An Exploratory Parametric Analysis of Production Cost Improvement in the Aerospace Industry*

Paul S. Killingsworth, M.S., CCE/A
Technical Expert, Tecolote Research, Inc. • pkillingsworth@tecolote.com

ABSTRACT

Decades after its observation and description by T. P. Wright (1936), unit cost improvement is a well-documented characteristic of production processes. Since that time, however, it has remained chiefly an empirical phenomenon. Cost analysts have few objective tools to apply to the prediction of cost improvement in future production programs. This problem is particularly acute in the aerospace industry. In this arena, prime contractor and government cost analysts must provide estimates for multi-billion dollar programs involving new products and technologies, which may not be entering production for years. Current methodologies involve the use of analogies and standards that are subjective and often difficult to justify. The research described in this paper, making use of a dataset comprised of 30 diverse aerospace programs, investigated whether parametric methods can be employed to model variation in cost improvement. Using a simple theory of cost improvement based on the scale, complexity, and design stability of the products, it found that parametric models may have utility in predicting the cost improvement of future programs.

INTRODUCTION

In the field of cost analysis, the terms "learning" and "cost improvement" refer to the observed reductions in unit cost that occur during a production run for a complex end item.[1] For many years, cost analysts have fit downward sloping curves to plots of unit production costs, and have used the slopes to characterize the rate of cost improvement. Cost analysts in training spend much time learning about these curves, including how to properly plot average lot costs, and the differences between the unit and cumulative average computational approaches. This training is useful for answering such questions as, "What will be the cost of the next production lot (or unit) to be produced on an established production line?" In this case, there is an empirical observation of the relevant improvement rate, and it is assumed that this rate will continue to be applicable to the given production facility with its known tooling, workers, and management (AFSC, 1986, 7–96). In addition, cost analysts working for established manufacturers often will have an extensive corporate cost experience on which to draw, and when a new product, or new model of a product, is to be produced they will have an expectation for unit cost improvement in the various factory departments (Teplitz, 1991, 48). To date, this is essentially the recommended curve selection approach—to extend the observed historical cost improvement experience on the same product or a very similar product (AFSC, 1986, 7–80).

[1] In the literature, the term "learning" is often applied to learning on the part of workers performing direct labor in a production process (Teplitz 1991, 8). The term "cost improvement" has a more general meaning, and is often applied to improvements in unit costs due not only to worker learning, but also to process improvements and greater management experience (Robbins 1986). It is the latter, broader aspect that will be addressed in this paper.

The underlying assumption of these methods is that similar products produced using similar production methods should have the same or nearly the same cost improvement experience. Unfortunately, there are common situations in which methods based on these assumptions are difficult to apply. These situations include the production of new products when there is little corporate experience to draw upon, or the setup of new production facilities. Even more problematic is a task encountered by cost analysts for government procurement agencies and aerospace contractors. Here they often must provide an estimate for a budget or corporate cost position on a prospective contract even before the request for proposals has been released. Later, during source selection, another related issue is encountered when bidders' estimates are being evaluated by the government. Is the improvement rate used in the proposal reasonable? Or is the bidder trying to "buy-in" with an overly optimistic rate?[2] As a result, the assumed improvement rates can become a source of controversy. Program advocates on one side press for "affordable" cost estimates (higher rates of learning), while on the other side independent reviewers and higher headquarters want "conservative" estimates (lower rates) in order to avoid overruns. The cost analyst is in the center of this discussion and faced with justifying one cost improvement rate that is acceptable to all. It is a difficult position, and it results from the fact that cost improvement remains essentially an empirical phenomenon. Current cost improvement "theory" can be more accurately described as a set of conjectures about the sources and effects of the various factors that influence the production process. But it is not objectively predictive. If it were, the theory would yield quantitative techniques that analysts could apply to their estimates. Such techniques could then be used to calculate "most probable," "conservative," or "optimistic" cost improvement rates, along with associated statistical confidence levels.

This paper addresses cost improvement from the perspective of government and prime contractor cost estimators who are faced with the task of specifying defensible cost improvement rates for future production programs. Much of this experience applies to the defense aerospace business and their products (space vehicles, missiles, aircraft, avionics, etc.). In this arena, analysis is often performed on programs that will not enter into production for years. They may still be in concept exploration or engineering development, with the production contractor yet to be identified. For these programs, the practical application of the concept of cost improvement is not well-supported by the descriptive nature of the underlying theory. In short, there is a deficit of cost estimating methodology in this area. After some further discussion and definition of the problem, this paper will describe the results of an exploratory parametric analysis of cost improvement for a number of aerospace products, which indicate the possibility of developing a parametric approach to the issue of choosing a cost improvement slope. Such an approach, based on broad assumptions about the sources of cost improvement, could provide cost analysts with a better basis for estimating cost improvement rates in analytic environments like that of the defense aerospace industry.

[2] The term "buy-in" refers to the practice of understating estimated costs in order to make a contract bid more attractive.

BACKGROUND

Cost Improvement Assumptions Can Be Decisive

The cost improvement rate applied to an estimate for a large production program can be important. Figure 1 shows the results of a sensitivity analysis performed on the cost improvement assumptions used for a recent space system acquisition program.[3] The graphs represent lines of equal first unit (T1) estimates, showing the change in total production cost due to change in the cost improvement assumption. The program proposed a constellation of more than 20 satellites, with a total production run of over 30. Program management maintained that although cost improvement in satellite programs is considered to be low or even non-existent, a slope of 90% would be feasible given the quantities involved. This resulted in a production estimate of $13.4 billion (FY01$), as indicated by the circle on the graph. Higher headquarters, however, took a position (backed by a longstanding industry standard) that cost improvement for military space systems production is 95%. This yielded a production estimate of $15.2 billion, shown by the square. This higher cost of production contributed to the program being subsequently cancelled.

