3-24-2016

A Comparative Study of Learning Curve Models and Factors in Defense Cost Estimating Based on Program Integration, Assembly, and Checkout

Brandon J. Johnson

A COMPARATIVE STUDY OF LEARNING CURVE MODELS AND FACTORS IN DEFENSE COST ESTIMATING USING PROGRAM INTEGRATION, ASSEMBLY, AND CHECKOUT

THESIS
MARCH 2016

Brandon J. Johnson, Captain, USAF

AFIT-ENV-MS-16-M-162

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-16-M-162

A COMPARATIVE STUDY OF LEARNING CURVE MODELS AND FACTORS IN DEFENSE COST ESTIMATING BASED ON PROGRAM INTEGRATION, ASSEMBLY, AND CHECKOUT

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

Brandon J. Johnson, BS

Captain, USAF

March 2016

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENV-MS-16-M-162

A COMPARATIVE STUDY OF LEARNING CURVE MODELS AND FACTORS IN DEFENSE COST ESTIMATING BASED ON PROGRAM INTEGRATION, ASSEMBLY, AND CHECKOUT

Brandon J. Johnson, BS

Captain, USAF

Committee Membership:

Dr. John Elshaw, Ph. D
Chair

Lt Col Brandon Lucas, Ph. D.
Member

Mr. Michael Seibel
Member

AFIT-ENY-12-M-05

Abstract

The purpose of this research was to investigate the flattening effect at tail end of learning curves by identifying a more accurate learning curve model. The learning curve models accepted by DOD are Wright's original learning curve theory and Crawford's Unit Theory. The models were formulated in 1936 and 1944 respectively. This analysis compares the conventional models to contemporary learning curve models in order to determine if the current DOD methodology is outdated. The results are inconclusive as to if there is a more accurate model. The contemporary models are the DeJong and S-Curve and they both include an incompressibility factor, which is the percentage of the process that includes automation. Including models that incorporate automation was important as technology and machinery plays a larger role in production. Wright's model appears to be most accurate unless incompressibility is very low. A trend for all models appeared. The trend is Wright's curve was accurate early in production and the contemporary models were more accurate later in production. Future research should have an objective of finding a heuristic for when the models are most accurate or comparative studies including more models.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Dr. John Elshaw for his guidance and support throughout the course of this thesis effort. The insight and experience was certainly appreciated. I would, also like to thank my sponsor, Mr. Mike Seibel, from the Life Cycle Management Center for both the support and latitude provided to me in this endeavor.

Capt. Brandon Johnson, USAF

Table of Contents

Page

Abstract iv

Table of Contents vi

List of Figures viii

List of Tables ix

I. Introduction 1

Purpose and Overview 1

Background 3

Problem Statement/Research Objectives 6

Investigative Questions 7

Methodology 8

Assumptions/Limitations 8

Preview 9

II. Literature Review 10

Chapter Overview 10

Description of Learning Curves 13

Relevant Research 22

Summary 26

III. Methodology 27

Chapter Overview 27

Data Collection 28

Learning Curve Models 30

Research Hypothesis 32

Method of Analysis 33

Conclusion..35

IV. Analysis and Results..35

Chapter Overview..35

Results of Simulation Scenarios..36

Results of Parameters..37

Results of Means Comparisons...40

Results of Sensitivity Analysis...42

Summary...47

V. Conclusions and Recommendations..48

Chapter Overview..48

Conclusions of Research...48

Significance of Research...52

Recommendations for Future Research...53

Assumptions and Limitations..55

Summary...55

Appendix A...57

Appendix B...58

Appendix C...58

Appendix D...59

Bibliography..60

Vita 63

List of Figures

Page

Figure 1: Air Force Program Actuals 12

Figure 2: Cumulative Unit Theory Learning Curve 14

Figure 3: Cumulative Unit Theory Log Scale 14

Figure 4: S-Curve Learning Model 19

Figure 5: WLC Frequency Distribution 39

Figure 6: APE Trends 43

Figure 7: All Data APE Trends 46

List of Tables

Page

Table 1 Cumav and Unit Cost Comparison 15

Table 2 Moore's F-15 Analysis 31

Table 3 Initial Analysis Parameters 37

Table 4 MAPE Analysis M=0.05 38

Table 5 Descriptive Statistics 40

Table 6 KW Test 41

Table 7 KW Test at M=0.10 42

Table 8 Parameters for All Data Points Analysis 44

Table 9 MAPE Values for All Data M = 0.05 44

Table 10 Descriptive Statistics All Data 45

Table 11 KW Test All Data 46

Table 12 MAPE Values M = 0.10 All Data 47

A COMPARATIVE STUDY OF LEARNING CURVE MODELS AND FACTORS IN DEFENSE COST ESTIMATING USING PROGRAM INTEGRATION, ASSEMBLY, AND CHECKOUT

I. Introduction

Purpose and Overview

Department of Defense (DOD) acquisition programs have frequent cost overruns that exceed billions of dollars every year. The effect of the overruns is detrimental to the DOD and limits the United States Government's ability to operate efficiently. Inaccurate cost estimates play a large role in the stated cost overruns. The final cost of a program when compared to the baseline, show DOD programs historically have underestimated costs. Cost estimates allow the DOD to assess the future of programs and the role that cost plays. The government budgets are based these cost estimates. The current state of the DOD includes shrinking budgets and large funding cuts for acquisition programs. An extra emphasis on and scrutiny of accurate cost estimates are the result of the current cuts and budget issues. There is a new standard for Financial Managers and Program Managers. They have to support and maintain a cost estimate like no time before. A balanced budget is a goal for the DOD, and the budget depends on the cost estimate. Gone are the days when the DOD had ever-growing budgets where the fiscal mentality was to spend. The fiscal mentality now involves saving and receiving as much value as possible for every budgeted dollar. In order to obtain reliable cost estimates, cost estimating models and tools within the DOD present the opportunity for an evaluation on their accuracy. The current learning curve methods within the DOD's cost estimating procedures are from the 1930s. As automation and robotics increasingly replace human touch-labor in the production process, a model that is 80 years old, and assumes constant learning, may no longer be appropriate for accurate learning curve estimates. Robotics and automation do not learn and they are inevitably a part of future production. New learning curve methods are available for cost estimators to utilize. The modern learning curve methods could be a useful tool for obtaining better cost estimates for the DOD. The purpose of this research is ultimately to investigate new learning curve methods, develop the learning curve theory within DOD, and pursue a more accurate model.

This thesis will examine whether different learning curve models are more accurate than Wright's Learning Curve model (the status quo) when comparing actual values to predicted. The current DOD learning curve methodology does not take into account available information and factors that contribute to learning. The information and factors are not taken into account largely because the status quo models are outdated and do not account for many parts of the learning processes. The theory of forgetting is a concept from contemporary learning curve models and it may provide an answer to providing accurate cost estimates. Forgetting is human error or the tendency to not show constant learning over time (Badiru, 2012). The point of emphasis for this research, and issue that needs to be resolved is that DOD agencies need to estimate more accurately. Prior thesis work and research on this subject shows that the learning curve methods show room for further development. The prior research found that models that are more accurate potentially exist. There is a chance to incorporate alternate learning curve models and more DOD programs into this area of research. Research found that an important factor (incompressibility) was not explicitly researched or known. An assumption on the incompressibility factor was the route of analysis for the certain learning curve models that incorporated it. Towill and Cherrington defined incompressibility as the percentage of the learning process that is automated (1994). Robotics and automation are not going away and will likely play a larger role in the future. Research on what that factor actually equals or how it relates to different airframes could be critical for obtaining a more precise model. Using integration, assembly, and checkout (IA&CO) processes instead of complete touch labor processes should provide an analysis that is more insightful and potentially lead to a more accurate model. IA&CO are specific work that occurs during production. Further chapters will provide more detail and explanation of IA&CO. The reason for this significance is that production, in general, is moving towards a more automated process. Machines are starting to take over parts of production. This research on the specific part of production assumes that humans will be doing the work. Since the IA&CO is a highly manual process, human learning and forgetting occurs. The rest of Chapter 1 includes background, a problem statement, objectives, investigative questions, methodology, and assumptions and limitations. Lastly, the conclusion will provide an overview of the remaining chapters of this comparative study.

Background

The learning concept itself is the theory that as a worker performs a task repeatedly, he or she will require less time to complete the same task due to familiarity with the process (Moore, 2015). The learning concept and learning curve models have been applied to manufacturing for over 70 years. From government to private industry, learning curve methodology has been applied and adopted. The earliest learning curve model, Wright's Model, is a mathematical representation that illustrates as the quantity produced doubles, a worker's performance will increase at a constant rate (Wright 1936). Tasks tend to be completed faster with practice; the task completion increase in speed is not surprising because we have all seen this and it is intuitive (Ritter, 2001). People and the organizations that include human labor tend to learn and complete tasks better and more efficiently over time if the tasks are repetitive in nature. Under a repetitive task, and certain conditions, a pattern that is usable for estimating time or costs emerges. If the conditions for learning vary or are not repetitive in nature, learning curve formulations and patterns do not emerge. One of the first and most recognizable learning curve formulas is y=ax^b. This is referred to as the Unit Learning Curve Model,

Equation 1:

$$y = ax^b \tag{1}$$

Where

y = the estimated production hours or cost
a = the production hours of the theoretical first unit
x = is the unit produced
b = is a factor of the learning = log *R* (learning rate)/log2

The theoretical first unit, a, is determined by analysis on obtainable historical data and is not to be confused with the actual first unit cost (T1) or hours. Chapter 2 will explain the model in further detail based on unit and cumulative average concepts that were developed. Cumulative Average theory is the same equation but different definitions for the variables "y" and "x." Cumulative average theory is attributed to Wright, but the equation differs in that "y" represents cumulative average costs and "x" represents cumulative average units.

The Air Force methodology on learning curves and guidance in their application is found in Chapter 8 of the *Air Force Cost Analysis Handbook* (AFCAH) and chapter 17 of the DOD *Basic Cost Estimating Guidebook* (BCE). These two guides focus on two theories: unit theory and cumulative average theory. The unit theory, equation 1, predicts a specific unit cost. Cumulative average theory focuses on the average of all units produced up to a certain point in the production process.

Wright's Model has been the standard in manufacturing. However, recent research has shown other models may provide a more accurate predictor of cost. Moore analyzed learning curve models and provided the foundation for this research. His research showed that two alternative learning curve models, the S-Curve and DeJong, were more accurate when an incompressibility factor was assumed to be somewhere between in the range 0.0 and 0.06 for production on the F-15 fighter aircraft. The research highlighted that results were inconclusive as to which model is the most accurate and whether or not the level of automation assumption was valid (Moore 2015). The equations and factors will be explained in further detail the equations in chapter 2. There are multiple learning curve models in present day production. The S-Curve and DeJong Models were models in the comparative study that emphasize the concept of forgetting in the form of automation. A 2013 *Journal of Aviation and Aerospace Perspective* Article titled "Half-Life Learning Curve Computations for Airframe Life-Cycle Costing of Composite Manufacturing" emphasized a forgetting concept that relates learning curves methodology. The forgetting concept includes the idea that workers experience forgetfulness and an associated decline in performance with time instead of constant learning. Even as a worker is making progress with time, the learning can decrease (Badiru, 2012). This decrease in learning can be as simple as forgetting processes over time. The article added, "Contemporary learning curves have attempted to incorporate forgetful components into learning curves" (Badiru et al, 2013). Forgetting theory was the concept and basis for Moore's (2015) research into learning curve models. Forgetting was the basis because Wright's Model does not account for it. Moore attempted to demonstrate that the S-Curve and DeJong were better predictors of manufacturing hours (or costs) than conventional models. Wright's Learning Curve, the Stanford B, S-Curve, and DeJong were the models used in Moore's research. The research was able to incorporate one aircraft for comparison. Incorporating other aircraft at the time of research did not happen because of multiple data issues on other fighter aircraft. The research found that there is evidence to support a hypothesis that a more accurate model exists. Forgetting and the future of production using a more automated manufacturing process could provide insight into cost estimating. The idea is to provide cost estimates that are more accurate. This thesis, sponsored by the Air Force Life Cycle Management Center (AFLCMC), will expand upon prior research, and attempt to research a more accurate model that can provide a more accurate cost estimating method for the DOD.

Problem Statement/Research Objectives

The problem is that there is flattening effect near the end of production runs, learning does not remain constant in aircraft production, and machinery is becoming more involved in the production process. Prior research sought to answer why the flattening effect happens, but was based on only one aircraft and a certain factor in the learning curve models was assumed. The research needs development and further comparisons in order to capture breadth. Through the efforts of Moore, there is evidence to support a hypothesis that a different model may be more accurate than Wright's model. Moore's research found that the contemporary models are more accurate than Wright's model given an incompressibility factor (M) that is somewhere between 0.0 and 0.1. M is a number between zero and one where zero indicates a completely manual process and one indicates a fully automated process. Wright's model is the most accurate predictor of cost if M is assumed to be greater than 0.1. Prior research results as to which alternative learning curve model is the more accurate were inconclusive (Moore, 2015). Results did not support nor disprove a change from Wright's Model based on one aircraft, the F-15. Prior research provides the foundation for further research into additional types of aircraft. Specifically further research and analysis using program integration, assembly, and checkout. Additional research on the impressibility factor may indicate a model that is more applicable to DOD methodology. Additional learning curve models that do not rely on an assumed M may provide strong results as well. Using all of the data points, every airframe produced instead of a lot data, should prove insightful and may be more robust than using lots. Lot data is typical for DOD programs. Further chapters will explain why lot data is typical for DOD programs. The unit data refers to actual individual production or manufacturing units whereas lot data refers to data for multiple units for an entire production. Lot data is harder to see what is happening in the learning process because there are less checkpoints.

