 32	$ -	$ -	$ -		$ 53
V6	Sustaining Engineering	$ 48	$ 23	$ 18	$ -	$ 82	$ 108	$ -	$ -	$ -		$ 279
V7	Program Mgt & Support	$ 38	$ 37	$ 29	$ -	$ 55	$ 60	$ -	$ -	$ -		$ 218

Figure 3. Example Disaggregation Output.

systems engineering cost to subsystem hardware. When an estimate is close to a base rate that should increase confidence that it is reasonable. When an estimate deviates significantly that should raise questions. Deviation from a base rate does not mean the estimate is incorrect; however, you must be able to explain the rationale for the deviation.

In his award-winning paper, “Bayesian Parametrics, How to Develop a CER with Limited Data, and Even without Data,” Christian Smart discusses how Bayesian approaches can be used to estimate cost. I will not attempt go into detail on Christian’s methods or findings, only to say that he puts forth an interesting and useful way to apply Bayes’ Theorem to the problem of cost estimating with very little data. The same approach could possibly be used to validate an estimate when there is little or no directly applicable historical data by analytically comparing related historical information to the estimate.

The use of validation to counteract biases should be obvious. Validation offers a powerful means to battle any bias by injecting a dose of reality through the requirement that the estimate be explainable relative to past experience. When you have a large historical data set of comparable projects, validation is easy. However, when historical data is non-existent or lacking in some way (like age), the estimator must be creative in finding ways to use whatever information, techniques, and tools are available to determine if the estimate is reasonable. Always be aware that the less historical truth available for validation, the greater the likelihood of a biased estimate.

Risk and Uncertainty

In his book The Signal and the Noise, Nate Silver tells the story of a flood in Grand Forks, North Dakota. The flood forced the evacuation of nearly all of the city’s 50,000 residents and caused billions of dollars in damage and clean-up costs. However, the flood was not a surprise. Due to unusually heavy snowfall (even by North Dakota standards) the National Weather Service had predicted that the nearby Red River would crest at 49 feet, two feet below the levee height of 51 feet. What the Weather Service did not tell the city leaders of Grand Forks is that the margin of error in their forecast was plus or minus 9 feet. In reality, the margin of error implied about a 35 percent chance of the river rising above the levee, and the actual crest was 54 feet, inundating the town.

Obviously, failing to communicate risk and uncertainty can have serious consequences when making predictions and cost estimates. Point estimates create a false sense of certainty and deprive decision makers of useful information. It is also hubris on our part to believe that we can forecast the financial outcome of a technically complex, multiyear project with such a high degree of accuracy.

For the purposes of this paper, I define risk as the chance of loss, the chance that something could go wrong or not work as planned. Uncertainty is the indefiniteness about the outcome, the margin of error. Most of the time cost estimators use the term Risk Analysis to cover both risk and uncertainty. However, I argue that the Confidence Level (CL) analysis is a broader and therefore more appropriate term.

One mistake estimators can make is to treat uncertainty as only model error and risk as only those items identified on the project’s risk list (or 5x5 matrix). Uncertainty includes not only model error but also uncertainty on the model inputs, especially those inputs that are subjective. When the estimator limits the scope of the confidence level analysis it can create a false sense of security which can lead to uninformed decisions.

The focus of a sensitivity analysis should be on the subjective inputs that are incorporated into the analysis. As we well know, numerous subjective judgments and assessments must be made in the development of a cost estimate. In many cases, these judgments have significant consequences on the final estimate value. A sensitivity analysis is an excellent way to quantify the impact of these judgments in a way that can be used to develop an uncertainty range around the point estimate, and identify which judgments have a significant impact on the cost.

By combining the results of the sensitivity analysis with a description of the conditions that describe the extreme points, the analyst is providing the decision maker with valuable context for understanding how different assumptions influence the possible cost outcomes. This helps the decision maker understand what is and what is not important to the estimate and what decisions could affect the outcome.

When we assign probability distribution functions to input or model settings, or when we assign probabilities to specific outcomes, we have moved into the realm of the confidence level analysis. A good confidence level analysis accounts for uncertainty in our inputs, uncertainty in our judgments, uncertainty in our estimating methods, and project risk. The assignment of probabilities (outside of those derived through statistical means) is often highly subjective. The human mind is not good with probabilities (if it were, then casinos and lotteries would not exist). We tend to overestimate the occurrence of low probability events and underestimate the occurrence of high probability events. Therefore, be clear with your customer or stakeholder on how you derived the uncertainties and probabilities.

Despite the sometimes highly subjective nature of a confidence level analysis, by providing a probabilistic range or s-curve, the estimator is giving the decision maker a sense of the risk exposure assumed when a project moves forward at a given funding level. Knowledge of this risk information will hopefully influence organizations to make more rationale, and less biased decisions around funding and project content.

One final note about risk and uncertainty. The NASA cost community has noticed that the implementation of a requirement for joint cost schedule confidence level analyses (JCL) on all major projects has changed the conversation about cost and schedule estimates. A shift has occurred from a discussion around the validity of the estimate to a discussion around the inputs to the JCL. This focus away from outcomes and towards the factors that affect the results appears to be leading to better overall management decisions on funding levels and project risk. Hopefully this represents a trend towards using cost estimating and analysis products to reduce cost overruns and schedule slips.

The Value of Expertise

The combination of simple mathematical models and expert opinion provides the best predictions. This observation has been made by experts in the field of human behavior and decision making such as Daniel Kahneman, Nate Silver, Malcolm Gladwell, and Douglas Hubbard. It has also been noted by leaders in the field of parametric estimating, such as Joe Hamaker. In his paper "What are Quality Cost Estimates or the 260 Hz Cost Estimate" Joe makes the following statement:

> "But my point is that many of us close to the practice do have some innate and intuitive ability, honed by years of being associated with the cost estimating game, that is usually pretty reliable when it comes to judging the quality of a cost estimate."

While this observation seems counterintuitive in light of all the previous discussions in this paper on biases in our thinking, there actually is strong evidence to support it. The key is that the models must be logical and the experts must be real experts. The models get you an answer, the expert evaluates that answer to see if it really fits the data.

You can apply your own expert judgment by simply asking the "why" question. For example, if you have an estimate that deviates significantly from the validation data, you need to ask "why?" In asking "why" you are looking for technical or programmatic explanations for the deviations. The danger in asking the "why" question is that it can lead to the confirmation bias: you overemphasize data that supports your result while underemphasizing or dismissing data that might lead to a different outcome. However, don't let that concern stop you from questioning your estimates. One approach to counterbalance the tendency to confirm our judgments is to have a different expert examine our estimate and ask the "why not" question.

Our subconscious can process a tremendous amount of data and when properly trained and combined with our natural ability to simplify complex situations, becomes a powerful tool. In his book Blink, Malcolm Gladwell tells the story of an ancient Greek statue that was purchased by the Getty Museum for millions of dollars. The museum had carefully checked the history of the statue and had performed a sophisticated scientific investigation to determine its legitimacy. However, experts in ancient Greek artifacts who saw the statue knew almost immediately that it was a fake.

Just like experts in ancient Greek artifacts, our minds can be trained to recognize good cost estimates and bad cost estimates. Malcolm Gladwell makes this point in Blink: "Just as we can teach ourselves to think logically and deliberately, we can also teach ourselves to make better snap judgments." So how do we, as cost estimators and analysts, teach ourselves?

As Joe Hamaker previously pointed out, it helps to have years of experience. Unfortunately, not everyone has years of experience. So if you are new to the profession, maximize the experiences that you have had. Learn as much as you can from them, not just how to use the models or develop and present estimates (and these are important skills), but also take the opportunity to learn something about technology, systems design, requirements, organizational behavior, management, etc. As Yogi Berra said, "you can observe a lot by just watching."

A great source of information for building up your expert knowledge are our databases and libraries. I have already covered the value of historical data, so I will not repeat it here. However, I do want to reiterate the value of using historical data to establish "base rates" that can be used to test the credibility of any estimate. You don't have to walk around with tables of numbers in your head, but you should be familiar enough with historical experience so that you can sense the value of a project to at least an order of magnitude.

My first job in the cost profession was in analyzing data and developing cost models. I quickly figured that to do that job, I had to learn something about the space systems I was analyzing. I did not have to know enough to actually design and build the systems, but I did need to have a layman's understanding of how they worked and what characteristics about them really drove cost. When I began working at NASA, I got the opportunity to spend more time talking to professionals in other technical disciplines. These experts filled in gaps in my knowledge and helped me understand the state of the art, and what was not difficult versus what was a real technical challenge. I strongly encourage every estimator to spend time talking to technical experts. Yes, they can be as biased as the rest of us. But by combining their knowledge and expertise with historical information, you can develop a good understanding of what is and is not important to the cost of a system.

Education is very beneficial and I certainly encourage everyone to get as much as possible. Obviously, science, math, and engineering are all extremely useful to the cost estimator. However, don't discount the value of non-technical courses, especially those that encourage creative thinking and develop communications skills. Also useful are courses that increase your technical knowledge in those systems that are the primary focus of your cost estimates and analyses. As part of your education, you should take training in cost estimating and analysis as well as in related disciplines such as earned value management and scheduling. Becoming certified is a great way to demonstrate your knowledge and expertise.

Reading is a great way to increase your general knowledge. This paper is the result of an intentional effort I undertook several years ago to read more non-fiction. I started by reading books on the development of space systems, then I read books about the development of other technologies. I also read popular books about human behavior and decision making, such as "Freakonomics" by Steven Levitt and Stephen Dubner; and "The Wisdom of Crowds" by James Surowiecki. I began to discuss what I was reading with other cost professionals, and their own readings and feedback led me to discover several of the books that became the foundation for my understanding of how biases affect cost estimates.

When you look for what to read let curiosity be your guide. Obviously, books and articles that pertain to cost estimating, cost analysis, mathematical modeling, and relevant technical subjects are good places to start. However, don't overlook biographies, books on organizational behavior, or other fields of science. I have found that while the particulars may be different, other disciplines have dealt with problems and issues that are similar to those that we cost professionals face. We may have different jobs and responsibilities, but at the end of the day we are all human.

An excellent way to grow the knowledge needed to be a cost expert is to become engaged in the greater professional community. Through organizations such as ICEAA you are exposed to new ideas and ways of doing estimates and analyses. You can learn from the experiences of other professionals and use that knowledge to develop better cost estimates. You can also develop relationships with leaders in the cost profession, contacts who will prove useful in helping you perform better estimates or improve your estimating capabilities. As mentioned previously, ICEAA and other professional societies also offer training and certification, both valuable ways by which you can grow professionally.

Finally, allow me to add this one final recommendation: be open to new data, thoughts, and ideas. Our human nature is to think we have it all figured out. Our biases enable us to fool ourselves into believing that we are being rational when in fact we are only responding to the loudest voice in the room, looking for plausible explanations, and confirming what we already believe. It takes courage to question ourselves and such questioning can be downright uncomfortable. But it is only through an openness to the possibility that we could be wrong that we can create the space that enables us to grow and mature as professionals.

Telling Your Story

As I discussed earlier, we are more influenced by stories and anecdotes than we are by facts and data. Therefore, to make a convincing argument that your cost estimate is good, you have to tell a good story. What is a good story? In my opinion, it is one that relates facts and data to logical action and speaks in a language the customer understands.

Part of being able to tell a good story is understanding your customer, especially how your customer views the world. The most basic understanding of your customer begins with knowing what is and is not important to them. For example, are they trying to design a program to fit within a predetermined budget profile? At they trying to sell a new start? Is the estimate in support of a Key Decision Point (KDP)? Or are you doing an analysis of alternatives?

Each of these scenarios places different requirements on the project or team lead, and thus different requirements on the estimator. If the goal is to fit a program within a predetermined budget, the estimating requirement becomes how much capability can be bought for the money. If the purpose is a new start, there will likely be pressure to keep the cost as low as possible. If the estimate is in support of a KDP, then the project manager will worry that an unfavorable outcome could derail the project or lead to cancellation. Knowing and understanding what your customer is facing can help you craft your story so that it aligns with your customer's goals. Your job is to help your customer succeed. But success is not telling them what they want to hear, but rather what they need to hear.

It is also important to try and have some idea of what your customer believes. Almost everyone who has had any experience in project management has stories of how cost was affected (or not affected) by something that happened. Often these experiences become beliefs, and once we believe in something it has a direct influence on how we respond to new information. In fact, psychological research shows that once we believe something, the confirmation bias kicks in, causing us to give more weight to information that agrees with our position and to discount or disregard information that disagrees with our belief (beliefs trump statistics). Thus, understanding your customer's belief system (at least with regards to project management and cost estimating) can help you prepare your story in a way that gives your estimate a higher probability of being understood.

So what do you do if your estimate runs counter to a customer's strongly held beliefs? Chris Mooney, in an article for *Mother Jones* titled "The Science of Why We Don't Believe Science" offers the following advice: "If you want someone to accept new evidence, make sure to present it to them in a context that doesn't trigger a defensive, emotional reaction." Thus it is important that your analysis is explained in the context of their worldview, using language and metaphors that they are comfortable with and therefore less likely to create a defensive reaction. Christie Aschwanden, writing for the website FiveThirtyEight gives similar advice in an article titled "Your Brain Is Primed To Reach False Conclusions." Here is what Christie has to say:

> If you want someone to accept information that contradicts what they already know, you have to find a story they can buy into. That requires bridging the narrative they've already constructed to a new one that is both true and allows them to remain the kind of person they believe themselves to be.

Therefore, to effectively tell your cost estimate story you must know what is important to your customer and understand and be sensitive to what your customer believes. With that knowledge you should begin your story with the facts and data used in your estimate. These should be the facts and data (including estimated parameters) on which everyone can agree. You can then introduce other pertinent information, such as base rates and historical experience. Be prepared to defend your decisions to include information that may contradict the customer's beliefs or worldview.

The next step in telling your story is to define the relationship between the objective information (facts, data, base rates, etc.) and your subjective assessments. My rule of thumb is to make this part transparent and keep it simple. At this point you introduce and review the ground rules and assumptions needed to facilitate the estimate.

Once you move into the subjective aspect of your analysis you have entered the realm where biases are more likely to occur. The best way to deal with possibility of biases and subjectivity is with an uncertainty analysis. As I stated earlier, by combining the results of the sensitivity analysis with a description of the conditions that describe the extreme points, the analyst is providing the decision maker with valuable context for understanding how different assumptions influence possible cost outcomes. By providing a range of possible outcomes that vary depending upon one's beliefs, you are giving the decision maker the space to adjust their belief system without putting them in an overtly defensive position.

I want to reiterate the importance of validating your results to the telling of your story. Bear in mind that your customer will either accept your validation, question the data you use in your validation, or try to introduce new (and sometimes irrelevant) data into the discussion. Always be prepared to defend your approach to validation and be willing to address any new information.

The goal of your cost analysis story is to show that your estimate is a logical outcome of the evidence. Much like a lawyer, you have built your case around facts and data, relevant historical experience, consistent subjective assessments, an examination of the uncertainties, and a comparison to valid, real world experience. As one of our senior analysts, Richard Webb, likes to say, show that your estimate is credible, supportable, and defendable.

Signs of a (Possibly) Overtly Biased Estimate

Even if you do everything right, bias is going to creep into your estimate. After all, everyone involved in the cost estimate is human, and the psychological evidence is overwhelming: humans are irrational, biased creatures. If we accept for the moment that no estimate can be totally unbiased, how do we know if our estimate is too biased? While I don't have a foolproof tried and tested approach, I have identified four things to be on the lookout for. The appearance of any of these signs does not automatically mean your estimate is overtly biased. Use them as warning flags that something could be wrong, and that further analysis may be needed. Be aware that when you are trying to find support for a decision, it is easy to fall into the trap of the confirmation bias.

The first red flag that you (or someone) may have biased an estimate is the discarding of applicable data. In this case, I am talking about data that could be used for validation or as an analogy or as part of a data set for developing a cost estimating relationship (CER). An example of this would be ignoring an appropriate analogy that has a higher cost than your estimate.

A red flag that is in some ways the opposite of the previous one is the placing of significant emphasis on a single expert opinion, data point, or other bit of information. An example of this is when you base your estimate on one or two pieces of data while either failing to validate the information or ignoring other data that gives a contradictory conclusion. Because sometimes we do not have good historical data, this sign can be the result of severe limitations in your data set or other available information, and not the result of a bias.

The use of an inappropriate analogy or extrapolation to support, perform, or validate an estimate is a sign that you are trying to bias the estimate to obtain a certain outcome. An analogy should always embody key characteristics of the system being estimated such as function, performance, and development approach. If you cannot establish clear similarities between the system you are estimating and an analogy, you should not use that data. It would be like basing the cost of a rocket engine on a robotic spacecraft. Inappropriate extrapolation can occur when you try to use the cost experience of a small, simple system to estimate or validate the cost of a larger, more complex system. Over the years I have observed that small teams, given a focused task and assured funding, can accomplish amazing things, such as the development of a new technology or the building of a small spacecraft. While these experiences provide useful information for cost estimating, these experiences are not appropriate for estimating larger, more complex missions.

Anytime your estimate deviates significantly from the historical trend and/or reasonable analogs that is a sign of possible bias. There may be good reasons for the deviation. However, if the deviation is significantly below historical experience your estimate should be examined closely. Once again, it is all too easy to find evidence to support your position when you are looking for it, so I recommend having an independent set of eyes review your work.

Any estimate that depends on changes in historical business practices should immediately raise a red flag. Examples do exist of projects or companies (i.e. SpaceX) who were able, by doing things differently, to achieve success for significantly less money than the historical record would predict. Whenever I am faced with a customer who says that they can do it for significantly less than it has been done before, I simply ask them "how?" If the customer or study team can build a story of how they are going to operate differently, I will use that in the basis of my estimate. But I will also do a sensitivity analysis showing what will happen if they actually do "business as usual."

We are rightfully proud of our hard work. When you have put significant time and energy into an estimate or analysis you are going to naturally be defensive if your judgment is questioned. However, being unwilling to accept or incorporate new data into your analysis is usually a sign that you are overconfident and therefore biased. Not all new information is useful or appropriate, but neither should it be dismissed simply because it does not support your results. Anytime someone questions your judgment with regards to the subjective part of the estimate treat it as an opportunity to test your reasoning by telling your story and seeing if the audience supports your conclusions. Being open to new ideas or perspectives is an excellent way to guard against biases.

Summary and Conclusion

I realize that I covered a tremendous amount of material in this paper. I like to believe that all of it is useful to the cost analyst, however; there are four key points that I want everyone to remember.

First, we are all biased, and these biases affect how we do are estimates and how our estimates are received. These biases cause us to put too much faith in our own abilities, carefully select information that supports our conclusions, and be more influenced by what people tell us than by facts and data. We are not rationale thinkers. While we cannot force ourselves to become unbiased, we can learn to recognize what biases look like and take steps to minimize their impact.

Second, you can control your behavior, but you can only influence others. Being proactive to minimize the impact of bias on your cost estimates and analyses is smart. Trying to force others to do the same is futile. Focus on building a credible, supportable, and defendable estimate. Understand what your customer needs and why. Use that information to be add value and make a difference.

Third, the cost community's greatest asset is our historical data and perspective. Historical data provides background information that can increase knowledge and understanding. Historical data provides ground truth that can be used to balance information from other sources. Use historical data to bound uncertainty, validate your estimates, and establish base rates.

My final point is simply this: a valuable cost estimate (or analysis) is not one that gives the customer the answer they want, but gives them the answer they need. Dare to go beyond providing numbers (as valuable as those numbers might be) to providing true value to enabling customer success. That does not mean the customer will always agree with your answer. But, if you can get the customer to respect what you have done, to recognize that even in disagreement you are working to further their objectives, they more likely to be influenced by your analysis.

Being a cost estimator is hard. Everything that goes into an estimate or analysis can be questioned. We are often pressured to reach a preconceived conclusion. It takes courage, it takes knowledge, and it takes experience to produce an objective result, a result that adds value. And it takes an understanding of how estimates can be biased and how to overcome those biases. I wish you well on your journey.

Bibliography

Ariely, Dan, *Predictably Irrational*, Revised and Expanded Edition, New York: Harper Perennial, 2009

Aschwanden, Christie, "Your Brain is Primed to Reach False Conclusions." *fivethirtyeight.* February 17, 2015. <http://fivethirtyeight.com/features/your-brain-is-primed-to-reach-false-conclusions/

Gladwell, Malcolm, *Blink, The Power of Thinking Without Thinking*, New York: Little, Brown and Company, 2005

Hamaker, Joseph W., "What Are Quality Cost Estimates? Or the 260 Hz Cost Estimate," *Journal of Parametrics* Vol. 25, Issue No. 1, 2007: 1 – 7

Hubbard, Douglas W., *How to Measure Anything*, New Jersey: John Wiley & Sons, 2010

Kahneman, Daniel, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011

Levitt, Steven D. and Dubner, Stephen J., *Freakonomics, a Rouge Economist Explores the Hidden Side of Everything*, New York: Harper Perennial, 2009

Mlodinow, Leonard, *The Drunkards Walk: How Randomness Rules Our Lives*, New York: Pantheon Books, 2008

Mooney, Chris, "The Science of Why We Don't Believe Science." *Mother Jones.* May/June 2011. <http:// www.motherjones.com/politics/2011/03/denial-science-chris-mooney>

Silver, Nate *The Signal and the Noise: Why most Predictions Fail but some Don't*, New York: The Penguin Press, 2012

Smart, Christian, "Bayesian Parametrics: How to Develop a CER with Limited Data and Even Without Data," *Proceedings of the 2014 International Cost Estimating and Analysis Association Professional Development and Training Workshop*, Colorado: June, 2014

Surowiecki, James, *The Wisdom of Crowds*, New York: Anchor Books, 2005

Andy Prince *is recently retired from NASA/Marshall Space Flight Center, where he managed the Program Planning & Control Office. He is perusing new career opportunities and recently accepted a job with New Horizons 5. A lifetime member of ICEAA, he has 38 years of cost estimating and analysis experience and is past ISPA Board Chair and ICEAA co-President. Andy has received numerous awards during his career, including the Frank Freiman Lifetime Achievement Award.*

Cost Estimating Relationships for Recurring T100 Flyaway Costs

Kyrie M. Rojo
Edward D. White, Ph.D.
Brandon M. Lucas, Ph.D.
Jonathan D. Ritschel, Ph.D.
Shawn M. Valentine

Abstract: This research investigates a dataset of over 80 Air Force and Navy aircraft and applies regression techniques to create two cost estimating relationships (CERs) for predicting recurring T100 flyaway costs, depending on where in the acquisition lifecycle the estimate takes place. The first CER explains 89 percent of the variability and can be applied prior to Milestone B (MS B). The second CER explains 88 percent of the variability and can be applied between MS B and MS C. Significant cost drivers identified include stealth, cohort, empty weight, the natural log of speed, legacy aircraft, fighter aircraft, and Engineering and Manufacturing Development costs. This research is the largest aircraft regression study to date for recurring T100 flyaway costs and can be used by cost analysts as a cross-check in early estimations.

Introduction

The Air Force is preparing for the future of air superiority with the introduction of new aircraft such as the B-21, T-7, E-7, and the Next Generation Air Dominance (Department of the Air Force, 2021). These programs need a credible and accurate life cycle cost estimate for the acquisition to be successful. In the Department of Defense (DoD), flyaway costs constitute most of the procurement costs in aircraft acquisition: prime mission equipment, systems engineering and program management, test and evaluation, warranties, engineering changes, nonrecurring startup costs, and government-furnished equipment (Department of Defense, 2022). Thus, accurately estimating flyaway costs is a key component in establishing a realistic acquisition program baseline.

Cost estimating relationships (CERs) for airframes or flyaway costs are typically derived using the 100th production unit. When the 100th production unit is not available, that value is derived via a cost improvement curve (e.g. a learning curve utilizing cost data rather than hours). This 100th production unit is therefore referred to as an UC100, the T100 unit cost, or simply T100 (Department of Defense, 1992). A T100 flyaway cost, therefore, looks specifically at the flyaway costs associated with the T100 unit.

This research is the largest aircraft regression study to date for recurring T100 flyaway costs. The study employs and analyzes historical data to create two CERs. These CERs utilize data prior to production and identify key cost drivers. The results from this paper can be used by program managers or estimators early in the aircraft acquisition life cycle as a cross-check to other methods that might estimate the T100 flyaway cost.

Background

Flyaway costs occur during the production phase of an aircraft, also known as the investment phase (Mislick & Nussbaum, 2015). During this phase, a build-up technique is often used for cost estimation because actual cost data is available. However, when calculating unit costs such as the T100, a cost estimator should use a cost improvement curve (CIC) (Government Accountability Office, 2020). A CIC addresses the phenomenon that as tasks are repeated, learning occurs making the task more efficient and therefore cost less (Department of Defense, 2022). A CIC measures the reduction in terms of cost, while the more colloquially known learning curve measures the reduction in terms of hours. This analysis employs the CIC construct.

There are two leading theories on CICs: Unit Theory and Cumulative Average (CUMAV) Theory. Both theories address the learning phenomenon previously mentioned, but unit theory assumes a reduction in unit costs while CUMAV assumes a reduction in cumulative average costs. Since T100 costs are unit costs and it is the predominant approach amongst Air Force practitioners, the unit theory cost improvement curve is the one we adopt. While a CIC is useful for determining the production unit costs of an aircraft, it should only include recurring costs to prevent skewing the results (Department of Defense, 2022). Therefore, the term flyaway cost is in reference to recurring flyaway costs as opposed to total flyaway costs.

To understand how the T100 flyaway cost is ascertained with actual aircraft production data, we guide the reader through the following process. First, normalize the data to remove the effect of escalation to constant price (CP$) via the Produce Price Index (PPI) 3364, which details price changes in aerospace products and parts (Bureau of Labor Statistics, 2022). Normalizing to CP$ for a CER is a best practice according to the 2021 OSD-CAPE *Inflation and Escalation Best Practices for Cost Analysis*. One then calculates the average unit cost (AUC) by dividing the lot's recurring flyaway costs by the total number of units produced (see Equation 1).

$$AUC\ of\ Lot_t = \frac{Recurring\ Flyaway\ Costs\ in\ Lot_t}{Number\ of\ Units\ in\ Lot_t} \quad (1)$$

Equation (2) shows how the lot midpoint (LMP) for Lot 1 is calculated.

$$LMP\ of\ Lot_1 = \begin{cases} For\ Lot\ Size < 10, Lot\ Size\ \div 2 \\ For\ Lot\ Size\ \geq 10, Lot\ Size\ \div 3 \end{cases} \quad (2)$$

For all subsequent lots, the LMP is calculated by adding the first (F) and last (L) unit number in a lot, plus two times the square root of F times L, then divide the total by four (Equation 3).

$$LMP\ of\ Lot_{t>1} = \frac{F + L + 2\sqrt{F * L}}{4} \quad (3)$$

A linear regression is then performed via the natural logs of the AUC and LMP to estimate the flyaway cost for any unit, taking into consideration cost improvement curve and economies of scale. Equation (4) shows this where ln(LMP) (the explanatory variable) is regressed onto ln(AUC) (the response variable).

$$\hat{Y}_X = \hat{\beta}_0 + \hat{\beta}_1 * X \quad (4)$$

Back-transforming from log space, we arrive at the customary cost improvement curve (5)

$$Y_X = A * X^b \quad (5)$$

Where:

Y_x = the flyaway cost of unit X

A = the theoretical cost of unit one (T1)

X = the unit number

b = the theoretical slope of the cost improvement curve

Once these calculations are made and the cost improvement curve equation is computed, one evaluates the equation at X = 100 or Y_{100}. This is the flyaway cost of unit 100 or the T100 flyaway cost. This process results in an approximation of the recurring flyaway cost at the theoretical 100th unit while considering the learning effect. This process is what generated the response data we obtained for our study.

Next, we turn to prior published sources to identify possible explanatory variables for the CERs. Unfortunately, we could find no prior studies that predicted recurring T100 flyaway costs (nor any type of flyaway cost for that matter). The most similar study conducted was published in a series of papers by RAND from 1972 to 2001 and investigated cost drivers for different elements of aircraft airframes. To cast a broader net for research related to flyaway costs, we looked for studies focused on production costs; but it resulted in only one report from 1991, which created cost models for production support elements. Altogether, we explored five prior studies. Table 1 lists and summarizes these studies. For our purposes, they serve as a reference to consider which explanatory variables might be predictive of T100 and the development of the CERs within this article.

Table 1. Summary of Previous Research.

Study	Number of Aircraft in Dataset	Costs Estimated	Dependent Variables	Independent Variables Selected for Equations	Synopsis
Levenson, Boren, Tihansky, and Timson: 1972	29	Aircraft Airframes	1. Engineering 2. Development Support 3. Flight Test Operations 4. Tooling 5. Manufacturing Labor 6. Manufacturing Material 7. Quality Control	1. Aircraft Quantity 2. Maximum Speed 3. AMPR Weight	An earlier set of CERs for development and production costs of airframes. All seven CERs included aircraft quantity, maximum speed, and AMPR* weight.
Large, Campbell, and Cates: 1976	25	Aircraft Airframes	1. Engineering 2. Flight Test Operations 3. Tooling 4. Nonrecurring Manufacturing Labor 5. Recurring Manufacturing Labor 6. Nonrecurring Manufacturing Material 7. Recurring Manufacturing Material 8. Quality Control 9. Total Airframe Program Cost	1. Maximum Speed 2. Airframe Unit Weight	Attempted to improve upon prior CERs from the 1972 study by investigating 17 new independent variables (IVs) and developing an additional CER for the total airframe costs. Ultimately, maximum speed and AUW* were still the only variables tested that could explain variations in cost.
Hess and Romanoff: 1987	34	Aircraft Airframes	1. Engineering 2. Development Support 3. Flight Test Operations 4. Tooling 5. Manufacturing Labor 6. Manufacturing Material 7. Quality Control 8. Total Airframe Program Cost	1. Maximum Speed 2. Empty Weight	A follow-up to the 1976 study with a larger dataset, which assessed 19 IVs from four categories: size, performance, construction, and program. Size (empty weight) and performance (maximum speed) were the only characteristics selected for the final set of CERs.
Owens, Allard, Ellison, Hofmann, Gahagan, and Valaika: 1991	8	Production Support Elements	1. Peculiar Support Equipment 2. Training 3. Data 4. Initial Spares	1. Maximum Speed 2. Airframe Unit Weight 3. Maintenance Man Hours Per Flying Hour 4. Time of Arrival 5. Aircraft Type 6. Avionics Type	This report created CERs for production support elements. No single IV was present in all four CERs, and weight is only present in the peculiar support equipment equation.
Younossi, Kennedy, and Graser: 2001	5	Aircraft Airframes	1. Recurring Engineering 2. Recurring Tooling 3. Recurring Manufacturing 4. Recurring Quality Assurance	1. Weighted Material Cost Factor 2. Lot Size 3. Cumulative Aircraft Quantity 4. Average Airframe Unit Weight per Lot 5. Recurring Labor Hours 6. EMD	The most recent study to create CERs for airframe costs, with an emphasis on the role of material properties. The final equations were in a complex exponential form that require a comprehensive knowledge of an aircraft's material mix and manufacturing techniques.

*Note: AMPR stands for Aeronautical Manufacturers' Planning Report, while AUW is Airframe Unit Weight.

METHODS

Data

We acquired the data analyzed in this article through the Cost Assessment Data Enterprise (CADE), as compiled by the Air Force Life Cycle Management Center (AFLCMC). Contractor, quantity, and cost data, such as the lot costs required to calculate T100 flyaway costs, were collected via the Cost Data Summary Reports (CDSRs), also known as 1921s, within CADE's Defense Automated Cost Information Management System (DACIMS). Aircraft weight data was obtained by accessing CADE's Data & Analytics application. Speed data was provided by the AFLCMC who compiled the data from past studies.

Once all the initial data was captured, the number of aircraft in the dataset was filtered based on availability of specific aircraft data. For an aircraft to have complete data and be included in the finalized dataset it had to contain at least one weight statement, aircraft cost data, and engine cost data. For aircraft, engines typically have their own production and 1921s separate from the aircraft itself, which was limited in CADE. The AFLCMC provided most of the engine cost data analyzed in this dataset, but this limitation excluded several aircraft, most of which are retired.

We developed two CERs. The first CER investigated all identified explanatory variables but excluded EMD (Engineering & Manufacturing Development) costs as a possible explanatory variable. The reason EMD costs were excluded is due to timing. By excluding EMD costs, the practitioner can use the CER prior to MS B. The EMD cost variable was reinstated for the second CER. However, inclusion of this variable for the second CER reduced the number of aircraft available for CER development. Consequently, we created a separate data inclusion criterion to investigate EMD costs as a cost driver. The total number of aircraft available for both the first and second CER is reflected in Table 2. Because they are a different commodity, helicopters are not considered in this article.

Regarding possible explanatory variables in the development of the CERs, these had to meet the following criteria:

1. Must be available pre-production (all variables have data available pre-EMD except for EMD costs).
2. Must be logically related to cost.
3. Must have accessible historical data.

Inspiration for these variables stemmed from the previous studies shown in Table 1, in addition to logically associated variables with reoccurring flyaway cost or variables that are speculated to perhaps affect these costs. Table 3 lists the potential independent variables considered along with their descriptions. Tables 4 and 5 further delineate some of these explanatory variables.

Table 2. Aircraft Inclusion and Exclusion Criteria.

Inclusion/Exclusion Criteria	Aircraft Removed	Remaining Aircraft
Aircraft in CADE with Weight Statements Available		516
Aircraft with Aircraft Cost Data Available	329	187
Aircraft with Engine Cost Data Available	105	82
Total Aircraft in Dataset for First CER		**82**
Aircraft with EMD Costs	23	59
Total Aircraft in Dataset for Second CER		**59**

Table 3. Potential Explanatory Variables.

Variable	Name	Description
ST	System Type	Ten dummy variables that represent the different system types of aircraft in this dataset. Table 4 provides a breakout of each one.
Qt	Quantified Units	Total number of aircraft in a lot production that was applied to calculate T100 flyaway cost.
AF	Air Force	Dummy variable where 1 = aircraft produced solely for the Air Force and 0 = it was not.
EC	Engine Count	The total number of engines on an aircraft.
Ct	Contractor	Six dummy variables that represent the current contractors who developed and produced the aircraft in this dataset. See Table 5.
EW	Empty Weight	The weight of the aircraft (in pounds) minus fuel, ordnance, and personnel.
AUW	Airframe Unit Weight	Empty weight (in pounds) minus propulsion, avionics, and government furnishings and equipment.
Speed	Max Speed	Maximum speed (in knots).
AD1	Aircraft Density 1	Airframe unit weight divided by empty weight: (AUW/EW)
AD2	Aircraft Density 2	Empty weight minus airframe unit weight then divided by empty weight: (EW-AUW)/EW
Stealth	Stealth	Dummy variable where 1 = aircraft has stealth technology and 0 = it does not.
Legacy	Legacy	Dummy variable where 1 = legacy aircraft and 0 = modern aircraft.
EMD*	EMD Costs	EMD costs for the mission design series (MDS) A-model
***Will not be tested in first regression analysis due to number of aircraft with this data, and when in a program's lifecycle this data is available.**		

Table 4. System Type by Aircraft.

System Type Variable	System Type	Number in Dataset	Aircraft in Dataset
ST1	Attack	11	A-10A, A-3A/B, A-4A, A-5A/RA-5C, A-6A, A-6E, A-7A/B, A-7D, EA-6B, S-3A, S-3B
ST2	Bomber	11	B-1B, B-2A, B-36A, B-47A, B-52A, B-52D, B-57A, B-58A, B-66B, RB-57D, RB-66B
ST3	Electronic Attack	1	ES-3A
ST4	Fighter	33	F-117A, F-22A, F-35A, F-35B, F-100A, F-101A, F-102A, F-104A, F -105A, F-106A, F-111A, F-14A, F-14D, F-15A, F-15C, F-15E, F-16A/B, F-16C/D, F-16C, F-4B, F-4C, F-4D, F4D-1, F-4E, F-4F, F-4J, F-5E, F-5F, F-80A, F-80C, RF-4B, RF-4C, RF-4E
ST5	Fighter/ Attack	4	EA-18G, F/A-18A, F/A-18C, F/A-18E/F
ST6	Patrol	2	P-3C, P-8A
ST7	Reconnaissance	2	E-3A, E-6A
ST8	Trainer	3	T-38A, T-39A, T-45TS
ST9	Transport/ Tanker	12	C-123B, C-130A, C-130J, C-131A, C-141A, C-17A, C-27J, C-5A, C-5B, HC-130J, KC-135A, MC-130J
ST10	UAV/Drone	3	MQ-1C, MQ-9A, RQ-4A

Table 5. Current Contractor Breakdown.

Contractor Variable	Contractor (Year Founded)	Number in Dataset	Aircraft in Dataset
Ct1	Boeing (1916)	8	B-47A, B-52A, B-52D, E-3A, E-6A, EA-18G, KC-135A, P-8A
Ct2	General Atomics Aeronautical Systems, Inc (1955)	2	MQ-1C, MQ-9A
Ct3	General Dynamics (1899)	4	F-111A, F-16A/B, F-16C/D, F-16C
Ct4	Leonardo Aviation (1948)	1	C-27J
Ct5	Lockheed Martin (1995)	6	C-130J, F-22A, F-35A, F-35B, HC-130J, MC-130J
Ct6	Northrop Grumman (1994)	1	RQ-4A

For the Stealth dummy (dichotomous) variable, five aircraft were considered to have stealth technology: B-2A, F-117A, F-22A, F-35A, and F-35B. These were coded as a '1', while the other 77 aircraft were coded a '0'. The reasoning was stealth aircraft might have higher reoccurring flyaway cost due to technological complexity.

For the Legacy dichotomous variable, a similar coding logic was employed. The Legacy variable is intended to capture the age and complexity of an aircraft and is defined by whether the weapon system is completely integrated or not. Legacy aircraft do not consist of an integrated weapon system, but rather separate components contained within an aircraft weapon system. If an aircraft at the Mission Design (MD) level was defined as a legacy aircraft, then all modifications of this aircraft were also defined as a legacy aircraft because their technology is based on legacy aircraft. For example, the C-5A was produced in the 1960s when weapon systems were not fully integrated and is therefore a legacy aircraft. The C-5B on the other hand was produced in the 1980s when weapon systems were being fully integrated, but this is still based on the same C-5A aircraft, and is therefore also a legacy aircraft.