It is interesting to note that the movement from a 90% to 95% improvement rate resulted in a production program increase of 13.4% (from $13.4 to $15.2 billion). An equivalent increase in cost while holding the cost improvement assumption steady at 90% would have required a 46% increase in the estimate for T1 (as shown by the upper curve). Most cost analysts will agree that the time and effort expended on the estimate for T1 in a program such as this far exceeds the time spent on the cost improvement assumptions. Changes in T1 as small as 5% are considered highly significant, yet the graph shows that a 5% change results in a very small change in total production costs compared to the effect of changes in the cost improvement rate (as shown by the middle curve). This indicates that cost improvement assumptions are at least as important in determining total program

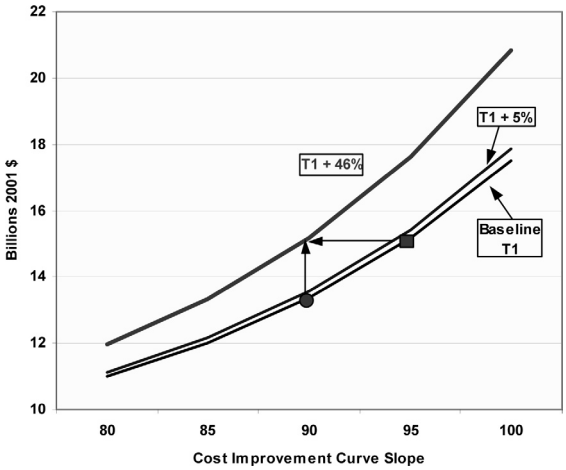

Figure 1. Satellite Program Life Cycle Cost as a Function of Cost Improvement Rate Assumption

[3] The specific costs shown in the figure have been modified without changing the relative effect of changes in the cost improvement assumptions.

cost as the T1 estimate, leading an observer to suspect that the analytic emphasis on T1 has more to do with the quality of the methodologies that can be brought to bear than with the relative importance of the two issues.

Current Cost Improvement Methodologies Are Subjective

Despite the importance of cost improvement rates, cost analysts have few quantitative tools to apply to their estimation. For the program described above, program management cited analogous space programs that had 90% or better improvement rates while the reviewers questioned the validity of the analogies and cited other programs with lower rates. These kinds of controversies are common when it comes to cost improvement.

Cost analysts currently use a number of approaches for specifying cost improvement rates, all of which have a substantial degree of subjectivity. Cost estimating handbooks and texts often cite the high levels of uncertainty associated with forecasting these rates, and the need for detailed knowledge of the product and its manufacturing processes (AFSC 1986, 7-4). In his reference on learning curves, Teplitz (1991, 47) cites prior experience as one of the guides available to cost analysts in specifying a cost improvement rate. Prior experience is a form of historical analogy, which draws on the cost history of specific firms manufacturing products similar to ones they have manufactured in the past. However, if no such highly specific analogies are available, this methodology becomes more difficult to apply. In addition, even a narrow product class such as ground-launched tactical missiles has experienced production improvement rates ranging from 75% to 85%.[4] Satellite production programs have had a similarly wide range with some experiencing as much as 82.5% improvement, and others experiencing none, or negative improvement. Cost analysts looking for analogies in this environment will have many diverse ones from which to choose, and most likely a difficult time justifying their choice in any definitive sense. The wide ranges of cost improvement within product categories make analogy a difficult methodology to objectively apply. It also indicates that the assumption that product type is a useful predictor of cost improvement rate is worthy of critical analysis and testing.

Another approach cited by Teplitz (1991, 47) is that of relying on industry standards. These can range from specific standards for individual manufacturing operations, to broad rules that apply to entire classes of products. Unfortunately, even for the simplest, most focused operations, cost improvement rates can vary widely. For example, in an early effort to provide a set of industry standards, Nadler and Smith (1963, 115) gathered detailed cost data on ten production processes being conducted in three companies. Table 1 summarizes their findings. The wide ranges around the mean for most of these operations suggest that there were a number of factors affecting the cost improvement rates besides the process itself that were unaccounted for.

At the other end of the spectrum there are cost improvement standards that have been applied as a matter of acquisition policy to the finished products of entire industries. Baloff (1966, 275) and Conway (1959, 39) referred to the "notorious 80% law," which was assumed for many years to apply to all airframe manufacturing. A similar example is the 95% assumption currently attached to military satellite production. These standards

[4] A range of 75% to 85% cost improvement rates was observed in the dataset of nine tactical missile programs used in the analysis that follows.

Table 1. Observed cost improvement rates for machine operations. (Nadler and Smith, 1963)

Process	Mean Rate (%)	Observed Range(%)
Power sawing	83.3	83.3–83.3
Milling	89.0	75.5–101.2
Drilling	87.4	70.5–101.6
Honing	100.5	91.1–117.5
Lathes	84.4	79.8–95.1
Deburring	91.9	76.8–115.2
Bench Inspection	89.8	75.3–107.1
Grinding	82.0	82.0–82.0
Assembly	80.4	79.9–80.8

are based on industry consensus about reasonable cost improvement rates for classes of products—rates that can be used in the absence of other data. Often the standard is established as a conservative estimate of the expected improvement rate, which will insure that budgets are set high enough to avoid overruns. For example, at the beginning of military satellite production in the 1960s there was skepticism on the part of many in the industry that any cost improvement could occur at all. However, it was concluded that some minimal amount must take place, so 95% was allowed in contract negotiations. This rate has remained a standard assumption ever since.

The "industry standard" approach to cost improvement has some significant disadvantages for both the contractor and the contracting authority. First, a contractor may firmly believe that it will achieve higher rates of cost improvement than its competitors, and bid its lower costs accordingly. The selection authority, however, will be rightly skeptical and ask for data supporting the higher rates of improvement. Unfortunately, the state of the art for estimating cost improvement is so limited that any deviation from the industry consensus rate is difficult to justify. The result is that the lower cost producer may not get the full benefit of its cost improvement processes in the source selection.

From the perspective of the contracting authority, the standard approach may help insure that the contracted cost fully covers the anticipated work, but it can also insure that the "worst case" production cost becomes a reality. As Young (1966, 410) observed years ago, although the contracting authority may think of the standard as a conservative "not-to-exceed estimate," it becomes a de facto goal for the contractor. The work expands to fill the budgeted hours, and an opportunity to achieve greater efficiency is lost by failing to incentivize the contractor to do better. Meanwhile, the industry standard rate becomes reinforced by another programmatic example.