Investigative Questions

The following investigative questions are the basis for this thesis.

1. How does the application of learning curve models using program integration, assembly, and checkout data effect learning curve models that incorporate an incompressibility factor?
2. How sensitive are IA&CO data to the incompressibility range?
3. Which learning curve model is the most accurate at predicting cost or time for IA&CO?
4. How can the individual airframe work codes prove beneficial for predicting cost or time?

Methodology

The method for this thesis begins by collecting data from the Air Force Life Cycle Management Center Cost Staff on Airframe, Integration, Assembly, and Checkout. Data are essential to predict costs and research the incompressibility factor. The basic method is to use learning curve models to compare actual costs to predicted costs. A comparison of actual costs to predicted costs will enable a calculation of error and further analysis of variance. Lastly, this research will attempt to determine a good estimate for an incompressibility factor through application into different models. The alternative models, the DeJong and S-curve, will be predictors of cost or hours for the Air Force Program. The Air Force Program is a multi-mission aircraft used for intelligence surveillance and reconnaissance. The alternative models will be compared to Wright's Model (Cumav or Unit based on which provides a more accurate estimate) in order to calculate a percent error. The error is the comparison point for accuracy. The comparison will use an Analysis of Variance (ANOVA) and potentially a Kruskal-Wallis means test if the results are not from a normal distribution. Chapter 3 will explain the methodology in detail.

Assumptions/Limitations

If one or more of the models is found to have a mean residual value (average difference between the observed values and the predicted values) that is significant, the next step is to determine which model is the best predictor of actual production costs ("Residual Analysis in Regression," n.d.). The smallest mean total for a model is reflective of the most accurate model. The incompressibility factor, explained further in Chapter 2, must have an accurate range in order to provide a more accurate model. Assuming the factor lies within a certain range could provide an estimator a basis for including an alternative model in an estimate. Data should provide constants and factors for model incorporation. Later chapters provide the assumptions on the factors and constants in the equations. The scope of this thesis will focus on aircraft production hours within the Air Force.

Preview

The goal of this thesis is to answer the research questions, expand prior research, and determine if contemporary learning curve models more accurately predict costs than the conventional models used. The analysis of data will compare the contemporary models to the conventional models (status quo) used in the DOD. Provided there are significant results, there is potential for adopting new learning curve methods. The next chapter will deliver an in depth look into literature surrounding learning curve concepts, the associated incompressibility factor, [and the influence of machinery in aircraft production]. Specifically, Chapter 2 will assess Air Force and DOD costs estimates. Chapter 2 will address how learning curves play a role in the cost estimation. Chapter 2 provides in depth explanations of the contemporary learning curves models. The following chapter, Chapter 3, will identify the methodology. The methodology provides the basis for, data collected, how the data application will work, and statistical techniques. Chapter 4 highlights the results from the analysis (tests and methodology) described in Chapter 3. Descriptions of results and graphical explanations are the basis of Chapter 4. The results of the analytical tests provide the basis for Chapter 5, which provides the impacts of the results, the conclusions, and implications for cost estimation.

II. Literature Review

Chapter Overview

The purpose of this chapter is to summarize previous published research as appropriate to learning curve methodology. As stated in Chapter 1, the current spending on DOD programs is declining. The Budget Control Act of 2011 set to DOD spending and sought to reduce the deficit by $1.5 trillion over a ten-year period. Since 2010, there have been three iterations of the Air Force's *Better Buying Power (BBP), which* focuses on efficiencies and affordability within the Air Force. BBP is an initiative from the Office of Acquisition, Technology and Logistics that pursues continuous improvement being the best approach to improving DOD acquisitions (Kendall, 2014). One pillar of BBP 3.0 is to achieve dominant capabilities while controlling lifecycle costs. A sub-pillar is to strengthen and expand "should cost" based cost management (Kendall, 2014). The concept of should cost based management includes identifying goals for cost reduction and implementing efforts to achieve cost reductions (Mueller, 2011). Essentially BBP 3.0 is attempting to achieve the efficiency of DOD acquisitions. Kendall states this by saying, "progression from BBP 1.0 to 2.0 reflected a change in emphasis from best practices to an increased emphasis on helping acquisition professionals think critically and make better decisions as they confront a myriad of complex situation encountered in defense acquisition" (Kendall, 2014). Finding a more accurate learning curve model (also known as an experience curve) for predicting cost can make the DOD acquisition process more efficient and control life cycle costs. The success of the budget control act relies on controlling costs. Controlling costs does not mean lowering the costs of programs; it means the DOD estimate will not have a better picture of what a program will actually cost. The DOD has a budget and the estimates drive where the budgeted dollars end up (which programs get what). Budgeting accuracy is achievable through accurate cost estimates.

Many learning curve models exist and learning curve theory has grown and evolved over time. However, the DOD practice and methodology has seen little change or adaptation to current technological trends. It is known that over time a worker and organizations learn (find efficiencies and process improvements) when performing repetitive tasks (Malashevitz & Williams, 2010). Efficiencies and improvements over time equate to the learning curve methodology. From learning curve methodology, the concept of forgetting emerged. Forgetting is the phenomenon that a worker's performance may fluctuate or decline with time. Forgetting, unlike Wright's theory that assumes constant learning, there is unlearning in the production process. Moore (2015) uses the example of the production of an automobile: "While the processes and parts are always the same, a savvy car buyer may want to avoid cars that were built on a Monday or Friday. The worker and even the entire assembly line may suffer a loss in performance due to working at the beginning or end of the week" (Moore, 2015. 10). It is human nature for people to lose focus or concentration at certain times when performing repetitive tasks. An example of this loss of focus would be a production line worker doing the same task repeatedly. Eventually that person might slow down because of a daydream or boredom. The same idea for the production of cars can occur within any production. A worker will show forms of forgetting and will not show constant learning over time. There is visual evidence of forgetting when plotting cost or hours to complete tasks. The forgetting is evident when the tail end of the learning curve does not exhibit a constant slope. The learning curve's slope changes, showing a flattening, or decrease in learning. The change might be an effect attributable to forgetting. Figure 1 highlights that constant learning was not the case for the Air

Force Program. The hours to complete individual units actually in some instances show an increase.

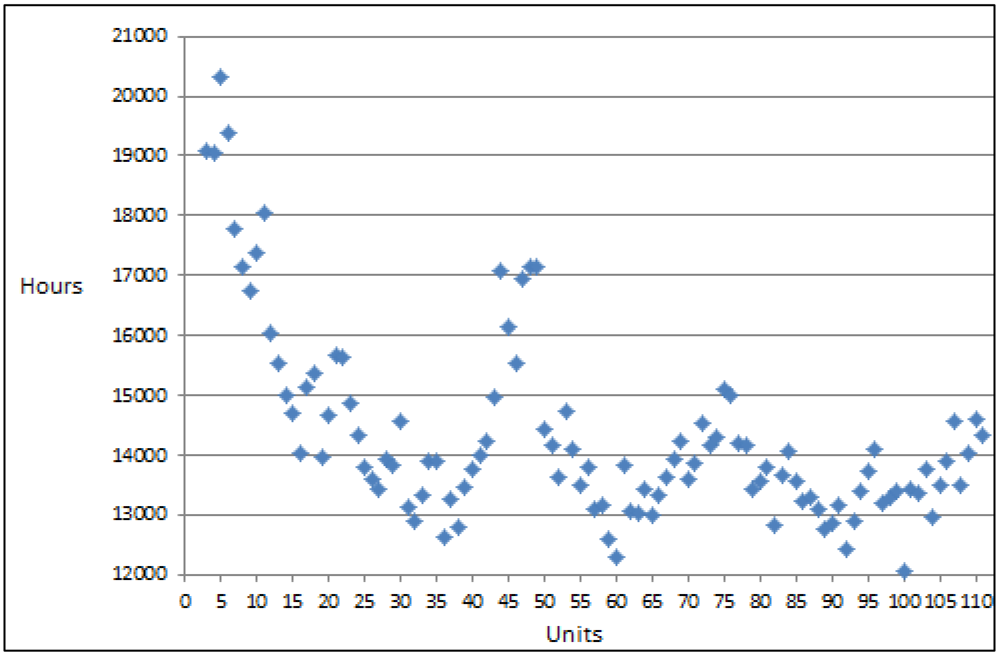


Figure 1 Air Force Program Actuals

The actuals leave a ground for questioning whether the flattening effect at the tail end of production attributable to forgetting or the influence of machinery. One might be able to hypothesize that machinery, which does not learn may have caused the flattening effects at the tail end of the curves. With the combination of organizational forgetting theory and automation, a more robust model may be present.

This chapter summarizes research appropriate to learning curve methodology. Chapter II includes concepts of forgetting theory, appropriate learning curve models, if and how private industry utilizes learning curves, and the background on the incompressibility factor. This chapter will provide background on contemporary learning curve theories, summarize prior research on learning and forgetting theories, investigate machinery and technology involved in production, limitations, and conclude with a description of methodology.

Description of Learning Curves

Theory

The theory of learning curves dates back over 70 years, when Theodore Paul Wright (T. P. Wright) first identified a relationship between a worker's performance on a single task and the time required to complete that task. The time to complete a task decreases at a constant rate over time. In 1936, T. P. Wright annotated the effect of learning on production costs in his article about the aircraft industry, "Factors Affecting the Cost of Airplanes." Wright developed the learning curve mathematical model and stated it in his article (Wright, 1936). The best way to think Wright's model is as the unit doubles from 1-2, 2-4, 4-8, etcetera, learning will happen at a constant rate. In the 1970s the Boston Consulting Group found that various industries experienced a range from 10 to 25 percent learning curve effect (Hax & Majluf, 1982), which was in line with what Wright observed in the aircraft industry years earlier (20 percent learning curve effect).

The effect of learning is evident by examining the slope of the line over time. Figure 2 is a graphical representation of Wright's learning curve. Figure 2 is an example with the first unit costing 100 per unit. Based on 90%, 80%, and 70% learning, the graph highlights when the number of units doubles, the *average* production cost decreases by 10%, 20%, and 30% respectively. The direct costs decrease because they are a function of direct labor costs. As the time to complete a unit decreases, so will the costs. The average is the sum of the values divided by the number of units (the mean).

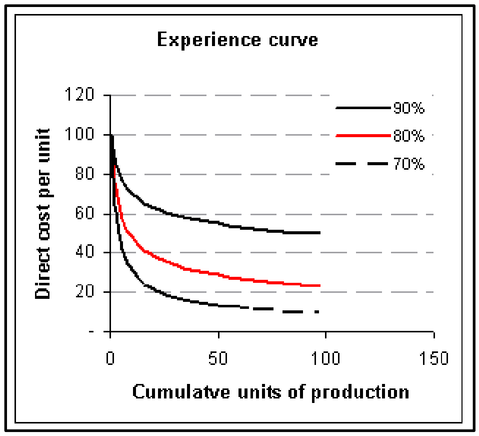


Figure 2: Cumulative Unit Theory Learning Curve (Wikipedia, 1936)

Figure 3 highlights the log scale learning curve rate. This log relationship results in a linear relationship for the experience curve. The log relationship is the result of an algebraic manipulation. The algebra enables regression analysis; the transformation makes the data linear. Practitioners find the log relationship useful in their efforts to best fit a straight line to the algebraically manipulated data.

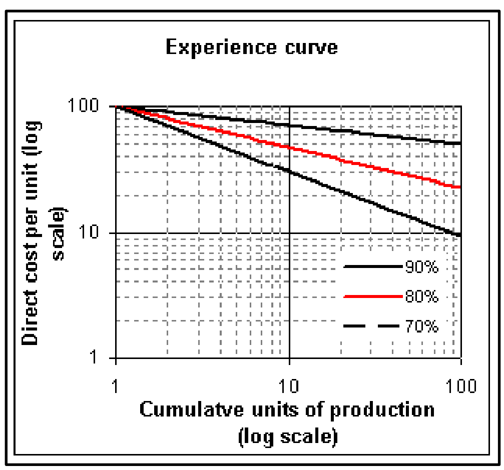


Figure 3: Cumulative Unit Theory Log Scale (Wikipedia, 1936)

The log scale's straight line for Wright's Model provides the basis that a constant learning rate over time is a straight line, and thus learning is constant. The most flexible and appropriate transformation is the logarithmic (LOG (base 10)) or the Natural LOG (LN (base e)) of hours and quantity (Malashevitz & Williams, 2010). Logarithms are a method to reduce long mathematical calculations. Graphing the values on log-log paper as seen in Figure 3 (or by mathematical calculation of the logs) is evident. A line, characterized as having a constant slope, emerges from this method. Slope is defined as change in Y given a change in X where Y=mX+B (equation of a straight line). A figure in log space highlights the slope and it relates it to a constant learning percentage. Wright's model originally took into account cumulative average costs, where the average cost is the average cost of the units. Crawford was able to create a model based on Wright's model and created an equation (Equation 1 found in Chapter 1) to predict a cost for a specific unit. The cumulative average equation is similar, but uses averages. For example, consider a hypothetical 80% learning rate with a theoretical first unit cost of $100. The second unit would cost $80. Whereas the Cumav cost would be the cumulative average of the first two units. The average of $100 and $80 equals $90. Cumav thus will always result in a more conservative estimate than the unit curve. Table 1 highlights the cumav and unit cost comparisons below.