There are 46 legacy aircraft in this dataset, with first flight dates that range from 1944 – 1968 at the MD level. Alternatively, modern aircraft are wholly integrated weapon systems whose production began in the 1970s. There are 36 modern aircraft in this dataset, with first flight dates that range from 1972 – 2007. Identification of whether an aircraft is legacy or modern was verified by a subject matter expert from the AFLCMC, and the breakdown between the two classifications is displayed in Table 6.

Table 6. Aircraft Breakdown by Legacy vs Modern.

Legacy vs Modern Aircraft	
Legacy Aircraft	A-3A/B, A-4A, A-5A/RA-5, A-6A, A-6E, EA-6B, A-7A/B, A-7D, B-36A, B-47A, B-52A, B-52D, B-57A, RB-57D, B-58A, B-66B, RB-66B, C-123B, C-130A, C-131A, KC-135A, C-141A, C-5A, C-5B, F-100A, F-101A, F-102A, F-104A, F-105A, F-106A, F-111A, F4D-1, F-4B, F-4C, F-4D, F-4E, F-4F, F-4J, RF-4B, RF-4C, RF-4E, F-80A, F-80C, P-3C, T-38A, T-39A
Modern Aircraft	A-10A, B-1B, B-2A, C-130J, HC-130J, MC-130J, C-17A, C-27J, E-3A, E-6A, ES-3A, EA-18G, F/A-18A, F/A-18C, F/A-18E/F, F-117A, F-14A, F-14D, F-15A, F-15C, F-15E, F-16A/B, F-16C, F-16C/D, F-22A, F-35A, F-35B, F-5E, F-5F, MQ-1C, MQ-9A, P-8A, RQ-4A, S-3A, S-3B, T-45TS

There are over 1,000 weight statements in the CADE library for approximately 516 different mission design series (MDS). This means for certain MDSs, such as the F-117A and P-3C, there is only one weight statement. While other MDSs, such as the A-10A and C-17A, have over a dozen weight statements. Out of the 82 aircraft in this dataset, 53 have only one weight statement in CADE and 29 have more than 1. For the EW and AUW variables listed in Table 3, if there was more than one weight statement available then the weight statement reflecting production units that occurred around the 100^{th} unit was selected. However, to investigate when in a program's life cycle weight is the most predictive of T100 flyaway costs, four additional variables are analyzed: EW1 and AUW1 which represents data from the first (or only) weight statement for an aircraft, and EW2 and AUW2 which represents the last.

Two other explanatory variables, Air Force (AF) and engine count (EC), are added for possible CER consideration for exploratory purposes. Of the 82 aircraft, 50 or approximately 61% were Air Force aircraft. Therefore, we were interested to see if there might be a difference between AF and non-AF aircraft with respect to reoccurring flyaway cost. With respect to engines, there are five different engine counts an aircraft can possess as observed in our dataset: 1, 2, 4, 6, or 8. The most common engine count is 2, representing exactly half of the dataset (41 out of 82). For the 1, 4, 6, and 8 engine aircraft, we observed counts of 21, 16, 2 and 2 occurrences, respectively.

Statistical Analysis

The descriptive and inferential analysis documented in this article was accomplished with JMP® Pro 15; and a 10% level of significance is used for most statistical tests. We adopt the method of ordinary least squares (OLS) to build the two CERs featured in this article and utilize a stepwise regression approach. Stepwise regression is an automatic process that screens potential independent variables to determine their best combination in predicting the dependent variable (McClave et al., 2017). If, while assessing the descriptive statistics, an independent variable appears to take on a different form (i.e., non-linear), then the alternative form is also examined in this stepwise process.

During this stepwise procedure, we utilize a mixed approach with the *p*-value threshold set to 0.1 for both inclusion and exclusion. To maintain the overall experimentwise error rate, we incorporated the Bonferroni Correction to set individual significance at 0.1/(number of significant explanatory variables). The response variable was recurring T100 flyaway costs. The possible explanatory variables consisted of all the dummy variables and continuous variables, as denoted earlier in this paper; as well any other noted patterns in the descriptive analysis, which preceded the inferential analysis.

To assess model validity, we assessed normality of residuals via the Anderson Darling test and constant variance via the Breusch-Pagan test. Both tests used a 0.05 level of significance. For model diagnostics, we assessed multicollinearity via the Variance Inflation Factor (VIF), outliers via studentized residuals, and overly influential datapoints via Cook's D. Although we recognize that OLS is robust against deviations from normality and constant variance (Kutner et al., 2004), we needed to determine if the finalized stepwise models were statistically sound and valid for practitioner usage.

In addition to testing assumptions and running diagnostics, the model must also be validated. The metrics employed in this article to explain the model's performance are the R^2, adjusted-R^2, and PRESS R^2 statistics. Because R^2 will always increase with the addition of a new independent variable, the adjusted R^2 corrects this drawback by considering the number of explanatory variables included in the model; and therefore will only increase if the new explanatory variable adds to the predictability of the model. The predicted residual error sum of squares (PRESS) R^2 statistic is recommended in evaluating a model's prediction ability (Naval Center for Cost Analysis, 2018). When PRESS R^2 is compared with the adjusted R^2, results can determine if the model is over-fitted and disproportionally reflecting model behavior.

Lastly, we performed a sensitivity analysis on the finalized stepwise CERs to investigate what would occur if we took an austere approach of removing any data point that might be an outlier, influential data point or cause residuals to deviate from normality and/or constant variance. The point of this sensitivity analysis was to not make the models 'appear' more significant than they are (as denoted by a very high R^2), but to ascertain if any other explanatory variable would be statistically significant, if we took such a myopic view. Our sensitivity analysis confirmed our finalized two CERs, which we now present.

Table 7. Summary Statistics of Recurring T100 Flyaway Costs.

Summary Statistics of Dependent Variable (in $K and CP$21)	
N	82
Median	$26,914.42
Mean	$51,297.87
Std Dev	$60,533.16
IQR	$44,118.01

RESULTS

Descriptive

Table 7 presents the summary statistics of the T100 flyaway costs for our sample. All dollar amounts are in Constant Price (CP)$21. Given the difference between the mean and median, we expected to see large flyaway costs associated with some aircraft. As shown in Figure 1, some aircraft with a 4-engine count generally have a much higher threshold of recurring T100 flyaway costs than any other engine count, including the four aircraft with six and eight engines. Delving deeper, Figure 2 depicts this is particularly true for heavier, 4-engine aircraft, as the seven heaviest aircraft had four engines. The highlighted datapoints in Figure 2 suggest a subgroup of heavy, 4-engine aircraft might have

Figure 1. Boxplots of Recurring T100 Flyaway Cost vs Engine Count.

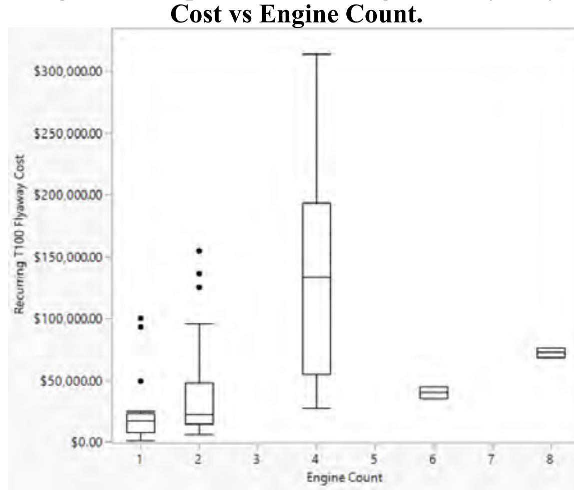
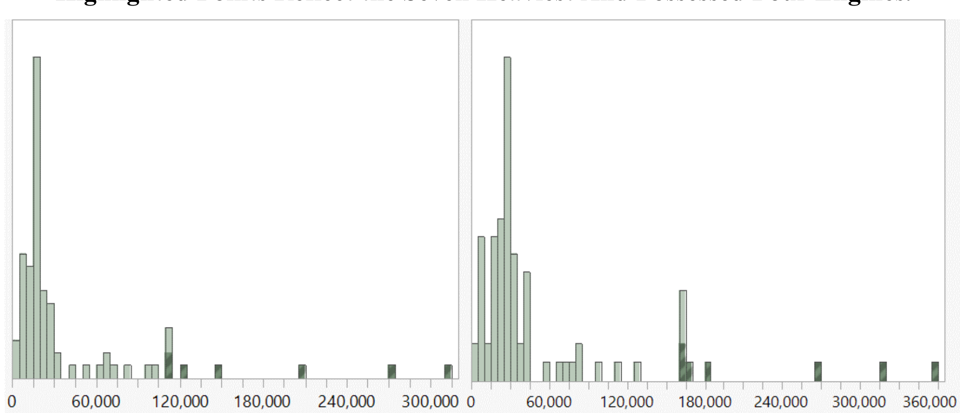


Figure 2. AUW (Left) and EW (Right) Distributions of Aircraft in Study Sample. Highlighted Points Reflect the Seven Heaviest And Possessed Four Engines.



high reoccurring flyaway cost compared to other aircraft in the sample study. Consequently, we created a dummy variable as another explanatory variable for stepwise regression to consider in building the two CERs. Table 8 lists the criteria used for an aircraft to be considered in this cohort, or subgroup. As we shall demonstrate shortly, this cohort became the most significant driver of reoccurring flyaway costs for aircraft. The aircraft in this cohort consisted of the E-3A, E -6A, B-2A, B-1B, C-17A, C-5A, and C-5B.

Table 8. Inclusion Criteria for Cohort.

Criteria (truncated)
1. AUW > 111,000 lbs
2. EW >162,000 lbs
3. Engine Count = 4

CER Model 1

Tables 9 and 10 present the statistically significant explanatory variables when not including or knowing EMD costs associated with an aircraft. All six variables are significant at the comparisonwise error rate with each *p*-value less than 0.0167 (0.1/6). Although all the various definitions of weight tested individually predictive, stepwise selected EW as the most significant, given the very high VIF scores (in excess of 5000) of the weight explanatory variables when included together. With PRESS R^2, adjusted R^2, and R^2 being relatively close to each other, this result gives an impression of a stable model and suggests CER1 is approximately 85-89% predictive of reoccurring flyaway costs. Equation 6 depicts this model for a practitioner to use, mindful of the ranges applicable to prevent model extrapolation. Those applicable ranges are given in Table 11. Note the coefficients are in $K.

Table 9. CER Model 1.

Variable	Estimate	*t Ratio*	*p-value*
Stealth	93115.58	9.03	<.0001
Cohort	90941.47	6.26	<.0001
Empty Weight	0.336918	5.49	<.0001
ln(Speed in knots)	23984.99	4.08	0.0001
Fighter Aircraft	-25872.6	-3.73	0.0004
Legacy	-18477.4	-3.68	0.0004

Table 10. Metrics for CER Model 1.

Metric	Value
R^2	0.8919
Adjusted R^2	0.8833
PRESS R^2	0.8529

Table 11. Boundaries for Applying CER Model 1.

Variable	Minimum	Maximum
Cohort – Airframe Unit Weight	111,899 lbs	310,484 lbs
Cohort – Empty Weight	162,228 lbs	356,797 lbs
Cohort – Engine Count	4	4
Empty Weight	2,183 lbs	356,797 lbs
Ln(Speed in knots)	Ln(150 knots) = 5.0106	Ln(1434 knots) = 7.2682

CER Model 2

The process by which we produced the second CER is identical to the first in both initial findings and the robustness check/diagnostics. The initial stepwise regression for the second model was analyzed with all the same explanatory variables from Model 1, plus EMD information. Tables 12 and 13 present our results. Both explanatory

$$\widehat{CER\ Model}\ 1 = -\$115{,}363.70 + \$90{,}941.47 * Cohort + \$93{,}115.58 * Stealth + \$23{,}984.99 * \ln(Speed) - \$25{,}872.60 * Fighter\ Aircraft - \$18{,}477.43 * Legacy + \$0.3369 * Empty\ Weight \quad (6)$$

$$\widehat{CER\ Model\ 2} = \$16{,}686.84 + \$142{,}033.49 * Cohort + \$0.004471 * EMD\ Costs \quad (7)$$

variables are significant at the comparisonwise error rate with each *p*-value less than 0.05 (0.1/2). PRESS R^2, adjusted R^2, and R^2 are relatively close to each other, which again gives the impression of a stable model and suggests the second CER is approximately 86-88% predictive of reoccurring flyaway costs. Equation 7 depicts this model for a practitioner to use, mindful of the ranges applicable to prevent model extrapolation. Those applicable ranges are given in Table 14. Note the coefficients are in $K.

CONCLUSION AND TAKEAWAY

We initially identified 13 explanatory variables (shown previously in Table 3) to be investigated in the development of two CERs for reoccurring flyaway costs. [Note that Table 3 has two umbrella variables: system type (ST) and contractor (Ct). The individual STs and Cts are not listed in Table 3. Rather, the full set of ST and Ct variables are provided in Tables 4 and 5 respectively]. Combining the full set of categorical variables, ST and Ct, with the initial explanatory variables of Table 3 resulted in 27 variables. Then, to account for the timeline of the different weight statements for empty weight (EW) and airframe unit weight (AUW), four additional weight variables were added: EW1, EW2, AUW1, and AUW2. Ultimately, after visually assessing the descriptive statistics for trends, two final variables were added, the natural log of Speed (ln(S)) and Cohort. Therefore, the total number of explanatory variables considered in developing the two CERs finalized at 33.

Table 12. CER Model 2.

Variable	Estimate ($K)	t Ratio	*p-value*
Cohort	142033.5	14.28	<.0001
EMD	0.004471	9.81	<.0001

Table 13. Metrics for CER Model 2.

Metric	Value
R^2	0.8814
Adjusted R^2	0.8771
PRESS R^2	0.8623

Table 14. Boundaries for Applying CER Model 2. Dollars are in $K and CP$21.

Variable	Minimum	Maximum
Cohort – Airframe Unit Weight	111,899 lbs	310,484 lbs
Cohort – Empty Weight	162,228 lbs	356,797 lbs
Cohort – Engine Count	4	4
EMD Costs	$36,793.92	$41,667,947.73

Recall that CER 1 does not include the total EMD cost variable. Out of the 32 remaining variables analyzed, six were selected for the final CER 1 model: Cohort, Stealth, ln(Speed in knots), Fighter Aircraft, Legacy, and Empty Weight. All the variables in this model have information available prior to Milestone B, making it applicable well before flyaway costs are incurred. With respect to interpretation of Equation (6), the intercept value of -$115,363.70 is simply a baseline and is uninterpretable for we never observed an instance where all the *x* variables took on the value zero.

The remaining coefficients describe how each explanatory variable effects recurring T100 flyaway costs. If an aircraft is a member of the cohort, it increases reoccurring flyaway costs by $90,941K on average. If an aircraft has stealth technology, it increases costs by $93,115K. For each unit increase in the natural log of an aircraft's speed (in knots), flyaway costs increase by $23,984K. If an aircraft is a fighter system type, it decreases costs by $25,872K on average. If an aircraft is identified as a legacy aircraft (which will not be the case for any future aircraft), then it decreases flyaway costs by $18,477K. Lastly, each pound increase in an aircraft's empty weight increases flyaway costs by $0.3369K (or $336.90). With respect to the explanatory variables' relative weighting and percentage effect on flyaway costs, Table 15 shows those details.

Table 15. Contribution Percentage by Explanatory

Variable	% Effect on CER 1
Cohort	24%
Empty Weight	22%
Stealth	21%
ln(Speed in knots)	12%
Fighter Aircraft	12%
Legacy	9%

All 33 explanatory variables (including total EMD cost) were analyzed for the development of CER Model 2. Of these, only two were selected for the final equation, Cohort and EMD costs. While Cohort can be determined near Milestone B in the acquisition lifecycle, EMD costs can only be incurred near Milestone C, which is still before the production phase when flyaway costs occur. However, this proximity does make the applicability of Model 2 more limited than Model 1. With respect to interpretation of Equation (7), again the intercept is simply a baseline. For the remaining two coefficients, if an aircraft is a member of the cohort, it increases the average reoccurring flyaway cost by $142,033K. Lastly, each dollar increase in EMD costs increases flyaway costs by $0.00471K (or $4.471) – four and a half fold. All dollars reflect CP$21 amounts. With respect to the explanatory variables' relative weighting and percentage effect on flyaway costs, Table 16 shows those details.

A significant discovery in this analysis was the identification of the variable Cohort, which was the only variable included in both CERs. Additionally, as seen in Tables 15 and 16, it has the greatest impact on the response for both models. This subgroup was initially identified in

Table 16. Contribution Percentage by Explanatory

Variable	% Effect on CER 2
Cohort	59%
EMD	41%

several scatter plots as a cluster of seven aircraft and included the E-3A, E-6A, B-2A, B-1B, C-17A, C-5A, and C-5B. While their complete criteria are shown in Tables 11 and 14, they are essentially amongst the heaviest aircraft in the dataset with four engines. Future aircraft that will likely be members of this cohort and whose flyaway cost estimate will benefit from this finding include the B-21.

Another major takeaway from this study is the identification of a proxy for complexity, and how strong a variable EMD is in predicting T100 flyaway costs. Yes, Stealth combined with Legacy were shown to be a significant proxy for complexity, but their effects are greatly diminished if total EMD costs are accessible. In fact, the moment EMD costs are introduced into stepwise regression analysis, five previously significant variables (Empty Weight, Stealth, ln (Speed), Fighter Aircraft, and Legacy) drop out, revealing the predictive power of EMD with respect to reoccurring flyaway costs. So, even if a practitioner chooses neither CER 1 nor CER 2 as a crosscheck for estimating flyaway costs, we advocate capturing complexity in their estimate and incorporating EMD costs, if available.

In summary, this paper fills a gap in the cost estimator toolkit. While previous efforts by RAND and others have developed useful CERs for airframes and other components, no CERs previously existed for recurring flyaway costs. With new aircraft, such as the B-21, T-7, E-7 and Next Generation Air Dominance on the horizon, accurate cost estimates will be of paramount importance. The CERs developed in this paper are a small step in helping achieve more awareness regarding flyaway costs. Thus, we humbly suggest practitioners employ them as a cross-check to their primary methodologies.

References

Bureau of Labor Statistics. (2022). *Producer Price Index 3364*. https://www.bls.gov/ppi/databases/

Department of Defense. (1992). *Cost Analysis Guidance and Procedures*. (DODI 5000.4-M). Assistant Secretary of Defense. https://www.acqnotes.com/Attachments/DoD%205000.4-M,%20DoD%20Cost%20Analysis%20Guidance%20and%20Procedures,%20Dec%201992.pdf

Department of Defense. (2022). *Cost Estimating Guide v2*. https://www.cape.osd.mil/files/Reports/DoD_CostEstimatingGuidev1.0_Dec2020.pdf

Department of the Air Force. (2021). *Fiscal Year 2019/2020 Acquisition Biennial Report*. U.S. Department of the Air Force. https://www.af.mil/Portals/1/documents/2021SAF/04_Apr/FY19_FY20_Dept_of_the_Air_Force_Acquisition_Biennial_Report_final.pdf

Government Accountability Office. (2020). *Cost Estimating and Assessment Guide*. (GAO-20-195G). U.S. Government Accountability Office. https://www.gao.gov/assets/gao-20-195g.pdf

Hess, R. W., & Romanoff, H. P. (1987). *Aircraft Airframe Cost Estimating Relationships*. The RAND Corporation.

Kutner, M., Nachtsheim, C., Neter, J., & Li, W. (2004). *Applied Linear Statistical Models* (5th Ed.). McGraw-Hill/Irwin

Large, J. P., Campbell, H. G., & Cates, D. (1976). *Parametric Equations for Estimating Aircraft Airframe Costs*. The RAND Corporation.

Levenson, G. S., Boren, H. E., Tihansky, D. P., & Timson, F. (1972). *Cost-Estimating Relationships for Aircraft Airframes*. The RAND Corporation.

Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying Cost and Schedule Uncertainty for Major Defense Acquisition Programs (MDAPs)*. (Report No. RR1723). RAND Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1723/RAND_RR1723.pdf

McClave, J. T., Benson, P. G., & Sincich, T. (2017). *Statistics for Business and Economics* (13th Ed.). Pearson.

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost Estimation Methods and Tools*. Hoboken, Nj: John Wiley & Sons, Inc.

Naval Center for Cost Analysis. (2018). *Joint Agency Cost Estimating Relationship (CER) Development Handbook*. https://www.ncca.navy.mil/references/CER_Dev_Handbook_Aug2018_Final.pdf

Owens, R. C., Allard, S. M., Ellison, M. C., Hoffman, J., Gahagan, L., & Valaika, J. R. (1991). *Estimating Relationships for Aircraft Production Support Elements*. Management Consulting & Research, Inc.

Younossi, O., Kennedy, M., & Graser, J. C. (2001). *Military Airframe Costs: The Effects of Advanced Materials and Manufacturing Processes*. The RAND Corporation.

Kyrie M. Rojo , *USAF, is a recent graduate of the Air Force Institute of Technology (AFIT) where she earned a Master's of Science in Cost Analysis. She is currently assigned as a cost analyst for the Aircraft and Weapons Division at the Air Force Cost Analysis Agency. (E-mail address: kyrie.rojo@us.af.mil)*

Dr. ***Brandon M. Lucas*** *is an Assistant Professor of Systems Integration & Cost Analysis in the Department of Systems Engineering and Management at AFIT. He holds a BA in History from the University of Texas at Austin, a MA in International Relations and ME in Teacher Education from the University of Oklahoma, a MS in Cost Analysis from AFIT, and a PhD in Economics from George Mason University. Dr. Lucas' research interests include profit analysis, cost & economic analyses, and incentive structures. (E-mail address: Brandon.Lucas@afit.edu)*

Dr. ***Edward D. White*** *is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and Ph.D. in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. (E-mail address: Edward.White@aft.edu)*

Dr. ***Jonathan D. Ritschel*** *is an Associate Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, his MS in Cost Analysis from AFIT, and his Ph.D. in Economics from George Mason University. Dr. Ritschel's research interests include public choice, cost analysis, and economic institutional analysis. (E-mail address: Jonathan.Ritschel@aft.edu)*

Mr. ***Shawn M. Valentine*** *is an Operations Research Analyst at AFLCMC/FZCE at Wright Patterson AFB, Ohio. Mr. Valentine graduated from Ohio University with a Bachelor of Science (BS) in Actuarial Science and a Master of Science (MS) in Financial Economics. (Email address: shawn.valentine@us.af.mil)*

Second Source Manufacturing: Lessons from the Second World War

Brent M. Johnstone

Abstract: Manufacturing defense systems at different sites is increasingly common due to foreign coproduction and international cooperative ventures. These situations challenge estimators, posing questions about the transfer of learning and relative efficiency of multiple production sites. This paper examines cost history from World War II, when U.S. bomber production lines were shared across multiple companies. The conclusions are tested against modern experience and guidance provided to estimators seeking help. .

Introduction

Estimators are sometimes confronted with situations where production on an already existing program begins at a second manufacturing site. Building the same product at multiple sites poses a challenge to conventional learning curve theory. How much, if any, learning can be transferred from the lead manufacturing site to the secondary site? How much learning curve improvement can be expected at the secondary site? Is it possible for the second source producer to become as productive as the lead site?

There are four common situations in the aircraft and missile industries where manufacturing of an item may occur at two or more sites simultaneously. These include:

- **Production at different facilities owned by the same firm.** A recent example is the commercial Boeing 787 manufactured simultaneously at its Everett, Washington and Charleston, South Carolina plants from 2011 until 2021 when final assembly was consolidated at Charleston. (Podsada, 2021)
- **Foreign coproduction.** A variety of U.S. military systems – fighters, helicopters, missiles, trainers, and anti-submarine warfare aircraft – have been coproduced simultaneously in the U.S. and foreign countries. The first military aircraft to do so was the F-86 in 1949. A list of U.S. military aircraft with foreign coproduced components or aircraft includes the F-86, T-33, T-34, S-2, P-2H, F-104, F-5, F-4, P-3C, F-16, AV-8B and F-35. (Rich, 1981)
- **International cooperative ventures.** Popular among European countries, these feature joint development projects with production and design authority split among the industries of different countries. A typical setup might have countries assigned to build specific aircraft components with some final assembly and flight test performed in each country. Examples include Jaguar, Tornado, Eurofighter Typhoon, and the Airbus family of commercial aircraft. (Svartman, 2018)
- **Competing companies producing the same end item.** This is more common in the missile industry where a second-source manufacturer competes with the developing company for a variable share of overall production. Examples include the AMRAMM, Hellfire, Maverick, Phoenix, Sidewinder, Sparrow, Standard, Stinger and Tomahawk missiles. (Lyon, 2006) The terminated Navy A-12 program would have required a price competition between General Dynamics and McDonnell-Douglas for a variable share of production after several production lots. (GAO, 1990)

Note that these situations are different from a true workshare, where two or more companies work together on the same end-product but each having build responsibilities which do not overlap.

Examples include Boeing and Northrop Grumman's split of the F/A-18E/F or Lockheed Martin and Boeing's split of the F-22 program. In these cases, there is not a question of learning transfer between firms. For instance, Boeing built F-22 aft fuselages and wings – neither Lockheed Martin nor any other company simultaneously built those components.

Learning Curve Theory – What Might We Expect?

In cases where two or more manufacturing sites build the same end-product, it is reasonable to expect learning can be transferred from the lead manufacturing site to the second source. In a typical contractual arrangement, both parties have a strong incentive for transfer to occur. If the lead company failed to provide manufacturing know-how to the second source, the second source will likely fail to make on-time deliveries, creating legal, contractual, manufacturing, and financial problems for the lead. Likewise, the second source is incentivized to accept technical assistance to bring its production up to speed and make it profitable as quickly as possible.

Technology transfer can cross multiple functions – engineering, planning, tooling, management – and come in many forms: data, training, on-site management, and assistance teams, furnishing start -up parts, et al. Success of the technology transfer program also depends on the capability of the second source. All else equal, the more capable the second source the more learning can be transferred to it.

On the other hand, it is unrealistic to expect that 100% of a firm's learning can be transferred. There are some things that only be learned by "hands-on" effort. If we think of Anderlohr's five elements of learning – shop personnel, supervision, continuity of production, tooling, and methods – it is apparent that no amount of formal or informal training can completely prepare a worker asked to work on a part he has never built before. (Anderlohr, 1969) Some things can only be learned by experience.

So how much learning should the estimator assume can be achieved by technology transfer, and how much left to experience? Let us construct a quick example to illustrate the complexities.

We begin with a company which is the original manufacturer of an item (the lead site). After several years, a second firm joins it in building the same product (the second source). Assume:

- The lead site builds 150 units with a first unit cost of 20,000 hours on an 80% learning curve slope before the second source builds its first unit.
- The second source retains 80% of the learning that the lead site accumulated up to that point (or equivalently, the second source will experience 20% learning loss).
- The second source also experiences an 80% learning curve slope beginning from the equivalent point on the lead site's learning curve after learning loss is applied.
- Both sites build an additional 350 units each.

Units Built by Lead Site Before Break-In	150
Hours per Unit (HPU) (Lead Site) at T-1	20,000
Unit Factor (UF) (Lead Site) at T-150	0.1993 [calculated as 150 ^ (ln (0.80) / ln (2))]
HPU (Lead Site) at T-150	3,986 [calculated as 20,000 x 0.1993]
Learned to Date	0.8007 [calculated as 1 - 0.1993)
Learning Lost	0.1601 [calculated as 0.8007 x 0.20 learning loss]
UF for Second Source's 1st Unit on Lead Site's Learning Curve	0.3594 [calculated as 0.1993 + 0.1601)
Second Source's HPU for its 1st Unit	7,188 [calculated as 20,000 x 0.3594]
Equivalent Unit for Second Source's 1st Unit on Lead's Learning Curve	24 [calculated as 2 ^ [(ln (0.3594) / ln (0.80)]
Unit Setback on Lead's Learning Curve	-84% [calculated as (24-150) / 150]

Example 1.

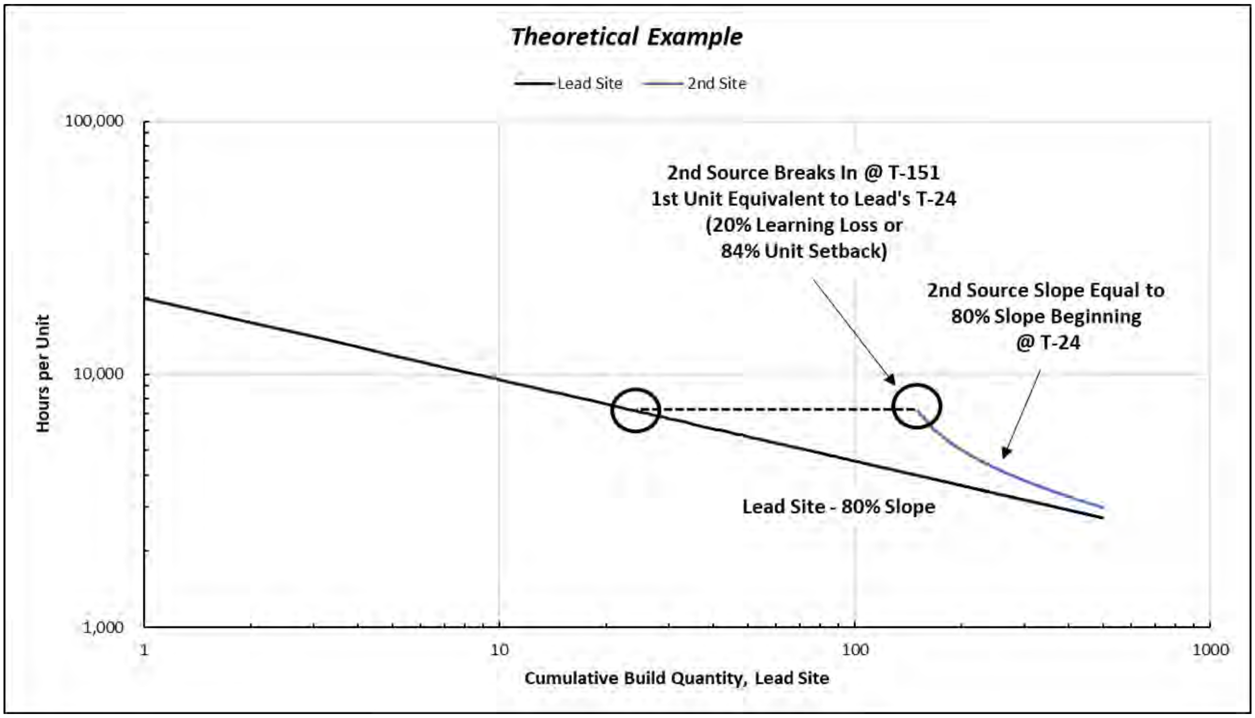


Figure 1. Theoretical Example – Hours per Pound Performance (I)

In this example, we have measured learning loss (or its inverse, learning gain) in two ways. The first is the percent of learning lost from the lead site's cumulative experience. The second is percent unit setback – that is, from the lead site's position on its learning curve, how far will the second source be set back on the learning curve when it builds its first unit? As we see, these two (learning loss/gain, unit setback) are not the same.

Figure 1 shows this case graphically. At the point of break-in, the lead site is building units at slightly under 4,000 hours per unit. The second source's first unit is 7,188 hours, which is equivalent to T-24 on the lead site's learning curve. This represents a 20% learning loss (or an 84% unit setback). As the second source continues to build, its HPU declines over time to its final unit at 2,973 hours. That is equivalent to T-373 on the lead site's learning curve. Meanwhile, the lead site's costs continue down the learning curve as well. Its final unit – the 500th – will cost 2,705 HPU.

Two things are apparent from the graph. First, the second source will asymptotically approach – but never intersect with – the lead site's cost performance. At no point will there be convergence, which will we define as the point the second source equals or exceeds the lead site's *historical* performance at some point on the curve. (It does not matter if the lead is currently producing the product at a lower cost, only that the second source matches where the lead formerly performed.) If there is no convergence to the lead site's learning curve, then it is also impossible for the second source to meet a second, stricter test: whether it can perform better than the lead's *current* performance.

Figure 2 shows the same information but uses a different method to plot the data. In it, the first unit of the second source's build is plotted as T-1. This method emphasizes the lower first unit cost for the second source because of learning gain. It also shows the second source's asymptotic cost performance as it approaches, but does not reach, the lead site's hours per unit. (Due to the peculiarities of the logarithmic scale, it may appear that the second source achieves the same cost as lead site. It does not – the plotted data is the same as that portrayed in Figure 1, where the gap is more visually apparent.) By treating the second source's first build as T-1, this would give an equivalent 86% learning curve for the second source.

Asymptotic non-convergence results from our assumption the second source will achieve the same learning slope as the lead site. If we assumed a flatter slope by the second source, the gap widens further. Only if the second source achieves a steeper rate of learning is it possible for the two slopes to achieve convergence.

In Figure 3 we have given the second source a steeper slope (76%) than the lead site beginning at the same break-in HPU. This allows the second source to achieve convergence with the lead – its cost performance intersects the learning curve of the lead site. Moreover, by the end of production it is actively producing units at a lower cost than the lead site can. When each site finishes its last unit – T-500 for the lead, T-350 for the second source – the second source's last unit costs is 2,426 hours versus 2,705 hours for the lead site. Under different learning curve assumptions, a second source could converge to the lead site's learning curve performance but at the same time does not produce the aircraft at a lower cost than the lead.

Nevertheless, theory cannot tell us whether the estimator should assume the second source's learning curve slope is shallower, steeper or the same as the lead site. A theoretical case could be made for any of these outcomes:

1. *The second source's learning curve slope will be the same as the lead site*. After the initial transfer of learning, the second manufacturing site will experience the same sources of future learning as the lead company – worker proficiency, supervisor familiarity with his crews, improvements in production layouts and improved part availability as the supply chain gears up. The second source experiences a "rerun" of the lessons the lead site learned

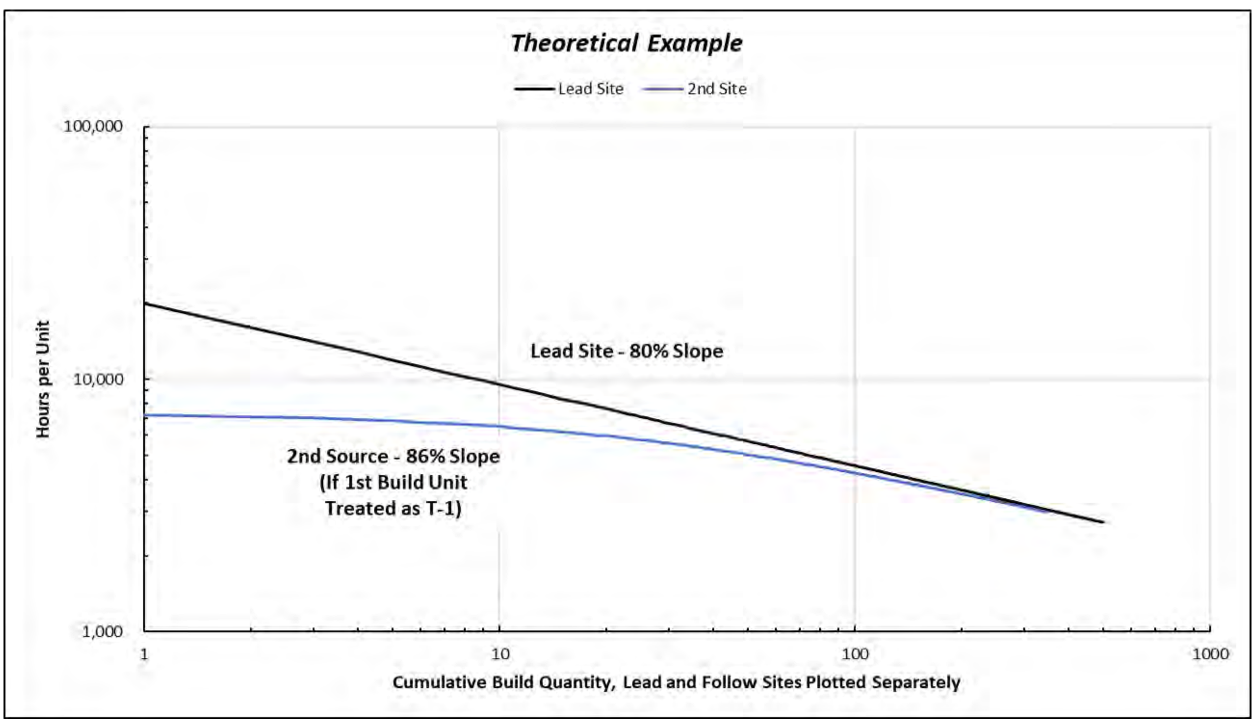


Figure 2. Theoretical Example – Hours per Pound Performance (II)

(but could not transfer to the second source). In such a case, the second source will improve its performance at the same rate of learning as the lead site at an equivalent point on the curve. This is the standard assumption in production gap literature, which also deals with the subject of lost and retained learning. (Anderlohr, 1969; DCAA, 1996)

2. *The second source's learning curve slope will be flatter than the lead site*. This argument looks at build rates and the phenomena of learning and forgetting. The longer the period between build units, the harder it is for the mechanic to retain what he has learned since the last time he completed a task. If the second source's production rates are lower than the lead, its shop floor mechanics will go longer between builds, potentially losing learning and creating a flatter learning curve slope relative to the lead site.

3. *The second source's learning curve slope will be steeper than the lead site*. Learning curve analyst E. B. Cochran wrote of the "time compression penalty," which encompasses many of the issues surrounding aircraft development and early production: late engineering releases, tooling errors, part shortages, manpower disruption, and high levels of scrap and rework, all of which conspire to force the learning curve to be flatter in its early phases. (Cochran, 1968) If technology transfer is successful, however, much of this early disruption endured by the lead site can be avoided by the second source. It too will have its growing pains, but they need not be as severe. That suggests the second source might be able to start its phase of rapid cost improvement sooner, rather than later, resulting in an overall steeper slope.

But which of these scenarios is the most likely to unfold?

The answer to that question lies in historical experience – after all, this is not a new situation in the aircraft business. However, such historical experiences are typically locked away in company vaults as proprietary information and not available for wider distribution. What can we do?