In short, production cost improvement remains the same empirical phenomenon that was described in the early papers based on aircraft production experience (Wright, 1936) (Crawford n.d.) (Asher, 1956). It was not long after the original descriptions of the phenomenon that research papers began citing the need for a theoretically-based calculation of cost improvement slopes and a movement away from empiricism (Levy, 1965, B-136)

(Baloff, 1966, 275). No such theory has been forthcoming, however, and the estimation of cost improvement for future programs remains a methodological weak link in the cost analysis process.

A Parametric Approach to Cost Improvement

A cost analyst can have a number of objectives when applying an improvement rate to a production estimate. At the direction of management, it could be a conservative, "not to exceed" rate in order to assure that the allocated hours are sufficient. On the other hand, applying a manufacturing engineer's knowledge of the value of production goals, the rate could be a difficult one to achieve that could serve to motivate continuous improvement in the production process. Clearly, neither rate is necessarily the actual one that will be experienced. What the analyst needs to know is the most likely rate along with the confidence levels associated with alternatives. Parametric methods are well suited to provide this type of information.

Most cost analysts are familiar with the use of statistical regression techniques to estimate the coefficients of cost estimating relationships (CERs). In this application, scale-, complexity-, or performance-related parameters are correlated with the historical costs of items. The resulting relationships are used to estimate the costs of future items that are similar. Production CERs commonly yield the cost of a single unit of production, such as the first, 100th, etc., to which a cost improvement curve is applied to estimate the cost of an entire production run. An analogous parametric approach to cost improvement would make the rate itself another dependent variable to be estimated, based on parameters hypothesized to influence that rate. Such an approach may have useful predictive power, but even if this does not turn out to be the case, a parametric analysis would be a useful means to investigate certain questions about cost improvement, such as the relevance of product type in predicting the phenomenon.

RESEARCH OBJECTIVE AND QUESTIONS

The objective of the research described here was to investigate whether parametric analysis is a feasible approach to estimating cost improvement rates. If so, the "cost improvement relationships," or CIRs, resulting from such an analysis could provide a useful addition to the cost analysis toolbox. With traditional CERs to estimate unit costs and CIRs to predict the rate of improvement of those costs during production, cost analysts would have a more complete set of analytic tools with which to conduct production estimates.[5]

The following research questions were used to pursue the research objective:

- Can a theory of cost improvement causality be incorporated into the design of CIRs?
- What independent variables can be used in CIRs?
- Are independent variables that are related to product type (e.g., satellite, aircraft, avionics, missiles) as predictive as those related to the production environment (level of automation, design stability, etc.)?

[5] In addition, another difficult analysis problem, that of characterizing the estimating risk associated with cost improvement assumptions, would have a more definitive answer due to the availability of statistics describing the distribution of cost improvement rates on which the CIRs were based.

As pointed out in the above discussion of problems associated with the analogy methodology, the wide variation within product type classes makes it worthwhile to investigate whether cost improvement should be estimated without reference to product type, i.e., by using characteristics of the product and production environment that apply to a range of product types. A parametric analysis such as the one envisioned in this research would serve to test this hypothesis.

DATASET DEVELOPMENT

To pursue the research objective it was necessary to compile data on cost improvement experienced by a number of diverse aerospace products. This was because the research did not assume that cost improvement is best estimated within the bounds of individual product classes. Instead, it remained open to the possibility that avionics production experience, for example, could be applied to missile or satellite production. As a result, data were sought on the cost improvement experience for spacecraft, missiles, avionics, and aircraft. The exploratory nature of the research precluded an extensive effort to compile data directly from aerospace contractors and government agencies. Previously conducted studies were drawn on to develop a dataset representing as many programs of each product type as possible. The effect of this constrained data collection process was that availability was a greater determinant of inclusion in the study than representativeness or comprehensiveness. Nevertheless, data were available on 31 significant aerospace programs representing a range of spacecraft, missile, avionics, and aircraft systems. Table 2 shows the programs that were part of the initial analysis.

The missile, avionics, and aircraft production data were obtained in the form of average lot costs. In each case, curves of the form $Y = Ax^b$ were fit to the cumulative average plot points using logarithmic transformations and ordinary least squares regression (Tecolote, 2004). The slopes of these curves in log space were used to derive the rates used in the analysis. In a number of cases, most notably the missile and avionics systems, prototype lot data were also available. These lots were not used in the estimation of the full rate production cost improvement curves.

Table 2. Programs included in cost improvement dataset.

Spacecraft	**Avionics**
DMSP 5D-2 FO & 5D-3	AN/AAR-47 EO TWS
DSP I (14-22)	Apache AH64-A TADS/PNVS
DSCS III (B4-B14)	AN/ALR-45D RWR
MILSTAR 1	AN/APQ-128 TFR
GPS II/IIA (14-40)	AN-ALQ-137
Additional Program	AN-APR-25
Missiles	AN-AAS-35V
Patriot	AN-APQ-134
HARM	AN-APQ-153
AMRAAM	AN-APX-76
Stinger	AN-APX-100
MLRS	**Fighter Aircraft**
Hellfire	AV8B
Phoenix	F14
Javelin	F15
ALCM	F16
	F/A18

Data representing a number of potential predictive parameters were also collected. System weight, total production quantity, average lot size, production rate, number of prototypes, and integration and assembly costs were obtained from various program documentation materials. Another parameter, the theoretical first unit cost (T1$), was obtained from the fitted improvement curves and normalized to FY2002 dollars using the standard inflation indices published by the Office of the Secretary of Defense (OSD).

ANALYSIS

The analysis consisted of a series of regressions of the cost improvement slopes associated with the programs in Table 2 against parameters of interest from the perspective of product type and production processes. The first step was to consider the theoretical specification of the relationships, in terms of the functional form whose coefficients would be estimated. In the second step, the broad classes of influences on cost improvement rates were proposed, and specific parameters were identified in each of the classes. Finally, the coefficients of the relationships were estimated along with their associated statistics, and the results evaluated.