Table 1 Cumav and Unit Cost Comparison

Unit	Unit Cost	Cumav Cost
1	$ 100.00	$ 100.00
2	$ 80.00	$ 90.00
3	$ 70.21	$ 83.40
4	$ 64.00	$ 78.55
5	$ 59.56	$ 74.75
6	$ 56.17	$ 71.66
7	$ 53.45	$ 69.06
8	$ 51.20	$ 66.82
9	$ 49.29	$ 64.88
10	$ 47.65	$ 63.15

The DOD typically receives reports from contractors based on lot or unit data. It is more common for a contractor to provide lot data, however it is far preferable for the estimator to have unit data. Having unit data provides the estimator with actual costs or hours and helps them avoid a Lot Midpoint or Lot Plot Point determinations. Based on the data available to DOD cost estimators, the unit curve method is not readily used. Regression is the primary method for predicting learning curves using lot data. The lot is the number of units in a certain production run. If the data are in lots, and the lots are consecutive, the approach is to use the last unit in the data as the plot point. The last unit of the lot approach is the Lot Plot Point (LPP) approach. LPP is similar to plotting individual units in order to determine learning from lot to lot over time (Hu and Smith, 2013). The Lot Midpoint (LMP) captures the unit curve and is the unit that corresponds with average unit cost for the lot. This is simply the lot average unit cost. The reason lots usage typically occurs in cost estimation is that there are limits on the availability of information. In general, the military procures equipment (planes, tanks, satellites, etc.) in lots greater than one and as a result receives cost data in the form of dollars or hours by lot and not on a unit basis (Everest, 1988). Inquiries show that defense acquisitions typically utilizes the cumulative average cost curve because of receiving data in lots Unit data is considered superior to Cumav data because individual points highlight learning and forgetting over time. An average (CUMAV) will not display learning/forgetting as readily because of a smoothing effect. By using an average, the curve is smoother and the effects of learning are less apparent (All, 2013).

If a program or estimator is able to obtain unit data in the form of actual hours there may be room to incorporate the unit curve. Having all of the proprietary data on a weapon system may prove beneficial to utilizing a unit cost curve. The learning curves may prove more accurate than using the LPP or LMP [this assumes there was not a major production break]. The trends are much more evident for hours than costs and lots. Hours provide a constant homogenous comparison from year to year. Dollars on the other hand can feels the effects of external factors such as inflation. For example, an hour in 1985 is equivalent to an hour in 2010. Lots include multiple airframes that are sometimes not equivalent. For example, a program could purchase 10 units in the first lot and 85 in the next. The actual labor hours for different work codes gives this analysis grounds to test the specific data. Loss of learning, machinery, and automation analysis using IA&CO may provide profound and insightful learning curve results.

A description of labor categories Integration, Assembly, and Checkout highlights that not all definitions are synonymous amongst manufacturers. The manufacturers largely consider what is involved in each category as proprietary information. For the purpose of this study, Final Test Integration, Electrical and Mechanical Assembly, Test/Integration, Composites (all locations), and Quality Control are considered IA&CO. Final Test Integration includes the direct labor for the final integration and test, which includes final assembly, system burn-in, payload integration and interface, autopilot checks, taxi tests, range tests and first flight support. Electrical Assembly is the direct labor required to assemble electronic components. Mechanical Assembly is the direct labor required to build servos for the aircraft, to build landing gear, build starter/alternators, to perform rework, and high-time maintenance on those components. Test/Integration is the direct labor for new build electronics, field repairs, integrating avionics, and testing them at the system level. Composites manufacturing is the direct labor required to lay up, cure, and finish components such as the fuselage, wings, tails, and landing gear. Quality Control is the direct labor required to provide inspection of electronics and mechanical components and assemblies, document discrepancies, and resolve problem areas (“Labor

Categories_2.pdf," n.d.). Of note, these labor categories involve mainly direct labor performed by humans where the learning process is observable.

A contemporary variation of learning curve models is DeJong's Learning Formula. This formula is a modification of Wright's model, and it takes into account the constant, M. M is the incompressibility factor, which is a constant between zero (fully manual operation) and one (fully automated or machine dominated operation) (Badiru ET. Al, 2013). Equation 2 highlights DeJong's Learning Model.

$$t_n = t_1\{M + (1 - M)n^{-b} \tag{2}$$

Where:

t_n = the cumulative average time after producing n units
t_1 = time required to produce theoretical first unit
n = cumulative unit number
b = log R/log 2 (learning index)
M = incompressibility factor (a constant)

A machine based production process would result in no learning, and thus an M value of one. It the basis of this thesis and belief that aircraft production, complex in nature, has an M value close zero because aircraft production is a highly manual process. Thus, M would be closer to zero for IA&CO. M does not have a specific value. This research will focus on the best M value for the particular aircraft production. A potential weakness of the DeJong model is that it does not take into account previous units produced much as the S-Curve model does.

The S-Curve Model takes into account both previous units produced and the incompressibility factor. Figure 4 below shows the effects of learning over time as hypothesized from the S-Curve Model. The linear nature of Wright's original learning curve model has been in question for many years (Everest, 1988). The Rand Corporation first sought to explain the progression of the learning curve used to estimate costs for both military and civilian airframes.

The report made an attempt to describe the relationship between units (quantity) and costs, and ultimately whether the relationship was linear on a log scale. The results of the Rand Corporation found that a convex curve may provide more accuracy if producing a large number of units (Asher, 1956). The results found that a convex model provides less error if there is a need for large extrapolation. Essentially, an estimation of significantly more units in production instead of fewer provides less error. For units where large extrapolation is the circumstance a non-linear model was more appropriate. The S-Curve model, convex in nature, presents a shape of learning. The S-Curve, when plotted on using a log scale relationship follows an S function. The experience over time (attempts at learning) may exhibit the S-Curve (Everest, 1988).

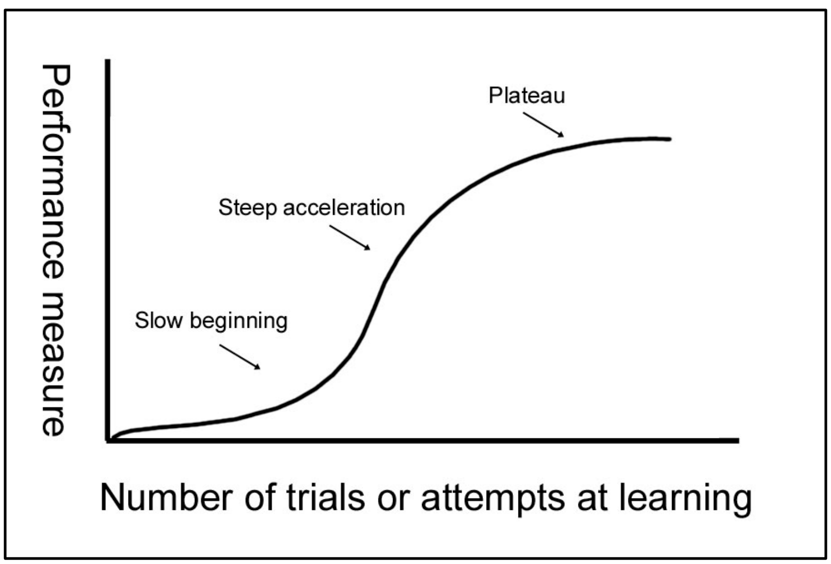


Figure 4: S-Curve Learning Model (Dewey, n.d.)

Initially there is a slow beginning as a worker learns the production process. The newness of the product is a characteristic of the slow initial beginning. New tools, methods, shortages of parts, reworks, and the challenge of developing a cohesive production team are all potential contributors to the slow beginning. The fact that the initial stage in production deals changes from tooling to even workers contributes to the gradual start (Badiru, 1992). From there, learning and familiarity of tools, methods, and workers occur. The learning enables a steep acceleration of production. Production improvement occurs with attempts on the process, or learning by doing. An example from literature is aircraft production. Aircraft production that includes workers that are more efficient, more efficient tools and organization results in an assembly that is more efficient. The efficiencies found result in less time to complete an aircraft (Asher, 1956). However, the improvement and efficiencies eventually begin to fade. The plateau at the trailing edge of the curve is the *slope of diminishing returns* where the curve begins to flatten out or in many occurrences at the end of production cycles, there is a "tailup" (Everest, 1988). After time inefficiencies can occur: forgetting, experienced workers focusing on new projects, failure to repair worn tooling at the normal rate, increase of machine disassembly, lack of key materials (safety stock), and workers taking more time to prolong their employment (Everest, 1988). The S-Curve equation is shown in Equation 3:

$$t_n = t_1 + M(n + B)^{-b} \tag{3}$$

Where:

t_n = the cumulative average time after producing n units
t_1 = time required to produce theoretical first unit
n = cumulative unit number
b = log R/log 2 (learning index)
M = incompressibility factor (a constant)
B = equivalent experience units (a constant)

From this equation and figure 4, the forgetting concept is evident. The S-Curve portrays that with time, some inefficiencies will occur. Use of the S-Curve and DeJong Models may provide more precision to learning curve and enable higher accuracy within DOD cost estimating because they include previously unaccounted for influences.

Many learning curve models exist. Badiru's (1992) paper *Survey on Univariate and Multivariate Learning Curve Models* that reiterates the abundance of models, "Learning in the context of operations management, refers to the improved efficiency obtained from repetition of a production operation. Workers learn and improve by repeating operations" (Badiru, 1992). Badiru listed the most notable univariate (one variable) models, which are perhaps the simplest forms. In regards to this work, the learning curves are both univariate and express the dependent variable (cost or hours) in terms of the independent variable (cumulative output). The first univariate model was Wright's Learning Curve. Other univariate models that Badiru identified: Stanford-B, S-Curve, DeJong's, levy's Adaptation Function, Glover's Learning Formula, Pegel's Exponential Function, Knect's Upturn Model, and Yelle's Product Model (Badiru, 1992). The models chosen for this research include the incompressibility factor which was found as very important in previous research, "results indicate that the S-Curve and DeJong models offer improvement over current estimation techniques, but more importantly-and unexpectedly-highlight the importance of incompressibility (the amount of a process that is automated) in learning curve estimating (Moore, Elshaw, Badiru, Ritschel, 2015). Multivariate models have the potential to examine numerous factors that could influence learning. They have not received the attention they deserve due most likely to the nature of their complexity.

Another potential problem is multicollinearity (Alchian, 1963). The definition is multicollinearity is where two or more independent variables (predictor variables) are highly correlated. From this correlation, multiple regression coefficient estimates may change substantially in response to small changes of variables. Essentially the predictor variables might be redundant ("Multicollinearity in regression," n. d.). Alchian experimented with multivariate models, specifically the Cobb-Douglas Multiplicative Model. The heavily researched Cobb-

Douglas model and other multivariate models exist. The chance of Multicolinearity is high. All of this considered this research has a goal to make the cost estimators jobs easier. Going from Wright's Model to a multivariate model where the estimators have to find variables and ask for contractor data will likely make the work more complicated and tedious. Learning curve model complexity adds analysis time for an analyst and there is no evidence to support a complex model provides any benefit. The models that taken into account incompressibility could provide an answer to a more accurate model. Ultimately, understanding the forgetting curve, the influence of automation, and relevant research on incompressibility factors may provide an answer to improved cost estimating.

Relevant Research

Prior research tested three models and compared the results to Wright's model. The results were inconclusive as to whether there is a more accurate model available. Moore did have some insights "results do provide some insight…findings also emphasize the importance of incompressibility in the learning process" (Moore, 2015). Moore found that slight changes in the assumed incompressibility within the DeJong and S-Curve Models lead to drastically different results as to which model is the most accurate. In addition to the analysis, Moore also found that incompressibility might affect learning models more than prior experience units. Moore's research showed that the DeJong and S-Curve models were more accurate. They were more accurate than Wright's model at predicting of cost when compared. This was under the assumption that the incompressibility was low. Because prior research did not provide a conclusive result based on the assumed factor value, there is a demand for continued research. Further research into more programs could prove beneficial to finding a more accurate learning curve model. Additionally, research into incompressibility and machinery for production could prove beneficial for learning curve methodology. Assessment on how changing incompressibility is difficult and it could wildly change results could also prove beneficial. Other relevant research has compared learning curves; however, there are limited comparative studies in defense cost estimating.