Fortunately, there is a public domain, nonproprietary database we can use to develop answers, and which has been used in several influential learning curve studies over the years. (Stanford Research Institute, 1949; Asher, 1956; Alchian, 1963) The database is the *Source Book of World War II Basic Data*. (*Source Book*, undated) This data, collected from Aeronautical Monthly Progress Reports (AMPR) provided by contractors during the war, provides manufacturing hours per month by model and facility as well as hours per pound against cumulative plane number. Moreover, this database contains several examples of the same aircraft model being produced at different facilities.

The obvious objection is that this data is 80 years old, and aircraft manufacturing processes have changed substantially over eight decades. That is entirely true; but the data can still provide important insights into the transfer of learning between manufacturing sites. We will use this data to test four propositions. After drawing conclusions from the wartime data, we will compare it (at a high level, to protect proprietary information) with modern-day data to determine if these conclusions still appear valid in today's environment.

The four propositions to be tested are as follows:

1. The second source will show some degree of learning transfer – that is, it will not begin back at the lead's T-1 cost – but it will not completely transfer all the lead's learning, either.

2. A concerted effort by the lead site to foster technology transfer should improve the learning gain achieved by the second source, resulting in a lower-cost break-in.

3. The second source will not fully converge to the lead company's learning curve – that is, the two lines will not intersect.
4. The second source will not be able to produce at a lower cost than the lead company – that is, the coproducer's best hours per pound performance will always be greater than the lead company's best hours per pound performance.

Approach of the Second World War

As war in Europe approached, the United States began preparing itself for possible conflict. The American aircraft industry was poorly prepared for a substantial expansion of deliveries. The industry had numerous manufacturers, each making aircraft in small quantities in an artesian "job-shop" environment. Most manufacturers did not build aircraft on an assembly line, but in one spot on the factory floor in their entirety. (Stoff, 1993) In 1938 the United States produced 900 military aircraft. The entire industry employed only 36,000 people – slightly less than the knit-hosiery industry. (Harr, 1965)

An executive for Consolidated-Vultee Aircraft described the aircraft manufacturing process in the prewar years:

> Under the pre-war production system, if an order for say 60 planes (a big order in those days) was received, groups of workers would concentrate on the various parts needed for the components and produce 60 units. As fast as these components were made they were stored in a central stockroom, there to remain until all the parts for certain subassemblies had been completed. Then they would be withdrawn and the 60 subassemblies fabricated. And as the 60 subassemblies were finished they, in turn, would be assembled into the completed unit until the 60 had been constructed, tested, and delivered. (Laddon, 1943)

This system worked fine for small orders, minimizing setup time and parts fabrication costs. (Laddon, 1943) However, production quantities started increasing as Europe grew closer to war. In June 1938 Lockheed received an order for 200 Hudson bombers for Great Britain, at the time the largest aircraft order received by a U.S. firm between the world wars. (Harr, 1965) But the watershed moment did not come until May 1940, when President Franklin Roosevelt declared before Congress:

> Our immediate problem is to superimpose on this [military aircraft] production capacity a greatly increased additional production capacity. I should like to see this nation geared up to the ability to turn out at least 50,000 planes a year. (*The New York Times*, 1940)

In response to Roosevelt's demand, the War Department began developing plans for a rapid expansion of aircraft production. Bomber production was a high priority of the United States Army Air Force. However, there was insufficient capacity to provide the needed quantities of any given aircraft model. In addition, there was a high concentration of aircraft manufacturers on the West Coast, which was considered vulnerable to enemy attack. The need was two-fold: (1) to increase production capacity by bringing on more suppliers and (2) build more aircraft production facilities in the interior of the United States – "behind the mountain chains" -- where they would be safe from enemy attack. (Holley, 1964)

The answer was to pool bomber production across multiple companies, each producing the same aircraft and sharing production knowledge. Douglas and Lockheed-Vega would build B-17s under license from the designer and lead producer Boeing. Similarly, Douglas, North

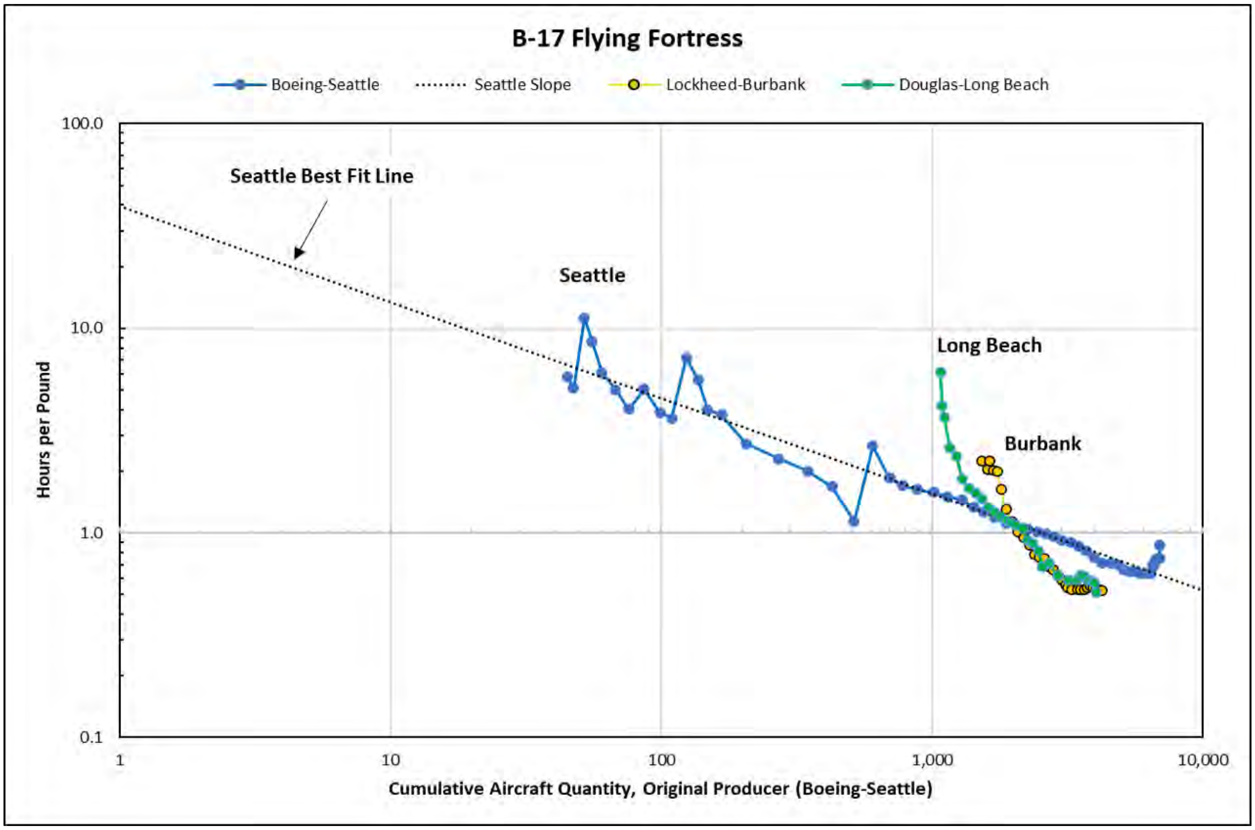


Figure 4. B-17 Flying Fortress Hours per Pound (I)

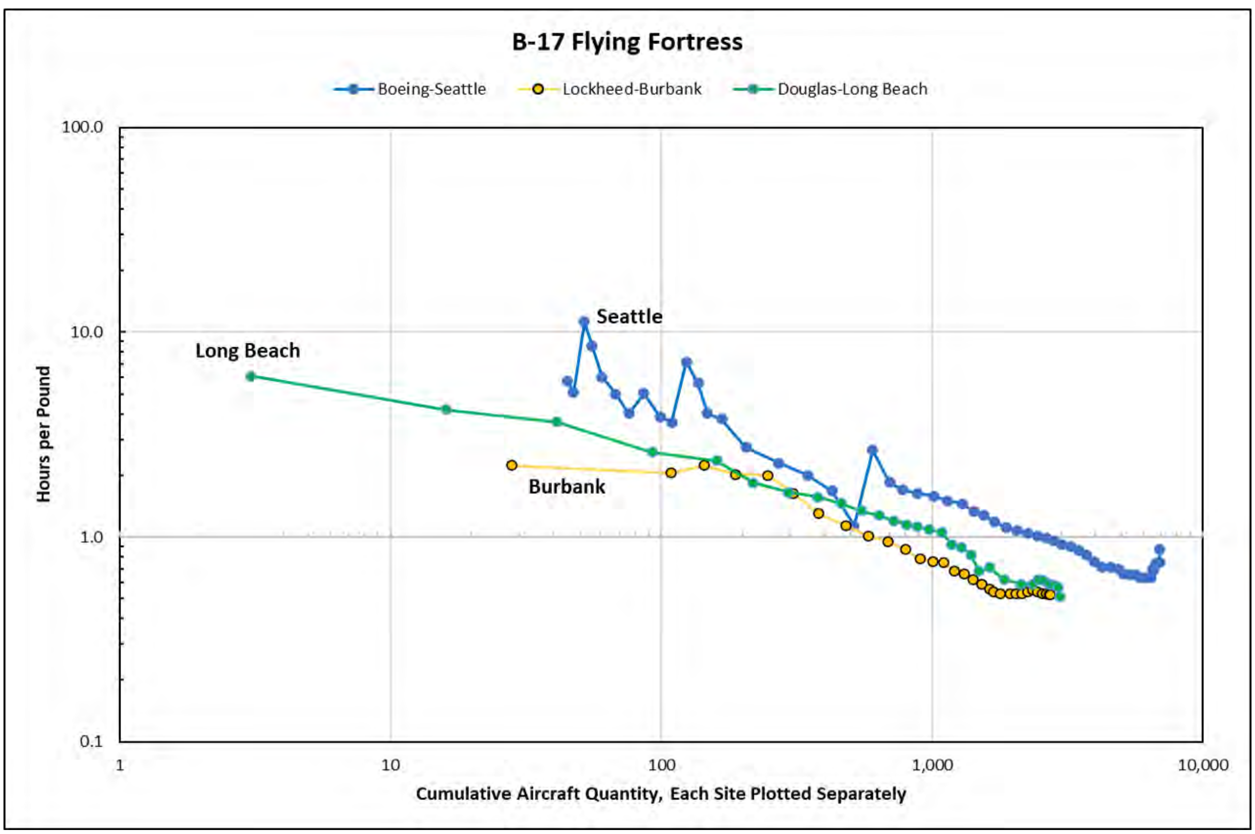


Figure 5. B-17 Flying Fortress Hours per Pound (II)

American and the automaker Ford would join Consolidated-Vultee to build B-24s. Finally, for the B-29 Superfortress, the Air Force's largest bomber, Bell and Martin would enter a licensing agreement with Boeing. As part of the capacity expansion, new aircraft facilities would be opened in Dallas (North American), Fort Worth (Consolidated-Vultee), Long Beach (Douglas), Marietta (Bell), Omaha (Martin), Tulsa (Douglas), Wichita (Boeing) and Willow Run, Michigan (Ford).

The sudden explosion in order sizes forced dramatic changes on the shop floor. Consolidated-Vultee soon discovered that its central warehouse could not stock millions of finished parts. Consequently, it eliminated the warehouse and installed smaller stock bins along the assembly line, working to a just-in-time inventory system. Building an entire aircraft in place was replaced by a moving line that transported the aircraft as it was built through successive stations manned by dedicated crews. Planes were stationed in final assembly at 45-degree angles, allowing 50% more aircraft to be worked in the same floor space. Complicated assemblies previously worked by highly skilled craftsmen were broken into simpler and more accessible subassemblies that could be more easily worked by inexperienced mechanics. Better, more precise tooling was introduced to simplify drilling and machining operations. (Laddon, 1943). These lessons learned by Consolidated were repeated across the aircraft industry.

This program was enormously successful. In the end, these eight companies delivered almost 35,000 bombers before the end of the war. Overall, the entire American aircraft industry not only met the President's goal of 50,000 aircraft per year, but almost doubled it, producing over 96,000 aircraft in 1944 alone. (Holley, 1964)

We will examine the cost performance of each of these bomber models in turn.

B-17 Flying Fortress

The B-17 had been in production at Boeing's Seattle plant as early as 1938, but at very low production rates. Only 53 aircraft were delivered in 1940. In addition to rapidly expanding Boeing production at Plant 2, Douglas Aircraft and Vega Aircraft (a wholly owned subsidiary of Lockheed) were brought on-line in 1942 and 1943 respectively. By 1944, the three facilities were delivering almost 5,400 bombers a year. In total, more than 12,600 B-17s were delivered. (Holley, 1964)

Figure 4 shows the cost performance of the three facilities. (*Source Book*, undated) The first units of the Long Beach and Burbank facilities are plotted beginning at the cumulative number of aircraft produced to date at the lead site in Seattle.

Figure 5. shows the same information except that the cumulative production of the Long Beach and Burbank facilities is plotted independent of the number of units produced at Seattle.

Table 1 summarizes the performance of the three sites in terms of learning curve slopes, percent learning loss and percent unit setback. It also answers if the coproducing sites were able to achieve convergence with the lead site's learning curve, and if they were able to produce at an eventual lower cost than the lead.

Figure 4 shows that the Long Beach and Burbank were not only able to converge to Seattle's learning curve but eventually produce the B-17 at a lower cost than Seattle, despite producing half as many aircraft as the Seattle plant. One reason the B-17 coproducers were so successful was the robust level of cross-company cooperation between the three contractors. In May 1941 a committee of company and government representatives was established, the so-called BDV (Boeing-Douglas-Vega) committee. The committee coordinated material purchases, master production schedules, release

B-17 Flying Fortress		*Lead*	*Coproducer*	*Coproducer*
		Boeing Seattle	**Douglas Long Beach**	**Lockheed Burbank**
Initial Build Plotted as T-1	Actual 1st Lot (Hrs/Lb)	5.79	6.12	2.24
	Theoretical First Unit (TFU) (Hrs/Lb)	39.57	13.79	16.31
	Unit Curve Coefficient	(0.4689)	(0.3886)	(0.4406)
	Unit Curve Slope	72.3%	76.4%	73.7%
	R-Square (R^2)	94.2%	95.4%	92.3%
	Minimum Hrs/Lb	0.63	0.51	0.52
Initial Build Plotted at Setback Unit #	Setback Unit on Lead's Learning Curve	N/A	54	457
	% Learning Loss	N/A	12.1%	2.5%
	% Unit Setback	N/A	95.0%	69.5%
	Unit Curve Slope	N/A	67.9%	55.1%
Additional Data	1st Delivery	1938	Oct-42	Jan-43
	Prior Units Produced by Lead	N/A	1,073	1,495
	Total Aircraft Built	6,981	3,000	2,750
	Achieve Convergence to Lead's Learning Curve?	N/A	Yes	Yes
	Achieve Lower Cost Than Lead?	N/A	Yes	Yes

Table 1. B-17 Flying Fortress Cost Performance by Manufacturing Site

of engineering drawings, inspection criteria and production lessons learned between the three companies. Ideas for improvement did not just flow from the lead to the second sources. If the second source or one of their lower-tier suppliers simplified a design, reduced the use of expensive materials, or improved performance, the committee recommended the revised design as the standard for all companies. The BDV committee became the template for other aircraft programs with multiple contractors, including the B-29. (Holley, 1964) It is not surprising, then, that learning loss was minimized (12% for Long Beach, 2% for Lockheed-Vega) – by far, the least amount of learning loss among all the bomber producers.

B-24 Liberator

While the B-17 is probably the most iconic World War II bomber, the Army Air Force purchased more B-24 Liberators than any other bomber model – over 18,000 aircraft. (Holley, 1964) Given such large procurement quantities, production was eventually split over five sites: San Diego, Fort Worth, Willow Run, Tulsa and Dallas. Consolidated-Vultee was the lead, beginning B-24 production in 1940 at its San Diego facility.

Figure 6 shows the cost performance of the five facilities. (*Source Book*, undated) The first units of the Fort Worth, Willow Run, Tulsa, and Dallas facilities are plotted beginning at the cumulative number of aircraft produced to date at the lead site in San Diego.

Figure 7 shows the same information except that the cumulative production of the Fort Worth, Willow Run, Tulsa, and Dallas facilities is plotted independent of the number of units produced at San Diego.

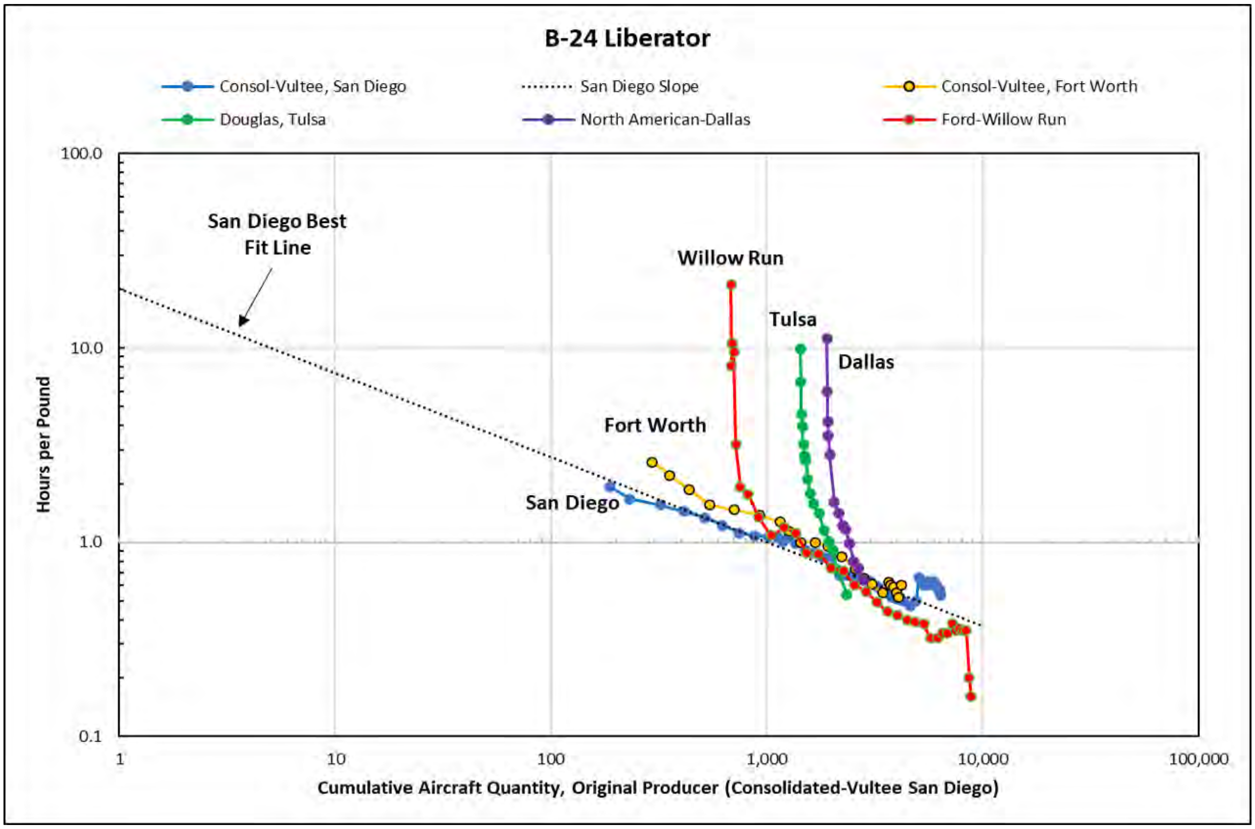


Figure 6. B-24 Liberator Hours per Pound (I)

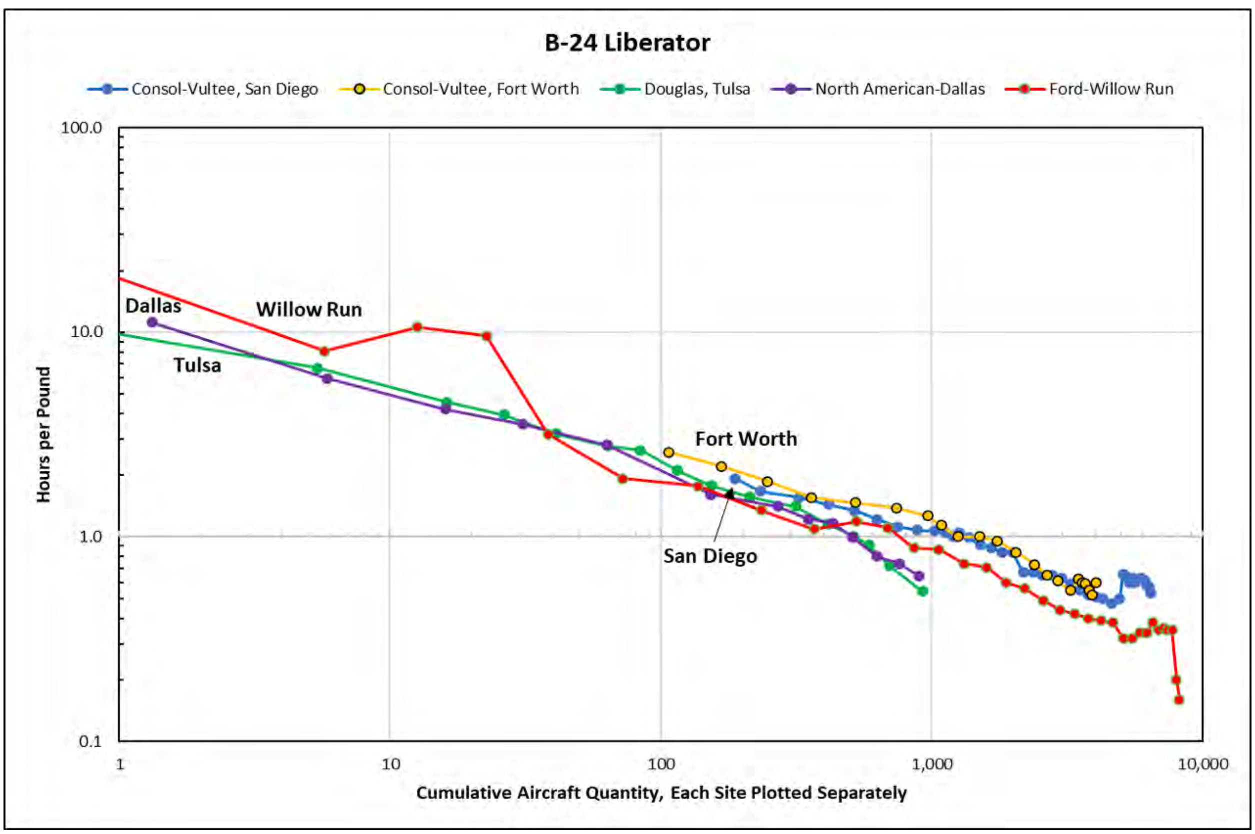


Figure 7. B-24 Liberator Hours per Pound (II)

Table 2 summarizes the performance of the five sites in terms of learning curve slopes, percent learning loss and percent unit setback. It also answers if the coproducing sites were able to achieve convergence with the lead site's learning curve, and if they were able to produce at an eventual lower cost than the lead.

The B-24 shows a wide variance in the degree of learning loss experienced by the coproducing companies. Consolidated-Vultee's Fort Worth facility experienced 3% learning loss while Ford experienced greater than 100% learning loss. All four coproducing sites were able to reach convergence with San Diego's cost performance. Only one – Ford's plant in Willow Run – was able to produce the B-24 at an eventual lower cost than Consolidated's San Diego plant.

Several factors explain the wide variance in learning loss. Unlike the B-17, there was no coordinating committee for B-24 production. At the low end, the minimal loss of learning from San Diego to Fort Worth is best explained that both facilities operated under the same company, and the new Fort Worth plant was operated by a cadre of management and engineers transferred from San Diego. At the high end, Willow Run decided to adopt a completely different manufacturing approach from the other sites. In 1940 Henry Ford's leading manufacturing expert, Charles Sorensen, was sent to the San Diego B-24 line, only to be dismayed by Consolidated's assembly approach:

> Inside the [Consolidated] plant I watched men putting together wing sections and portions of the fuselage…. [W]hat I saw reminded me of nearly thirty-five years previously when we were making Model N Fords…before we achieved the orderly sequence of the assembly line and mass production.
>
> The nearer a B-24 came to its final assembly the fewer principles of mass production there were as we at Ford had developed and applied over the years. Here was a custom-made plane, put together as a tailor would cut and fit a suit of clothes.
>
> The B-24's final assembly was made out of doors under the bright California sun and on a structural

B-24 Liberator		*Lead*	*Coproducer*	*Coproducer*	*Coproducer*	*Coproducer*
		Consol-V San Diego	**Consol-V Fort Worth**	**Ford Willow Run**	**Douglas Tulsa**	**N. American Dallas**
Initial Build Plotted as T-1	Actual 1st Lot (Hrs/Lb)	1.93	2.59	21.25	9.91	11.21
	Theoretical First Unit (TFU) (Hrs/Lb)	20.10	22.18	22.95	13.53	13.74
	Unit Curve Coefficient	(0.3641)	(0.4393)	(0.4882)	(0.4146)	(0.4249)
	Unit Curve Slope	77.7%	73.7%	71.3%	75.0%	74.5%
	R-Square (R^2)	91.8%	96.7%	96.4%	95.5%	98.5%
	Minimum Hrs/Lb	0.47	0.52	0.16	0.54	0.64
Initial Build Plotted at Setback Unit #	Setback Unit on Lead's Learning Curve	N/A	113	1	5	4
	% Learning Loss	N/A	2.8%	106.1%	47.0%	54.0%
	% Unit Setback	N/A	39.9%	99.9%	99.6%	99.8%
	Unit Curve Slope	N/A	72.6%	71.0%	70.2%	71.3%
Additional Data	1st Delivery	Early 1940	Apr-42	Sep-42	Apr-43	Jul-43
	Prior Units Produced by Lead	N/A	188	680	1,433	1,897
	Total Aircraft Built	6,435	4,105	8,233	1,052	1,000
	Achieve Convergence to Lead's Learning Curve?	N/A	Yes	Yes	Yes	Yes
	Achieve Lower Cost Than Lead?	N/A	No	Yes	No	No

Table 2. B-24 Liberator Cost Performance by Manufacturing Site

> steel fixture. The heat and temperature changes so distorted this fixture that it was impossible to turn out two planes alike without further adjustment….[I]t was obvious that if the wing sections had uniform measurements, the way we made parts for automobiles, they would not fit properly under out-of-doors assembly conditions.
>
> All this was pretty discouraging, and I said so. Naturally, and quite properly, the reply was "How would you do it?" I had to put up or shut up. "I'll have something for you tomorrow morning," I said.

Sorensen retreated to his hotel room and overnight produced a plan for a new manufacturing facility based on automotive build principles. Sorensen's rough sketches became the blueprint for Ford's massive Willow Run facility, designed to roll out a B-24 every hour at maximum capacity. (Sorensen, 1956)

Realizing Sorensen's dream was more difficult than he or the other Ford executives imagined. Ford was forced to re-do 30,000 drawings it received from Consolidated because it could not resolve discrepancies between loft boards and detailed part designs, discrepancies which Consolidated simply left to their skilled production workers to reconcile on the shop floor. Likewise, Ford built 21,000 jigs and fixtures, but eventually only used 11,000 of them – the rest scrapped due to errors in the source drawings or rendered obsolete by the stream of engineering design changes flowing from the Air Force and Consolidated. (Holley, 1964) Willow Run struggled to accelerate initial production – a commonly asked question by journalists of the day was: "Will It Run?" (Baime, 2015) By March 1944, though, Willow Run had wrung out its production inefficiencies and was producing over 400 bombers per month – short of Ford's stated goal of a B-24 every hour, but still more than the Air Force could absorb in the field. (Holley, 1964) In the end, Ford's automotive-based process was able to produce the B-24 at a lower cost per pound than Consolidated-San Diego or the other sites.

The Tulsa and Dallas plants represent learning loss in between the extremes of Fort Worth and Willow Run, losing 47% and 54% of learning respectively in their first build. The reason for North American's higher loss of learning was, ironically, poor liaison between the Dallas plant and Ford. Ford was slow to notify North American of engineering design changes, thus creating downstream tooling and production problems; and the drawings Ford provided were inadequate. Eventually North American redrew all the engineering drawings Willow Run provided. (Holley, 1964)

These widely varying experiences on the B-24 confirm that the ability of the lead contractor to successfully transfer its technology and lessons learned is the predominant factor on the degree of learning loss.

B-29 Superfortress

The Air Force's heaviest bomber, the long-range B-29 Superfortress, began production in 1943 at Boeing's new Wichita facility. In short order, production lines at Marietta (Bell), Renton (Boeing) and Omaha (Martin) were opened. Almost 3,900 Superfortresses were eventually delivered, over half at the two Boeing facilities.

Figure 8 shows the cost performance of the four facilities. (*Source Book*, undated) The first units of the Marietta, Renton, and Omaha facilities are plotted beginning at the cumulative number of aircraft produced to date at the lead site in Wichita.

Figure 9 shows the same information except that the cumulative production of Marietta, Renton and Omaha is plotted independent of the number of units produced at Wichita

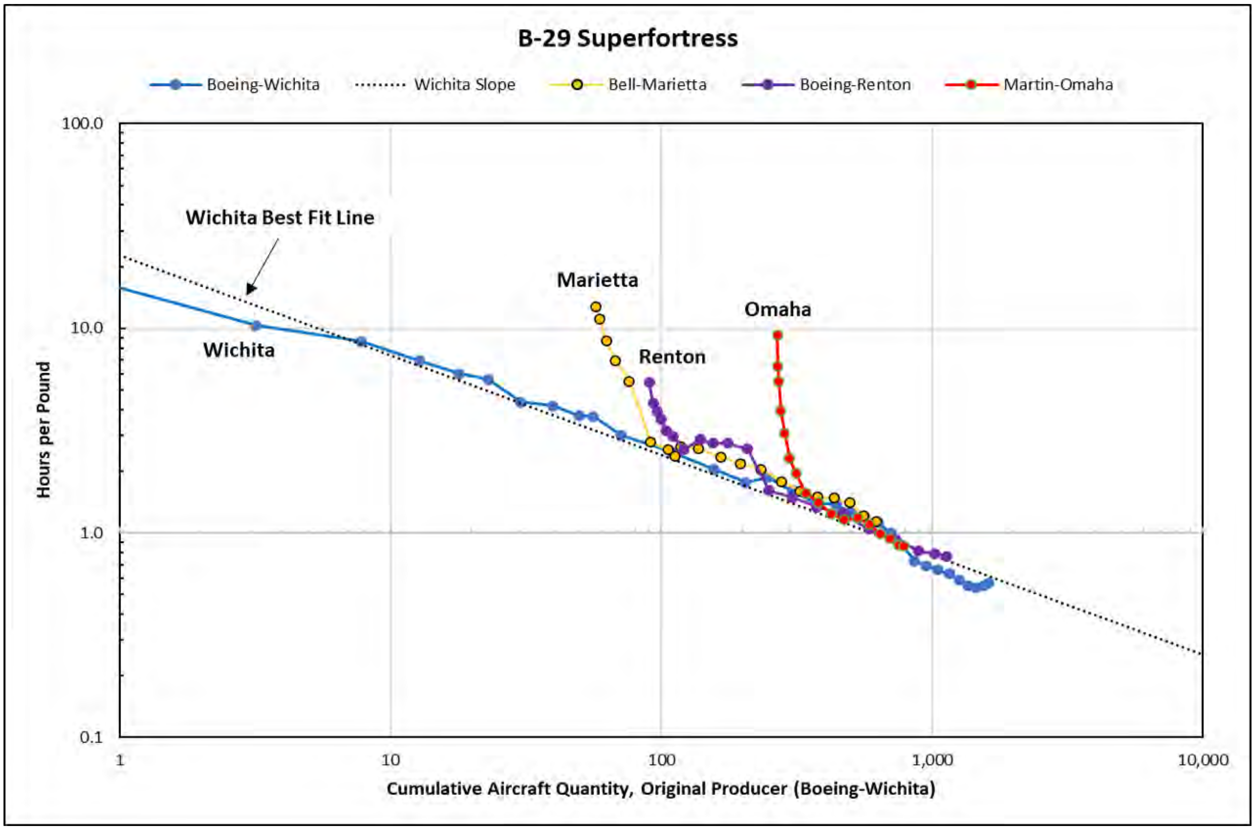


Figure 8. B-29 Superfortress Hours per Pound (I)

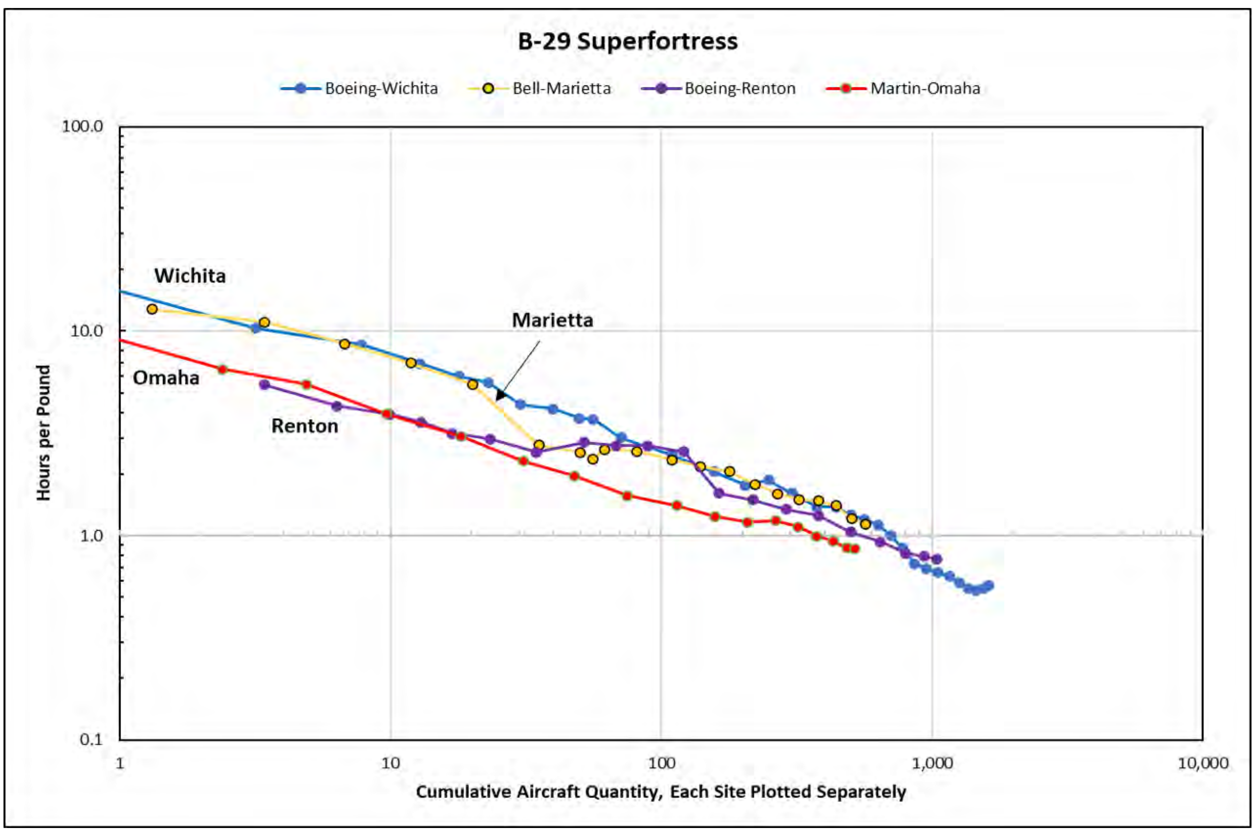


Figure 9. B-29 Superfortress Hours per Pound (II

Table 3 summarizes the performance of the four sites in terms of learning curve slopes, percent learning loss and percent unit setback. It also asks if the coproducing sites were able to achieve convergence with the lead site's learning curve, and if they were able to produce at an eventual lower cost than the lead.

Like the B-17, the B-29 program had a coordinating committee among the build companies. However, several factors kept the B-29 committee from performing as successfully as its B-17 predecessor. First, the B-29 program was originally intended to pair Boeing with North American and the Fisher Body Division of General Motors. However, these companies eventually dropped as prime contractors and were replaced by Martin and a second Boeing plant in Renton. In addition, the B-29's design was highly experimental, resulting in a high degree of engineering changes. Finally, five other companies –Chrysler, Hudson, Goodyear, McDonnell, and Republic – provided major components and assemblies to the prime contractors. These factors significantly complicated production coordination and the sharing of knowledge. Historian Irving Holley writes, "The B-29 program was the most complex joint production undertaking of the war." (Holley, 1964)

The B-29's prime coproducers experienced between 14% to 49% learning loss. Boeing's Renton plant showed the lowest degree of learning loss. Like Consolidated's San Diego and Fort Worth B-24 plants, the Renton plant was initially staffed with a management and engineering cadre from Seattle and Wichita. (Mishina, 1999) For the other two coproducers, Omaha achieved 36% learning loss while Marietta experienced 49% loss. Only two coproducers (Renton, Omaha) reached convergence with Boeing-Wichita's learning curve, and none of the coproduction sites achieved a lower hours per pound than the lead Wichita site.

B-29 Superfortress		*Lead*	*Coproducer*	*Coproducer*	*Coproducer*
		Boeing Wichita	**Bell Marietta**	**Boeing Renton**	**Martin Omaha**
Initial Build Plotted as T-1	Actual 1st Lot (Hrs/Lb)	16.15	12.77	5.45	9.25
	Theoretical First Unit (TFU) (Hrs/Lb)	22.81	16.30	9.09	9.05
	Unit Curve Coefficient	(0.4883)	(0.4168)	(0.3382)	(0.3793)
	Unit Curve Slope	71.3%	74.9%	79.1%	76.9%
	R-Square (R^2)	97.8%	96.6%	94.0%	99.4%
	Minimum Hrs/Lb	0.54	1.14	0.77	0.86
Initial Build Plotted at Setback Unit #	Setback Unit on Lead's Learning Curve	N/A	3	19	6
	% Learning Loss	N/A	48.8%	14.2%	36.4%
	% Unit Setback	N/A	94.1%	78.4%	97.6%
	Unit Curve Slope	N/A	72.2%	73.4%	71.7%
Additional Data	1st Delivery	Feb-43	Dec-43	Feb-44	May-44
	Prior Units Produced by Lead	N/A	56	87	267
	Total Aircraft Built	1,642	636	1,096	531
	Achieve Convergence to Lead's Learning Curve?	N/A	No	Yes	Yes
	Achieve Lower Cost Than Lead?	N/A	No	No	No

Table 3. B-29 Superfortress Cost Performance by Manufacturing Site

Like the B-17, the B-29 program had a coordinating committee among the build companies. However, several factors kept the B-29 committee from performing as successfully as its B-17 predecessor. First, the B-29 program was originally intended to pair Boeing with North American and the Fisher Body Division of General Motors. However, these companies eventually dropped as prime contractors and were replaced by Martin and a second Boeing plant in Renton. In addition, the B-29's design was highly experimental, resulting in a high degree of engineering changes. Finally, five other companies –Chrysler, Hudson, Goodyear, McDonnell, and Republic – provided major components and assemblies to the prime contractors. These factors significantly complicated production coordination and the sharing of knowledge. Historian Irving Holley writes, "The B-29 program was the most complex joint production undertaking of the war." (Holley, 1964)

The B-29's prime coproducers experienced between 14% to 49% learning loss. Boeing's Renton plant showed the lowest degree of learning loss. Like Consolidated's San Diego and Fort Worth B-24 plants, the Renton plant was initially staffed with a management and engineering cadre from Seattle and Wichita. (Mishina, 1999) For the other two coproducers, Omaha achieved 36% learning loss while Marietta experienced 49% loss. Only two coproducers (Renton, Omaha) reached convergence with Boeing-Wichita's learning curve, and none of the coproduction sites achieved a lower hours per pound than the lead Wichita site.