Function Specification

Although the steps of an analysis such as this can be laid out sequentially as described above, in execution the process is often an iterative one. The log-linear functional form was settled on after a consideration of the candidate parameters along with an examination of the curve fits that were achieved in the early exploratory regressions. It was clear that for those parameters correlated with cost improvement, such as production rate, the response of the dependent variable to changes in the parameters was proportionate and had negative returns to scale. That is, as production rate increased, the cost improvement rate also increased, but not indefinitely. Cost improvement increased at a decreasing rate. Since "higher" rates of cost improvement are actually represented by lower percentage values, the correlation between independent and dependent variables described here is a negative one.6 The function representing this response is a negative exponential form, with the exponent ranging between –1 and 0. These exponents are represented in Figure 2 as $\mathbf{b_1}$, $\mathbf{b_2}$, and $\mathbf{b_3}$.

Also under investigation in the analysis was the role of product type in determining cost improvement rates. This was modeled by using indicator variables signifying missiles, avionics, and fighter aircraft. A zero setting for all three of these product types represented the fourth type, spacecraft. The hypothesized form of the function tested in the parametric analysis and CIR development is summarized in Figure 2.

Identification of Independent Variables

Consideration was next given to the nature of the underlying causality between possible independent variables and cost improvement rates. The purpose of this preparatory step was to propose the major dimensions of cost improvement potentiality, in order that parameters could be selected as surrogates for each. In this way, the CIRs would attempt to encompass the major influences on cost improvement. Three broad product charac-

[6] An 80% improvement rate is "higher" than a 90% rate, since a doubling of production quantity with an 80% rate represents a 20% unit cost reduction compared with a 10% reduction for the 90% rate.

$$CIC\% = b_o X_1^{b1} X_2^{b2} X_3^{b3} b_4^{X4} b_5^{X5} b_6^{X6} \varepsilon$$

Where: CIC% = cost improvement curve rate (%)
X_1, X_2, X_3 = independent variables related to scale, complexity, and design stability
X_4 indicator variable = 1 for missile, 0 otherwise
X_5 indicator variable = 1 for avionics, 0 otherwise
X_6 indicator variable = 1 for fighter aircraft, 0 otherwise
b_o = constant coefficient derived from regression
b_n = coefficients to be estimated by regression
ε = multiplicative error term

Figure 2. Log-linear functional specification used for analysis.

teristics were identified to represent these dimensions. They were scale, complexity, and design stability.

Scale, as exemplified by parameters such as weight and volume, is proposed with the straightforward assertion that larger products have greater potential for cost improvement than smaller ones, assuming other factors such as complexity are held equal. Larger products generally require more hours of fabrication, assembly, and integration labor, as well as more production steps. With more workers and more production steps, the opportunity for both autonomous worker learning and management-directed process improvements increases with product scale.

Complexity is more difficult to define and measure than product scale. A portion of the literature relates product complexity to the number of subassemblies and lower components that must be integrated to manufacture the complete product. Taylor (1985) uses this definition of complexity in specifying a range of standard cost improvement rates for electronics assemblies, characterizing them as components, subassemblies, rack chassis, and full racks. Conway and Shultz (1959) used an alternative definition of complexity in their study of a complicated, low production rate product. They observed that the complexity of the product was indicated by the high degree of specialized labor and large proportion of effort involved with integration, assembly and test. Yet another aspect of product complexity involves the processes used during manufacturing. For example, in the space business exotic materials such as beryllium and composites are considered more complex and time consuming to implement than standard materials such as aluminum. It is clear that a definition addressing every aspect of product complexity would be problematic. However, for the purposes of this analysis the presumed effects of complexity on unit cost were focused on rather than delineating an all-encompassing definition. Indirectly defined, the analysis assumed that more complex products would be more costly, have greater labor content, or require greater integration effort than other products of the same size and weight (scale). More complex products were assumed to have greater potential for cost improvement during production due to their specialized labor content and larger required integration efforts. This would be reflected by such metrics as dollars-per-pound or estimated T1 cost.

The last dimension of cost improvement used in the analysis was design stability. One characteristic of the government acquisition process for defense aerospace products is that technology can change rapidly and operational requirements can be volatile. These

factors are often translated into design changes, even during production. Although systems with high production rates such as tactical missiles usually have relatively stable designs by the time they enter full-rate production, spacecraft, which have very low rates, often must incorporate significant design changes from one unit to the next. These design changes generally do not have the objective of enhancing manufacturability, but usually implement new requirements or more effective technologies. The result is to "reset" the cost improvement curve for the affected components or subassemblies. Therefore, it is assumed that products with stable designs have a greater potential for cost improvement than those that must incorporate many design changes during production.

In each of the three areas, specific parameters were identified to serve as surrogates for the general causal effects described above. Figure 3 lists these parameters along with their hypothesized effect on cost improvement. Larger total production quantity, lot size, and production rate were considered indicative of higher levels of design stability, since these aspects of production would tend to motivate stable designs to reduce cost. Larger numbers of pre-production prototype units would result in the identification of design problems before entering production, and would therefore also tend enhance the stability of the produced design. Complexity was assumed to be closely related to unit costs, resulting in the investigation of T1$; integration, assembly, and test (IA&T) cost; as well as the ratio of the two as possible complexity surrogates. A dollars-per-pound metric, T1$/lb, was also investigated, since more complex products were assumed to have higher unit costs than less complex products of the same weight. Finally, weight was used to represent the scale (size) of the products in the dataset.[7]

Pairwise Analysis

As an initial step, a pairwise analysis was performed to examine the strength of the direct linear correlations between potential independent variables and the dependent variable.

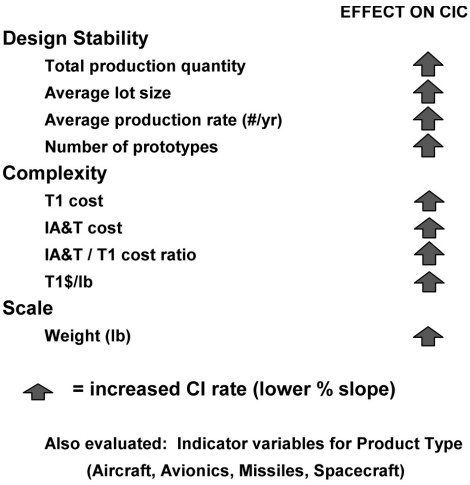

Figure 3. Parameters investigated in analysis.