Relevant research has highlighted an important point in why military programs have not adapted a contemporary learning curve model, “Because of the regularity of production in military programs, organizational forgetting, and spillovers of production experience are less apparent. If forgetting is present, it may be very difficult to identify (e. g., data could be consistent with either a 20 percent learning rate or a 25 percent learning rate with 5 percent forgetting). And, in most cases there are not many model variants, so spillovers are not important” (Benkard, 1999, p. 4). The newest fighter weapon in the US military arsenal will be the F-35. The F-35 has 3 variants and the Pentagon plans to spend over $390 billion on these aircraft (Luce, 2014) Five percent of $390 billion attributed to learning/forgetting processes is still a staggering number. The point is that there is room for improvement. Many of the fighter aircraft in use today have had multiple models. The F-15 had models A-E and the F-16 had models A-F. The DOD can use the hypothesized five percent forgetting to save millions of taxpayer’s dollars. The accurate estimates result in less spending, or savings that could go into other taxpayer needs or public works. The estimates enable the DOD to truly forecast a budget and spending levels.

So one may consider question what does the future actually look like in regards to machinery and automation. Are people still going to be relevant for production? In the Defense acquisitions realm the basis is that with low purchase quantities for state of the art machines will not rely on technological or machine dominated production. This idea really comes down to the machine verses machines argument. Asking whether a robot with take the jobs of people is key. Experts say yes and no. In the past, machines were used to replace manual labor that is an intensive and repetitive task. According to a study by the Bank of America, robots are likely to be performing 45% of manufacturing tasks by 2025, versus 10% today (Madigan, 2011). The price of a computer, a robot, a chip, etc. is falling and it is speculated that they will fall even more in the future. However, the jobs that require human interaction are least likely to be replaced by a robot. Maybe DOD acquisitions are in the clear and 100% of learning is still realizable for learning curve methodology. Experts agree that the future does include a significant presence of machinery because the prices of robotics and computers are decreasing while the cost of a human employee is increasing (Aeppel, 2012).

One method that is popular amongst program offices within the DOD is the Rate Effect for estimating costs. The principal idea behind the Rate Effect is modeling a production rate decrease due to a decrease in personnel-related expense. The reason behind the decrease is that fixed costs, when spread over fewer units is less expensive. The opposite happens for an increase in production. Personnel-related expenses and costs generally decrease. The Rate Effect, first proposed by RAND in 1974 attempted to incorporate production rate changes in the learning curve model (All, 2013). The result was a modified learning equation within the unit theory that incorporates a rate adjustment. Two key variables to the learning curve model: R and c are now a part of the model shown below in equation 4.

Equation 4:

$$y = ax^b R^c \tag{4}$$

Where

y = the estimated production

hours or cost
a = the production hours of the theoretical first unit
x = is the unit produced
b = is a factor of the learning = log R (learning rate)/log2
R = annual production rate
C = log2 (rate slope)

The Rate Effect is popular amongst Program Offices, but unpopular with oversight organizations such as the DOD because the rate term is rarely found to be significant (All, 2013). Another reason for this models unpopularity is the effect of multicollinearity and its likelihood for multivariate models. Multicollinearity refers to a correlation between dependent variables, meaning the variables are predicting the same thing. The rate effect could play a more prevalent role in the future if machinery becomes more prevalent. Oversight offices such as OSD Cost Assessment and Program Evaluation (CAPE) may start to consider Rate Effect Model if humans are less involved in the production process.

The principal idea behind the future incorporating more machinery and levels of automation in production is that people are expensive. If companies, who work for profit, want to increase profit, an easy route is to have low skilled employees. An employer would have to pay high skilled employees more than the low skilled employees did. If you have high skilled employees they will be expensive and thus the price you have to charge for the product you are selling will have to increase in order for the business to have the same level of profit. The definition of profit is revenues minus cost. The definition of revenue is the price charged multiplied by the quantity sold. The idea behind increasing profit is that a company should have the lowest skilled employees needed to complete the production process. Most employees offer a broad range of skills that sometimes are not necessary on a production run. The simplest, lowest skilled employee would have to be a machine, because for the life of the machine costs will be low (Godin & Warner, 2010). Machinery relatively has a high initial cost, but the machine will repeat every task and the company does not have to pay for its medical insurance or retirement. The idea that machinery will dominate the future is highly likely; in the immediate future, machines or robotics could be working next to humans. Prior generation machines might have been limited to the a certain skill and people to service them, the future of machinery in production will be cheaper and could provide a renaissance ("Making the future | The Economist," n.d.). Machines have replaced workers in production lines and people answering telephones. In my opinion, competitive pressures encourage organizations to turn their workers into machines. Machines can be measured and if the organizations can measure it, they can produce faster and increase profit (Godin & Warner, 2010).

Summary

Chapter 2 provided a summary of previously published research as appropriate to learning curves. Learning theory and modeling is a known process in production and plays a role in cost estimating. The forgetting theory, automation, and the learning curve models that take forgetting/automation into consideration have not been adapted into the DOD cost estimating process. Research on the models that take forgetting into account and the incompressibility factor are the basis of this thesis. The methodology for this thesis follows in Chapter 3 and it will describe data, methods, and assumptions.

III. Methodology

Chapter Overview

The purpose of this chapter is to explain the methods behind the research that modern learning curve models provide better accuracy than conventional models when estimating production costs using program integration, assembly, and checkout data. Modern learning curve models take into account factors such as automation that conventional learning curve models do not. Conventional models assume a constant learning rate over time, which seems like an unrealistic assumption. Intuitively, learning fluctuates and does not remain constant because humans tend to forget and machines do not learn. Contractors are looking to lower expenses because lowering expenses leads to more profit. From large organizations that are comparable to large DOD contractors, it is known that people are expensive; total human capital costs (cost of workforce) average approximately 70% of expenses (Higgins & Cooperstein, 2010). One way to lower costs is to have a machine be a part of the process. This move towards automating processes is the future of production. Production estimates within the DOD utilize learning curve models. As previously stated, DOD budgets for acquisition programs are shrinking. Cost estimates associated with acquisition programs are a high priority within the DOD. Finding methods to increase the accuracy of the estimates are of value. They are particularly of value to estimators because increased accuracy leads to less error. If a modern learning curve model is more accurate than what is currently being used (Wright's model), it has the potential to add significant value to the cost-estimating field. The method is to test which learning curve model is the best predictor of cost. Wright's learning curve equation, $t_i = t_1 x^b$ does not take into account the performance decline of forgetting or the influence of machinery. Using program integration, assembly, and checkout data as a basis for this research is useful because it involves manual processes that will have a low incompressibility factor. The hypothesis is that using this level of data instead of all of the production data, will demonstrate that the modern learning curve models are more accurate predictors of cost than the status quo. Essentially, the analysis will be more sensitive to program integration, assembly, and checkout data in hours than the entire lot data in dollars. The reason behind this is having individual unit labor hours will provide an accurate picture of learning, forgetting and the influence of machinery. Using the specific work codes identified, and hypothesized to have more forgetting/learning processes involved, should have results that are more sensitive because more manual processes are involved.

Comparing the modern learning curve models to the conventional method is essential in determining which is more accurate. The models chosen are relevant to this research because they incorporate incompressibility and the application of only IA&CO can be input. Moore states “the conventional method lacks the application of key factors that affect learning: prior experience and incompressibility. Accounting for these factors can reduce the amount of estimating error for airframe costs" (Moore, 2015, p.14). The incompressibility factor played a large role in his research. The S-Curve and DeJong model were more accurate predictors of cost than Wright’s model. Those models only exhibited more accuracy if incompressibility was low. Using integration, checkout, and assembly may make this assumption valid. Chapter 3 clarifies how the application of the models, methods for comparison, and data analysis.

Data Collection

The Air Force Life Cycle Management Center Cost Staff (AFLCMC/FCZ) at Wright-Patterson Air Force Base (WPAFB) provided learning curve data for 17 Major Defense Acquisition Programs (MDAPs). A MDAP, classified as a major program that exceeds a certain dollar threshold. There are 80 MDAPs in the DOD as of 2014. The numbers have decreased slightly over the years. The data files consist of average Learning Curve Reports of Annual Unit Cost (AUC) in addition to the MDAPs estimate methods using Wright's conventional learning curve model. Only one program provided was broken into the specific work codes that include the needed data (IA&CO). That program was the for the purpose of this study, an Air Force Program. The data are primary data that is proprietary. Primary data are data that is from the contractor and has not been normalized, inflated, or altered (Ellis, n.d.). When comparing models based on airframe's integration, assembly, and checkout data the assumption of incompressibility close to zero is reliable. The airframes for this analysis are the Air Force Program. Moore researched one fighter aircraft, the F-15, and found incomplete or too few data on other airframes to make a valid comparison. It is important to have data span several years for comparison purposes. The historical data is important because the effects of learning are evident. Having data spread over a period of time shows patterns. The Air Force Program data are adequate and spans Fiscal Years (FY) 2005 to FY 2013. If data from AFLCMC and the Air Force Program are not adequate, the Joint Cost Analysis Research Database (JCARD) system may have cost data for programs or asking for specific Work Breakdown Structure level IV data from a contractor may be a route to take.

Once data compilation is complete, the data will have to be standardized. Standardization will occur by converting prior year's values into a Base Year (BY) to take into account the effects of inflation. This standardization assumes the data are in dollars and not hours. If the data are in hours, no conversion is necessary. Comparing the cost of a program in 1980 dollars and the cost of program in 2000 dollars will not suffice. Every year the Office of the Secretary of Defense (OSD) publishes OSD Inflation Tables. Values for this research will be standardized to Base Year 2015 (BY$15) using the tables. This conversion will display dollars that are relevant to present time. Once data conversion to BY$15 for all programs is complete, the next step will be plotting the average unit costs and the cumulative average unit costs. Trends might be evident for the production such as initial learning, period of peak performance, and then diminishing returns or forgetting. In order to highlight the data points, a log-log graph will provide the visual representation. The log-log graph is a two-dimensional graph of numeric data that uses logarithmic scales on both axes. Relationships appear in a straight line on the graph which is useful in estimating relationships and parameters (Ritter, 2001)

Learning Curve Models

Wright's model, which takes the form of $t_i = t_1 x^b$, has parameters t_1 and b that need to be determined in order to obtain an estimate. The statistical method for determining these parameters is linear regression. Linear regression is an approach for modeling relationships between variables. The natural log of a cumulative units (x) against their associated costs (y) will provide t_1 and b. The regression line that explains the most variability will determine whether unit cost or cumulative average theory is the best method for predicting cost. Determining the regression that explains the most variability comes from a comparison of the line's R^2 values. R^2 is a goodness of fit percentage the represents variance between independent and dependent variables (McClave, Benson, and Sincich 2011). Besides R^2 there is an adjusted R-squared which is an adapted description of R-squared that adjusts for the number of predictors in the model. Essentially, the adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors (Frost, 2013). R-Squared is sufficient for this analysis because the number of predictors are low. From the regression, b is the slope of the line and t_1 is the natural log of the y-intercept. With these two parameters solved, they remain constant for the other models.

The DeJong Model and S-Curve models are the modern models considered for comparison in this research. The incompressibility factor in Dejong's model highlights the effects of learning. M is the constant between zero and one where zero is a fully manual development and one is a fully machine development. Aircraft production, specifically the program integration, assembly, and checkout portion, is close to fully manual. The S-Curve Model in addition to *M* uses a previous unit experience constant, *B*. Incompressibility is undefined and led to subjectivity in Moore's research (Moore 2015). Moore tested the F-15 assuming a low *M*. Moore tested the models in increments of .05 from zero to 0.20. Moore's research hypothesis was that one or more of the modern learning curve models would have a mean average percent error (MAPE) significantly different from the conventional model. He found that when the incompressibility factor was zero to 0.05 the DeJong and S-Curve Models were statistically different and ultimately more accurate than Wright's Model. However, for all other values there was no statistical difference or Wright's Model was more accurate. Table 2 below highlights these findings.

Table 2 Moore's F-15 Analysis

	M =0.0	*M* =0.05	*M* =0.10	*M* =0.15	*M* =0.20
WLC	*N/A*	*N/A*	*N/A*	*N/A*	*N/A*
Stanford-B	*X*	*X*	*X*	*X*	*X*
DeJong	*X*	-	*X*	+	+
S-Curve	*X*	-	*X*	+	+

X indicates model is not significantly different from WLC
(+) indicates model is statistically less accurate than WLC (Higher MAPE)
(-) indicates model is statistically more accurate than WLC (Lower MAPE)

Research Hypothesis

This comparative study's theory is that using program integration, assembly, and checkout data will provide a more realistic assumption for low incompressibility and ultimately a more accurate predictor of actual costs based on a modern learning curve model. The hypothesis is that the MAPE is significantly different between the predicted labor hours for the modern models when compared to the conventional model. MAPE is the measure of variation that takes the error both positive and negative as an average. The positive and negative errors, taken as absolute values, or the average distance from the actual values is the measure. Small MAPE values relate to values that are more accurate. Therefore, smaller MAPES are better. The research hypotheses based on comparing the models using program integration, assembly, and checkout data are as follows:

H1: One of the compared models will have MAPE significantly different from the others.

H2: One of the modern learning curve models will be significantly more accurate than Wright's model in predicting costs or hours.

H3: The S-Curve model with IA&CO will have a significantly lower MAPE than Wright's and DeJong's Learning Curve Models.