Summary of World War II Experience

Table 4 summarizes the experience of the B-17, B -24, and B-29 coproducers.

We can summarize our conclusions from the World War II data as follows:

- Some degree of learning transfer from the lead to the coproducer occurred in eight of nine cases. Learning loss shows a wide variation from as little to 2% to as much as 106%. The reasons for these extremes have

Aircraft	Coproducer Company/Site	% Learn Loss	Setback Unit	% Setback	Converge to Lead's Cost at Equiv Position?	Best Cost Lower Than Lead's Best?
B-17	Douglas-L. Beach	12%	53.6	95.0%	Yes	Yes
	Lockheed-Burbank	2%	456.7	69.5%	Yes	Yes
B-24	Consolidated-Ft. Worth	3%	113.1	39.9%	Yes	No
	Ford-Willow Run	106%	0.9	99.9%	Yes	Yes
	Douglas-Tulsa	47%	5.1	99.6%	Yes	No
	N. American-Dallas	54%	3.8	99.8%	Yes	No
B-29	Bell-Marietta	49%	3.3	94.1%	No	No
	Boeing-Renton	14%	18.8	78.4%	Yes	No
	Martin-Omaha	36%	6.3	97.6%	Yes	No
Statistics	Mean	36%	73.5	86.0%	N/A	N/A
	Median	36%	6.3	95.0%	N/A	N/A
	Minimum	2%	0.9	39.9%	N/A	N/A
	Maximum	106%	456.7	99.9%	N/A	N/A

Table 4. Summary of Bomber Coproducer Experience

already been discussed, but in general the more successful the lead company's technology transfer, the lower the learning loss. On average, 36% learning loss (or alternatively, a 64% learning gain) was achieved during World War II coproduction.

- Percent setback varies from a minimum of 40% to a maximum of 100% with an average of 86% (mean) and 95% (median)
- In eight of nine cases, the coproducer converged to the lead company's learning curve.
- In three of nine cases, the coproducer eventually produced at a lower cost than the lead company.

Comparison to Current Experience

Fast forward 80 years. Military aircraft today are manufactured using advanced materials (titanium and composites) unknown to World War II designers. Fighters and bombers perform at supersonic speed with jet engines, not in the subsonic environment with turboprops. Aircraft are stuffed with electronic computers which can fly and maneuver the aircraft, operate its weapons systems, and allow a fighter pilot to engage his target far beyond visual range. Parts and assemblies are manufactured to previously unachievable tolerances to appear nearly invisible on enemy radar screens. So, are these conclusions – drawn from a war our grandparents and great -grandparents fought – still valid?

Ideally, we could test this hypothesis by looking at postwar data with smaller production runs. However, there is limited data for military aircraft to be built in two locations inside the United States. There are only three such cases, all of them from the 1950s, the North American F-86 and F-100 and the Boeing B-52. However, the published data provides little insight into the questions we are considering. (Rich, 1981, Cook, 2002)

However, if we look not at the total aircraft level, but at individual components and consider either foreign coproduction or cases where work was incrementally transferred from one site to another, the available dataset begins to expand.

Due to the proprietary nature of this data, it can only be discussed at a high level without any program identification. All these cases, however, represent components with a lead manufacturing site and a coproducing second source brought in later during the program life cycle. All have occurred within the past 30 years. In addition, all had robust technology transfer programs to reduce program risk and enable the second source to come up to speed as quickly as possible by sharing production and tooling lessons learned.

	% Learn Loss	% Setback	Converge to Lead's Cost at Equiv Position?	Best Cost Lower Than Lead's Best?
Component A	28%	64%	Yes	Yes
Component B	23%	71%	Yes	Yes
Component C	40%	82%	Yes	Yes
Component D	31%	94%	Yes	No
Component E	44%	88%	No	No
Component F	56%	95%	No	No
Mean	37%	82%	N/A	N/A
Median	35%	85%	N/A	N/A
Minimum	23%	64%	N/A	N/A
Maximum	56%	95%	N/A	N/A

Table 5. Modern-Day Manufacturing Coproduction.

Table 5 shows the mean learning loss in our modern sample is almost identical to the World War II experience – 37% versus 36%. The range of learning loss in the modern sample is substantially narrower. This percentage is not surprising since in all these cases the lead site pushed hard to make a successful learning transfer. There is no modern-day equivalent of Ford's Willow Run experience.

As a secondary data point, in its 2002 analysis of F-35 final assembly alternatives RAND assumed that learning transfer in a work split was analogous to a production gap. The analogy assumes that after a production gap learning gains attributable to shop personnel would be lost but gains attributable to methods improvements could be retained. The retained learning is the same kind of knowledge which could be transferred from a lead to a second manufacturing site. Based on prior research, RAND calculated learning retention of 30-88%, with an average of 64% retained learning after a production gap. (Said alternately, RAND observed 36% lost learning). (Cook, 2002). Coincidentally, that 36% learning loss assumption exactly matches the observed World War II learning loss in Table 4.

Like the World War II experience, in four of the six modern-day cases in Table 5, the second source was able to converge to the lead site's learning curve. Less often, the second source was able to produce at lower hours per unit than the lead site. A discussion of why that occurred might potentially disclose sensitive information: therefore, we only note that it happened and leave the "How?" and "Why?" to a different forum.

Conclusions

How might this data assist an estimator dealing with a second-source manufacturing situation? Let us revisit our four propositions:

Proposition 1: The second source will show some degree of learning transfer – that is, it will not begin back at the lead's T-1 cost – but it will not completely transfer all the lead's learning, either.

True. In all but one of the World War II cases, there was learning gain from the lead site. In the only case where there was not – Ford's B-24 Willow Run plant – Ford explicitly rejected Consolidated's manufacturing and tooling philosophy in lieu of its automotive-based approach. This rejection was an unusual situation unlikely to be repeated in a modern second-source case study. Exactly how much learning transfer should be assumed by the estimator depends however on the strength of the technology transfer program, leading us to our second proposition.

Proposition 2: A concerted effort by the lead site to foster technology transfer should improve the learning gain achieved by the second source, resulting in a lower-cost break-in.

True. The World War II data shows a wide variation in learning loss experience. For the B-17, B-24 and B-29, successful technology transfer depended on the lead's ability to communicate engineering and production knowledge to the second sources. Learning loss was minimized in cases where there was successful cross-company coordination (the B-17's BDV committee) or the second source happened to be a sister plant which absorbed a cadre of engineers and management from the lead site (B-24 Fort Worth, B-17 Renton). Learning loss was greater when there were difficulties in the engineering handoff (B-24 Willow Run, B-24 North American), when the lead site was poorly prepared for the transfer (B-24 San Diego) or the second source rejected the lead company's manufacturing approach and instead struck out on their own (Willow Run, again).

In a world of Computer Aided Three-Dimensional Interactive Application (CATIA) and other 3-D modeling tools, the engineering handoff should be much easier compared to the primitive design tools of 80 years ago. But even in a modern era, the handoff can pose difficulties. In the shipbuilding industry, where production at multiple shipyards is more common, the use of incompatible design and analysis tools for CAD/ CAM at different sites has posed significant problems. (Cook, 2002)

Other factors can influence the transmission of manufacturing and tooling lessons learned. Amicable business relationships between the two companies were cited as another significant factor in Navy shipyard learning transfers. (Cook, 2002) Contractual arrangements can weigh heavily – for instance, if the two companies are direct competitors fighting over a share of production, there may be a strong *disincentive* to cooperate.

It is tempting for the estimator to use 36% as a default assumption. In the end, the estimator must make a careful analysis of the technology transfer program and the experience and capabilities of the companies involved to determine how successful he believes the learning transfer will be – a decision difficult to quantify, and largely judgmental.

Proposition 3: The second source will not fully converge to the lead company's learning curve – that is, the two lines will not intersect.

Frequently untrue. The World War II data suggests our theoretical construct of second source learning is partially incorrect. Theory suggests a coproducer can only asymptotically approach the lead's cost performance. The World War II data shows under the right circumstances, the second source can intersect the lead company's learning curve. This occurred primarily because the second source's learning curve slope was slightly steeper than the lead site's. However, it is important to note that all the second source bomber manufacturers had large production runs (ranging from 500 to 8,000 aircraft) which gave them an opportunity for convergence. If those production runs had been smaller -- say, only 50 or 100 units – such performance would probably have been impossible.

Choosing a learning curve slope for projection is always treacherous and adding a second source does not make it any less so. Without a better appreciation for why the slopes were steeper – difficult to ascertain after eight decades – it is difficult to provide guidance. The estimator's tolerance for risk also comes into play. Assuming the second source will perform at the same learning curve slope as the lead company is a conservative choice, but it may serve where a more risk-adverse estimate is desired.

Proposition 4: The second source will not be able to produce at a lower cost than the lead company – that is, the coproducer's best hours per pound performance will always be greater than the lead company's best hours per pound performance.

Usually, but not always true. In most cases, the second source will not perform at a lower hours per pound than the lead site. Yet successful instances appear in the World War II data if the degree of learning loss is low (B-17 assembly lines at Long Beach and Burbank) or if the second source's manufacturing and tooling approach proves superior (Willow Run). Either instance would require an extended production run by the second source to play out, however. Another possible scenario where a second source might provide lower costs could arise from an aircraft with multiple models. If the second source is permitted to concentrate on a single model while the lead site must build more than one variant – and experience the attending loss of learning and disruption – it is conceivable the second source could demonstrate better cost performance. That scenario did not appear in the World War II data, however, so it remains untested.

This proposition provides a lower bound for learning curve slopes. If the second source's learning curve slope creates projections where the coproducer has a lower cost than the lead site, the estimator should recognize this for a lesser probability scenario and possibly alter his choice.

Caution:

A final word. It is important to note what we have *not* attempted to prove here: that the use of a second-site manufacturing facility is cost-effective in terms of the total program cost. That would require an analysis of relative labor rates, overhead impacts, production capacity, additional overseas sales generated by the inclusion of foreign industry, domestic industrial base considerations, etc. and is far beyond the scope of this paper. The analysis of second-site manufacturing is only one piece of a much larger puzzle.

References:

Alchian, Armen (1963, October). Reliability of Progress Curves in Airframe Production. *Econometrica*, 31(4), 679-693.

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

Asher, Harold (1956). *Cost-Quantity Relationships in the Airframe Industry*. RAND.

Baime, A. J. (2015) *The Arsenal of Democracy: FDR, Detroit, and an Epic Quest to Arm an America at War*. Mariner Books.

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

Cook, Cynthia R.; Arena, Mark V.; Graser, John C., et al. (2002) *Final Assembly and Checkout Alternatives for the Joint Strike Fighter.* MR-1559-OSD. RAND.

DCAA Contract Audit Manual (1996, July). DCAAM 7640.1. Government Printing Office.

General Accounting Office (GAO) (1990, December). *Navy A-12: Cost and Requirements*. GAO/NSIAD-91-98. https://www.gao.gov/assets/nsiad-91-98.pdf

Harr, Karl G., Jr. (1965, September) Industry and World War II – Embryo to Vigorous Maturity. *Air Force/ Space Digest*, 54-64.

Holley, Irving Brinton, Jr. (1964). *Buying Aircraft: Materiel Procurement for the Army Air Forces, Center of Military History*. United States Army. https://history.army.mil/html/books/011/11-2/CMH_Pub_11-2.pdf

Laddon, I. M. (1943, May). Reduction of Man-Hours in Aircraft Production. *Aviation*, 170-173, 356-360.

Lyon, Thomas P. (2006, June). Does Dual Sourcing Lower Procurement Costs? *The Journal of Industrial Economics*, 54(2), 223-252.

Mishina, Kazuhiro (1999). Learning by New Experiences: Revisiting the Flying Fortress Learning Curve. In *Learning by Doing in Markets, Firms, and Countries*, 145-184. University of Chicago Press. http://www.nber.org/chapters/c10232

Mlinar, Anthony J. (1978, September). *Application of Learning Curves of Aircraft Produced at More than One Location to the F-16 Lightweight Fighter*. Master's Thesis, Air Force Institute of Technology.

The New York Times (1940, May 17). Text of the President's Address Asking a Great Defense Fund, 10.

Podsada, Janice (2021, March 5). Last Everett-built Boeing 787 Rolls Off the Assembly Line. *Everett HeraldNet*. https://www.heraldnet.com/business/last-everett-built-boeing-787-rolls-off-the-assembly-line/

Rich, Michael; Stanley, William; Birkler, John; Hesse, Michael (1981, October). *Multinational Coproduction of Military Aerospace Systems*. R-2861-AF. RAND. https://www.rand.org/content/dam/rand/pubs/reports/2006/R2861.pdf

Sorensen, Charles E. (1956). *My Forty Years with Ford*. W. W. Norton.

Source Book of World War II Basic Data, Airframe Industry Vol. I Direct Manhours, Progress Curves (Undated). Headquarters Air Materiel Command, Wright Field. https://apps.dtic.mil/sti/citations/ADA800199

Stanford Research Institute (1949). *Relationships for Determining the Optimum Expansibility of the Elements of a Peacetime Aircraft Procurement Program*.

Stoff, Joshua (1993). *Picture History of World War II American Aircraft Production*. Dover Publications.

Svartman, Eduardo Munhoz; Teixeira, Anderson Matos (2018). "Coproduce or Codevelop Military Aircraft? Analysis of Models Applicable to USAN," *Brazilian Political Science Review*, 12(1). https://www.redalyc.org/journal/3943/394357143004/html/

***Brent Johnstone** is a Lockheed Martin Fellow and production air vehicle cost estimator at Lockheed Martin Aeronautics Company in Fort Worth, Texas. He has 34 years' experience in the military aircraft industry, including 31 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.*

Schedule and Cost Estimations Through the Decades: Are They Improving?

Captain Sammantha Jones

Edward D. White, Ph.D.

Jonathan D. Ritschel, Ph.D.

Shawn M. Valentine

Acquisition professionals strive to provide the best estimation of schedule and cost to deliver war-winning capability. Numerous reforms and improvement initiatives have been implemented towards improving these cost and schedule outcomes in Major Defense Acquisition Programs (MDAPs). This leads to the following question: Are schedule and cost outcomes improving over time? We use both descriptive and inferential techniques to investigate schedule and cost trends in MDAPs from the 1970s to 2010s. We find schedule growth does not exhibit statistically significant improvement across the decades; all decades indicated a consistent schedule slippage for a typical MDAP. In contrast, the analysis of Cost Growth Factors (CGFs) detected statistical differences in some instances. The most novel finding, however, is found in the *standard deviations* of CGFs. We identified a statistically significant decreasing trend in the standard deviations of total program CGFs throughout the decades. This lowering variability trend also appeared for Program Acquisition Unit Cost (PAUC) CGFs from the 1980s onward. The decrease in variability of cost estimates suggest to us that cost estimators and/or the process behind them might be improving over time.

This article identifies macro-level trends of cost and schedule growth for Major Defense Acquisition Programs (MDAPs) from the 1970s to the 2010s. Specifically, we investigate overall program cost growth, program acquisition unit cost (PAUC) growth, and schedule growth for the Department of the Defense's (DoD) largest program acquisitions. The inspiration for this study came from Arena et al. (2006) and Younossi et al. (2007). Both papers provide insights into cost growth of MDAPs mainly prior to 2000. This article can be considered an extension of these often-cited works with a few key differences.

We do not delineate between development and procurement costs; we consider these together as total program cost as reported in the Selected Acquisition Reports (SARs). Although dividing cost growth into development and procurement is a common practice when analyzing MDAPs, we wanted to look at the overall cost and schedule growth holistically. There are other deviations between the analyses, such as how the data is presented and the type of inferential analyses utilized, but the overall goal of this paper is to investigate cost and schedule growth from the 1970s to the 2010s and to determine statistically whether the DoD has seen a change in cost or schedule growth over this timespan.

Background

MDAPs are essential for the development and production of military aircraft, satellites, missiles, and other large investment items that U.S. military operations require. By statute, MDAPs are categorized as Acquisition Category I (ACAT, 2021) programs if they have either total expenditure of research, development, test and evaluation (RDT&E) costs greater than $525 million (fiscal year 2020 constant dollars), total expenditure of procurement costs greater than $3.065 billion (fiscal year 2020 constant dollars), or specifically designated by milestone decision authority as special interest (MDAP, 2020). MDAPs are the DoD's largest investments and constitute a large proportion of the DoD portfolio relative to their program numbers. These investments often entail large economic risks.

Currently, the Government Accountability Office (GAO) reports annually on DoD weapon systems based on their total cost and acquisition status. Of the 107 programs evaluated in their 2021 report, 84 were MDAPs. These 84 MDAPs have a total planned investment of 1.79 trillion Fiscal Year (FY) 2021 dollars. The GAO has reported consistent cost growth in the DoD's MDAP portfolio for the last 15 years. They attribute the most dramatic cost changes to quantity changes (Government Accountability Office, 2021). Other studies have also noted historical precedent for underestimating program costs (Arena et al., 2006; Younossi et al., 2007) and schedules (Monaco & White, 2005; Riposo, McKernan, & Kaihoi, 2014). Light et al. (2017) even recommended that the acquisition community approach early cost estimates with skepticism.

Cost growth in MDAPs appears common; however, dramatic growth within programs can lead to a Nunn-McCurdy Breach. From 1997 to 2016, 58 out of 189, or 36% of MDAPs experienced cost growth large enough to cause such a breach. Out of these 58 breaches, 18 were significant and 40 were critical (USD(AT&L), 2016, p. 65). Significant breaches occur when current cost estimates meet or exceed 15% of the current baseline estimate or 30% of the original baseline estimate of an acquisition program. Critical breaches occur at the 25% and 50% levels respectively (Nunn-McCurdy Breach, 2021).

MDAPs that experience Nunn-McCurdy breaches are extreme examples of cost growth. But due to their programmatic costs, even a small cost growth percentage can add millions of dollars worth of additional funding needs for the programs. Schedule growth in MDAPs can also lead to readiness issues and apprehension for military and congressional leadership. Because of these funding and readiness issues, there have been efforts over the last several decades to reduce cost and schedule growth within MDAPs (Fox et al, 2011). These efforts include sweeping reforms, changes in business practices, updates to record keeping requirements, and adjustments in the overall structure of how MDAPs are executed, and their records maintained (Fox et al., 2011, Dwyer et al., 2020).

Over the last few decades there have been extensive analyses on DoD MDAPs. Various organizations such as the Congressional Research Service, the DoD itself, GAO, or even contracted organizations such as RAND or the Institute for Defense Analyses (IDA) have conducted these studies. In 2016 the DoD published an annual acquisition system performance report. In this report they analyzed MDAPs through a variety of different lenses to include cost and schedule growth, cost performance overall, cost performance broken out by development and production, cost growth by military departments, cost growth by contractors, and a few other viewpoints (USD(AT&L), 2016).

The 2016 report claims there has been a continuing improvement in the field of defense acquisitions, however their analyses concentrate on several various micro-level insights into the cost and schedule growth of DoD MDAPs. While these micro-level assessments are extremely important to understanding what is happening in specific MDAPs, their study does not provide a macro-level analysis truly examining whether the overall cost and schedule growth of MDAPs have changed over time (USD AT&L, 2016). Thus determining changes, if any, to cost and schedule growth is the intent of this article.

Data and Methods

Data

We utilized the Cost Assessment Data Enterprise (CADE) system to obtain the data for this article's analyses. Available since February of 2019, the CADE Selected Acquisition Report (SAR) database is a consolidation of DAMIR (Defense Acquisition Management Information Retrieval) SAR data and non-DAMIR legacy SARs. Using the SAR Unit Cost Report along with the Current and Baseline Estimate report and the CADE SAR Data listing, we identified 409 potential programs to analyze as of October 2021. From there, we excluded programs. Note that the dataset only includes MDAPs. Major Automated Information Systems (MAIS) are not part of the analysis. Table 1 lists the reasons for program exclusion and rationale. For programs categorized as transitioned or restructured, if these actions led to the creation of a new MDAP, then that new program remained in the database. For example, the WIN-T, after being broken into three separate programs, drove the creation of one MDAP that met the requirements to be included into our final dataset: the WIN-T increment 2.

SAR Sample Inclusion & Exclusion Table	
Total Number of SARS Available in CADE	409
Programs Classified as Terminated	26
Transitioned or Restructured Programs	11
SAR not Classified as MDAPs*	17
SAR w/no data available in CADE**	25
SARs with Missing Milestone B Data***	129
Programs < 5 years since MS B	7
Final MDAP SAR Sample	**194**

Table 1. Selected Acquisition Report (SAR) inclusion and exclusion table.

**This includes Pre-MDAP, Other, Special Interest, MAIS Major System, and DoE Program Classifications*

*** These programs were listed in CADE but had no cost or schedule data available for analysis*

**** These programs did not have any MS B data available as a starting point for the cost and schedule growth analysis*

We use Milestone (MS) B as the starting point for collecting program data, as this is typically considered the official start of a program (AcqNotes, 2021). Additionally, many previously published studies have used MS B as the starting point of their analyses on MDAP cost or schedule variations. These include studies by Younossi et al. (2007), McNicol (2018), and Dwyer et al. (2020).

The final exclusion criteria for our analysis involved accounting for the low maturity level of modern MDAPs. Programs that were less than five years old (and had yet to meet Initial Operating Capability (IOC)) were omitted from the analysis. This is because of the increased likelihood of these less than mature programs not having yet realized their schedule and cost changes compared to programs further along in development/ production. Within our schedule database (described later) the mean time for a MDAP to move from MS B to IOC is 8.6 years with 97 of the 120 taking more than five years to reach IOC. This maturity requirement led to the exclusion of seven MDAPs that reached MS B in 2017 or later. Younossi et al. (2007) adopted a similar exclusion criterion. Table A.1 in the Appendix lists all the programs by decade we analyzed for this article.

After the completion of this initial 194 MDAP database, we parsed the data into three separate databases to explore schedule growth, total program cost growth (this is just RDT&E plus procurement costs), and PAUC cost growth individually. To calculate the change in MDAP schedule growth, we used two main milestones: MS B and IOC. From the starting 194 programs, 74 did not have IOC estimates available in CADE and were subsequently not used in the schedule analysis. This left us with 120 programs for comparing schedule growth across the decades.

For investigating the Cost Growth Factor (CGF) – we define this shortly – for overall program total, a program was required to have cost data at MS B as well as on the last reported SAR. Eleven programs were missing cost data, reducing the initial 194 to 183 MDAPs for analyzing the CGF with respect to total program growth. For analyzing changes in PAUC, a program also needed quantity data. This 183 was further reduced to 165 since 18 MDAPs were missing quantity data. Table A.1 highlights these three databases used for comparing schedule, total program, and PAUC growth over the decades.

Completed Vs. Ongoing Programs	
Schedule Difference (Yr) - Completed vs. Ongoing	
Completed Programs	70
Ongoing Programs	50
Total	**120**
Overall CGF - Completed vs. Ongoing	
Completed Programs	118
Ongoing Programs	65
Total	**183**
PAUC CGF - Completed vs. Ongoing	
Completed Programs	102
Ongoing Programs	63
Total	**165**

Table 2. Completed vs. ongoing program breakout by response.

Besides initially analyzing all MDAP data together (completed and ongoing), we also split the completed and ongoing programs into separate categories. We do this to compare any aggregate statistical trends detected. Table 2 highlights the number of programs with respect to completed and ongoing. We define completed as any MDAP that no longer reports any SAR information. Ongoing is just the opposite. Those ongoing MDAPs still report SAR data even for programs that might have had a MS B date decades ago. This is because of ongoing contracts still reporting on those MDAPs.

After finalizing our three databases, we standardized all the cost data. Since these MDAPs can take many years to complete, there are instances where their costs are re-baselined to a different Fiscal Year (FY). There were several programs that had their estimates at MS B set to an earlier FY, while the current estimates were in a different FY. To ensure internal consistency for a program, we used the current base years for that program and standardized all cost data to that particular year. We used the Secretary of the Air Force raw inflation indices to perform these calculations.

Responses

In our analyses, we compared how three responses have changed from 1970s to the 2010s. These three responses consist of changes in schedule, total program cost, and PAUC. Equation (1) defines the percent schedule growth utilized. The denominator reflects the time from MS B to the last reported IOC date, while the numerator reflects the time from the estimated IOC date provided at MS B to the last reported IOC date. A value of 0 indicates no schedule growth. A positive percentage highlights a schedule slippage, while a negative percentage indicates a program reaching IOC quicker than expected at MS B.

$$(\text{IOC date}_{\text{Last Reported}} - \text{IOC date}_{\text{Estimated at MS B}}) \,/\, (\text{IOC date}_{\text{Last Reported}} - \text{MS B date}_{\text{Actual}}) \quad \textbf{(1)}$$

$$\text{Total Program Cost}_{\text{Last Reported}} \,/\, \text{Total Program Cost}_{\text{Estimated at MS B}} \quad \textbf{(2)}$$

$$\text{Total \# of Units}_{\text{Estimated at MS B}} \,/\, \text{Cost Estimate}_{\text{Estimated at MS B}} \quad \textbf{(3)}$$

$$\text{Total \# of Units}_{\text{Last Reported}} \,/\, \text{Cost Estimate}_{\text{Last Reported}} \quad \textbf{(4)}$$

To analyze total program cost growth, we took the last reported total cost value and divided it by the estimated total program cost at MS B (or equivalent from acquisition programs from earlier time periods). Equation 2 displays this calculation that generated the CGFs for our analysis. A CGF of 1 equates to a program experiencing no change in total program cost from MS B to the latest SAR. A value less than 1 suggests the program costs less than estimated at MS B, while a value greater than 1 shows an increased total program growth. This CGF calculation has been utilized in previous cost growth studies (Arena et al., 2006; Younossi et al., 2007; Kozlak et al., 2017).

The last response analyzed focused on the unit level, specifically at the PAUC. Quantity changes could drive some cost growth within MDAPs. To analyze the PAUC changes, we divided the total number of units estimated on the MS B SAR by the total cost estimate on the same SAR. [Note: The total number of units includes development and production units.] Then we calculated the current PAUC by taking the quantity reported on the latest SAR and dividing that by the latest program cost. Equations 3 and 4 highlight these calculations. After those two values were determined, we then divided (4) by (3) to arrive at the PAUC CGF, similar to the logic of (2).

Statistical Analysis

The goal for our analysis is to compare the decades, 1970s to 2010s, with respect to schedule growth, total program CGF, and PAUC CGF. We conduct these analyses with all MDAPs (completed and ongoing), then with only completed programs, and finally just ongoing programs. These analyses consist of a combination of descriptive and inferential statistics. The descriptive statistics include reporting means, medians, standard deviations, coefficient of variations (CVs), and interquartile ranges (IQRs) by decade.

Regarding inferential analyses, the standard F-test conducted under an Analysis of Variance was originally thought to be the best methodology to compare the responses across the decades. However, the non-normality pattern of the data indicated a non-parametric approach would be more appropriate given such inferential techniques have no distributional assumptions. Consequently, we utilized the non-parametric Kruskal-Wallis (K-W) test to determine statistically significant differences in the responses across the decades (Laerd Statistics, 2018). The specific null hypothesis tested is that the responses across the decades are equivalent versus the alternative hypothesis that at least one decade performs differently than the others. If the null hypothesis is rejected, then we use the non-parametric Steel-Dwass (S-D) pairwise comparison to isolate the specific decade(s) that is/are different.

The K-W and S-D inferential non-parametric tests are concerned with the typical response of a variable of interest. To assess how the variability of our responses (schedule growth, total program CGF, and PAUC CGF) might change across the decades, we employed the Brown-Forsythe (B-F) test. The B-F tests whether the response standard deviations/variances are equal or different across the decades. The B-F analyzes deviations based on the medians rather than the means of the data to minimize the effect of outliers or skewness in

Completed Vs. Ongoing Programs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	21	21%	20%	0.20	0.36	0.93
1990	27	32%	21%	0.28	0.44	0.87
2000	35	32%	21%	0.50	0.31	1.58
2010	25	13%	8%	0.31	0.28	2.34

Table 3. Schedule growth percentage summary statistics – Ongoing and completed programs. Median and mean values converted to percentages

the data (Brown & Forsythe, 1974, Stephanie, 2015). Since our data is not normally distributed, utilizing the B-F test provides more robust results versus the Levene Test, which uses means in its calculation. A level of significance of 0.05 was the default value that we used for all inferential hypothesis tests.

Analysis and Results

Total

The first analysis entailed all data, combining completed and ongoing programs. Table 3 presents the descriptive statistics for schedule growth by decade. All the means and medians are positive indicating consistent schedule slippage throughout the 1970s to 2010s for the typical MDAP program. The K-W and B-F tests returned *p*-values of 0.2123 and 0.6198, respectively, indicating no statistical difference among the decades with respect to the typical amount of schedule growth.

Table 4 presents the total program CGF by decade with one MDAP removed from the 1980s. This program is the DDG 51, the Arleigh Burke-class guided missile destroyers. Originally this MDAP had an initial purchase quantity of 14 ships; however, the most recent SAR shows the program acquiring 95. Such a dramatic change in units is more indicative of a scope change and not an issue with development/production issues. Thus the data are insufficient to parse cost/ schedule increases from the quantity increase.

The K-W and B-F tests returned p-values of 0.0812 and 0.0006, respectively. The p-value of 0.08, although not significant at the 0.05 level, does suggest that there may be evidence that the 1970s possessed higher total program CGFs than the other decades if one was willing to increase the level of a Type I error to 0.10. The very low p-value of 0.0006 for the B-F strongly suggests statistical differences among the standard deviations of total program CGF by decade. As seen in Figure 1, the standard deviation has been

Overall CGF Summary Statistics - Ongoing & Completed (Excluding DDG 51)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	29	2.83	1.44	3.62	2.88	1.28
1980	45	1.54	0.98	2.11	1.24	1.37
1990	37	1.66	1.26	1.17	1.51	0.71
2000	42	1.33	1.12	0.71	0.53	0.54
2010	29	1.14	1.02	0.37	0.25	0.32

Table 4. Overall CGF summary statistics – Ongoing and completed programs (Excluding the DDG 51 MDAP).

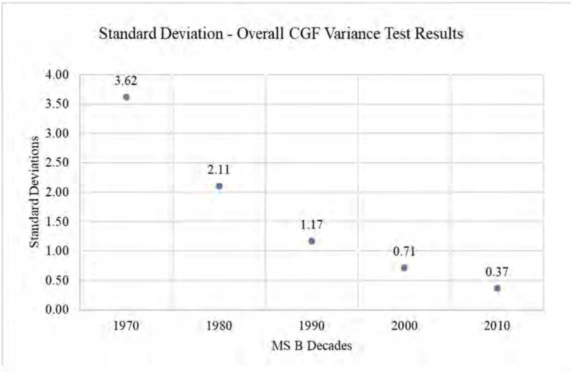


Figure 1. Standard deviations - Overall CGF of ongoing and completed programs.

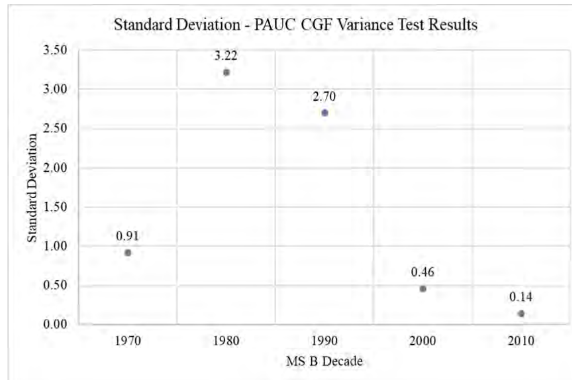


Figure 2. Standard deviations - PAUC CGF of ongoing and completed programs.

decreasing over the years. This appears to be a novel finding we haven't seen before in the literature. From our perspective, we have read many studies that have documented the patterns of cost and schedule growth. However, we have seen none that documented the actual variability of this process.

Table 5 presents summary statistics of the PAUC CGF by decade with one MDAP removed from the 2000s. This program is the C-130 AMP, which originally planned to acquire 519 units, but reported only purchasing nine (see Defense Industry Daily (2014) for some background on the decreasing number of units) on its most recent SAR. This drove PAUC from $7.26 million dollars per unit to $255.18 million dollars per unit. Since this outlier is markedly different from any other programs analyzed, we removed this program prior to conducting any inferential analysis.

The K-W and B-F tests returned p-values of 0.0302 and 0.0101, respectively. The K-W test concludes at least one decade is statistically different than the others with respect to PAUC CGF . The subsequent D-W test reveals that the 1970s and 1990s are statistically different than the 2010s with p-values of 0.0505 and 0.0411, respectively. These two decades have higher median PAUC CGFs than the lowest median PAUC CGF of the 2010s. With respect to the variability of PAUC CGF, Figure 2 highlights the standard deviations by decade; specifically, the 1980s and 1990s have statistically higher standard deviations than the other three decades while the 2000s and 2010s are decreasing.

Completed

We now duplicate the prior analysis but restrict it to just completed MDAPs. Table 6 presents the descriptive statistics for schedule growth by decade prior to any exclusions. The 2010s have only two MDAPs, indicating too few data to draw

PAUC CGF Summary Statistics - Ongoing & Completed (Excluding C-130 AMP)

MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	25	1.55	1.36	0.91	1.10	0.59
1980	35	2.39	1.10	3.22	0.94	1.35
1990	37	2.15	1.26	2.70	0.99	1.25
2000	38	1.17	1.11	0.46	0.38	0.39
2010	29	1.01	1.02	0.14	0.19	0.14

Table 5. PAUC CGF summary statistics – Ongoing and completed programs (excluding C-130 AMP).

Schedule Growth Percentage Summary Statistics - Completed MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	20	20%	20%	0.19	0.34	0.97
1990	23	27%	18%	0.26	0.23	0.99
2000	13	34%	11%	0.66	0.38	1.95
2010	2	4%	4%	0.06	0.08	1.41

Table 6. Schedule growth percentage summary statistics – Completed programs. Means and medians converted to percentages.

Schedule Growth Percentage Summary Statistics - Completed MDAPs (Excluding Joint MRAP)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	12	24%	24%	0.15	0.27	0.65
1980	20	20%	20%	0.19	0.34	0.97
1990	23	27%	18%	0.26	0.23	0.99
2000	12	17%	11%	0.22	0.30	1.35

Table 7. Schedule growth percentage summary statistics – Completed programs (excluding Joint MRAP). Means and medians converted to percentages.

any meaningful conclusions about this decade. In addition, there is a noticeable outlier in the 2000s belonging to the Joint Mine Resistant Ambush Protection (MRAP) MDAP; its schedule growth was approximately 240%. Table 7 presents the remaining descriptive data after removing these three programs and remain excluded for the K-W and B-F tests. The K-W and B-F tests returned p-values of 0.5208 and 0.3340, respectively, indicating no statistical difference among the decades with respect to the amount of schedule growth. This conclusion is consistent with using both completed and on-going MDAPs.

Table 8 presents the total program CGF by decade with the 2010s again removed (just two MDAPs completed). The K-W and B-F returned p-values of 0.1302 and 0.0270, respectively. The p-value of 0.1302 suggests that the decades are similar with respect to total program CGF, but the 0.0270 for the B-F suggests that the variability is not equal. As seen in Figure 3, there appears to be a decreasing trend in total program CGF variability by decade; a trend we witnessed in Figure 1.

Table 9 presents summary statistics of the PAUC CGF by decade with again the 2010 MDAPs (only two) removed and the exclusion of the C-130 AMP MDAP from the 2000 decade. The K-W and B-F tests returned p-values of 0.3508 and 0.4275, respectively. These results suggest no statistical differences with respect to the PAUC CGF (values or standard deviations) for the 1970s to the 2000s. This result is contradictory to what we concluded with all the programs, completed and ongoing. This suggests that the next section may reveal that PAUC CGF mainly varies between the decades for just ongoing programs.

Overall CGF Summary Statistics - Completed						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	28	2.83	1.37	3.69	2.91	1.30
1980	41	1.50	0.98	2.15	1.31	1.43
1990	29	1.41	1.01	1.00	1.27	0.71
2000	18	1.14	1.11	0.44	0.25	0.39

Table 8. Overall CGF summary statistics – Completed programs.

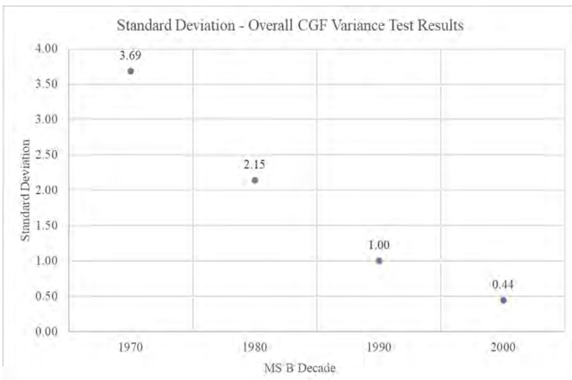


Figure 3. Standard deviations - Overall CGF of completed programs.

Ongoing

This section analyzes just the ongoing MDAPs. The 2000s and 2010s contained the bulk of our ongoing programs, but there are a couple of programs from the 1980s and 1990s that are still active and ongoing (e.g., reporting development/ production SARs). Because the K-W test needs at least five observations per group for statistical validity (Kruskal & Wallis, 1952), we removed from consideration any decade that did not meet the sample size criteria for either the schedule, total program CGF, or PAUC CGF analysis.