	Ln CI Rate
Ln Tot Qty	-0.65
Ln Av Lot Size	-0.61
Ln Prod Rate	-0.56
Ln Prototype Qty	-0.70
Ln T1 $	-0.09
Ln IA&T $	0.23
Ln IA&T$ / T1$	0.14
Ln T1$/lb	-0.25
Ln Weight (lbs)	0.04
AIRCRAFT	0.05
MISSILE	-0.73
AVIONICS	0.43
SPACECRAFT	0.28

Figure 4. Correlation coefficients.

[7] Volume was unavailable for most of the data points.

Since a log-linear functional specification was assumed, the coefficients were calculated using the logarithmically transformed values. The results are displayed in Figure 4. The parameters related to design stability (total production quantity, average lot size, production rate, and number of prototypes) all had promising levels of correlation with production cost improvement. The parameters related to complexity (T1$, IA&T cost, the IA&T/T1$ ratio, and T1$/lb) had less direct correlation with the cost improvement rate. T1$/lb, at –0.25 was the most correlated, and in the proper direction. Finally, weight, the sole parameter related to product scale, showed little direct correlation in this first inspection. Pairwise analysis, however, does not indicate whether a parameter explains a significant portion of the variation left unexplained by the other parameters. This indeed was the case when weight was later included with other parameters in the specified model.

Of the indicator parameters related to product type, MISSILE was highly correlated with cost improvement; indicating that the tactical missile data points for the most part had higher rates of cost improvement than other products.

Regression Analysis of the Specified Model

In the next step of the parametric analysis, a number of regression estimates of the specified model were run. Various combinations of the independent variables related to design stability and complexity were tested, along with weight and the three product-related indicator parameters. During the course of the analysis, one data point, MLRS, was found to be based on incomplete data and was excluded.[8] This was the only outlier removed from the dataset.

Of the parameters related to design stability, prototype quantity had the most significant coefficient (1-p value of 0.98). However, this parameter was not available for all of the data, including the majority of the avionics data points. Average lot size was also significant (1-p value of 0.81), and was available for all of the data. For this reason, average lot size was retained in the model for subsequent analysis. It also had a lesser degree of multicollearity with the other parameters at later stages of the modeling activity.

Among the complexity-related parameters, IA&T cost was significant, but indicated a substantial degree of multicollinearity with the weight parameter and reduced its significance when included in the model. T1$ and T1$/lb both had significant coefficients in these early model runs (1-p values of 0.99). T1$/lb served to enhance the significance of the weight parameter, and so was retained in the model. The weight parameter and product-related indicator variables were retained in these early estimates as part of the initial specification.

Figure 5 provides summary statistics and an actual vs. predicted plot of the full-specified model resulting from the initial trials. Although the results at this stage were not impressive from an overall modeling perspective, it was notable that both the technical parameters (average lot size, weight, T1$/lb) and the product type indicator parameters were significant, as shown by the t-statistics of their coefficients. None of the parameters could be considered candidates for exclusion, since all the t-statistics exceeded the 0.7 (1-p) value criterion. However, the relative predictive power of the parameters was hard to judge, due to the high degree of multicollinearity between the technical and product type parameters. As might be expected, weight was positively correlated with the

[8] The MLRS datum was found to be based on non-concurrent lots with the intervening quantities unknown.

aircraft in the database, and negatively correlated with the avionics boxes. Lot size was positively correlated with the missiles, which are generally produced in greater quantities. The next step of the analysis explored the predictive power of the technical parameters in the absence of the product type indicator variables.

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	30
Equation in Unit Space:	CIRate = 127.5 * LotSz ^ (-0.01919) * Wt ^ (-0.01766) * T1$/lb ^ (-0.05557) * 0.9151 ^ AC * 0.9076 ^ MISS * 0.886

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	4.8482	0.1603		30.2485	1.0000
LotSz	-0.0192	0.0141	-0.4651	-1.3584	0.8125
Wt	-0.0177	0.0148	-0.5285	-1.1961	0.7562
T1$/lb	-0.0556	0.0166	-0.7861	-3.3408	0.9972
EXP_AC	-0.0887	0.0713	-0.4179	-1.2448	0.7743
EXP_MISS	-0.0969	0.0708	-0.5419	-1.3701	0.8162
EXP_AVI	-0.1211	0.0651	-0.7376	-1.8592	0.9243

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0623	52.50%	40.11%	0.7246

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	6	0.0986	0.0164	4.2375	0.9949
Residual (Error)	23	0.0892	0.0039		
Total	29	0.1878			

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	Indiv R-Sqr/Model R-Sqr	Variance Inflation Factor
LotSz	82.38	1.5691	5.7
Wt	89.42	1.7032	9.5
T1$/lb	62.70	1.1943	2.7
AC	81.68	1.5556	5.5
MISS	86.80	1.6532	7.6
AVI	86.88	1.6547	7.6

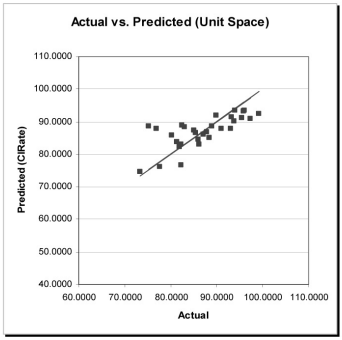

Figure 5. Regression results for the specified model.

Regression Analysis of a Non-Product-Specific Model

Figure 6 summarizes the results when the model was estimated using the same dataset, with the product-related indicator variables excluded. In the absence of these parameters, the degree of multicollinearity decreased substantially. Generally, the remaining technical parameters (average lot size, T1$/lb, and weight) showed good significance and stabil-

I. Model Form and Equation Table

Model Form:	Weighted Log-Linear model
Number of Observations Used:	30
Equation in Unit Space:	CIRate = 108.4 * LotSz ^ (-0.02691) * Wt ^ (-0.006937) * T1$/lb ^ (-0.03189)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	4.6857	0.0735		63.7613	1.0000
LotSz	-0.0269	0.0070	-0.6522	-3.8679	0.9993
Wt	-0.0069	0.0061	-0.2077	-1.1366	0.7340
T1$/lb	-0.0319	0.0118	-0.4512	-2.6946	0.9878

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0638	43.70%	37.20%	0.6611

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	3	0.0820	0.0273	6.7269	0.9994
Residual (Error)	26	0.1057	0.0041		
Total	29	0.1878			