H4: The incompressibility factor will have a significant influence on the accuracy of the DeJong and S-Curve Models

The first hypothesis' null (H_0) is $\mu_1 = \mu_2 = \mu_3$. This means that the MAPE for each learning curve model are the same. The alternative hypothesis (H_a) is that one of the model's MAPE is statistically different. A rejection of null hypothesis in favor of the alternative hypothesis supports significant finding. The significant finding means that testing each contemporary learning curve model against Wright's model is the next phase. The second hypothesis (H2) has a null $\mu_1 = \mu_i$. Where i will equal models 2 and 3 (DeJong's and the S-Curve). The H_a is that the contemporary learning curve models will have a lower MAPE than the conventional model ($\mu_1 > \mu_i$). H3's null hypothesis is $\mu_2 = \mu_3$. The H_a is that $\mu_2 > \mu_3$ meaning that the S-Curve will have the lowest MAPE and thus be the most accurate predictor of cost or hours. The last hypothesis is that small changes in the incompressibility factor will have a large influence of the MAPE of each model.

Method of Analysis

A Microsoft Excel Spreadsheet is the means to housing the data collected. Standardization (if needed) to BY$15 averages if it is in dollars is the second step. The spreadsheet will include actual costs and the predicted costs using one of the learning curve models. Once calculation of the predicted costs is complete, the error is simply the difference between the actuals costs and the predicted costs. To provide a comparison, a difference calculation in the absolute value and absolute value percent error are the means of analysis. The next step is to perform an analysis to test the hypotheses. ANOVA or Kruskal-Wallis test [with IBM® SPSS statistics software] with Microsoft Excel will provide the basis for comparing the percent errors. The tests will produce an F-statistic (a test statistic) that falls within a Chi-distribution and a p-value. This comparative study will produce results based on a 95% confidence level (an $\propto$ of 0.05). In order for the test to result in significant findings, the test's p-value has to be less than 0.05. If the P-value is less than 0.05, rejection of the null hypothesis in favor of the alternative hypothesis will occur. Rejecting the null hypothesis for this study will represent that there is a 95% chance that the tested populations are different. The conditions for ANOVA are as follows: the samples must be from a random selection of the population, normally distributed, and population variances must be equal (McClave, Benson, Sincich 2011).

Chapter 4 will state the conditions for this ANOVA. If the conditions for ANOVA fail to meet the needed criteria, the Kruskal-Wallis test (non-parametric equivalent to ANOVA) will be the test to determine if multiple samples arise from the same distribution and have the same parameters (Kruskal & Wallis, 1952). The ANOVA or Kruskal-Wallis f-test provide insight into the first hypothesis. The Kruskal-Wallis test is beneficial since the one-way ANOVA is usually robust based on the assumptions for ANOVA. The Kruskal-Wallis test becomes useful in particular when group sample strongly deviate from normal (sample size is small and unequal and data are not symmetric) and variances are different (potential outliers exist). The assumptions for the Kruskal-Wallis test are that no assumptions are made about the underlying distribution, however, assume that all groups have a distribution with the same shape, and no population parameters are estimated (no confidence intervals in the data) (Zaiontz, 2015). If the F-statistic is significant, then rejection of the null hypothesis in favor of the alternative that at least one of the sample means is different is the outcome.

The t-Test for two samples test will evaluate the second hypothesis that one or more of the models is a better fit to the data than Wright's Model. The control for this comparison is Wright's learning curve model. Since Wright's model is the control for this study, a comparison to the other model's MAPEs is the method. If the assumption for equal variance is not met, the t-Test for two samples assuming unequal variances will be used. The next analysis that corresponds to H3 will be testing which model is most accurate given significant results for more than one model from the H2. Once again, the paired difference t-test is the next step. A paired difference experiment uses a probability distribution when comparing two sample means and produces a t-statistic that falls within a student-t distribution that can either reject or fail to reject the null hypothesis depending on the desired confidence level (McClave et al, 2011). Lastly, H4 will require reiterations of the tests in order to determine a good estimate for incompressibility factor based on the airframes.

Of note, the reader may question why the means cannot provide the basis for the analysis. This lies in the variation of the means. If the coefficient of variation (standard deviation as a percentage of the mean) is low, the mean may be a good predictor of the better model. However, as a rule of thumb, if the coefficient of variation (CV) is greater than 15% the mean indicates a looser distribution. Analyst would like a tighter distribution with less variability. In practice, a low CV (say, 5%) would indicate that the average (mean) of the cost data is a useful description of the data set. On the other hand, if the CV is much higher (say, greater than 15%), there should be a cost driver in the data set that causes the cost to vary (citation needed). The CV's for the analysis will provide insight into the dispersion of the data points.

Conclusion

The methods for this comparative study are quantitative and require statistical analysis as a means of comparison. If any rejection of the null hypotheses in favor of the alternate hypotheses occurs, the findings could result in a valuable tool for cost estimating with more accuracy. At a minimum, the results can be an opportunity for further analysis. The following section will highlight the results from the analysis. The final chapter will interpret the results and discuss the impacts of the comparative study.

IV. Analysis and Results

Chapter Overview

This chapter highlights the results from the analysis (tests and methodology) described in Chapter III. This chapter answers the research questions by describing the results and highlighting relevant analytical information. This chapter includes graphical representations and in depth description of charts. The Air Force Program provided the initial analysis and investigation into whether or not integration, assembly, and checkout data provide robust results on the comparative learning curve model study. After analysis and results, the limitations, implications, and the future research areas will conclude in Chapter V.

Results of Simulation Scenarios

The first step in the analysis was to determine whether the unit theory or the cumav for the Air Force Program data was the best representation. The next step was a log-linear regression; performed on the early-fielded units of the Air Force Program. The early units, treated as historical values, predicts hours for the remaining Air Force Program units used in the analysis. Both the unit and cumav data were plotted using a scatter plot. The cumav regression resulted in an R^2 (coeffeicient of determination) value of 0.9525. This 0.9525 value means that roughly 95% of the variance for the cumav regression is explainable. The unit curve regression resulted in a R^2 value of 0.8174. Since the cumav theory resulted in a higher R^2value, the unit theory was no longer useful in this study to estimate the remaining points (regression graphs can be seen in Appendix A). The remaining points to that need an estimation are the remaining Air Force Program units produced and provided.

The cumav regression from the early units of the Air Force Program provided the factors and parameters for the equations used in this analysis. The analysis needs to transform the log space results. The theoretical first unit (T_1) is found by taking the log-linear regression's Y-intercept (a) from the equation lnY=lna + blnx and raising it to the constant Euler's number, "*e*". Euler's Number is the base of the natural logarithm (Weisstein, n.d.). The theoretical first unit in hours for this analysis is 29,971 hours. The learning curve slope, based on the equation $e^{b*\ln(2)} = 2^b$. The b value was -0.1726 and resulted in an 88.7% learning curve slope. The remaining factors for the initial analysis were B (prototypes) of two and an M value of 0.05. The table below highlights these initial values.

Table 3 Initial Analysis Parameters

Theoretical First Unit (Hours)	29,971
B	2
M	0.05
b	-0.1726
LCS	88.72%

Results of Parameters

From the initial analysis, the data was used to build a Microsoft Excel equation for all three learning curve models. Each respective model's equation provided a predicted value for each Air Force Program. A comparison for the predicted values to the actual values was the next step. Appendix B shows an example model calculation for the DeJong Model. This resulted in an APE for all of the units. The MAPE of the three models is below. The table highlights that given the specific parameters, the S-Curve and the DeJong Models both have lower MAPE values. Performing statistical analysis is the next step in order to identify if there is a significant difference between the models occurs.

Table 4 MAPE Analysis M=0.05

WLC	DeJong	S-Curve
4.11%	3.00%	2.64%

The coefficient of variation ($CV = \frac{Standard\ Deviation}{Mean}$) calculations were performed as a cross check. The mean may be a good predictor of a better model. However, as a rule of thumb, if the coefficient of variation (CV) is greater than 15% the mean indicates a looser distribution. The CVs for the three models exceeded the 15% rule of thumb. The large CVs indicate that the mean is a poor estimator of a more accurate model. An analyst would like a tighter distribution with less variability when using means. On the other hand, if the CV is much higher (say, greater than 15%), there should be a cost driver in the data set that causes the cost to vary (Dameron, Megan E., n.d.). The CV's for the analysis will provide insight into the dispersion of the data points. In order to compare models further statistical methods were the basis of evaluation.

The next step was to test for normality of the predicted values of each learning curve model. If normality passes, ANOVA will be the method. The assumption of normality was not meet because of kurtosis and skewness. An important part of the statistical analysis is to characterize the variability of the data set. The description of the data includes skewness and kurtosis. Skewness is a measure of symmetry. Typically, the skewness is important because of the lack of symmetry. The distribution is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution (Filliben, 2012). Data when plotted tend to have heavy tails or outliers if there is high kurtosis. Low kurtosis tend to have light tails, or lack of outliers. The histogram is an effective method for showing both the skewness and kurtosis.

Wright's Learning Curve Model (Figure 5) when graphed appears to follow a normal curve. Kurtosis equaled -1.164 and skewness equaled 0.320.

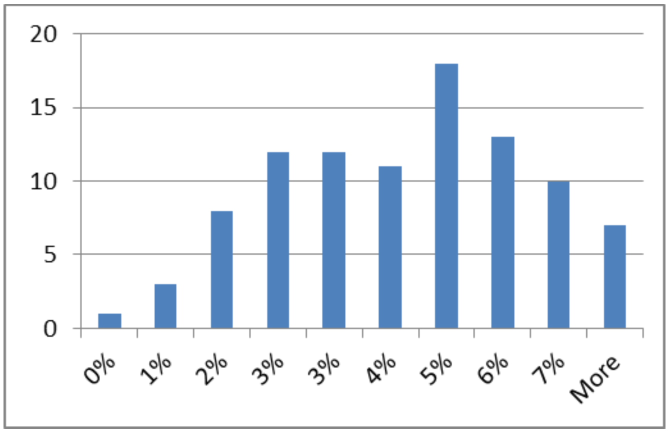


Figure 5 WLC Frequency Distribution

The descriptive statistics provided both of the measures that Table 5 below highlight. The kurtosis absolute values for the DeJong and S-Curve Models are both greater than one. Both of these models have moderate skewness. Appendix C graphically captures the frequency of both the DeJong and S-Curve's distributions. It is evident that the distributions do not follow a normal curve. The Kruskal-Wallis test, used to determine if the predicted samples are significantly different, is the next statistical test.

Table 5 Descriptive Statistics

MLC		Dejong		S-Curve	
Mean	0.041	Mean	0.030	Mean	0.026
Standard Error	0.002	Standard Error	0.003	Standard Error	0.002
Median	0.032	Median	0.020	Median	0.017
Standard Deviation	0.021	Standard Deviation	0.024	Standard Deviation	0.020
Sample Variance	0.000	Sample Variance	0.001	Sample Variance	0.000
Kurtosis	-1.164	Kurtosis	-1.328	Kurtosis	-1.245
Skewness	0.320	Skewness	0.595	Skewness	0.626
Range	0.080	Range	0.069	Range	0.060
Minimum	0.002	Minimum	0.000	Minimum	0.001
Maximum	0.082	Maximum	0.069	Maximum	0.061
Sum	3.904	Sum	2.854	Sum	2.512
Count	95	Count	95	Count	95
Confidence Level(95.0%)	0.004	Confidence Level(95.0%)	0.005	Confidence Level(95.0%)	0.004

The variances are the next statistic to test. This will determine whether performing a T-Test with equal variances is a possibility. The rule of thumb for testing variance is to divide the largest sample standard deviation by the smallest. From the descriptive statistics, S-Curve has the smallest (0.020) and DeJong has the largest (0.024). If this value is greater than one, the assumption of equal variances will not be acceptable. The rule of thumb value equals 1.231 so equal variance cannot be assumed and a T-Test with unequal variances has to be performed to compare sample means.

Results of Means Comparisons

The test for non-normal distributions, the KW test is useful to predict if values are significantly different. The KW test, performed at an M value of 0.05, resulted in significantly different distributions. The null hypothesis was that the distributions of the APE values for the three models was the same based on a significance level of 0.05. The table below shows these results. Since the K value is greater than the critical value, rejecting the null hypothesis that the APE values are same is necessary. Table 6 also highlights that the p-value is less than the 0.05 significance level. The p-value echoes rejecting the null hypothesis.

Table 6 KW Test

K	34.56
Critical Value	5.99
P-Val	3.1E-08

Once there are indications that one or more of the models are significantly different, it is necessary to determine which model captures the difference. We held WLC as the status quo in order to determine if the DeJong Model or S-Curve Model were different. The conclusion for WLC and DeJong t-Test assuming unequal variance: Perform a two-tail test (inequality), if the t Stat < -t Critical two-tail or t Stat > t Critical two-tail, then reject the null hypothesis. T Stat (3.329 is > 1.97) so results reject the null hypothesis. The means are different for WLC and DeJong. The conclusion for WLC and S-Curve t-Test assuming unequal variance: Perform a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, reject the null hypothesis. T Stat (4.912 is > 1.97) so results reject the null hypothesis, the means are different. Because both MAPES for the DeJong and S-Curve model, the next step was comparing the two alternative model. Conclusion: Performing a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. T Stat (1.118 is not > 1.97) so we fail reject the null hypothesis. There is no reason to believe the S-Curve is a more accurate predictor over the DeJong model. The results for The DeJong and S-Curve t-Test resulted in failing to reject the Null hypothesis. As to which model is most accurate, the results were inconclusive. Appendix D highlights all results from the t-Tests. The S-Curve is statistically different from WLC. However, the S-Curve it is not statistically different from the DeJong model. This highlights that WLC is a less accurate predictor with a MAPE value of 4.1% compared to DeJong and S-Curve MAPES of 3.0% and 2.6% respectively. There was not a need for an M value of zero test because at an M value of zero, DeJong's Model reverts to WLC Model.