Table 10 presents the descriptive statistics for schedule growth by decade. As we have seen previously, both the means and medians are positive indicating consistent schedule slippage throughout the years for a typical MDAP program. The K-W and B-F tests returned p-values of 0.1067 and 0.9398, respectively, indicating no statistical difference among the decades with respect to the amount of schedule growth. This conclusion has been consistent throughout our analysis.

Table 11 presents the total program CGF for decades that had five or more MDAPs reporting development/production SARs. The K-W and B-F returned p-values of 0.0069 and 0.0020, respectively. The p-value of 0.0069 suggests that the decades are different with respect to total program CGF. The S-D test returned a p-value of 0.0096 when comparing the 1990s and 2010s, indicating that the 1990s total program CGF were

PAUC CGF Summary Statistics - Completed (Excluding C130 AMP)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1970	25	1.55	1.36	0.91	1.10	0.59
1980	31	2.03	1.05	2.83	0.9	1.4
1990	29	2.29	1.26	2.88	1.37	1.26
2000	14	1.19	1.07	0.66	0.19	0.56

Table 9. PAUC CGF summary statistics – Completed programs (excluding C130 AMP).

Schedule Growth Percentage Summary Statistics - Ongoing MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
2000	22	30%	39%	0.39	0.31	1.29
2010	23	14%	32%	0.32	0.30	2.29

Table 10. Schedule growth percentage summary statistics – Ongoing programs. Means and medians converted to percentages.

statistically higher than those of the 2010s. The 2000s were statistically equivalent to both decades. The 0.0020 p-value for the B-F test suggests that the standard deviations associated with total program CGF is not equal across the decades. As seen in Figure 4, there appears to be a decreasing trend in total program CGF variability by decade; a trend we witnessed in Figures 1 and 3.

Table 12 presents summary statistics of the PAUC CGF by decade with one outlier removed from the 1990s, the National Security Space Launch (NSSL) MDAP. This program possessed approximately a 6 PAUC CGF, while the next highest was around 1.6. The K-W and B-F tests returned p-values of 0.2564 and 0.0001, respectively. The p-value of 0.2564 suggests PAUC CGF through the three decades investigated are statistically equivalent. The p-value for the B-F test suggests that the variability associated with PAUC CGF is not equal. As seen in Figure 5, there appears to be a decreasing trend in PAUC CGF variability for the last three decades; a trend also shared by total program CGF. The next section discusses the significance of the statistical findings from our analysis.

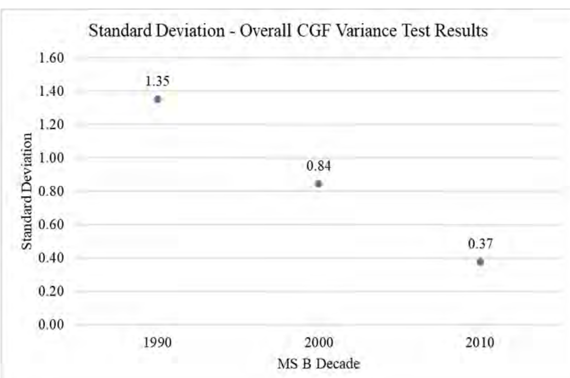


Figure 4. Standard deviations - Overall CGF of ongoing programs.

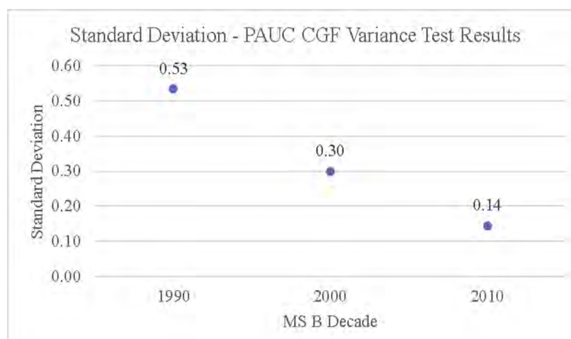


Figure 5. Standard deviations - PAUC CGF of ongoing programs.

Overall CGF Summary Statistics - Ongoing MDAPs						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1990	8	2.56	2.79	1.35	2.24	0.53
2000	24	1.47	1.21	0.84	0.64	0.57
2010	27	1.16	1.04	0.37	0.24	0.32

Table 11. Overall CGF summary statistics – Ongoing programs.

PAUC Summary Statistics - Ongoing MDAPs (Excluding NSSL)						
MS B Decade	Observations	Mean	Median	Std Deviation	IQR	CV
1990	7	1.00	1.21	0.53	0.91	0.54
2000	24	1.16	1.15	0.30	0.44	0.26
2010	27	1.01	1.02	0.14	0.19	0.14

Table 12. PAUC CGF summary statistics – Ongoing programs (Excluding NSSL).

Discussion and Conclusion

This article investigated whether cost and schedule estimations are improving over the decades. Despite numerous reforms and initiatives enacted to improve cost and schedule performance, our analysis found very few instances where schedule growth, total program CGF, or PAUC CGF statistically differed across the decades. Rather, our finding corroborated previous studies such as Arena et al. (2006) and Younossi et al. (2007) where schedule and cost growth are consistently positive across the decades.

Although the initial purpose of this study was to examine average cost and schedule trends, the most novel and exciting results were found elsewhere. This novel finding was found through an examination of the standard deviations of the CGFs across the decades. As shown throughout the analysis, even when the CGFs themselves were not statistically different across decades, there were differences detected in the variances of the CGFs themselves. This observation was seen for both overall CGF and PAUC CGFs for ongoing and completed programs. Perhaps most exciting is that these variances were generally decreasing. The overall CGF variance decreased through the five decades reviewed, while the PAUC CGF variance has decreased in every decade since the 1980s. Similarly, for completed programs the overall CGF variance has decreased since the 1970 while the PAUC CGF variance of on-going programs has decreased since the 1990s. See Figure 6.

To reiterate, although there were no identifiable statistical trends pointing to the DoD improving its schedule or cost estimation accuracy, the variances of the cost estimates have been noticeably decreasing from the 1980s onward. This decrease appears to be a new finding not seen in the literature previously. MDAPs are very expensive and time-consuming programs. Frequently, these programs are pushing technology capabilities. That alone suggests that cost growth and schedule slippage might just be endemic to MDAPs. However, the decreasing variability of cost estimates suggest to us that

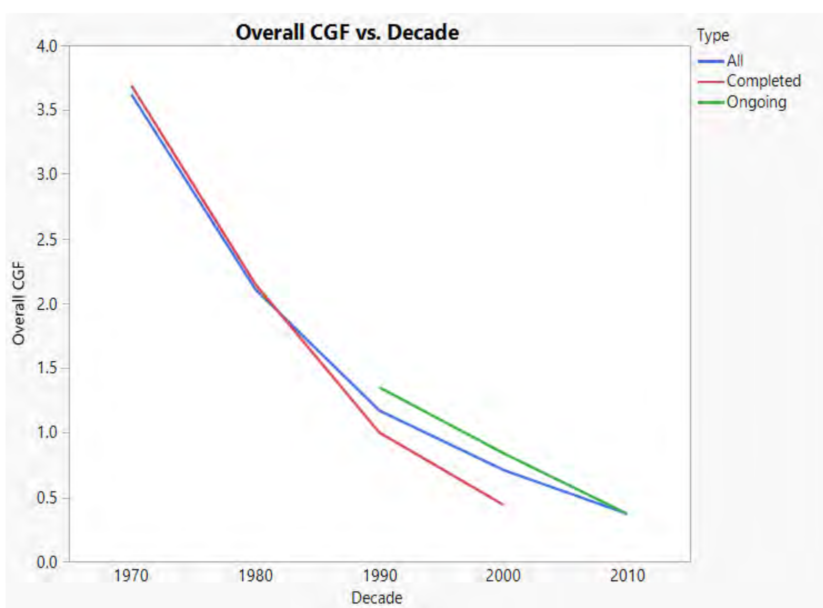


Figure 6. Standard deviations of overall total program CGFs.

cost estimators and/or the process behind them might be improving over time.

In statistics, those combinations speak to a bias outcome with minimal variance. In our opinion based on our experience in analyzing MDAPs over the years, that suggests perhaps the continued systematic bias of keeping initial cost estimates on the smaller side to make budgets more palatable. But eventually, the inherent risks of MDAPs are realized and true costs start accumulating. That is when cost growth appears. However, the variability of this cost growth difference has been reducing over the decades. That is the good news story. This is not to say that cost estimating cannot improve. However, we believe this cost growth is more of an artifact of keeping cost appearances low and not a reality of poor cost estimating. Thus, we humbly suggest program managers and executives consider this information when selecting the confidence level for budgetary inputs from a MS B estimate.

References

Acquisition Category. (2021). In AcqNotes [Online encyclopedia]. Retrieved from https://acqnotes.com/acqnote/acquisitions/milestone-b

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (Report No. TR-343). Santa Monica, CA: RAND. https://apps.dtic.mil/sti/pdfs/ADA455020.pdf

Brown, M. B., & Forsythe, A. B. (1974). Robust tests for the equality of variances. *Journal of the American Statistical Association*, *69*(346), 364–367. https://doi.org/10.1080/01621459.1974.10482955

Defense Industry Daily. (2014). Retrieved from https://www.defenseindustrydaily.com/c-130-amp-program-receives-contract-heads-into-uncertainty-06046/#:~:text=The%20C%2D130%20AMP%20program%20was%20originally%20supposed%20to%20cost,come%20under%20a%20separate%20contract.

Dwyer, M., Tidwell, B., Blivas, A., & Hunter, A. (2020, April 8). *Cycle times and cycles of acquisition reform*. Acquisition Research Program, Naval Postgraduate School. https://dair.nps.edu/bitstream/123456789/4183/1/SYM-AM-20-046.pdf

Fox, J. R., Allen, D., Moody, W., & Shiman, P. (2011). *Defense acquisition reform 1960-2009: An elusive goal.* U.S. Government Printing Office. https://history.defense.gov/Portals/70/Documents/acquisition_pub/CMH_Pub_51-3-1.pdf

Government Accountability Office. (2021, June). *Weapon systems annual assessment: Updated program oversight approach needed* (GAO-21-222). United States Government Accountability Office. https://www.gao.gov/assets/gao-21-222.pdf

Kozlak, S., White, E., Ritschel, J., Lucas, B., & Seibel, M. (2017). Analyzing cost growth at program stages for DoD aircraft. *Defense Acquisition Research Journal*, *24*(3), 386–407. https://doi.org/10.22594/dau.16-763.24.03

Kruskal, W.H. & Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*, 583-621.

Laerd Statistics. (2018). *Kruskal-Wallis H Test in SPSS Statistics | Procedure, output and interpretation of the output using a relevant example.* Lund Research Ltd. https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php

Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for Major Defense Acquisition Programs (MDAPs)* (Report No. RR1723). Santa Monica, CA; RAND. https://www.rand.org/pubs/research_reports/RR1723.html

Major Defense Acquisition Program. (2020). In AcqNotes [Online encyclopedia]. Retrieved from https://acqnotes.com/acqnote/acquisitions/major-defense-acquisition-program

McNicol, D (2018). *Acquisition policy, cost growth, and cancellations of Major Defense Acquisition Programs* (IDA Publication No. R-8396). Institute for Defense Analysis. https://apps.dtic.mil/sti/pdfs/AD1122550.pdf

Monaco, J. V. & White, E. D. (2005). Investigating schedule slippage. *Defense Acquisition Review Journal*, *12*(2), 176–193. https://doi.org/10.21236/ada441770

Nunn-McCurdy Breach. (2021). In DAU Glossary [Online encyclopedia]. Retrieved from https://www.dau.edu/glossary/Pages/GlossaryContent.aspx?itemid=28040

Riposo, J., McKernan, M., & Kaihoi, .Chelsea. (2014). *Cycle times and schedule growth in defense acquisition: A literature review* (RR-455-OSD). Santa Monica, CA: RAND. http://www.rand.org/pubs/research_reports/RR455.html

Stephanie, G. (2015, March 11). *Brown-Forsythe test: Definition*. Statistics How To. Retrieved December 15, 2021, from https://www.statisticshowto.com/brown-forsythe-test/

Under Secretary of Defense, Acquisition, Technology, and Logistics (USD(AT&L). (2016). *Performance of the Defense Acquisition System: 2016 Annual Report*. Washington, DC: Department of Defense. Available at https://apps.dtic.mil/sti/pdfs/AD1019605.pdf

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jain, A., & Sollinger, J. M. (2007). *Is weapon system cost growth increasing? A quantitative assessment of completed and ongoing programs* (Report No. MG-588-AF). Santa Monica, CA: RAND. https://doi.org/10.7249/mg588

Captain ***Sammantha Jones****, USAF, is a recent graduate of the Air Force Institute of Technology (AFIT) where she earned a Master's of Science in Cost Analysis. She is currently assigned as a cost analyst for the Air Force Life Cycle Management Center Enterprise IT and Cyber Infrastructure Division at Hanscom AFB. (E-mail address: sammantha.jones@us.af.mil)*

Dr. ***Edward D. White*** *is a Professor of Statistics in the Department of Mathematics and Statistics at AFIT. He received his BS in Mathematics from the University of Tampa, MAS from The Ohio State University, and PhD in Statistics from Texas A&M University. His primary research interests include statistical modeling, simulation, and data analytics. (E-mail address: Edward.White@afit.edu)*

Dr. ***Jonathan D. Ritschel*** *is an Associate Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in Accountancy from the University of Notre Dame, MS in Cost Analysis from AFIT, and PhD in Economics from George Mason University. His research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. (E-mail address: Jonathan.Ritschel@afit.edu)*

Mr. ***Shawn M. Valentine*** *is an Operations Research Analyst at AFLCMC/FZCE at Wright Patterson AFB, Ohio. Mr. Valentine graduated from Ohio University with a Bachelor of Science (BS) in Actuarial Science and a Master of Science (MS) in Financial Economics. (Email address: shawn.valentine@us.af.mil)*

8D Cost Trades with Entanglement

Douglas K. Howarth

Abstract: Many products have sizable components that comprise much of their costs. In such cases, it is crucial for the suppliers and the offerors of the ultimate products to work together to achieve common goals. It is possible to display such interactions in as few as two dimensions, and many firms might seek such answers as they are easily constructed and understood. However, as this paper demonstrates, seeking easy answers by artificially reducing the scope of the problem can lead producers astray. It is possible to get all the costs right and still sink a project. This paper proposes a construct with more dynamic elements, as it uses eight dimensions to understand how jets and their engines can work in tandem to enhance sales. This specific example generalizes to other markets.

Introduction

Jets and jet engines. You can't sell one without the other. What happens when the market interactions between them are not fully understood? This issue is not a hypothetical question, nor one without an answer. Recently, an example occurred in the business jet market, with more than $1B lost on a single project.

Texas billionaire Robert Bass founded Aerion in 2003 and began developing the Aerion AS2 in 2004. In December 2020, I wrote on LinkedIn that the plane was worth every penny of its $120M price tag, but there were not enough pennies in the world to hit its demand target. Aerion wrote a firm retort days later, claiming new orders. I repeated my position, citing my evidence.

The company halted development in May 2021 and went into liquidation that September.

They're not writing me anymore.

How can we validate both its cost and price but confidently invalidate the project in advance? It is possible to describe the business jet market in four dimensions, and that for its engines with the same number, less the one common price dimension both share, for a total of seven. Time adds the eighth dimension.

This paper studies how these eight dimensions interact as they entangle.

Historic Context

Paul Samuelson, considered by many the father of modern economics and the 1970 winner of the Nobel Memorial Prize in Economic Sciences, had definite thoughts about price determination. He wrote that the law of supply and demand meant that "the equilibrium price, i.e., the only price that can last…must be at this intersection point of supply and demand curves."

Every introductory text in economics has this paradigm in one form or another, though those examples are uniformly hypothetical. Where do we find these relationships in the real world?

We can see a modern example in Figure 1, in the market for iron ore, where costs rise from mine to mine. After adding a profit margin above their costs, the mines collectively form an upward-

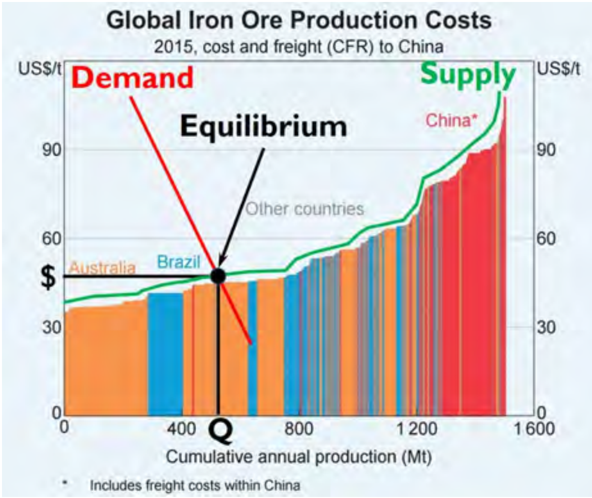


Figure 1 – Iron Ore Market Equilibrium

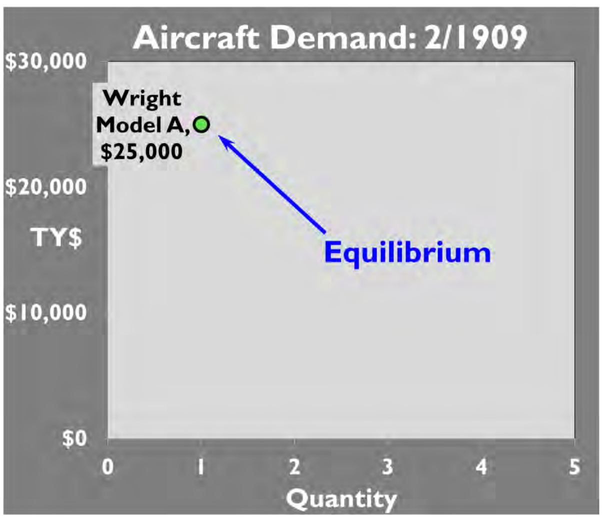


Figure 2 – Aircraft Market Equilibrium

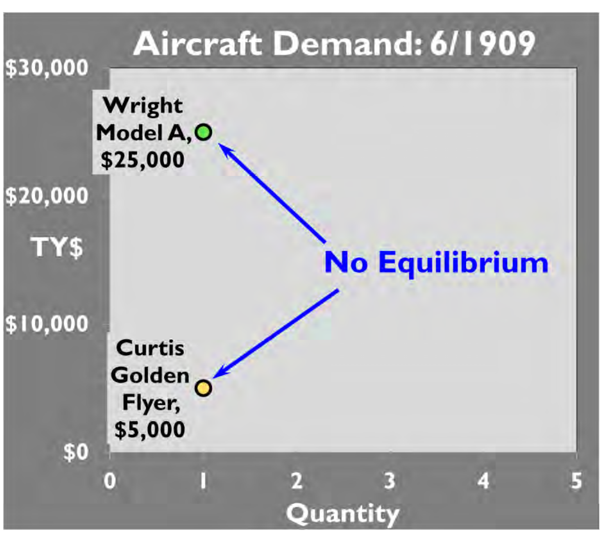


Figure 3 – Aircraft Market Disequilibrium

sloping demand curve, a hallmark of modern economic analysis.

But the market for a commodity such as iron differs from other markets that use it. Aircraft have iron in them. The Wright Brothers made the first aircraft sale in February 1909, when they contracted with the US Army to provide one Model A Flyer for $25,000. As in Figure 2, classical economics would say the market is in single-point equilibrium with this lone exchange in the market. But months later, when Glenn Curtis sold a second airplane in June 1909 for $5,000, that put two distinct points in that market, as Figure 3 reveals. This observation is sufficient to negate the law of supply and demand. So, the question becomes: What replaces it?

Observing the market is in disequilibrium does not suggest that it is in disarray. It merely notes that we have not accounted for other forces at work. We'll investigate those in a bit. Right now, we'll study demand in more detail.

The Known Twin

With any kind of luck, identical twins know their sister or brother their entire life. Typically, each of them would have a solid bond with the other. It would be hard for one to comprehend the other being unknown to them. Despite their close bonds in this modern world, it might be possible for one of them to have a much higher media presence than the other, making one of them effectively invisible in that realm, unknown to the public.

Let's give the twins names and jobs to make this example more tangible.

Figure 4 – Cristina Models Learning & Demand

We get a unique type of estimator in Cristina, from Figure 4, as she's studying aerospace learning and demand curves. Based in Argentina, Cristina got a bad taste in her mouth when she heard about the law of supply and demand. No amount of locally sourced Argentinian Malbec could rinse it out. She couldn't shake it out of her limbs with a brisk ride across the Pampas, dancing the tango, or jumping in a stadium watching Lionel Messi playing football.

It stuck in her head as an anomaly. It was certainly something she wanted to explore in detail.

She never believed in the single-point equilibrium theory. She reasoned that there are dozens of business aircraft, from the smallest that can squeeze a few people into them to converted airliners that seat hundreds. There were private turboprops for a few million dollars at the low end of this market and converted jumbo jets for nearly a quarter billion dollars up at its top, with speeds ranging from a couple of hundred miles per hour up to high subsonic models.

There couldn't be a sole point that described them all in a mathematically valuable way.

She plots the market's quantities sold by model and the prices they command. To complete the analysis, she adds turboprops to all the business jets she collects in her database in Figure 5, assembling data on 95 models over ten years.

When Cristina plots her data, she can refute the

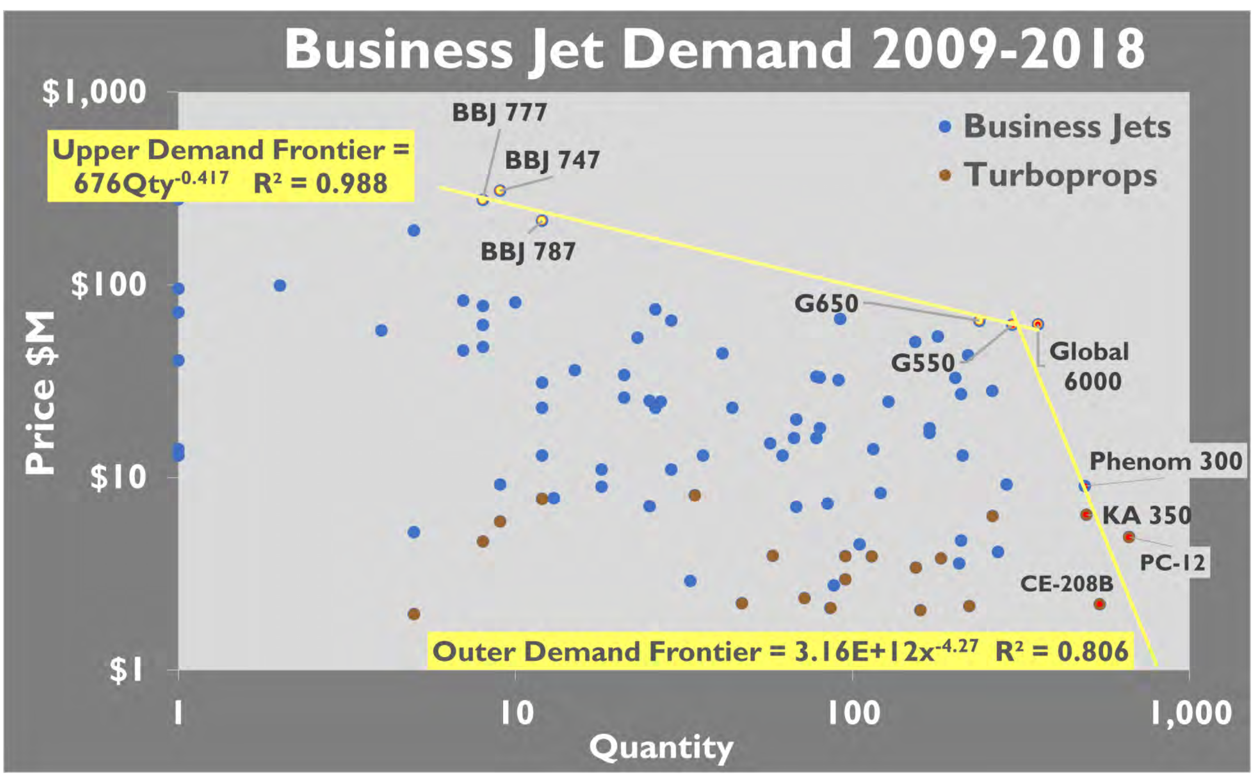


Figure 5 – The Business Aircraft Market has well-defined Upper and Outer Demand Frontiers

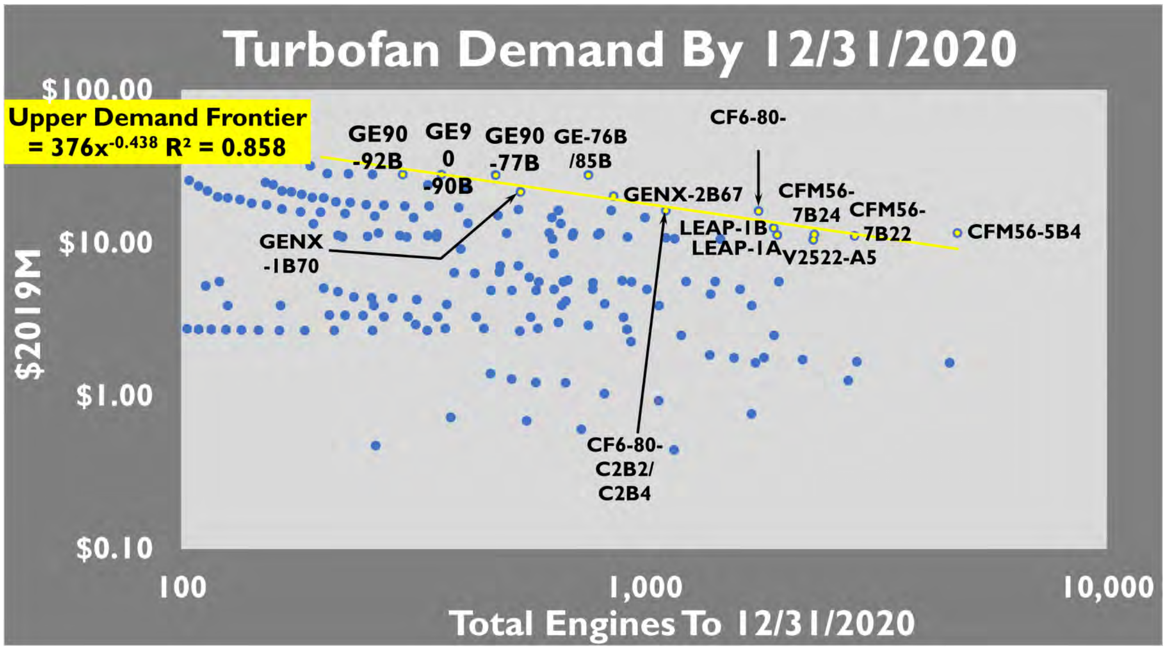


Figure 6 – The Upper Demand Frontier for Turbofan Engines is statistically significant

hypothesis of a single-point equilibrium and discover some other interesting phenomena. Figure 5 reveals that the market has at least a pair of self-organizing features which are statistically significant. Along the higher reaches of the business aircraft market, there is an Upper Demand Frontier, as Equation 1.

$$2019\$M = \$676.4 * Quantity^{-0.417} * \varepsilon \qquad (1)$$

Where:

2019$M = Predicted Upper Demand Frontier $ for business aircraft

Quantity = Aircraft sales 1/1/2009 to 12/31/2018

ε = Error term for the equation

Adjusted for bias using the Ping Factor (as all equations are in this piece, thus, that factor won't be noted again), the Equation 1 curve represents a limit to how much money the market has within it to buy the highest-priced aircraft. With a P-Value of 5.39E-05, an adjusted R2 of 98.5%, a standard error of $20.5M, and an especially low Mean Absolute Percentage Error (MAPE) of 6.9%, players in the market should not ignore it. Manufacturers operating in that region should also take note of its relatively flat slope of -0.417. That indicates that price reductions offer the chance for more than proportional increases in sales. Cases in point include the Global 6000 (which sold 355 units * $62.31M/unit = $22.1B over a decade), found at the right end of this curve, generated more than ten times the revenue than the B777 (8 units * $275.96M/Unit = $2.2B for those ten years) near the left end of this curve.

Cristina finds this market bounds itself concerning quantities sold with an Outer Demand Frontier, as represented by Equation 2.

$$2019\$M = \$3.16E{+}12 * Quantity^{-4.27} * \varepsilon \qquad (2)$$

Where:

2019$M = Predicted Outer Demand Frontier $ for business aircraft

Quantity = Aircraft sales 1/1/2009 to 12/31/2018

ε = Error term for the equation

Equation 2, with a P-Value of 1.52E-02, is statistically significant, though less well correlated than Equation 1. It has an adjusted R2 of 75.7%, a standard deviation of $14.3M, and a

MAPE of 63.1%. This line means that at its limit, the market has absorbed as much product as possible for a given period, which is ten years for the case at hand.

Cristina finds the slope of the Upper Demand Frontier especially intriguing and decides to concentrate on it. If the manufacturer's cost structure can support the potential increase in revenue due to price decreases, the firm could improve profits. Of course, airframers will look to their suppliers to help offer such prices. As engines make up a significant cost component of business aircraft (from 17% to 40%), she decides to study the market for turbofan engines.

The engine manufacturers' prices are costs to the airframers.

The turbofan engine market has many more models, and Cristina found 186 distinct models that were active at the time of the compilation of her database. Those points form the blue dots in Figure 6. She observes another self-organizing Upper Demand Frontier for the turbofan engine market, which we can characterize as Equation 3.

$$2019\$M = \$376.3 * Quantity^{-0.437} * \varepsilon \qquad (3)$$

Where:

2019$M = Predicted Upper Demand Frontier $ for turbofan engines

Quantity = Aircraft sales 1/1/2009 to 12/31/2018

ε = Error term for the equation

Equation 3 mimics the one for business aircraft, as it, too, has a flat angle across log-log space. At the same time, its adjusted R2 of 84.9% is less well-correlated than the same curve for business aircraft. We need to recognize its deeper meaning with its P-Value of 1.76E-06, MAPE of 8.3%, and standard deviation of $2.72M. In this market, as we found in the one for business aircraft, price reductions may be met with proportionally more significant revenue increases, as long, that is, as those engines can find willing airframers to use the models in question.

After all her work, Cristina found that no single point equilibrium exists for the business aircraft or turbofan engine markets. Such a curve would mean costs increase with the number of units (see Figure 1). However, if that condition is applied to business jets and their turbofan engines, the builders of such devices take more time with successive units; they lose learning as they go along, essentially becoming dumber.

Surely, Cristina reasoned, that could not be the case. She knows people get smarter over time; that's what learning curves confirm.

But, if upward-sloping supply curves intersecting downward-sloping demand curves do not determine prices, what does, she wonders? She decides to contact her twin expatriate sister.

The Unknown Twin

Just as the southern hemisphere appealed to her sister, Sheila found herself drawn to the other side of the globe. In her case, she landed in the land down under. No wonder, then, when Cristina asked for help, Shiela piped back with a quick "no worries."

Figure 7 – Sheila studies features and their Value

Sheila works in what we could reasonably call the unknown realm of economics. She had a hunch that the product features have something to do with sustainable prices. That means Sheila doesn't believe in upward-sloping supply curves for products that are not commodities. She studied a 1987 RAND Corporation aircraft cost model and found costs increased with weight and speed, as Equations 4 and 5.

$$Labr_{100} = 0.141EW^{0.820} * SP^{0.484} \qquad (4)$$

$$Matl_{100} = 0.241EW^{0.921} * SP^{0.621} \qquad (5)$$

Where:

$Labr_{100}$ = Cumulative Manufacturing Labor Hours for 100 Aircraft (in thousands)

EW = Aircraft Empty Weight (in pounds)

SP = Maximum Speed (in knots)

Matl100 = Cumulative Manufacturing Material Dollars for 100 Aircraft (in thousands of 1977$)

RAND built Equations 4 and 5 on 13 observations. (Note: While this database is small, it offers something commercial aircraft do not – some supersonic examples. The top speeds in the RAND database exceed the projected top speed of emerging supersonic airliners and business jets, thus bounding the problem of figuring out the impact of speed). The labor equation, number 4, had an R2 of 88%, with P-Values for empty weight and a maximum speed of less than 0.001 and 0.013, respectively. Aircraft material in equation 5 had a better correlation at 91%, and its P-Values for weight and knots were less than 0.001 and 0.003, in that order.

As someone who studies the business aircraft market, it makes sense to Sheila that cost and Value should go up with speed. But, while cost models use weight to measure size or capacity, she reasons that the Value of space of business aircraft would be better estimated using some other metric. If adding weight were the best way to increase Value, all one need do is add lead to a plane to increase the sales price. That's not the best option. She notes she could use maximum passenger limits but realizes larger planes offer more space per traveler. Instead, she decides to see how the value changes with the cubic feet offered in each craft's cabin. After all, who doesn't want the room to spread out?

In Figure 8, we see her results. She finds that as cabin sizes increase, the sustainable prices do as well, according to Equation 6:

$$2019\$M = \$0.0463 * Cab\ Vol^{.897} * \varepsilon \qquad (6)$$

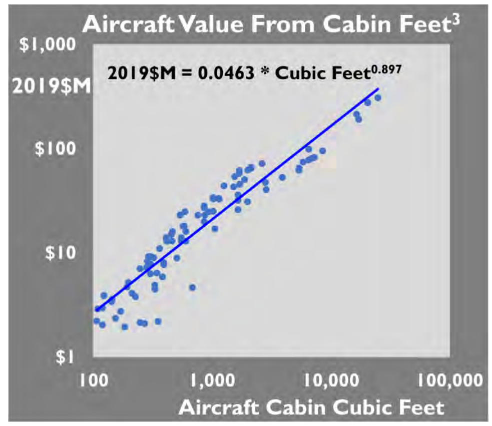


Figure 8 – Aircraft price goes up with cabin size.

Where:

2019$M = Predicted price for business aircraft

Cab Vol = Aircraft cabin volume (in cubic feet)

ε = Error term for the equation

Sheila finds Equation 6 is an excellent price estimator, with an Adjusted R2 of 89.8%, but is concerned with the MAPE of 38.5%. She observes that many of the smaller turboprop cabins fall below the line of best fit and decides to see how the market reacts to speed, as the prop-driven planes are slower.

Maximum Miles Per Hour, Shiela discovers, provides a viable estimator for the price of business aircraft, as shown in Equation 7.

$$2019\$M = \$7.82E\text{-}08 * MaxMPH^{\,0.897} * \varepsilon \qquad (7)$$

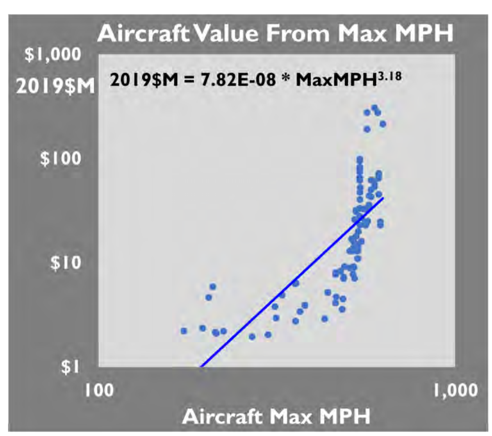


Figure 9 – Speed adds Value to business aircraft

Where:

2019$M = Predicted price for business aircraft

MaxMPH = Max aircraft speed (in miles per hour)

ε = Error term for the equation

Equation 7 is statistically significant (with a P-value of 1.85e-17). Its adjusted R^2 of 53.7% and MAPE of 78.8% aren't as good as those for Equation 6. She also notes that the speed exponent is extraordinarily high. She combines the analyses, using cabin volume and maximum MPH simultaneously.

When she does, she gets Figure 10, expressed by Equation 8.

$$\$2019M = 3.65E\text{-}05 * Cab\ Vol^{0.736} * MxMPH^{1.33} * \varepsilon \quad (8)$$

Where:

2019$M = Predicted price for business aircraft

Cab Vol = Aircraft cabin volume (in cubic feet)

MaxMPH = Max aircraft speed (in miles per hour)

ε = Error term for the equation

With an adjusted R2 of 96.4% and a MAPE of 19.3%, Equation 8 is a better predictor than either Equations 6 or 7. Sheila notes that the speed exponent is still high, at 1.33. She also remembers Cristina wants to study jets, not turboprops, so she removes the latter group from the dataset and reruns her analysis in Figure 11, which uses Equation 9.

$$\$2019M = 2.46E\text{-}10 * Cab\ Vol^{0.671} * MxMPH^{3.29} * \varepsilon \quad (9)$$

Where:

2019$M = Predicted price for business aircraft

Cab Vol = Aircraft cabin volume (in cubic feet)

MaxMPH = Max aircraft speed (in miles per hour)

ε = Error term for the equation

There are even better statistics for Equation 9, as its adjusted R2 is 97.5%, while the MAPE falls to 13.7%, using 75 observations, compared to the 95 used for Equations 6, 7, and 8. Note the dramatic difference in the slope for the speed component in Figure 11 compared to Figure 10. In the business jet market, buyers pay dearly for added speed.

Since she did so well with aircraft, Sheila decides to see how the engines that power them behave. When she does, she discovers Figure 12.

Since she did so well with aircraft, Sheila decides to see how the engines that power them behave. When she does, she discovers Figure 12.