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	Indiv R-Sqr/Model R-Sqr	Variance Inflation Factor
LotSz	0.24	0.5457	1.3
Wt	0.35	0.8037	1.5
T1$/lb	0.23	0.5207	1.3

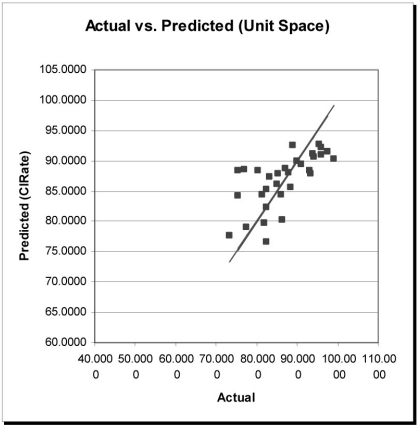

Figure 6. Regression results for specified model without product indicator variables.

ity with little multicollinearity between them. The overall predictive power of the model improved slightly, as indicated by a higher F-statistic; however, the adjusted coefficient of determination declined, and the standard error of the estimate remained essentially the same.

By decreasing the multicollinearity, excluding the product-specific indicator variables improved the quality of the model. Nevertheless, the fit of the model to the data using the remaining technical parameters was still not especially good. Further analysis showed that the exclusion of a small number of data points would improve the model fit substantially. However, with the exception of the previously-discussed MLRS data point, none of these additional possible outliers were judged to be anomalous to the extent that exclusion could be justified. At this stage, although the value of parametric analysis to predict cost improvement rates appeared promising, it could not be said that this had been demonstrated with a robust model. In addition, although the product-specific indicator variables did not significantly enhance the fit of this particular model to the data, the value of product-specific analysis of the phenomenon was not disproved.

Chow Test on Product Data Subsets

In a further effort to objectively gauge whether cost improvement should be modeled separately by product, a Chow Test (Pindyck, 1998, 133) was conducted on each product subset of the data. For each product, the following statistic was calculated,

$$F_{k,\, N+M-2K} = \frac{(ESS_R - ESS_{UR}) / k}{ESS_{UR} / (N + M - 2k)}$$

Where:

ESS_R = Error sum-of-squares of restricted model
ESS_{UR} = Error sum of squares of unrestricted model
$N + M$ = Number of observations (30)
k = Number of estimated coefficients (4),

which is F-distributed when the coefficients of the unrestricted and restricted models are jointly equal. The restricted model was the model shown in Figure 6, since its coefficients are in actuality restricted to being equal for all product subsets of the data. The error sum-of-squares of the unrestricted model (ESS_{UR}) is the ESS of an identically-specified product-specific model with degrees of freedom $N - k$, plus the ESS of the same model estimated using the rest of the data with degrees of freedom $M - k$. The null hypothesis (H_0) tested by this statistic is:

H_0: The coefficients of the restricted and unrestricted models are equal, i.e., that they are the same model.

The results of this test for each product regression are shown below in Table 3.

	N	M	ESS (N)	ESS (M)	F	p
Spacecraft	6	24	0.0019	0.0781	1.7669	0.1715
Missile	8	22	0.0011	0.0816	1.5296	0.2282
Avionics	11	19	0.0450	0.0250	2.8050	0.0507
Aircraft	5	25	0.0006	0.0917	0.7985	0.5390
Pooled Data	30		0.1057			

Table 3. Chow test results.

While the test does not conclusively reject the null hypothesis at the usual 0.05 or 0.10 levels of significance, for three of the four products there is strong evidence that the product models are based on different populations. For the spacecraft model, the test result could be expected only 17% of the time if the spacecraft data were in actuality consistent with the pooled data. For the missile model, the result could be expected with only 23% probability, and for the avionics model, with only 5%. The aircraft data appear to be consistent with the pooled data; however this result is based on the smallest sample size.

In the next part of the research, further analysis was focused on the separate product

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	6
Equation in Unit Space:	CIRate = 181.4 * LotSz ^ (-0.05212) * Wt ^ (-0.07736)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	5.2005	0.2155		24.1319	0.9998
LotSz	-0.0521	0.0116	-1.1424	-4.5104	0.9796
Wt	-0.0774	0.0265	-0.7407	-2.9246	0.9388

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0254	87.23%	78.71%	0.9340

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	2	0.0132	0.0066	10.2438	0.9544
Residual (Error)	3	0.0019	0.0006		
Total	5	0.0151			

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	Indiv R-Sqr/Model R-Sqr	Variance Inflation Factor
LotSz	33.63%	0.3856	1.51
Wt	33.63%	0.3856	1.51

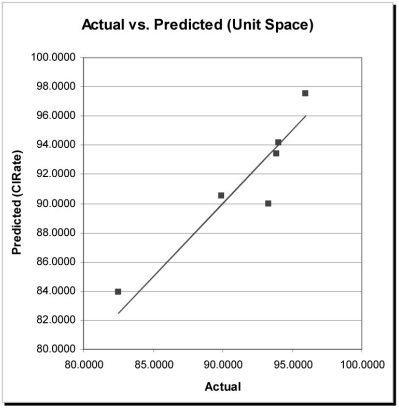

Figure 7. Regression results for spacecraft model.

data subsets. Regression models were independently fit to each of these subsets, without the restriction of using the same specification or independent variables for all.