Results of Sensitivity Analysis

The sensitivity analysis includes changing the M to a value of 0.1 and 0.15. Results dramatically changed by changing the incompressibility factor. The KW test at an M=0.10 is shown below in Table 7.

Table 7 KW Test at M=0.10

K	104.54
Critical Value	5.99
P-Val	2E-23

Since the K value is greater than the critical value, reject the null hypothesis that there is not a difference. The next step was holding WLC as the status quo and it to each of the alternative models. WLC when tested using the t-Test for two samples assuming unequal variance rejected the null hypothesis against the S-Curve and DeJong found there was a significant difference. The two-tail test (inequality) conclusion found that if t Stat < -t Critical two-tail or t Stat > t Critical two-tail, reject the null hypothesis. T Stat (-9.31 is < -1.97) so we reject the null hypothesis. The means are not the same. WLC is more accurate than Dejong and S-Curve. This means that WLC is the most accurate predictor with a MAPE value of 4.1% compared to DeJong and S-Curve MAPES of 7.3% and 6.7% respectively. The results of the APE values highlights that WLC has less error earlier on, but as the sample size increase DeJong and the S-Curve have less APE. Chapter V will provide insight into why the error occurs for WLC and the contemporary models. Figure 6 shows the results at an M value of 0.05 and 0.10.

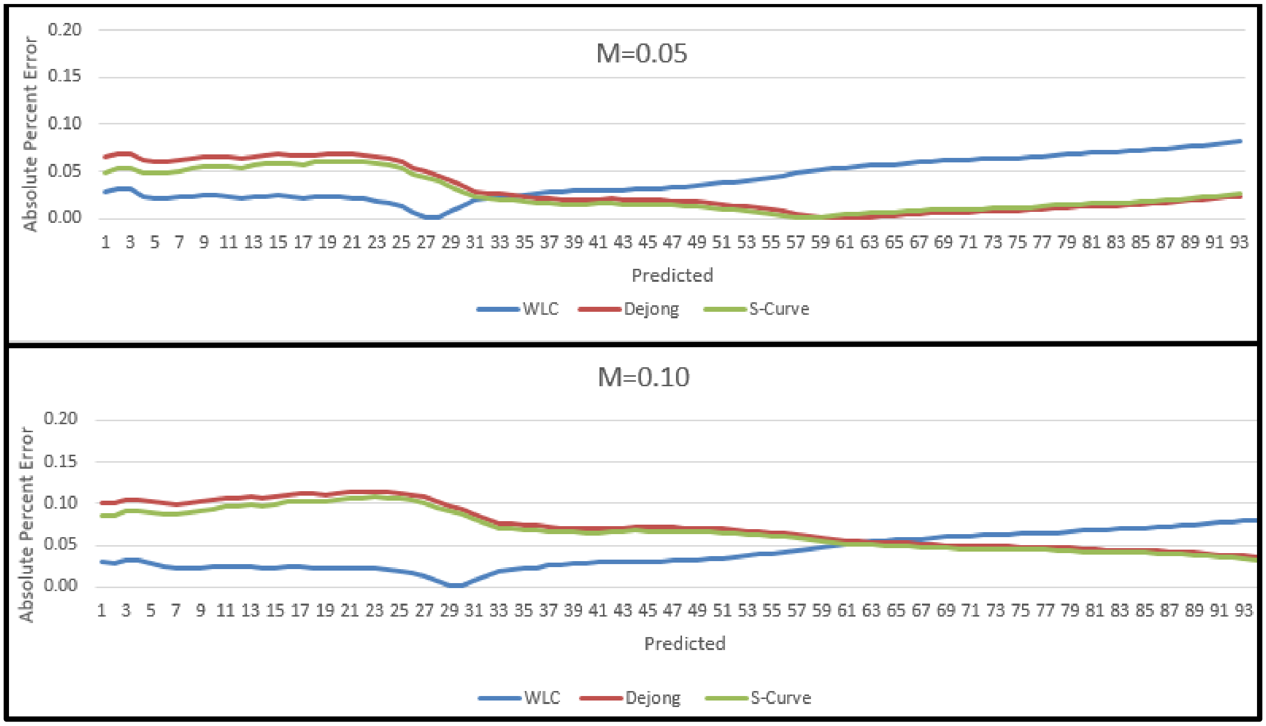


Figure 6 APE Trends

Of note is how the incompressibility factor plays a role in the estimates. It is evident that the Absolute percent errors swing in favor of the modern learning curve models when predicting hours using IA&CO. WLC provides less percent error initially, but WLC has more percent error in both cases for an M-Value of 0.05 and 0.10.

The next test for sensitivity was running the method again using all of the data on the Air Force Program. All of the data includes Test Support, Machine Shop. Program Support and Design. These work codes are not repetitive in nature like IA&CO. The initial parameters found that the cumav regression was the best fit for the data. The parameter analysis values are in

Table 8 below. Of note and as expected the theoretical first unit would be higher with more data points.

Table 8 Parameters for All Data Points Analysis

Theoretical First Unit (Hours)	41440
B	2
M	0.05
b	-0.2517
LCS	83.99%

The three learning curve models calculations all for a MAPE analysis with the parameters in place. The MAPE analysis found the values in Table 9 below. WLC has the lowest Mean Absolute Percent Error, followed by the S-Curve, and lastly DeJong's model.

Table 9 MAPE Values for All Data M = 0.05

WLC	DeJong	S-Curve
7.02%	6.59%	5.92%

The coefficient of variation calculations were performed as a cross check. The mean may be a good predictor of a better model. However, as a rule of thumb, if the coefficient of variation (CV) is greater than 15% the mean indicates a looser distribution. The CVs for the three models exceeded the 15% rule of thumb. The large CVs indicate that the mean is a poor estimator of a more accurate model. In order to compare models further statistical methods were the basis of evaluation.

The next step was to test for normality of the predicted values of each learning curve model. If normality passes, ANOVA will be the method. The assumption of normality was not met because of kurtosis and skewness. The variances are the next statistic to test. This will determine whether performing a T-Test with equal variances is a possibility. The rule of thumb for testing variance is to divide the largest sample standard deviation by the smallest. From the descriptive statistics, WLC has the smallest (0.034) and DeJong has the largest (0.059). If this value is greater than one, the assumption of equal variances will not be acceptable. The rule of thumb value equals 1.734 so equal variance cannot be assumed and a t-Test with unequal variances has to be performed to compare sample means. Table 10 shows the descriptive statistics for the data analysis including all points.

Table 10 Descriptive Statistics All Data

WLC		*Dejong*		*S-Curve*	
Mean	0.071	Mean	0.067	Mean	0.060
Standard Error	0.003	Standard Error	0.006	Standard Error	0.005
Median	0.073	Median	0.050	Median	0.043
Standard Deviation	0.034	Standard Deviation	0.059	Standard Deviation	0.052
Sample Variance	0.001	Sample Variance	0.003	Sample Variance	0.003
Kurtosis	-1.043	Kurtosis	-1.108	Kurtosis	-1.171
Skewness	-0.180	Skewness	0.613	Skewness	0.606
Range	0.127	Range	0.188	Range	0.162
Minimum	0.001	Minimum	0.000	Minimum	0.000
Maximum	0.127	Maximum	0.188	Maximum	0.163
Sum	6.719	Sum	6.378	Sum	5.728
Count	95.000	Count	95.000	Count	95.000
Confidence Level(95.0%)	0.0069	Confidence Level(95.0%)	0.01204	Confidence Level(95.0%)	0.0107

Table 11 below shows the results for the KW test for non-normal distributions. KW test predicts if values are significantly different. The null hypothesis was that the distributions of the APE values for the three models was the same based on a significance level of 0.05. The table below shows these results. Since the K value is less than the critical value, rejecting the null hypothesis that the APE values are same is not necessary. Table 10 also highlights that the p-value is 0.063 is greater than the 0.05 significance level. The p-value echoes failing to reject the null hypothesis. These values are close to being significant.

Table 11 KW Test All Data

K	5.534
Critical Value	5.991
P-Val	0.063

When plotting the Absolute Percent Errors a trend similar to the analysis using IA&CO was evident. WLC starts as a more accurate predictor of cost and then becomes less and less accurate. Whereas the DeJong and S-Curve models become increasingly more accurate predictors. Figure 7 below shows the results of APE.

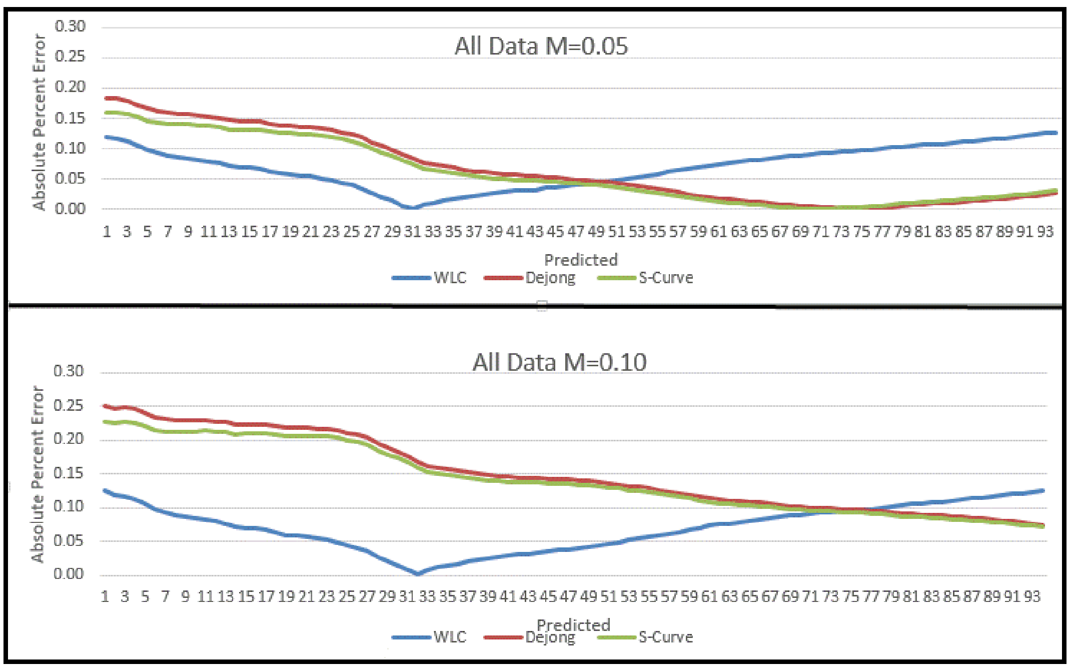


Figure 7 All Data APE Trends

If there are indications that one or more of the models are significantly different, it is necessary to determine which model captures the difference. The t-Test will not take place because there is no reason to believe any of the models were different. When changing the M Value to 0.10, the S-Curve and DeJong MAPE values increased more than double. WLC is a better predictor of with all of the data at an M value of 0.10. Table 12 highlights the MAPE findings.

Table 12 MAPE Values M = 0.10 All Data

WLC	DeJong	S-Curve
7.01%	15.04%	14.21%

Summary

Chapter IV provided results to the analytical tests performed as described in Chapter III. A description of the results for the Air Force Program, based on the assumption of low incompressibility values of 0.05 and 0.1, highlights the effects of learning. The results changed between these values and indicated that the S-Curve model may be a more accurate predictor at 0.05, but WLC is more accurate at 0.10. After 0.10 WLC becomes significantly less error prone for both IA&CO and the analysis with all of the work codes. WLC is a better predictor of cost when the incompressibility is 0.10 and higher. Analysis on all of the data points, not just IA&CO, showed no difference between the models at an M of 0.05. However increasing the M value results in WLC becoming an increasingly more accurate predictor of cost.

V. Conclusions and Recommendations

Chapter Overview

Chapter V will conclude the thesis and provide insight into whether or not there is reason to believe a more accurate learning curve model exists. The thesis was a comparative study on learning curve models in defense cost estimating. The study focused on the two conventional models (Wright's Cumav and Crawford's Unit Curves) used in defense cost estimating and two other contemporary models. DeJong's Learning and the S-Curve model were the contemporary models in the comparison. Chapter IV shows the results of the comparative study analysis which led to the conclusion and recommendations in this chapter. Chapter V highlights the conclusions of the research and the impacts of it. Specifically Chapter V will draw inferences on IA&CO data analysis on the Air Force Program. The significance of the results will provide a potential conclusion to the questions posed by this comparative study. Chapter V also addresses the recommendations for action within the Air Force and DOD. In Chapter V there is a section detailing the limitations of the research study. Lastly, Chapter V will highlight recommendations for future follow on research.