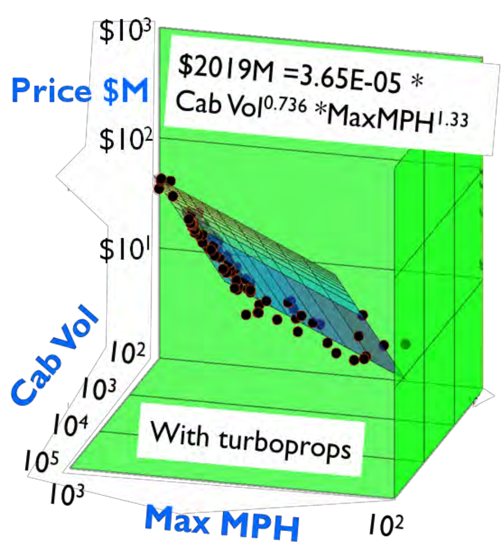


Figure 10 – Aircraft value (with turboprops) from Cabin Volume and Maximum Miles Per Hour

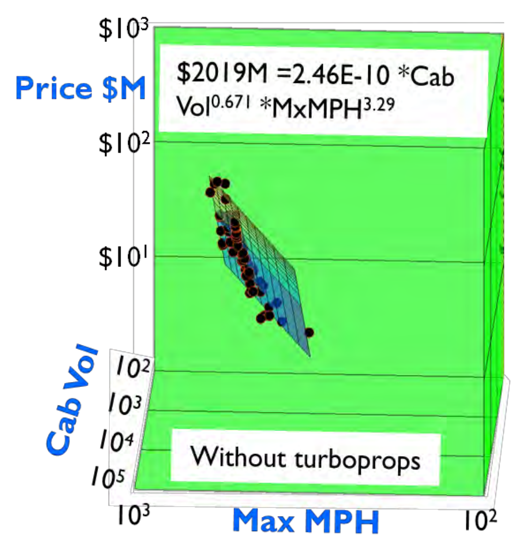


Figure 11 – Aircraft value (w/o turboprops) from Cabin Volume and Maximum Miles Per Hour

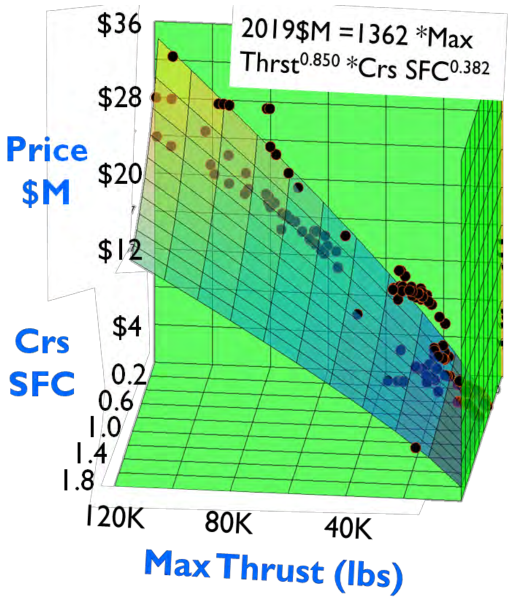


Figure 11 – Aircraft value (w/o turboprops) from Cabin Volume and Maximum Miles Per Hour

She describes Figure 12 with Equation 10.

$$2019\$M = 1362 * Max\ Thrst^{0.850} * Crs\ SFC^{-0.382} * \varepsilon \quad (10)$$

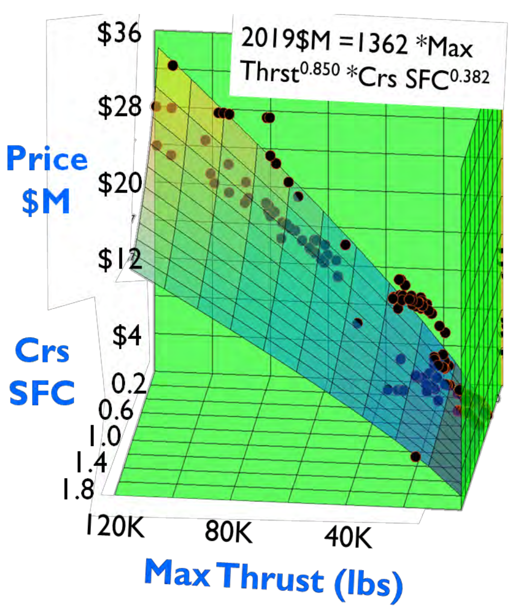


Figure 12 – Turbofan value from maximum thrust (in lbs) and Specific Fuel Consumption (lb/lbf/h)

Where:

2019$M = Predicted price for turbofans

Max Thrst = Max turbofan thrust in pounds

Crs SFC = Specific Fuel Consumption in lb/lbf/h at cruise speed

Equation 10, derived from 186 observations, is well-correlated with an adjusted R2 of 94.6% and a MAPE of 18.2%, with P-values of 9.51E-116 and 2.52E-06 for Max Thrust and Specific Fuel Consumption, respectively.

Pole Position

Now that they've completed some deep analysis of their problems, Cristina and Sheila wonder how they might be able to extend it. They remember how they used to share adjoining rooms as young girls living side by side and think about how they might recreate a similar environment for their work. At first blush, it appears problematic if they remain in their adopted countries, with Cristina and Sheila living on widely distant continents.

Then they ask themselves this: what if Australia and Argentina touched? After some reflection, they changed the question: Where do parts of Australia and Argentina meet? With Australia as the world's only continental country surrounded by ocean, that looks to be a trick question.

And it is.

But it's one with some hidden mathematical meaning buried in its geography, which may need only the slightest tweaks to offer a new, beneficial structure that is not widely known.

With a little recall and research, they rediscovered those countries, specifically their territorial claims on Antarctica, touched at the South Pole, as shown in Figures 13, 14, and 15. [6,7,8]

It shows us that point is near the Amundsen-Scott South Pole Station, and we can get a close-up view of it in Figure 14.

Figure 13 – The Argentine and Australian claims in Antarctica meet at the South Pole, with their airspaces separated by the Earth's axis

Australia has a couple of Antarctic slices in its claim, while Argentina lies entirely within the UK's.

If Sheila were to go to the South Pole and walk into the Australian claim, one could not say that she was in a negative Argentinian space. The same could be said for Cristina, as any movement of her part into the Argentine claim says nothing about the one Australia has. The twins note that, by convention, we call the point where all these claims meet 90° South latitude the South Pole. They wonder: What would happen if we called it something else?

3 + 2 =4

Given their analyses of business aircraft and observations about the South Pole, the twins decide to place the axes of their graphs near each other, as shown in Figure 15. Cristina's Demand Plane, at right, needs only a vertical plane of two axes and is easily accommodated by Argentina's claim. Again, it has a horizontal quantity axis and a vertical price axis.

At the same time, in the exact Figure, Sheila's three Value axes to the left, consisting of a pair of valued features, cabin volume, and maximum miles per hour, are plotted on horizontal axes and a vertical price axis.

The twins have a thought. What, they wonder, would happen if we placed our horizontal axes in line with the Earth's axis at the South Pole?

Figure 14 – The South Pole marker with a coffee cup; Amundsen–Scott South Pole Station behind

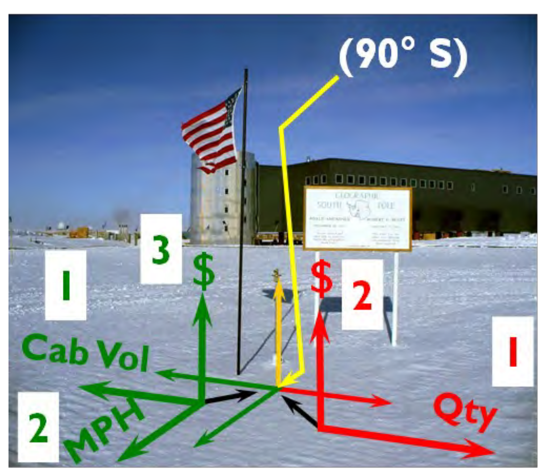


Figure 15 – What do the 3D Value Space and the 2D Demand Plane for Biz A/C have in common?

In Figure 16, they do just that and note that since the Value Space and the Demand Plane share the common vertical price axis, they need not replicate it in their axis count. Thus, the 3D Value Space and the 2D Demand Plane combine to form a 4D system, or 3 + 2 = 4. With this display system initiating at the South Pole, the twins think to rename its origin to (0, 0, 0, 0), with four axes representing (Value Feature 1, Value Feature 2, Price, Quantity).

In Figure 17, they additionally note that while the South Pole drew their systems together, there is no need to depict them starting there, so they drop the geographic reference. They recognize that all 4D market models form their unique systems.

Simply getting the axes in the right place is only the beginning of the analysis – they realize they must populate these systems in Figures 18 and 19.

In Figure 18, with both sides of the system fully populated with business aircraft, they make an added insight. The Value points on the left and those on the right for Demand are not separate.

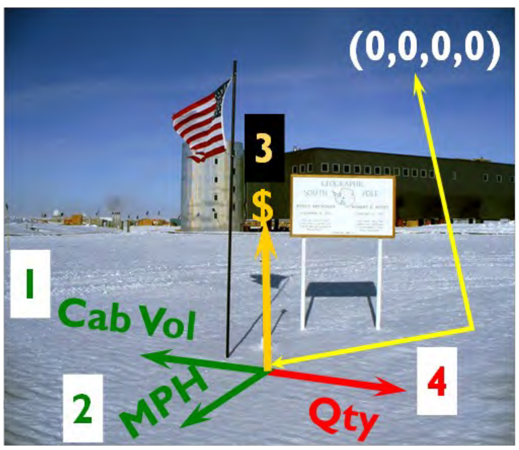


Figure 16 – 3D + 2D = 4D, since Value Space and Demand Plane share the Price Dimension

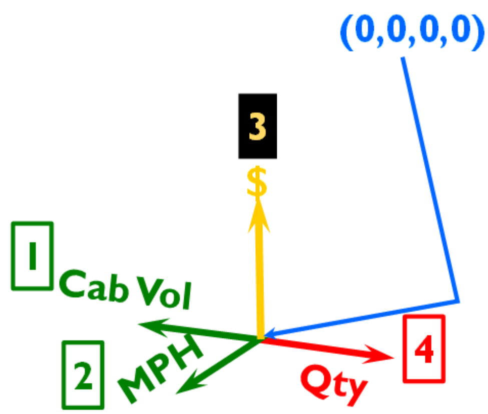


Figure 17 – 4D markets reside in their own space

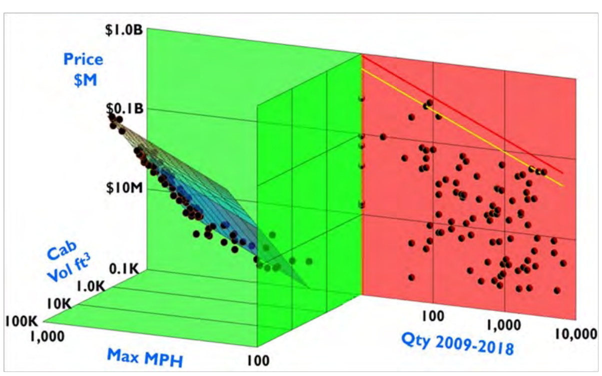


Figure 18 – The 4D market for business aircraft, with Value Space to the left, the Demand Plane at right

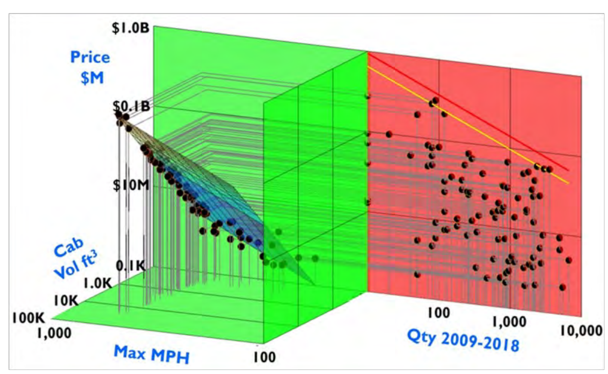


Figure 19 – All points in Business Aircraft Value Space have matches on the Demand Plane; Value and Demand entangle with each other in every market

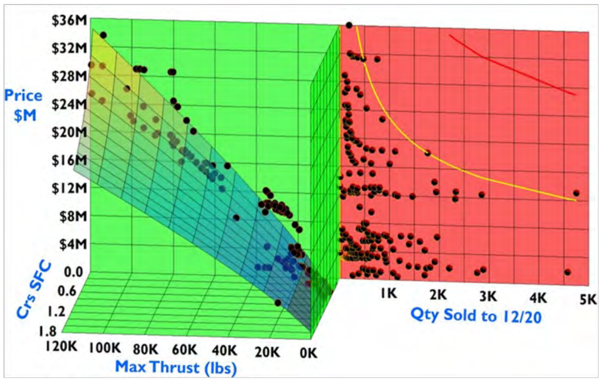


Figure 20 – The 4D market for turbofan engines, with Value Space to the left, its Demand Plane at right

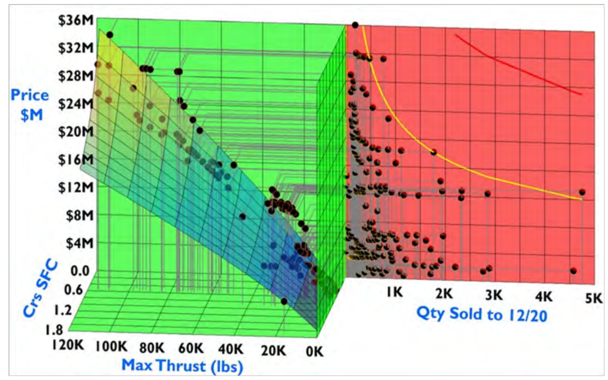


Figure 21 – Every point in the turbofan engine Value Space has a match on its Demand Plane as they entangle with one another

Instead, they entangle with one another through the common price axis.

To drive this thought forward, they draw point lines that connect each Value Space position with its mate on the Demand Plane in Figure 19.

Figures 20 and 21 copy their business aircraft methodologies for turbofan engines and draw similar conclusions. They begin to wonder if they may be able to reveal even more entanglement.

4 + 4 = 7

Once they break the convention of traditional land-based geometries, the twins realize there is little to prevent them from expanding their analyses. To that end, they observe that the turbofan market, as depicted, takes 180° of arc, as does the one for business aircraft; there is nothing to prevent them from pairing them together, as shown in Figure 22.

There, we find the four dimensions of the turbofan market combined with the same number in the business aircraft market. Since each market shares the same price axis, we portray both markets simultaneously with only seven axes.

But there is more to Figure 22 than meets the casual eye. As we noted at the beginning of this piece, turbofan engines mate with the jet aircraft that use them. The markets entangle with one

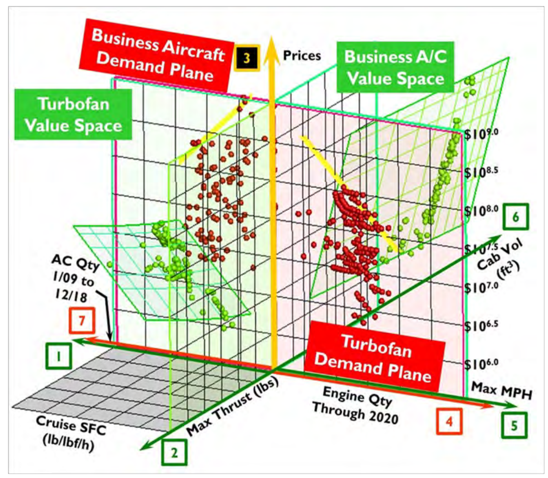


Figure 22 – Since the 4D turbofan and 4D business aircraft markets share a common price axis, they combine to form a 7D system (so, here, 4D + 4D = 7D)

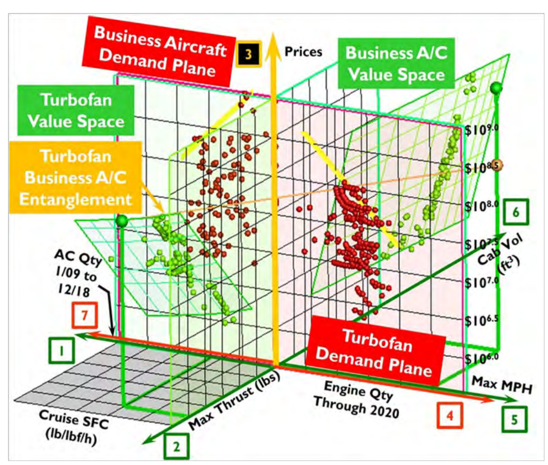


Figure 23 – Turbofans only sell jet aircraft. Thus, every point (i.e., engine model) in the turbofan market entangles with one or more business aircraft models that use it; here, a new engine (the largest sphere at left) finds a need in the business aircraft market (the sphere partway up the rightmost vertical line, forming part of a new aircraft cost)

another. Rather than show how every engine matches with one or more aircraft with which it may be paired, we can see this mutual relationship in Figure 23.

In Figure 23, at the behest of a business aircraft manufacturer, a maker of turbofan engines has built a new engine, shown as the largest sphere on the left-hand side of the figure, placed on the turbofan Value Surface. Note: it has a connecting line that runs to a matching sphere partway up the rightmost vertical line in the diagram, representing the cost of a single engine. That rightmost vertical line represents all the value components of a new business aircraft. The total Value of the new aircraft model is the like-sized sphere above the turbofan engine component, lying on the business aircraft Value Surface. (Observe there is some distortion in the apparent contribution of the engine component of the aircraft due to the log-scaling – here, the engine portion of cost appears to exceed 50%; in practice, it typically runs from about 17% to 40%, depending on the paired models).

Thus, the engine manufacturers depend entirely upon their aircraft manufacturers to buy their products, and airframers face a significant cost component in their engines. It makes the twins wonder how to perform cost trades between these markets.

A 7D Entangled Trade

Far from a hypothetical construct, Cristina finds real-world issues suppliers from both markets could alleviate with benefits accruing to both manufacturers.

In Figure 5, she found twelve (12) sales of the Boeing BBJ 787 over a decade, while in Figure 6, her work showed that one of its engines, the GEnx-1B had sales of 341 over the same length of time (the BBJ 787 also uses the Rolls-Royce Trent 1000). What if Boeing wanted to push to sell 20 of these units in a decade? What would have to happen?

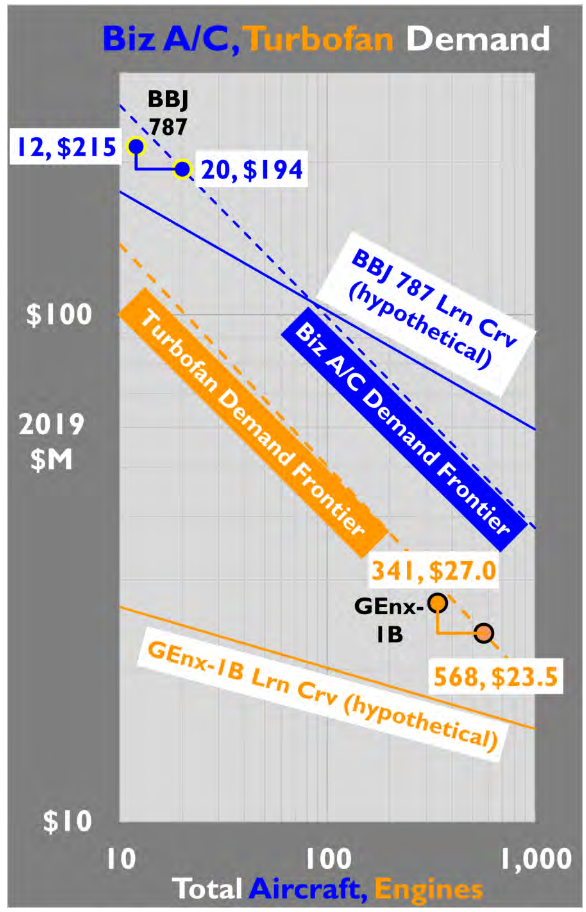


Figure 24 – The Boeing BBJ 787 might sell as many as 20 units in a decade, but only if its price falls in with its Demand Frontier; GE, with its GEnx-1B engine, might be able to offer relief

According to her calculations, if Boeing wants 20 units to sell in a decade, the price must fall nearly 10% to $194M as the Demand Frontier limits sales. Importantly, from Boeing's perspective, their GEnx-1B engines (they need two per plane) represent about 25% of the sales price at the current figures, and to get to this potential target, they'll want some supplier help. In all cases, Boeing will have to ensure the BBJ 787 price does not drop below its recurring cost, shown by its hypothetical BBJ 787 learning curve, for, in that case, they would be losing money.

General Electric, for its part, finds itself in a similar position. Its GEnx-1B is close to hitting its Demand Frontier. To get more sales, it will have to drop prices. They must compare their prices to their learning curve to verify that they can do that.

Crucially, if either firm were to keep its prices high, despite their demonstrated abilities to do so comfortably, sales for both would not attain their ideal, maximized level. Intractability on either side could lead to decreased profits for both parties. (Qualifier: If GE decided not to drop its price for its GEnx-1B, Boeing could try to work with Rolls Royce with their Trent 1000, another engine qualified for the BBJ-787). Current profits may mask this condition – one firm or the other may believe they are doing well enough while not realizing they are not doing as well as possible.

Two Out Of Three Ain't Bad – Or Is It?

The recently deceased rock and roll singer Meat Loaf told us in a song, "Two out of three ain't bad."

A primary hypothesis held by the twins is that for any project to succeed, producers have 1) cost, 2) value (as sustainable prices), and 3) demand working in concert with one another.

If the cost were to exceed the price, a program would stop. No one can stand to build at a loss consistently.

Values, again sustainable product prices, must align with the markets' view, as determined in the abovementioned methods. Overpricing leads to decreased sales; underpricing results in monies left on the table.

A vital result that follows is that producers need to abide by the Demand Frontiers they face. These limits have error terms, and roughly speaking, about half of the products that form them will lie outside their bounds, the others within them. But, as producers begin their quest

Figure 25 – The Aerion AS2

to launch a new product into their market, they must measure their markets' limits. Not doing so can lead to economic disaster.

With that in mind, in 2020, Cristina and Sheila analyzed the Aerion AS2, shown in Figure 25

The Aerion AS2 is a supersonic business jet, and the twins first want to know if its costs align with history. While it was, in 2020, yet to be in production, the firm did offer its development cost at $4B.

After some research, they construct Figure 26.

Despite the low number of observations in Figure 26, they note the P-Value for the line of best fit is well below their criterion of 0.05, at 4.65E-04. At 70,000 pounds empty weight, the AS2, with its projected $4B development cost, is 62% higher

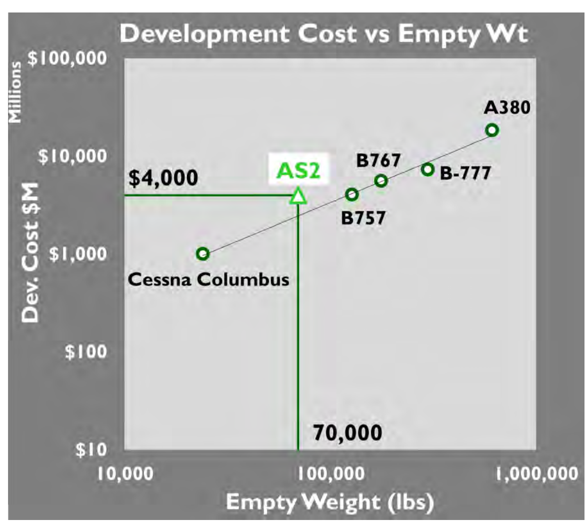


Figure 26 – Jet empty weight versus development costs, AS2 cost is 62% higher than that for subsonic models.

than the projection for all subsonic aircraft in this database. They strive to see if that should be sufficient additional monies for development.

Returning to the RAND model, they find a complete set of development equations and decide to compare the ratio of the projected cost components of the AS2 by discipline to each of those of the next fastest model in the dataset, the Boeing 777. When they do, they derive Figure 27.

Given Boeing's historical cost breakdowns, the added cost of going from 511 knots (the Boeing 777's top speed) to 805 nautical miles per hour (the AS2's maximum), the added 62% cost above subsonic development programs seems to be reasonable.

Discipline	RAND Exponent	Base Factor	AS2 Factor	Ratio
Engineering	1.03	616.13	983.94	1.60
Tooling	0.609	44.61	58.83	1.32
Mfg. Labor	0.429	14.52	17.64	1.22
Material	0.811	157.23	227.30	1.45
Design Supt.	1.28	2929.40	5241.06	1.79
Flight Test	1.27	2752.29	4901.86	1.78
Program	0.745	104.18	146.16	1.40

Figure 27 – Ratio of cost differences due to speed by discipline for AS2 relative to Boeing 777

Turning her attention to the Value of AS2, Sheila decides to put its projected cabin volume (1146 cubic feet) and top speed (925 miles per hour) into Equation 8; she finds the projected AS2 Value at $57 million. Realizing there is a premium for speed among business jet owners, she runs the same variables into Equation 9 (the one that removes turboprops); Sheila finds the market might support a price of $160M.

Since the company priced its AS2 at $120M and received some firm orders, the market proved it was worth that much.

So, Aerion had passed the initial 1) cost and 2) value tests, or two out of three key measures. Meatloaf would say that's already not bad. But what, the twins wondered, could they say about demand?

7 + 1 = 8

The twins hypothesize that Demand is something quite different from Value or Cost. They discovered that cost falls upon manufacturers, and they incur additional charges if they build larger or more complicated products or have a newer, inexperienced labor force. Costs fall with learning or added experience over time.

Despite manufacturers setting initial prices, thereby putting their stamp on Value, they've found that the buyers will set ultimate prices based on how they assign Value to all the features offered in goods and services. Value often falls over time; some new buyers only enter the market through lower prices.

The girls examine other markets to see the broader effects of Value, Cost, and Demand to see what phenomena might be ubiquitous.

Learning, which drops costs over time, enables price reductions, expanding the reach of disparate markets. For example, when it comes to consumer electronics, when she acquires the data forming Figure 28, Sheila discovers that a television price of $300 in 2000 dropped to below $10 in 2019 (adjusting for resolution, refresh rate, warranty, etc.).

The twins take note of a physical trajectory in Figure 29. They wonder if they might track market movements in like fashion.

Specifically, the market with which they began their analysis, that of business aircraft, has been around for decades. What might we discover? Do they wonder if we look at an emerging market?

They find that the market for mass-produced electric cars began relatively recently and has proliferated. In Figure 30, they depict the direction of the electric car market for a decade, from January 1, 2009, through December 31, 2018.

With a single sale, the market for modern mass-produced electric cars began the same way aircraft did. As shown in the upper left chart in Figure 30, a single sale of the Mitsubishi i-MiEV in July 2009 launched the market, as it was the only entrant in the field that year. As 2009 ended, Mitsubishi sold more of its ground-breaking machine. But by 2012 (center right), many

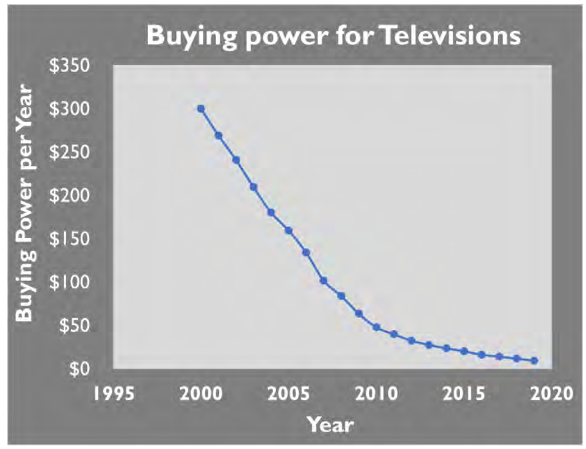


Figure 28 – Television buying power, 2000-2019

Figure 29 – An object placed in motion (here, a soccer ball) stays in motion unless acted up by other forces. Market forces mimic physical forces.

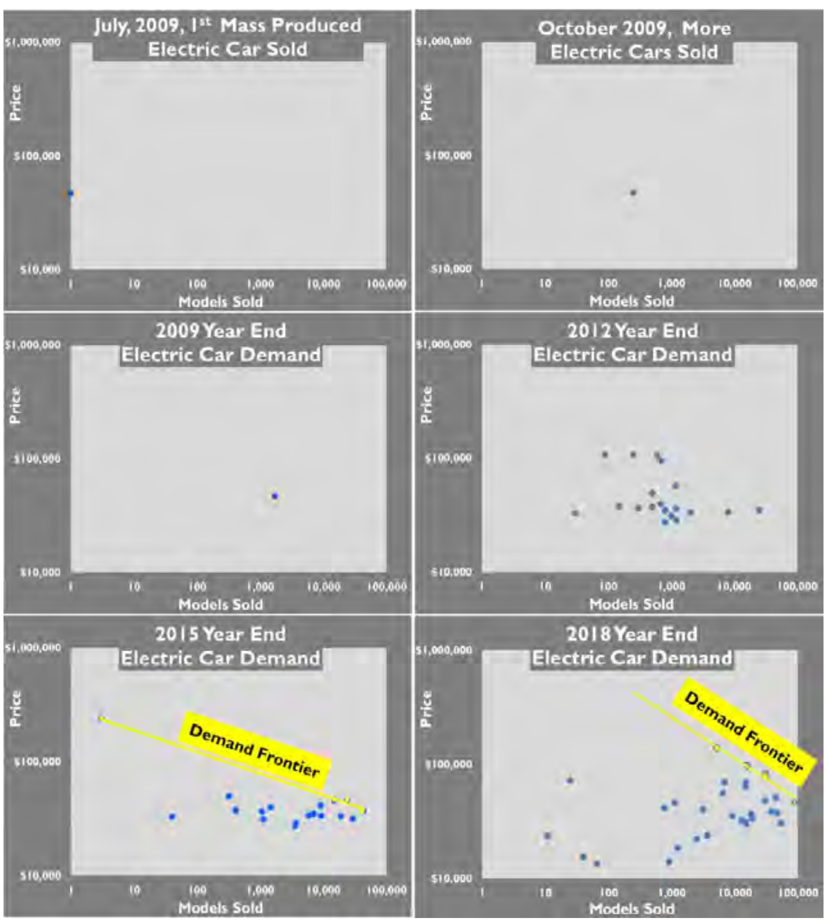


Figure 30 – The electric car market demand changed rapidly, beginning with a single model in 2009 and growing to dozens of market entrants by 2018 when a clear Demand Frontier formed. Here, each point represents the quantity of a model sold (the horizontal component) and its price (the vertical part)

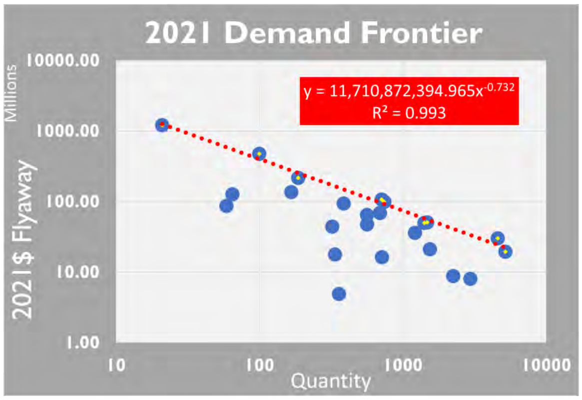


Figure 31 – The United States Government market for bomb-dropping aircraft is very stable. Its Demand Frontier in 1996 (the gray line) changed by about 2% in 25 years as it reached its 2021 Frontier (in red). Not seeing this limit confounded the B-2 program, as the United States Air Force only received 21 of the 132 units it wanted when the US Congress stopped their buy. At its Demand Frontier, this market is effectively at rest.

more models made their way into the market – observe too, with the constant scaling between all six graphs, that Demand had soared. By 2015, this market had organized at its limit into a Demand Frontier, which moved dramatically to the chart in the lower right-hand corner in the next three years.

But do all markets expand so quickly?

In Figure 31, they find that the United States Government has a self-imposed limit on the number of fighters, bombers, and attack aircraft they can buy. The amount they purchased in 1996 (the gray points and line) moved only about 2% in the 25 years up to the 2021 Frontier. Failure to observe this limit led planners to assume more buys of the B-2 bomber than the market would support (the same type of action applied to the F-22, which started with 750 units but settled for 187).

Figure 32 adds time to the seven dimensions we used in Figure 23 for 8. Since half of the right-hand side of the chart uses half of the left-hand side, these analyses are necessarily entangled. It is as if the soccer ball in Figure 28 had only advanced half of its diameter. For the period left at the target price of $120M, the market could support 47 units; five years later, at right, it might absorb 63. The market was going in the right direction for the Aerion AS2.

Crucially, though, the standard deviation of the 47-unit projection in the 2004-2013 projection was more extensive than that for 2009-2018. That meant the chance of Aerion making their targeted quantity of 300 units over ten years fell

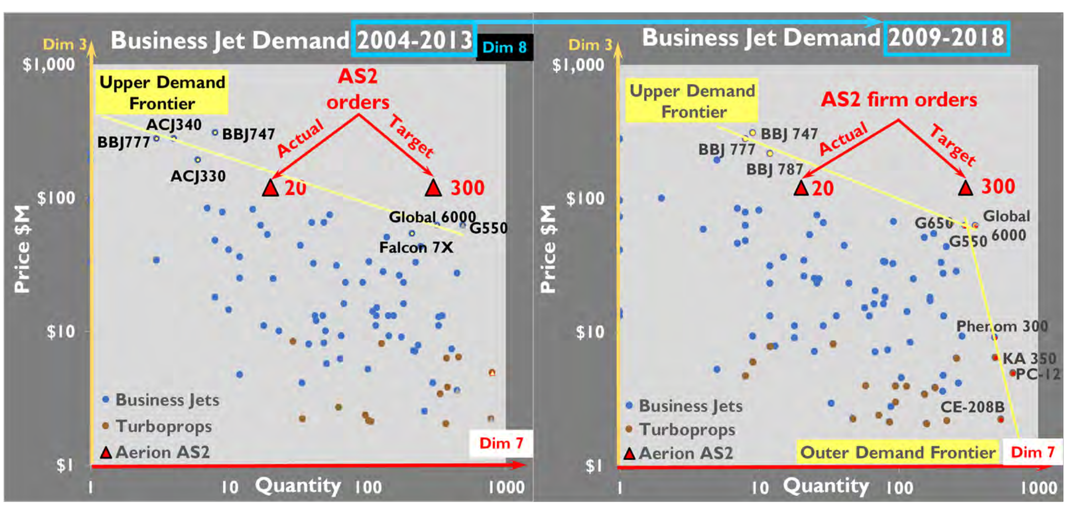


Figure 32 – At right, Dimensions 3 and 7 comprise the Business Aircraft Demand Plane from 2009 to 2018 from Figure 23; other dimensions are removed for clarity). As we add another dimension, Time, to the mix, we go back five years into the market's history to the figure at left while looking at the identical dimensions. Five years before, the market supported 47 models over ten years at $120 million each at the Demand Frontier. Five years later, the market could carry 63 models, indicating slow growth in this market.

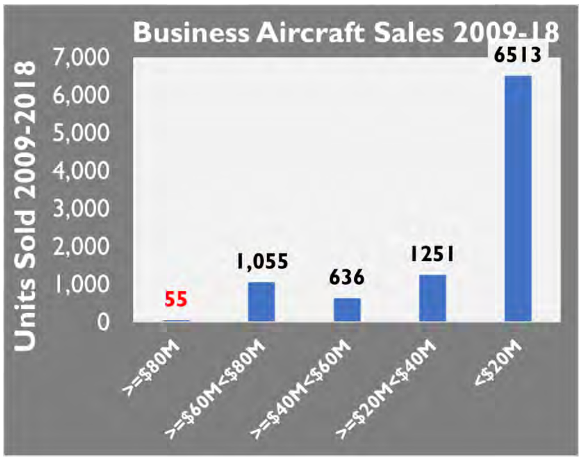


Figure 33 - Aerion wanted to sell 300 AS2s at $120M in a decade, but the market only supported 55 units at $80M or more over ten years

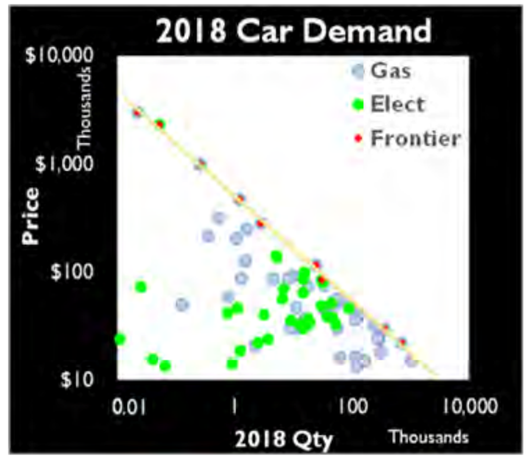


Figure 34: In 2018, gas and electric cars abided by the same Demand Frontier

from one in ten in the left-hand chart to about one in 40 in the latter.

Another way to look at market capacity comes in Figure 33. This chart presents a different view of the ability of the business aircraft market to absorb expensive models.

Some may argue that the AS2 is so unique that it forms its own market with its rules. But, when we compare electric cars to those with internal combustion engines, as shown in Figure 34, we find that both equipment classes abide by the same Demand Frontier.

Figure 34 plots the 2018 quantities sold and prices for all 36 electric car models then in production, compared to 43 gas-powered designs (a fraction of those on offer in 2018). Purposely included in the gasoline group were that year's most famous (the Toyota Corolla) and expensive (the Bugatti Chiron) cars to help discover market limits. Interestingly, several electric and gas vehicles combined to form a relatively flat and highly correlated curve: The Demand Frontier $ = 14.2M* Qty-0.484, Adjusted R2 = 99.8%, MAPE = 6.0%). While this does not comprise the entire market, this study, by design, attempts to model its Demand. What is clear here is that both gas and electric models abide by the same Demand Frontier.

Thus, for these reasons, the Aerion AS2 would not make its targeted quantity of 300 units in a decade and 500 overall. That's what the twins concluded. I did too.

So, in December 2020, I said the vehicle was worth every penny (which came from the Value Analysis), but there weren't enough pennies in the world to make its target (the Twins' Demand Analysis conclusion). They went into receivership in September 2021. Its final tally seems much like that of the Concorde, as Figure 35 reveals.

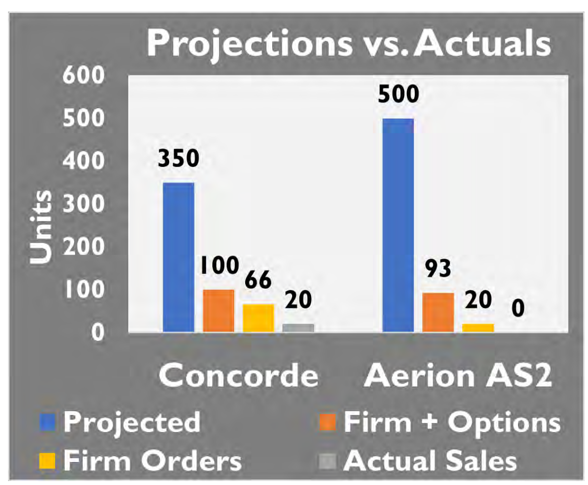


Figure 35: The Aerion AS2's sales history mimicked the Concorde. Both supersonic programs were overly optimistic.