Regression Analysis of Product-Specific Models

In the development of separate models by product type, each model became a separate analysis. That is, each could have its own theory of causality, functional specification, and combinations of product-specific parameters. For the purposes of the exploratory analysis, however, the original range of parameters related to scale, complexity, and design stability was used in the separate models. The chief problem encountered was the limited

I. Model Form and Equation Table

Model Form:	Weighted Log-Linear model
Number of Observations Used:	8
Equation in Unit Space:	CIRate = 97.6 * Wt ^ (-0.01754) * T1$/lb ^ (-0.0334)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	4.5809	0.0424		108.1210	1.0000
Wt	-0.0175	0.0051	-1.0248	-3.4508	0.9818
T1$/lb	-0.0334	0.0076	-1.3041	-4.3913	0.9929

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0168	79.51%	71.32%	0.8917

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	2	0.0055	0.0027	9.7027	0.9810
Residual (Error)	5	0.0014	0.0003		
Total	7	0.0069			

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	Indiv R-Sqr/Model R-Sqr	Variance Inflation Factor
Wt	53.54%	0.6733	2.152389152
T1$/lb	53.54%	0.6733	2.152389152

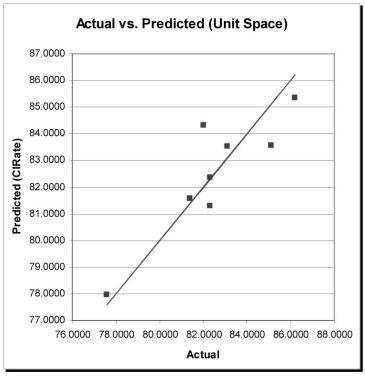

Figure 8. Regression results for missile model.

degrees of freedom available in the data subsets. Six spacecraft data points were available, eight missiles, eleven avionics programs, and only five aircraft. For this reason, models with three parameters were judged to be unrealistic and possibly misleading. At most two parameters were used in the estimation of the product-specific models. Figures 7 through 11 summarize the results.

The fit of the product-specific models to the data is closer than for the models based on all the data, with substantially higher adjusted R^2 values, and, except for the avionics model, lower standard errors. The scatterplots reveal the sparseness of the data, however, sometimes resulting in highly influential single data points. Definitive models of cost improvement for these products would clearly need to be supported by more data. Figure

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	11
Equation in Unit Space:	CIRate = 101.2 * T1$/lb ^ (-0.08403)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	4.6169	0.0476		97.0537	1.0000
T1$/lb	-0.0840	0.0240	-0.7598	-3.5059	0.9933

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0723	57.73%	53.03%	0.7598

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	1	0.0642	0.0642	12.2915	0.9933
Residual (Error)	9	0.0470	0.0052		
Total	10	0.1112			

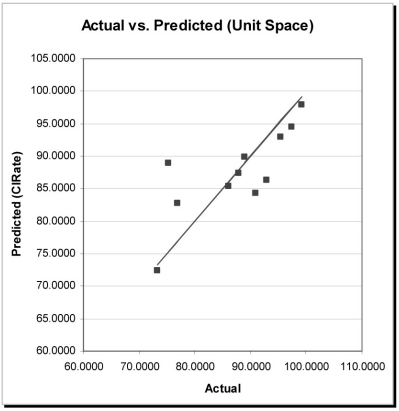

Figure 9. Regression results for avionics model.

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	5
Equation in Unit Space:	CIRate = 91.58 * T1$/lb ^ (-0.08329)

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	1-p
Intercept	4.5173	0.0159		284.9011	1.0000
T1$/lb	-0.0833	0.0184	-0.9342	-4.5359	0.9799

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.0264	87.27%	83.03%	0.9342

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	1-p
Regression	1	0.0144	0.0144	20.5745	0.9799
Residual (Error)	3	0.0021	0.0007		
Total	4	0.0165			

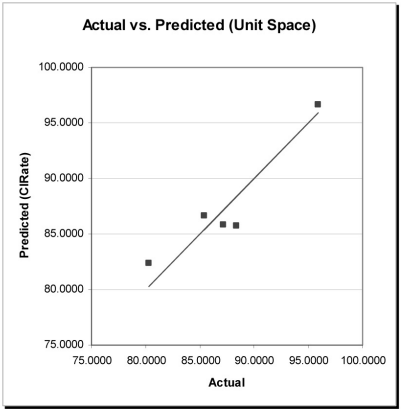

Figure 10—Regression Results for Aircraft Model

	Specified Model (Fig. 5)	Non-Product Specific Model (Fig. 6)	Spacecraft (Fig. 7)	Missiles (Fig. 8)	Avionics (Fig. 9)	Aircraft (Fig. 10)
R2	0.53	0.44	0.87	0.80	0.58	0.87
R2 (Adj)	0.40	0.37	0.79	0.71	0.53	0.83
Standard Error	0.06	0.06	0.03	0.02	0.07	0.03
F-Statistic	4.24	6.73	10.24	9.70	12.29	20.57
1 - P(F)	0.99	1.00	0.95	0.98	0.99	0.98
Number of Obs	30	30	6	8	11	5
Regression DOF	6	3	2	2	1	1

Figure 11. Summary of model results.

11 below summarizes for comparison the statistics of the principal models estimated in the exploratory analysis.

OBSERVATIONS AND FINDINGS

The purpose of the research was to investigate the feasibility of predicting cost improvement in future production programs using parametric methods. Toward this end, three research questions were posed.

- Can a theory of causality be incorporated into the design of CIRs?

While this is an unavoidably subjective question, it is a necessary one when investigating whether statistically-based parametric relationships can be applied to an empirical phenomenon that has not been previously modeled. Any statistical relationship can only be said to have predictive qualities to the extent that it incorporates a theory of causality that makes sense to the users of the relationship. This research proposed that the potential for production cost improvement can be encompassed by three dimensions; product scale, product complexity, and design stability. This scheme seemed to have predictive power with respect to cost improvement rates, both for the models based on the pooled data and for the product-specific models. The fact that this was observed in a relatively small dataset indicates that future research efforts using more extensive and representative data could find success in modeling more detailed descriptive theories.

- What independent variables can be used in CIRs?

In CERs, the validity of using technical and performance parameters to predict the development or production costs of products is usually not questioned. However, at this early stage of thinking about CIRs, it is worthwhile asking whether "cost improvement drivers" can be found. In this research, although the use of weight to represent product scale was straightforward, trying to exemplify product complexity and design stability was more problematic. Dollars-per-pound was settled on as the surrogate for complexity, while the average lot size was used for design stability. While the analysis showed that there was a degree of statistical correlation between these parameters and the cost improvement rates of diverse aerospace products, whether or not they actually represent the desired aspects of the causality theory remains a subjective analytic judgment. As research and consensus-building moves forward on how to embody cost improvement in statistical relationships, a key area of discussion and agreement clearly must be on the identification of appropriate parameters.

- Are independent variables related to product type (e.g., satellite, aircraft, avionics, missiles) as predictive as those related to the production environment (level of automation, design stability, etc.)?