Conclusions of Research

The results of the comparative study were inconclusive. The answer to whether or not there is a more accurate learning curve model did not produce a significant result. The status quo, Wright's Learning Curve, was just as accurate of a predictor of learning when compared to contemporary learning curve models. The results using Integration, Assembly, and Checkout data provides awareness on automation and the effects that it plays in the learning curve models. In comparison to the analysis on all of the data, the results were, as theorized, more sensitive to

IA&CO. The conclusions for the hypotheses for the analysis are next. They hypotheses were one of the compared models will have MAPE (Mean Absolute Percent Error) significantly different from the others. One of the modern learning curve models will be significantly more accurate than Wright's model in predicting costs or hours. The S-Curve model with IA&CO will have a significantly lower MAPE than Wright's and DeJong's Learning Curve Models. Lastly, the incompressibility factor will have a significant influence on the accuracy of the DeJong and S-Curve Models

In regards to the first hypothesis, that one of the compared models will have MAPE (Mean Absolute Percent Error) significantly different from the others the following holds. At all incompressibility factors of 0.05, 0.10, and 0.15 the learning curve models were statistically different. However, when using all of the data there was no reason to believe there was a difference at 0.05. At values of 0.10 and 0.15 there was statistical difference between the models. The results of the first hypothesis are essential in the comparative study. The first hypothesis establishes the basis for testing the remaining hypotheses.

The second hypothesis, one of the modern learning curve models will be significantly more accurate than Wright's model in predicting costs or hours held when IA&CO analysis was held at an incompressibility factor of 0.05. WLC proved to be more accurate of a predictor for IA&CO at incompressibility factors of 0.10 and 0.15. When including all of the data points, there was not a statistical difference at an incompressibility of 0.05. However, the analysis held that WLC was more accurate for IA&CO at incompressibility factors of 0.10 and 0.15. Of interest, the analysis portrays that Wright's Learning Curve performs with more accuracy when the automation increases. Results did not support that that the modern learning curve models, DeJong and S-Curve are more accurate (lower percent error) than WLC.

The third hypothesis, the S-Curve model with IA&CO will have a significantly lower MAPE than Wright's and DeJong's Learning Curve Models did not hold. The reason behind this hypothesis was that the S-Curve accounts for more aspects of the learning process by including both prior unit experience and incompressibility. This hypothesis was not found to be significant at the chosen parameters (alpha of 0.05) for the analysis. However, the MAPE for the S-Curve was lower (more accurate and less error prone) than DeJong's Model. The MAPES of 2.64% and 3.00% respectively reflect the results. The DeJong and S-Curve Models were more accurate than WLC, but they did not show a statistical difference amongst themselves. Because of no statistical difference, the third hypothesis is inconclusive as to which model will predict with the lowest percent error.

Lastly, the hypothesis that incompressibility factor will have a significant influence on the accuracy of the DeJong and S-Curve Models held. Small changes in incompressibility did result in drastically different results. Changes in increments of 0.05 to 0.15 led to significantly different results. The figures of the Absolute Percent Errors in Chapter IV highlight the result of the comparison. With time and increasing trials, the S-Curve and DeJong models become increasingly less error prone. WLC is more accurate initially, however with time the contemporary models become less error prone. The result of this leads to a theory that the influence of machinery in this instance could be more sensitive to longer production cycles. It may support an expanded incompressibility range as well.

In summary of the hypotheses, results support the first hypothesis that there is a significant difference between the models. Results are inconclusive as to whether any models are significantly more accurate than Wright's model. Between an incompressibility of 0 to 0.1 DeJong and S-Curve models were more accurate (less error prone). Nevertheless, at an incompressibility of 0.10 and beyond Wright's model most accurate. The third hypothesis has results that are inconclusive as to which model is most accurate at an incompressibility of 0.05. Both DeJong and S-Curve models were more accurate than WLC, but no difference between the two. Finally yet importantly, incompressibility was highly influential as hypothesized. The results of the findings lead to questioning why for the program chosen incompressibility would become increasingly more error prone when more automation is present. In addition, the findings put into question how the DOD can draw a conclusion about the application of contemporary learning curve models in acquisitions and specifically cost estimation.

In regards to the questions posed in the last paragraph, the first question as to why the analysis for program chosen, would incompressibility became increasing more error prone when more automation is present may be due to the data. The Absolute Percent Error figures highlight that WLC is accurate initially and eventually becomes increasingly less accurate. The opposite that the S-Curve and DeJong Models are not as accurate initially, but become increasingly accurate over trials. The MAPE analysis averages all of the errors. If the data set included more units, results may trend towards results in favor of the contemporary models. That theoretical question answer is based on the visual trend from the APE figures. Of interest when the incompressibility factor is 0.10, the models portray that 90% of learning is obtainable. Because the data set was small, changes in incompressibility may not be as evident to the significance of the comparative study. The findings also put into question how the DOD can draw a conclusion about the application of contemporary learning curve models in acquisitions and specifically cost estimation. If the production cycle is long, and many trials will be realized, there is potential the contemporary models may capture a more accurate picture of learning. Aircraft production may provide starkly different results from a missile production run where more units are produced over time. The results support there is potential for a more accurate model. However, it may not be in the realm of aircraft production. Aircraft production may include some automation. It is not implausible that aircraft production is 95% manual and supports an incompressibility factor of 0.05. The contemporary models may support a more automated process such as a production line much like the automobile industry. Prior studies and subject matter expert opinion support that aircraft production is manual. However, there is the belief that more automation is going to be present in the future.

Significance of Research

Results from the analysis show that there is reason to believe Wright's Learning Curve may not be the best method for estimating costs. By extrapolating from actuals, the method for Wright's model may not incorporate enough of the variability of learning. The results provide evidence that Wright's Model is accurate initially, but with attempts at learning (trials) the amount of error increases. The comparative analysis on learning curve models provides a standalone analysis of program actuals. The fact that the analysis includes proprietary data in labor hours is distinctive. The conclusions from that study are there is potential for a more accurate cost-estimating model and that the conventional learning curve models become increasingly less error prone over trials. The DeJong and S-Curve models show promise as a way to improve DOD cost estimating.

The results of the research do not support all of the hypotheses. Results did confirm that the incompressibility factor was highly influential for both the S-Curve and DeJong models. The results of the comparison changed drastically with a small change in the incompressibility factor. The DeJong and S-Curve models were both more accurate than WLC, but there was no difference between the two. This finding makes it challenging to simplify the results given the uncertainty of incompressibility. Findings do provide a representation that further investigation on learning curve methodology and potential change is near as automation becomes increasingly more relevant in every aspect of life. The influence of machinery in a longer production cycle is a valid assumption for the future. The influence of automation in this comparative study was evident by the absolute percent error graphs.

Recommendations for Future Research

The comparative study tried to find answers in regards to the learning curve methodology within Department of Defense acquisitions. The effects of learning have shown a flattening trend at the tail end of production. This investigation sought to investigate whether the current methodology was significantly less accurate than contemporary learning curve models. Wright's Learning Curve Model is the method that the DOD uses to extrapolate from actuals and estimate. The contemporary learning curve models incorporate a percentage of the process that has automation. The study found that the Dejong Model and S-Curve Model may be more accurate is automation is low for aircraft production. The research did not find support as to which model is most accurate or find evidence that there is a given incompressibility factor. Given the findings from this research, recommendations for future research include expanding findings to determine which model is most accurate, incorporating a different aircraft platform, incorporating the Rate Effect into the models, and testing the different models for separate production runs or lots to see which is most accurate.

Future research should incorporate more platforms and look at a more automated production such as missiles where more automation may be more present. This research focused on one platform. This platform was the only program with data broken out into specific work codes. The recommendation for future research is to look at additional platforms with a longer production. Aircraft with low total units in production make it challenging to see the effects of automation. The Absolute Percent Error graphs highlight that with time the contemporary learning curve models become more accurate. With this analysis, there is impending possibility of a more accurate model.

Analysis of additional models, such as the Rate Effect may provide strong results. The potential to add another variable to the Rate Effect may be a unique and exclusive study. Adding the incompressibility to the Rate Effect and performing the same analysis would be an analysis that sheds light on automation, production runs, and learning within the DOD or Air Force acquisitions realm. This idea for future research would mean making of modifying a learning equation similar to what RAND did with the Rate Theory. Rate Theory essentially attempts to include production rate as part of the ULC model. Consider it the ULC with a rate adjustment. Of note, the rate term (variable) is infrequently found to be statistically significant even though it popular with Program Offices. The nest step in the future could be the ULC, with rate adjustment, and incompressibility.

Given that trend for Wright's Learning Curve is initially to have less error and the contemporary models becoming increasingly less error prone, finding where the switching point for the models may provide profound insight into the production. For instance if WLC is only accurate for predicting the initial units and the contemporary models are more accurate for production that exceeds where WLC is accurate is a recommendation for future research. Finding a defendable break point or threshold to give the cost estimator such as use WLC for the initial 25% of production and a contemporary for the remainder of production. Fundamentally, if a heuristic for a defendable break point were found, it would be of value to defense cost estimating and estimators using extrapolation from actuals and learning curve analysis. Finding a defendable range as to where this break occurs is another one of the recommendations for future research.

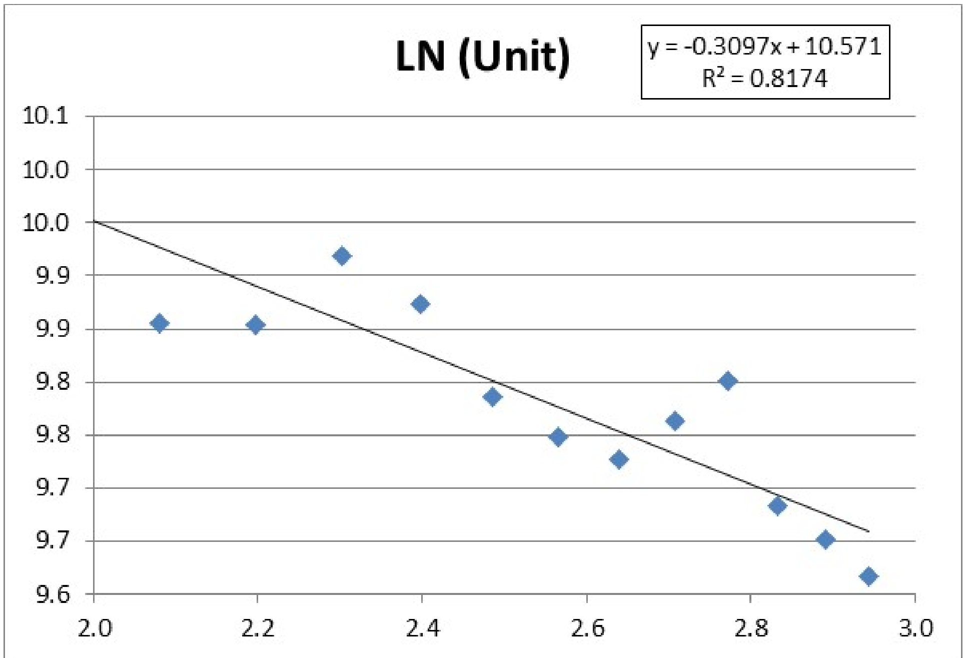
Assumptions and Limitations

The assumptions and limitations to this comparative research study are a part of any analysis. Assumptions for the research are necessary in order to test the hypotheses, which is a limitation to the validity of the results. The first limitation to the study is that only one aircraft was available with the specific work codes, Integration, Assembly, and Checkout. The data available to analyze was small. The results cannot be generalized to other aircraft acquisitions or DOD acquisitions since the sample size was a small percentage of production. Analysis uses Wright's CUMAV theory, which can artificially smooth data into the appearance of better fit. Crawford's Unit theory may provide different results. Use of just the one aircraft production makes it difficult to assume results would hold across multiple platforms. Essentially different results may be found for other airframes. In general, a limitation was that most available data was inefficient because it was not broken out into the specific work codes. Results are very dependent on the incompressibility factor. The assumption that incompressibility is low was a basis of expert opinion, a 1993 Bureau of Labor and Statistics report, and prior research. As a whole, the analysis assumes no major change in product design, production processes, workforce composition, and interval between units.