Conclusions

Life is intricate. Market interactions reflect and define the complexities of our economic life. Significantly, however, we demonstrate self-organization in all facets of the economy as makers and buyers of goods and services.

RAND and others have derived statistically significant models which reflect these self-organizing features as they apply to costs. Producers across different companies and industries work in much the same way. While there are variations, in the aggregate, the result of their efforts is a series of cost relationships that become predictable over time. Estimators can assemble broad paradigms that forecast how manufacturers will behave based on their work in the past.

Buyers ultimately set Value, the sustainable prices of products based on their specifications and Demand, and the relationship between quantities sold of products and their prices. At first, a new market such as that for the first airplane or mass-produced electric car will not reveal any organization. But, when those markets gain new models, they form collective Demand limits or Demand Frontiers and reactions to the features offered: Value Response Surfaces. These reactions are often more highly correlated than their corresponding cost equations. The market effectively dismisses goods and services that are too expensive or bids up the prices of too cheap products.

Based on that, the Aerion AS2 had a reasonable development cost estimate and likely had a defensible recurring cost number. It was worth the $120 million they charged, as evidenced by the firm order for 20 units they received at that price. It had two of three critical parameters nailed down.

But Aerion completely misjudged Demand. Often Demand projections for new products take one of two routes.

In the first method, producers poll potential buyers and ask them to lay down the required amount of cash as a down payment if they want to make a purchase. If they agree in principle, that will form part of their basis for the Demand estimate, against which the firm in question would apply some form of discount in the total, perhaps taking away as many as half of those who paid from their projection based on their historical records.

A second way would be to form an operating cost model to flesh out the new system's efficiencies over the old ones, thus providing a method by which they could forecast how many of the latest models the market would want.

As shown here, existing markets *always* reveal what they want and, when it comes to Demand, how many new products they can absorb. Getting the data that enables predictions is time-consuming but typically costs only a tiny fraction of the money a firm can lose by not doing so.

Estimators need to do the same exercises concerning Value – DeLorean thought he could sell his car based on looks. Still, initially, he neglected to put in the requisite horsepower, which cost him his company.

Entangled markets, such as those for jets and jet engines, move in concert. Only by recognizing collective benefits might one firm convince its partner of the usefulness of dropping prices to increase revenue and profits for both parties.

Cost trades across eight or more dimensions occur every day. This paper provides some methods to uncover the intricacies of those details.

It is incumbent upon estimators to address the programs they work on, analyzing their market and key suppliers to enable maximum possible profitability. Analysts should study cost, schedule, and risk, as they've done for decades, but they need to add Value, Demand, and Time effects to gain a broader grasp of their markets.

References

1. Samuelson, Paul A., Economics,9th Edition, McGraw Hill, 1971, p. 63
2. https://www.rba.gov.au/publications/smp/2015/feb/graphs/graph-a2.html
3. https://www.britannica.com/technology/aerospace-industry/History
4. https://en.wikipedia.org/wiki/Glenn_Curtiss
5. Hess, R.W. and Romanoff, H.P, Aircraft Airframe Cost Estimating Relationships: All Mission Types, a RAND NOTE, 1987, p. 4
6. https://en.wikipedia.org/wiki/Aerion_AS2
7. RAND, op.cit., Table 14, p. 88
8. Willcox, Karen, 16.885 Aircraft Systems Engineering Cost Analysis, MIT, slides 16-17, September 19, 2004

Doug Howarth *discovered Hypernomics, the study of market actions across four or more dimensions. He's written 16 peer-reviewed papers on the subject and its implications. His company, Hypernomics Inc., founded on his ideas, has worked for NASA, Virgin Galactic, Lockheed Martin, Raytheon Technologies, and Northrop Grumman. Along with two others in his company, he holds US Patent 10,402,838 for the world's first 4D market analytic software. Wiley will publish his book,* Using Hidden Dimensions to Solve Unseen Problems*, later this year.*

The BS in BoeS:

Oh, the Games That Are Played

Sandy Burney, CCEA®

Prolouge

The young engineer stopped to watch the evening thunderstorm. Outside the rain washed some of the oppressive southern summer air away. Inside the modern office building, the air conditioning chilled the few remaining occupants. The engineer looked back at the computer monitor to finish updating the last of the BoEs. It had been a long three weeks. The engineer reflected upon the memory of the boss coming into the vast corporate office cubical space with this BoE assignment to this southern hinterland office. The boss had told the young engineer not to worry about never having written a BoE before, as it would be a simple stretch assignment. This assignment with its 12-hour days had been anything but simple. The only activity that made this assignment bearable was the nightly trip to the beverage establishment that contained 40 flavors on tap and 300 bottles scattered around the rooms. The engineer finished the last BoE update, turned off the computer, and walked toward the exit, thinking about what flavor to try tonight. Approaching the exit, the engineer passed the normally locked door to a special room. The door was open with bright lights and a lot of activity within. The engineer was glad to not be one of those Pricers working in that special room. The young engineer knew that they would be up all night to incorporate all the BoE changes, just completed, and publish the final cost volume to be delivered tomorrow.

Basis of Estimates (BoEs) Are a Story

So, why does a paper on BoEs start off sounding like a short story? The reason is to emphasize that a good BoE should be crafted like a well-constructed story that leads the reader step by step from the beginning to the end. The story should be logical, without plot twists and math errors, and should lead to a simple conclusion. To emphasize that a BoE should be a story, this paper is purposely constructed in story form, with a prologue, chapters, and an epilogue. As part of a good BoE story, the BoE writer must clearly present an acceptable estimating methodology and its supporting data. A BoE should also be considered a sales brochure that entices the buyer to purchase your product. A **glossy** BoE may have fancy statistics and sound historical data that proclaims a great result, but the buyer still needs to consider the seller's agenda, contained within the BoE. The next section describes the BoE agenda – the BS.

Disclaimer

This paper presents observations, analysis, and opinions of the author only, and does not claim any endorsement or agreement from the author's previous, current, or future employers. This paper considers all mentioned persons performing their job to be hard-working good employees. Any job function that appears to cause conflict should not be considered bad but should be looked at as reflecting differing institutional or organizational incentives. If there were no conflict or differing incentives between the various job functions described within this paper, there would be no games to be played or reason to continue this discussion.

Bias Selectivity (BS)

The lurking sales part of a BoE is the BS in the title of this paper, which is defined as Bias Selectivity. Bias Selectivity is not an official statistical term but is defined in this paper as the systemic favoritism injected into the BoE by the author's choice of such elements as a particular program, comparable to the proposed program, from which historical data can be drawn; the program's period of performance; any complexity factors; the use of engineering judgment; and/or any other estimating methodology to develop the BoE's work scope estimate. In simple terms, the BoE author fits the estimate justification to the author's preferred outcome by selectively choosing the justification inputs. Bias Selectivity is not bad – it is just part of a sales strategy. The buyer, also having a strategy, should be aware of the seller's potential strategies. Hence, when two players – a buyer and a seller – employ strategies, game playing occurs.

This paper will attempt to explain why and where Bias Selectivity appears in most but not all BoEs. A BoE, at its core, is a written explanation of a seller's estimated cost to a buyer. If a buyer selects the seller's proposal, negotiations usually follow, focusing on the costs estimated in the seller's BoEs. Markets where sellers provide BoEs to buyers tend not to be for commodities but rather for complex programs, hence the opportunity for negotiations.

This paper focuses on the unique government market for non-commodity products. In this market, the government is the sole buyer, and it procures products and services under specified rules and regulations. The sellers of products and services in this paper are Original Equipment Manufacturers (OEMs). The paper also looks at the buying and selling of intragovernmental budgets. In all cases the buyer requires the seller to provide some type of BoE that will be used for some type of negotiation between the buyer and seller.

Frequently[1], in the government market a negotiation takes place between the buyer and seller, setting up the conditions for game-playing behavior, as described by John von Neumann (1903–1957). This paper will demonstrate how BoEs are used in the game playing between the buyer and seller. Bias Selectivity in the seller's BoEs plays a role in this game.

Prior BoE Research by the International Cost Estimating and Analysis Association (ICEAA)

In the OEM world BoEs play a significant role, as they are a key input to the OEMs' pricing process for proposals to obtain new government contract work. A search through the ICEAA archives from 2007 through 2021 found only nine presentations and no written papers on the topic of BoEs. A summary of these nine presentations appears below. One presentation, by Frank R. Flett in 2016 (Number 6 below), is relevant to this paper. Mr. Flett's presentation emphasizes impression management for BoEs. He presents two cardinal rules: "Never Make an Evaluator Work!" and "Never Make an Evaluator Think!" Three of the nine presentations argue for using parametric tools in writing BOEs, three presentations provide mythologies for reviewing BoEs, and the three remaining presentations respectively offer a tool to help write a BoE, discuss how to write a schedule BoE, and describe how to use better organization and word choice in a BoE.

[1]Even when a government agency buys off a government catalog, usually the agency creating the catalog has negotiated with the seller for some type of discount. The cases where emergency executive powers are invoked to procure goods and services without negotiations probably border on violations of government procurement rules.

1. "A Basis of Estimate (BOE) Tool for Project Estimates," Bob Fairbairn, James Miller, and Rosemary Baize, 2008. A presentation of a BoE tool developed at NASA for use in formulating and documenting project build-up estimates as part of an effort to improve the quality and documentation of proposals.

2. "Risk Based BOE Analysis – PMAG Approach," Imran Ahmed, David Wang, and Mun Kwon, 2010. A presentation on a top-down approach to analyzing OEM BoEs.

3. "Time Is Money: The Importance and Desired Attributes of Schedule Basis of Estimates," Justin Hornback, 2013. Applying cost BoE properties to schedules.

4. "Analysis of Large O&S Proposal: Lessons Learned!" 2013. A presentation on the process used and the lessons learned by the evaluation team of an OEM proposal.

5. "BOE Development: Scope Evaluation and Criteria," Michael Butterworth and Demetrius Prado, 2014. A presentation focusing on improving the BoE process and the criteria for grading BoEs.

6. "Footprints in the Sand: A Conversational Approach to Basis of Estimate," Frank R. Flett, 2016. A presentation giving tips on writing more persuasive BoEs.

7. "Generating a Semi-Automated 'Similar To' Basis of Estimate from a Complex Parametric Hardware Cost Model for Antennas," Danny Polidi and David Bloom, 2016. A presentation that discusses the development of a "Similar To" BoE generation tool used in conjunction with a complex parametric antenna cost model.

8. "The Journey from 'Bottom-up' to Predictive Modelling BOE," Lori Saleski, 2017. A presentation by an OEM that looks to using a parametric COTS tool for BoEs instead of traditional methodologies.

9. "The Beginning of the End of Traditional Analogous 'Bottom-up' Estimating," Chris Price, 2019. A presentation on the benefits of using a parametric tool for BoEs instead of the traditional approaches.

The Protagonists: Govy, SETA, OEM

Like a good story, this paper has three protagonists: the Government Employee (Govy), the Systems Engineering and Technical Assistance contractor (SETA) and the OEM. These labels are not meant to be pejorative, but are simply a means to model the three key players in the government procurement game into representative categories. The Govy, generally, plays the buyer, with support from the SETA, and the OEM plays the seller role.

Any person employed by a governmental organization falls into the Govy category. SETAs encompass all the traditional SETA companies, Federally Funded Research and Development Centers (FFRDCs), tool providers (such as the esteemed ICEAA sponsors), and consultants. The OEM category contains all companies trying to provide goods and services to the government. Since OEMs create the bulk of the detailed and complex BoEs submitted to the government as a buyer, this paper gives them top billing over companies that are predominantly labor service providers. Although SETAs are service providers, they play a special role in the procurement game that will be further detailed in their own chapter.

Protagonists Are Vectors

This paper uses concepts of game theory, which is a branch of mathematics, throughout. In addition, it needs some enhanced math bona fide, accomplished by establishing the three protagonist categories as multidimensional vectors. In other words, the Govy is not a monolithic worker, but a vector of many different types of workers, denoted by subscripts. The Govy can be expressed as a function of its elements, e.g., Govy = f(Govy$_{DM}$, Govy$_{ME}$, Govy$_{PM}$, …). A list of all the vector definitions will be found at the beginning of each chapter for that vector type.

The paper begins with the Govys and continues with a short chapter on the SETAs, followed by a description of the complex OEM gamesmanship. The epilogue summarizes the discussion and recommends some ways of reducing Bias Selectivity and improving BoE quality.

Let the Games begin!

Chapter 1 – The Govy

Based on the number of BoEs it reads versus the number of BoEs it creates, the government mostly acts as a buyer in the procurement game. However, within the government, a significant amount of buying- and selling-like activity occurs between different governmental organizations. Two categories of goods are being bought and sold: budget requests, and authorizations for procuring Research and Development (R&D), Production, and Operations and Maintenance products and services from OEMs. This chapter will first discuss the intragovernmental BoEs and their game play, and then the Govy's role in reviewing OEM BoEs.

The Govy vectors:

Govy$_{APR}$ – Approvers: this large group includes elected members of Congress, their staffers, the professional committee staffers, and the congressional researchers and auditors.

Govy$_{BFO}$ – Budget/Finance Offices: the offices responsible for creating and managing a government agency's budgets and finances.

Govy$_{CE}$ – Cost Estimator: a government employee, working in a budget or program office, who is educated and trained in the disciplines and methodologies required for cost estimating.

Govy$_{CO}$ – Contract Officer (CO): a government employee who oversees the procurement and execution of government contracts. The CO has the sole authority to award and issue contract modifications.

Govy$_{DM}$ – Decision Maker: any government person who approves budget requests and/or authorizes funds.

Govy$_{ME}$ – Mission Effector: the footwear-on-the-ground Govy who executes a government agency's mission. This includes military soldiers, airmen, and sailors; Social Security claims representatives; and tax auditors.

Govy$_{PM}$ – Program Manager: the person responsible for executing an authorized program.

Budget Formulation and Approval

The most common type of BoE created by a Govy is a budget justification for use in the budget formulation process, which is an annual event for the US government. Typically, budget formulation starts at low levels of an agency, with each level of the organization trying to sell its budget request to the next-higher level. The budget formulation process within the government often includes the gamesmanship of the requesting (seller) organizations asking and justifying requests for sums greater than their needs, knowing that the approvers (buyer) will not budget them for their full request. Since this is a repeatable game, the buyers know that the sellers are asking for more than they need in their budget justification documents. This game gets resolved in the end by collaboration between the Govy$_{BFO}$s and Govy$_{Dm}$s, based on politics, policy, and the Govy$_{BFO}$'s evaluation of all the budget justifications.

How much Bias Selectivity goes into these budget justifications? Somewhere between none and a lot. When an agency requests a budget large enough to cover only its authorized staffing needs, then no bias will be present in its justification. When an agency requests a significantly increased budget over its previous year's budget, then the justification may contain Bias Selectivity or optimistic assumptions. It is up to the Govy_{BFO}s and Govy_{DM}s to decide how much of this increase to include in the final budget formulation.

In the US, after executive agencies have completed the budget formulation process, the government budget goes through the process of approval by the Govy_{APR}s. This process requires a congressionally approved appropriation and agreement by the President. This process involves professional congressional staffers, researchers, auditors, and, of course, the elected members of Congress reviewing submitted budget documentation, analyzing non-budgetary data, and reading polling data to modify the budget submittal to their preferences. During the approval process, negotiations between the Govy_{APR}s and the ultimate Govy_{DM} occur until an agreement is reached.

The budget formulation and approval process can be characterized as a multi-player, repeatable, non-zero-sum game, where there is no equilibrium solution. It is unlikely that negotiations between the Govy_{APR}s and the Govy_{DM} involve disagreements on the routine operational BoEs contained in budget documentation. Disagreements arise over the estimates for large new investments in R&D and expensive production items. A discussion of the BoEs for these expensive items follows in the next sections.

Cost Estimates for Large Budget Items

In the budget formulation process, it is large new investments in R&D and the procurement of expensive production items that particularly attract attention – specifically, as to the validity of their cost estimates. The cost estimates for these expensive items are usually performed by a dedicated staff of trained cost estimators, which are denoted as professional Govy_{CE}. Often for these high-valued cost estimates, the government will procure SETA support to augment their staff of Govy_{CE}s. The professional Govy_{CE} will use available historical data and various analytical techniques to formulate the budgetary cost estimate. These estimates are usually done at high levels of the product's work breakdown structure (WBS), since that is the level of detail contained in their data sources. Interestingly, the Govy_{CE} can perform more accurate cost estimates than their OEM counterparts at this phase of requirement specificity, since the Govy_{CE} has access to cost databases covering multiple OEMs, such as the Cost Assessment Data Enterprise (CADE) database[2]. As part of doing these cost estimates the Govy_{CE} may develop a **should-cost** model for use in negotiating the budget request and for later use during the procurement process after budget approval. However, sometimes the Govy_{CE} may not understand the technical and schedule challenges in their unbiased cost estimate.

The game that is played during this formulation process primarily involves underestimating the true expected costs, since Govy_{DM}s may be concerned that too high a cost will lead to non-acceptance by Govy_{APR}s. Does this mean that there is Bias Selectivity in the BoEs for these cost estimates? Not necessarily, as there are multiple ways to underestimate the costs: assumptions that are too optimistic or pessimistic; immature requirements; and lack of similar-to historical data. These three potential problems are inherent in BoEs for large new investments in R&D, since what is being estimated is mostly just a concept. The James Webb Space Telescope is an example of gamesmanship via underestimating the

[2] For more information see the following website: https://cade.osd.mil/.

technical complexities, as the original cost estimate of $1.6 billion has grown to an actual cost of almost $10 billion today after finally launching.

Procurement Support

After budget approval, the government is allowed to begin the process of buying goods and services. For simplicity this paper uses the term OEM as the seller to the government of goods or services. The vast majority of procurements to the US government are for relatively small dollar values. In government fiscal year 2021, 5,549,307 contracts were awarded to business[3]. Of those awards, 98% were for less than one million dollars, and 92% were for less than one hundred thousand dollars. While some game playing may occur with these small-dollar-value contracts, the interesting analysis and game playing occurs in the higher-value contracts; in fiscal year 2021 there were only 9,126 contracts awarded for greater than 25 million dollars, or 0.16% of the total. Many of these large-value contacts are sole sourced to one OEM. This option occurs when the Govy_{CO} can justify the conclusion that only one OEM can provide that product or service at a reasonable cost. Sole source procurement will be examined from the OEM perspective later in the OEM chapter.

For a high-value competitive procurement, the Govy_{CE} supports the Govy_{PM} and Govy_{CO} in preparing the formal Request for Proposals (RFP) and reviewing the submitted proposals. During the RFP preparation, the Govy_{CE} may include instructions in the RFP as to the level of the WBS the OEMs should use in constructing their BoEs. However, this situation does not occur often. It is during the RFP evaluation phase that the Govy_{CE} provides the most support to the procurement process. The Govy_{CE} gets the pleasure of reading and evaluating the OEM BoEs for reasonableness. Often the Govy_{CE} is supported by a SETA contractor, as discussed in the next chapter.

During the RFP evaluation phase, not much game playing occurs due to the Govy's role in reading and evaluating BoEs for reasonableness. Sometimes during this phase, however, a significant game is played between the Govy_{CO} and the Govy_{PM} in which the Govy_{CE} plays the honest broker. This game takes two different forms, depending on if the procurement is competitive or sole source. In a competitive procurement, this situation occurs when the Govy_{CO} wants to award a contract to a lower-priced proposal that the Govy_{PM} thinks will not meet their needs. In a sole source procurement, this conflict can take the form of not being able to reach a contract agreement during negotiations.

The root cause of this conflict between Govys is a differing of personal incentives. Sometimes the Govy_{CO} is evaluated and promoted based on their ability to obtain the lowest price or negotiate large reductions in price during contract negotiations. The Govy_{PM}, on the other hand, is incentivized to deliver their program capabilities at or under budget. In a competitive evaluation, the Govy_{PM} may prefer a higher-priced proposal over the lowest-price proposal preferred by the Govy_{CO}. The Govy_{PM} may believe the higher-priced proposal carries less execution risk and provides significantly more capabilities for the extra price. Ultimately, this is resolved by the source selection authority, an executive Govy_{DM}. In a sole source negotiation, the Govy_{PM} may want to settle quickly to begin execution due to a crucial need, while the Govy_{CO} may want to continue negotiations to extract more cost concessions from the OEM. Again, an executive Govy_{DM} will make the final decision.

This intragovernmental game can get more complicated when the Govy_{ME}s disagree with what the Govy_{PM} wants to procure for them, meaning that the requirements in the RFP do not satisfy the Govy_{ME}s' needs. Unfortunately, this misalignment of needs happens when the Govy_{PM}

[3] Data from US government web site: https://www.usaspending.gov/search.

is biased by outdated knowledge, budget constraints, or external influences. When the Govy$_{MEs}$ have a strong advocate in the procurement process, their critical needs will be included in the final contract.

In this chapter about the role of the Govy we have observed only modest Bias Selectivity, mostly occurring during budget formulation, and game playing primarily occurs between Govys.

Chapter 2 – The SETA

The role of the SETA requires less elaboration. SETA companies, FFRDCs, and firms providing specialized analysis tools have been formed to be professional independent expert advisors. As an advisor, a SETA has no decision-making responsibilities. SETA analysts may incorporate some Bias Selectivity into the cost estimates they provide. However, this Bias Selectivity becomes the responsibility of the organization that hired the SETA, since the hiring organization is the cost estimate owner. Since there are no differentiating elements within the SETA, the vector contains only one element.

The single element SETA vector:

SETA – A person or organization that provides analysis and engineering services. Since cost estimation is part of the Systems Engineering discipline, companies that specialize in providing cost estimation support to the government are SETAs.

The SETA Role

A SETA primarily supports the Govy in developing requirements, creating RFP solicitations, and evaluating submitted proposals. This support can range from small to significant. The Govy uses SETA advice to develop requirements, prepare RFPs, and/or provide analysis during RFP evaluation. Sometimes a SETA may work for or with an OEM during execution of a contract, but not on a contract that may provide information leading to requirement development for future RFPs. In the past, OEMs had business units that primarily acted like a SETA company. These business units were generally acquired during past market consolidation of government contractors through mergers and acquisitions. Despite the firewalls set up by the large OEMs to prevent organizational conflicts of interest, the government encouraged the OEMs to divest these SETA-like business units. Today, a few legacy SETA-like contracts may exist in the OEMs' portfolios, with the bulk of SETA contracts going to external SETA services companies.

Govy organizations hire SETAs to help perform cost estimation with supporting BoEs that justify budgetary estimates when they have too few resources to do the work themselves. Since SETAs are independent experts, they have little incentive to insert Bias Selectivity into their estimates. Bias Selectivity may, however, be inserted at the direction of the Govy in support of the Govy budget game (see previous chapter on the Govy).

With respect to BoEs, the only game the SETAs play on their own account is the **gotcha** game. SETAs are hired to support the Govy review of OEM BoEs. To demonstrate their value SETAs can be aggressively critical of the OEM BoEs. As shown in the next chapter on OEMs, aggressive criticism of OEM BoEs is easy, like shooting the metaphorical fish in a barrel.

SETAs and Independent Cost Estimates (ICEs)

Sometimes SETAs are hired to perform an ICE by one Govy element to review a different Govy element's cost estimate. One purpose of the ICE review is to find any Bias Selectivity in the cost estimate and its BoEs. If the Govy element creating the cost estimate knows ahead of time that an ICE will be performed, a simple game may be incorporated into the cost estimate. Knowing that SETAs play the **gotcha** game, the Govy creating the cost estimate and BoEs may insert some small (red herring) errors into the estimates, hoping the ICE team focuses on these small errors and not on other more significant cost elements. Sometimes this game works, especially if it is not part of a repeated game between the same two Govy organizations (players). This game is also played by the OEM and will be discussed in the next chapter.

The SETA Acting Like an OEM

Finally, SETAs may act like OEMs when they are bidding on contracts for their services. For simplicity, this paper classifies a SETA as an OEM service provider when it is bidding on a new contract for its services. The SETA behavior in its BoEs for this instance will be discussed in the OEM chapter.

Chapter 3 – The OEM

OEMs produce the bulk of written BoEs in response to government solicitations or RFPs. Also, OEMs are in business to earn a profit. The OEM's profit incentive, the government's rules governing the acquisition process, and the organizational structure of the OEMs all cause game playing and the use of Bias Selectivity in OEM BoEs. It is time to **play ball**.

The OEM vector:

OEM_{CE} – Cost Estimator: an OEM employee who is educated and trained in the disciplines and methods required for cost estimating. The OEM cost estimator may work in the Systems Engineering function, the Estimating and Pricing function, on a specific program, or in an overhead staff organization.

OEM_{DM} – Decision Maker: an OEM person, generally an executive, who has authority to commit the OEM to the terms of a submitted proposal or contract.

OEM_{EP} – Estimating and Pricing: an OEM employee that may perform cost estimation, pricing, or both functions in response to government RFPs.

OEM_{FE} – Functional Estimator: an OEM assigned to estimate a function's work scope in response to an RFP. A function is defined later in this chapter.

OEM_{FM} – Functional Management: OEM executives and managers assigned to lead functional organizations. They have the responsibility to ensure programs are fully staffed and performing to functional standards.

OEM_{PM} – Program Manager: the person responsible for the execution of an OEM program.

Functional Versus. Program Organization

As stated above, the organizational structure of the OEM contributes to BoE game playing and Bias Selectivity. The two leading OEM organizational structures are **functional** and **programmatic**. A functional organization assigns resources (employees) to a specific skill category or function within an OEM. A functional organization is also called a matrix organization. The OEM_{FM} is responsible for staffing various revenue programs and other non-revenue work scope that leads to positive OEM profits. Examples of functional organizations are Systems Engineering, Mechanical Engineering, Software Engineering and Development, and Business Management. Some functions – for example, the Systems Engineering function – have multiple subfunctions such as Logistics, Configuration Management, and Modeling and Simulation. The OEM_{PM} function has a significant characteristic difference from the other functions, since it is responsible for delivering the OEM's key financial metrics, such as profit (or margin), revenues, sales, and awards. In a functional organization the OEM_{PM} does not own most of the resources supporting its operational needs, as these are owned by the OEM_{FM}. Development and manufacturing OEMs generally align in a functional structure.

In a programmatic OEM structure, the OEM_{PM} owns most of the staff supporting its program. In this structure the managers of different functional disciplines report to the Program Manager. Note that support functions such as Business Management, Human Resources, and Business Development still exist in this organizational construct. In either structure the term **owned** means that the management organization has direct authority to hire, terminate, promote, and review its employees. Service provider OEMs tend to align in a programmatic structure.

An OEM must weigh the benefits and costs to choosing either the functional or programmatic operating model. Operational efficiency is the key benefit of a functional organization, since resources can be matrixed. A functional OEM structure happens when the OEM_{FM} can successfully allocate its resources to minimize program costs while delivering a quality product. On the other hand, a programmatic structure benefits from customer intimacy and employee cohesion. Employees tend to start with the program from the beginning and continue with the program either until it ends or until they transfer to another program. In a programmatic organization the program personnel build strong relationships with the customer's staff, gaining valuable insights into their needs and wants for follow-on work during program execution. Since most of the employees work within the OEM_{PM}'s organizational chain, their goals become aligned with the goals of the OEM_{PM}. This alignment of goals and customer intimacy helps the OEM when it attempts to capture additional follow-on and new business with the same customer or in a similar market area.

These diverging costs and benefits present OEMs with a dilemma. On the one hand, operational inefficiency costs drive many manufacturing OEMs away from the programmatic structure. On the other hand, OEMs that are organized by functions incur an extra cost in capturing follow-on and new business, which can be described as a **functional dysfunctionality with respect to capture**. This dysfunctionality, which can be attributed to a misalignment of goals between the OEM_{PM} and the OEM_{FM}s, contributes significantly to poor-quality OEM BoEs that contain Bias Selectivity. To be clear, neither of the OEM_{PM} nor the OEM_{FM}s are villains; it is the OEM's incentive system that drive this observed behavior, which ranges from mild to extreme.

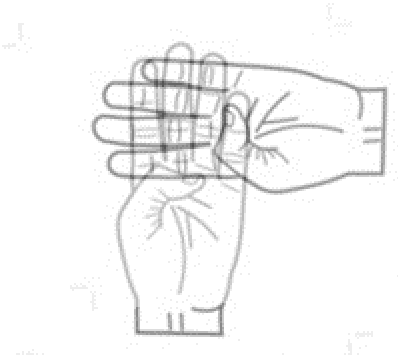


Figure 1. Example of orthogonal forces

To visualize functional dysfunctionality, follow these simple steps as shown in Figure 1. First place both hands in front of your face with your palms facing you. Next, align each hand's fingers upward, with your thumb tucked into your palms. Finally rotate your right hand 90 degrees and crisscross your fingers. The left hand (with the vertical fingers) represents the OEM_{PM}, who has profit, revenue, and award incentives. The right hand (with horizontal fingers that are orthogonal[4] to the left hand) represents the OEM_{FM}s, who are incentivized to deliver program execution under budget. The OEM_{PM} wants to create an affordable proposal to the customer by keeping costs low to maximize awards or new business. The OEM_{FM}s want the proposal to include costs as high as possible for their functional area, so they can deliver a large underrun on their execution budget. How this orthogonality of forces affects BoEs and causes game playing will be explored in the following sections.

[4] Orthogonal is a mathematical term for perpendicular in multidimensional space.

Sole Source BoEs

In this section, assume the OEM is aligned functionally. In the government acquisition process, there are two distinct types of RFPs: sole source or competitive. A few of the rules applied by the Federal Acquisition Regulation (FAR) for submitting sole source proposals differ from those that apply to competitive proposals. In general, however, the process that OEMs use for responding to the RFPs do not significantly differ between the two proposal types. This section will explore the OEM process for developing sole source BoEs.

FAR Section 15.403-4 requires OEMs to provide certified cost and pricing data in responding to high-dollar-value sole source RFPs. Far Section 31.205-7 on contingencies forbids OEMs to propose management reserve (MR) costs for most cost objectives in their proposals. The only exception is specified in FAR Section 31.205-7(c)(1)[5]. This rule allows OEMs to bid quantifiably objective MR. An example would be an allowance for material scrap, which can be calculated from historical manufacturing costs. The inability to include MR for risk mitigation in proposals drives OEMs to include Bias Selectivity in their BoEs.

WBS #	Level 1	Level 2	Level 3	Level 4
1	Hypersonic Widget			
1.1		Widget		
1.1.1			Subsystem W.1	
1.1.1.1				Component 1.1
....				
1.1.1.8				Component 1.8
1.1.2			Subsystem W.2	
1.1.2.1				Component 2.1
....				
1.1.2.6				Component 2.6
1.2		Propulsion		
1.2.1			Subsystem P.1	
1.2.1.1				Component 3.1
....				
1.2.1.7				Component 3.7
1.2.2			Subsystem P.2	
1.2.2.1				Component 4.1
....				
1.2.2.9				Component 4.9
1.3		Systems Engineering		
1.4		Program Management		
1.5		System Test & Evaluation		
1.6		Support Equipment		

Figure 2. Example of hypersonic widget WBS with 2 subsystems and 30 components

For the most part, OEM BoEs, using historical functional data, are developed by OEM_{FE}s that report to their functional organization, and approved by their OEM_{FM}s. It should be emphasized that the OEM_{FE}s try to find relevant similar historical data on which to base their estimate but sometimes fail. Several reasons contribute to the failure to use historical data in BoEs: (1) no data may be available from a similar project if the OEM is attempting to get into a new market; (2) the OEM_{FE} may take the easy path by asserting engineering judgment; (3) the OEM_{FE} may attempt to estimate costs in the proposed WBS at too low a level when all the historical data was collected at a higher WBS level[6], and (4) a combination of the first three cases. These four reasons can be attributed to the OEM_{FE} estimating at too low a WBS level, since it may take many levels of WBS indenture (decomposition) to get to individual functional work scope.

[5] FAR Section 31.205-7(c)(1) defines allowable MR costs as "[t]hose that may arise from presently known and existing conditions, the effects of which are foreseeable within reasonable limits of accuracy, e.g., anticipated costs of rejects and defective work. Contingencies of this category are to be included in the estimates of future costs so as to provide the best estimate of performance cost."

[6] If the OEM_{FE} provides too many estimates at a WBS level that is lower than the OEM collected historical costs, then the OEM could have a compliance violation under Cost Accounting Standards 401.

To help clarify how OEMs estimate costs for a proposal, this section will use an example of an OEM bidding on the development of a hypersonic widget using a WBS, detailed in Figure 2, that contains 2 subsystems and 30 components. In this example, if the OEM estimated costs at the lowest indentured level there would be 34 BoEs – 30 components at Level 4 plus four Level 2 WBSs. However, a functionally aligned OEM will add at least one more WBS level in Figure 2 for each function that contributes to each of the 34 lowest indentured WBS lines. If each of these 34 WBS lines covered, on average, three different functions, then there would be 102 functional BoEs. The number of functional level BoEs can grow substantially as the complexity of the WBS grows.

The Figure 2 example WBS is based upon the United States Department of Defense *Standard Practice for Work Breakdown Structures*, also known as MIL-STD 881E[7]. The OEM_{FE}s face the challenge that the historical data available to them may not have been collected using a standard WBS such as the one in Figure 2. When historical data the OEM_{FE} selects to use does not exactly align with what they are estimating, the OEM_{FE} must normalize the historical data, which involves making assumptions and choices. This normalization process is part of the **selectivity** in Bias Selectivity.

The bias in the OEM_{FE} estimate comes from the game played between the OEM_{FM} and the OEM_{PM}. The strategy or goal of the OEM_{FM} is to have an estimate large enough to be confident that the work scope can be executed for less than what was bid, meaning that the estimate should include some MR. For the OEM_{FE} to get estimate approval from the OEM_{FM} (their boss), they must find an estimate basis with some MR in it. The discussion above showed that functional estimating occurs at low levels of the WBS, so when all the functional estimates are added together, the MR from Bias Selectivity might be larger than necessary.

The Govys often criticize the OEMs for providing poor quality BoEs, giving several reasons for their observations. One criticism relates to the number of engineering judgment BoEs. The use of the engineering judgement BoE methodology mostly occurs when the OEM_{FE} estimates a WBS element where the work scope is for a small number of hours, often as a result of estimating at too low a WBS level. An example of this would be a BoE for 26 hours that had a rationale for attending a one-hour weekly meeting with the customer over a six-month period of performance[8]. The next major contributor to poor quality OEM BoEs comes from the inexperience of the OEM_{FE}. The OEM_{FE} often is the most junior employee in the functional organization, since the more senior members get to prioritize their other functional activities over BoE writing. The senior functional members may provide suggestions to the OEM_{FE} on how to insert Bias Selectivity in the BoEs. The final common contributor to poor quality OEM BoEs is a lack of historical data. The absence of historical data mostly occurs when the OEM is bidding on developing new technology, or on applying capabilities the OEM has not supplied in the past. In this case, the Govy might have better historical data to use for estimating than the OEM, as the Govy has access to cost data from all OEMs in a single database, such as the Cost Assessment Data Enterprise database previously mentioned.

After the OEM_{PM} receives the total cost from the OEM_{EP}, the functional estimates will be reviewed. In the sole source case, if the OEM_{PM} feels that this total cost is so high that it will tarnish the relationship with the customer, the OEM_{PM} will push back against the estimates to get the OEM_{FM} to agree to present lower ones. If the OEM_{PM} feels

[7] MIL-STD-881E, Work Breakdown Structures for Defense Materiel Items, 6 October 2020.

[8] Unfortunately, the author of this paper has seen way too many of these BoEs, and they are most likely a Cost Accounting Standards 401 violation.

that the total cost is high but acceptable, they will choose to submit it as is, knowing that total costs will be reduced during the negotiation game with the Govy. Sometimes the OEM_{PM} arranges for an OEM_{CE} who is independent from the functions to assess the functional estimates. It may also be possible for the OEM_{PM} to use the OEM_{CE}s to estimate most or all the BoEs instead of using OEM_{FE}s. This case can occur under certain conditions, i.e., if enough OEM_{CE} resources and time are available, if OEM policies and culture allow for non-OEM_{FE} estimating, or if the OEM_{DM} overrides the OEM_{FM}'s objections.

After a sole source proposal is submitted, the negotiation game between the OEM and the Govy can take several forms. Often negotiations start with the OEM and Govy bargaining over every BoE. If the OEM has submitted a big stack of BoEs, then negotiations can take a long time. If time and negotiating energy begin to run low, the OEM and Govy will try to find a methodology applicable to a higher cost level to finish negotiations. Finally, if negotiations get to an impasse at the working level, an OEM executive and a Govy executive will negotiate an agreement on a few top-line values.

Here are some of the gaming strategies the OEM may use, knowing that negotiations will commence with the Govy. If the Govy has limited time and resources, the OEM may choose to estimate at low levels of the WBS to overwhelm the Govy with BoEs. If this is a repeatable game, e.g., the first of many lot purchases, then the Govy may counter this strategy by specifying what level of the WBS they want to see in BoEs. If the OEM knows that the Govy and/or the Govy's SETA like playing the **gotcha** game as described in the SETA chapter, then the OEM may purposely insert errors or glaring overestimates in some BoEs that the OEM can knowingly sacrifice during negotiations. As stated in the Govy chapter, the

Figure 3. Example of competition aligning forces

OEM may be caught in the middle of a game between the $Govy_{PM}$ and $Govy_{CO}$. When this game becomes obvious, the OEM strategy is to try to raise the negotiations to the executive level as quickly as possible.

Competitive BoEs

Initially, for competitive RFPs functionally aligned OEMs tend to use an estimating process similar to the one they use on sole source RFPs. However, in a competitive environment there is another force that can drive OEMs to produce better BoEs. To illustrate this, reimagine from the section above (Figure 1) on functional organization that your hands and fingers are aligned in an orthogonal orientation. Then imagine an invisible pair of hands cupping over your hands and squeezing your fingers into a single ball. This forces your fingers (or vectors) to align more closely, as shown in Figure 3. This competitive force (cloud-like) is akin to Adam Smith's invisible hand[9].

Figure 4 is an actual picture of cost proposals submitted by three different large manufacturing OEMs[10] on a competitive bid for a US Navy radar program. All three OEMs are functionally aligned and use a combination of OEM_{CE}s and OEM_{FE}s for their large RFPs. Divergence from sole source estimating begins with the elimination of the cost

[9] Smith, Adam, An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. 2, 1778 (2nd ed.).