When this question was posed at the outset of the research it was anticipated that CIRs would look like the Specified Model shown in Figure 5. This model incorporated a combination of technical parameters and product-related indicator variables. It was assumed that if the indicator variables were not significant, then this would be evidence that cost improvement could be modeled without regard to product type. The analysis showed, however, that all the variables in the Specified Model were significant, albeit multicollinear. The exclusion of indicator variables in the "non-product specific" model shown in

Figure 6 reduced the multicollinearity, but the technical variables alone did not result in a close fit to the data. At this point, the results were inconclusive about the value of product type as a predictor of cost improvement.

The results of the Chow Test conducted on the product-specific subsets of the data indicated that neither the Figure 5 nor the Figure 6 models might be appropriate. It indicated that the products could be independent in terms of the intercepts and slopes of the models to be fit to their data. The relatively good fit statistics of the product-specific models shown in Figures 7 through 10 provided support to this hypothesis, although these models were based on fewer observations than is desirable.

The research question was intended to address the value of product-related indicator variables, and the analysis results must be considered inconclusive with respect to this issue. However, this inconclusiveness led to additional analysis providing indications that product type is indeed an important element of cost improvement analysis, not as a stratifying variable, but as a subject for separate analysis.

LIMITATIONS OF THE RESEARCH

The limitations of this work derive chiefly from its exploratory nature. In any cost research, data collection and normalization are the most time consuming tasks, and it was in these areas that the effort described here was chiefly constrained. In order to streamline the process of dataset development, previously conducted studies were drawn on to compile data on 31 production programs; six spacecraft, nine tactical missiles, 11 avionics systems, and five fighter aircraft. A more conclusive effort clearly would require a broader range of data in each product area.

Despite the diverse sources, efforts were made to insure consistency of the data across product types in terms of computational approach (cumulative average) and the use of a single base year for the dollar related parameters (FY2002). However, only limited information was available to insure that the content of the unit costs to which the improvement curves were fit was consistent in terms of the treatment of direct and indirect costs, overhead, and so on. Future research efforts on this topic need to address the content of each data point in detail, as well as the strength and consistency of the cost improvement trends in each program.

Finally, future efforts should go beyond the limited number of independent parameters available in this study for prediction of cost improvement rates. Parts counts, numbers of workers per shift, levels of automation, numbers of engineering design changes, and so on could be useful parameters to investigate.

CONCLUSIONS

This exploratory analysis was aimed at investigating whether a parametric approach to predicting production cost improvement is feasible. Keeping in mind the previously discussed limitations of the research, it appears that CIRs would represent an improved methodology for estimating cost improvement for future programs. The model specification and parameters investigated in this research resulted in significant fits to the cost improvement data, particularly when the data was segregated by product type. Furthermore, although the product-specific models were hampered by fewer degrees of freedom, the improvement in the fit statistics for these models compared with the pooled data was significant. The research indicates that formulating cost improvement models within the

context of aerospace product types would probably yield the best results. An additional advantage of this approach is that it allows for the use of functional specifications and parameters that would be appropriate to the production environment, technology, or physical characteristics of the product. Nevertheless, the research did not demonstrate that non-product-specific models are infeasible. It may be that more general models, especially those that combine product types that have much in common, may have merit. Techniques such as the Chow Test could assist researchers in identifying products that should be modeled jointly. It is through more parametric analysis of the unit cost improvement phenomenon that cost analysts will be able to move beyond the current subjective methodologies to a more theory-based, quantitative treatment.

Cost research aimed at extending the work begun in this study should focus on the following:

- Discussion and formulation of an improved theory of cost improvement that can be incorporated into CIRs
- The compilation of a larger and more consistent database of aerospace programs that have experienced cost improvement, along with a wider array of possible descriptive parameters is clearly needed to support more definitive parametric analysis of cost improvement.
- Formulation of CIRs to provide better predictors of future cost improvement than the current approaches used by the cost analysis community

ACKNOWLEDGEMENT

The author would like to thank and acknowledge the contributions of the reviewers to the quality and usefulness of the research. The reviewers made important suggestions that improved the presentation of the research and challenged its conclusions. A key criticism by them led to additional research showing that parametric analysis of cost improvement within the context of product types could have significant predictive value. The author is indebted to the reviewers for broadening the scope of the analysis and ultimately strengthening its conclusions.

REFERENCES

Air Force Systems Command (AFSC). 1986. *AFSC Cost-Estimating Handbook Series*, prepared for AFSC by The Analytic Sciences Corporation, Reading, Massachusetts.

Asher, H. 1956. *Cost-Quantity Relationships in the Airframe Industry,* R-291, The Rand Corporation, Santa Monica, California.

Baloff, N. 1966. The Learning Curve—Some Controversial Issues. *The Journal of Industrial Economics* (July), pp. 275–282.

Book, S. and E. Burgess. 2003. A Way Out of the Learning Curve Morass: Quantity as an Independent Variable (QAIV). Presentation to ISPA Southern California Chapter Workshop, International Society of Parametric Analysts (January).

Conway, R. and A. Schultz 1959. The Manufacturing Progress Function. *Industrial Engineering* (January–February), pp. 39–54.

Crawford, J. No Date. Learning Curve, Ship Curve, Ratios, Related Data. Lockheed Aircraft Corporation, Burbank, California.

Levy, F. 1965. Adaptation in the Production Process. *Management Science* (April), pp. B-136—B-154.

Nadler, G. and W. Smith. 1963. Manufacturing Progress Functions for Types of Processes. *International Journal of Production Research* (June), pp. 115–135.

Pindyck, R. and D. Rubinfeld. 1998. *Econometric Models and Econometric Forecasts,* Irwin McGraw-Hill, Boston.

Robbins, J. 1986. *Cost Improvement Analysis*, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

Tecolote Research, Inc. 2004. *COSTAT User Guide*, Santa Barbara, California

Teplitz, C. 1991. *The Learning Curve Deskbook,* Quorum Books, New York.

Wright, T. 1936. Factors Affecting the Cost of Airplanes. *Journal of Aeronautical Science* (February), pp. 1–10.

Young, S. 1966. Misapplications of the Learning Curve Concept. *Journal of Industrial Engineering* (August), pp. 410–415.