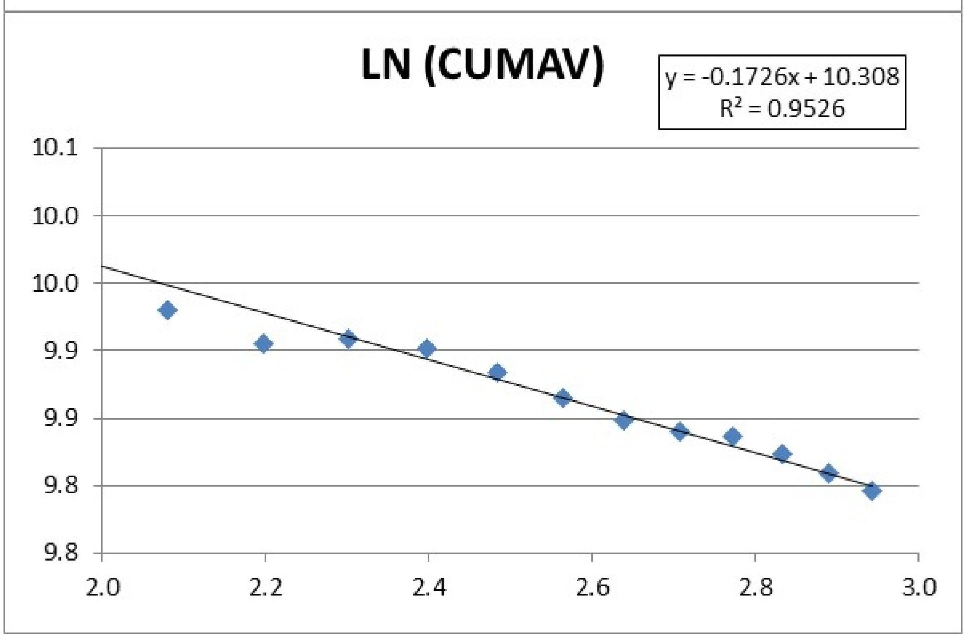
Summary

In summary, the goal of this research was to find out if a more accurate learning curve model exists. Help from AFLCMC/FCZ) to attain data on aircraft in order to test DeJong and S-Curve learning models against Wright's model was the basis of finding a way to increase the accuracy of the current learning curve methodology within DOD and Air Force acquisition. The results were inconclusive as to which model is most accurate, but indicate that DeJong and S-Curve models may be more accurate than WLC if incompressibility is assumed low. If the incompressibility is 0.10 or greater, Wright's Learning Curve is most accurate having a Mean Absolute Percent Error lower than the contemporary Dejong and S-Curve Models. However, the trend of the Absolute Percent Errors shows divergent tendencies for both the conventional status quo and the contemporary models. The results did not show which contemporary model is most accurate. This thesis paves the way for future research on learning curve methodology and the significance of Integration, Assembly, and Checkout Data. IA&CO was less error prone than the entire learning curve data. Combining models and broadening the scope of the analysis, at a minimum, provides a groundwork for the learning curve methodology. AFLCMC/FZC backed research into flattening effect at tail end of learning curves. Wright's original learning curve theory (CUMAV) was formulated in 1936 and Crawford's theory (Unit) was adopted a decade after. The learning curve models accepted by DOD are over 70 years old and both models are in use within the DOD. This study sought to determine if the current DOD methodology is outdated and if a contemporary model is more accurate.

Appendix A



LN (Unit)
y = -0.3097x + 10.571
R² = 0.8174
10.1
10.0
10.0
9.9
9.9
9.8
9.8
9.7
9.7
9.6
2.0
2.2
2.4
2.6
2.8
3.0

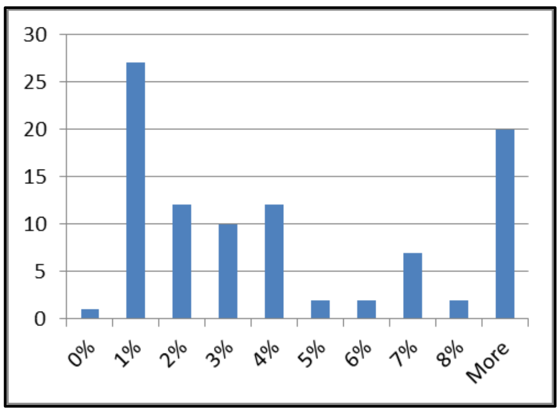


LN (CUMAV)
y = -0.1726x + 10.308
R² = 0.9526
10.1
10.0
10.0
9.9
9.9
9.8
9.8
2.0
2.2
2.4
2.6
2.8
3.0

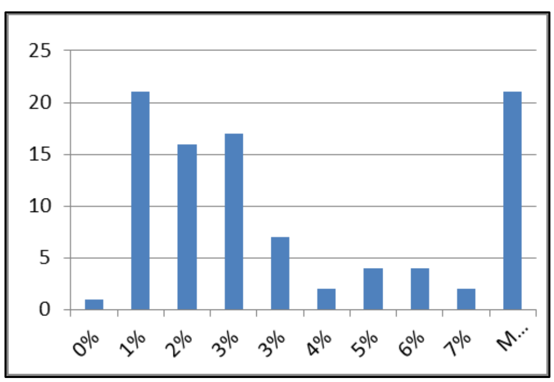
Appendix B

Dejong								
Unit (x)	Actual	Predicted	Error	Abs Error	Abs PE	ln(x)	ln(Actual)	ln(predict)
20	17336.8	18476.1	-1139.24	1139.24	0.07	3.00	9.76	9.82
21	17221.8	18333.7	-1111.86	1111.86	0.06	3.04	9.75	9.82
22	17041.3	18199.1	-1157.72	1157.72	0.07	3.09	9.74	9.81
23	16916.1	18071.4	-1155.31	1155.31	0.07	3.14	9.74	9.80
24	16854.6	17950.1	-1095.52	1095.52	0.06	3.18	9.73	9.80
25	16797.5	17834.6	-1037.13	1037.13	0.06	3.22	9.73	9.79
26	16710.3	17724.4	-1014.12	1014.12	0.06	3.26	9.72	9.78

Appendix C



DeJong 1



S-Curve 1

Appendix D

Conclusion: We do a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. T Stat (3.32 is > 1.97) so we reject the null hypothesis. The means are not the same.

t-Test: Two-Sample Assuming Unequal Variances

	MLC	Dejong
Mean	0.041	0.030
Variance	0.000	0.001
Observations	95	95
Hypothesized Mean Difference	0	
df	185	
t Stat	3.329	
t Critical two-tail	1.973	

Conclusion: We do a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. T Stat (4.91 is > 1.97) so we reject the null hypothesis. The means are not the same.

t-Test: Two-Sample Assuming Unequal Variances

	MLC	S-Curve
Mean	0.041	0.026
Variance	0.000	0.000
Observations	95	95
Hypothesized Mean Difference	0	
df	187	
t Stat	4.912	
t Critical two-tail	1.973	

Conclusion: We do a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. T Stat (1.117 is not > 1.97) so we fail reject the null hypothesis. There is no reason to believe the S-Curve is a more accurate predictor over the DeJong model

t-Test: Two-Sample Assuming Unequal Variances

	Dejong	S-Curve
Mean	0.030	0.026
Variance	0.001	0.000
Observations	95	95
Hypothesized Mean Difference	0	
df	180	
t Stat	1.118	
t Critical two-tail	1.973	

Bibliography

Aeppel, T. (2012). It's Man vs. Machine and Man Is Losing - Real Time Economics - WSJ. Retrieved February 28, 2016, from http://blogs.wsj.com/economics/2011/09/28/its-man-vs-machine-and-man-is-losing/

Alchian, A. (1963). Reliability of Progress Curves in Airframe Production on JSTOR. Retrieved October 28, 2015, from http://www.jstor.org/stable/1909166?origin=crossref&seq=1#page_scan_tab_contents

All, A. (2013). 2002-2013 ICEAA. All rights reserved. 1.

Asher, H. (1956). Cost-quantity relationships in the airframe industrie. *Project Rand*. Retrieved from http://www.rand.org/content/dam/rand/pubs/reports/2007/R291.pdf

Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Transactions on Engineering Management*, *39*(2), 176–188. http://doi.org/10.1109/17.141275

Badiru, A. B. (2012). Half-Life Learning Curves in the Defense Acquisition Life Cycle. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA582714

Benkard, C. L. (1999). *Learning and Forgetting the Dynamics of Aircraft Production*. *National Bureau of Economic Research*.

Dameron, Megan E., B. L. C. (n.d.). Module06-Data-Analysis. ICEAA.

Dewey, R. (n.d.). The Learning Curve | in Chapter 07: Cognition | from Psychology: An Introduction by Russ Dewey. Retrieved March 2, 2016, from http://www.psywww.com/intropsych/ch07_cognition/learning_curve.html

Ellis, D. S. (n.d.). Module04-Data-Collection. ICEAA.

Everest, J. D. (1988). *Measuring Losses of Learning Due To Breaks In Production*.

Filliben, J. J. (2012). e-Handbook of Statistical Methods. http://doi.org/NIST/SEMATECH

Frost, J. (2013). Multiple Regression Analysis: Use Adjusted R-Squared and Predicted R-Squared to Include the Correct Number of Variables. Retrieved March 1, 2016, from http://blog.minitab.com/blog/adventures-in-statistics/multiple-regession-analysis-use-adjusted-r-squared-and-predicted-r-squared-to-include-the-correct-number-of-variables

Godin, S., & Warner, A. (2010). How To Produce Like A Linchpin By Understanding Your Lizard Brain – with Seth Godin. *Mixergy*. Retrieved from http://mixergy.com/linchpin-lizard-seth-godin/

Hax, A. C., & Majluf, N. S. (1982). Competitive Cost Dynamics: The Experience Curve. *Interfaces*, *12*(5), 50–61. http://doi.org/10.1287/inte.12.5.50

Higgins, J., & Cooperstein, G. (2010). Managing an organization ' s biggest cost : The workforce. *Human Capital Management Institute*, 1–9.

Kendall, H. F. (2014). Better Buying Power 3.0, (September).

Labor Categories_2.pdf. (n.d.).

Luce, D. De. (2014). Why Is The US Military Spending So Much Money On The F-35 Fighter Jet? Retrieved October 28, 2015, from http://www.businessinsider.com/f-35-cost-2014-2

Madigan, K. (2011). Man vs. Machine: Behind the Jobless Recovery - WSJ. Retrieved February 28, 2016, from http://www.wsj.com/articles/SB10001424052970204468004577164710231081398

Making the future | The Economist. (n.d.). Retrieved December 17, 2015, from http://www.economist.com/node/21552897

Malashevitz, S., & Williams, B. (2010). Handout for three day Learning Curve Workshop. *DAU*

Midwest, 1–16.

Moore, J. R. (2015). A COMPARATIVE STUDY OF LEARNING CURVE MODELS IN DEFENSE, (March).

Mueller, J. (2011). Should Cost Management: Why? How?, (October), 14–18.

Multicollinearity in regression. (n.d.). Retrieved from http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/regression/supporting-topics/model-assumptions/multicollinearity-in-regression/

Residual Analysis in Regression. (n.d.). Retrieved December 16, 2015, from http://stattrek.com/regression/residual-analysis.aspx?Tutorial=AP

Ritter, F. (2001). The learning curve. *International Encyclopedia of the Social*, 8602–8605. http://doi.org/10.1067/mhn.2002.121517

Weisstein, E. W. (n.d.). e. Wolfram Research, Inc. Retrieved from http://mathworld.wolfram.com/e.html

Wikipedia, F. (1936). The Experience Curve. *Group*, 1–4. http://doi.org/10.1016/0305-0483(83)90010-5

Zaiontz, C. (2015). Real Statistics Using Excel. Retrieved from http://www.real-statistics.com/

Vita

Captain Brandon J. Johnson was born in Duluth, Minnesota in October of 1985 and graduated from Duluth Marshall High School in 2004. He entered the United States Air Force in 2006 and attended the United States Air Force Academy. Upon completion of a Bachelor's of Science in Management at the Air Force Academy, he transitioned into the Financial Management Career field. Captain Johnson was stationed at the Space and Missiles System Center in California for four years. While there, he contributed to financial management and analytical services for a $20B Space Based Infrared System satellite program. During his time in California, Captain Johnson deployed as an Executive Officer in support of 544 personnel from eight nations at 23 sites. In September of 2014, Captain Johnson became a Cost Analysis Graduate Student in the Department of Engineering and Management, Air Force Institute of Technology, OH. Upon graduation, he will be assigned to Wright Patterson Air Force Base.

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704–0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202–4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD–MM–YYYY)	2. REPORT TYPE	3. DATES COVERED (*From — To*)
24-03-2016	Master's Thesis	Sept 2014 – March 2016

4. TITLE AND SUBTITLE
A Comparative Study of Learning Curve Models and Factors in Defense Cost Estimating Based on Program Integration, Assembly, and Checkout

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Johnson, Brandon J, Capt

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-MS-16-M-162

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AGENCY: AFLCMC/FCZ
ADDRESS: 1865 4th St. Bldg. 14 WPAFB, OH 45433
PHONE: 986-5478 EMAIL: seibelmjs@gmail.com
ATTN: Mr. Michael Seibel

10. SPONSOR/MONITOR'S ACRONYM(S)
AFLCMC/FCZ

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMTED.

13. SUPPLEMENTARY NOTES
This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT
The purpose of this research was to investigate the flattening effect at tail end of learning curves by identifying a more accurate learning curve model. The learning curve models accepted by DOD are Wright's original learning curve theory and Crawford's Unit Theory. The models were formulated in 1936 and 1944 respectively. This analysis compares the conventional models to contemporary learning curve models in order to determine if the current DOD methodology is outdated. The results are inconclusive as to if there is a more accurate model. The contemporary models are the DeJong and S-Curve and they both include an incompressibility factor, which is the percentage of the process that includes automation. Including models that incorporate automation was important as technology and machinery plays a larger role in production. Wright's model appears to be most accurate unless incompressibility is very low. A trend for all models appeared. The trend is Wright's curve was accurate early in production and the contemporary models were more accurate later in production. Future research should have an objective of finding a heuristic for when the models are most accurate or comparative studies including more models.

15. SUBJECT TERMS
Learning curves, automation, cost estimating, incompressibility

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. John J Elshaw, AFIT/ENV
U	U	U	UU	64	19b. TELEPHONE NUMBER (Include Area Code) (937)255-3636, ext. 4650 John.elshaw@afit.edu

Standard Form 298 (Rev. 8–98) ***Prescribed by ANSI Std. Z39.18***