[10] The names of the OEMs have been obscured to protect the dignity of the Big Stack OEM.

Figure 4. Three cost proposals from large OEMs

and pricing data requirement and the elimination of negotiations on proposed costs[11]. These two changes from sole source contracting encourage the OEMs to write BoEs at higher levels of the WBS.

Assuming the three OEMs in Figure 4 used a WBS similar to the one in Figure 2, the picture in Figure 4 implies that the "big stack" OEM estimated its BoEs at Level 5 or lower of the WBS, while the two "little stack" OEMs estimated their BoEs at higher WBS levels. The culture within each OEM defines whether the orthogonal functional force is counteracted by the competitive force, resulting in a proposal that is integrated across functions (little stacks), or the functional force is stronger than the competitive force, in which case the functional sole source process (big stack) dominates. The observed behavior in Figure 4 shows that the "big stack" OEM allowed the OEM_{FM} review to keep the additive MR embedded in the numerous low-level BoEs. For the "little stack" OEMs in Figure 4, the competitive force appears to have been stronger than the functional force, resulting in fewer BoEs. Having fewer BoEs does not mean that the "little stack" OEMs do not have piles of draft backup BoEs that are equal in size to the "big stack" OEM, but rather they chose a methodology that supplied them with less BoEs to submit.

Some Bias Selectivity may appear in competitive BoEs, as the OEMs still need to include some level of MR in their RFP response. Since competitive BoEs are evaluated for reasonableness, the Govy and/or SETA review of the BoEs will not be scrutinized as much as sole source BoEs. It is assumed that competition will drive down costs. However, sometimes an OEM will try to play the Engineering Change Proposal game. This game occurs when an OEM sends in an RFP at below expected costs (**buying in**) and tries to recoup costs and gain additional fees by submitting Engineering Change Proposals to the original contract. The Govy sometimes counteracts this game in their evaluation of cost proposals by risk -adjusting any proposals that seem to understate the Govy's cost assessment.

The OEM Executability Review

Some OEMs require an executability review by executives before they submit a proposal that commits them to contractually binding terms and conditions. To be effective, the review must be carried out by OEM employees who are independent from the proposal process and outcome. This review has the independent evaluators assess the likelihood that the BoEs and other costs, such as the Bill of Material, can be executed at a threshold profit level. This type of review can lead to a game – similar to the one that is played between the OEM_{PM} and OEM_{FM} – where the independent reviewers may suggest adding costs to the proposal to inflate the expected profits. In this case, the executives have an incentive not to be part of a review team for a program that has poor execution financials. The review teams suggest cost increases more often for sole source proposals than for competitive proposals, as the force of competition will again push back on suggested cost increases.

[11] Sometimes on competitive awards the government will negotiate costs with the awardee over minor scope changes and ask for a lower fee percentage.

Epilogue

Summary

As our BoE story comes to a close, the three protagonists – the Govy, the SETA, and the OEM – have been actively writing, reading, and analyzing BoEs, while playing strategic games that further their self-interest. As part of the strategic game playing, Bias Selectivity is incorporated into BoEs. Bias Selectivity is neither good nor bad; it is just a tool used in a sales document to further the seller's objectives. Buyers know that the BoEs they read may contain some Bias Selectivity. Buyers must decide how much to accept based on their own analysis. It is in this context that the game between sellers and buyers commences.

This paper provides some additional observations about the differences between Govy and OEM cost estimating. The Govy estimates costs during the early concept and design phase of a program. These cost estimates can vary widely, as there are many unknowns. Rarely will a Govy suffer consequences for a cost estimate that was significantly wrong. The OEMs mostly estimate costs in response to an RFP[12] for a specified product or service. These estimates vary less, since they happen later in the program lifecycle. If an OEM cost estimate that becomes contractually binding ends up costing the OEM money because the estimate was too low, some of the OEM vector elements may lose their jobs. To end on a more positive note, the next section suggests solutions for improvement.

How to Reduce Bias Selectivity and Improve BoE Quality

This paper has highlighted the problem of Bias Selectivity and the games played in writing and reviewing BoEs. This story will end with some ideas on reducing Bias Selectivity and improving BoE quality. It should be noted that Bias Selectivity cannot be fully eliminated from BoEs, since there will always be game playing when noncommodity products and services are bought from a seller trying to maximize profits. These four suggestions, however, can improve BoE quality: (1) use professional cost estimators; (2) encourage estimating at higher levels of the WBS; (3) change the FAR to allow contingency or MR to be included in proposals; and (4) reduce OEM functional oversight of BoE inputs.

Use Professional Cost Estimators

Generally, the government and SETAs hire and train employees who perform cost estimation as their primary responsibility, and thus are considered professionals. Since there are few academic programs specifically designed to educate and train cost estimators, most estimators learn on the job, with some supplemental training. Some government organizations and many SETAs strongly encourage their employees to obtain cost estimating certification.

In the OEM world, there are far fewer professional cost estimators relative to the number of cost estimates produced. OEMs produce different internal and external cost estimates across their multiple functional organizations. Resource constraints prevent the OEMs from adequately training all the personnel involved in all the high- and low-level cost estimates produced. The professional cost estimator deficiency becomes evident in the continual use of junior staff to write BoEs. A few large OEMs have solved this problem by installing a centralized cost estimating function, while the remainder rely on inexperienced functional estimators. OEMs would benefit from solving the resource misalignment problem by adding more professional cost estimators. The government

[12] Sometimes OEMs provide a Rough Order of Magnitude (ROM) cost to the Govy in support of their budget planning process.

could encourage OEMs to use professional cost estimators by including inducements to their use in RFP solicitations.

Encourage Higher-level Estimation

Estimating at higher levels of the WBS has multiple benefits, discussed above, and is an easy fix for the government to achieve. In final RFP packages, the government can provide a WBS and specify at what level they want to see BoEs. With proper planning and discussions with all potential bidders, the government can develop a WBS structure and estimation level that represent a reasonable compromise across the potential bidders. If the government were to include these specifications in the RFP instructions, then large discrepancies in the number of cost volume pages submitted, as seen in Figure 2 above, would become far less likely.

Include Contingency or MR in Cost Proposals

Changing the FAR to allow contractors to bid contingency in their cost proposals would be a big change. This change would reduce the pressure on OEMs to include Bias Selectivity in as many BoEs as possible to allow for MR in execution. In fact, some non-US countries allow MR to be bid in their cost proposals. One simple way to make this change would be to align the contingency costs with the Risk Register. Cost estimates and their associated BoEs could be created for each major risk identified in the proposal. These costs could be included as contract options in the signed contract value, but not executed unless the risk occurs. Once the risk occurs, the Govy_{CO} could execute the risk option value to the contract without having to negotiate an Engineering Change Proposal.

Reduce Functional Oversight

This last improvement can happen only if the OEM executive leadership implements a top-down cultural change to reduce the functional oversight of BoEs. This paper is not suggesting that functional input and review be eliminated, but that functional personnel work collaboratively with non-functional estimators to provide less biased BoEs.

The Last Game

Readers who are aficionados of game theory must be wondering why the most famous game has not been mentioned. This game is probably played out daily on cable TV on one of the many crime shows running 24 hours a day. Since this game, the Prisoner's Dilemma, involves two criminal suspects in two separate interrogation rooms, it is not applicable to BoE writers and this paper. It is unimaginable and unfathomable that honest, hardworking cost estimators could ever face the Prisoner's Dilemma game. However, if, as a cost estimator, you ever find yourself in an interrogation room about some cost estimate you provided, this author's advice; - Take the deal and blame your boss.

List of Acronyms:

APR	Approver
BFO	Budget/Finance Office
BoE	Basis of Estimate
BS	Bias Selectivity
CE	Cost Estimator
CEO	Chief Executive Officer
CO	Contract Officer
DM	Decision Maker
ECP	Engineering Change Proposal
EP	Estimating and Pricing
FAR	Federal Acquisition Regulation
FE	Functional Estimator
FM	Functional Manager
FFRDC	Federally Funded Research and Development Center
Govy	Government Employee
ICE	Independent Cost Estimate
ICEAA	International Cost Estimating and Analysis Association
ME	Mission Effector
MR	Management Reserve
O&M	Operations and Maintenance
OCI	Organizational Conflict of Interest
OEM	Original Equipment Manufacturer
PM	Program Manager
R&D	Research and Development
RFP	Request for Proposal
SETA	Systems Engineering and Technical Assistance contractor
US	United States
WBS	Work Breakdown Structure

Sandy Burney provides cost strategy consulting for a large OEM. In his 30+ year career, Sandy has provided cost estimating analyses as a Government, SETA, and OEM employee. Across these three types of employers, Sandy has performed forecasting, cost estimation, and modeling within specific cost groups, system engineering, capture, and pricing organizations. From this experience he has developed a unique perspective on capture incentives to the Industrial base.

Simplifying Software Sizing with Simple Function Points

Carol Dekkers
Dan French

Abstract: Professional software cost estimators recognize that one of the most elusive, yet fundamental components of parametric software cost estimation is the size of the software under development. While many methods have been proposed over the years to quantify software size, none has been as stable or independent of changing technologies as functional size measurement (FSM), first introduced at IBM in the late 1970's. FSM and its unit of measure, function points, derives software size based on a standardized assessment of its functional requirements. Today, the most popular and globally accepted FSM approach is the International Function Point Users Group (IFPUG) Function Point Analysis (FPA) version 4.3.1. In October 2021, the IFPUG released a new and standardized approach called Simple Function Points (SFP) version 2.1, based on an IFPUG 4.3.1 compatible approach developed by Dr. Roberto Meli of Italy in 2010.

This paper introduces the SFP methodology, demonstrates its use, and highlights the challenges and opportunities for software cost estimators who need to estimate software size from high level software requirements. We will also explore the key differences between SFP and traditional IFPUG FP, including guidance for cost estimators using Function Point measures as the basis for their software cost estimates .

Introduction and brief history of IFPUG function points:

The IFPUG function point analysis methodology was developed by IBM in the 1970's in response to customer concerns that newer, more efficient software languages (such C, SQL and Pascal) resulted in a smaller volume of computer code (quantified at the time by the number of Source Lines of Code or SLOC) and thus, appeared to be of less "value" to their customers. With the advent of higher-level languages, developers found they were increasingly experiencing cost and schedule overruns for software projects based on SLOC and sought to find a better means of assessing software size, independent of development technology.

To address this issue, IBM assembled a team of software engineers, led by Allan Albrecht, with the goal of developing an alternative software size measure, agnostic of programming language and platform. The first iteration of "Function Points" was formally presented in Albrecht's paper "Measuring Application Development Productivity," at an IBM Guide/Share conference. The industry response was so positive that the rest, as they say, is history.

In 1984, the International Function Point Users Group (IFPUG) was founded as the not-for-profit custodians of Function Point sizing methodology and the first IFPUG Function Point Counting Practices Manual (CPM) version 1.0 followed in 1986.

The IFPUG Function Point methodology has slowly evolved, and became standardized, over the years; but the original Albrecht-based components and rules still apply.

Following the 1998 publication of the International Standardization Organization/ International Electrotechnical Commission (ISO/ IEC) functional size measurement framework standard ISO/IEC 14143-1: Concepts of Functional Size Measurement (FSM), et al, IFPUG's Function Point Analysis method became the first ISO/IEC standardized Functional Size

Measurement Method: ISO/IEC 20926, of which the current instantiation is known as ISO/IEC 20926: IFPUG Functional Size Measurement Method version 4.3.1.

Over the years, the function point methodology has matured and is now codified into the ISO/IEC standard (30 pages) supported by a formal counting practices manual (CPM) with several hundred pages of terms, application guidelines, and examples of practical implementation FP counts.

Today, Functional Size Measurement (FSM) is well accepted (by the International Cost Estimating and Analysis Association (ICEAA) and other leading software cost estimating experts within the US government and internationally), that software size is a main driver of software development cost and schedule. Additionally, as more organizations cope with tighter Information Technology (IT) budgets, coupled with increases in project overruns and failures, there is a major need to develop better, fact-based, and reliable software estimates. While IFPUG FPA holds promise to revolutionize the software cost estimating industry due to its technology-independent approach to software sizing, there are a number of barriers to widespread adoption. These include the investment of time and resources to properly train analysts to implement function point-based estimating, and the challenge of applying function point counting rules to early requirements, (which is all that is available when the cost and schedule estimates are needed.)

The emergence of Simple Function Points (SFP)

For some of the reasons stated above, as well as European market demands for functional-size-based estimates from early requirements documents, a group of Italian researchers, led by Dr. Roberto Meli, set out to develop a simplified approach to functional size measurement, specifically designed to work with high-level software requirements. In 2009, Meli et all debuted their Early and Quick Function Point method (E&Q FP), based on the IFPUG method. E&Q FP replaced the detailed IFPUG FP steps of function identification and complexity with a more generic and simplified FP scoring system thereby reducing the dependence on detailed software requirements (such as the number of data fields or files involved in countable components) and enabled quicker functional size estimates. E&Q FP also allowed analysts to apply the traditional IFPUG formal counting rules when such details were available.

E&Q FP methodology eventually evolved into the Simple Function Point method (SiFP) in 2010 and was subsequently acquired by the IFPUG in 2019. In October 2021, the IFPUG standardized the Simple Function Point terminology and formally released IFPUG Simple Function Points (SFP) version 2.1.

SFP simplifies the traditional IFPUG FP method by simplifying the functional size measurement process to the assessment of two IFPUG-compatible base functional components: Elementary Processes (EP) and Data Groups (DG), each with a single function point value: 4.6 FP for EPs and 7 for DGs. As such, SFP eliminates the traditional IFPUG FP steps of determining the primary intent, identify and categorize five distinct types of functions, and the subsequent step of categorizing them based on their relative complexity (low, average, or high) before assigning function point values.

Currently, the authors are participating in the full rollout and development of formal IFPUG SFP training and an accompanying SFP-based certification program.

Functional Size Concepts and Terminology

Both IFPUG FP and IFPUG FP use the same concepts and definitions pertaining to functional size measurement and functional size. This section provides an overview of salient terms for readers unfamiliar with this software sizing approach.

There are a few key terms and definitions applicable when discussing IFPUG FP and IFPUG SFP. Note: All terms and definitions are in accordance with the IFPUG Counting Practices Manual (CPM) version 4.3.1 and the Simple Function Point (SFP) Manual version 2.1. Those taken directly from the official IFPUG documents are included in italics below.

According to the IFPUG CPM, **functional size** is the *"measure of the functionality that a <software> application provides to the user, determined by the application function point count."* (IFPUG, 2010). Functionality or functions, in turn, are the user specified functions or business practices and procedures that the software performs, as specified by the **Functional User Requirements (FUR).**

Functional User Requirements (FUR) *- A sub-set of the user requirements; requirements that describe what the software shall do, in terms of tasks and services.* **FUR** are those requirements that describe what the software will do: for example, what data to store, what reports to produce, which data to display, what information to send to other systems, to name a few.

Functional size measurement (FSM) is a methodological approach to determining the Functional Size from evaluating a software's FUR and assigning a specified number of function points to each.

Note that FUR is distinct from, and should not be mistaken for, other types of software requirements: technical, quality, or **non-functional requirements (NFR)**, that describe other aspects of the software including how the software must perform (NFR), the quality of the software (also NFR), the development environment (technical) or the programming language. A few further examples of software requirements that are NOT functional requirements include: the hardware or hosting platform(s), quality requirements, response time (to meet service level agreements), data capacity, industry or organizational standards and policies, and processing loads. Many of these requirements can be measured using a different methodology and units of measure, such as the IFPUG Software Non-Functional Assessment Process (SNAP) and associated SNAP points.

(Application or software) ***Boundary*** *- The boundary is a conceptual interface between the software under study and its users.*

User *- A user is any person or thing that communicates or interacts with the software at any time.* A user could be a physical person, other software or hardware, or anything that sends or receives data that crosses the software's application boundary.

Elementary process (EP) - *"An Elementary Process is the smallest unit of activity, which is meaningful to the user, that constitutes a complete transaction, it is self-contained and leaves the business of the application being measured in a consistent state". The term* ***transaction*** *here does not mean a physical collection of software instructions grouped according to a technical criterium (a Non-Functional Requirement). An elementary process is, instead, a logical aggregation of processing steps which is meaningful from a logical user perspective, and it is fulfilling a Functional Requirement.*

Logical file (LF) "*A Logical File represents functionality provided to the user to meet internal and external data storage requirements. It is a user recognizable group of logically related data or control information maintained and/or referred within the boundary of the application being measured." The term file here does not mean physical file or table. In this case, file refers to a logically related group of data and not the physical implementation of those groups of data.*

Additionally for the formal IFPUG FP methodology, the following definitions apply:

*An **internal logical file (ILF)** is a user recognizable group of logically related data or control information maintained within the boundary of the application being measured. The primary intent of an ILF is to hold data maintained through one or more elementary processes of the application being measured.*

*An **external interface file (EIF)** is a user recognizable group of logically related data or control information, which is referenced by the application being measured, but which is maintained within the boundary of another application. The primary intent of an EIF is to hold data referenced through one or more elementary processes within the boundary of the application measured. This means an EIF counted for an application must be in an ILF in another application.*

*An **external input (EI)** is an elementary process that processes Data or control information sent from outside the boundary. The primary intent of an EI is to maintain one or more ILFs and/or to alter the behavior of the system.*

*An **external output (EO)** is an elementary process that sends data or control information outside the application's boundary and includes additional processing beyond that of an external inquiry. The primary intent of an external output is to present information to a user through processing logic other than or in addition to the retrieval of data or control information. The processing logic must contain at least one mathematical formula or calculation, create derived data, maintain one or more ILFs, and/or alter the behavior of the system.*

*An **external inquiry (EQ)** is an elementary process that sends data or control information outside the boundary. The primary intent of an external inquiry is to present information to a user through the retrieval of data or control information. The processing logic contains no mathematical formula or calculation and creates no derived data. No ILF is maintained during the processing, nor is the behavior of the system altered.*

***Data Element Type (DET)** - A unique, user recognizable, non-repeated attribute.*

***File Type Referenced (FTR)** - A data function read and/or maintained by a transactional function.*

***Record Element Type (RET)** - A user recognizable sub-group of data element types within a data function*

Evolution of Simple Function Points (SFP)

With the initial introduction of function points in the mid-and late1980's, many software development organizations, who had been struggling with delivering high fidelity software estimates and metrics using SLOC, were quick to adopt the new approach to software sizing based on the functionality provided to its users. There were adjustments made to the methodology in the 1990's and early 2000's resulting in new versions of the counting practices manual to address issues and concerns users had in the application of the rules. Additionally, the implementation of non-mainframe software platforms provided challenges to applying the rules, in particular interpretation of the application boundary. However, with the release of version 4.1 of the CPM in 1996 the rule set was stabilized.

While IFPUG worked to address shortfalls in the process, there were still challenges with developing function point counts when the requirements were not detailed and the ability to identify key components such as DETS or FTRs was not possible. There were also claims (false) that the use of function points was not possible until detailed design requirements were available, function points could not be counted until the software was in production, or that certain platforms, application types, or some software development methodologies could not be counted.

The assertions, mostly incorrect, did demand that there be a way to address the concerns around lack of details needed to properly identify and classify functions, particularly in the early phase of software development. IFPUG and others promulgated differing approaches which primarily consisted of assuming the average complexity for all functions.

In 2007, Data Processing Organization (DPO) in Italy introduced the more refined concept of Early and Quick Function Points (E&Q FP). Based on the IFPUG methodology, they replaced the "assume average complicity" concept with a more refined approach. While IFPUG-based, the need for detailed requirements was not necessary, the analyst is still required to have a reasonable amount of detail in the requirements to properly discern which of the various levels of aggregation is appropriate as well as which functional type it is, which correlates highly with the IFPUG functions.

The range of the function sizes used are determined based on the aggregation level employed. For 1st aggregation level it uses the traditional function point sizes and ranges from 3 FP to 15 FP. The 2nd aggregation level ranges from 4.0 to 8.1 FP, the 3rd aggregation level from 14.1 to 101.8 FP, and the 4th aggregation level spans 111.5 to 617.4 FP. See appendix A for E&Q FP aggregation levels and sizing table.

While the E&Q FP approach provided a mechanism for counting function points based on the level of detail of the functional user requirements, including a way to count FP where the FURs were at a high level, some function point analysts still had difficulty with determining how to count to the appropriate level of size, aggregation, and determining the appropriate sizes for Typical Processes (TP), General Processes (GP), General Data Groups (GDG) and Macro Processes (MP). Given the ranges of these functions, misclassification could still lead analysts to over- or under-counting the software size.

This led Dr. Meli to further refine and simplify the methodology and develop Simple Function Points (SFP). The method approximates the IFPUG function point methodology but does not require the identification of DETS, RETS, or FTRs and consists of only two types of functions: Elementary Process (EP) replacing EI, EO, and EQ and Logical File (LF) replacing ILF and EIF.

IFPUG FP and IFPUG SFP – Similarities and Differences

The SFP concept embraces the same concepts and definitions that the traditional IFPUG method does with regards to the definition of boundary, functional and technical requirements, maintenance, enhancement, user, logical file, and elementary process but removes the need for the analyst to identify and classify the different transactional and data function types into EI, EO, EQ, ILF and EIF.

Rather, in SFP, functional user requirements are identified and classified as transactional Elementary Process (EP) functions or logical (data) file (LF) functions. SFP also eliminates the complexity rating of each function (as Low, Average or High) based on their component range of DETS, FTRS or RETS. This omittance allows for the functional size to be quantified more easily based on high-level, not-yet-detailed functional requirements, and also speeds up the assessment process by eliminating the need to assess the functionality based on the various transaction and date types, and their component DETS, FTRS and RETS.

When to use IFPUG SFP vs IFPUG FP

The determination of which method to use can be influenced by a number of factors: skill level; expertise and training of the analyst; fidelity and availability of detailed functional requirements; and the business need for the count. It is always advisable, when a count is being performed that the analyst(s) conducting the count are properly trained and preferably IFPUG certified, regardless

Concept	IFPUG Traditional FP	IFPUG SFP
IFPUG standardized glossary	Yes	Yes, same
Intent to measure functional size based on FUR	Yes	Yes, same
Method owned by IFPUG	Yes	Yes
IFPUG FP measurement steps: 1. Gather available documentation 2. Purpose/scope/boundary, identify FUR 3a. Measure data functions 3b. Measure transactional functions 4. Calculate functional size 5. Document and report	Yes, but steps 3a and 3b involve additional sub-steps: subclassification into 3 types of transactional functions and 2 types of data functions, and a complexity classification (into Low, Average, or High) to get FP values	Yes
Base functional components (BFC): transactional functions and data functions	Yes: Transactional functions are subdivided into EI, EO, EQ, and Data functions are subdivided into ILF, EIF	Yes: Transactional functions are called "Elementary Processes" and Data Functions are called "Logical Files"
Number of different FP values allocated across function types	3 FP values allocated as Low, Average or High across 5 function types (total of 8 different values)	2 SFP values allocated, one each to two function types
Range of FP values by category	Transactional functions are worth between 3 and 7 FP depending on type and complexity. Logical files are worth 7 to 15 FP depending on type and relative complexity	All transactional functions are considered to be EP and assigned 4.6 SFP. All data functions are considered to be logical files and assigned 7 SFP
Unit of measure	Function Points (FP)	Simple Function Points (SFP)
Convertibility	1 FP = 1 SFP	1 SFP = 1 FP

Table 1: IFPUG FP compared to IFPUG SFP

of method used. Having a count performed by untrained or poorly trained analysts will likely result in a function point count significantly over- or under-counted. Ideally, the analyst is a Certified Function Point Specialist (CFPS) or Certified Function Point Practitioner (CFPP). While IFPUG currently does not have training or certification available for the SFP methodology, a task force has been formed and current plans are to deliver these by the end of 2023.

If the analyst(s) is/are not trained then it is advisable, regardless of the phase of the project or requirements state, to use the SFP method. Likewise, if the requirements and supporting documentation (Entity Relationship Diagrams (ERD), Data Schema, Interface Requirements Documents (IRD)) are not defined to the point where DETS, FTRS or RETS can be confidently identified, SFP should be used. Typically, this is the case early in the software development lifecycle such as at the proposal or project definition phase. If there are cost or time constraints, or there is only a need for a Rough Order of Magnitude (ROM) estimate, then the SFP method can be used. If the sizing will be updated as the project progresses throughout the life cycle, it is recommended – when there are sufficient details available – to use the traditional IFPUG method; particularly, if doing a baseline or application count.

Where there are trained analysts, sufficiently detailed requirements, other documentation available and sufficient time and resources, it is recommended to use the traditional IFPUG Function Point methodology. It is advisable, as well, to use the traditional method when a high degree of accuracy and fidelity for the sizing and estimate are required.

Can IFPUG FP or IFPUG SFP be used to Size Agile Software Development?

With regards to Agile, DevOps, and other non-waterfall development methodologies and frameworks, there is a misconception that function points cannot be used -- either SFP or traditional IFPUG. In addition to function points being language, platform, and technology agnostic, they are also agnostic to development methodology. It is likely the requirements, typically documented as use cases in the Agile world, are not of sufficient detail; The IFPUG SFP method can be used to size product backlogs, use cases, epics and features. Functional size measurement provides the advantages of using objective rule-based sizing over the subjective sizing approaches typically employed in Agile software development, such as story point estimations. As a standardized unit of measure, function points are particularly useful for providing more accurate metrics such as sprint velocity, productivity, and cost/FP.

Dos and Don'ts of Functional Size estimation (using IFPUG FP or IFPUG SFP)

There may be various circumstances which determine the function point sizing method used by the practitioner, but regardless of whether simple function points or traditional IFPUG function points are used, the following provide guidance on the dos and don'ts of function point analysis:

Do:

- Use properly trained analysts, if at all possible, even if it requires hiring an outside analyst
- Properly document the function point count and all source documentation
- Use traditional IFPUG function points if a high degree of accuracy in sizing is required for estimating or legal reasons and there are sufficiently detailed requirements to support it
- Use SFP when it is necessary to develop a quick sizing estimate with little documentation available

Don't:

- Use SFP just because it is easier or quicker; make sure that it will also meet other business needs for the count
- Use SFP if using a parametric estimating tool to develop cost and schedule estimates as none currently on the market support native SFP sizing
- Don't use traditional IFPUG function point sizing when there is limited time or lack of resources to properly conduct the count
- If sizing a waterfall method project and the early phase sizing estimates are done using SFP, it is recommended to transition to traditional IFPUG function points sizing when available documentation becomes available.
- Depending on the business need, it is not recommended to use SFP for application counts, because all of the prerequisite details to do a formal IFPUG FP count should be available and known.

Example Case Study to Demonstrate Functional Size Estimation

Consider: We have a high-level CONOPS (Concept of Operations) document that outlines the following hypothetical functional requirements for a simple online book sales system:

a. Create, read, update, delete (CRUD), and store customer records.
b. System administrator functions to create, read, update, delete (CRUD), and store book catalog entries for available books.

Functional Requirement	IFPUG avg functions	IFPUG FP value	IFPUG SFP functions	IFPUG SFP value
CRUD, store customer records.	3A EI, A EQ 1A ILF	26 FP	4 EP, 1 LF	25.4 SFP
CRUD, store book catalog	3A EI, A EQ 1A ILF	26 FP	4 EP, 1 LF	25.4 SFP
Display books by author or title	1A EQ	4 FP	1 EP	4.6 SFP
Select and display book details	1A EQ	4 FP	1 EP	4.6 SFP
Select books to create order	1A EI, 1A ILF	14 FP	1 EP, 1 LF	11.6 SFP
Display order summary (calcs)	1A EO	5 FP	1 EP	4.6 SFP
Pay for order with credit card	1A EI	4 FP	1 EP	4.6 SFP
Order summary to customer	1A EO	5 FP	1 EP	4.6 SFP
Order request to sales staff	1A EO	5 FP	1 EP	4.6 SFP
TOTAL	**8A EI, 3A EO, 4A EQ, 3A ILF**	**93 FP**	**15 EP, 3 LF**	**90 SFP**

Table 2: Comparison of IFPUG FP (avg) and IFPUG SFP for CONOPS Case Study

c. Customers can display and browse book catalog by author or title.

d. Customers can select and see details about an individual book.

e. Customers can create an order for one or more books by selecting them from the catalog and placing them in a shopping cart, saved as an order.

f. The system will display order summary with the total amount calculated from the prices of all books.

g. Customers can complete their order by paying with a credit card.

h. Software will generate an order summary to the customer.

i. Software will generate an order request to the sales staff at the store.

Table 2 presents the high-level summary of using both IFPUG FP (assuming all functions are average complexity) and IFPUG SFP. The total over the entire case study came out to be close for the IFPUG avg FP estimate and the IFPUG SFP estimate, respectively being 90 FP and 93 SFP. If there were more detailed requirements, such as complex reports, that would be scored as a high complexity EO (External output), there would be a larger variation between the methods because the value of a H EO is 7 FP versus the IFPUG SFP single EP score of 4.6 SFP.

Note that the following acronyms are used in Table 2:

For IFPUG avg (average) functions:

- *A EI or A EQ= average External Input or average External Query worth 4 FP each*
- *A EO = average External Output worth 5 FP*
- *A ILF = average complexity Internal Logical File worth 10 FP*

For IFPUG Simple Function Point (SFP) functions:

- *EP = elementary process worth 4.6 SFP*
- *LF = logical file worth 7 SFP*

Conclusion

The IFPUG function point methodology is a tried-and-true software sizing method, recognized as an ISO/IEC Functional Size Measurement standard, and is especially suitable to sizing software when detailed functional requirements are known. The evolution of the Simple Function Point methodology (IFPUG SFP V2.1) presents a simplified approach to functional sizing that is especially useful for early estimation when functional requirement details are not yet specified or available. IFPUG SFP facilitates using IFPUG FP concepts when conditions and circumstances warrant the use of a rules-based sizing method, while simultaneously providing one that can be readily used quickly for high-level requirements. IFPUG SFP provides such a method, in lieu of, but true to IFPUG FP, with the added benefits that it is easier to learn and provides a reasonable level of accuracy in a more timely and efficient manner than using the formal IFPUG FP methodology.

Appendix A
Early and Quick Function Point Aggregation Levels and Values (DPO):

BFC IFPUG	E&QFP components
ILF	ILFL – low ILFA – average ILFH – high
EIF	EIFL – low EIFA – average EIFH – high

BFC IFPUG	E&QFP components
EI	EIL - EI low EIA - EI average EIH - EI high
EQ	EQL - EQ low EQA - EQ average EQH - EQ high
EO	EOL - EO low EOA - EO average EOH - EO high

Table 3: Early and Quick 1st level aggregation (DPO)[2]

If less mature requirements are available, then the analyst can genericize the functions to the 2nd aggregation level:

Transactions:
classified as UEP - Unclassified Elementary Process:

GEI – Generic EI EI-type process with undetectable level of complexity.

GEO – Generic EO EO-type process with undetectable level of complexity.

GEQ – Generic EQ EQ-type process with undetectable level of complexity.

UGO - Unspecified Generic Output (EO/EQ) "doubtful" or "uncertain" output process for which there are no details available to differentiate between EO and EQ.

UGEP - Unspecified Generic Elementary Process (EI/EO/EQ) "doubtful" or "uncertain" elementary process for which there are no details available to single out the primary goal, namely the presence of EI, un EO or u n EQ.

GILF – Generic ILF Sets of data recognizable by users as ILF-type of an uncertain complexity

GEIF - Generic EIF Sets of data recognizable by users as EIF-type of an uncertain complexity.

UGDG - Unspecified Generic Data Group Unspecified logical file (either ILF or EIF) of uncertain complexity.

Table 4: Early and Quick 2nd level aggregation (DPO)[3]

Transactions

When it is not possible to accurately identify a specific UBFC or the precise amount of UBFC that makes up a specific software component it is possible to use a 3rd level component.

Typical Process (TP) It consists of a set of four typical functional processes: Insert, Edit, Delete, Display a record data, recognised as CRUD – (Create, Read, Update & Delete) and generally centred around a specific data store. Normally it corresponds to the general definition "Management of a data store", "Management of …".

When detectable, the Typical Process helps save measurement time without losing out in accuracy in the four base components shortlisted.

There are three TP classes:

TPS – Typical Process - Small: CRUD

TPM – Typical Process - Medium: CRUD + List (EQ)

TPL – Typical Process - Large: CRUD + List (EQ) + Report (EO)

General Process (GP)

It consists of a general set of Unclassified Elementary Process (UEP). If they fail to be detected with accuracy a General Process component is detected instead.

It is a more general type of "unspecified" BFC aggregation which differs from CRUD.

There are three different GP components that depend on the amount of UEP put together.

GPS – General Process - Small: 6 -10 UEP's

GPM– General Process - Medium: 11 -15 UEP's

GPL– General Process - Large: 16 -20 UEP's

Data

General Data Group (GDG) For the data component, three General Data Group (GDG) typologies are identified at three different aggregation levels which depend on the amount of ULF taken into account in the GDG, in particular:

GDGS - General Data Group - Small: 2-4 ULF

GDGM - General Data Group - Medium: 5-8 ULF

GDGL - General Data Group - Large : 9-13 ULF

Table 5: Early and Quick FP 3rd level aggregation (DPO)[4]

Group of GP's (General Processes)

The fourth level of aggregation applies when user requirements are such as to be described at a summary level and measured as a functional area of a medium or large application. This level of aggregation can be used for subsets of large and functionally complex applications.

Aggregations are functional components of the General Process type (third aggregation level) that are grouped together as MP-type components (MP= macro process).

Transactions

MP – Macro Process If the level of detail is insufficient, instead of the numerous General Processes (GP) it is possible to detect a Macro Process (MP) of small, medium and large scale.

MPS – Macro Process – Small: 2-4 GP's

MPM – Macro Process – Medium: 5-7 GP's

MPL – Macro Process – Large: 8-10 GP's

A Macro Process can amount to a large system segment, a sub-system or even an entire small scale application.

Table 6: Early and Quick FP 4th level aggregation (DPO)[5]

1st aggregation level: components and values

Type of functional component	Function Type	Min	ML most likely	Max
Transactions				
Base Functional Component (IFPUG)	EIL - EI low	3,0	3,0	3,0
	EIA - EI average	4,0	4,0	4,0
	EIH - EI high	6,0	6,0	6,0
	EQL - EQ low	3,0	3,0	3,0
	EQA - EQ average	4,0	4,0	4,0
	EQH - EQ high	6,0	6,0	6,0
	EOL - EO low	4,0	4,0	4,0
	EOA - EO average	5,0	5,0	5,0
	EOH - EO high	7,0	7,0	7,0
Data				
Base Functional Component (IFPUG)	ILFL - low	7,0	7,0	7,0
	ILFM - medium	10,0	10,0	10,0
	ILFH - high	15,0	15,0	15,0
	EIFL - low	5,0	5,0	5,0
	EIFM - medium	7,0	7,0	7,0
	EIFH - high	10,0	10,0	10,0

2nd aggregation level: components and values

Type of functional component	Function Type	Min	ML most likely	Max
Transactions				
UEP (Unclassified Elementary Process)	GEI - Generic EI	4,0	4,2	4,4
	GEQ - Generic EQ	3,7	3,9	4,1
	GEO - Generic EO	4,9	5,2	5,4
	UGO - Unspecified Generic Output (EQ/EO)	4,1	4,6	5,0
	UGEP - Unspecified Generic Elementary Process (EI/EQ/EO)	4,3	4,6	4,8
Data				
ULF (Unclassified Logical File)	GILF-Generic ILF	7,4	7,7	8,1
	GEIF-Generic EIF	5,2	5,4	5,7
	UGDG – Unspecified Generic Data Group	6,4	7,0	7,8

3rd aggregation level: components and values

Type of functional component	Function Type	Min	ML most likely	Max
Transactions				
TP Typical Process	TPS – small (CRUD)	14,1	16,5	19,0
	TPM – medium (CRUD+List)	17,9	21,1	24,3
	TPL – large (CRUD+List+Report)	22,3	26,3	30,2
Data				
GP General Process	GPS – small 6-10 UEP's	26,4	35,2	44,0
	GPM – medium 11-15 UEP's	42,9	57,2	71,5
	GPL – large 16-20 UEP's	59,4	79,2	98,9
Data				
GDG General Data Group	GDGS – small 2-4 ULF	15,0	21,4	27,8
	GDGM – medium 5-8 ULF	32,4	46,3	60,2
	GDGL – large 9-13 ULF	54,8	78,3	101,8

4th aggregation level: components and values

Type of functional component	Function Type	Min	ML most likely	Max
MP Macro Process	MPS – small 2-4 Generic GP's	111,5	171,5	231,5
	MPM – medium 5-7 Generic GP's	185,8	285,9	385,9
	MPL - large 8-10 Generic GP's	297,3	457,4	617,4

Table 7: Early and Quick FP Range Values by Aggregation Level (DPO)[6]

***Carol Dekkers** is a management consultant and president of Quality Plus Technologies, specializing in workings with organizations who want to keep their promises to their clients through realistic, data-founded estimates. In 2022, she was named the ICEAA 2022 Educator of the Year for her work as the lead author for soon-to-be-released Software Cost Estimating Body of Knowledge (CEBoK-S).*

She holds a BSc in Mechanical Engineering and is an IFPUG Certified Function Point Specialist (Fellow), a PMP, Certified Scrum Master and a P.Eng. (Canada). Ms. Dekkers has authored several textbooks on functional size measurement (function Points) and project management and is a frequent presenter and instructor on these topics internationally. She currently consults under contract to several U.S. federal government agencies and to the private sector. In Feb 2023, Ms. Dekkers was named one of Consulting Magazine's 2023 Global Leaders in Consulting for her leadership in pro bono for her contributions to the IFPUG, ASQ, PMI, and the International Organization for Standardization (ISO).

***Daniel (Dan) B. French** has twenty years of experience in a variety of IT roles including analyst, developer, tester, project management, software metrics and software estimation. He has developed CMMi and ISO compliant process improvement and software estimation processes.*

Dan holds a B.S. in Economics from Virginia Tech and holds certification as a Lifetime CFPS, PMP and CSM. He recently served as the Certification Director on the IPFUG Board of Directors and is currently the chair. Previously he served as the Chair of the Functional Software Sizing Committee (FSSC). He's authored numerous white papers/presentations on software metrics, Agile, and estimation.

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